Application of Neural Networks for Foreign Exchange Rates Forecasting with Noise Reduction

Wei Huang^{1,3}, Kin Keung Lai^{2,3}, and Shouyang Wang⁴

¹ School of Management, Huazhong University of Science and Technology
WuHan, 430074, China
whuang@amss.ac.cn

² College of Business Administration, Hunan University
Changsha 410082, China

³ Department of Management Sciences, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong
{weihuang, mskklai}@cityu.edu.hk

⁴ Institute of Systems Science, Academy of Mathematics and Systems Sciences
Chinese Academy of Sciences, Beijing, 100080, China
swang@amss.ac.cn

Abstract. Predictive models are generally fitted directly from the original noisy data. It is well known that noise can seriously limit the prediction performance on time series. In this study, we apply the nonlinear noise reduction methods to the problem of foreign exchange rates forecasting with neural networks (NNs). The experiment results show that the nonlinear noise reduction methods can improve the prediction performance of NNs. Based on the modified Diebold-Mariano test, the improvement is not statistically significant in most cases. We may need more effective nonlinear noise reduction methods to improve prediction performance further. On the other hand, it indicates that NNs are particularly well appropriate to find underlying relationship in the environment characterized by complex, noisy, irrelevant or partial information. We also find that the nonlinear noise reduction methods work more effectively when the foreign exchange rates are more volatile.

1 Introduction

Foreign exchange rates exhibit high volatility, complexity and noise that result from the elusive market mechanism generating daily observations [1]. It is certainly very challenging to predict foreign exchange rates. Neural networks (NNs) have been widely used as a promising alternative approach for a forecasting task because of several distinguishing features [2]. Several design factors significantly affect the prediction performance of neural networks [3]. Generally, NNs learn and generalize directly from noisy data with the faith on ability to extract the underlying deterministic dynamics from the noisy data. However, it is well known that the model's generalization performance will be poor unless we prevent the model from over-learning.

Given that most financial time series contain dynamic noise, it is necessary to reduce noise in the data with nonlinear methods before fitting the prediction models. However, not much work has been done [4]. In this study, we employ two nonlinear noise reduction methods to alleviate these problems. The remainder of this paper is organized as follows. In Section 2, we give a brief introduction to the two nonlinear noise reduction methods. Section 3 presents the experiment design. Section 4 discusses the empirical experiment result. Finally, Section 5 offers some concluding remarks.

2 Nonlinear Noise Reduction

Conventional linear filtering in the time or Fourier domain can be very powerful as long as nonlinear structures in the data are unimportant. However, nonlinearity can not be fully characterized by second order statistics like the power spectrum. Nonlinear noise reduction does not rely on frequency information in order to define the distinction between signal and noise. Instead, structure in the reconstructed phase space will be exploited. Here we want to concentrate on two nonlinear noise reduction methods that represent the geometric structure in phase space by local approximation. The former does so to constant order, while the latter uses local linear subspaces plus curvature corrections. Interested reader can refer to the articles of review character [5-7].

2.1 Simple Nonlinear Noise Reduction (SNL)

SNL replaces the central coordinate of each embedding vector by the local average of this coordinate:

$$z_n = \frac{\sum_{x_k \in U_n^\delta} x_k}{\|U_n^\delta\|} \quad (1)$$

where U_n^δ is neighborhood formed in phase space containing all points like x_k , such that $|x_k - x_n| < \delta$. This noise reduction method amounts to a locally constant approximation of the dynamics and is based on the assumption that the dynamics is continuous.

2.2 Locally Projective Nonlinear Noise Reduction (LP)

LP rests on hypotheses that the measured data is composed of the output of a low-dimensional dynamical system and of random or high-dimensional noise. This means that in an arbitrarily high-dimensional embedding space the deterministic part of the data would lie on a low-dimensional manifold, while the effect of the noise is to spread the data off this manifold. If we suppose that the amplitude of the noise is

sufficiently small, we can expect to find the data distributed closely around this manifold. The idea of the projective nonlinear noise reduction scheme is to identify the manifold and to project the data onto it.

Suppose the dynamical system forms a q -dimensional manifold $\mathbf{M}$ containing the trajectory. According to the embedding theorems, there exists a one-to-one image of the attractor in the embedding space, if the embedding dimension is sufficiently high. Thus, if the measured time series were not corrupted with noise, all the embedding vectors S_n would lie inside another manifold $\tilde{\mathbf{M}}$ in the embedding space. Due to the noise this condition is no longer fulfilled. The idea of the locally projective noise reduction scheme is that for each S_n there exists a correction Θ_n , with $\|\Theta_n\|$ small, in such a way that $S_n - \Theta_n \in \tilde{\mathbf{M}}$ and that Θ_n is orthogonal on $\tilde{\mathbf{M}}$. Of course a projection to the manifold can only be a reasonable concept if the vectors are embedded in spaces which are higher dimensional than the manifold $\tilde{\mathbf{M}}$. Thus we have to over-embed in m -dimensional spaces with $m > q$.

The notion of orthogonality depends on the metric used. Intuitively one would think of using the Euclidean metric. But this is not necessarily the best choice. The reason is that we are working with delay vectors which contain temporal information. Thus even if the middle parts of two delay vectors are close, the late parts could be far away from each other due to the influence of the positive Lyapunov exponents, while the first parts could diverge due the negative ones. Hence it is usually desirable to correct only the center part of delay vectors and leave the outer parts mostly unchanged, since their divergence is not only a consequence of the noise, but also of the dynamics itself. It turns out that for most applications it is sufficient to fix just the first and the last component of the delay vectors and correct the rest. This can be expressed in terms of a metric tensor P which we define to be

$$P_{ij} = \begin{cases} 1 & 1 < i = j < m \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where m is the dimension of the over-embedded delay vectors.

Thus we have to solve the minimization problem

$$\begin{aligned} \min \quad & \sum_i \Theta_i P^{-1} \Theta_i \\ \text{s.t.} \quad & a_n^i (S_n - \Theta_n) + b_n^i = 0, \quad \text{for } i = q+1, \dots, m \\ & a_n^i P a_n^j = \delta_{ij} \end{aligned} \quad (3)$$

where a_n^i are the normal vectors of $\tilde{\mathbf{M}}$ at the point $S_n - \Theta_n$.

3 Experiments Design

3.1 Neural Network Models

In this study, we employ one of the widely used neural networks models, the three-layers back-propagation neural network (BPNN), for foreign exchange rates forecasting. The activation function used for all hidden nodes is the logistic function, while the linear function is employed in the output node. The number of input nodes is a very important factor in neural network analysis of a time series since it corresponds to the number of past lagged observations related to future values. To avoid introducing a bias in results, we choose the number of input nodes as 3, 5, 7 and 9, respectively. Because neural networks with one input node are too simple to capture the complex relationships between input and output, and it is rarely seen in the literatures that the number of input nodes is more than nine. Generally speaking, too many nodes in the hidden layer produce a network that memorizes the input data and lacks the ability to generalize. Another consideration is that as the number of hidden nodes in a network is increased, the number of variables and terms are also increased. If the network has more degrees of freedom (the number of connection weights), more training samples are needed to constrain the neural network. It has been shown that the in-sample fit and the out-of-sample forecasting ability of NNs are not very sensitive to the number of hidden nodes [8]. Parsimony is a principle for designing NNs. Therefore, we use four hidden nodes in this study.

3.2 Random Walk Model

The weak form of efficient market theory describes that prices always fully reflect the available information, that is, a price is determined by the previous value in the time series because all the relevant information is summarized in that previous value. An extension of this theory is the random walk (RW) model. The RW model assumes that not only all historic information is summarized in the current value, but also that increments—positive or negative—are uncorrelated (random), and balanced, that is, with an expected value equal to zero. In other words, in the long run there are as many positive as negative fluctuations making long term predictions other than the trend impossible. The random walk model uses the actual value of current period to predict the future value of next period as follows:

$$\hat{y}_{t+1} = y_t \quad (4)$$

where $\hat{y}_{t+1}$ is the predicted value of the next period; y_t is the actual values of current period.

3.3 Performance Measure

We employ root of mean squared error (RMSE) to evaluate the prediction performance of neural networks as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_t (y_t - \hat{y}_t)^2}{T}} \tag{5}$$

where y_t is the actual value; $\hat{y}_t$ is the predicted value; T is the number of the predictions.

3.4 Data Preparation

From Pacific Exchange Rate Service provided by Professor Werner Antweiler, University of British Columbia, Canada, we obtain 3291 daily observations of U.S. dollar against the British Pound (GBP) and Japanese Yen (JPY) covering the period from Jan 1990 to Dec, 2002. We take the natural logarithmic transformation to stabilize the time series. First, we produce the testing sets for each neural network models by selecting 60 patterns of the latest periods from the three datasets, respectively. Then, we produce the appropriate training sets for each neural networks model from the corresponding left data by using the method in [9].

4 Experiments Results

Table 1 show the prediction performances of the random walk model, which are used as benchmarks of prediction performance of foreign exchange rates. In Table 2 and 3, the RMSE of noisy data is largest, and RMSE of LP is least in the same row. It indicates that noise reduction methods actually improve the prediction performance of NNs. Further, LP is more effective than SNL in reducing noise of exchange rates.

We also apply the modified Diebold-Mariano test [10] to examine whether the two noise reduction methods can improve NNs financial forecasting significantly. From the test statistic values shown in Table 4 and 5, only in the prediction of JPY, the NNs with 9 input nodes using data filtered by LP outperform those using noisy data significantly at 20% level (the rejection of equality of prediction mean squared errors is based on critical value of student's t-distribution with 69 degrees of freedom, namely 1.294 at 20% level). Perhaps there is still noise in exchange rates time series after using the noise reduction methods. We look forward more effective noise reduction methods in the future. On the other hand, it also implies that NNs are useful to extract the underlying deterministic dynamics from noisy data at present stage. In addition, the test statistic value of SNL for GBP is less than that of SNL for JPY in the same row. Such a pattern is observed between LP for GBP and LP for JPY. JPY is more volatile than GBP, that is, there are more noises in JPY than in GBP to be filtered. So the improvement for JPY after noise reduction is more significant than GBP.

Table 1. The prediction performance of the random walk models

RMSE of GBP	RMSE of JPY
0.005471	0.007508

Table 2. The prediction performance of NNs using data with noise, filtered by SNL and filtered by LP, respectively (GBP)

#Input nodes	Noisy data	Data filtered by SNL	Data filtered by LP
3	0.005471	0.005465	0.005461
5	0.004491	0.004473	0.004457
7	0.004496	0.004494	0.004467
9	0.0054671	0.005464	0.00541

Table 3. The prediction performance of NNs using data with noise, filtered by SNL and filtered by LP, respectively (JPY)

#Input nodes	Noisy data	Data filtered by SNL	Data filtered by LP
3	0.007438	0.007018	0.006719
5	0.006348	0.00616	0.005989
7	0.006358	0.006279	0.006067
9	0.007293	0.006811	0.006399

Table 4. The test statistic value of equality of prediction errors between noisy and filtered data (GBP)

#Input nodes	Data filtered by SNL	Data filtered by LP
3	0.012	0.021
5	0.043	0.093
7	0.005	0.087
9	0.007	0.129

Table 5. The test statistic value of equality of prediction errors between noisy and filtered data (JPY)

#Input nodes	Data filtered by SNL	Data filtered by LP
3	0.904	1.122
5	0.391	0.733
7	0.162	0.509
9	0.879	1.605

5 Conclusions

In the paper, we apply the two nonlinear noise reduction methods, namely SNL and LP, to the foreign exchange rates forecasting with neural networks. The experiment results show that SNL and LP can improve the prediction performances of NNs. The improvement is not statistically significant based on the modified Diebold-Mariano test at 20% level. Especially, LP performs better than SNL in reducing noise of foreign exchange rates. The noise reduction methods work more effectively on JPY than

on GBP. In the future work, we will employ more effective nonlinear noise reduction methods to improve prediction performance further. On the other hand, it indicates that NNs are particularly well appropriate to find underlying relationship in an environment characterized by complex, noisy, irrelevant or partial information. In the most cases, NNs outperform the random walk model.

Acknowledgements

The work described in this paper was supported by Strategic Research Grant of City University of Hong Kong (SRG No.7001806) and the Key Research Institute of Humanities and Social Sciences in Hubei Province-Research Center of Modern Information Management.

References

1. Theodossiou, P.: The stochastic properties of major Canadian exchange rates. *The Financial Review*, 29(1994) 193-221
2. Zhang, G., Patuwo, B.E. and Hu, M.Y.: Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14(1998) 35-62
3. Huang, W., Lai, K.K., Nakamori, Y. and Wang, S.Y.: Forecasting foreign exchange rates with artificial neural networks: a review. *International Journal of Information Technology & Decision Making*, 3(2004) 145-165
4. Soofi, A., Cao, L.: Nonlinear forecasting of noisy financial data. In *Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Soofi, A. and Cao, L. (Eds), Boston: Kluwer Academic Publishers, (2002) 455-465
5. Davies, M.E.: Noise reduction schemes for chaotic time series. *Physica D*, 79(1994) 174-192
6. Kostelich, E.J. and Schreiber, T.: Noise reduction in chaotic time series data: A survey of common methods. *Physical Review E*, 48(1993) 1752-1800
7. Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C. and Schreiber, T.: On noise reduction methods for chaotic data, *CHAOS*. 3(1993) 127
8. Zhang, G. and Hu, M.Y.: Neural network forecasting of the British Pound/US Dollar exchange rate. *Journal of Management Science*, 26(1998) 495-506
9. Huang, W., Nakamori, Y., Wang, S.Y. and Zhang, H.: Select the size of training set for financial forecasting with neural networks. *Lecture Notes in Computer Science*, Vol. 3497, Springer-Verlag Berlin Heidelberg (2005) 879-884
10. Harvey, D., Leybourne, S. and P. Newbold: Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting* 13(1997) 281-91