Modeling for Predicting Sales Revenue Changes with Uncontrollable Economic Factors: A Case for Korean Companies

Taeheung Kim and Hoo-Gon Choi

Department of Systems Management Engineering, Sungkyunkwan University, Suwon, Republic of Korea

Abstract: *Uncontrollable economic variables such as the interest rate, commodity price, oil price, currency exchange rate, and other macroeconomic factors affect the sales volume, sales cost and revenue, and profitability of a firm. Therefore, it is necessary to observe the influences of such variables on sales figures through an accurate quantitative model. In this study, two outstanding linear models—an autoregressive (AR) model and a support vector regression (SVR) model—are adapted to the prediction of sales revenue changes under the influence of 15 exogenous variables for 315 Korean companies. After collection of raw datasets of quarterly sales revenues and exogenous variables from the real world, predictions are made at the given quarters by using training datasets for the companies. The performances of both the models are evaluated based on the prediction accuracy defined by the average and standard deviation. As a result, the SVR model is found to perform better in 66,150 experiments. The findings of this study suggest that accurate prediction models should be developed for observing the behavior of the exogenous factors, for managing them more effectively and efficiently in new product development, and for developing an optimal decision-making process.*

Keywords: *uncontrollable economic variables, sales revenue, AR model, SVR model, prediction accuracy*

1. Introduction

A firm can gain effective competitive advantage through successful new product development (NPD) irrespective of the types of industries being considered. The developed new products might contribute to an increment in sales revenues. However, the success of an NPD project is difficult to achieve because these projects involve unexpected risks, uncertainties, and difficulties. It has been reported that around 66% of all NPD projects fail before the target date and more than 50% of all projects make no returns on investment; further, more than 40% of respondents to a survey stated that NPD projects require more money and time than existing products [2]. Various internal and external risks or uncertainties, such as lack of funds, government policy changes, market segmentation, economic changes, rapid technology changes, unpredictable customer needs, oil price changes, commodity price changes, and currency exchange rate changes, cause project failure. Therefore, the sales revenue increment resulting from the development of new products is affected by both risks and uncertainties. Furthermore, the revenue figures fluctuate either positively or negatively depending on macroeconomic factors throughout the lifecycle of the new products.

Any firm may wish to examine how significantly either risks or uncertainties affect its sales revenue. In this study, quantitative models are developed for predicting the quarterly changes of sales revenues based on variations of core external uncertainties. For this purpose, sales revenue data are collected from 315 companies registered in Korea’s two major stock markets, covering the quarters in years 2002 to 2014. Furthermore, 15 core economic uncontrollable exogenous factors are collected for the same quarters. The autoregressive (AR) model and the support vector regression (SVR) model are adapted to this study.

2. Related Work

Many macroeconomic studies have examined how uncontrollable factors such as exchange rates, oil prices, and interest rates operate in various economic environments. For example, Ahn [1] investigated how exchange rate fluctuations affected 6,801 Korean manufacturing firms between 1989 and 2006. This empirical study showed that the firms' profitability declined when the Korean won/US dollar exchange rate increased and that it improved when the Japanese yen/US dollar rate increased. The LG Economic Research Institute [3] has reported that fluctuations in the Korean won/US dollar exchange rate have affected Korean industries both positively and negatively depending on how well the firms have managed their own exchange-rate risk systems. Kim and Choi [5] considered financial indices such as the interest rate, commodity price, oil price, currency exchange rate, and stock price to determine their effects on the sales revenue of electronics companies. A dynamic time warping method classifies sales revenue data collected from 50 electronics companies covering 2000 to 2012 into two clusters, and a linear regression determines three critical exogenous factors: interest rate, oil price, and commodity price. Nevertheless, only a few studies have examined how multiple uncontrollable exogenous factors synthetically affect sales measures, perhaps because each factor's behavior is complex because of the confusedly interlaced economic situations in different countries.

3. Methodology

The purpose of this study is to predict the sales revenue fluctuation of Korean companies registered in the Korean stock markets. It is assumed that the latest sales revenue is related to the previous sales revenues and to 15 uncertainties or uncontrollable exogenous factors. From this assumption, a functional relationship between the sales revenue and the exogenous factors is given as in Eq. (1):

$$Y_n = f\left(Y_{n-1}, Y_{n-2}, \cdots, Y_{n-k}, X_{n-1,1}, X_{n-1,2}, \cdots, X_{n-1,15}\right) + \varepsilon_n \qquad (1)$$

where

$\hat{Y}_n$ = predicted sales revenue in n^{th} quarter
$Y_{n-1}, Y_{n-2}, \cdots, Y_{n-k}$ = actual sales revenue
k = number of quarters to be used for predicting the sales revenue
$X_{n-1,i}$ = value of i^{th} exogenous factor in $(n-1)^{th}$ quarter
ε_n = error term of n^{th} quarter with standard normal distribution N(0, 1)

Two different linear models—the AR model and the SVR model—are adapted to this study for predicting the sales revenue in the n^{th} quarter. Both these models are developed by defining training datasets, and the actual prediction at a specified time is made using k. The prediction accuracy is measured by the range of k and the amount of errors for both models.

3.1. AR Model (AR(k))

In the AR model, Y_n is assumed as the k^{th}-order linear function in which the coefficients of independent variables are estimated by the least-squares method, as given in Eq. (2):

$$\hat{Y}_n = \beta_0 + \beta_1 Y_{n-1} + \beta_2 Y_{n-2} + \cdots + \beta_k Y_{n-k} + \alpha_1 X_{n-1,1} + \alpha_2 X_{n-1,2} + \cdots + \alpha_{15} X_{n-1,15} + \epsilon_n \qquad (2)$$

The least-squares method determines all the coefficients by minimizing the total amount of errors defined by $\sum_{j=1}^{M}\left(\hat{Y}_n - Y_n\right)$, where M is the number of models applied for a company.

3.2. SVR Model (SVR(k))

The SVR model is another version of the support vector machine (SVM), which is one of the most well-known supervised linear classifiers, for regression. In this model, the coefficients of independent variables are estimated by minimizing the cost function plus the penalty term, as given by Eq. (3) [4], instead of estimation by minimizing the amount of errors as done in the AR model. $Y_{n-1}, Y_{n-2}, \cdots, Y_{n-k}, X_{n-1,1}, X_{n-1,2}, \cdots$, and $X_{n-1,15}$ are training samples and the k^{th}-order regression is established in a manner similar to that in the AR(k) model. It should be noted that the function f has an identical form to that in the AR(k) model but with different coefficient values.

$$\mathrm{H} = \sum_{i=1}^{N} V_c\left(Y_n - f\left(Y_{n-1}, Y_{n-2}, \cdots, Y_{n-k}, X_{n-1,1}, X_{n-1,2}, \cdots, X_{n-1,15}\right)\right) + \frac{\lambda}{2}\left(\sum_{j=0}^{k} \beta_j^2 + \sum_{j=1}^{15} \alpha_j^2\right) \quad (3)$$

where

$V_c(\cdot)$ = cost function defined by

$$V_c(t) = \begin{cases} 0 & \text{if } |t| < \varepsilon \\ |t| - \varepsilon & \text{otherwise} \end{cases}$$

Y_n = target value or the actual sales revenue at quarter i

λ = weight coefficient for penalty ($\lambda > 0$; high weight results in low β values)

N = total number of quarters

In this study, $\lambda = 1$ and $\varepsilon = 0.1$.

4. Data Collection and Experimental Results

The purpose of this study is to predict the sale revenue for future periods (quarters) under variations of the exogenous economic factors by using the AR and SVR models. Both these models are established for the same company. The prediction accuracy is the key criterion for evaluating the model performances. It should be noted that the models predict the sales revenue values at a given quarter ($\hat{Y}_n$), and the predicted values are compared with the actual values (Y_n) in terms of the corresponding values of the previous quarters, (($\hat{Y}_n, \hat{Y}_{n-1}$) and (Y_n, Y_{n-1})). If both pairs of revenues have an identical trend—either positive or negative, the corresponding model is said to perform accurately.

4.1. Datasets

There are two major stock price indices in the Republic of Korea: KOSPI (Korea Composite Stock Price Index) and KOSDAQ (Korea Securities Dealers Automated Quotation). KOSPI is the index for major companies, and KOSDAQ is the index for either small-medium-sized or venture companies. In this study, a total of 315 companies registered in both the stock markets are selected for collecting the quarterly sales revenue data covering 2002 to 2014 (i.e., a total of 52 quarters). Table I presents the number of companies classified under eight different categories.

TABLE I: Number of companies according to industry categories for collecting sale revenue data

Category	Chemistry	Electronics & Electric	Textile	Mechanical	Medical Devices	Pharmacy	Retail, Logistics, & Distribution	Software
Number of Companies	82	57	21	41	14	50	32	18

The core economic uncontrollable factors are classified into four groups: interest rate, commodity prices, oil price, and currency exchange rate. Each group has several associated factors, as listed in Table II.

TABLE II: Uncontrollable exogenous factors

Exogenous Factor	Quarterly data collected
Interest rate	USA, EU, Japan, Korea, and China
Commodity price	Producer price index Consumer price index Raw material price index
Oil price	Dubai, Brent, WTI
Currency exchange rate	Dollar, Euro, Yen, Yuan

The values of both the sales revenue and the exogenous factors are time series data and would be characterized by trends and seasonal factors. In this study, the collected dataset is preprocessed to remove the seasonal variations and is used for predicting the sales revenues in future periods.

4.2. Experimental Method

The experimental sequence for using both the AR(k) and the SVR(k) models includes the following four steps:

Step 1 – The candidate quarters for predicting the sales revenues are set as the 31^{st} to 51^{st} quarters after developing both models using the revenue data for the past 25 quarters. This means that the training dataset used for developing the models is the 25-quarter data for each company. For example, the training dataset for predicting the revenue for the 31st quarter includes past data of n = 6–30. This procedure leads to the development of 21 different models for the 31st to 51st quarters.

Step 2 – With the models obtained in Step 1, the *k* value in Eqs. (2) and (3) is changed from 1 to 5. This means that five different models each for AR(k) and SVR(k) are developed for predicting the sales revenue at a given quarter for a company.

Step 3 – If the revenue change predicted at a given quarter is the same as the actual change at the corresponding quarter in a given model, the model performs accurate prediction for that quarter. In order to determine the prediction accuracy for a company, this step is repeated for the 31^{st} to 51^{st} quarters for each of the five models developed by either AR(k) or SVR(k).

Step 4 – As a result from Step 3, the number of accurate models from among the 21 models for predicting the 31^{st} to 51^{st} quarters can be counted for each company under either AR(k) or SVR(k). Then, the proportion of accurate models is computed and it is added to the total proportion of all the companies in a given industry category. The total proportion of accurate models is the final accuracy of either the AR(k) or the SVR(k) model. For example, the category "Chemistry" includes 82 companies, for which 21 models are applied under AR(1). Suppose that the total number of accurate models is 1,061. Then, the average accuracy of AR(1) becomes 0.61644 (1061/(82*21)).

4.3. Experimental Results

The total number of experiments for determining a model's accuracy is 66,150 (total number of companies × number of orders (*k*) × number of quarters to be predicted × number of model types = 315 × 5 × 21 × 2).

Table III summarizes the average accuracy (%) along with the standard deviation for each category under either AR(k) or SVR(k). The prediction accuracy improves when the order *k* is increased, especially in the SVR(k) models. Further, a larger standard deviation is produced when the model complexity increases in each model. The SVR(k) models produce lower standard deviations than the AR models, which leads to the SVR(k) models showing more stable performance than the AR(k) models. Therefore, SVR(k) models with high *k* values are recommended for future related studies irrespective of the industry categories being considered for predicting the sales revenue.

TABLE III: Averages and standard deviations by prediction models

	Chemistry	Electronics & Electric	Textile	Mechanical	Medical Devices	Pharmacy	Retail, Logistics, & Distribution	Software
AR(1)	61.6144 (8.0903)	60.3174 (9.7847)	59.4104 (7.723)	63.1823 (6.57)	61.2244 (7.6056)	63.9047 (6.5354)	63.3928 (7.0868)	58.4656 (9.0914)
AR(2)	57.0847 (7.4921)	56.1403 (9.596)	58.5034 (8.8348)	60.2787 (7.5556)	56.4625 (8.4145)	60.4761 (7.4535)	58.7797 (7.8615)	54.2328 (8.7141)
AR(3)	79.6167 (14.8973)	67.4185 (15.1056)	76.6439 (19.656)	77.8164 (13.5754)	70.7482 (17.4367)	87.4285 (13.859)	84.6726 (15.2331)	73.5449 (17.8331)
AR(4)	82.8106 (15.345)	71.5956 (16.9703)	78.458 (22.2684)	83.5075 (14.7278)	73.1292 (19.6838)	89.238 (15.4394)	87.9464 (13.7284)	76.1904 (18.7812)
AR(5)	80.7781 (15.6958)	69.6741 (16.9625)	78.458 (21.3272)	81.5331 (15.5355)	74.4897 (15.5461)	87.619 (15.9931)	85.5654 (15.1749)	73.8095 (20.711)
SVR(1)	71.835 (7.2562)	70.0918 (7.9158)	68.9342 (10.0338)	70.2671 (9.2157)	72.1088 (5.045)	74 (4.8757)	73.0654 (8.3)	68.5185 (12.4113)
SVR(2)	71.6027 (7.0334)	70.7602 (7.0483)	69.6145 (8.6287)	71.0801 (8.9483)	73.1292 (4.2618)	74.3809 (4.2474)	72.4702 (8.7266)	69.5767 (11.2394)
SVR(3)	81.0685 (10.9545)	77.0258 (11.4966)	76.8707 (11.7826)	80.8362 (13.2404)	80.2721 (8.0202)	82.1904 (7.5478)	82.738 (8.5639)	75.1322 (14.4222)
SVR(4)	84.3205 (11.5088)	79.4486 (13.3249)	80.7256 (12.1606)	83.6236 (14.2294)	82.9931 (9.475)	87.3333 (7.4145)	86.4583 (11.111)	80.6878 (14.7603)
SVR(5)	84.2044 (11.8027)	79.6157 (12.9493)	80.4988 (12.1097)	83.8559 (14.2616)	84.0136 (8.1703)	87.4285 (7.5448)	86.1607 (11.7465)	82.2751 (12.4788)

5. Conclusion

Various uncontrollable economic factors such as the interest rate, commodity price, oil price, currency exchange rate, and other macroeconomic variables affect the sales volume, sales cost and revenue, and profitability of a company. The impacts of such variables should be managed effectively and efficiently to optimize the decision-making processes of a firm. The optimized process is significantly important in the development of new products in order to extend the firm's competitive advantage, because the success of NPD projects is heavily dependent on risks and uncertainties. Successful NPD through the use of an optimized management process would increase the sales revenue of a firm.

Predicting the sales revenue in a given period by developing highly accurate models is necessary for making a go/no-go decision on multiple NPD projects and for managing both controllable and uncontrollable economic variables. In this study, linear functions are developed to determine the relationship between the sales revenue changes and 15 uncontrollable exogenous factors and to predict the sales revenue changes for 315 Korean companies on the basis of time series data collected for 2002 to 2014. Two outstanding linear models—the autoregressive (AR) model and the support vector regression (SVR) model—are adapted to this study. The performance of the SVR model is found to be better than that of the AR model in terms of the accuracy defined by the average and standard deviation.

The findings of this study suggest that accurate prediction models should be developed for observing the behavior of the exogenous factors and for managing them more effectively and efficiently, rather than determining which model performs better.

6. Acknowledgements

This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIP) (NRF-2012R1A1A2045906).

7. References

[1] S. K. Ahn, *The effect of exchange rate fluctuation on the Korean firm's performance and its implication for the strategy to face exchange rate problems*, Report, Korea Economic Research Institute, 2007.

[2] J. Coppendale, "Manage risk in product and process development and avoid unpleasant surprises," *Engineering Management Journal*, vol. 5, no. 1, pp. 35-38, 1995.

[3] LG Economic Research Institute, *How much the profitability is deteriorated by the rise of dollar exchange rates*, Report, 2002.

[4] J. Friedman, T. Hastie and R. Tibshirani, *The elements of statistical learning* (Vol. 1), Springer, Germany, 2001.

[5] T. H. Kim and H. G. Choi, "Critical exogenous factors affecting sales revenues of Korea's electronics companies," presented at *The 2015 International Conference on Management Engineering and Information Technology Application*, Hong Kong, 2015.