



Fuzzy dual-factor time-series for stock index forecasting

Hsing-Hui Chu^a, Tai-Liang Chen^{b,*}, Ching-Hsue Cheng^c, Chen-Chi Huang^c

^a Department of Accounting and Information Technology, Ling Tung University 1, Ling Tung Road, Nantun, Taichung 408, Taiwan, R.O.C.

^b Department of Information Management and Communication, Wenzao Ursuline College of Languages, 900, Mintsu 1st Road, Kaohsiung 807, Taiwan, ROC

^c Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan, ROC

Abstract

There is an old Wall Street adage goes, “It takes volume to make price move”. The contemporaneous relation between trading volume and stock returns has been studied since stock markets were first opened. Recent researchers such as Wang and Chin [Wang, C. Y., & Chin S. T. (2004). Profitability of return and volume-based investment strategies in China’s stock market. *Pacific-Basin Finance Journal*, 12, 541–564], Hodgson et al. [Hodgson, A., Masih, A. M. M., & Masih, R. (2006). Futures trading volume as a determinant of prices in different momentum phases. *International Review of Financial Analysis*, 15, 68–85], and Ting [Ting, J. J. L. (2003). Causalities of the Taiwan stock market. *Physica A*, 324, 285–295] have found the correlation between stock volume and price in stock markets. To verify this saying, in this paper, we propose a dual-factor modified fuzzy time-series model, which take stock index and trading volume as forecasting factors to predict stock index. In empirical analysis, we employ the TAIEX (Taiwan stock exchange capitalization weighted stock index) and NASDAQ (National Association of Securities Dealers Automated Quotations) as experimental datasets and two multiple-factor models, Chen’s [Chen, S. M. (2000). Temperature prediction using fuzzy time-series. *IEEE Transactions on Cybernetics*, 30 (2), 263–275] and Huarng and Yu’s [Huarng, K. H., & Yu, H. K. (2005). A type 2 fuzzy time-series model for stock index forecasting. *Physica A*, 353, 445–462], as comparison models. The experimental results indicate that the proposed model outperforms the listing models and the employed factors, stock index and the volume technical indicator, VR(*t*), are effective in stock index forecasting.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Fuzzy time-series model; Dual time-series; Stock index forecasting; Fuzzy linguistic variable

1. Introduction

Time-series models have utilized the fuzzy theory to solve various domain forecasting problems, such as university enrolment forecasting (Chen, 1996, 2002; Chen & Hsu, 2004; Song & Chissom, 1993; Song & Chissom, 1994), financial forecasting (Huarng, 2001a, 2001b; Huarng & Yu, 2005; Lee, Wang, Chen, & Leu, 2006) and temperature forecasting (Chen, 2000). However, to deal with the forecasting problems with complex influencing factors such as weather forecasting (Chen, 2000) and stock price forecasting (Huarng & Yu, 2005; Kim, Min, & Han, 2006; Lee

et al., 2006), more factors should be considered to derive more accurate results.

In stock markets, stock trading volume plays an important role in stock technical analysis theories. Recently, several researchers such as Ting (2003), Kitt and Kalda (2005) have revealed that the relation between trading volume and price fluctuation is obviously strong. Hence, in this paper, we propose a new time-series model, which considers stock index and a technical analysis indicator of trading volume in forecasting process, to forecast stock index. In the proposed model, two forecasting approaches are employed to process these two factors separately and a dual-factor forecasting equation is used to generate conclusive forecasts. By employing the TAIEX and NASDAQ as experimental datasets, the results show that our model surpasses in accuracy the models advanced by Chen (2000) and Huarng and Yu (2005).

* Corresponding author. Tel.: +886 920975168.

E-mail addresses: g9320817@yuntech.edu.tw (T.-L. Chen), chcheng@yuntech.edu.tw (C.-H. Cheng), g9423730@yuntech.edu.tw (C.-C. Huang).

The remaining content of this paper is organized as follows: Section 2 introduces the related literature; Section 3, demonstrates the proposed model and algorithm; Section 4 evaluates the performance of the proposed model; and Section 5 concludes this paper.

2. Related works

This section includes two parts of literature reviews: fuzzy time-series and causality in stock markets.

2.1. Fuzzy time-series

Song and Chissom (1993) first applied the fuzzy theory (Zadeh, 1965; Zadeh, 1975a; Zadeh, 1975b; Zadeh, 1976) in time-series and provided the definitions and framework of fuzzy time-series, where time-series data are represented by fuzzy sets instead of crisp values. The definitions and research processes are introduced as follows (Song & Chissom, 1993; Song & Chissom, 1994).

Definition 1. Let “ $Y(t)$ ($t = \dots, 0, 1, 2, \dots$)”, a subset of real numbers, be the universe of discourse by which fuzzy sets $f_j(t)$ are defined. If $F(t)$ is a collection of $f_1(t), f_2(t), \dots$, then $F(t)$ is called a fuzzy time-series defined on $Y(t)$.

Definition 2. Assume that $F(t)$ is a fuzzy time-series and $F(t) = F(t-1) * R(t-1, t)$, where $R(t-1, t)$ is a fuzzy relation and “ $*$ ” is the max–min composition operator. Then $F(t)$ is caused by $F(t-1)$ and it is denoted as “ $F(t-1) \rightarrow F(t)$ ”, where $F(t-1)$ and $F(t)$ are fuzzy sets.

Definition 3. Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive observations, $F(t)$ and $F(t-1)$, referred to as a fuzzy logical relationship (FLR), can be denoted by $A_i \rightarrow A_j$, where A_i is called the left-hand side (LHS) and A_j the right-hand side (RHS) of the FLR.

Definition 4. All fuzzy logical relationships in the training dataset can be further grouped together into different fuzzy logical relationship groups according to the same left-hand sides of the fuzzy logical relationship. For example, there are two fuzzy logical relationships with the same left-hand side (A_i): $A_i \rightarrow A_{j1}$ and $A_i \rightarrow A_{j2}$. These two fuzzy logical relationships can be grouped into a fuzzy logical relationship group.

Definition 5. Suppose $F(t)$ is caused by $F(t-1)$ only, and $F(t) = F(t-1) \times R(t-1, t)$. For any t , if $R(t-1, t)$ is independent of t , then $F(t)$ is named a time-invariant fuzzy time-series, otherwise a time-variant fuzzy time-series.

2.2. Forecasting processes of fuzzy time-series

The forecasting processes of Song and Chissom’s model include six steps: (1) define and partition the universe of discourse; (2) define fuzzy sets for the observations; (3)

fuzzify the observations; (4) establish the fuzzy relationship, R ; (5) forecast; and (6) defuzzify the forecasting results.

To simplify the computations of fuzzy time-series, Chen (1996) proposed an arithmetic approach to Song and Chissom’s model. In subsequent research, Chen proposed several methods, such as high-order fuzzy relationships (Chen, 2002) and more precise linguistic intervals with sub-intervals (Chen & Hsu, 2004) to improve his initial model.

The fuzzy time-series models above employ one factor to forecast enrollments. However, to deal with more complex forecasting problems such as weather forecasting and stock price forecasting, recent fuzzy time-series models employ multiple factors in forecasting processes such as Chen (2000), Huarng and Yu (2005) and Lee et al. (2006). Therefore, to derive better performance, more factors which will improve forecasting accuracy should be considered in fuzzy time-series models.

2.3. Causality in Stock Markets

There is an old Wall Street adage goes, “It takes volume to make price move”. Most of earlier empirical works focused on the contemporaneous relation between trading volume and stock returns. However, recent studies have begun to address the dynamic relation between daily stock returns and trading volume, which is called “causality” (Granger & Wiener, 1969). The correlation can be found in stock markets and is not deniable (Marc & Hiraki, 1999; Hodgson, Masih, & Masih, 2006; Shen & Wang, 1998; Wang & Chin, 2004).

The causality has been analyzed in the econometric literature for GNP (Gross National Product) and money supply relation (Gujarati, 1995; Maddala, 1992) and can be expressed by a dual time-series model which contains mutual influencing factors. In stock markets, the regression of the dual time-series (Ting, 2003) can be defined by a linear functional form (Eqs. (1) and (2), where I_t is the future index, and V_t is the future trading volume)

$$I_t = \sum_i a_i V_{t-i} + \sum_j b_j I_{t-j} + u_t \quad (1)$$

$$V_t = \sum_i l_i V_{t-i} + \sum_j k_j I_{t-j} + v_t \quad (2)$$

Besides the causality in stock markets, it is well known to the practitioner, as an empirical rule, that “volume changes come before price change.” Therefore, stock trading volume should be considered in stock price forecasting as a prevision.

3. Fuzzy dual time-series model

In this section, to improve forecasting accuracy, we propose a new fuzzy time-series model using stock index and volume as two forecasting factors (see Fig. 1) since both these two factors influence the future stock index (Ting, 2003). However, instead of stock trading volume, the

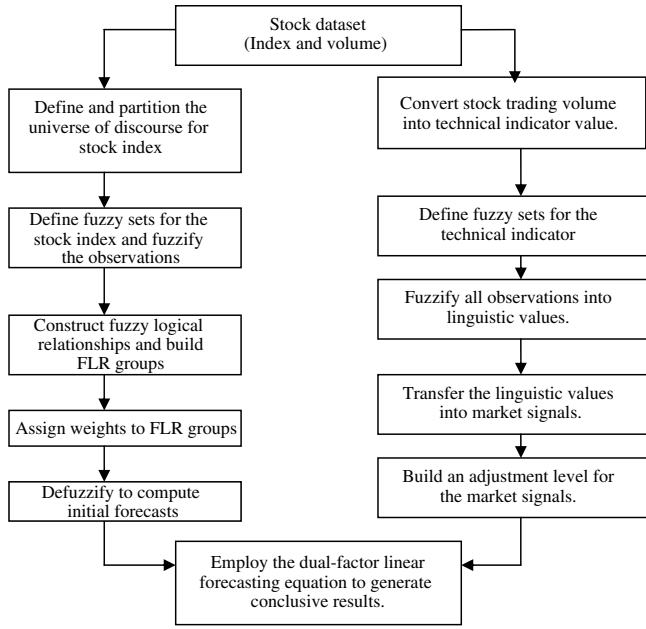


Fig. 1. Forecasting processes of fuzzy dual-factor time-series model.

technical analysis indicator of stock trading volume, $VR(t)$, is employed, because it is a significant indicator (Eq. (3), where $V_{up}(t)$ means the trading volume when the stock index closes up and $V_{down}(t)$ means the trading volume when the stock index closes down) (Kim et al., 2006)

$$VR(t) = \frac{\sum_{i=1}^n V_{up}(t) - \sum_{i=1}^n V_{down}(t)}{\sum_{i=1}^n V(t)} \quad (3)$$

In this model, we provide two forecasting approaches to deal with the two factors separately and integrate these processes to produce a conclusive result by a dual-factor linear forecasting equation (defined in Eq. (4)) (Cheng, Chen, & Huang, 2007). In this forecasting equation, $I(t)$ is the stock index at t ; $forecast(t+1)$ is a forecast for the stock index at $t+1$; b is an adjustment parameter for the initial forecast based on stock index; $L_{VR(t)}$ is a linguistic value of the indicator; M is a signal transferring function which converts a linguistic value into the corresponding market signal; a is an adjustment parameter for the initial forecast based on the technical analysis indicator.

$$\begin{aligned} \text{dual_factor_forecast}(t+1) &= a * M(L_{VR(t)}) * I(t) + b \\ &\quad * (\text{forecast}(t+1) \\ &\quad - I(t)) + I(t) \end{aligned} \quad (4)$$

By the way of summarization, we crystallize the proposed model into a detailed algorithm and specify the algorithm in the next subsection.

3.1. The algorithms of the proposed model

The proposed model consists of three *algorithms*: (1) the forecasting processes using stock index; (2) the forecasting processes using stock trading volume; and (3) the dual-

factor linear forecasting equation to generate a conclusive forecast.

Algorithm 1. The forecasting processes based on the factor of stock index

Step 1: Define and partition the universe of discourse reasonably. Define the universe of discourse, $U = [\text{low}, \text{up}]$, which can cover all observations of stock index in the training dataset. And initially partition the universe of discourse into seven linguistic intervals, u_i .

Step 2: Establish fuzzy sets for the stock index and fuzzify the observations. Define the fuzzy set, A_1, A_2, \dots, A_k , on the universe of discourse by Eq. (5). The value of a_{ij} indicates the grade of membership of u_j in fuzzy set A_i , where $a_{ij} \in [0, 1]$, $1 \leq i \leq k$ and $1 \leq j \leq m$. Find out the degree of each stock index belonging to each A_i ($i = 1, \dots, m$). If the maximum membership of the stock index is under A_k , then the fuzzified stock index is labeled as A_k . The fuzzy logical relationships are generated based on the fuzzified stock index. In this paper, the seven fuzzy linguistic values, $A_1 = (\text{very low price})$, $A_2 = (\text{low price})$, $A_3 = (\text{little low price})$, $A_4 = (\text{normal price})$, $A_5 = (\text{little high price})$, $A_6 = (\text{high price})$ and $A_7 = (\text{very high price})$, are applied (Chen, 1996)

$$\begin{aligned} A_1 &= a_{11}/u_1 + a_{12}/u_2 + \dots + a_{1m}/u_m \\ A_2 &= a_{21}/u_1 + a_{22}/u_2 + \dots + a_{2m}/u_m \\ &\vdots \\ A_k &= a_{k1}/u_1 + a_{k2}/u_2 + \dots + a_{km}/u_m \end{aligned} \quad (5)$$

Step 3: Construct fuzzy logical relationships and build FLR groups. There are two sub-steps included in this process as follow: (1) Establish a relationship between two consecutive linguistic values, $A_i(t-1)$ and $A_j(t)$. Then, represent the relationship into a FLR such as $A_i \rightarrow A_j$, and (2) classify all FLRs with the same LHSs (left-hand sides) to form a FLR group. For example, $A_i \rightarrow A_j$, $A_i \rightarrow A_k$, $A_i \rightarrow A_m$ can be group as $A_i \rightarrow A_j, A_k, A_m$.

Step 4: Assign weights to FLR groups. Assign weights to all FLR groups based on trend-weighted method. And transfer these weights into a normalized weight matrix, $W_n(t)$, which is defined in the following equation (Cheng, Chen, & Chiang, 2006):

$$\begin{aligned} W_n(t) &= [W'_1, W'_2, \dots, W'_j] \\ &= \left[\frac{W_1}{\sum_{k=1}^j W_k}, \frac{W_2}{\sum_{k=1}^j W_k}, \dots, \frac{W_k}{\sum_{k=1}^j W_k} \right] \end{aligned} \quad (6)$$

Step 5: Defuzzify to compute initial forecasts. In this process, the normalized weight matrix, $W_n(t)$, and the defuzzified matrix, $L_{df}(t)$, defined in Eq. (7), are employed (where m_i is the midpoint of each linguistic interval, u_i)

$$L_{df}(t) = [m_1, m_2, \dots, m_i] \quad (7)$$

The process of “defuzzify” is defined in Eq. (8) to generate the initial forecasts based on the factor of stock index.

$$\text{Forecast}(t+1) = L_{df}(t) \cdot W_n(t)_T \quad (8)$$

Algorithm 2. The forecasting processes based on the factor of stock trading volume

Step 1: Convert daily stock trading volume into an observation of the technical indicator of volume by using Eq. (3) (Kim et al., 2006; Meyers, 1994).

Step 2: Define and partition the universe of discourse. Define the universe of discourse, $U = [\text{low}, \text{high}]$, which can cover all observations of the technical indicator in the training dataset. And initially partition the universe of discourse into three linguistic intervals, L_i , based on the properties of stock markets. In practical stock markets, the stock index in the future has better chances to go up when the indicator is under low level. Otherwise, the stock index probably downturn when the indicator is under high level.

Step 3: Define fuzzy sets for the technical indicator. Apply Eq. (5) to define fuzzy sets for the technical indicator. Three linguistic values are defined as follows: $L_1 = (\text{over sold})$, $L_2 = (\text{normal})$ and $L_3 = (\text{over bought})$.

Step 4: Fuzzify all observations into three linguistic values. By applying the fuzzy sets, L_1 , L_2 and L_3 , defined in step 4, convert all observations into linguistic values.

Step 5: Transfer linguistic values into market signals. Based on the meaning of the technical indicator, we define a signal transferring function which converts all linguistic values into corresponding market signals. The function is defined in the following equation:

$$M(L_{VR(t)}) = \begin{cases} +1, & \text{if } L_{VR(t)} \text{ is } L_1 \\ 0, & \text{if } L_{VR(t)} \text{ is } L_2 \\ -1, & \text{if } L_{VR(t)} \text{ is } L_3 \end{cases} \quad (9)$$

Step 6: Build an adjustment level for the defined market signals (bull or bear market). Based on the fluctuation ratio of stock index in training dataset, define a reasonable adjustment, a , for the market signals (the adjustment, a , ranges between 0 and 0.07 because of the fluctuation limit in the TAIEX).

Algorithm 3. Dual-factor linear forecasting equation

In this paper, we propose the dual-factor linear forecasting equation (defined in Eq. (10), where $I(t)$ is the stock index at t ; $\text{forecast}(t+1)$ is a forecast for the stock index at $t+1$; b is an adjustment parameter for the initial forecast based on stock index; $L_{VR(t)}$ is a linguistic value of the

indicator; M is a signal transferring function which converts a linguistic value into the corresponding market signal; a is an adjustment parameter for the initial forecast based on the technical analysis indicator), which integrates two forecasting processes dealing with different factors, to produce one conclusive forecast. With two adjustments, a and b , the conclusive forecast can be adapted to meet the patterns of stock index in the past. Besides, using the technical analysis indicator, $VR(t)$, as the second forecasting factor, the forecasts by the proposed model are more reliable than those derived by one-factor forecasting models

$$\begin{aligned} \text{dual_factor_forecast}(t+1) \\ = a * M(L_{VR(t)}) * I(t) + b * (\text{forecast}(t+1) - I(t)) + I(t) \end{aligned} \quad (10)$$

4. Verifications and comparisons

In verification section, we employ two financial datasets as experimental datasets: (1) a five-year period of the TAIEX (Taiwan stock exchange capitalization weighted stock index); and (2) an eleven-year period of the NASDAQ (National Association of Securities Dealers Automated Quotations). Additionally, to measure forecasting performance, we employ the root mean square error (RMSE¹) as performance indicator, and two multiple-factor fuzzy time-series models, Chen's (2000) and Huarng and Yu's (2005), as comparison models.

4.1. Forecasting for TAIEX

In this experiment, a five-year period of TAIEX stock index, from 2000/1/4 to 2004/12/31, is selected to initially test the proposed model. There are two simulating methods for the TAIEX dataset: (1) previous three-year of the dataset, from 2000/1/4 to 2002/12/31, is used for training and the rest, 2003/1/4 to 2004/12/31, is for testing (see Fig. 2); and (2) a half-year moving-window testing approach is used to go through the dataset (the ratio of training to testing is 1:0.5 (year)).

With seven linguistic values and two adjustments (a and b), we produce the forecasting results for the experimental dataset. Figs. 3 and 4 illustrate the forecasts for the testing period using simulating method 1. To demonstrate the forecasting performance of the proposed model, we produce two comparison tables with the reduplicated algorithms of the comparison models which are listed in Table 1 (simulating method 1) and Table 2 (simulating method 2). The comparison tables both show that the proposed method surpasses in accuracy the two comparison models in different simulating methods. Based on the excel-

¹ RMSE = $\sqrt{\sum_{i=1}^n (\text{actual}(i) - \text{forecast}(i))^2 / n}$.

lent performance, a further experiment with a larger scale of stock index data is warranted.

4.2. Forecasting for NASDAQ

To verify the improvement of the proposed model in the initial experiment, a longer period of another financial dataset is used as the second experimental dataset. An eleven-year period of NASDAQ, from 1996/1/4 to 2006/12/31, is selected and a one-year moving-window testing approach is used to go through the whole dataset (the ratio of training to testing is 2:1 (year)). With the reduplicated algorithms of the comparison models, we produce a comparison table which is listed in [Table 3](#). It is obvious that

the proposed model outperforms the [Chen's \(2000\)](#) and [Huang and Yu's \(2005\)](#) models except 1999.

5. Findings and conclusions

In this paper, a new time-series model using dual factors has been proposed to improve the forecasting accuracy of fuzzy time-series models. In the experiments, there are three findings as follows:

- (1) The forecasting factors should be correlated to the future stock index. The proposed model performs the listing models because it uses the effective factors, the stock index and $VR(t)$, which is significant related to the future stock index ([Kim et al., 2006](#); [Ting, 2003](#)) in stock price forecasting, rather than the subjective factors such as the daily maximum and minimum values employed by [Chen's \(2000\)](#) and [Huang and Yu's \(2005\)](#) models.
- (2) The proposed dual-factor linear forecasting equation makes our forecasts more reliable than those derived from conventional models, which only mine fuzzy logical relationships from stock index time-series and ignore the relationship between stock volume and stock index. In practical stock markets, there is no denying that the correlation exists between stock volume and stock price ([Marc & Hiraki, 1999](#); [Hodgson et al., 2006](#); [Shen & Wang, 1998](#); [Wang & Chin, 2004](#)). Based on the excellent performance of the proposed model, we argue that stock volume is one of important factors which are requisite in forecasting.
- (3) With two optimal values, a and b , the dual-factor linear forecasting equation can produce an optimal-adapted prediction to meet the past patterns of stock price and, then, generate a more accurate forecast.

From the experimental results, [Tables 1–3](#), it is apparent that the proposed model performs better than the listing fuzzy time-series models employing multiple subjective factors. Based on this clear evidence, we conclude the research goal has been reached. However, there is room of improvement for the proposed model as follows:

1. Training experimental data sets with different dynamic moving windows and updating the training data sets immediately.
2. Simulating the proposed model to trade in stock market, and sum up the profits of these trades to evaluate the performance of profit making.
3. Employing multiple technical analysis indicators in forecasting processes to improve the dual-factor model.

References

- Chen, S. M. (1996). Forecasting enrollments based on fuzzy time-series. *Fuzzy Sets and Systems*, 81, 311–319.
- Chen, S. M. (2000). Temperature prediction using fuzzy time-series. *IEEE*

- Hodgson, A., Masih, A. M. M., & Masih, R. (2006). Futures trading volume as a determinant of prices in different momentum phases. *International Review of Financial Analysis*, 15, 68–85.
- Huarng, K. (2001a). Effective lengths of intervals to improve forecasting in fuzzy time-series. *Fuzzy Sets and System*, 123(3), 387–394.
- Huarng, K. (2001b). Heuristic models of fuzzy time-series for forecasting. *Fuzzy Sets and System*, 123, 369–386.
- Huarng, K. H., & Yu, H. K. (2005). A type 2 fuzzy time-series model for stock index forecasting. *Physica A*, 353, 445–462.
- Kim, M. J., Min, S. H., & Han, I. G. (2006). An evolutionary approach to the combination of multiple classifiers to predict a stock price index. *Expert Systems with Applications*, 31, 241–247.
- Kitt, R., & Kalda, J. (2005). Scaling analysis of multi-variants intermittent time-series. *Physica A*, 353, 480–492.
- Lee, L. C. H., Liu, A., & Chen, W. S. (2006). Pattern discovery of fuzzy time-series for financial prediction. *IEEE Transactions on Knowledge and Data Engineering*, 18(5), 613–625.
- Lee, L. W., Wang, L. H., Chen, S. M., & Leu, Y. H. (2006). Handling forecasting problems based on two-factor high-order fuzzy time-series. *IEEE Transactions on Fuzzy Systems*, 14(3), 468–477.
- Maddala, G. S. (1992). *Introduction to econometrics* (2nd ed.). New York: Macmillan Publishing Co.
- Marc, B., & Hiraki, T. (1999). Volume and individual security returns on the Tokyo stock exchange. *Pacific-Basin Finance Journal*, 7, 351–370.
- Meyers, A. T. (1994). *The technical analysis course – A winning program for investors & traders*. Chicago, 1994: PROBUS Publishing Co.
- Shen, C. H., & Wang, L. R. (1998). Daily serial correlation, trading volume and price limit: evidence from the Taiwan market. *Pacific-Basin Finance Journal*, 251–273.
- Song, Q., & Chissom, B. S. (1993). Forecasting enrollments with fuzzy time-series – Part 1. *Fuzzy Sets and Systems*, 54, 1–9.
- Song, Q., & Chissom, B. S. (1994). Forecasting enrollments with fuzzy time-series – Part 2. *Fuzzy Sets and Systems*, 62, 1–8.
- Ting, J. J. L. (2003). Causalities of the Taiwan stock market. *Physica A*, 324, 285–295.
- Wang, C. Y., & Chin, S. T. (2004). Profitability of return and volume-based investment strategies in China's stock market. *Pacific-Basin Finance Journal*, 12, 541–564.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338–353.
- Zadeh, L. A. (1975a). The concept of a linguistic variable and its application to approximate reasoning I. *Information Science*, 8, 199–249.
- Zadeh, L. A. (1975b). The concept of a linguistic variable and its application to approximate reasoning II. *Information Science*, 8, 301–357.
- Zadeh, L. A. (1976). The concept of a linguistic variable and its application to approximate reasoning III. *Information Science*, 9, 43–80.