Quest Journals
Journal of Research in Agriculture and Animal Science
Volume 4 ~ Issue 8 (2017) pp: 23-26
ISSN(Online) : 2321-9459
www.questjournals.org

Research Paper

Predicting Agricultural Products Volume With APC ERP Data

Minje Jeong[1], Young Jin Lee[2], Youngchan Choe[*]
[1*] *Department of Agricultural Economics and Rural Development, Seoul National University.*
[2] *Ezfarm Corporation.*

Received; 08 Mar. 2017 Accepted; 23 Mar. 2017; © The author(s) 2017. Published with open access at **www.questjournals.org**

ABSTRACT: *The purpose of the study is to predict agricultural product volumes using Agricultural product processing center (APC) enterprise resource planning (ERP) data from Korea. So far, great attention by the government has been shown to predict agricultural product volumes to stabilize price, especially high volatility products such as cabbage. In the past, it was hard to predict volumes precisely due to the lack of useful data. Recently, useful data has been accumulated from various sources such as sensors in greenhouses, information systems and public areas. This makes it possible to predict agricultural product volumes more precisely. For this study, we employ the support vector machine (SVM) to predict cabbage volumes. SVM is a semiparametric technique with origins in the machine-learning literature of computer science and its prediction performance is well known. We explore results using SVM against three other methods: ordinary least square (OLS), auto regression (AR), and vector auto regression (VAR). The results show that the prediction performance of SVM is better than that of the other three methods. We expect that the results can be applied to predict domestic cabbage volumes and ERP dashboard for top management at the APC.*
Keywords: *Support vector machine, APC ERP data*

I. INTRODUCTION

The government of South Korea (hereafter, Korea) has implemented efforts in stabilizing agricultural product prices. The volatility of agricultural product prices, especially cabbage, is large. Due to the high volatility of agricultural product prices, oversupply happens frequently, which deteriorates rural household incomes.

Prediction value of agricultural products volume is needed to stabilize agricultural product prices. For predicting future value, data of farm productivity volume or mid-distribution is required. In the past, this kind of data was rare due to the lack of information systems. During the recent development of information technology (IT), useful data has been accumulated from information systems such as farm sensors. ERP in agriculture has enabled us to predict agricultural product volumes more precisely when compared to the past.

In this study, we develop a model for predicting domestic agricultural product volumes monthly using the agricultural processing center (APC) enterprise resource planning (ERP) data. The APC is the mid-distribution agent in agriculture. The APC contracts with farms and purchases their original products. Once, they purchase agricultural products, they distribute the products to agricultural wholesale markets or hypermarkets. For that reason, APC plays an important role in agricultural distribution. Thus, their data can be crucial for predicting agricultural product volumes.

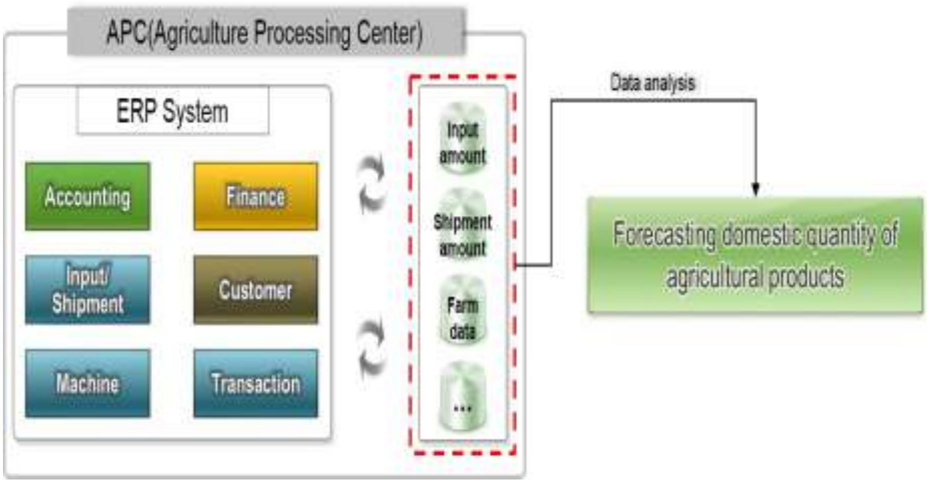
We use the support vector machine (SVM) model with APC ERP data to predict cabbage volumes. SVM is a semi-parametric technique with origins in the machine-learning literature of computer science and its prediction performance is well known (Cui & Curry, 2005). Additionally, we explore results using SVM against three other econometrics methods such as ordinary least square (OLS), autoregression (AR), and vector auto regression (VAR).

This study has two implications. For academia, this study can be useful to find which algorithm has an effective prediction performance among OLS, SVM, and time series. For the practitioner, the results of the study enable the APC to develop a flexible strategy for shipment.

II. APC DATA FROM INTERNAL INFORMATION SYSTEM

The APC is the mid-distribution agent in agriculture. They provide a contract with regional farms and purchase their original products. As mentioned, their role is crucial for domestic agriculture. APC contributes to the farms' benefits and they have valuable data for predicting agricultural product volumes.

The APC uses information systems such as ERP because there are many things they have to control. In the ERP server, there is valuable data related to contracts with regional farms and shipments (i.e.,product location and volumes distributed). Figure 1 shows the architecture of an ERP system at the APC. There are a lot of useful data for predicting agricultural product volumes. Thus, we can estimate future agricultural product volumes using APC ERP data.



III. SVM(SUPPORT VECTOR MACHINE)

The support vector machine combines concepts from abstract Hilbert spaces with modern optimization techniques (Cui & Curry, 2005). SVM is well known for one of the effective solutions to solve out classification issues (Heo, 2013).

$$\frac{1}{2}\parallel \omega \parallel 2 + C\sum_{i=1}^{n}\gamma_i$$

subject to 1) $\omega_t x_i + b \geq 1 - \gamma_i, \text{ if } y_i = 1,$

subject to 2) $\omega_t x_i + b \leq -1 + 1 + \gamma_i, \text{ if } y_i = -1$

And it uses kernel transformations that take the general form -x□z = -x□-z(Cui & Curry, 2005). Kernels used frequently are below,

Radial kernel : $K(x_1, x_2) = \exp(-\gamma \parallel x_1 - x_2 \parallel {}^2), \gamma > 0$

Polynomial kernel : $K(x_1, x_2) = (\gamma x^t + c_0)^d, \; d = 1, 2, \ldots$

Logistic kernel : $K(x_1, x_2) = \tanh(\gamma x_1^t + c_0)$

The result of the support vector machine depends on unit cost, slack, and kernel, so it is necessary to find optimized unit cost and slack(Lantz, 2014). In this study, we set slack(Gamma = 1), unit cost(Cost = 9) and radial kernel.

IV. DATA

Our data set is constructed from two sources. We use ERP data from the APC, which is a leading mid-distribution processing center for cabbage. We extract their monthly shipment variable from their database. Another source is the public agricultural wholesale market database. From that database, we can extract monthly transaction volumes from the Garak wholesale market (response variable). The Garak wholesale market is the largest agricultural wholesale market in Korea. Our data period runs from January 2013 through December 2015. We use samples from January 2013 to September 2015 because the remains are used for out-of-sample prediction. The information of our data set is provided in Table 1.

Our data set is constructed from two sources. We use ERP data from the APC, which is a leading mid-distribution processing center for cabbage. We extract their monthly shipment variable from their database. Another source is the public agricultural wholesale market database. From that database, we can extract monthly

transaction volumes from the Garak wholesale market (response variable). The Garak wholesale market is the largest agricultural wholesale market in Korea. Our data period runs from January 2013 through December 2015. We use samples from January 2013 to September 2015 because the remains are used for out-of-sample prediction. The information of our data set is provided in Table 1.

Variable	Description(monthly base)
output_lag	1 lag for shipment amount of APC
input_garak	Input amount of Garak wholsale market
garak_lag	1 lag for Input amount of Garak wholsale market

Table 1. Description of variables

Each variable correlates strongly with each other. Table 2 shows the correlation matrix for each variable.

	input_garak	output_lag	garak.lag
input_garak	-		
output_lag	0.61	-	
garak_lag	0.33	0.59	-

Table 2. Correlation of variables

The Linear form of our prediction model is shown below. Response variable Y is the input amount from the Garak wholesale market. Explanatory variable X is the shipment amount of APC.

$$Y_t = \alpha + \theta Y_{t-1} + \theta X_{t-1} + \epsilon,$$

V. RESULT

One-step-ahead forecasting performances of the various models were subsequently compared against a benchmark, which is defined as a ‘no-change’ forecast (Smith, 2016). We compare our model using SVM to a few models such as AR(1), VAR, and OLS. The period of our prediction model is based on January 2013 through September 2015 . We compare each model for out-of-sample (October 2015 to December 2015) prediction.

Relative accuracy of the individual models is assessed through the root mean square error (RMSE) and mean absolute percentage error (MAPE). We set OLS as a bench mark (BM) model and we compare the other models based on a ratio calculated from MAPE.

Table 3 presents the results of the out-of-sample prediction. The results show that the prediction performance of SVM is better than that of the other three methods. SVM has the lowest RMSE and MAPE. We expect that the results can be applied to predict domestic cabbage volumes and ERP dashboard for the top management at the APC.

(unit : kg)

Period	Garak_M	OLS	SVM	AR(1)	VAR
2015/10	40,534,094	41,250,358	41,178,180	46,388,069	48,914,499
2015/11	66,382,652	45,822,283	62,251,317	48,248,476	47,268,290
2015/12	60,540,662	44,103,219	60,673,832	47,613,885	45,958,139
RMSE	-	15,203,414	2,415,265	13,294,346	14,699,660
MAPE	-	0.199	0.027	0.210	0.245
MAPE ratio to OLS	-	1	0.136	1.055	1.231

Table 3. Out-of-sample performance

VI. CONCLUSION

Given the government’s high interests in controlling agricultural product prices, this study aimed to provide assessment of usefulness of data from the APC and prediction performance of SVM. APC plays an important role in regional distribution of agricultural products. Thus, APC data represents regional productivity of agricultural products and is useful for predicting agricultural products volumes. Actually, some success was found in our study regarding the prediction of agricultural products volumes using APC ERP data.

This study has two limitations. First, the period of data is only for thirty-six months (36 rows). Thus, it is not sufficient to assess out-of-sample performance (only three months). Second, market share of the APC in the domestic cabbage market is less than 10%. To predict cabbage volumes more precisely, we need APCs that have more than 20% market share.

For future research, we suggest two topics. First, it is necessary to use weather data as explanatory variables. Weather conditions are the most important factors for productivity of agricultural products. Thus, if weather condition data adds to the predictors, the accuracy of predicting performance can be more improved. Second, text mining with agricultural news and reports could be useful for predicting volumes. Media has been proven as useful source for prediction (e.g., Trusov et al., 2009). Using latent semantic analysis (LSA) and latent Dirichlet allocation (LDA), which are powerful methods of finding patterns in a large text data set, we expect to predict agricultural product prices/volumes.

ACKNOWLEDGMENTS

This research was supported by 'Agricultural Biotechnology Development Program', Ministry of Agriculture, Food and Rural Affairs.

REFERENCES

[1]. B. Lantz, Editor, Machine learning with R(Packt Publishing Ltd., Birmingham, 2015).
[2]. D Cui and D Curry, Prediction in Marketing Using the Support Vector Machine, Marketing Science, 24(4), 2005, 595-615.
[3]. M. H.Heo, Applied data analysis using R(FreeAcademy, Seoul, 2014).
[4]. M Trusov, R. E. Bucklin and K. Pauwels. Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site, Journal of Marketing, 73(5), (2009, 90-102.
[5]. P. Smith, Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data, Journal of Forecasting, 35(3), 2016, 263-284.