Quest Journals
Journal of Research in Business and Management
Volume4 ~ Issue12 (2017) pp: 05-10
ISSN(Online) : 2347-3002
www.questjournals.org

Research Paper

Forecasting Agri-food Consumption Using Web Search Engine Indices

Ikhoon Jang[1], Young Chan Choe[1]

Received 29Jan, 2017; **A**ccepted18Feb, 2017© The author(s) 2017. **P**ublished with open access at **www.questjournals.org**

ABSTRACT : *In this study, we examine Naver Trend of South Korea (similar to Google Trends), which is a real-time weekly index of the volume of queries that users enter into Naver. We aim to find that these search engine data improves accuracy in forecasting consumption of agri-foods. For a more detail explanation, we classify empirically agri-food items into several specific groups.The results are different by agri-food groups. The agri-food items often used for main or single dishes tend to show significant improvement for both in-sample estimation and out-of-sample forecast. However, agri-food items used mainly as minor ingredients are not significant. Even if the item is a main or single dish, agri-food items frequently consumed out of the home have no significant relationship. However, agri-food items mainly consumed at home are significant. Meanwhile, the significant difference between common products and brand products are not found. Rather, the more important criteria are macro trends. Agri-food items with gradual growth macro trends or long-term fluctuation have a significant relationship. On the contrary, the steady selling of ramen products or dumpling products is not significant.*

Keywords -*Agri-food consumption, Forecasting, Search engine query*

I. INTRODUCTION

Forecasting in the agri-food sector is an important topic. Accurate yield prediction improves farmers' revenue and management stability. Agile demand forecasting helps agri-food producers adjust their production properly. Also, appropriate price forecasting directly affects the profit of agri-food suppliers and distributors, and can be important information for decision making by policy makers. However, previous studies often use only a price variable as a predictor without critical variables such as yield or demand because collecting instant yield or demand data is difficult. Recently, with the progress of technology, the performance of production yield forecasting using climate data has reached a high level. For example, Monsanto has acquired Climate Corporation, a climate data research company for $930 million to maximize farmers' yields. This event means that yield forecasting techniques have matured. However, studies on forecasting agri-food demands are not vibrant due to lack of real-time consumer data.

In this study, we examine Naver Trends of South Korea (similar to Google Trends), which is a real-time weekly index of the volume of queries that users enter into Naver. We aim to discover whether these search engine data improve accuracy in forecasting the consumption of agri-foods. For a more detailed explanation, we classify empirically agri-food items into several specific groups based on four measures: main dish or ingredient, dine out or in, with macro trends or not, and having a brand or not. The prediction model of this study is based on the AR(1) model. For the evaluation of prediction models, we conduct in-sample estimation and out-of-sample forecasting by the rolling window method. The results show that search engine data on agri-foods with specific characteristics can be regarded as an important predictor of agri-food consumption. Furthermore, the results suggest that using search engine data as a substitute for the consumption amount when collecting actual agri-food consumption data is difficult.

[1]Department of Agricultural Economics and Rural Development, Seoul National University, South Korea

II. PREVIOUS RESEARCH

Forecasting is a long-standing research topic in the agri-food sector. Notably, forecasting studies on commodity price have been steadily conducted. The many studies of them focus on various forecasting methods (Allen, 1994). Several studies have used AIDS model reflecting demand system (Kastens&Brester, 1996; Sang &Tonsor, 2015). Meanwhile, most forecasting studies have adopted time series model such as ARIMA, GARCH, and VAR considering lagged endogenous variables (Xu &Turman, 2015; Guney, 2015). However, it is necessary to explore significant variables to enhance prediction model even if to improve prediction accuracy through developing a methodology is important.

Presently, there are several sources of data on real-time economic activity available from private sector companies such as Google, MasterCard, Federal Express, UPS, and Intuit (Choi & Varian, 2012). Many forecasting studies with significant results have used search engine data as an important indicator to improve prediction accuracy. Yahoo’s search query predicted stock market trade volume (Preis et al., 2013). Google Trends data forecasted the number of movie viewers (Hand & Judge, 2011), which is used as an important predictor and indicator of monthly consumption (Vosen& Schmidt, 2011; Kholodilin et al., 2010). In the commodity markets, automobile purchases (Choi & Varian, 2012; Carrière-Swallow &Labbé, 2013) and house prices (Wu &Brynjolfsson, 2014) can be explained by Google Trends data. However, there are few studies on forecasting using Google Trends data in the agri-food and agribusiness sectors.

III. INDENTATIONS AND EQUATIONS

We cannot claim that keyword search behavior has a correlation with agri-food consumption because there are no empirical studies which can explain the relationship between them. Thus, it is not appropriate to investigate the correlation between two variables on the assumption that all agri-food product consumption is equally related with keyword search behavior. In this study, we classify agri-food products into several groups based on four arbitrary measures. In order to derive more meaningful results, the four measures are defined as follows, though these classification measures are not based on the theoretical background.

Web portal services such as Google.com provide the requested information through the search engine. Also, they can link portal users with online shopping malls and facilitate actual purchases in the shopping mall. The search behavior engaged in to get information on agri-food products can be classified into two categories. The first category is the search behavior to collect the information on ingredients or effects of specific agri-food products. The second category is the search for recipes by internet users to do the cooking for themselves.

Plenty of recipe information is provided through users’ blogs as well as the knowledge sharing service of the web portal. In the digital age, printed recipes have rapidly been substituted by online recipe information (Teng et al., 2012). Consequently, recipe information on main dishes is more likely to be searched than recipes related to side dishes because main dishes are more complicated to cook than side dishes. Therefore, main dishes are expected to have a more significant relationship with the amount of keyword searches than side dishes or ingredients. In this context, we define the first agri-food classification measure as whether the search is for main dishes or not.

Most agri-food products purchased in grocery shops, except for gifts, are consumed by households, whereas agri-foods consumed outside of the household are purchased in quantity by restaurants. Thus, agri-food products used for frequently cooked dishes in households are more related with web search behavior than those consumed in restaurants. Thus, we define the second classification measure as whether to dine out or in.

Third, we consider agri-food products with increasing or decreasing trends as a classification measure. With the recent well-being trends, consumption of healthy foods such as chicken breasts and nut products is increasing steadily. Also newly launched agri-food products generally attract attention from consumers. These kinds of agri-foods are likely to be searched in web portal services. Therefore, we can expect that agri-food products with macro trends have a significant relationship with keyword searching volume in a portal service.

The last classification measure is defined as whether the product names are a brand or not. Food manufacture companies such as Sunkist or Delmont want to predict consumption of their own brand products in order to make informed decisions for future management. However, agri-food brands do not require as much information to evaluate their qualities unlike brands of search goods such as DVDs, smartphones, and laptop computers, which probably do not have a meaningful association with keyword search volume in portal services. On the other hand, in the case of the recent popular agri-food brands, many consumers are likely to frequently search for the product in portal services as mentioned earlier. Therefore, keyword searching volume may influence actual consumption differently according to the characteristics of the agri-food brand. Table 1 shows the list of agri-food products selected by group based on the four measures. We investigate the correlation between the keyword search volume and actual consumption of agri-food products by each group. The expected correlation results are suggested in Table 1.

Table 1Theclassification criteria of agri-foods

Measure	Groups	Products	Correlation
1	Main Dish	chicken, sweet pumpkin, green pumpkin, sweet potato, and shiitake mushroom	Positive
	Ingredient	carrot, cucumber, canned tuna, potato, and enoki mushroom	No sig
2	Dine-out	roasted pork belly, fried chicken, and dumpling	No sig
	Dine-in	apple, tangerine, tomato, cherry tomato, strawberry, blueberry, kiwi	Positive
3	Macro Trends	instant rice products, healthy food such as chicken breast and nut products, newly released ramen and dumpling products	Positive
	No Trends	the steady selling of ramen products and dumpling products	No sig
4	Brand	fruit product brands (Sunkist, Delmont, Zespri), rice brands (Kyeong-gi), chichen brand (Harim)	Mixed
	No Brand	fruits in the dine-in group, chicken in main dish group	Positive

IV. DATA AND PREDICTION MODEL

Consumption data is collected from a panel survey of housewives from the Rural Development Administration in South Korea. The data includes five years of daily household food consumption records based on receipt from December 2009 through November 2014. We finally obtained 3.6 million purchase data of 732 panels continuously sustained during five years. The data set for the estimation is processed weekly based on the amount of consumption by targeted agri-food items for 261 weeks.

For search data, we used the relative weekly time series search volume index displayed on the Naver Trends website (http://trend.naver.com/) for each of the products. Naver, Korea’s largest portal site with a 70~80% market share, has provided keyword search volume trend services since 2007.

The prediction model of this study is based on the AR(1) and seasonal AR(1) model of Choi and Varian (2012). The baseline model (1) only consists of the consumption of t period and lagged consumption of t-1 period. And the baseline seasonal model (2) has an additional AR term (lagged consumption of the t-52 period). The test model (3) and the test seasonal model (4) have an additional predictor as a keyword volume index of t period.

$$\cdot\, baseline\ model: \quad y_t = b_1 y_{t-1} + e_t (1)$$

$$\cdot\, baseline\ seasonal\ model: \quad y_t = b_1 y_{t-1} + b_{12} y_{t-52} + e_t \qquad (2)$$

$$\cdot\, test\ model: \quad y_t = b_1 y_{t-1} + b_0 x_t + e_t \qquad (3)$$

$$\cdot\, test\ seasonal\ model: \quad y_t = b_1 y_{t-1} + b_{12} y_{t-52} + b_0 x_t + e_t \qquad (4)$$

It was not certain whether the agri-food products tested in this study had a seasonal characteristic or not. Thus, the model selection between the basic model and the seasonal model is decided by the estimation results of the baseline seasonal mode (2). If the coefficient of the seasonal AR variable of the target agri-food product is significant, then that product can be considered to have a seasonal characteristic. And, we regard that seasonal model as more appropriate to out-of-sample forecasting as well as in-sample estimation.

For the evaluation of prediction models, we first examine that the test model is a significantly improved in-sample fit, and we next investigate whether the trends variables improve out-of-sample forecasting by the rolling window method, which is consistent with the method of Choi and Varian (2012). The rolling window forecast is to estimate the model using the data for periods k through t - 1 and then forecast y_t using y_(t-1), y_(t-12), and the contemporaneous values of the keyword search volume variables as predictors (Choi & Varian, 2012). In data processing, a 3-week moving average is used for the model estimation in order to smooth the short-term fluctuation of the measured value.

The model estimation and prediction performance evaluation is achieved by the agri-food items when considering the classified groups. The first classification measure is whether the agri-food item is a main (or single) dish or minor ingredients of a main dish. The second measure is whether the agri-food item is consumed at home or dining out. The third measure is whether the agri-food item has macro sales trends or no trends. The fourth measure is whether the agri-food item is a common product or brand product.

V. RESULTS

The results of the test model of the in-sample estimation and out-of-sample forecasting are summarized in Table 2-5. The results are different by agri-food groups. The agri-food items often used for main or single dishes including chicken, sweet pumpkin, and shiitake mushroom tend to show positive influences as a predictor for both in-sample estimations and out-of-sample forecasts. However, sweet potato does not show meaningful results in either case, and sweet pumpkin is only significant in in-sample estimations. These unexpected results may be caused by the variety in agri-food dishes. In Korean dishes, sweet potato and sweet pumpkin are not used in various kinds of dishes due to their sweet taste; thus we can assume agri-food consumers do not frequently use web portal services for information searches in the case of dishes cooked by those materials.

Whereas, most agri-food products used mainly as minor ingredients such as carrot, cucumber, canned tuna, and enoki mushroom do not show a significant correlation with keyword search volume. Also there is generally no improvement in the prediction error in out-of–sample forecasting. In the case of the potato, prediction performance is improved; however, the amount of improvement is not large. Therefore, we can consider that keyword search volume of agri-food products used for a main dish's major ingredients can be a meaningful predictor for the prediction of the consumption.

Table 2. The estimation results and prediction error improvement in the first measure

Group	Product	y_{t-1}	sig	y_{t-52}	sig	x_t	sig	R^2	Improve
Main Dish	chicken	0.897	***	-		0.00046	**	0.806	1.7%
	sweet pumpkin	0.760	***	0.049		0.00219	*	0.745	-1.9%
	green pumpkin	0.828	***	0.098	*	-0.00067	***	0.758	5.6%
	sweet potato	0.780	***	0.189	***	0.00035		0.934	-1.5%
	shiitake mushroom	0.633	***	0.109	*	0.00537	***	0.868	12%
Ingre-dients	carrot	0.858	***	-		0.00023		0.734	-0.4%
	cucumber	0.898	***	0.072	***	0.00047		0.961	-6.6%
	canned tuna	0.759	***	-		0.00002		0.572	-0.6%
	enokimushroom	0.784	***	0.082	*	0.00044		0.703	-1.7%
	potato	0.780	***	0.237	***	-0.00072		0.942	1.1%

Table 3 shows the results of the dine-out and dine-in groups. Even if the product is a main dish or ingredient, agri-food products frequently consumed out of the home such as roasted pork belly, fried chicken, and dumpling have no significant relationship. Meanwhile, agri-food items mainly consumed at home such as most fruits except strawberry are significantly related with keyword search volume; however, the improvement in prediction performance in the out-of-sample forecast is not consistent among in-group products. From the time series plot, we can find that the consumption amount of the products such as apple, tangerine, and strawberry has a relatively larger seasonal fluctuation compared with other fruit products. We guess that this characteristic of the product may attenuate the effect of the keyword search volume on the prediction model.

Table 3. The estimation results and prediction error improvement in the second measure

Group	Product	y_{t-1}	sig	y_{t-52}	sig	x_t	sig	R^2	Improve
Dine	roasted pork belly	0.904	***	-		0.000449		0.808	-2.2%

out	friedchicken	0.851	***	-		0.000485		0.772	-1.2%
	dumpling	0.901	***	-		0.000443		0.829	-1.4%
Dine in	apple	0.769	***	0.063		0.002914	***	0.837	-4.4%
	tangerine	0.746	***	0.193	***	0.001933	*	0.965	-1.8%
	tomato	0.811	***	0.135	***	0.001396	***	0.971	4.8%
	cherrytomato	0.792	***	0.119	***	0.002095	***	0.961	7.6%
	strawberry	0.555	***	0.446	***	-0.00073		0.982	-1.3%
	blueberry	0.735	***	0.030	*	0.008801	***	0.788	2.0%
	kiwi	0.869	***	-		0.001017	*	0.817	1.4%

The most consistent results are identified in the measure related with macro trends. Agri-food products with gradual growth macro trends or those with long-term fluctuations have a significant relationship as shown in Table 4. The recent popular products such as healthy food such as chicken breast and nut products, and most popular newly released instant rice products and ramen products have growing trends in the time period of our data. The time series plot shows that the consumption trends have a similar pattern with keyword search volume, which can be interpreted that the early consumer interest toward these products is reflected in web search behavior. On the contrary, the most popular steady selling ramen products and dumpling products are not significant in either the in-sample estimation or out-of-sample forecasting, which means that already well-known products are not searched frequently in web portal services in proportion with the consumed amount in grocery shops.

Table 4. The estimation results and prediction error improvement in the third measure

Group	Product	y_{t-1}	sig	x_t	sig	R^2	Improve
Macro Trends	chicken breast	0.713	***	0.00097	*	0.547	1.7%
	nut products	0.772	***	0.00082	*	0.646	-1.7%
	instant rice product A	0.723	***	0.00313	***	0.723	4.1%
	instant rice product B	0.655	***	0.00194	**	0.518	4.1%
	rice wine	0.819	***	0.00041	*	0.719	1.1%
	new instant ramen A	0.604	***	0.06032	**	0.537	-5.4%
	new instant dumpling A	0.782	***	0.00484	**	0.733	3.7%
No Trends	existing instant ramen A	0.863	***	0.00034		0.755	-0.4%
	existing instant ramen B	0.782	***	0.00019		0.632	-0.7%
	existing instant ramen C	0.850	***	0.00059		0.757	0.9%
	existing instant dumpling A	0.782	***	0.00031		0.620	-1.6%

Finally, we tested the correlation between the actual consumption and keyword search volume of major agri-food brands in Korea. Sunkist, Delmont, and Zespri are the most famous fruit brands. Kyeong-gi rice is the leading rice product brand, and Harim chicken is the leading chicken product brand in Korea. Table 5 reports mixed results by each brand. Brands, except Sunkist and Harim, with no macro trends show no correlation between consumption and keyword search, whereas the consumption amount of Sunkist with gentle decreasing trends and Harim with long-term fluctuations has a significant relationship with keyword search volume. These are consistent with the results of the macro group agri-food products. In other words, the most important factor is whether the consumption amount has macro trends or not in the cases of agri-food brands too.

Table 5. The estimation results and prediction error improvement in the fourth measure

Group	Brand	y_{t-1}	sig	x_t	sig	R^2	Improvement
Major Brand	Sunkist	0.823	***	0.00208	*	0.726	-1.4%
	Delmont	0.825	***	0.00005		0.654	-0.7%
	Zespri	0.719	***	0.01141		0.541	-6.3%
	Kyeong-gi rice	0.764	***	0.00228		0.614	0.0%
	Harim chicken	0.821	***	0.00078	*	0.714	2.5%
No Brand	kiwi	0.869	***	0.001017	*	0.817	1.4%
	chicken	0.897	***	0.00046	**	0.806	1.7%

VI. COUNCLUSION

Although the keyword search volume data from Naver Trends for all of the agri-food items do not improve the performance of consumption forecasting, the search engine data is correlated with actual consumption in several agri-food groups such as the ingredients of main dishes, agri-foods mainly consumed at home, and products with macro consumption trends. In the cases of the first and second groups, we can identify that real time data from search behavior on agri-foods consumption in households is directly linked with actual consumption amounts. As well, the results show that recent interest in food trends or newly launched products can synchronize with actual agri-food consumption as shown in the results of the third group. Therefore, the search engine data of agri-foods with these characteristics can be considered as an important predictor in agri-food consumption forecasting research. In other cases, it is expected that we can utilize the keyword search volume as the explanatory variable to substitute for the consumption amount data if it is difficult to collect actual consumption data on agri-food products.

This study addresses the lack of research on consumer internet search activity data in the agri-food sector. Also we identify the possibility of using search engine data to predict more accurate actual consumption. Although this study has many limitations such as a weak theoretical background, sample selection problems, and a non-rigorous methodology, we expect to extend many further relevant studies, and to help agri-food producers and distributors when they forecast the sales of their agri-products.

Acknowledgements

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2017-H8601-17-1007) supervised by the IITP(Institute for Information & communications Technology Promotion)

REFERENCES

[1] M Ozaki, Y. Adachi, Y. Iwahori, and N. Ishii, Application of fuzzy theory to writer recognition of Chinese characters, International Journal of Modelling and Simulation, 18(2), 1998, 112-116. (8)

[2] P. Allen,Economic Forecasting in Agriculture,International Journal of Forecasting, 13, 1994, 81-135.

[3] H. Choi and H. Varian, Predicting the Present with Google Trends,The Economic Record, 88, 2012, 2–9.

[4] S. Guney, An evaluation of price forecasts of the cattle market under structural changes, *2015 AAEA & WAEA Joint Annual Meeting,* 2015

[5] C. Hand and G. Judge, Searching for the picture: Forecasting UK cinema admissions using Google Trends data, *Applied Economics Letters, 19(11)*,2012, 1051-1055.

[6] K. A. Kholodilin, M.Podstawski, and B. Siliverstovs, Do google searches help in nowcasting private consumption?: A real-time evidence for the US.,*Discussion Papers of DIW Berlin 997*. DIW Berlin, German Institute for Economic Research, 2010

[7] T. Preis, H.S. Moat, and H.E. Stanley, Quantifying Trading Behavior in Financial Markets Using Google Trends, *Scientific Reports3*, 2010, 1-6.

[8] X. Shangand G. T. Tonsor, Comparing Forecasting Ability of Demand System Using Different Data Sources: the Case of U.S. Meat Demand with Food Safety Recalls, *2015 AAEA & WAEA Joint Annual Meeting*, 2015

[9] S. Vosen and T. Schmidt.Forecasting private consumption: Survey- based indicators vs. Google Trends, *Journal of Forecasting 30(6)*, 2011, 565-578.

[10] L. Wu and E. Brynjolfsson.The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales, *Economic Analysis of the Digital Economy*,2014, 89-118.

[11] X. Xu and W. Thurman,Forecasting Local Grain Prices: An Evaluation of Composite Models in 500 Corn Cash Markets, *2015 AAEA & WAEA Joint Annual Meeting*, 2015