A Knowledge Integration Model for Corporate Dividend Prediction

Jinhwa Kim[a], Chaehwan Won[b] and Jae Kwon Bae[c]

[a, b, c] *School of Business, Sogang University*
#1 Sinsu-dong, Mapo-gu, Seoul 121-742, Korea
Tel: +82-2-705-8860, Fax: +82-2-705-8519, E-mail: {jinhwakim, chwon, baejaekwon}@sogang.ac.kr

Abstract

Dividend is one of essential factors determining the value of a firm. According to the valuation theory in finance, discounted cash flow (DCF) is the most popular and widely used method for the valuation of any asset. Since dividends play a key role in the pricing of a firm value by DCF, it is natural that the accurate prediction of future dividends should be most important work in the valuation. Although the dividend forecasting is of importance in the real world for the purpose of investment and financing decision, it is not easy for us to find good theoretical models which can predict future dividends accurately except Marsh and Merton (1987) model. Thus, if we can develop a better method than Marsh and Merton in the prediction of future dividends, it can contribute significantly to the enhancement of a firm value. Therefore, the most important goal of this study is to develop a better method than Marsh and Merton model by applying artificial intelligence techniques.

Keywords:

Dividend Policy; Marsh and Merton; Artificial Intelligence; Knowledge Integration

Introduction

Previous studies regarding dividend payouts show that dividend policy is irrelevant in all instances regardless of the existence of growth or corporate taxes. It has no effect on shareholder's wealth. Only when personal taxes are introduced do we have a result that dividend payouts matter. For shareholders who pay high taxes on dividends than on capital gains, the preferred dividend payout is zero; they would rather have the company distribute cash payments via the share repurchase mechanism.

Yet, it is well-known fact that corporations do pay dividends in reality. Since there appear to be strong cross-sectional regularities in dividend payout, there may be optimal dividend policy which results from a trade-off between the costs and benefits of paying dividends as Rozeff (1982) suggests. The list of possible costs includes tax advantages of receiving income in the form of dividends rather than capital gains and the cost of raising external capital if dividends are paid out. On the other hand, the possible benefits of dividend payouts are higher perceived corporate value because of the signaling content of dividend, the lower agency costs of external equity, and the ability of dividend payments to help complete markets. In other words, there is the possibility that we can derive the optimal payout policy of an individual firm under some conditions imposed by the financing and investment policy of the firm. As Brealey and Myers (1991) point out, however, the dividend policy of a firm still remains one of the most controversial subjects in the modern finance theory. They argue that three major theories about determination of the optimal dividend policy are competing as follows: First, Hansen et al. (1994) argue that agency cost of dividends is one of the major factors affecting decision making of payouts. According to this hypothesis, dividend payouts can serve as a way to reduce agency costs. By applying dividends equal to the amount of "surplus" cash flow, a firm can reduce management's ability to squander the firm's resources. Since dispersion of ownership among shareholders is a basic measure of agency costs, it would be expected that firms with high dispersed ownership would have high dividends. Second, Bhattacharya (1979), Miller and Rock (1985), and Nissim and Ziv (2001) suggest a 'information content hypothesis' that dividends serve to signal to shareholders the firm's current and future performance. Third, Kim, et al. (1979) propose a 'clientele-effect hypothesis' that those individuals in high tax brackets are likely to prefer either no or low dividends, and vice versa. In addition, Gordon (1959, 1962) and Lintner (1962), Litzenberger and Ramaswami (1979), Black and Scholes (1974), Miller and Scholes (1978, 1982), Hess (1983), Eades et al. (1984), Benartzi et al. (1997), DeAngelo et al. (1996) and Nissim and Ziv (2001), Shiller (1981), Marsh and Merton (1987), Hakansson (1982) and others argue about the relevancy of dividend policy, implying the importance of dividends in the business policy.

In addition, dividend is one of essential factors determining the value of a firm. According to the valuation theory in finance, discounted cash flow (DCF) is the most popular and widely used method for the valuation of any asset. Since dividends play a key role in the pricing of a firm value by DCF, it is natural that the accurate prediction of future dividends should be most important work in the valuation. Although the dividend forecasting is of importance in the real world for the purpose of investment and financing decision, it is not easy for us to find good theoretical models which can predict future dividends accurately except Marsh and Merton (1987) model. Thus, if we can develop a better method than Marsh and Merton (1987) in the prediction of future dividends, it can contribute significantly to the enhancement of a firm value. Therefore, the most important goal of this study is to develop a better method than Marsh and Merton (1987) model by applying artificial intelligence techniques.

Neural networks are considered the most powerful classifier for their low classification error rates and robustness to noise. But neural networks have two obvious shortcomings when applied to data mining problems (Anandarajan et al. 2001; Atiya 2001, Charalambous et al. 2000; Craven and Shavlik 1997; Lu et al. 1996; Pendharkar 2005). The first is that neural networks require long time to train the huge amount of data of large databases. Secondly, neural networks lack explanation facilities for their knowledge. The knowledge of neural networks is buried in their structures and weights. It is often difficult to extract rules from a trained neural network (Li and Wang 2004).

A decision tree is a non-linear discrimination method, which uses a set of independent variables to split a sample into progressively smaller subgroups. The procedure is iterative at each branch in the tree; it selects the independent variable that has the strongest association with the dependent variable according to a specific criterion (Curram and Mingers 1994; Kass 1980; Michael and Gordon 1997; Quinlan 1986, 1993). Classification and Regression Tree (CART) is a recursive partitioning method to be used both for regression and classification. CART is constructed by splitting subsets of the data set using all predictor variables to create two child nodes repeatedly, beginning with the entire data set. The best predictor is chosen using a variety of impurity or diversity measures. The goal is to produce subsets of the data which are as homogeneous as possible with respect to the target variable (Breiman et al. 1984).

A regression tree is a decision tree in which the target variable takes its values from a continuous domain (numeric). For each leaf, the regression tree associates the mean value and the standard deviation of the target variable. On the other hand the classification tree is a decision tree in which the target variable takes its values (e.g., class) from a discrete domain.

The vast majority of rule induction algorithms have a bias that favors the discovery of large disjuncts, rather than small disjuncts. This preference is due to the belief that it is better to capture generalizations rather than specializations in the training set, since the latter are unlikely to be valid in the test set.

To summarize the primary goal of this paper, we need to emphasize the importance of future dividends, because dividends are essential parts of the valuation of a company via DCF method. It is a basic principle in business that knowing the exact value of a firm is just a starting point of an investment decision in the equities of the firm, mergers and acquisitions decision, and many other important business decisions. Therefore, the accurate prediction of future dividends is of great importance, and hence our major goal in this study is to develop a better dividend forecasting method than existing financial models by utilizing knowledge integration approaches. The effectiveness of our approach was verified by the experiments comparing with Marsh and Merton model, Neural Networks, and CART approaches.

The remainder of this paper is organized as follows. Section 2 gives some research background about dividend prediction model. Knowledge Integration (KI) model is described and the algorithms to generate classification rules from decision table are proposed in Section 3. Section 4 explains the processes of data collections and variables selections. Some experimental results are presented and analyzed in Section 5, and finally concluding remarks are given in Section 6.

Dividend Prediction Model

In this section, we briefly describe a Marsh and Merton (1987)'s dividend prediction model which is one of the most widely used ones in both academic circles and business fields. By applying general equilibrium framework, Marsh and Merton succeeded in deriving an equilibrium dividend prediction model which seems to significantly outperform any other prediction models. Since the model uses just stock price and dividend in the previous periods to forecast next period's dividend, it is much easier to apply than any other models that use many accounting related variables. In addition, since it uses only market values instead of book values, the estimated dividends can be directly used for the valuation of a firm in the DCF framework and for the decision making of future dividend policy. Therefore, most of the researchers and experts in finance areas use the model to predict future equilibrium dividends. Although the model is one of the best theoretical models, the accuracy of prediction is not so satisfactory to the decision makers in the business field.

Marsh and Merton (1987) derived the following structural regression model from its original equilibrium model to estimate expected dividends:

$$\ln\left(\frac{D_{t+1}}{D_t}+\frac{D_t}{P_{t-1}}\right)=a_0+a_1\ln\left(\frac{P_t+D_t}{P_{t-1}}\right)+a_2\ln\left(\frac{D_t}{P_{t-1}}\right)+\varepsilon_{t+1}, \quad (1)$$

where D_t and P_t are dividend and stock price in period t, respectively, ε_{t+1} is the disturbance term in period (t+1), and 'ln' denotes the natural logarithm. After we determine regression parameters, a_0, a_1, and a_2 and rearrange the equation for D_{t+1}, we can obtain next period's expected dividend, D_{t+1}, from the equation (1) by using past dividend, D_t, and past stock prices, P_{t-1} and P_t. That is, the predicted dividend can be obtained using the following equation:

$$D_{t+1}^{\mathrm{Pr}edicted} = D_t\left[e^{a_0}\cdot\left(\frac{P_t+D_t}{P_{t-1}}\right)^{a_1}\cdot\left(\frac{D_t}{P_{t-1}}\right)^{a_2}-\frac{D_t}{P_{t-1}}\right]. \quad (2)$$

Equation (2) is the Marsh and Merton (1987) model that will be used as a benchmark model in this study to compare with our new model.

Knowledge Integration Model Development

The concept of knowledge integration is illustrated in Figure 1. This figure shows a part of a decision tree induced from the dividend data sets.

Knowledge integration is a methodology in accordance with rules derived from CART algorithm with all data sets including missing values. As shown in Figure 1, training data is classified into two categories of complete data, data with one missing variable. There are 4 combinations in type of records where one variable has missing value in P_{t-1}, P_t, D_t, or complete data set.

The method discovers rules in two training phases. In the first phase it runs CART, a well-known decision tree induction algorithm. The induced, pruned tree is transformed into a set of rules.

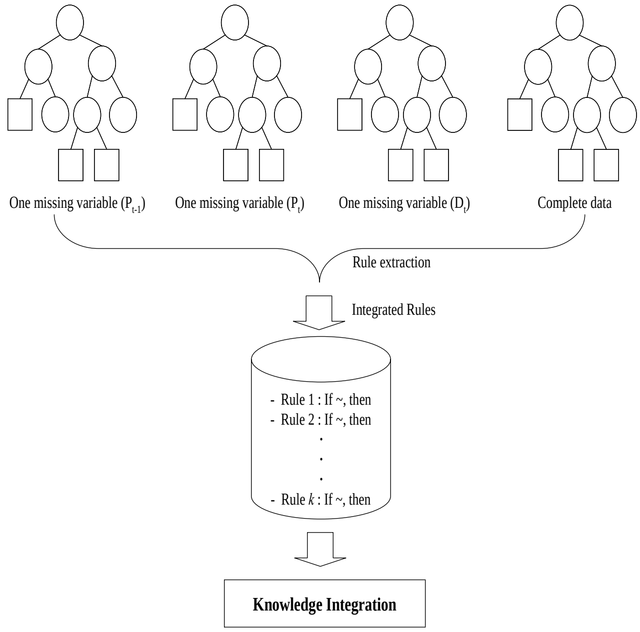

Figure 1 - The procedures of knowledge integration model

CART is developed from each of 4 data sets. Each of these 4 decision trees are transformed into *k* different rule sets. These 39 rules are consolidated into one. By combining separate knowledge in the form of *If-then* rules induced from different data set, it builds one meta model. Each rule in this model serves as an agent predicting dividends. Algorithm and interpretation of knowledge integration model is represented in equation (3) and Figure 2.

Ⅰ. P_{t-1}, P_t, D_t, : Independent variable, t = 1999
Ⅱ. $D_{i,\,t+1}^{\,Actual}$: Dependent variable, i = 1, 2, …185
Ⅲ. R_k : Knowledge in the form of *If-then* rules induced from the dividend data set
Ⅳ. $\sum R_k$: Cumulative rule sets
Ⅴ. $D_{i,\,t+1}^{\,Predicted}$: Predicted dividend (= $\sum R_k$/the number of decision tree)
Ⅵ. $\varepsilon_{i,\,t+1} = D_{i,\,t+1}^{\,Actual} - D_{i,\,t+1}^{\,Predicted}$ (3)

ID	P_{t-1}	P_t	D_t	$D_{t+1}(Act)$	Rule 1 (R_1)	Rule 2 (R_2)	Rule 3 (R_3)	Rule 4 (R_4)	Rule 5 (R_5)
1	37000	19900	750	576	0	0	1156.154	0	0
2	14350	9420	1500	1000	0	0	0	0	0
3	0	15550	1000	1200	0	0	0	937.3333	0
4	5530	6650	0	100	485.9238	0	0	0	0
5	41000	6000	1000	1000	0	0	0	0	0
6	9550	10400	1000	600	0	0	0	937.3333	0
7	3800	11600	800	800	0	752.2647	0	0	0
8	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
185	12000	11850	700	600	0	752.2647	0	0	0

Rule 6 (R_6)	Rule 7 (R_7)	Rule 8 (R_8)	Rule 9 (R_9)	…	Rule 39 (R_k)	ΣR_k	$D_{t+1}(Pred)$	ε_{t+1}
0	0	0	0	…	0	3300.616	825.154	249.154
0	1690	0	0	…	0	4480	1120	120
0	0	0	0	…	0	4801.332	1200.333	0.333
0	0	0	135	…	0	1003.695	250.9238	150.9238
1303.33	0	0	0	…	0	2553.332	638.333	38.333
0	0	0	0	…	0	4109.332	1027.333	427.333
0	0	0	0	…	0	3329.059	832.2647	32.2647
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	…	0	2921.059	730.2647	130.2647

Figure 2 - The structure of knowledge integration model

Data Collection

In this study, we consider all available companies listed in the Korea Exchange market for the periods from 1980 to 2000. Although the number of listed companies varies a little over time, the average number was about 680 during those periods, and we obtained necessary data, such as dividends and stock prices, from the database *KIS-SMAT* available from Korea Investors Service, Inc. which is a Moody's Korean affiliate. To predict the dividend of year 2000 for each company, i.e., t+1=2000 in equations (1), (2) and (3), we use data from 1980 to 1999. Since the database contains many missing data in dividends, we use 137 companies which have more than 15 years of dividends records in order to estimate regression parameters in equation (1).

Each data set is split into two subsets, a training set and a validation (holdout) set. The training data set is used to train the prediction models. The validation data set is used to test the model's prediction performance with the data which have not been used in developing the classification models. For each set of data set, a training subset and validation subset, consisting of 73%(500/685) and 27%(185/685) of the data respectively, are randomly selected. We replicate ten times (Set 1 to 10) of data set selection, estimation and testing process to reduce the impact of random variation in data set composition (Weiss and Kulikowski 1991).

Incomplete data sets are divided into two cases. One is a data set including independent variable with missing value. The other is a data set including dependent variable with missing value. This paper uses a data set including independent variables with missing values.

Table 1 - The proportion of missing data

Missing Variables	Missing Ratio
P_{t-1}	48/685 (7.01%)
P_t	22/685 (3.21%)
D_t	92/685 (13.43%)
Complete data	523/685 (76.35%)

Experiment Results

Marsh and Merton (1987) model

Before we apply the Marsh and Merton (1987) model to predict future dividends for the sample firms, we tested the validity of the model in Korean stock markets. Using the 137 sample companies that have more than 15 dividend records during the period 1980-2000, we run multiple regressions in equation (1) and get the results as in Table 2. The major results can be summarized as follows: First, we tested the statistical significance of regression parameters, a_0, a_1, a_2, in equation (1) and we found that they were all significant at the 1% level of significance as we can see from the table. In addition, the signs of parameters are exactly consistent with the results by Marsh and Merton (1987). Second, in order to test whether or not the Marsh and Merton (1987) model can be used as a prediction model, we use the null hypothesis that the mean of actual dividends is equal to the mean of the predicted dividends by Marsh and Merton (1987). From the data, we find that we can not reject the null hypothesis, because the value of t-statistic is 0.4543 and its p-value is 0.6499. These two results strongly support that we can use the Marsh and Merton (1987) model in Korean capital markets for the purpose of predicting future dividends.

Since Marsh and Merton (1987) model was proved to be a good prediction model, we employ it to predict dividends for the year of 2000 using equation (2). We provide the prediction results in Table 4. For the prediction, we sampled 185 companies and from these samples the values of regression parameters were 0.0160, 0.2297 and -0.0309, respectively.

Knowledge Integration model

Each rule is presented in Table 3. In Table 3, since the number of the terminal nodes is 8, 8 classification rules are generated. D_t variable plays the most influential role in predicting dividends.

Table 2 - Estimation results for Marsh and Merton Dividend Model

Parameter	Mean[a]	Mode	Std. Dev.	Min.	Max.	M&M[b]
a_0	**-1.0648*****	-1.0128	1.1216	-4.4644	4.2129	**-0.1010**
a_1	**0.2656*****	0.2722	0.4127	-1.1825	1.4038	**0.4370**
a_2	**-0.2906*****	-0.2900	0.2763	-1.0888	0.9555	**-0.0420**
R^2 (%)	19.88	17.17	13.61	0.37	75.52	47.00

Note: a) *** means 'significant' at 1% significance level.
b) M&M denotes the results by 'Marsh and Merton (1987)'.

Table 3 - An example of KI-complete model rules from data set 5

Rule	Description	Instance
1	If D_t is less than 1850, D_t is less than 950 and D_t is less than 560, predicted dividend is 485.9238.	105/400 (26.25%)
2	If D_t is less than 1850, D_t is less than 950, D_t is equal to or more than 560 and P_t is less than 18750, predicted dividend is 752.2647.	68/400 (17%)
3	If D_t is less than 1850, D_t is less than 950, D_t is equal to or more than 560 and P_t is equal to or more than 18750, predicted dividend is 1156.1538.	13/400 (3.25%)
4	If D_t is less than 1850, D_t is equal to or more than 950, D_t is less than 1450, P_{t-1} is less than 31950 and D_t is less than 1081, predicted dividend is 937.3333.	69/400 (17.25%)
5	If D_t is less than 1850, D_t is equal to or more than 950, D_t is less than 1450, P_{t-1} is less than 31950 and D_t is equal to or more than 1081, predicted dividend is 1154.08.	50/400 (12.50%)
6	If D_t is less than 1850, D_t is equal to or more than 950, D_t is less than 1450 and P_{t-1} is equal to or more than 31950, predicted dividend is 1303.3333.	15/400 (3.75%)
7	If D_t is less than 1850, D_t is equal to or more than 950 and D_t is equal to or more than 1450, predicted dividend is 1690.	40/400 (10%)
8	If D_t is equal to or more than 1850, predicted dividend is 2959.5.	40/400 (10%)

Comparison experiments

Table 4 and Figure 3 compare the prediction performances of MM, NN, CART, and KI model from ten-fold cross validations. The result of this cross-validation procedure is the average accuracy rate in the test set over the ten iterations.

Among these models, KI model has the highest level of average accuracy (12.65%) with given data sets, followed by NN (10.59%), MM (10.16%), and CART (7.51%) next at the 5% tolerance level. At the 10% tolerance level, KI model has the highest level of average accuracy (22.65%) with given data sets, followed by NN (19.30%), MM (18.43%), and CART (15.84%) next. At the 20% tolerance level, KI model has the highest level of average accuracy (36.27%) with given data sets, followed by CART (33.78%), MM (33.57%), and NN (32.43%) next. At the 30% tolerance level, KI model has the highest level of average accuracy (48.97%) with given data sets, followed by CART (46.86%), NN (45.24%), and MM (44.76%) next. At the 40% tolerance level, KI model has the highest level of average accuracy (58.05%) with given data sets, followed by CART (56.43%), NN (54.81%), and MM (53.14%) next. At the 50% tolerance level, KI model has the highest level of average accuracy (68.11%) with given data sets, followed by CART (64.59%), NN (61.89%), and MM (60.38%) next.

Table 4 - Comparison of prediction models*

Tolerance	MM	NN	CART	KI
1%	1.89	2.11	2.43	2.27
5%	10.16	10.59	7.51	12.65
10%	18.43	19.30	15.84	22.65
15%	27.19	26.27	24.22	29.14
20%	33.57	32.43	33.78	36.27
25%	38.97	39.57	40.70	44.38
30%	44.76	45.24	46.86	48.97
35%	49.19	50.38	52.97	54.70
40%	53.14	54.81	56.43	58.05
45%	57.19	58.59	61.03	62.76
50%	60.38	61.89	64.59	68.11

Note: * Average Accuracy Rate (%).
MM: Marsh and Merton, NN: Neural Networks
CART: Classification and Regression Tree
KI: Knowledge Integration.

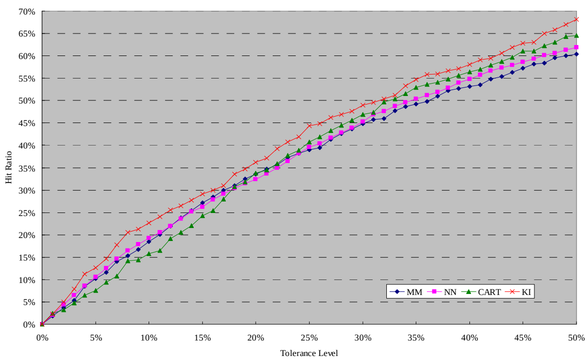

Figure 3 - Average accuracy of prediction models

Conclusions

Mining classification rules is an important task of data mining. In this paper from a new perspective we have presented a knowledge integration model integrating cumulative rule sets to mine classification rules from dividend data sets.

To summarize the primary goal of this paper, we need to emphasize the importance of future dividends, because dividends are essential parts of the valuation of a company via DCF method. It is a basic principle in business that knowing the exact value of a firm is just a starting point of an investment decision in the equities of the firm, mergers and acquisitions decision, and many other important business decisions. Therefore, the accurate prediction of future dividends is of great importance, and hence our major goal in this study is to develop a better dividend forecasting method than existing financial models by utilizing knowledge integration approaches.

Two different techniques were also applied to the same data sets and used as benchmarks: Classification and Regression Trees, known as CART, which represented a well-established statistical method, and back propagation Neural Networks which represented the (currently large) family of neural network algorithms.

Before we apply the Marsh and Merton (1987) model to predict future dividends for the sample firms, we tested the validity of the model in Korean stock markets. The major results can be summarized as follows: First, we tested the statistical significance of regression parameters, a_0, a_1, a_2, in equation (1) and we found that they were all significant at the 1% level of significance as we can see from the table. In addition, the signs of parameters are exactly consistent with the results by Marsh and Merton (1987). Second, in order to test whether or not the Marsh and Merton (1987) model can be used as a prediction model, we use the null hypothesis that the mean of actual dividends is equal to the mean of the predicted dividends by Marsh and Merton (1987). From the data, we find that we can not reject the null hypothesis. These two results strongly support that we can use the Marsh and Merton (1987) model in Korean capital markets for the purpose of predicting future dividends.

The experiments of this study shows that KI model cumulating rules from missing data sets can improve overall performance as it can reduce error-term and increase R-square. This architecture of this KI model is also called an ensemble approaches. The basic idea of this is that collaboration of many experts produces better performance than any single expert. Another advantage of KI model is that it needs extracted information only in the form of rule, not the raw data. In case of stream data, the data is created at real time. The transaction data from stock market, for an example, is created every second. Saving only extracted information from this real-time data set can reduce storage for the data.

The KI model suggested in this study needs to be refined for better prediction. Firstly, it has a problem of redundant rules. Redundant rules are considered to be relatively important rules compared to other rules. Different weights can be assigned to these rules following the degrees of redundancy. Secondly, there can be inconsistencies among rules. These inconsistent rules can be deleted in a simple way. There, however, can be better way such as a voting algorithm to solve this inconsistency problem. Thirdly, there are problems of general rules and specific rules. Some rules can be subset of other rules.

Despite the many findings from this study, it has some limitations. Firstly, the results from the study should be generalized. It would be better to investigate other data sets in order to generalize the results of this study. Secondly, integrating with other rule generation techniques, such as CHAID and genetic algorithm, is also an important issue in future studies.

References

[1] Anandarajan, M., Lee, P. and Anandarajan, A. (2001) 'Bankruptcy Prediction of Financially Stressed Firms: An Examination of the Predictive Accuracy of Artificially Neural Networks', *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 10, No. 2, pp.69-81.

[2] Atiya, A. (2001) 'Bankruptcy prediction for credit risk using neural networks: a survey and new results', *IEEE Transactions on Neural Networks*, Vol. 12, No. 4, pp.929-935.

[3] Benartzi, S., Michaely, R., and Thaler, R., (1997) 'Do changes in dividends signal the future or the past?', *Journal of Finance*, Vol. 52, No. 3, pp.1007-1034.

[4] Bhattacharya, S. (1979) 'Imperfect information, dividend policy, and 'The bird in the hand' fallacy', *Bell Journal of Economics*, Vol. 10, No. 1, pp.259-270.

[5] Black, F. and Scholes, M.S. (1974) 'The effects of dividend yield and dividend policy on common stock prices and returns', *Journal of Financial Economics*, Vol. 1, No. 1, pp.1-22.

[6] Brealey, R.A. and Myers, S.C. (1991). *Principles of Corporate Finance*, 4th ed., NY: McGraw-Hill.

[7] Breiman L., Friedman, J., Olshen, R., and Stone, J. (1984). *Classification and Regression Trees*, Belmont, California: Wadsworth Inc.

[8] Charalambous, C., Chartious, A. and Kaourou, F. (2000) 'Comparative Analysis of Artificial Neural Network Models: Application in Bankruptcy Prediction', *Annals of Operations Research*, Vol. 99, pp.403-425.

[9] Craven, M.W. and Shavlik, J. (1997) 'Using neural networks for data mining', *Future Generation Computer Systems*, Vol.13, No.2/3, pp.211-229.

[10] Curram, S.P. and Mingers, J. (1994) 'Neural Networks, Decision Tree Induction and Discriminant Analysis: an Empirical Comparison', *Journal of the Operational Research Society*, Vol. 45, No. 4, pp.440-450.

[11] DeAngelo, H., DeAngelo, L. and Skinner, D.J. (1996) 'Reversal of fortune: dividend signaling and the disappearance of sustained earnings growth', *Journal of Financial Economics*, Vol. 40, No. 3, pp.341-371.

[12] Eades, K.M., Hess, P.J. and Kim, E.H. (1984). 'On interpreting security returns during the ex-dividend period', *Journal of Financial Economics*, Vol. 13, No.1, pp.3-34.

[13] Gordon, M.J. (1959) 'Dividends, earnings and stock prices', *Review of Economics and Statistics*, Vol. 41, No. 2, pp.99-105.

[14] Gordon, M.J. (1962) 'The savings investment and valuation of a corporation', *Review of Economics and Statistics*, Vol. 44, No. 1, pp.37-51.

[15] Hakansson, N. (1982) 'To pay or not to pay dividends', *Journal of Finance*, Vol. 37, No. 2, pp.415-428.

[16] Hansen, R., Kumar, R. and Shome, D.K. (1994) 'Dividend policy and corporate monitoring:
Evidence from the regulated electric utility industry', *Financial Management*, Vol. 23, No. 1, pp.16-22.

[17] Hess, P.J. (1983) 'Tests for tax effects in the pricing of financial asserts', *Journal of Business*, Vol. 56, No. 4, pp.537-554.

[18] Kass, G.V. (1980) 'An exploratory technique for investigating large quantities of categorical data', *Applied Statistics*, Vol. 29, pp.119-127.

[19] Kim, E.H., Lewellen, W. and McConnell, J. (1979) 'Financial leverage clienteles: Theory and evidence', *Journal of Financial Economics*, Vol. 7, No. 1, pp.83-110.

[20] Li, R., and Wang, Z. (2004) 'Mining classification rules using rough sets and neural networks', *European Journal of Operational Research*, Vol. 157, pp. 439-448.

[21] Lintner, J. (1962) 'Dividends, earnings, leverage, stock prices and the supply of capital to corporations', *Review of Economics and Statistics*, Vol. 44, No. 3, pp.243-269.

[22] Litzenberger, R.H. and Ramaswamy, K. (1979) 'The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence', *Journal of Financial Economics*, Vol. 7, No. 2, pp.163-195.

[23] Lu, H., Setiono, R., and Liu, H. (1996) 'Effective data mining using neural networks', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp.957-961.

[24] Marsh, T.A. and Merton, R.C. (1987) 'Dividend behavior for the aggregate stock market', *Journal of Business*, Vol. 60, No. 1, pp.1-40.

[25] Michael, J.A., and Gordon, S.L.,(1997). *Data mining techniques: for marketing, sales, and customer support*, 1st ed., NY: Wiley.

[26] Miller, M.H. and Modigliani, F. (1961) 'Dividend policy, growth and the valuation of shares', *Journal of Business*, Vol. 34, No. 4, pp.411-433.

[27] Miller, M.H. and Rock, K. (1985) 'Dividend policy under asymmetric information', *Journal of Finance*, Vol. 40, No. 4, pp.1031-1051.

[28] Miller, M.H. and Scholes, M.S. (1982) 'Dividends and taxes: Some empirical evidences', *Journal of Political Economy*, Vol. 90, No. 6, pp.1118-1141.

[29] Nissim, D. and Ziv, A. (2001) 'Dividend changes and future profitability', *Journal of Finance*, Vol. 56, No. 6, pp.2111-2133.

[30] Pendharkar, P.C. (2005) 'A Threshold-varying Artificial Neural Network Approach for Classification and its Application to Bankruptcy Prediction Problem', *Computers and Operations Research*, Vol. 32, No. 10, pp.2561-2582.

[31] Quinlan, J.R. (1986) 'Induction of Decision Trees', *Machine Learning*, Vol. 1, pp.81-106.

[32] Quinlan, J.R. (1993) *C4.5: Programs for Machine Learning*, San Mateo, CA: Morgan Kaufmann.

[33] Rozeff, M., (1982) 'Growth, beta and agency costs as determinants of dividend payout ratios', *Journal of Financial Research*, Vol. 5, No. 3, pp.249-259.

[34] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) "Learning internal representation by error propagation," in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, MA.

[35] Shiller, R.J. (1981) 'Do stock prices move too much to be justified by subsequent changes in dividends?', *American Economic Review*, Vol. 71, No. 3, pp.421-436.

[36] Weiss, S. and Kulikowski, C. (1991) *Computer Systems That Learn*, Morgan Kaufmann.