Timing of Resources Exploration in the Behavior of Firm-Comparative Approaches on ARMAX, BPNN, ANFIS, and ASVR

Bao Rong Chang
Department of Computer Science and
Information Engineering
National Taitung University, Taiwan
brchang@cc.nttu.edu.tw

Hsiu Fen Tsai
Department of International Business
Shu Te University, Taiwan
soenfen@mail.stu.edu.tw

ABSTRACT

We have insight into the importance of resource exploration derived from the quest for sustaining competitive advantage as well as the growth of the firm, which are well-explicated in the resources-based view. However, we really do not know when the firm will seriously commit to this kind of activities. Therefore, this study proposes comparative approaches using auto-regressive moving-average regression (ARMAX), back-propagation neural network (BPNN), adaptive neuro-fuzzy inference system (ANFIS), or adaptive support vector regression (ASVR) to constitute the relationship among five indicators, the growth rate of long-term investment, the firm size, the return on total asset, the return on common equity, and the return on sales. In such a way, the methods we build can explain the timing of resources exploration in the behavior of firm. Meanwhile, the performance between these methods is compared quantitatively.

Keywords: resources exploration, auto-regressive moving-average regression, back-propagation neural network, adaptive neuro-fuzzy inference system, adaptive support vector regression.

1. INTRODUCTION

When we think about the firm as a collective of resources [1], it drives the different aspects of research directions to answer two fundamental strategic questions: the sources of competitive advantage and the growth trajectory of firm. There is fairly general agreement that the accumulation of heterogeneous resources can explain the success of firm for a period of time [2][3][4][5][6][7]; it also shapes the path of firm's growth [1][8]. However, resources or capabilities, like product, have life cycles [9]. Thus, researchers always remind us the importance of exploring new resources due to the pressure of external changing environment [1][8][10]. We know the importance of resource exploration derived from the quest for sustaining competitive advantage as well as the growth of the firm that are well-explicated in the resources-based view. It is worth understanding that the idea of balance between exploration and exploitation will be solely achieved under the assumption of calculated rationality. However, we in fact do not know when the firm takes it into account and commits itself to the exploring activities. In each occasion of decision-making, decision makers are constrained by bounded rationality [11][12][13] which is raised from two facts: (a) managers have limited absorptive capacity, and (b) managers acquired finite information subjected to the external changing environment. Therefore, the managers might miss or postpone the exact timing of exploring activities because of their inability on controlling the future uncertainty under the situation of limited rationality. Accordingly, the reinforcement of the precedent tendency mentioned above will be stressed by the conservative personality of managers. All of these will leads the managers to persist on the exploiting the existing resources rather than exploring new ones.

So, we concern the timing problem, namely when the risky attitude of managers will be shifted from risk-avoiding exploiting activities to relative risk-taking exploring activities. In the basis of the prospect theory [14], we argue that the turning point will be triggered by the negative prospects. That is, when the firm is framed by positive performance, it will incline the managers to utilize the existing resources and neglect the need of exploring new ones. On the other hand, when the firm suffers from loss or decline in performance, it will reverse the risky attitude of decision makers to approach risk-taking considerably that will ignite more exploring activities. And then, we can observe that the trajectory of the growth of firm is emerging with the exploration and the following exploitation and so on [1][15]. In the mean time, we also proposed that large firm holds much more resources than small one [16] in this case that leads to the large firm's 'value function' [14] is flatter than small firm. Therefore, large firm has revealed low risk-aversion so as to potentially proceed to higher exploring activities in positive frames; in contrast, small firm with the emergence of high risk-seeking in negative frames will then undertake more exploring activities.

There are five indicators that are the growth rate of long-term investment, the firm size, the return on total asset, the return on common equity, and the

return on sales. The relationship among these indicators indeed can be used to analyze the timing of resources exploration in the behavior of firm. Several quantitative methods, such linear time series models with single-output, multi-input structure as AR, MA, ARX, ARMA, and ARMAX [17][18], are applicable for modeling the dynamics of the interaction between five indicators mentioned above. Once the trained model has achieved, it interprets the timing of resources exploration in the behavior of firm based on the coefficients of explanatory variables. However, a trained ARMAX model cannot reach optimal one with respect to the performance criterion of mean absolute percent error. Thus, three more nonlinear models, back-propagation neural network (BPNN) [19][20], adaptive neuro-fuzzy inference system (ANFIS) [21][22], and adaptive support vector regression (ASVR) [23][24], are also provided in this study so that the performance of each model can be compared quantitatively.

2. METHODS

Several linear and nonlinear models are described in this section. Three major models: autoregressive moving-average regression, back-propagation neural network, and segmented adaptive support vector regression are introduced here.

2.1. Auto-regressive moving-average regression (ARMAX)

A general polynomial black-box model [19] allows a flexible model description of the response variables $y_j(t); j = 1,2,...,ny$, the explanatory variables $u_i(t); i = 1,2,...,nu$, and the disturbances $e_l(t); l = 1,2,...,ne$. The most typical single-output, multi-input system is written as follows.

$$A(q)y(t) = \frac{B_1(q)}{F_1(q)}u_1(t-nk_1)+...+\frac{B_{nu}(q)}{F_{nu}(q)}u_{nu}(t-nk_{nu})+\frac{C(q)}{D(q)}e(t) \quad (1)$$

where q is the shift operator, $nk_1,...,nk_{nu}$ are delays for input signals, as well as A, B_1,..., B_{nu}, C, D, and F_1,..., F_{nu} are polynomials in the lag operator q^{-1}. Before estimating the appropriate parameters for the model, both the orders of polynomials A, B_1,..., B_{nu}, C, D, and F_1,..., F_{nu} and the delays of input signals $nk_1,...,nk_{nu}$ must be determined to make sure whether the selected structure is the best. This model can perform simulation, forecasting, and parameter estimation of univariate time series in the presence of auto-regressive moving-average regression (ARMAX) [19], especially in financial time series applications like asset return problem. ARMAX model is described below.

$$A(q)y(t) = B_1(q)u_1(t-nk_1)+...+B_{nu}(q)u_{nu}(t-nk_{nu})+C(q)e(t) \quad (2)$$

where polynomials A, B_1,..., B_{nu}, C indicated by the lag operator q^{-1} are expressed below.

$$A(q) = 1 + a_1 q^{-1} + ... + a_{na}q^{-na} \quad (3)$$

$$B_1(q) = b_{11} + b_{21}q^{-1} + ... + b_{nb1}q^{-nb1+1} \quad (4)$$

$$...$$

$$B_{nu}(q) = b_{1nu} + b_{2nu}q^{-1} + ... + b_{nbnu}q^{-nbnu+1}$$

$$C(q) = 1 + c_1 q^{-1} + ... + c_{nc}q^{-nc} \quad (5)$$

In this study, five indicators suitably plug in ARMAX structure for modeling their relationship in order to analyze the timing of resources exploration in the behavior of firm. More precisely, the growth rate of long-term investment is designed as output $y(t)$, as well as the others, the firm size, the return on total asset, the return on common equity, and the return on sales, are assigned to be input signals $u_1(t)$, $u_2(t)$, $u_3(t)$, and $u_4(t)$, respectively. To select the best structure that has the smallest loss function in the validation data set is a good way to choose the delays for input signals $nk_1,...,nk_{nu}$ in Eq. (2). Alternatively, a good idea is to select the structure with the best fit determined by the number of parameters that will be used in modeling. In such a way, the appropriate order of polynomial in lag operator q^{-1} for $A(q)$, $B_1(q)$,..., $B_{nu}(q)$, $C(q)$ on Eq. (3),(4),and (5) can be obtained. After the best structure of ARMAX model is selected, it will estimate the parameters of the designated structure based on the training and validation data sets.

2.2. Back-propagation neural network (BPNN)

A well-known intelligent computing machine, three-layer back-propagation neural net (BPNN) [20] is used for modeling nonlinear structure to analyze the timing of resources exploration in the behavior of firm. For a three-layer BPNN, a structure of $4\times10\times1$ multilayer-perceptron is used that the input layer has 4 input neurons to catch the input patterns, the hidden layer has 10 neurons to propagate the intermediate signals, and the output layer has 1 neuron to display the computed results (output $y(t)$) as shown in Fig. 1. We arrange the input pattern in the following four input signals: the firm size $u_1(t)$, the return on total asset $u_2(t)$, the return on common equity $u_3(t)$, and the return on sales $u_4(t)$. Only an appropriate growth rate of long-term investment $y(t)$ is designed as output. For more training assignments in this three-layer BPNN, the tangent-sigmoid transfer function is applied as the activations in the hidden layer, the symmetric saturating linear transfer function is employed to the output layer as the activations, and Bayesian regulation derived from Levenberg-Marquardt training method is used as the learning algorithm for three-layer BPNN. Moreover,

2

training epochs is assigned to 1000 as well as training stop criteria is set to be 10^{-6} in this case.

2.3. Adaptive neuro-fuzzy inference system (ANFIS)

The acronym ANFIS derives its name from adaptive neuro-fuzzy inference system [21]. Using a given input/output data set, the ANFIS constructs a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a backpropagation gradient descent algorithm alone, or in combination with a least squares type of method [22]. This allows your fuzzy systems to learn from the data they are modeling. It actually is of Sugeno-type fuzzy inference systems in which ANFIS uses a hybrid learning algorithm to identify parameters of Sugeno-type fuzzy inference systems. ANFIS can also be invoked using an optional argument for model validation. The type of model validation that takes place with this option is a checking for model overfitting, and the argument is a data set called the checking data set. We may think of a learning data set as a training data set plus another checking data set. ANFIS has been shown to be very useful for modeling nonlinear system concerning the complex dynamic behavior [21], for instances, the applications of non-periodic short-term forecasting. This is because the feature of ANFIS with inherent distributive architecture and the efficient learning algorithms for adapting system's parameters. However, ANFIS must follow four constraints: (a) Each output is specified to be first or zeroth order Sugeno-type function. (b) All output membership functions must be a single output with the same type and either being linear or constant. Defuzzification employs weighted average method (c) the number of output membership functions must be equal to the number of rules. (d) Unity weight for each rule is applied.

A typical rule in a Sugeno fuzzy model has the form

$$If\ Input1 = x\ and\ Input2 = y,\ then\ Output\ z = ax+by+c \quad (6)$$

For a zero-order Sugeno model, the output level z is a constant $(a = b = 0)$.

The output level z_i of each rule is weighted by the firing strength w_i of the rule. For example, for an AND rule with Input 1 $= x$ and Input 2 $= y$, the firing strength is

$$w_i = AndMethod(F_1(x), F_2(y)) \quad (7)$$

where $F_{1,2}(.)$ are the membership functions for Inputs 1 and 2. The final output of the system is the weighted average of all rule outputs, computed as

$$Final\ Output = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \quad (8)$$

A Sugeno rule operates [21] like the designated diagram as shown in Fig. 2.

2.4. Adaptive support vector regression (ASVR)

Support Vector Machines along with neural networks as one of the standard tools for machine learning and data mining [23][24]. Initially developed for solving classification problems, SV technology can also be successfully applied in regression, i.e. functional approximation, problems. Unlike pattern recognition problems, where the desired outputs are discrete values like Booleans, here there are real-valued functions [25]. We consider approximating functions solved by support vector regression (SVR) as the form of

$$f(x,w) = \sum_{i=1}^{l} w_i \phi(x) , \quad (9)$$

where $\phi(x)$ are denoted by features. In order to introduce all relevant and necessary concept of SV regression in a gradual way, a simple linear regression is considered first.

$$f(x,w) = w^T \phi(x) + b \quad (10)$$

Furthermore, Vapnik introduced a general type of loss function, namely, error, the linear loss function with ε-insensitivity zone:

$$|y - f(x,w)|_\varepsilon = \begin{cases} 0 & if\ |y - f(x,w)| \le \varepsilon, \\ |y - f(x,w)| - \varepsilon & otherwise \end{cases} \quad (11)$$

A new empirical risk is introduced for performing support vector regression.

$$R_{emp}(w,b)_\varepsilon = \frac{1}{l} \sum_{i=1}^{l} |y_i - w^T \phi(x_i) - b|_\varepsilon \quad (12)$$

According to the learning theory of SVMs, the objective is to minimize the empirical risk and norm-squared of weight vector simultaneously. Thus, estimate a linear regression hyperplane $f(x,w) = w^T \phi(x) + b$ by minimizing

$$R(w,\xi,\xi^*) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l} \xi_i + \sum_{i=1}^{l} \xi_i^*\right), \quad (13)$$

under constrains

$$y_i - w^T \phi(x_i) - b \le \varepsilon + \xi_i,\ i = 1,...,l \quad (14)$$

$$w^T \phi(x_i) + b - y_i \le \varepsilon + \xi_i^*,\ i = 1,...,l \quad (15)$$

$$\xi_i \ge 0,\ i = 1,...,l \quad (16)$$

$$\xi_i^* \ge 0,\ i = 1,...,l \quad (17)$$

where the constant C influences a trade-off between an approximation error and an estimation error decided by the weight vector norm $\|w\|$, and this design parameter is chosen by the user. ξ_i and ξ_i^* are slack variables as the measurement upper bound and lower bound of outputs. This quadratic optimization is equivalence to apply Karush-Kuhn-Tucker (KKT) conditions for regression in which maximizing dual variables Lagrangian $L_d(\alpha, \alpha^*)$:

3

$$L_d(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi(x_i)^T \phi(x_j) \quad (18)$$

$$-\varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) - \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) y_i,$$

subject to constraints

$$\sum_{i=1}^{l} \alpha_i = \sum_{i=1}^{l} \alpha_i^*, \quad (19)$$

$$C \geq \alpha_i \geq 0, \quad i = 1,...,l \quad (20)$$

$$C \geq \alpha_i^* \geq 0, \quad i = 1,...,l \quad (21)$$

After calculating Lagrange multipliers α_i and α_i^*, find an optimal desired weights vector of the regression hyperplane as

$$w_0 = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\phi(x_i) \quad (22)$$

and an optimal bias of regression hyberplane as

$$b_0 = \frac{1}{l} \left(\sum_{i=1}^{l} (y_i - \phi(x_i)^T w_0) \right). \quad (23)$$

In nonlinear cases for regression, the kernal function, for typical instances, polynomial, RBF, or sigmoid function, will be adopt to replace the scale product $\phi(x_i)^T \phi(x_j)$ with $K(x_i, x_j)$ in Eq. (18).

If the term $\beta_i = (\alpha_i - \alpha_i^*)$ is defined in training data set, the output of SVR can be obtained with new input pattern z_i [25].

$$y = g\beta + b_0 \quad (24)$$

where the vector g is constructed by

$$g = z_i^T x,$$

and matrix x stands for patterns in training data set as well as vector z_i represents new input pattern.

$$x = [x_1 \, x_2 \, ... x_l]$$

$$z_i = [z_{i1}, z_{i2}, ..., z_{iN}]^T$$

The algorithm applied to the constrained optimization of support vector regression (SVR) is called adaptive support vector regression (ASVR) and designed for exploring two free parameters C and ε such that the computation burden for quadratic programming (QP) reduce a lot, and it converges to the near-optimal solution soon.
The unit-step function $u(t)$ is written by

$$u(t - t_0) = \begin{cases} 0, & if \ t < t_0 \\ 1, & if \ t \geq t_0 \end{cases}, \quad (25)$$

the delta function is represented as

$$delta(t - t_0) = \begin{cases} 0, & if \ t \neq t_0 \\ 1, & if \ t = t_0 \end{cases}, \quad (26)$$

and the sign function is defined as

$$sign(h) = \begin{cases} 1, & if \ h \geq 0 \\ -1, & if \ h < 0 \end{cases} \quad (27)$$

Let the original data sequence as expressed below.

$$X_N = \{x^{(0)}(1), x^{(0)}(2),..., x^{(0)}(n)\} \quad (28)$$

The algorithm of adaptive support vector regression (ASVR) is proposed in the following several steps from Eq. (29) to Eq. (46).

Step 1: normalization of data sequence on Eq. (25) as follows.

$$X_N = \begin{bmatrix} x^{(0)}(1) \\ x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \ \widetilde{X}_N = \frac{X_N}{\max_i |x^{(0)}(i)|} = \begin{bmatrix} \widetilde{x}^{(0)}(1) \\ \widetilde{x}^{(0)}(2) \\ \widetilde{x}^{(0)}(3) \\ \vdots \\ \widetilde{x}^{(0)}(n) \end{bmatrix} \quad (29)$$

Step 2: constructing a simple linear regression among the most recent normalized data points, $\widetilde{x}^{(0)}(1), \widetilde{x}^{(0)}(n),..., \widetilde{x}^{(0)}(n)$.

$$\widetilde{x}^{(0)}(k) = \varphi k + \psi , \quad k = 1,2,...,n \quad (30)$$

where φ is the slope and ψ is the bias in this line.

Step 3: Eq. (30) turns out to be a normal equation and its solution Θ to this least squared problem is obtained.

$$\widetilde{X} = \Omega\Theta , \quad (31)$$

$$\Theta = (\Omega^T \Omega)^{-1} \Omega^T \widetilde{X} , \quad (32)$$

where $\widetilde{X} = [\widetilde{x}^{(0)}(1), \widetilde{x}^{(0)}(2),..., \widetilde{x}^{(0)}(n)]^T$

$$\Omega = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ \vdots & \vdots \\ n & 1 \end{bmatrix}, \text{ and } \Theta = [\varphi, \psi]^T \quad (33)$$

Step 4: taking difference on Eq. (29) led to Δ_{N-1} followed by normalizing Δ_{N-1} to Λ_{N-1}.

$$\Delta_{N-1} = \begin{bmatrix} x^{(0)}(2) - x^{(0)}(1) \\ x^{(0)}(3) - x^{(0)}(2) \\ x^{(0)}(4) - x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) - x^{(0)}(n-1) \end{bmatrix} = \begin{bmatrix} \delta^{(0)}(2) \\ \delta^{(0)}(3) \\ \delta^{(0)}(4) \\ \vdots \\ \delta^{(0)}(n) \end{bmatrix}, \Lambda_{N-1} = \frac{\Delta_{N-1}}{\max_i |\delta^{(0)}(i)|} = \begin{bmatrix} \widetilde{\delta}^{(0)}(2) \\ \widetilde{\delta}^{(0)}(3) \\ \widetilde{\delta}^{(0)}(4) \\ \vdots \\ \widetilde{\delta}^{(0)}(n) \end{bmatrix} \quad (34)$$

Step 5: according to the above sequences both Δ_{N-1} and Λ_{N-1}, the total deviations κ, the coefficient of oscillation τ, the ratio of final deviation to mean deviation ρ, the ratio of last two deviation in average to mean deviation ϑ, and a coefficient of weighted-oscillation at final deviation σ are calculated.

$$\kappa = \sum_{i=2}^{n} |\widetilde{\delta}^{(0)}(i)| \quad (35)$$

$$\mu = \frac{\left| \sum_{i=2}^{n} \widetilde{\delta}^{(0)}(i) \right| + e}{\kappa + e} , \quad e = 10^{-6} \quad (36)$$

$$\rho = \frac{|\widetilde{\delta}^{(0)}(n)|}{\dfrac{\kappa}{n-1}} \quad (37)$$

$$\vartheta = \frac{|\widetilde{\delta}^{(0)}(n-1) + \widetilde{\delta}^{(0)}(n)|}{2\dfrac{\kappa}{n-1}} \quad (38)$$

$$\sigma = \tanh(\rho^2 / \mu) \quad (39)$$

Step 6: After that, a brief formula based on the several expressions on Eq. (35)-(39) to determine the

value of ε is established in the following Eq. (40)-(42).

$$q = \exp\left\{-\tanh(|\varphi|)^{1/2}\sigma^{(1-u(\mu-1))} - u(\mu-1)\cdot\rho^{\vartheta}\right\} \quad (40)$$

$$v = \begin{cases} q^{(1-delta(\kappa))} & if \ sign(\varphi)=1 \\ \left[q\mu^{(1-u(\mu-1))}\vartheta^{2u(\mu-1)}\right]^{(1-delta(\kappa))} & if \ sign(\varphi)=-1 \end{cases}, \quad (41)$$

$$\varepsilon = v\times\frac{|\max(X_N)-\min(X_N)|}{2}, \quad (42)$$

Once the value of v has been determined, the ε in Eq. (42) is set, and then the constrained optimization on Eq. (18)-(21) will start for several iterations to search the optimal w_0 and b_0 on Eq. (22)-(23). In support vector regression, an increase value of parameter C will highly penalizes the big empirical error while an increase value of ε will reduce the support vectors to loose the bound of empirical error [26]. Therefore, how to deal with a trade-off between C and ε in SVR is become a very important issue. In this research, the relationship between ε and C, we proposed, can be constructed in the basis of modified Bessel function of second kind with the order n [27] as expressed below. A specific integer number n is obtained from a function $\lceil \ \rceil$ of the coefficient of the oscillation μ as described on Eq. (33).

$$n = \lceil 1/\mu \rceil \quad (43)$$

where the operator $\lceil 1/\mu \rceil$ is represented as a smallest integer bigger than $1/\mu$.

$$C = K_n(\varepsilon) = (-1)^{n+1}\left\{\ln(\varepsilon/2)+\gamma\right\}I_n(\varepsilon)+\frac{1}{2}\sum_{k=0}^{n-1}(-1)^k(n-k-1)!(\varepsilon/2)^{2k-n}$$
$$+\frac{(-1)^n}{2}\sum_{k=0}^{\infty}\frac{(\varepsilon/2)^{n+2k}}{k!(n+k)!}\left\{\Phi(k)+\Phi(n+k)\right\} \quad (44)$$

$$I_n(\varepsilon) = \sum_{k=0}^{\infty}\frac{(\varepsilon/2)^{n+2k}}{k!\Gamma(n+k+1)} \quad (45)$$

where $\gamma = 0.5772156...$ is Euler's constant and $\Phi(p) = 1+\frac{1}{2}+\frac{1}{3}+...+\frac{1}{p}$, $\Phi(0)=0$.

The tunable free parameters in SVR can be done automatically and referred it to as adaptive support vector regression (ASVR). In addition, another free parameter σ of radial basis kernel function has exploited here for optimizing SVR learning. As we know, radial basis function is usually utilized as a sort of activation function while training an intelligent machine. Furthermore, the radial basis function can not only be applied to linear or nonlinear dynamics, but it also frequently acted as a kernel function for SVR learning more than polynomial or tangent-sigmoid function do. Thus, it is important that how to select an optimal one from the given training data set is a critical issue. However, there is no any unified thumb of rule to search an appropriate scale for free parameter σ of radial basis kernel function, even though several literatures [28][29] has proposed some method to explore the near optimal free parameter σ. Accordingly, the optimal learning in SVMs becomes an challenge

problem. In this study, we proposed a novel scheme to compute a near optimal value of free parameter σ, denoted by σ_{rbkf}, automatically so that it can be obtained fast and succinctly according to provided training data sequence $\{x_1, x_2,..., x_l\}$ in the following formula on Eq. (46)

$$\sigma_{rbkf} = \upsilon\cdot\left(\frac{\sum_{i=1}^{l}(x_i-\bar{x})^2}{l-1}\right)^{1/2}, \ \bar{x} = \frac{\sum_{j=1}^{l}x_j}{l} \quad (46)$$

In such a way, the calculation burden for searching an appropriate free parameter σ indeed reduces a lot.

3. EMPIRICAL SIMULATION AND DISCUSSIONS

A collection of data about five indicators, (i) the growth rate of long-term investment (GRLTI), (ii) the firm size (FS), (iii) the return on total asset (ROA), (iv) the return on common equity (ROE), and (v) the return on sales (ROS), from TEJ [30] including 30 corporations have been cited herein for explaining the timing of resources exploration in the behavior of firm to fit in with the real world dynamics in changing environments. In order to accomplish data manipulation easier, data preprocess is required to transform indicator GRLTI linearly with appropriate bias, and a natural logarithm applied to indicator FS. The first phase designed as training/learning stage for modeling linear structure of ARMAX as well as nonlinear structure of BPNN and SASVR, which is of the posterior analysis from observed 383 historical data for a period of 10 years from 1995 to 2004. Next, the second phase, the prior validation stage proceeded to simulate the empirical results for examining the system performance employing interpolation from trained model. Estimated ARMAX with a bias of scalar (bias2 = 2000.5) simulated from computer:

$$y(t) = 0.3023\cdot y(t-1)+0.1651\cdot y(t-2)+0.04129\cdot y(t-3)$$
$$+0.2023\cdot y(t-4)+0.04312\cdot u_1(t-3)-0.004114\cdot u_2(t-2)$$
$$-0.0023\cdot u_3(t-1)-0.0056\cdot u_4(t-3)+e(t)-0.2227\cdot e(t-1)$$
$$-0.1648\cdot e(t-2) \quad (47)$$

Thus, we can construct an estimated ARMAX model as expressed in Eq. (47). In this estimated ARMAX model, we can check directly from Eq. (47) to explain the current GRLTI is related to it's the most recent four lags of GRLI, as well as coupled to the second lag of FS, the second lag of ROA, the second lag of ROE, and the second lag of ROS. We can interpret that in the respect of autoregressive GRLTI is definitely auto-correlated to a few of most recent historical (the past) GRLTI. Furthermore, there are three indicators ROA, ROE, and ROS playing the roles to affect the current GRLTI negatively. In other words, increasing on ROA, ROE, or ROS will depress the current GRLTI. This development can meet the typical theory in the prospect theory

perspective [14]. Nevertheless, an indicator FS can promote the current GRLTI such that the larger FS is, the higher GRLTI will be. It is also noted that strictly speaking the residual terms indicated by $e(t)$, $e(t-1)$, and $e(t-2)$ in MA part have small values usually. The MA part cannot affect GRLTI significantly, even though they have relatively larger coefficients with respect to those terms, $e(t)$, $e(t-1)$, and $e(t-2)$ as shown in Fig. 3. Obviously, MA part is trivial in this ARMAX model as a result of small residuals; in contrast, AR and X part of estimated ARMAX model are used to determine GRLTI predominantly and their corresponding coefficients is displayed as shown in Fig. 4. As a matter of fact, the most recent lags of GRLTI are essentially related to the performance of GRLTI, and secondly we must also take FS into account when we examine the changes in GRLTI.

The performance criteria [31] derived by mean square error (MSE), mean absolute deviation (MAD), and mean absolute percent error (MAPE) will be used to compare the empirical simulation for every competitive models (ARMAX, BPNN, ANFIS, and ASVR) applied to the same sample data set, as listed in Table 1. After the comparison between three models, nonlinear models have better performance on MSE, MAD, and MAPE than the linear one, namely, ARMAX model. This is because MAPE criterion of the trained ARMAX model with different biases always cannot less than 5%. It implies that the accuracy of empirical simulation is not enough for the trained ARMAX model. On the contrary, nonlinear models like BPNN, ANFIS, and ASVR have achieved higher effectiveness in modeling. Furthermore, the goodness of fit for the proposed methods, as listed in Table 1, is also tested by Q-test [17], and null hypothesis cannot be rejected due to all p-value greater than level of significance (5%). In other words, all of trained models are significant on the test of fitting problem. As for model validation, ASVR has achieved the best Akaike information criterion (AIC) and Bayesian information criterion (BIC) [18], that is, the better reliability for ASVR modeling. However, the individual interaction between any single input indicator and the output indicator cannot be presented in the nonlinear structures of BPNN, ANFIS, or ASVR. In linear structure of ARMAX model, we can view directly which input indicator affecting the output indicator to explain the timing of resources exploration in the behavior of firm. Even though SVR takes a large longer time to train its model, ASVR we proposed takes much less computation time for modeling than ordinary SVR as listed in Table 1.

4. CONCLUSING REMARKS

The following statements summarize the accomplishment of the proposed methods.

(a) In the linear structure of ARMAX model, four input indicators (the firm size, the return on total asset, the return on common equity, and the return on sales) can actually affect the changes in the output indicator (the growth rate of long-term investment) significantly over the different levels. That is, the resulting ARMAX model can explain the growth rate of long-term investment that can help decision-maker to explore new resources due to the pressure of external changing environment. This dynamics also can be considered as the timing when the risky attitude of managers will be shifted from risk-avoiding exploiting activities to relative risk-taking exploring activities.

(b) In ARMAX model, the firm size affects the growth rate of long-term investment positively whereas the return on total asset, the return on common equity, and the return on sales influence the growth rate of long-term investment negatively. Besides, 1-lag and 2-lag of innovations are also involved in the resulting output. There are four lags within auto-regression of ARMAX model, and the auto-regression of ARMAX model dominates the growth rate of long-term investment largely since the coefficients with respect to four lags of auto-regression are relatively bigger than the other parts. Secondly, we cannot neglect the contribution of 3-lag of firm size to the output due to its big coefficient.

(c) For a given long length data stream, the performance of estimation using ASVR not only reduces the MSE a lot to attain better generalization, but it also improves MAPE highly to boost the localization as comparing with the ordinary SVR. Clearly, the nonlinear structure of ASVR can get the satisfactory results better than the linear structure of ARMAX model or the nonlinear structure of BPNN and ANFIS. However, the nonlinear models, BPNN, ANFIS, or ASVR, cannot tell us the exact the contribution of individual factor to the output, the growth rate of long-term investment, because those factors are hidden in the nonlinear system.

6. REFERENCES

[1] E. Penrose, The Theory of the Growth of the Firm*, Oxford University Press: Oxford, 1959.

[2] R. P. Rumelt, Toward a Strategic Theory of the Firm. In Competitive Strategic Management, pp.

556-570, Lamb RB (Ed), Prentice-Hall: Englewood Cliffs NJ, 1984.

[3] J. B. Barney, "Types of Competition and the Theory of Strategy: Towards an Integrative Framework," Academy of Management Review**,** vol.11, no.4, pp.791-800, 1986b.

[4] J. B. Barney, "Firm Resources and Sustained Competitive Advantage," Journal of Management, vol.17, no.1, pp. 99-120, 1991.

[5] J. B. Barney, "Looking Inside for Competitive Advantage," Academy of Management Executive, vol. 9, no. 4, pp. 49-61, 1995.

[6] I. Dierickx and K. Cool, "Asset Stock Accumulation and Sustainability of Competitive Advantage," Management Science, vol. 35, no. 12, pp. 1504-1514, 1989.

[7] M. A. Peteraf," The Cornerstones of Competitive Advantage: A Resourced-Based View," Strategic Management Journal, vol. 14, pp. 179-191, 1993.

[8] B. Wernerfelt," A Resource-Based View of Firm," Strategic Management Journal**,** vol. 5, pp. 171-180, 1984.

[9] C. E. Helfat and M. A. Peteraf, "The Dynamic Resource-Based View: Capability Lifecycles," Strategic Management Journal, vol. 24, pp. 997-1010, 2003.

[10] J. G. March, "Exploration and Exploitation in Organizational Learning," Organization Science, vol. 2, pp. 71-87, 1991.

[11] R. M. Cyert and J. G. March, A Behavioral Theory of the Firm, U.S.: Prentice-Hall, 1963.

[12] D. A. Levinthal and J. G. March, "The myopia of learning," Strategic Management Journal (1986-1998), vol. 14, pp. 95-112, 1993.

[13] H. A. Simon, Administrative Behavior: A Study of Decision-Making Process in Administrative Organization, Free Press: New York, 1997.

[14] D. Kahneman, &, A. Tversky, "Prospect Theory: An analysis of Decision Under Risk," Econometrica, pp. 263-291, 1979.

[15] F. T. Rothaermel and D. L. Deeds, "Exploration and Exploitation Alliances in Biotechnology: A System of New Product Development," Strategic Management Journal, vol. 25, no. 3, pp.201-221, 2004.

[16] J. A. Schumpeter, The Theory of Economic Development, Transaction Publishers: New Brunswick, NJ, 1934.

[17] Hamilton, J.D., *Time Series Analysis*, New Jersey: Princeton University Press, 1994.

[18] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting & Control,* New Jersey: Prentice-Hall, 1994.

[19] B. L. Bowerman and R. T. O'Connell, *Forecasting and Time Series: An Applied Approach*, Belmont: Duxbury Press, 1993.

[20] S. Haykin, Neural Network: A Comprehensive Foundation, 2nd Ed, Prentice Hall, New Jersey, 1999.

[21] J.-S. R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No. 3, pp. 665-685, May 1993.

[22] J. Neter, W. Wasserman, and M.H. Kutner, Applied linear statistical models, 2nd Ed, Homewood, IL: Irwin, 1985.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[24] N. Cristianini, and J. Shawe-Taylor, An Introduction to Support Vector Machines (and other kernel-based learning methods), London: Cambridge University Press, 2000.

[25] S. R. Gunn, Support Vector Machines for Classification and Regression. Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

[26] V. Kecman, Learning and Soft Computing, Massachusetts : MIT Press, 2001.

[27] E. Kreyszig, Advanced Engineering Mathematics, 8th Edition, Wiley, New York, 1999.

[28] Chih-Chung Chang, Chih-Jen Lin "LIBSVM: a library for Support Vector Machines", http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf

[29] Y. L. Murphey, Z. H. Chen, M. Putrus, and L. Feldkamp, "SVM Learning from Large Training Data Set", Proc. IEEE IJCNN, pp.860-2865, 2003.

[30] TEJ database, Taiwan Economic Journal Co. Ltd.,Taiwan, 2004. http://www.tej.com.tw/

[31] B. R. Chang, "Hybrid BPNN-Weighted Grey-CLMS Forecasting," *Journal of Information Science and Engineering*, Vol. 21, No. 1, pp. 209-221, January, 2005.

Figure 2. A fuzzy inference through a single rule is conducted by ANFIS.

TABLE 1

The performance criteria is asserted by using mean square error (MSE), mean absolute deviation (MAD), and mean absolute percent error (MAPE) for five indicators to analyze the timing of resources exploration in the behavior of firm with 383 sampled data from 1995 to 2004. Data preprocess has been manipulated linearly with two different biases (bias1=1000.5 and bias2=2000.5). The goodness of fit for the proposed methods is also test by Q-test with the level of significance (α =0.05). Model validation is evaluated by Akaike information criterion (AIC) or Bayesian information criterion (BIC) and the computation time is done as well.

Criteria	ARMAX (bias1)	ARMAX (bias2)	BPNN (bias2)	ANFIS (bias2)	SVR (bias1)	ASVR (bias2)
MSE	0.0025	0.0023	0.0021	0.0021	0.0021	0.0000215
MAD	0.0063	0.0085	0.0066	0.0068	0.0457	0.0046
MAPE	0.0532	0.0511	0.0349	0.0368	0.3856	0.0392
p-VALUE	0.4415	0.1583	0.4823	0.6262	0.0733	0.4797
AIC	-1205.0	-1230.6	-1258.5	-1261.2	-2520.9	-2293.5
BIC	-1189.2	-1214.8	-1242.7	-1245.4	-2505.1	-2285.8

Method abbreviation
1. ARMAX- Auto-regressive Moving-average Regression
2. BPNN- Back-propagation Neural Network
3. ANFIS- Adaptive Neuro-fuzzy Inference System
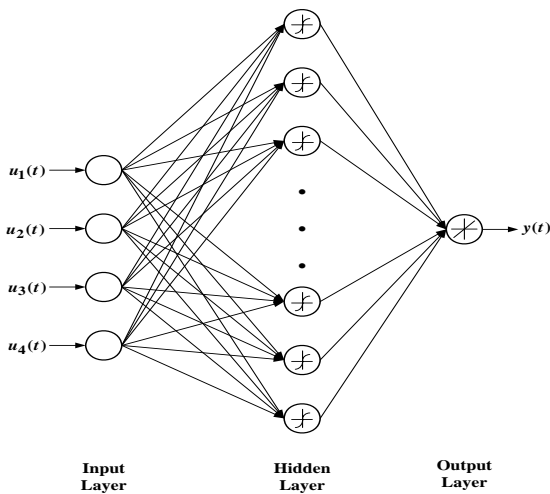4. ASVR- Adaptive Support Vector Regression

Figure 3. The coefficients with respect to the corresponding explanatory variables of AR, MA and X parts of ARMAX are displayed.

Figure 1. A typical three-layer BPNN architecture and Tangent-sigmoid and Pure-line transfer function used as the output of activations in the hidden layer and output layer, respectively.
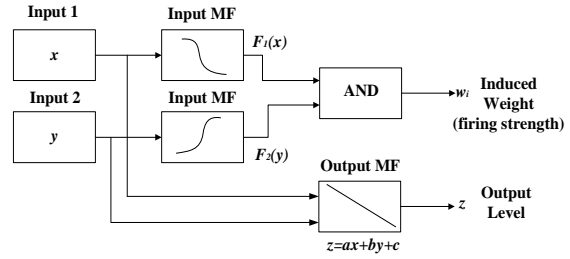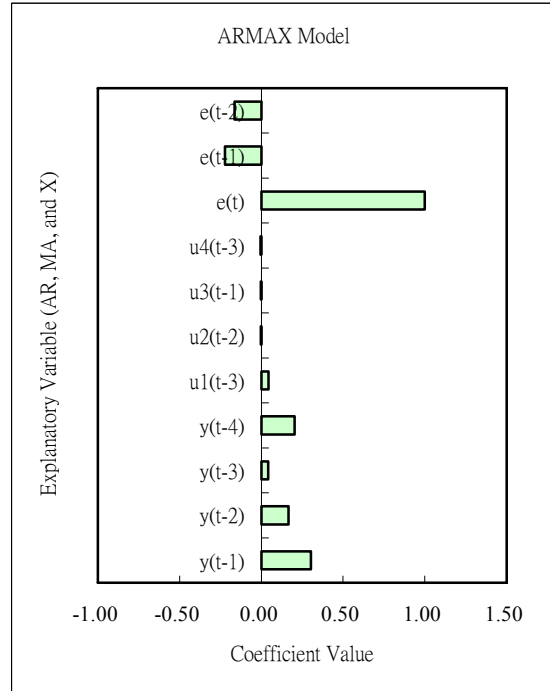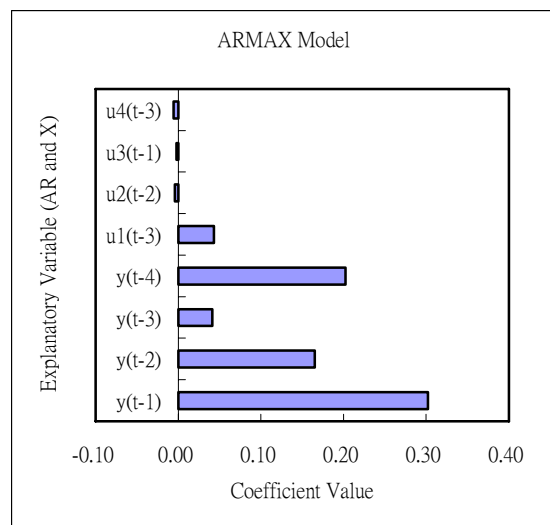

Figure 4. The coefficients with respect to the corresponding explanatory variables of AR and X parts of ARMAX model are emphasized here.