Developing Forecasting Model of Vegetable Price based on Climate Big Data

Do-il Yoo
Department of Agricultural Economics
Chungbuk National University
e-mail: d1yoo@chungbuk.ac.kr

Selected Poster prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Joint Annual Meeting, San Francisco, CA, July 26-28

Developing Forecasting Model of Vegetable Price based on Climate Big Data

Do-il Yoo
Department of Agricultural Economics, Chungbuk National University
E-mail: d1yoo@chungbuk.ac.kr

CHUNGBUK NATIONAL UNIVERSITY

Research Background

Big data is one of the most discussed topics in recent economic and business sectors with explosive applications of information and communication technologies (ICT). Literally, it requires a tremendous capacity to restore and process data gathered from various on- and off-line sources such as social network services (SNS), commercial websites, meteorological satellites, global positioning system (GPS), and so on. Efforts to apply Big Data to industry are made in every field. Marketing is a typical field in utilizing Big Data processing to predict consumer tastes and to develop appropriate strategies. In terms of agricultural production, Big Data is mainly applied to management of agricultural machinery, warehouse, and weather information. However, such applications are at an early stage in the field of agriculture. Besides, there are little studies applying Big Data to the field of agricultural economics. Thus, it would be timely to develop economic models applying Big Data.

Research Object

The object of this study is to develop a forecasting model based on a Big Data processing. This study focuses on the forecasting of vegetable price. The reason for selecting a vegetable as the analysis target is as follows; first, vegetable is consistently and frequently purchased by consumers rather than other processed goods or fruits. Second, vegetable price is unstable because of seasonality, import from foreign countries, and unbalanced market supply-demand. Third, current forecasting system for agricultural products is well constructed for vegetable supplies. In Korea, KREI (Korea Rural Economic Institute) Agricultural Outlook Center (http://aglook.krei.re.kr) plays a key role to forecast production and price trend for vegetables with extensive data sources. Fourth, on-line marketing for vegetable is actively being made as delivery is relatively easier than other agricultural products. Those reasons facilitate gathering data from on-line sources like SNS or websites, helping construct a forecasting model reflecting Big Data concerning vegetable price. Specifically, we focus on forecasting short-term and long-term prices of onion and napa cabbage based on climate big data.

Previous Literature

Previous literature concerning Korean vegetable price forecasting can be categorized as two branches. One is based on traditional partial equilibrium model (Kim et al., 2013). The other is to use time series models like ARMA, ARIMA, GARCH, and so on (Kim, 2005; Yoon and Yang, 2004). Here, we develop models considering both economic structure from partial equilibrium model and time-trend from time-series models. We adopt a Bayesian structural time series model (BSTS) suggested by Scott and Varian (2013) and a vector autoregressive model (VAR).

Model

Bayesian Structural Time Series (BSTS) Model

(1) Observation Equation

Observation equation links observed data with unobserved latent state variables

Observed data — Unobserved latent state

$$y_t = Z_t^T \alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H_t)$$

y_t: observation t in a real-valued time series

Daily price of vegetable (onion and napa cabbage)

Z_t: observation vectors

Daily climate factors (temperature, precipitation, sunshine duration) in chief producing districts

(2) Transition Equation

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \qquad \eta_t \sim N(0, Q_t)$$

α_t: unobserved latent state variable; trend and seasonality

T_t: transition (square) matrix

R_t: a mix of known values(e.g., 0 and 1) or unknown parameters

η_t: independent components of Gaussian random noise

Q_t: variance matrix

Economic structural equation

(3) Default Model

$$y_t = \mu_t + \tau_t + \beta^T x_t + \varepsilon_t$$

$$\left.\begin{aligned} \mu_t &= \mu_{t-1} + \delta_{t-1} + u_t \\ \delta_t &= \delta_{t-1} + v_t \\ \tau_t &= -\textstyle\sum_{s=1}^{S-1} \tau_{t-s} + w_t \end{aligned}\right\} \text{typical time series model}$$

μ_t: current level of the trend
δ_t: current slope of the trend
τ_t: seasonality
$\beta^T x_t$: regression effects based on climate factors
$(u_t, v_t, w_t)=\eta_t$: independent components of Gaussian random noise
Parameters: β, σ_ϵ^2, σ_u^2, σ_v^2, σ_w^2

Following Durbin and Koopman (2002), associated parameters are estimated by Spike and Slab regression. Price forecasting is performed by Markov Chain Monte Carlo Simulation (MCMC).

(4) Empirical Model

Model	Trend	Seasonality	Regressors	Note
BSTS I	O	X	X	Pure time series
BSTS II	O	O	X	Pure time series
BSTS III	O	O	O temperature, precipitation, sunshine duration	Structural time series
BSTS IV	O	O	O temperature, precipitation, sunshine duration, lowest temperature	Structural time series

Vector Autoregressive (VAR) Model

(1) Model considering Quantity and Climate Factors

$$\begin{bmatrix} y_t^i \\ \mathbf{x}_t^i \end{bmatrix} = A_1^i \begin{bmatrix} y_{t-1}^i \\ \mathbf{x}_{t-1}^i \end{bmatrix} + \cdots + A_p^i \begin{bmatrix} y_{t-p}^i \\ \mathbf{x}_{t-p}^i \end{bmatrix} + u_t^i$$

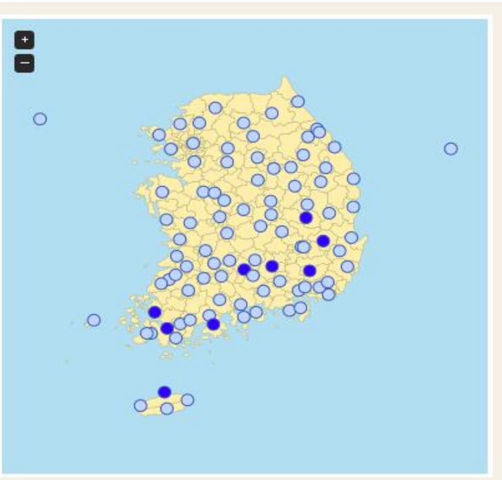
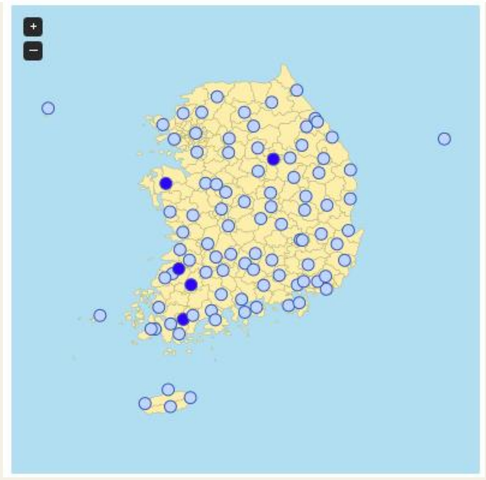
$\mathbf{x}_t^i$: quantity, temperature, precipitation, sunshine duration, lowest temperature

(2) Empirical Models

Model	Regressors
VAR I	quantity, temperature, precipitation, sunshine duration
VAR II	quantity, temperature, precipitation, sunshine duration, lowest temperature

Data

	Onion & Napa Cabbage
Daily Price (KRW/kg)	Wholesale price from KAMIS https://www.kamis.co.kr
Daily Quantity (kg)	Carry-in amount to Seoul Agro-Fisheries & Food Cooperation http://oasis.krei.re.kr
Climate Factors	Daily big data on temperature, precipitation, sunshine duration, and lowest temperature in chief producing district for each vegetable http://sts.kma.go.kr

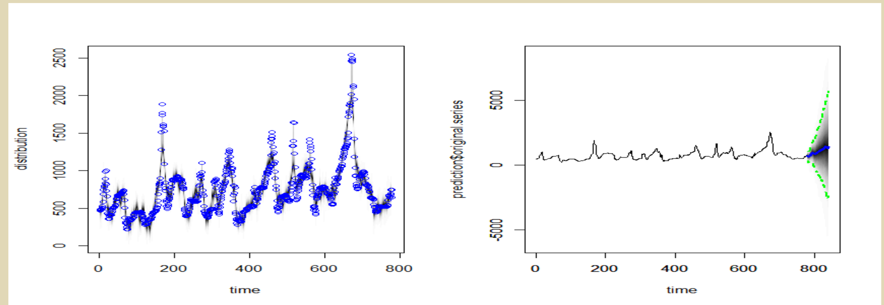
Chief Producing District of Onion for Climate Factors	Chief Producing District of Napa Cabbage for Climate Factors
	

Vegetable products: Onion & napa cabbage
Period: 2000.01.01.~ 2015.05.05. (daily data)
Weekly data: 780 weeks (onion) & 792 weeks (napa cabbage)
Forecasting:2015.05. ~ 2016.06.

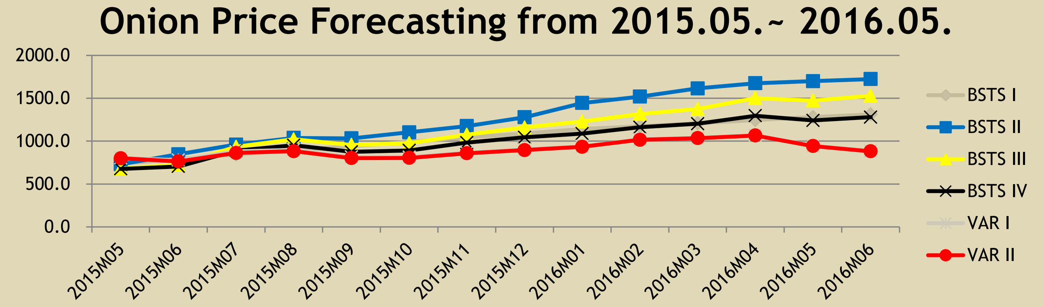
Results

Onion Price Forecasting

Forecasting based on BSTS IV

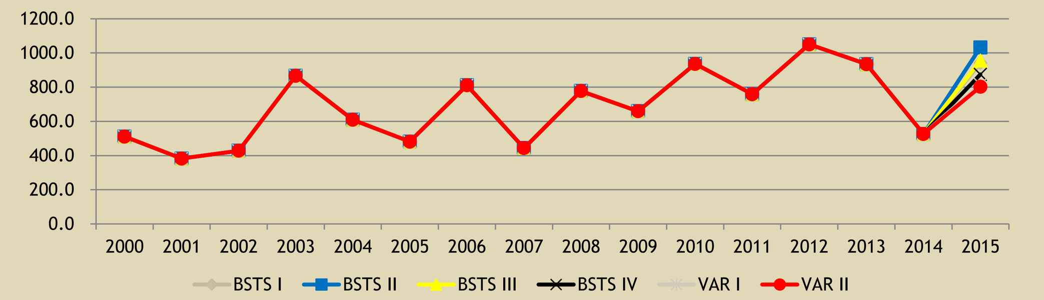


Onion Price Forecasting from 2015.05.~ 2016.05.



Both BSTS and VAR show similar results in the short-term onion price forecasting. However, in the long-term forecasting, VAR seems to be more appropriate in terms of virtual price cycle.

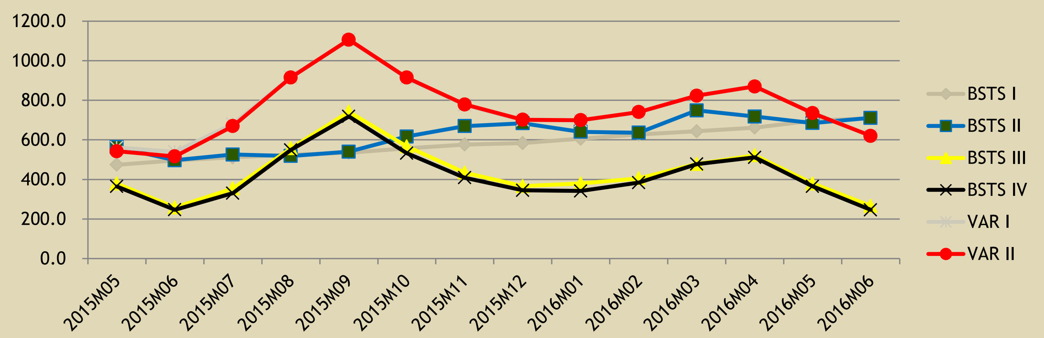
Onion Price for Sept. from 2000 ~ 2015



Both BSTS and VAR forecast onion price for September would increase in 2015 compared to that in 2014. Increments seem to be different across models with higher trends in BSTS and lower trends in VAR.

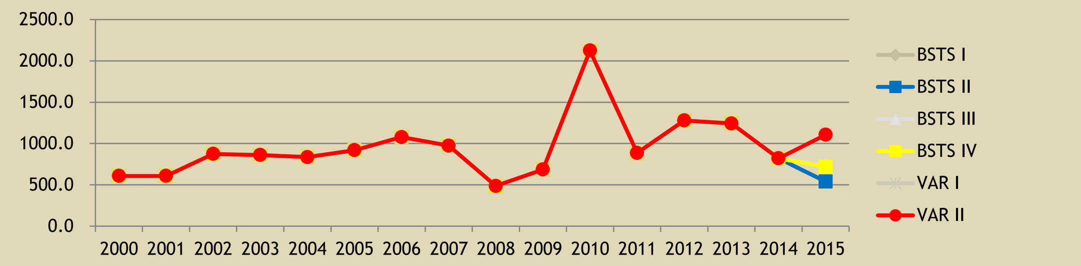
Napa Cabbage Price Forecasting

Napa Cabbage Price Forecasting from 2015.05.~ 2016.05.



Both BSTS and VAR show similar price forecasting patterns. However, except for pure time series models (BSTS I & II), BSTSs forecast price lower than VARs do.

Napa Cabbage Price for Sept. from 2000 ~ 2015



Compared to September price in 2014, pure time series models (BSTS I & II) forecast a sharp drop in Sep. price in 2015. BSTS III and IV predict a nudge down. However, VARs show opposite results of price increment.

Summary and Conclusions

As for onion price forecasting, BSTSs seem to be more appropriate for the short-term prediction, but VARs for the long-term prediction. Considering the short-term forecasting, onion price for September in 2015 would be higher than it was in 2014. For napa cabbage price forecasting, both BSTS except for pure time series models and VAR show similar patterns both in the short-term and in the long-term prediction. However, as for price levels, BSTSs tend to forecast price relatively lower than VARs predict.
Through our analysis, we conclude that whether big data concerning climate factors are considered or not makes significant difference in vegetable price forecasting. Also, it is necessary to apply various forecasting models across products and weather conditions. Future research could be done by considering SVAR, VECM, SVECM, and so forth. Atypical data by text-mining could also be introduced in our models.

References

Durbin, J. and Koopman, S. J., 2002, "A Simple and Efficient Simulation Smoother for STate Space Time Series Analysis," Biometrika, 89: 603-616.

Kim, M., S. H. Han, J. S. Cho, T. W. Kim, and C. S. Lee, 2013, Management and Development Study of Forecast Model for Agricultural Field "KREI-KASMO 2013", Korea Rural Economic Institute.

Kim, B. S., 2005, "A Comparison on Forecasting Performance of the Application Models for Forecasting of Vegetable Prices," Korean Journal of Agricultural Economics, 46(4): 89-113.

Scott, S. and H. Varian, 2013, "Bayesian Variable Selection for Nowcasting Economic Time Series," NBER Working Paper 19567.

Yoon, B. S. and S. Y. Yang, 2004, "A Study on the Day-of-the-Week, Monthly and Intra-Month Effects in Spicy Vegetable Prices," Korean Journal of Agricultural Economics, 45(2): 187-210.