15th Triennial World Congress, Barcelona, Spain

ON NONLINEAR MODELING FOR THE PREDICTABILITY OF EQUITY RETURN

K.P. Lam *

* *Department of Systems Engineering & Engineering Management*
The Chinese University of Hong Kong, Sha Tin, Hong Kong
Email: kplam@se.cuhk.edu.hk

Abstract: Non-stationary time series are commonly found in financial applications. Added to the complexity are the time-varying nature and non-linearity of accurate models for describing the dynamic behavior of these financial time series. We extend the techniques of cointegration to handle time-varying, non-linear relationship between a time series ("news") and its causally affected time series. The predictability of daily return, as related to the NASDAQ indexes and to a possible NASDAQ-GEM relationship, is investigated based on a proposed "news" model for dynamic changes. The effectiveness and robustness of neural network models for handling non-linearity is compared with linear least-squares estimation.

Keywords: Cointegration time series, equity return, news modeling, neural networks

1. INTRODUCTION

Recently there has been much interest in non-linear modeling (Campbell *et al.*, 1996), especially in using neural network techniques (Refenes *et al.*, 1996), for financial applications that are broadly encountered across diverse discipline of finance, econometrics, and engineering. Practical problems abound for more effective use of intelligent system technology, given the tremendous growth of Internet-based financial information for investment decision by both individual and institutional users. This paper considers some important issues related with the use of econometrics (Campbell *et al.*, 1996; Enders, 1995) and system identification techniques (Ljung, 1999), neural networks (Haykin, 1994; White, 1989; Anders *et al.*, 1998) for predicting equity return. Two empirical studies are performed on the relevance of these issues on the NASDAQ indexes, and on the causal effect of NASDAQ on a smaller new market called the Growth Enterprise Market (GEM).

Research on cointegration (Enders, 1995; Enders, 1996) is rather extensive in the econometrics literature, mostly related with statistical significance testing using the Engle & Granger procedure (Engle and Granger, 1987) and Johansen methodology (Johansen, 1988). Nonlinear cointegration has also been studied by Burgess & Refenes (Burgess and Refenes, 1996), in particular the use of neural network for "conditional cointegration". In Section 2 we focus mainly on a simple two time-series case to clarify the important role of error correction, and consider various model extensions to deal with nonlinearity, current and expected "news", and equity return prediction.

Section 3 describes our first empirical study on the NASDAQ indexes. As the NASDAQ composite index must be related with the other NASDAQ indexes to different extent, regression analysis is relevant in classifying any possible linear equilibrium relationship. The effectiveness of using linear and nonlinear models, including neural network, for predicting equity and its daily return is described. Particular emphasis is paid on the reduction of error variance and the relative sensitivity between in-sample and out-of-sample data. The Growth Enterprise Market (GEM) is a recently introduced second board of the Stock Exchange of Hong Kong. It is widely believed that the NASDAQ index has been one of the major "news" affecting the GEM stock prices and the Growth Enterprise Index (GEI). In Section 4, an empirical study is performed on testing a possible NASDAQ-GEM relationship for equity return prediction.

2. COINTEGRATION MODELING: LINEARITY AND NONLINEARITY

The concept of cointegration, as thoroughly discussed in (Enders, 1995), applies to many economic and financial time series, where a linear equilibrium relationship exists among non-stationary variables. For the simple case of two time series, y_t and p_t, the long-run relationship can be expressed simply as

$$y_t = \beta_0 + \beta_1 p_t + e_t \tag{1}$$

where e_t is a zero-mean normal innovation process. The significant contribution of Engle & Granger (Engle and Granger, 1987), in deriving an equivalent error correction formulation (Granger representation theorem), avoids the misspecification error when regression analysis is performed on the differences of the non-stationary variables. The error correction term (the residual estimate at $t-1$ is $\hat{e}_{t-1} = y_{t-1} - \beta_0 - \beta_1 p_{t-1}$) contributes to the regression analysis of the difference equation

$$\Delta y_t = \alpha_1 + \alpha_y \hat{e}_{t-1} + \sum_{i=1}^{m} \alpha_{11}(i)\Delta y_{t-i} + \sum_{i=1}^{n} \alpha_{12}(i)\Delta p_{t-i} + \epsilon_{yt} \tag{2}$$

where ϵ_{yt} is a white-noise disturbance, Δ is the backward shift operator $(= 1 - z^{-1})$, and α's are constants to be determined. There are several limitations on using linear regression analysis for cointegration. Firstly, the possibility of a nonlinear long-run relationship $f(.)$ should be addressed, i.e.,

$$y_t = f(p_t) + e_t \tag{3}$$

where the residual estimate $\hat{e}_{t-1}$ at $t-1$ becomes $y_{t-1} - f(p_{t-1})$.

Secondly, a more general nonlinear difference equation model can be used instead of Eqn. (2):

$$\Delta y_t = g_1(\Delta y_{t-1}, \ldots, \Delta y_{t-m}, \hat{e}_{t-1}, \Delta p_{t-1}, \ldots, \Delta p_{t-n}) \tag{4}$$

Thirdly, the correction term in Eqn. 2 has not taken account of the possible causal effect of p_t on y_t at time t. In addition, the availability of the future expectation of p_t (denoted by $\hat{p}_{t+1}$) can contribute to a more precise determination of Δy_t. We can classify both p_t and $\hat{p}_{t+1}$ as "news" items, and consider the following as a special case of "news" modeling:

$$news(t) = \Delta p_t \tag{5}$$

$$news(t+1) = \Delta \hat{p}_{t+1} \tag{6}$$

Thus, "news" modeling of p_t can be incorporated into Eqn. (2), as

$$\Delta y_t = \alpha_1 + \alpha_y \hat{e}_{t-1} + \sum_{i=1}^{m} \alpha_{11}(i)\Delta y_{t-i} + \sum_{i=1}^{n} \alpha_{12}(i)\Delta p_{t-i} + \gamma_1 news(t+1) + \gamma_2 news(t) + \epsilon_{yt} \tag{7}$$

Eqn. (7) can also be adjusted to handle nonlinearity:

$$\Delta y_t = g_2(\Delta y_{t-1}, \ldots, \Delta y_{t-m}, \hat{e}_{t-1}, \Delta p_{t-1}, \ldots, \Delta p_{t-n}, news(t+1), news(t)) \tag{8}$$

By drawing parity with Eqns. (7)-(8) we can study the predictability of the daily return r_t of the underlying equity by considering the modeling of $\Delta \log(y_t) = r_t$ and $\Delta \log(p_t) = s_t$, either linearly as

$$r_t = \Delta log(y_t) = \alpha'_1 + \alpha'_y \hat{e}_{t-1} + \sum_{i=1}^{m} \alpha'_{11}(i) r_{t-i} + \sum_{i=1}^{n} \alpha'_{12}(i) s_{t-i} + \gamma'_1 news'(t+1) + \gamma'_2 news'(t) + \epsilon'_{rt} \tag{9}$$

or nonlinearly as,

$$r_t = g_3(r_{t-1}, \ldots, r_{t-m}, \hat{e}_{t-1}, s_{t-1}, \ldots, s_{t-n}, news'(t+1), news'(t)) \tag{10}$$

where $news'(t) = s_t$ and $news'(t+1) = \hat{s}_{t+1}$. Eqn. (9)-(10) form the basis for predicting equity return in subsequent sections.

3. DYNAMIC MODELING OF NASDAQ COMPOSITE INDEX

A study was performed on two years of daily NASDAQ market indexes, including Composite, 100, Financial-100, Computer, Industrial, Telecommunication, and Biotech. For each index, this corresponds to a data file of 482 matching values from June 15, 1988 to June 12, 2000 by eliminating inconsistent records. The first half (241 records) is used for in-sample learning, while the latter half is used for out-of-sample testing. The six market indexes are ranked by normalized correlation, which is obtained by performing a linear regression on normalized data of zero-mean and unit variance. A high correlation close to 1 indicates that the Composite index/ Computer index has a long-run linear equilibrium relationship (Eqn. (1)). The correlation ranking is found as follows: (Composite/Computer: 0.9969), (Composite/100: 0.995), (Composite/Industrial: 0.9920), (Composite/ Telecom: 0.9874), (Composite/ Biotech: 0.9631), (Composite/ Financial-100: -0.5697). It is thus observed that the long-run equilibrium relationship between the Composite index and Financial-100 index is far from being linear.

3.1 *Linear model for* r_t

For each NASDAQ composite/ component index pair, we measure the significance of using that component index (p_t) for predicting the Composite index (y_t), and the sensitivity between in-sample and out-of-sample performance. All modeling is based on Eqn. (9) using a single Adaline neuron (Widrow and Sterns, 1985; Demuth and Beale, 1998) for batch least-squares estimation (Ljung, 1999). For simplicity, α_1', α_y', α_{12}', and γ_1' are taken as zero, $m = 3$, and Eqn. (9) then reduces to

$$r_t = \sum_{i=1}^{3} \alpha_{11}'(i) r_{t-i} + \gamma_2' news'(t) + \epsilon_{rt}' \quad (11)$$

Based on the predicted value of the daily return ($\hat{r}_t$) from Eqn. (11), the predicted composite index can also be readily obtained

$$\hat{y}_t = y_{t-1} \exp(\hat{r}_t) \quad (12)$$

The error variances for both the predicted composite index and its daily return are tabulated with regard to the presence or absence of $news'(t)$ in the model. In general, we observe that the use of component index information, $news'(t)$ at time t, contributes positively to the reduction of the prediction error variance of the composite index and its daily return. The reduction is rather significant when there is a high correlation between the composite index and the component index. As measured by the percentage of improvement (= reduction of error variance due to $news'(t)$ / error variance without using $news'(t)$), it ranges from less than 8% for the case of Financial-100 to greater than 96% for NASDAQ-100. For the best case of NASDAQ-100, its use for out-of-sample testing can yield a low prediction error for the NASDAQ composite (roughly 1 standard deviation = 16.88 which is less than 1% of the NASDAQ composite index value).

As expected, the out-of-sample error variance is higher than that obtained from in-sample learning using the single-neuron Adaline model (batch least squares) for most cases. However, the sensitivity is within an acceptable range (1 to 5 for y_t and around 1 for r_t). There is an indication that this out-of-sample sensitivity for r_t is less than that for y_t. The use of $news'(t)$ appears to contribute slightly on the reduction of this sensitivity. Other alternatives for Eqn. (9) have also been investigated, such as the use of $\hat{e}_{t-1}$, $news'(t+1)$, and s_{t-i}. There are no marked differences in the prediction error variances by including these terms if $news'(t)$ has already been included. This indicates the significant role of $news'(t)$ on the improvement of predictability of r_t and y_t, apparently due to the availability of information of p_t up to time t.

3.2 *Nonlinear model for* r_t

For a closer look on the details of dynamic changes, a smaller sample of the NASDAQ data was selected in a comparative study on using linear and nonlinear models. The first 51 data records out of a total of 124 are used for in-sample learning, with the remaining being tested for out-of-sample performance. NASDAQ-100 is chosen as the $news(t)$ that can be used for predicting the NASDAQ composite index (y_t) and its daily return (r_t). We consider mainly two types of neural network for nonlinear modeling: backpropagation network and radial basis function network. For a comparative evaluation with the linear model of Eqn. (11), the nonlinear model assumes a reduced structure from Eqn. (10):

$$r_t = g_3(r_{t-1}, r_{t-2}, r_{t-3}, news'(t)) \quad (13)$$

The same Adaline neuron (batch least squares) (Demuth and Beale, 1998) discussed in the previous subsection provides a reference linear model benchmark. The linear model (using $news(t)$) gives good prediction of r_t and y_t for in-sample data fairly well, but its dynamic performance noticeably deteriorates for out-of-sample data. The results observed previously, most notably the effect of $news(t)$ on the reduction of error variance, are again demonstrated.

(a) Backpropagation network (Hagan *et al.*, 1996; Haykin, 1994; Demuth and Beale, 1998)

The backpropagation network is using a conventional feedforward multilayer architecture with local gradient-descent optimization procedure. It has been found that the use of standard function ("newff") and default parameter values in MATLAB's Neural Network (NN) toolbox (Demuth and Beale, 1998) is quite adequate for our purpose. In this particular case, we choose a 4-input ($\{r_{t-1}, r_{t-2}, r_{t-3}, news'(t)\}$), 3-layer $S = [5, 3, 1]$ structure with single output (r_t). The converged model after 500 epochs is used. Other choices for S and initial parameter setting have been attempted but give largely similar results. The backpropagation-trained nonlinear model gives a significant in-sample improvement over linear model, though at the expense of a worsened out-of-sample performance. Hence, the sensitivity of this nonlinear model is substantially higher. In regard to the effect of $news(t)$ on error variance reduction, the percentage of improvement for the nonlinear model case is significant (92% versus 96% for the linear model case).

(b) Radial basis function network (Demuth and Beale, 1998; Hutchinson *et al.*, 1994)

Another neural network model based on radial basis funtion is also used for nonlinear modeling. The MATLAB's NN standard function ("newrbe") is adequate for our testing. Different spreads (as specified in $newrbe(P, T, spread)$) for the Gaussian point function have been tried to obtain different degree of model "fitness" to the data. As compared with the backpropagation network, there is an extremely high sensitivity of radial-basis-function network for out-of-sample data. Despite the near-perfect match for in-sample prediction, the out-of-sample performance is not promising.

4. TIME-VARYING NASDAQ-GEM RELATIONSHIP

The Growth Enterprise Market (GEM) in Hong Kong started trading on November 25, 1999. Given the market size disparity, it is apparent that a "one-way" causal relationship exists between NASDAQ and GEM. However, this relationship is observed to be non-trivial, indicating that a simple functional representation is not adequate. Unlike the case for NASDAQ indexes as discussed in the previous section, the NASDAQ-GEM relationship is more difficult to establish. However, it is a more practical case concerning the influence of NASDAQ on GEM as "news". This is in contrast with deriving relationships among NASDAQ indexes only, as these indexes should be known at the same time and hence one index could not "practically" affect another index as "news". Having said so, our study described in the previous section is useful for quantifying the role of the "artificial" $news'(t)$, in regard to the use of linear and nonlinear modeling for return prediction.

It is important to treat the NASDAQ-GEM relationship as a time-varying one, because the "news" effect of NASDAQ on GEM can come as future expectation or speculation, immediate reaction, or delayed response at different period of time. In additon, the magnitude of the effect may also be time varying. We consider 68 data samples of the NASDAQ composite index and Hong Kong's Growth Enterprise Index (GEI) between March 17, 2000 (the first date when GEI data is available) and June 30, 2000. A careful dichotomy of the similarities between the two normalized indexes gives more details on the time-varying nature of the relationship. The correlation between NASDAQ and GEI over the whole period is 0.8192. For the first half of the data samples, the correlation is much higher (0.935) than that for the second half (-0.18). For accurate modeling, some types of time-varying model that can follow the changes would be needed. Ideas of using multiple models with appropriate switching, and recursive estimation procedures have been explored; but preliminary results are not very supportive for significant improvemnet over more simpler models. This may be attributed to the lack of a priori information and the unknown nature of the changes, especially when the changes occur over fairly short time scale that makes it very difficult for any adaptive scheme to follow accurately.

In the follwing we only consider a simple single-model approach with batch-mode learning. However, additional terms for the future expectation, immediate reaction, and delayed respnse to news are included to handle possible time variation. The robustness of the model to out-of-sample data will be discussed.

4.1 *Linear model for the NASDAQ-GEM relationship*

While accepting the complexity of the time-varying and non-linear nature of the relationship, we first look at the advantages and limitation of using the following linear model for in-sample representation and out-of-sample prediction of GEI.

$$r_t = \sum_{i=1}^{3} \alpha'_{11}(i) r_{t-i} + \alpha'_{12}(i) s_{t-1} + \gamma'_1 news'(t+1) + \gamma'_2 news'(t) + \epsilon'_{rt} \quad (14)$$

where r_t is the daily return of GEI at t, $news'(t+1) = s_{t+1}$ is the future expectation of the NASDAQ-composite return at $t+1$. The current and delayed effect of NASDAQ are represented by $news'(t) = s_t$ and $news'(t-1) = s_{t-1}$. respectively.

(a) Without using NASDAQ information

By neglecting s_{t-1}, $news'(t+1)$, and $news'(t)$, Eqn. (14) reduces to an Autoregressive (AR) model (Ljung, 1999) which is quite effective to capture the dynamics of GEI without taking advantage of any available information on NASDAQ. We used a third-order AR model in our evaluation. The model is learned using the batch mode (equivalent to the batch least squares) of an ADALINE neuron based on the GEI in-samples from 1 to 48. Prediction accuracy is evaluated based on the out-of-sample data from 49 to 68. The estimated model shows an interesting characteristic of using AR model: the predicted value of y_t is somewhat delayed when compared with the actual value. In fact, this should be expected from an AR model which is using information only up to $t-1$ for predicting into time t.

As observed in the GEI daily return r_t, the estimated AR model appears to be ineffective in capturing the changes in the actual return for both in-samples and out-of-samples. In comparing with in-samples, out-of-sample error variance is better despite the model inaccuracy. This is mainly attributed to the less turbulent stock changes for the chosen out-of-samples in this particular case. However, it also indicates the robustness of linear least-squares estimation for making effective prediction over a broad range of data samples.

(b) Using NASDAQ information

The NASDAQ information or "news" is incorporated into the AR model as exogenous input. The input can either be based on future expectation (as $news'(t+1)$), current information up to time t (as $news'(t)$), time-lagged response or delayed information up to time $t-1$ (as $news'(t-1) = s_{t-1}$), for a better prediction on the GEI, y_t, at time t. It is assumed that $news'(t)$ can causally affect y_t (i.e., the news of the NASDAQ composite index in a trading day in U.S. will affect the GEI in the following trading day in Hong Kong due to the time difference which allows the spread of "news").

Apparently, it is observed that the predicted values of y_t are not "delayed" when compared with the actual values. The predictability of the GEI daily return is substantially improved. The inclusion of NASDAQ information as exogenous input can improve the performance by reducing the error variance on GEI. It is especially so when timely NASDAQ information on the most recent day (as $news'(t)$) preceding the GEM trading is available. The advantage of including $news'(t)$ is noted for error variance reduction (roughly 20% versus 10% when using $news'(t-1)$ only. We observe that the inclusion of future expectation $news'(t+1)$ is not very significant in reducing error variance.

It is useful to make a comparative evaluation between empirical results on NASDAQ indexes (in section 3) and on NASDAQ-GEM. For both studies, the positive effect of $news'(t)$ in improving predictability of the daily return r_t and actual index y_t (by reducing error variance) is established. However, the effect is comparatively less significant for NASDAQ-GEM with around 20% improvement (versus close to 96% for NASDAQ indexes). This may be attributed to several factors, such as the relatively low correlation (0.8192 versus 0.9969) and the time-varying nature of the NASDAQ-GEM relationship. These factors further indicate a nonlinear long-run equilibrium (Eqn. (3)) between the GEI and NASDAQ, compared with a fairly linear relationship (Eqn. (1)) among most of the NASDAQ indexes.

4.2 *Nonlinear model for the NASDAQ-GEM relationship*

We consider the possible improvement using nonlinear models for representation and prediction. While the previous study indicates that linear model based on least squares learning can give reasonable in-sample performance, there is much ground for reducing the error variance if more accurate modeling is applied. A restricted form of the nonlinear model given by Eqn. (10) has been considered:

$$r_t = g_3(r_{t-1}, r_{t-2}, r_{t-3}, s_{t-1}, news'(t+1), news'(t)) \qquad (15)$$

(a) Back-propagation (BP) neural network

Compared with the use of linear model, a much better in-sample performance in error variance reduction (around 89%) is obtained. However, the out-of-sample behavior is substantially worser than that of linear model, indicating that the nonlinear model is very sensitive to modeling error. The overall error variance in using nonlinear back-propagation NN model is higher than that of the linear model despite its more accurate in-sample representation.This agrees well with our previous observation (in section 3) on using nonlinear modeling based on BP for NASDAQ indexes. The sensitivity of the nonlinear model is more pronounced here, as the linear model gives a better behavior due to the less turbulent nature of the out-of-sample data.

While the positive effect of using $news'(t)$ is less drastic for linear model, it is not the case for using nonlinear BP network. As given in Table 4(b), the percentage improvement in reducing in-sample error variance for r_t and y_t is around

87%. This is quite a significant improvement and compares favourably with the 91% as reported for the case of NASDAQ indexes. Despite this in-sample improvement, there is little gain in using $news'(t)$ to reduce out-of-sample sensitivity.

(b) Radial-basis-function (RBF) network

It is readily seen that the nonlinear RBF network's behavior is similar to that of the BP, with close to 87% improvement in error variance reduction over the linear model. Again, the nonlinear NN gives very good in-sample performance but bad out-of-sample prediction. The high sensitivity of the nonlinear NN to modeling error is clearly demonstrated. The positive effect of $news'(t)$ in nonlinear modeling is confirmed with the RBF network. The results are in close resembance with that of the NASDAQ indexes, where near-perfect in-sample match can easily be attained with the RBF. An almost 100% improvement in error variance by using $news'(t)$ serves little to reduce the out-of-sample sensitivity of the RBF network.

5. CONCLUSION

Two different approaches have been proposed for studying the predictability of equity return using linear and nonlinear modeling. We extend the techniques of cointegration time series to deal with both linear and nonlinear models, including long-run equilibrium, correlation of "news" effect, and time-varying model changes. Two reference practical cases are discussed in detail: one is related with the largely linear relationships among the NASDAQ indexes, and the other is related with the modeling of a more difficult time-varying NASDAQ-GEM relationship. Linear least-squares estimation has been found to give good in-sample performance, and is robust in dealing with model inaccuracy for out-of-sample tests. Nonlinear models using neural network techniques, in particular the back-propagation and radial basis function networks, are found to give much improved in-sample performance over linear models, but at the expense of high sensitivity towards model inaccuracy for out-of-sample tests. Through the use of proper linear and nonlinear modeling, more accurate prediction of equity return and the actual price can be obtained by effective incorporation of timely "news" effects. Extensive empirical results are provided for the two pracitical cases.

6. REFERENCES

Anders, U., O. Korn and C. Schmitt (1998). Improving the pricing of options: A neural network approach. *Journal of Forecasting* **17**, 369–388.

Burgess, A.N. and A-P. N. Refenes (1996). Modeling nonlinear cointegration in international equity index futures. *Neural Networks in Financial Engineering (Edited by A-P. Refenes et. al.)* pp. 50–63.

Campbell, J.Y., A.W. Lo and A. MacKinlay (1996). *The Econometrics of Financial Markets.* Princeton University Press.

Demuth, H. and M. Beale (1998). *Neural Network Toolbox: User's Guide Version 3.0 (Fifth Printing).* The MathWorks, Inc.

Enders, W. (1995). *Applied Econometric Time Series.* New York: Wiley.

Enders, W. (1996). *RATS Handbook for Econometric Time Series.* New York: Wiley.

Engle, R.F. and C. Granger (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica* **55**, 251–276.

Hagan, M.T., H.B. Demuth and M.H. Beale (1996). *Neural Network Design.* Boston: PWS Publishing.

Haykin, S. (1994). *Neural networks: A comprehensive foundation.* New York : Macmillan College Publishing Company.

Hutchinson, J.M., A.W. Lo and T. Poggio (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance* **49**, 851–889.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.

Ljung, L. (1999). *System Identification: Theory for the User (Second Edition).* Prentice Hall.

Refenes, A-P. N., Y. Abu-Mostafa, J. Moody and A. Weigend (Editors) (1996). *Neural networks in Financial Engineering.* Singapore: World Scientific.

White, H. (1989). Learning in neural networks: A statistical perspective. *Neural Computation* **1**, 425–464.

Widrow, B. and S.D. Sterns (1985). *Adaptive Signal Processing.* Prentice-Hall.