

Forecasting the U.S. Stock Market via Levenberg-Marquardt and Haken Artificial Neural Networks Using ICA&PCA Pre-Processing Techniques

Golovachev Sergey

National Research University, Higher School of Economics, Moscow
Department of World Economics and International Affairs

Abstract. Artificial neural networks (ANN) is an approach to solving different tasks. In this paper we forecast U.S. stock market movements using two types of artificial neural networks: a network based on the Levenberg-Marquardt learning mechanism and a synergetic network which was described by German scientist Herman Haken. The Levenberg-Marquardt ANN is widely used for forecasting financial markets, while the Haken ANN is mostly known for the tasks of image recognition. In this paper we apply the Haken ANN for the prediction of the stock market movements. Furthermore, we introduce a novation concerning pre-processing of the input data in order to enhance the predicting power of the abovementioned networks. For this purpose we use Independent Component Analysis (ICA) and Principal Component Analysis (PCA). We also suggest using ANNs to reveal the “mean reversion” phenomenon in the stock returns. The results of the forecasting are compared with the forecasts of the simple auto-regression model and market index dynamics.

Keywords: artificial neural network, back-propagation, independent component analysis, principal component analysis, forecast.

1 The Levenberg-Marquardt Network

Artificial neural networks are a modern approach to various problem-solving tasks. For example, they are used for image recognition and in different biophysics researches. One of the possible applications of ANNs is forecasting and simulation of financial markets. The idea is the following: a researcher tries to construct such an ANN, so that it can successfully imitate the decision-making process of the “average” stock market participant. This hypothesis results from the fact that ANNs, in turn, try to imitate the design of the biological neural networks, in particular the ones which exist in human brain.

A market participant is an investor whose individual actions have no influence on the price fluctuations, for example a trader operating with insignificant sums of money. Moreover, we argue that the market participant makes his decisions solely on the analysis of the previous dynamics of the stock – thus we assume

endogenous price-making mechanism. Furthermore, we set the homogeneity of the investors so that they all have the same decision-making algorithms (that is why we call them “average”).

While designing the Levenberg-Marquardt ANN it is essential to set some of the key parameters of the network. Firstly, we must set the architecture of the network (number of layers, number of neurons in each, including number of input and output neurons). In our research we use simple three-layer ANN with 2 input neurons, 2 neurons in the hidden layer and 1 output neuron. The results show that such architecture is quite effective while it does not lead to lengthy computational procedure. Secondly, we determine the activation function in the hidden layer which performs a non-linear transformation of the input data. We use a standard logistic function with the range of values [0;1].

The key feature of the Levenberg-Marquardt ANN is using of back-propagation of the errors of the previous iterations as a learning mechanism. The idea of back-propagation rests on the attempt of communicating the error of the network (of the output neuron, in particular) to all other neurons of the network. As a result, after a number of iterations the network optimizes the weights with which neurons in different layers are connected, and the minimum of error is reached. Propagation of the error through the network also requires usage of the Jacobian matrix which contains first derivatives of the elements of the hidden and input layers.

The computational mechanism is as follows (1):

$$w_{new} = w_{old} - (Z^T Z + \lambda I)^{-1} * Z^T * \varepsilon(w_{old}), \quad (1)$$

where

w_{old} – weight vector of the previous iteration;

w_{new} – weight vector of the current iterations;

Z – Jacobian matrix with the dimensionality $m \times n$; m – is the number of learning examples for each iteration, n – total number of weights in the network;

λ – learning ratio;

I – identical matrix with the dimensionality $n \times n$;

ε – vector of n elements, which contains forecast errors for each learning example.

To enhance the predicting power of our model we introduce here pre-processing techniques of the Independent Component Analysis (ICA). This is a method of identifying the key and important signals in the large, noisy data. ICA is often compared with another useful processing tool – Principal Component Analysis (PCA). However, the general difference of ICA from PCA is that we obtain purely independent vectors on which a process can be decomposed, whereas PCA requires only non-correlatedness of such vectors. Moreover, ICA allows non-Gaussian distributions, which is quite useful and realistic assumption, especially for financial data.

The ICA stems from the so-called “cocktail party” problem in acoustics. The problem is the following: assume that we have i number of people(s) talking in the room and j number of microphones(x) which record their voices. For two people and two microphones signals from the microphones are as follows in (2):

$$\begin{aligned}
x_1 &= a_{11} * s_1 + a_{12} * s_2 \\
x_2 &= a_{21} * s_1 + a_{22} * s_2
\end{aligned}
\tag{2}$$

Consequently, we should set mixing matrix A which transforms voices into the recordings, (1):

$$A = \begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{li} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ a_{j1} & a_{j2} & \cdots & a_{ji} \end{Bmatrix}
\tag{3}$$

The task for the researcher consists then in finding a demixing matrix A^{-1} which enables to get the vector of voices s knowing only the vector of the recordings x , (4):

$$s = A^{-1} * x
\tag{4}$$

When we apply ICA for the stock market we assume that the empirical stock returns are the “recordings”, the noisy signals of the original “voices” which determine the real process of price movements. Consequently, when we obtain a de-mixing matrix A^{-1} then we get a powerful tool for extracting the most important information about the price movements. Furthermore, ICA allows us to reduce the dimensionality of the empirical data without losing significant information. It is very important while using ANNs, because, on the one hand we should present the network as much relevant information as possible, but, on the other hand, too much input information leads to lengthy computational procedures and problems with convergence to a nontrivial solution.

As it was mentioned above, we use two types of inputs in the Levenberg-Marquardt ANN. First input is the logarithmic return of the stock for the day which precedes the day of the forecast. Second input is derived from the processing of ten previous logarithmic returns with ICA algorithm: we get the de-mixing matrix A^{-1} and the subsequent vector of independent components s . Then we transform this vector to the scalar value considering the most influential independent component.

In the section “Results” we show that such pre-processing turns out to be very useful for stock market forecasting. Moreover, it is worth mentioning that ICA can be used as a self-sufficient forecasting tool for various financial markets.

2 The Haken network

The second ANN which is used for forecasting U.S. stock market is quite different from the Levenberg-Marquardt network. It is the network of Herman Haken, German scientist and the founder of synergetics.

This ANN is self-learning and uses a “library” of pre-set values which by default represent all possible states of the process. Therefore, during a number of iterations the network converges to one of these values. The Haken ANN is

widely used for image recognition. For example, when we set a task of recognizing letters of the alphabet we use the whole alphabet as a pre-set “library”. It is obvious, because any letter which is presented to the network is essentially a part of the alphabet.

However, we aim to apply the Haken ANN for the stock market forecasting, and the situation here is much more complicated. We must choose the “library” which contains all possible states of the market. To solve this task we resort to two important assumptions. Firstly, we argue that all necessary information which is needed for the forecast is contained in the returns of the stock during ten trading days before the day for which the forecast is calculated. Secondly, we assume that the using of processing techniques of the ICA and PCA, which eventually reduces the information dimensionality, allows us to extract most important and valuable information signals.

Thus, to obtain the “library” of pre-set values we use the eigenvectors of the covariance matrix of the subsequent empirical vectors of stock returns (PCA) or the de-mixing matrix of the empirical vectors obtained from ICA.

The network functions as follows, (5):

$$q^* = q + \sum_{k=1}^M \lambda_k v_k^T q v_k + B \sum_{k=1}^M (v_k^T q)^2 (v_k^T q) v_k + C(q^T q)q, \quad (5)$$

where

q – vector of M elements which the network tries to optimize. Initially this vector is deliberately made noisy to ignite the process of learning, thus we assume that the real data on the stock market is also noisy in the similar way;

q^* - vector which the network finally reconstructs;

V – matrix which plays a role of the “library” and contains pre-set values which are obtained from PCA or ICA;

λ – learning ratio;

B – computational parameters which calibrating has an influence on the convergence of the network and the speed of learning.

The final forecasting signal is obtained by subtracting the empirical vector from the reconstructed one.

3 Trading rules

Now we present trading rules which were used while working with the Levenberg-Marquardt and the Haken ANNs.

Firstly, we should specify the data which we forecast. We predict price movements of the 30 liquid stocks of the U.S. index S&P 500¹¹ in the period from November, 7th, 2008 to May, 2nd, 2010.

For each trading day t we make forecast via our ANNs for each stock. When the forecasts are made we range them according to their absolute value. The final selection of the stocks in the virtual portfolio is based on two opposing trading rules. According to the Rule A we select from 1 to 5 stocks with the highest forecast value²² (note that at this step we do not know the real return of day t which makes our forecast truly out-of-sample). According to the Rule B we select from 1 to 5 stocks with the lowest forecast values.

The reason for using Rule B is widely recognized phenomenon of “mean reversion” in the financial data. Thus, if Rule B is successful, then our ANN is capable of detecting this property of the market.

The dynamics of our trade portfolio will be compared to the dynamics of the S&P 500 index and the dynamics of the portfolio if the decision-making was based on the simple auto-regression model (while the trading rules A and B are retained), (6):

$$r_t^* = \alpha_t + \beta_t * r_{t-1}, \quad (6)$$

where

r_t^* - forecast value of the logarithmic return of the stock for the trading day t ,
 α_t, β_t - auto-regression coefficients,
 r_{t-1} - logarithmic return of the stock for the trading day $t-1$.

4 Results

Now we present some of the results of the forecasting using the Levenberg-Marquardt and the Haken ANNs. Due to the limited space of this paper we demonstrate here only most successful examples.

Figure 1 demonstrates relative dynamics of our virtual portfolio (red line) using the Levenberg-Marquardt ANN and trading Rule B (five stocks with “worst” forecasts). Blue line is a portfolio when the decision-making is based on the auto-regression model. Blue line is the S&P 500 index. The horizontal axis is time and t indicates trading days. The vertical axis displays the value of the portfolio with the initial value of 1.

Figure 2 demonstrates relative dynamics of our virtual portfolio (red line) using the Haken ANN with the PCA pre-processing and trading Rule B (one stock with the “worst” forecast). Blue line is a portfolio when the decision-making is based on the auto-regression model. Green line is the S&P 500 index.

¹ We use closing prices of the following stocks: ExxonMobil, Apple, Microsoft, General Electric, Procter&Gamble, Johnson&Johnson, Bank of America, JPMorgan Chase, Wells Fargo, IBM, Chevron, Sisco Systems, AT&T, Pfizer, Google, Coca Cola, Intel, Hewlett Packard, Wal Mart, Merck, PepsiCo, Oracle, Philip Morris International, ConocoPhillips, Verizon Communications, Schlumberger, Abbott Labs, Goldman Sachs, McDonalds, QUALCOMM.

² *2 Not that in this model we use only long positions, short selling is not allowed.*

The horizontal axis is time and t indicates trading days. The vertical axis displays the value of the portfolio with the initial value of 1.

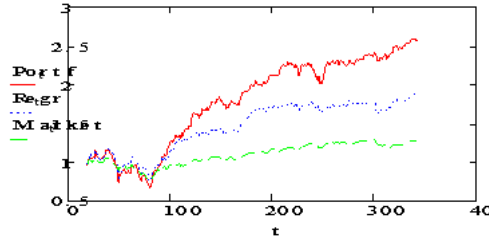


Fig. 1.

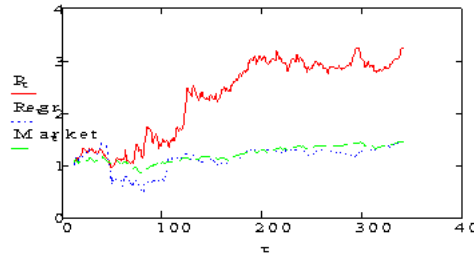


Fig. 2.

Figure 3 demonstrates relative dynamics of our virtual portfolio (red line) using the Herman Haken ANN with the ICA pre-processing and trading Rule A (one stock with the “best” forecast). Blue line is a portfolio when the decision-making is based on the auto-regression model. Green line is the S&P 500 index. The horizontal axis is time and t indicates trading days. The vertical axis displays the value of the portfolio with the initial value of 1.

5 Conclusions

Using of pre-processing techniques of the ICA and PCA with the ANNs proved to be a reliable decision-support mechanism for trading on the liquid stock market. Dynamics of the subsequent portfolios outperform portfolios which follow simple auto-regression forecast or linked to the stock index. Furthermore, the Levenberg-Marquardt and Haken ANNs displayed the ability to reveal the “mean reversion” phenomenon in the complex market data and use it for future forecasts.

However, despite the success of the Levenberg-Marquardt and the Haken ANNs and proper pre-processing techniques we still face difficulties in making up a strategy which will guarantee robust and stable growth of the portfolio over continuous period of time. Moreover, more theoretical research is needed to

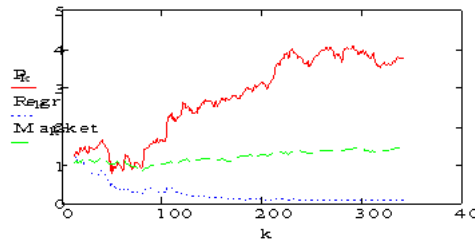


Fig. 3.

justify the argument that it is the neural network decision-making mechanism which is used by traders in real life. It is also obvious that more in-depth study is needed to explain the phenomenon of “mean reversion”. Some of these issues will be the topics of the future research.

References

1. Back A.D., Weigend A.S. A First Application of Independent Component Analysis to Extracting Structure from Stock Returns// International Journal of Neural Systems, Vol. 8, No.5 (October, 1997).
2. Bishop C.M. Neural Networks for Pattern Recognition. Oxford University Press, 1995 – 483 p.
3. Bell J.I., Sejnowski T. J. An information-maximisation approach to blind separation and blind deconvolution//Neural Computation, 7, 6, 1004-1034 (1995).
4. Grriz J.M., Puntonet C.G., Moiss Salmern, E.W. Lang Time Series Prediction using ICA Algorithms//IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 8-10 September 2003, Lviv, Ukraine.
5. Hyvrinen A., Oja E. Independent Component Analysis: Algorithms and Applications//Neural Networks, 13(4-5):411-430, 2000.
6. Krse B., van der Smagt P. An Introduction To Neural Networks, Eight Edition, November 1996.
7. Lu C.-J., Le T.-S., Chiu C.-C. Financial time series forecasting using independent component analysis and support vector regression//Decision Support Systems 47 (2009) 115–125.