Internet Search Volume Data as a Predictor of Consumers’ Daily Food Consumption

Taeyoung Kim¹, Sung-Yong Kim², Kyunsik Lee³

1. Dept. of Food and Resource Economics (Inst. of Agri. & Life Sci.), Gyeongsang National University, Jinju, Republic of Korea, tykim74@gmail.com
2. Dept. of Food and Resource Economics (Inst. of Agri. & Life Sci.), Gyeongsang National University, Jinju, Republic of Korea, sungyong@gnu.ac.kr
3. Dept. of Food and Resource Economics, Gyeongsang National University, Jinju, Republic of Korea, lilycard@gnu.ac.kr

Selected Poster prepared for presentation at the 2018 Agricultural & Applied Economics Association Annual Meeting, Washington, D.C., August 5-August 7

Copyright 2018 by Taeyoung Kim, Sung-Yong Kim, and Kyunsik Lee. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Internet Search Volume Data as a Predictor of Consumers’ Daily Food Consumption

Taeyoung Kim¹ · Sung-Yong Kim¹ · Kyunsik Lee²

Department of Food and Resource Economics (Inst. of Agri. & Life Sci.), Gyeongsang National University, Jinju-si, Republic of Korea
Department of Food and Resource Economics, Gyeongsang National University, Jinju-si, Republic of Korea

Introduction

Need for short-term forecast of food consumption
✓ Food consumption fluctuates by various factors
  • The demand for food is generally stable, while it can change drastically in the short term because of accidents related to food safety, media reports, predictable trends, and social atmosphere.
  • When a consumer demands food at the retail stage is estimated incorrectly, a bullwhip effect occurs in the food supply chain and increases the variation of supply and price in the upper part of the chain.
  • A consumer demand for food is an essential task for those engaged in the food industry, including agricultural producers, food manufacturers, food distributors, and policymakers.

Challenges and Limitations of Household Survey
• The demand for food has been obtained through a survey on households.
  • However, it is becoming difficult to conduct such surveys due to an increase in the number of single-person households and dual-earner households as well as the demand for safeguarding personal information.
  • Moreover, the process of officially deriving statistical survey results requires several months, which has led to a lower short-term estimation.

Internet search data can be used for quick and accurate decision-making in the short term.
• Big data is receiving attention as a new type of tool that can enhance survey-based data by capitalizing on the development of information technology (the internet in particular).
• Problems with using big data: Big data (e.g. scanned data in retail stores, web scraping data in online retail stores, and social media data) cannot be easily obtained, as they are possessed by private firms and can be purchased and utilized for decision making only at a high price.
• Internet search data are considered an alternative to such big data, which can be used free of charge in real time, making it easy to use for short-term prediction of food demand.

Objectives
✓ This study analyzes the effectiveness of Internet search data to speedily estimate food consumption in a retail scenario.
  • This study examines the effectiveness or availability of Internet search data for predicting the consumer demand for food at the retail stage.
  • Particularly, there is a lack of studies that analyze the effectiveness of Internet search data for quickly predicting consumers’ daily food consumption.

Literature reviews
✓ The usefulness of Internet search data proved through numerous studies
  • Studies on applying the data obtained from Internet search engines such as Google for the analysis or estimation of the real economy have been carried out for several years in various fields.
  • However, only a few studies have examined the effectiveness of Internet search data for predicting the consumer demand for food at the retail stage.

Methods

A consumer panel survey conducted by Korean Rural Development Administration
• Daily food consumption records such as purchased items, cost of purchase, quantity purchased, price of purchased unit, and purchase channel
  • Recorded in the transaction receipts of consumer panels between January 2017 and December 2017

Socio-demographic characteristics of individual households such as age, income, marital status, number of household members, and number of food earners

Internet Search Index (ISI) from NAVER search engine
• NAVER is a Korean Internet search website, accounted for 84.03% of the entire volume of Internet searches (85.7% of searches for food & beverage) in Korea as of 2015 (internet trends.co.kr).
• NAVER has been releasing Internet search index (0~100) by date since January 2016.

Analysis items
• Selected items that have high production value or short-term supply and demand fluctuations, and those subject to the government’s intensive management of agricultural prices.
  • Rice (Korean staple), beef, pork, chicken (Westernized diet), important share of agricultural spending, cabbage, radish, garlic, onion (major spicy vegetables of the season of kimchi).

Model selection procedure
✓ Time series model: used to check prediction accuracy and to verify the precedence of Internet Search Index (ISI).
  • Prediction accuracy of the models with/without ISI has been compared based on RMSE and MAPE.
  • Proceeding of ISI over food consumption can verify the usefulness of ISI to exert the causal relationship between ISI and food consumption

Panel model: used to elicit the effectiveness of ISI to consumers’ food consumption

Results & Discussion

Time series model: Usefulness of Internet Search Index

1. Check for stationarity in time series
   • Use unit root tests for the expenditure and Internet search indices across items

2. Prediction Accuracy
   • Decide number of lags
     2.1 Using non-ARIMA
     2.2 Using ARIMA
   • Using VEC Lag Length Selection Criteria

3. Test for causality
   • With the critical values between variables determined through Granger causality test
   • With the number of lags selected, a certain time period is chosen to change the actual expenditure

4. Comparison on RMSE and MAPE

Step 2: Test an effectiveness of internet search index using panel model

Impact of internet search index on Food Consumption
✓ Panel Regression (Random Effect)
✓ Panel Instrumental Variables Regression (Random Effect)
✓ Panel Instrumental Variables Tobit (Mixed Effect)

1. Check for stationarity in time series
   • Use unit root tests for the expenditure and Internet search indices across items

2. Prediction Accuracy
   • Decide number of lags
     2.1 Using non-ARIMA
     2.2 Using ARIMA
   • Using VEC Lag Length Selection Criteria

3. Test for causality
   • With the critical values between variables determined through Granger causality test
   • With the number of lags selected, a certain time period is chosen to change the actual expenditure

4. Comparison on RMSE and MAPE

1. Check for stationarity in time series
   • Use unit root tests for the expenditure and Internet search indices across items

2. Prediction Accuracy
   • Decide number of lags
     2.1 Using non-ARIMA
     2.2 Using ARIMA
   • Using VEC Lag Length Selection Criteria

3. Test for causality
   • With the critical values between variables determined through Granger causality test
   • With the number of lags selected, a certain time period is chosen to change the actual expenditure

4. Comparison on RMSE and MAPE

Panel model: Effectiveness of Internet Search Index

1. Check endogeneity of ISI: Test the Validity of Instrumental Variables.
   • ISI can be an endogenous variable with possible bias such as price & quantity of each item, and seasonal effects.

2. Test for causality
   • Internet search index is significantly and positively affect to consumers’ food consumption in the most cases.
   • ISI can be an effective measure to estimate consumers’ food consumption

Conclusion
✓ ISI can improve the prediction accuracy of the model, and can play a role as a useful measure to predict a short-term food demand
✓ ISI affects positively to consumers’ food consumption regardless of its endogeneity.
✓ ISI can be a useful measure to build the marketing strategies of food industry, and can also contribute to stabilization of supply and demand of government.

Acknowledgement
• This study was supported by Agricultural Technology Research in 2018 (PJ01322302) funded by the Korea Rural Development Administration.