Journal of Applied Finance & Banking, Vol. 12, No. 4, 2022, 73-97
ISSN: 1792-6580 (print version), 1792-6599(online)
https://doi.org/10.47260/jafb/1244
Scientific Press International Limited

Application of the Real-Time Tourism Data in Nowcasting the Service Consumption in Taiwan

Chien-jung Ting[1]*, Yi-Long Hsiao[2]* and Rui-jun Su[3]

Abstract

In this paper, we examined the relationship between tourism and service consumption in Taiwan. The service consumption in Taiwan is nowcasted with the real-time tourism data in Google Trends database. We used the high-frequency internet-searching tourism data to predict the low-frequency service consumption data, for the real-time data with rich information could enhance prediction accuracy. Applying the Principal Components Analysis (PCA), we used the internet-searching tourism keywords in Google Trends database to construct the diffusion indices. Following the classification of the tourism keywords in Matsumoto et al. (2013), we classified those keywords into five groups and twenty-nine classifications. We focused on the reciprocal reactions between those diffusion indices with service consumption to conclude which component has higher influence on service consumption in Taiwan. Our empirical results indicated that the keywords in "Recreational areas, and Travel-related" group have significant effects on service consumption in Taiwan via nowcasting. Among the components of those diffusion indices, "Farm, Travel insurance, and Visitor center" are important variables with higher weights in common.

JEL classification numbers: C60, C80, E01, E2, E60.
Keywords: Nowcasting, the Principal Components Analysis (PCA), Service Consumption, Tourism.

[1*] Department of Leisure, Recreation, and Tourism Management, Southern Taiwan University of Science and Technology, Taiwan. *Corresponding Author
[2*] Department of Finance, National Dong Hwa University, Taiwan. *Corresponding Author.
[3] Department of Leisure, Recreation, and Tourism Management, Southern Taiwan University of Science and Technology, Taiwan.

Article Info: *Received:* March 9, 2022. *Revised:* April 5, 2022.
Published online: April 11, 2022.

1. Introduction

Tourism industries play an important role in Taiwan's economy. The Taiwan Tourism Bureau (MOTC) proposed two major axes in new tourism strategies, Tourism 2020, which focused on exploiting markets and revitalizing tourism. The tourism statistics in Taiwan Tourism Bureau (MOTC) demonstrates that the number of domestic-travel visitors amounts to 150 million times, showing that traveling has already been part of daily life for people. In Taiwan, the domestic major recreational activities are "Nature appreciation activities, Other leisure activities, and Gourmet events," which accounts for 65.7%, 54.6%, 48.5%, respectively.

Since the outbreak of COVID-19, the accommodation industry, transport industry, and travel agency in Taiwan have suffered from huge adverse effects. People reduced their overseas tourism, and the local tours were in turn promoted. Taiwan government put forward several subsidy programs, which appealed people to participate the local tours. In July 2020, Taiwan Tourism Bureau (MOTC) proposed "Feeling-Safe Travel Subsidy" to stimulate traveling consumption. Over four million citizens had participated in this project by mid-August 2020, and the anticipated tourism receipts reached more than seven times. During 2020, the local tours were marketable until May 2021.

Preparing for traveling in advance, people are used to looking up travel-related keywords online, such as searching information of travel agency, travel insurance, or route planning with popular attractions. From the statistical data of Google Trends database, domestic-travel related keywords account for 74% of the internet-searching tourism keywords in Taiwan. The popular and wide range of traveling attractions cover "Zoo, Amusement Park, Places of interest, Hot spring, Night market, Farm, Historic sites, and Cultural old street." Besides, "Backpacker, Free travel, and Working holiday" are also such a big hit on the internet-searching traveling attractions.

In this paper, we focused on exploring the preferred internet-searching tourism keywords which stimulate tourists to expand their consumption in tours. It is anticipated to figure out which information has prominent effects on service consumption in Taiwan. Following the important internet-searching tourism keywords in literatures, we have collected numerous real-time high frequency data from Google Trends database to proceed the empirical tests. Recently, big data issue has received lots of attention, and the high-frequency internet-searching keywords are applied to nowcast low-frequency dependent variables in literatures for well forecasting economic data and providing multiple decision-making suggestions. In this paper, we adopted high-frequency tourism data to nowcast low-frequency service consumption[4] in Taiwan with numerous real-time data in Google Trends database. The main purposes lie in using abundant real-time information to enhance predictability of service consumption in Taiwan.

---

[4] Based on the classifications of BEA (Bureau of Economic Analysis in U.S. Department of Commerce) and Vosen, and Schmidt (2011), "Private Final Consumption Expenditure" covers "Durable consumption," "non-durable consumption," and "Service consumption." In this paper, we put stress on the "Service consumption."

We used the Principal Components Analysis (PCA) to extract diffusion indices from internet-searching tourism keywords in Google Trends database. Following the classifications in Matsumoto et al. (2013), we classified the internet-searching tourism keywords into five main groups and twenty-nine classifications to construct the diffusion indices. Verifying the predictability of diffusion indices for service consumption in Taiwan, we explored which tourism classification has higher impact on service consumption in Taiwan. For robust check, we used both quarterly "Chinese searching tourism keywords" and "English searching tourism keywords"[5] to nowcast annual service consumption data in Taiwan, and treated service consumption and service consumption -travel class[6][7] as the dependent variables in two different models.

There is currently a lack of literature focusing on nowcasting service consumption in Taiwan via real-time tourism keywords. We look forward to well explaining the correlation between tourism keywords and service consumption in Taiwan with abundant high-frequency information.

2. Literature

Recently, big data issue has been in the spotlight in numerous literatures, and high-frequency internet-searching keywords have been applied to nowcast low-frequency dependent variables for well forecasting economic and providing multiple decision-making suggestions. Since Klein and Park (1994), the high-frequency statistical data have been applied to nowcast the low-frequency data for reducing the predictability difficulties. Then, GDP has been nowcasted with dynamic models and other related techniques in literatures to make researchers to extract useful information (Evans, 2005; Barhoumi et al. 2010; Marcellino et al. 2003; Boivin and Ng, 2005; Bragoli and Fosten, 2016; Chernis and Sekkel, 2017; Chikamatsu et al. 2018; Kabundi et al. 2016; Luciani et al. 2018).

Following Giannone et al. (2005), many literatures used lots of data to nowcast GDP in different countries. Yiu and Chow (2011) used sixteen categories of variables to nowcast GDP in China to conclude that interest rate could be predicted effectively. Using the Kalman Filter in State-Space model, Lahiri and Monokroussos (2013) used the bridge equation and the dynamic factor model to nowcast GDP with diffusion indices, and put stress on its marginal effectiveness and real-time characteristics. Banbura (2011) mixed twenty-four categories of data to construct

---

[5] As shown in the Table 1, those keywords are classified into twenty-nine classifications. In each classification, both "Chinese searching tourism keywords" and "English searching tourism keywords" have exactly the same definition, and the only thing that is changed is people searching tourism keywords in English or Chinese on the internet.

[6] Based on BEA (Bureau of Economic Analysis in U.S. Department of Commerce), we used the definition of "service consumption" covering six classifications. They are separately "Housing, Water, Electricity, Gas and Other Fuels," "Transport," "Communication," "Recreation and Culture," "Restaurants and Hotels," and "Miscellaneous Goods and Services."

[7] In this paper, we also adopted the "service consumption-travel class" to do the robust check. "Service consumption-travel class" includes three classifications, which are separately" Recreation and Culture," "Restaurants and Hotels," and "Miscellaneous Goods and Services."

forecasting model to predict GDP, and those data frequencies were daily, weekly, monthly, and quarterly, respectively. Since the mixed-frequency data in the same model would result in dimension disasters and imprecise estimates, they applied Kalman filter to reduce the uncertainties. The results showed that the mixed-frequency data indeed improved the forecast accuracy for GDP. Mazzi and Montana (2009) used simultaneous indicators to build model and offered real-time information of economic activities in their paper, which aimed at nowcasting the GDP growth rate.

As supplementary tool, Notini et al. (2012) used monthly data of energy demand, steel production, cement, vehicles, industrial production, and sales to forecast GDP. They used Kalman filter to assess three-month summation data to be 1 quarterly data for acquiring new quarterly data in advance. Their research figured out that in-sample forecasting results were better than central bank's estimating data. Liebermann (2012) used the bridge equation and dynamic factor model to forecast quarterly GDP data in Ireland, and they used fourteen monthly variables in domestic and external economies. Luciani and Ricci (2013) used the Bayesian Dynamic Factor Model (BDFM) to apply monthly data including "PMI, Unemployment Rate, Industrial Production, Employee, Retail Sales, New Orders, Import and Export, and Consumer Confidence Index" in nowcasting the annual GDP growth of Norway. Their research showed that the forecasting performance was equal to the survey of Bloomberg Research. In addition, their nowcasting annual GDP data results were better than the forecast of Bank of Norway, and the results had significant smaller MSE and implied that Bayesian Dynamic Factor Model (BDFM) could effectively forecast via newer information. Summing up the research results in Notini et al. (2012), Liebermann (2012), and Luciani and Ricci (2013), the high-frequency monthly macroeconomic indicators could improve the low-frequency quarterly forecasting.

Some nowcasting literatures had focused on employment rate and stock issue, which were as follows. Askitas and Zimmermann (2009) used the real-time data on internet to forecast economic behavior, and they found the strong correlation between Germany internet-searching keywords and unemployment rates. The internet-searching data could be applied to forecast economic behavior with their abundant information, which showed the forecasting method with higher explanatory power. D'Amuri, and Marcucci, (2010) searched for job-related data in Google Insight for Search database, and transformed those weekly data into monthly ones, and then arranged those data into quarterly data as leading indicator for forecasting US employment rates. They concluded that the increasing searching times of job might reflect the growing unemployment rate. Their result represents that the forecasting effects of internet-searching keywords model are better than the traditional forecasting model for unemployment rate. Takeda and Wakao (2014) explored the relationship between stock names of Japan Nippon index and stock market behavior. They showed the insignificant positive relationship between searching times and stock prices, and the significant relationship between searching times and stock market trading volume. They pointed out the increasing searching times might expand stock market trading volume, but the expanded stock market trading volume cannot represent the stock price being rising.

Some nowcasting literatures focusing on tourism issues are as follows. Matsumoto et al. (2013) examined the influence of internet-searching tourism data on Japan's service consumption around the "Japan's 311 Earthquake." Choi and Varian (2012) explored the tourism issues with internet-searching keywords and compared with the traditional model to conclude which model has higher explanatory power.

Some nowcasting literatures focused on nowcasting service consumption with internet-searching data, such as Vosen and Schmidt (2011, 2012) and Kholodilin et al. (2010). They all found that new style indicators have higher predicting accuracy than traditional ones.

In this paper, the advantages of real-time data are elaborated, and the low-frequency service consumption data in Taiwan is nowcasted with high-frequency tourism data. We look forward to the more precise prediction of service consumption in Taiwan. We therefore concluded five main categories of internet-searching tourism data, namely Transport, Rest, Recreational areas, Grand tour, and Travel-related groups. Those descriptions are as follows.

1) *Transport*. The Transport-related tourism keywords cover six classifications, sequentially "Cruise, Bus, Train, Car rental, Taxi, and Airplane." When traveling in one single county, "Bus, Car rental and Taxi" are highly used. When traveling across counties, "Train" is the best choice. "Airplane" covers both external travel line and domestic travel line. "Cruise" is making a round-the-ocean passenger liner.
2) *Rest*. Rest includes "Hotel, Accommodation, and Resort." In Taiwan, there are nearly one hundred "Hotels" offering tourists commercial facilities to take a break. "Accommodation" is necessary for tourism and the main keywords. "Resort" includes recreation area, combining hotel, restaurant, amusement park, and indoor and outdoor leisure and recreation facilities, which focuses on designated travel and offers tourists large scale and hotel facilities.
3) *Recreational areas*. We divided Recreational areas into seven classifications, which are "Amusement Park, Zoo, Traveling attractions, Places of interest, Hot spring, Night market, and Farm," respectively.
4) *Grand tour*. We divided Grand tour into three classifications, which are "Backpacker, Free travel and Working holiday."
5) *Travel-related.* We divided Travel-related affairs into ten classifications, including "Guidebook, Travel insurance, Subsidy, Travel agency, Travel, Guide, Souvenir, Luggage, Visitor center, and Package tour."

3. Empirical model

In this paper, we nowcasted low-frequency annual service consumption in Taiwan with high-frequency quarterly internet-searching tourism keywords data. The empirical steps are as follows.

3.1 STEP 1: Transforming high-frequency data into low-frequency data

In this paper, we referred to the approach of Klein and Park (1994) to rearrange high-frequency data through ARIMA model. Then, following Giannone et al. (2008), we transformed the quarterly internet-searching tourism keywords data into the annual data. The transformation procedure is as follows.

The information set is composed of n variables,

$\Omega_{v_j} = \left\{X_{it|v_j}; t = 1, \dots . T_{iv_j}; i = 1, \dots, n\right\}.$

Among them, $X_{it|v_j}$ is the individual time series data. i represents n variables. t represents the data frequency, which is quarterly from the first observation to the last one (T_{iv_j}).

At first, we nowcasted annual service consumption data with quarterly data. Assume $\mathrm{y} = \text{the last quarter of each year}$. There are four quarters each year, and it could be represented as $\mathrm{y} = 4\mathrm{k}, \mathrm{k} = 1,2, \dots.$, and k is the observed year. The quarterly data (j) is announced four times each year, and there are four data collections, sequentially the 1st to the 4th quarter. They are represented as Ω_{v_j}, $v = 4k - l$, $l = 0, .\,., 3$

According to the information collection, the estimated Service consumption forecast is the nowcasting estimation. $\mathrm{y}_{4\mathrm{k}}$ is the estimated Service consumption in Taiwan, which is evaluated based on quarterly data information.

$$\mathrm{y}_{4k|v_j} = E\left[y_{4k}|\Omega_{v_j}\right],\ v = 4k - l;\ l = 0, .\,., 3, \tag{1}$$

Equation (1) is the bridge equation, nowcasting annual Service consumption in Taiwan with quarterly data. As aforementioned, we have four quarterly data each year to nowcast four current year data, which are treated as four methods. The four methods are separately Q1 to Q4, sequentially the 1st quarter to the 4th quarter of each year chosen as the quarterly data.

3.2 STEP 2: Index construct reduction-construct diffusion index

Diffusion indices are Principal Components Analysis (PCA) in statistics, which conclude several groups of series to have strongest correlation with variables. Then the series estimated by X are called diffusion indices or common factors. The methods are as follows.

Assume X is the $\mathrm{T} \times \mathrm{N}$ matrix composed of N time series variables, and T is the number of samples. Assume F is a $\mathrm{T} \times \mathrm{k}$ matrix, representing k diffusion indices estimated by X, and the relationship between X and F is as follows.

$$\mathrm{X} = \mathrm{F}\beta + \varepsilon \tag{2}$$

β represents the $\mathrm{k} \times \mathrm{N}$ coefficients matrix, which results from regressing X by the estimated F, or called as the factor loading matrix. ε is the vector of residuals, independent and identical distribution with white noise.

In other words, the distribution of each ε satisfying with $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{V}[\mathrm{vec}(\varepsilon)] = \Omega$, Ω is the symmetric positive-definite matrix of positive diagonal term, with mutually independent ε in different periods. Equation (2) describes the linear relationship between X and F, and its regression error terms satisfying the basic assumption of residuals, as the basis of forecasting single series with diffusion indices in the future. To ensure the asymptotic distribution, all the time series in the X vector should be series without unit roots.

F and β are all the estimated variables in Equation (2), which could not be identified in one estimation method. We have to use the two-step method to estimate F and β, which is to estimate F in the first step, and the best parameter estimator $\hat{\beta}$ will be estimated in the second step. As mentioned above, the estimation in the first step, F could be treated as a set of k diffusion factor series having the strongest correlation with X vector. That is the vector satisfying this condition, which is the solution of minimizing the objective function as follows.

$$\min_F \sum_{i=1}^{N} (X_i - F\beta_i)' (X_i - F\beta_i) \tag{3}$$

X_i and β_i are separately the ith element in X and the ith element in β vector. For solving F and β, we assumed temporarily F is known in equation (2), and then the least square estimator β could be represented as $\widehat{\beta_i} = (F'F)^{-1}(F'X_i)$. Treating $\widehat{\beta_i}$ as β_i of equation (3), we could then rewrite the objective function as

$$\min_F \sum_{i=1}^{N} X' [I - F(F'F)^{-1}F'] X_i,$$

or further

$$\min_F \{trace[X'X - X'F(F'F)^{-1}F'X]\} \tag{4}$$

Trace (•) represents the function of dimensional elements summation in square matrix. Because X'X comes from the sample series, not the estimated parameter, the solution to equation (3) would be the same with which in the equation (5).

$$\max_F \{trace[X'F(F'F)^{-1}F'X]\} \tag{5}$$

Stock and Watson (1998a, 1998b) adopted the proof in Connor and Korajczyk (1986, 1993), and concluded that the solution to F in equation (5) was the eigenvector corresponded by the maximizing k eigenvectors. In this way, the matrix $\hat{\mathrm{F}}$ is composed of k eigenvectors in $\mathrm{T} \times 1$, which was the estimated diffusion indices. Taking the estimated $\hat{\mathrm{F}}$ into equation (2), the least square estimator of factor loading matrix β was derived as follows.

$$\hat{\beta}_{OLS} = (F'F)^{-1}(F'X) \tag{6}$$

In this paper, we estimated several Taiwan tourism diffusion indices. The procedure is as follows.

At first, all the time series data were examined with Augmented Dickey-Fuller unit root tests, identifying these variables satisfying transformation without unit root. If the original series should be differenced to be stationary, the log difference will be adopted to build matrix X. The maximized k eigenvalues of XX' and its corresponding eigenvector would be estimated by programming-calculation, which were the estimated diffusion indices.

In this paper, we applied the tourism-related variables to construct the tourism diffusion indices, which is the "internet-searching tourism data indices." The n diffusion indices extracted from the tourism-related variables are the common factors of service consumption in Taiwan. According to Stock and Watson (1998a,1998b, 2002a, 2002b), the n diffusion indices extracted from the tourism related variables, which are named separately DF1 to DFn, are the possible common factors for service consumption in Taiwan.

We want to explore the prediction performance of those tourism-related diffusion indices. Through decomposing the weights of diffusion indices, we could find out which variable have much more impact on service consumption in Taiwan and conclude which tourism common factor has prominent impact on service consumption in Taiwan via the prediction performance.

3.3 STEP 3: Constructing the bridge equation

We set up the bridge equation related to diffusion indices with service consumption in Taiwan, and then added up the ARMA terms to solve the serial correlation problem in error terms to raise up the explanatory power. Following Giannone et al. (2008) and Giannone et al. (2010), we used the Kalman filter and nowcasted service consumption in Taiwan through the bridge equation and the dynamic factor model. At first, there are many variables in the information sets, which might have the curse of dimensionality and imprecise estimates. Hence, the Principal Components Analysis (PCA) was applied to estimate the common factors, which are

$$x_{i,t|v_j} = \mu_i + \lambda_{i1} f_{1,t} + .. + \lambda_{i,r} f_{r,t} + \xi_{i,t|v_j}, i = 1, \dots, n$$

μ_i is the intercept, and $\chi_{it} \equiv \lambda_{i1} f_{1k} + \cdots + \lambda_{ir} f_{rk}$ are the common factors. Represented by matrix forms, $x_{t|v_j} = \mu + \Lambda \mathrm{F}_t + \xi_{t|v_j}$. Among them, $\mathrm{x}_t = (x_{1t|v_j}, \dots, x_{nt|v_j})'$, $\xi_{\mathrm{t}|v_j} = (\xi_{1t|v_j}, \dots, \xi_{nt|v_j})'$, $\mathrm{F_t} = (f_{1t}, \dots, f_{rt})'$, Λ is a $\mathrm{n} \times \mathrm{r}$ factor loading matrix.

$$\hat{y}_{4k|v_j} = \alpha + \beta' \hat{F}_{4K|v_j} \text{ , } \hat{F}_{4K|v_j} = \mathrm{E}[\mathrm{F}_{4k} | \Omega_{v_j}; \mu] \text{ for } v = 4k - l; l = 0, .., 3 \qquad (7)$$

$$\mathrm{F}_t = \mathrm{A}F_{t-1} + \mathrm{Bu}_t, \quad u_t \sim WN(0, I_q) \tag{8}$$

B is the $r \times q$ matrix of full rank; A is the $r \times r$ matrix with eigenvalues larger than 1; u_t is the white noise of common factors.
Equation (7) is the bridge equation. $\hat{y}_{4k|v_j}$ is the estimated service consumption in Taiwan, which is the linear relationship between service consumption nowcasting estimates with the common factors. Giannone et al. (2008) assumed the common factor dynamics satisfies the VAR form, which is equation (8). They used the Kalman filter to estimate the common factors in two steps and brought them into equation (7) to get the nowcasting estimates. The two-step procedures are as follows.

STEP1: Use the Principal Components Analysis (PCA) to find the common factors, and then regress the common factors and dependent variables to get the estimated parameter in the state-space model.
STEP2: Use the Kalman filter to re-estimate the common factors and dependent variable.

Bai (2003), Bai and Ng (2002), Forni et al. (2005) and Stock and Watson (2002a) had already found the diffusion indices which were estimated from observable variables, and they were consistent with the unobservable common factors estimated from the two-step procedures. That is why we adopted the way of estimating diffusion indices in Stock and Watson (2002a), which was proxied as the common factor for nowcasting service consumption.

3.4 STEP 4: VAR model for diffusion indices

The estimated diffusion indices represent the k sequences having the strongest correlation with the original series components of X. In this paper, we followed the approach in previous section to estimate the diffusion indices for four individual quarters. Although the diffusion indices estimation is a purely statistical approach without economical granger causality background, we could further explore the relationship between service consumption patterns with different diffusion indices and understand the implications of nowcasting annual service consumption in Taiwan via diffusion indices.
Using diffusion indices to construct the VAR model to predict the future periods of objective variables, we concluded that X_{it} is the observation of the ith economic variable in t period, N is the number of economic variables, and γ is the estimating parameter originated from maximizing the DF_t variations. Under the limitation of variable standardization, satisfying with $\sum_{i=1}^{N} \gamma^2 = 1$, $DF_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + \cdots + \gamma_N x_{Nt}$ is called the first Principal Component, which is DF_t.
In this step, we used the VAR model to measure the relationship between the diffusion indices and objective variables, representing the VAR (p) model as AR (1) form,

$$Z_t = AZ_{t-1} + \varepsilon_t \tag{9}$$

Let $\binom{x_t}{y_t} \in R^2$, and assume $(x_t y_t)'$to be the VAR (p) form. Hence, $E_t(Z_{t+j}) = A^j Z_t$ and $E_t(y_{t+j}) = [010 \dots .0]A^j Z_t$, among them, $e_2 = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{2p\times 1}$

In addition, $x_t = [010 \cdots 0]Z_t = e_2' Z_t$. Combining the terms before, we can conclude,

$$e_2' Z_t = x_t = \sum_{j=1}^{\infty} \beta^j E_t(y_{t+j}) = \sum_{j=1}^{\infty} \beta^j e_2' A^j Z_t \tag{10}$$

Arranging them, we get $e_2' Z_t = e_2' \left(\sum_{j=1}^{\infty} \beta^j A^j \right) Z_t$.
And $e_1' = e_2' \beta A (I - \beta A)^{-1}$, that is $e_1'(I - \beta A) = e_2' \beta A$. The hypothesis testing for coefficient matrix A in VAR (p) could use the Wald test,

$$\begin{aligned} x_t &= \Phi_{11} x_{t-1} + \Phi_{12} y_{t-1} + \varepsilon_{xt} \\ y_t &= \Phi_{21} x_{t-1} + \Phi_{22} y_{t-1} + \varepsilon_{yt} \end{aligned} \tag{11}$$

Then,

$$\underbrace{\begin{bmatrix} x_t \\ y_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix}}_{Z_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \end{bmatrix}}_{\varepsilon_t}$$

$$\underbrace{e_1'(I - \beta A)}_{1\times 2} = [1 \quad 0](I - \beta A) = [I - \beta\Phi_{11} \quad -\beta\Phi_{12}]$$

Therefore,

$$\underbrace{e_2' \beta A}_{1\times 2} = [0 \quad 1]\beta A = [\beta\Phi_{21} \quad \beta\Phi_{22}]$$

And the null hypothesis is,

$$\begin{cases} 1 - \beta\Phi_{11} = \beta\Phi_{21} \\ -\beta\Phi_{12} = \beta\Phi_{22} \end{cases}, \text{ or } \begin{cases} \Phi_{11} + \Phi_{21} = \frac{1}{\beta} \\ \Phi_{12} + \Phi_{22} = 0 \end{cases}$$

If the null hypothesis is rejected, then the model is failed.

We aimed at understanding the correlation between the diffusion indices of four individual quarterly data and annual service consumption data in Taiwan. We used the VAR model to examine the individual diffusion index of each individual quarter to verify their lead, lag, or feedback relationship. The model is as follows.

$$
\begin{aligned}
y_t &= \alpha_1 + \beta_1 y_{t-1} + \beta_2 DF1_{t-1} + \beta_3 DF2_{t-1} + \beta_4 DF3_{t-1} + \varepsilon_{yt} \\
DF1_t &= \alpha_2 + \beta_5 y_{t-1} + \beta_6 DF1_{t-1} + \beta_7 DF2_{t-1} + \beta_8 DF3_{t-1} + \varepsilon_{DF1t} \\
DF2_t &= \alpha_3 + \beta_9 y_{t-1} + \beta_{10} DF1_{t-1} + \beta_{11} DF2_{t-1} + \beta_{12} DF3_{t-1} + \varepsilon_{DF2t} \\
DF3_t &= \alpha_4 + \beta_{13} y_{t-1} + \beta_{14} DF1_{t-1} + \beta_{15} DF2_{t-1} + \beta_{16} DF3_{t-1} + \varepsilon_{DF3t}
\end{aligned}
$$

$$
\begin{bmatrix} y_t \\ DF1_t \\ DF2_t \\ DF3_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \beta_5 & \beta_6 & \beta_7 & \beta_8 \\ \beta_9 & \beta_{10} & \beta_{11} & \beta_{12} \\ \beta_{13} & \beta_{14} & \beta_{15} & \beta_{16} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ DF1_{t-1} \\ DF2_{t-1} \\ DF3_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{DF1_t} \\ \varepsilon_{DF2_t} \\ \varepsilon_{DF3_t} \end{bmatrix} \tag{12}
$$

y_t is the annulled data of service consumption in Taiwan. $DF1_t$,$DF2_t$,$DF3_t$ are the diffusion indices for each individual quarter.

4. Data and Empirical results

4.1 Data description

In this paper, we used the high-frequency internet-searching tourism keywords data to nowcast the low-frequency annual service consumption in Taiwan. The data descriptions and sources are listed in Table 1. Owing to the fact that the data in Google Trends database starts from January 2004, the data in this paper covers from 2004 to 2020. We adopted quarterly internet-searching tourism keywords data from Google Trends database. For robust check, we compared two kinds of "Private Final Consumption Expenditure" in Taiwan, namingly service consumption and service consumption-travel class, which are sourced from "Directorate General of Budget, Accounting and Statistics, Executive Yuan, Taiwan."

In Table 2, the description statistics results of "Chinese searching tourism keywords" data showed that "Bus" has highest internet-searching times, followed by "Resort," "Attraction," and "Travel agency." The description statistics results of "English searching tourism keywords" data demonstrated that "Bus" has highest internet-searching times, followed by "Airplane." In brief, whether people search the tourism information in Taiwan via "Chinese searching keywords" or "English searching keywords," they care about the transport most.

Table1: Data description and classifications

Variables				Source	Period	Frequency
Dependent variables	Service consumption, Service consumption-travel class			Directorate General of Budget, Accounting and Statistics, Executive Yuan, Taiwan	2004~2020	Annual
Real-time Tourism-related variables	Groups	Classifications		Google Trends database	2004Q1~2020 Q4	Quarterly
		Chinese searching tourism keywords	English searching tourism keywords			
	Transport	Cruises	Cruises			
		Bus	Bus			
		Rail	Rail			
		Car rental	Car rental			
		Taxi	Taxi			
		Airplane	Airplane			
	Rest	Hotel	Hotel			
		Accommodations	Accommodations			
		Resort	Resort			
	Recreational areas	Amusement Park	Amusement Park			
		Zoo	Zoo			
		Attraction	Attraction			
		Places of interest	Places of interest			
		Hot spring	Hot spring			
		Night market	Night market			
		Farm	Farm			
	Grand tour	Backpacker	Backpacker			
		Free travel	Free travel			
		Working holiday	Working holiday			
	Travel-related	Guide book	Guide book			
		Travel insurance	Travel insurance			
		Subsidy	Subsidy			
		Travel agency	Travel agency			
		Travel	Travel			
		Guide	Guide			
		Souvenir	Souvenir			
		Luggage	Luggage			
		Visitor center	Visitor center			
		Package tour	Package tour			

Table 2: Description Statistics

Chinese searching tourism keywords								**English searching tourism keywords**							
Classifications	Mean	Median	Maximum	Minimum	Standard Deviation	Jarque-Bera	Probability	Classifications	Mean	Median	Maximum	Minimum	Standard Deviation	Jarque-Bera	Probability
Cruises	28.67	26	100	5	16.34	51.11	0	Cruises	22.82	17	100	0	18.89	90.06	0
Bus	***63.27***	62	100	28	16.06	9.44	0.01	***Bus***	***62.62***	60	100	32	15.44	9.95	0.01
Rail	38.39	30	100	12	20.74	52.57	0	Rail	43.13	35	100	14	19.61	21.85	0
Car Rental	26.1	23	100	0	18.5	88.1	0	Car Rental	41.31	33	100	13	21.26	18.23	0
Taxi	42.51	34	100	10	21.92	17.29	0	Taxi	45.57	45	100	14	13.87	43.5	0
Airplane	45.59	44	100	0	19.44	4.36	0.1	***Airplane***	***61.45***	59	100	43	9.24	57.75	0
Hotel	43.28	43	100	14	11.29	128.35	0	Hotel	70.68	68	100	28	12.22	18.2	0
Accommodation	36.4	31	100	10	18.78	15.02	0	Accommodation	26.38	23	100	0	14.38	199.21	0
Resort	***51.16***	48	100	24	12.62	61.2	0	Resort	39.74	36	100	12	14.17	98.81	0
Amusement park	21.78	21	100	3	10.74	3082.32	0	Amusement park	16.52	13	100	0	15.99	420.6	0
Zoo	43.27	41	100	17	12.4	173.94	0	Zoo	38.83	36	100	14	12.83	89.85	0
Places of interest	9.39	4	100	0	14.1	2074.25	0	Places of interest	7.47	3	100	0	15.6	1602.91	0
Attraction	***51.38***	50	100	14	15.71	1.94	0.38	Attraction	19.28	15	100	0	13.95	1124.16	0
Hot Spring	47.62	39	100	24	18.82	21.62	0	Hot Spring	16.43	16	100	0	9.91	5180.23	0
Night Market	45.75	50	100	4	28.17	17.87	0	Night Market	24.96	20	100	0	19.65	73.56	0
Farm	21.1	18	100	13	11.76	5127.76	0	Farm	20.74	19	100	0	12.41	3550.56	0
Backpacker	44.98	48	100	0	22.44	2.06	0.36	Backpacker	32.1	28	100	0	22.12	11.6	0
Free travel	46.42	41	100	6	19.62	9.39	0.01	Free travel	12.38	6	100	0	20.19	521.58	0
Working holiday	38.49	40	100	0	24.87	9.19	0.01	Working holiday	23.72	22	100	0	15.07	139.5	0
Guidebook	12.65	6	100	0	15.64	1075.26	0	Guidebook	10.5	5	100	0	17.24	952.9	0
Travel insurance	4.15	4	100	0	7.13	221993.7	0	Travel insurance	7.84	4	100	0	11.46	4224.77	0
Subsidy	18.08	13	100	3	15.78	774.62	0	Subsidy	12.27	7	100	0	15.74	466.1	0
Travel agency	***56.1***	51	100	7	19.1	2.85	0.24	Travel agency	17.51	10	100	0	18.28	243.07	0
Travel	20.12	20	100	11	6.88	72102.8	0	Travel	47.73	39	100	22	20	19.27	0
Guide	46.38	44	100	13	14.32	25.45	0	Guide	20.54	19	100	6	13.13	591.7	0
Souvenir	14.48	10	100	0	13.25	1406.85	0	Souvenir	19.12	16	100	0	15.38	500.66	0
Luggage	39.52	27	100	2	31.22	23	0	Luggage	43.44	44	100	0	17.44	11.12	0
Visitor Center	30.84	32	100	0	13.76	119.96	0	Visitor Center	7.09	3	100	0	14.64	2452.72	0
Package tour	28.79	22	100	0	20.04	48.42	0	Package tour	10.07	6	100	0	14.34	1057.86	0

Table 3: The component factor loadings of diffusion indices-Chinese searching tourism keywords

Dependent variables	**Part A**						**Part B**	
	Service consumption						**Service consumption-travel class**	
Quarterly data	Q2		Q3		Q4		Q1	
Diffusion index	DF1	DF3	DF1	DF2	DF1	DF2	DF2	DF3
Components of Diffusion index (keywords)	Ranking of components (by weight)							
Cruises								
Bus (3)	**2**			**2**	**1**			
Rail								
Car Rental								
Taxi (3)	**1**					**1**	**1**	
Airplane								
Hotel								
Accommodation (1)								**2**
Resort								
Amusement park								
Zoo								
Places of interest								
Attraction								
Hot Spring								
Night Market								**1**
Farm (1)								
Backpacker								
Free travel (2)		**3**				**1**		
Working holiday								
Guidebook (3)	**2**			**2**		**2**		
Travel insurance (2)		**1**			**2**			
Subsidy								
Travel agency								
Travel (2)		**2**		**1**				
Guide (1)								**1**
Souvenir								
Luggage								
Visitor Center (2)			**1**				**2**	
Package tour								
First r eigenvalues of the correlation matrix:	14.11	4.11	9.63	8.18	14.16	5.21	4.76	2.63
Variability explained	0.81	0.81	0.71	0.71	0.74	0.74	0.75	0.75

Source: The authors. 1. In the first column, the number in the parentheses after those components of diffusion index represents how many times that the keyword has ever been the top 3 components in each diffusion index. 2. In table 3, we choose those diffusion indices having significant effects on dependent variables in VAR tests in Table 5. 3. Based on the ranking of components, we conclude the higher-weight variables in bold italics, which are "Bus, Taxi, Accommodation, Farm, Free travel, Guidebook, Travel insurance, Travel, Guide, and Visitor Center."

Table 4: The component factor loadings of diffusion indices-English searching tourism keywords

Dependent variables	Part A		Part B	
	Service consumption		Service consumption-travel class	
Quarterly data	Q4		Q4	
Diffusion index	DF2	DF3	DF2	DF3
Components of Diffusion index (keywords)	Ranking of components (by weight)			
cruises				
Bus				
Rail				
Car Rental				
Taxi				
Airplane				
Hotel				
Accommodation				
Resort				
Amusement Park				
Zoo				
Places of interest				
Attraction				
Hot Spring (2)	**2**		**2**	
Night Market				
Farm (2)		**2**		**2**
Backpacker				
Free.travel				
Working holiday				
Guidebook				
Travel insurance (2)		**1**		**1**
Subsidy				
Travel agency				
Travel				
Guide				
Souvenir				
Luggage (2)	**1**		**1**	
Visitor Center (4)	**3**	**2**	**3**	**2**
Package tour				
First r eigenvalues of the correlation matrix:	4.81	3.62	4.81	3.62
Variability explained	0.73	0.73	0.73	0.73

Source: The authors. 1. In the first column, the number in the parentheses after those components of diffusion index represents how many times that the keyword has ever been the top 3 components in each diffusion index. 2. In table 4, we choose those diffusion indices having significant effects on dependent variables in VAR tests in Table 5. 3. Based on the ranking of components, we conclude the higher-weight variables in bold italics, which are “Hot Spring, Farm, Travel insurance, Luggage, and Visitor Center.”

Table 5: VAR results

Model 1, $y_t = \alpha_1 + \beta_1 y_{t-1} + \beta_2 DF1_{t-1} + \beta_3 DF2_{t-1} + \beta_4 DF3_{t-1} + \varepsilon_{yt}$
Model 2, $DF1_t = \alpha_2 + \beta_5 y_{t-1} + \beta_6 DF1_{t-1} + \beta_7 DF2_{t-1} + \beta_8 DF3_{t-1} + \varepsilon_{DF1t}$
Model 3, $DF2_t = \alpha_3 + \beta_9 y_{t-1} + \beta_{10} DF1_{t-1} + \beta_{11} DF2_{t-1} + \beta_{12} DF3_{t-1} + \varepsilon_{DF2t}$
Model 4, $DF3_t = \alpha_4 + \beta_{13} y_{t-1} + \beta_{14} DF1_{t-1} + \beta_{15} DF2_{t-1} + \beta_{16} DF3_{t-1} + \varepsilon_{DF3t}$

Dependent variables		**Part A**				**Part B**				**Part C**				**Part D**			
		Chinese searching tourism keywords- Service consumption				Chinese searching tourism keywords- Service consumption travel class				English searching tourism keywords- Service consumption travel class				English searching tourism keywords- Service consumption travel class			
Model	Coefficients	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	α_1	0.10 (3.26)	-5.55 (3.54)	7.39† (1.41)	3.52† (0.97)	8.40 (6.08)	-12.68 (19.34)	17.41 (19.90)	2.84† (1.24)	1.86 (2.32)	0.04 (2.27)	0.36 (1.81)	0.75 (1.88)	3.48 (6.02)	-12.02† (5.17)	-6.23 (4.60)	0.84 (2.91)
	β_1	0.79† (0.12)	0.92† (0.13)	0.61† (0.06)	0.89† (0.08)	0.93* (0.54)	0.43 (0.69)	-0.05 (0.66)	0.88† (0.11)	0.99† (0.14)	0.87† (0.17)	1.04† (0.18)	0.60† (0.14)	0.50* (0.30)	0.49 (0.32)	0.86† (0.39)	0.45† (0.19)
	β_2	-9.22 (10.94)	***27.50†*** ***(11.41)***	***19.11†*** ***(4.90)***	***7.91†*** ***(2.69)***	29.76 (21.74)	68.64 (62.29)	36.32 (62.91)	4.24 (3.37)	4.43 (10.00)	-6.82 (12.27)	-0.20 (7.43)	-11.42 (9.27)	-3.96 (25.68)	-67.35 (27.23)	-31.03 (19.78)	-16.13 (14.28)
	β_3	1.76 (3.21)	-4.15 (3.71)	***9.30†*** ***(1.41)***	***3.58†*** ***(1.27)***	***-15.13**** ***(8.78)***	-7.37 (20.28)	18.73 (19.35)	2.74 (1.71)	0.09 (1.55)	0.45 (1.77)	1.10 (2.22)	***3.87†*** ***(1.57)***	-1.87 (4.28)	-2.95 (4.28)	5.66 (6.17)	***6.68†*** ***(2.32)***
	β_4	1.47 (1.69)	***5.01†*** ***(1.94)***	0.31 (0.29)	-0.47 (0.36)	***-16.06†*** ***(6.11)***	10.42 (10.57)	-8.57 (10.46)	0.42 (0.41)	-0.13 (1.68)	0.83 (1.09)	-1.12 (1.19)	***-1.23**** ***(0.73)***	-5.73 (4.43)	3.75 (2.49)	-1.24 (2.94)	***-2.50†*** ***(1.07)***
2	α_2	-0.21* (0.11)	-0.04 (0.08)	-0.16† (0.04)	-0.25† (0.08)	-0.19† (0.05)	-0.04 (0.08)	-0.22† (0.08)	-0.23† (0.08)	-0.18† (0.07)	-0.05 (0.05)	-0.03 (0.04)	-0.05* (0.03)	-0.18† (0.07)	-0.06 (0.04)	-0.04 (0.04)	-0.05 (0.03)
	β_5	0.004 (0.004)	-0.0004 (0.003)	0.01† (0.001)	0.01 (0.01)	0.002 (0.005)	-0.0004 (0.003)	0.0003 (0.003)	0.004 (0.01)	-0.002 (0.004)	-0.001 (0.003)	-0.01† (0.004)	-0.0004 (0.002)	-0.0002 (0.003)	-0.002 (0.003)	-0.01* (0.003)	0.00004 (0.002)
	β_6	0.29 (0.37)	1.18† (0.26)	0.51† (0.12)	0.22 (0.22)	0.33* (0.20)	1.18† (0.26)	0.14 (0.26)	0.24 (0.21)	0.22 (0.30)	0.77† (0.24)	0.62† (0.16)	0.77† (0.16)	0.25 (0.30)	0.69† (0.23)	0.63† (0.17)	0.79† (0.17)
	β_7	-0.10 (0.11)	-0.12 (0.09)	-0.08† (0.04)	-0.17 (0.11)	0.19† (0.08)	-0.12 (0.09)	-0.07 (0.08)	-0.09 (0.11)	-0.01 (0.05)	0.05 (0.04)	-0.13† (0.05)	0.02 (0.03)	0.003 (0.05)	0.06* (0.04)	-0.12† (0.05)	0.01 (0.03)
	β_8	0.10* (0.06)	0.04 (0.04)	0.0005 (0.01)	0.02 (0.03)	0.14† (0.06)	0.04 (0.04)	0.16† (0.04)	0.05* (0.03)	0.08 (0.05)	-0.02 (0.02)	-0.05* (0.02)	0.09† (0.01)	0.09* (0.05)	-0.02 (0.02)	-0.06† (0.02)	0.09† (0.01)
3	α_3	0.02 (0.33)	-0.53 (0.34)	-0.41* (0.23)	-0.06 (0.24)	-0.19 (0.56)	-0.53 (0.34)	-0.23 (0.31)	-0.16 (0.19)	0.32 (0.25)	-0.19 (0.30)	0.40 (0.26)	-0.34 (0.26)	0.31 (0.30)	-0.19 (0.30)	0.35 (0.26)	-0.32 (0.27)
	β_9	0.02 (0.01)	0.01 (0.01)	0.004 (0.01)	0.03 (0.02)	-0.02 (0.05)	0.01 (0.01)	0.02† (0.01)	0.02 (0.02)	-0.07† (0.01)	0.03 (0.02)	-0.04* (0.03)	0.02 (0.02)	-0.06† (0.01)	0.02 (0.02)	-0.03 (0.02)	0.02 (0.02)
	β_{10}	0.46 (1.11)	1.79 (1.10)	-1.73† (0.82)	0.11 (0.66)	-1.20 (2.01)	1.79 (1.10)	-0.37 (0.98)	-0.32 (0.50)	-1.01 (1.07)	-0.08 (1.61)	0.31 (1.05)	-0.83 (1.29)	-0.69 (1.26)	-0.14 (1.57)	0.33 (1.10)	-0.84 (1.33)
	β_{11}	0.75† (0.33)	0.35 (0.36)	0.51† (0.24)	0.68† (0.31)	1.64† (0.81)	0.35 (0.36)	0.55* (0.30)	0.66† (0.26)	-0.01 (0.17)	0.52† (0.23)	0.26 (0.32)	0.60† (0.22)	-0.03 (0.21)	0.50† (0.25)	0.28 (0.34)	0.62† (0.22)
	β_{12}	-0.21 (0.17)	0.12 (0.19)	0.01 (0.05)	-0.10 (0.09)	1.43† (0.56)	0.12 (0.19)	-0.25 (0.16)	-0.08 (0.06)	0.03 (0.18)	0.24* (0.14)	-0.07 (0.17)	-0.05 (0.10)	0.03 (0.22)	0.24* (0.14)	-0.11 (0.16)	-0.07 (0.10)

4	α_4	0.84 (0.53)	1.51† (0.41)	2.85† (1.27)	1.55† (0.75)	0.61 (0.79)	1.51† (0.41)	-0.07 (0.40)	1.44† (0.61)	0.65 (0.47)	-0.06 (0.45)	-1.09 (0.48)	1.07 (0.68)	0.65 (0.47)	-0.13 (0.45)	-0.98† (0.47)	1.09 (0.72)
	β_{13}	-0.01 (0.02)	-0.04† (0.01)	-0.04 (0.05)	-0.04 (0.06)	0.06 (0.07)	-0.04† (0.01)	0.02 (0.01)	0.05 (0.05)	0.01 (0.03)	-0.01 (0.03)	0.09* (0.05)	-0.04 (0.05)	0.004 (0.02)	-0.02 (0.03)	0.07* (0.04)	0.02 (0.05)
	β_{14}	2.88 (1.77)	-4.55† (1.31)	10.78† (4.42)	5.08† (2.09)	4.02 (2.81)	-4.55† (1.31)	0.30 (1.25)	4.36† (1.65)	2.87 (2.03)	-0.50 (2.45)	-1.68 (1.96)	3.12 (3.38)	2.83 (2.01)	-1.15 (2.36)	-1.67 (2.04)	3.70 (3.51)
	β_{15}	1.08† (0.52)	1.96† (0.43)	3.40† (1.27)	2.08† (0.99)	-1.68 (1.13)	1.96† (0.43)	0.15 (0.39)	1.57* (0.84)	-0.35 (0.31)	0.41 (0.35)	1.23 (0.59)	0.96* (0.57)	-0.35 (0.33)	0.54 (0.37)	1.22* (0.64)	0.82 (0.57)
	β_{16}	0.31 (0.27)	-0.01 (0.22)	-0.53† (0.26)	-0.11 (0.28)	-1.71† (0.79)	-0.01 (0.22)	0.46† (0.21)	0.40† (0.20)	0.08 (0.34)	-0.37* (0.22)	-0.61 (0.31)	-0.30 (0.26)	0.07 (0.35)	-0.33 (0.22)	-0.52* (0.30)	-0.27 (0.26)
	det (SSE)	3e-10	3e-12	3e-11	4e-10	2e-09	1e-10	1e-10	1e-10	1e-08	4e-09	2e-09	2e-09	1e-07	1e-08	2e-08	4e-09
	AIC	-20.10	-24.76	-22.00	-19.59	-17.95	-21.57	-21.18	-21.44	-16.28	-17.43	-18.13	-18.30	-14.62	-16.25	-15.89	-17.31
	BIC	-19.32	-23.98	-21.25	-18.84	-17.17	-20.79	-20.40	-20.67	-15.50	-16.65	-17.34	-17.53	-13.83	-15.46	-15.10	-16.54
	HQ	-20.02	-24.68	-22.01	-19.60	-17.91	-21.49	-21.10	-21.40	-16.20	-17.35	-18.05	-18.26	-14.54	-16.17	-15.81	-17.27

Source: The authors. Robust t statistics in brackets. * significant at 10%; ** significant at 5%;
† significant at 1%

4.2 Empirical results

4.2.1 Part 1 The factor loading of components in diffusion indices

At first, we examined the significant diffusion indices, in which those components with highest weights are treated as the prominent important internet-searching tourism data. Those results are sequentially showed in Table 3 to 4[8,9].

Table 3 showed the results with "Chinese searching tourism data." Among those statistically significant diffusion indices in Table 3-Part A, the higher-weight components are separately "Bus, Taxi, Free travel, Guidebook, Travel insurance, Travel, and Visitor Center." Among those statistically significant diffusion indices in Table 3-Part B, the higher-weight components are separately "Taxi, Accommodation, Night market, Guide, and Visitor center." Those important internet-searching tourism data represent that the keywords in "Transport, Rest, Recreational areas, Grand tour, and Travel-related" groups have prominent effects on service consumption in Taiwan.

Table 4 demonstrated the results with "English searching tourism data." Among the statistically significant diffusion indices in Table 4-Part A, the higher-weight components are separately "Hot Spring, Farm, Travel insurance, Luggage, and Visitor Center." Among those statistically significant diffusion indices in Table 4-Part B, the higher-weight components are separately Hot Spring, Farm, Travel insurance, Luggage, and Visitor Center." Those important internet-searching tourism data represent that the keywords in "Recreational areas, and Travel-related" groups have prominent effects on service consumption in Taiwan. The five important internet-searching tourism data represent that the keywords in "Recreational areas, Travel-related" groups have prominent effects on service consumption in Taiwan.

The possible explanations of those components having higher weights are as follows. "Hot spring" has higher weight, owing to the recent popular combination of hotel and hot spring resources in the same industry in Taiwan. "Farm" has higher weight, for the Taiwan leisure farms have been gradually transformed into leisure industry configuration to meet customer's requirements. "Travel insurance" has higher weight, for preventing emergencies and reducing the adverse effect of tourists. "Luggage" has higher weight, for the internet information of tips for storing luggage, the rules of check-in luggage, and carry-on luggage all enhance the internet-searching motivation. "Visitor center" has higher weight for its merchandise, food supply, and information enquiry. "Bus" has higher weight, for its cheap fare and the well-developed bus network in Taiwan. "Taxi" has higher weight, for its convenience to hail a taxi via mobile APP in Taiwan. "Accommodation" has higher weight, for its important role in one trip. "Night market" has higher weight, for it is

---

[8] Before we estimated the tourism diffusion indices, all the time series data were examined with Augmented Dickey-Fuller unit root tests, identifying these variables satisfying transformation without unit root. If the original series should be differenced to be stationary, the log difference will be adopted.

[9] In this paper, the log difference is adopted for all original data, including two kinds of "Private Final Consumption Expenditure" and the tourism internet-searching keywords data. The empirical results of Augmented Dickey-Fuller unit root tests show that all data are stationary without unit root.

the cultural characteristics in Taiwan, also the tourists' favorite tourist attractions in Taiwan.

4.2.2 Part2 VAR model

4.2.2.1 The results with Chinese searching tourism data

Table 5-Part A to Part B are the results of VAR model with "Chinese searching tourism data." We extracted three diffusion indices, in which these components covered quarterly internet-searching tourism data. At first, the low-frequency annual data were nowcasted by high-frequency quarterly data, and then those diffusion indices were adopted to proceed VAR test with service consumption and service consumption-travel classification separately. Those results are sequentially shown in Table 5-Part A to Part B.

Table 5-Part A is the results of treating service consumption as the dependent variable. The lagged DF1 and lagged DF3 constructed by the 2nd quarter data (Q2), the lagged DF1 and lagged DF2 constructed by the 3rd quarter data (Q3), and the lagged DF1 and lagged DF2 constructed by the 4th quarter data (Q3) all have significant predictability for service consumption. Also, the lagged service consumption has significant predictability for DF3 constructed by the 2nd quarter data (Q2), and for DF1 constructed by the 3rd quarter data (Q3).

As mentioned above, the DF3 constructed by the 2nd quarter data (Q2) and the DF1 constructed by the 3rd quarter data (Q3) have granger causality reciprocal reactions with service consumption. And the lagged information of those diffusion indices significantly affects service consumption in the next period, representing that the tourism information could enhance the motivation of consuming in tours and further push up the service consumption in Taiwan.

Among the components of those diffusion indices, the higher-weight variables are separately "Bus, Taxi, Free travel, Guidebook, Travel insurance, Travel, and Visitor Center." In other words, among tourism-related keywords in this VAR model, the "Transport, Grand tour, and Travel-related" groups have prominent predictability.

Table 5-Part B is the results of treating service consumption-travel class as the dependent variable. The lagged DF2 and lagged DF3 constructed by the 1st quarter data (Q1), the lagged DF3 constructed by the 2nd quarter data (Q2), and the lagged DF2 constructed by the 3rd quarter data (Q3) all have significant predictability for service consumption-travel class.

Although service consumption-travel class has no granger causality reciprocal reaction with diffusion indices constructed by quarterly data, the lagged information of those diffusion indices significantly affects the service consumption-travel class in the next period, representing that the interested tourism information could enhance the motivation of consuming in tours and further increase the service consumption-travel class in Taiwan.

Among the components of those diffusion indices, the higher-weight variables are separately "Taxi, Accommodation, Night market, Guide, and Visitor center." In other words, among those tourism-related keywords in the VAR model, the "Transport, Rest, Recreational areas, and Travel-related" groups have prominent predictability.

In conclusion, from the results of VAR model in Table 5-Part A to Part B [10,11], both service consumption and service consumption-travel class are robustly affected by the "Chinese searching tourism data" in "Transport, Rest, Recreational areas, Grand tour, and Travel-related" groups, and their higher-weight components are separately "Bus, Taxi, Accommodation, Farm, Free travel, Guidebook, Travel insurance, Travel, Guide, and Visitor Center."

4.2.2.2 The results with English searching tourism data

Table 5-Part C to Part D are the results of VAR model with "English searching tourism data." We extracted three diffusion indices, in which these components covered quarterly internet-searching tourism data. At first, the low-frequency annual data are nowcasted by high-frequency quarterly data, and then those diffusion indices are adopted to proceed VAR test with service consumption and service consumption-travel class separately. Those results are sequentially showed in Table 5-Part C to Part D.

Table 5-Part C is the results of treating service consumption as the dependent variable. The lagged DF2 and lagged DF3 constructed by the 4th quarter data (Q4) have significant predictability for service consumption. Also, the lagged service consumption has significant predictability for DF2 constructed by the 1st quarter data (Q1), and for DF1, DF2, DF3 constructed by the 3rd quarter data (Q3).

Although service consumption has no granger causality reciprocal reaction with diffusion indices constructed by quarterly data, the lagged information of those diffusion indices significantly affects the service consumption in the next period, which represents that the interested tourism information could enhance the motivation of consuming in tours and further increase the service consumption in Taiwan.

Among the components of those diffusion indices, the higher-weight variables are separately "Hot Spring, Farm, Travel insurance, Luggage, and Visitor center." In other words, among those internet-searching tourism data in this VAR model, the keywords in "Recreational areas, and Travel-related" groups have prominent predictability.

Table 5- Part D shows the results of treating service consumption-travel class as the dependent variable. The lagged DF2 and lagged DF3 constructed by the 4th quarter data (Q4), the lagged DF2 constructed by the 1st quarter data (Q1), and the lagged

---

[10] In Table 5, we found whether the dependent variables are Service consumption, or Service consumption-travel class with "Chinese searching tourism keywords," the det (SSE), AIC, BIC, and HQ are almost the same. And whether the dependent variables are Service consumption, or Service consumption-travel class with "English searching tourism keywords," the det (SSE), AIC, BIC, and HQ are almost the same. However, the model with "English searching tourism keywords" has lower det (SSE), AIC, BIC, and HQ, which means that the model with "English searching tourism keywords" has better explanatory power.

[11] We choose VAR model to lag 1 period to be VAR (1), based on the AIC and SC criteria.

DF1 and lagged DF3 constructed by the 3rd quarter data (Q3) have significant predictability for service consumption-travel class.

Although service consumption-travel class has no granger causality reciprocal reaction with diffusion indices constructed by quarterly data, the lagged information of those diffusion indices significantly affects the service consumption-travel class in the next period, representing that the interested tourism information could enhance the motivation of consuming in tours and further increase the service consumption-travel class in Taiwan.

Among the components of those diffusion indices, the higher-weight variables are separately "Hot Spring, Farm, Travel insurance, Luggage, and Visitor center." In other words, among those tourism-related keywords in this VAR model, the keywords in "Recreational areas, and Travel-related" groups have prominent predictability.

In conclusion, from the results of VAR model in Table 5-Part C to Part D, both service consumption and service consumption-travel class are robustly affected by the keywords in "Recreational areas, and Travel-related" groups, and their higher-weight components are separately "Hot Spring, Farm, Travel insurance, Luggage, and Visitor center."

To sum up the results of VAR model in Table 5-Part A to Part D, as for the results of "Chinese searching tourism data," the "Bus, Taxi, Accommodation, Farm, Free travel, Guidebook, Travel insurance, Travel, Guide, and Visitor Center" are all important internet-searching tourism data having prominent effects on both service consumption and service consumption-travel class. As for the results of "English searching tourism data," the "Hot Spring, Farm, Travel insurance, Luggage, and Visitor center" are all important internet-searching tourism data having prominent effects on both service consumption and service consumption-travel class.

That is to say, the keywords in "Recreational areas, and Travel-related" groups have significant effects on service consumption in Taiwan through nowcasting annual service consumption with "Chinese searching tourism keywords" and "English searching tourism keywords." Among the components of diffusion indices, "Farm, Travel insurance, and Visitor center" are important variables with higher weights in common.

However, the empirical results in Table 5 are different from the description statistics results in Table 2. The former results tell us that people care about the tourism information of "Recreational areas, and Travel-related" groups most, which also further affects the service consumption in Taiwan. However, the latter result tells us that people care about the transport information most. That is to say, searching tourism keywords with highest searching times are not exactly the most important information that could enhance the motivation of consuming in tours and further increase the service consumption in Taiwan.

5. Conclusions

In this paper, we examined the effects of tourism on service consumption in Taiwan via nowcasting the low-frequency annual service consumption with the high-frequency quarterly internet-searching tourism keywords. Using the Principal Components Analysis (PCA), we extracted the diffusion indices from tourism keywords, which are classified into five groups and twenty-nine classifications. We aimed at the reciprocal reactions between those diffusion indices and service consumption in Taiwan.

With regards to the results of nowcasting annual data of service consumption and service consumption-travel class with quarterly “Chinese searching tourism data,” the keywords in “Transport, Rest, Recreational areas, Grand tour, and Travel-related” groups have significant effects. Among those keywords, “Bus, Taxi, Accommodation, Farm, Free travel, Guidebook, Travel insurance, Travel, Guide, and Visitor Center” have highest weight. With regards to the results of nowcasting annual data of service consumption and service consumption-travel class with quarterly “English searching tourism data,” the keywords in “Recreational areas, and Travel-related” groups have significant effects. Among those keywords, “Hot Spring, Farm, Travel insurance, Luggage, and Visitor center” have highest weight. We concluded that the keywords in “Recreational areas, and Travel-related” groups have significant predictability for service consumption in Taiwan. Further, whether through “Chinese searching tourism data” or “English searching tourism data,” the “Farm, Travel insurance, and Visitor center” are important keywords having significant effects on service consumption in Taiwan. That is to say, people care about the tourism information of “Recreational areas, and Travel-related” groups most, which could further affect the service consumption in Taiwan. It also could be regarded as important policy suggestions for the authority to enhance the tourists’ motivation of consuming in tours and further stimulate the service consumption in Taiwan.

References

[1] Bai, J. (2003). " Inferential Theory for Factor Models of Large Dimensions.", Econometrica 71(1), January,135-171.

[2] Bai, J., and Ng., S. (2002). "Determining the number of factors in approximate factor models." Econometrica,70.1: 191-221

[3] Banbura, M. (2011), "Nowcasting with daily data," European Central Bank, Working Paper.

[4] Barhoumi, K., Darné, O., and Ferrara, L. (2010). Are Disaggregate Data Useful for Factor Analysis in Forecasting French GDP? Journal of Forecasting, 29(1-2), 132–144.

[5] Boivin, J. and Ng, S., 2005. "Understanding and Comparing Factor-Based Forecasts," International Journal of Central Banking, International Journal of Central Banking, 1(3).

[6] Bragoli, D. and Fosten, J., (2016). "Nowcasting Indian GDP," University of East Anglia School of Economics Working Paper Series 2016-06, School of Economics, University of East Anglia, Norwich, UK.

[7] Chernis, T. and Sekkel, R., (2017). "A Dynamic Factor Model for Nowcasting Canadian GDP Growth," Staff Working Papers 17-2, Bank of Canada.

[8] Chikamatsu, K., Hirakata, N., Kido, Y., and Otaka, K., 2018. "Nowcasting Japanese GDPs," Bank of Japan Working Paper Series 18-E-18, Bank of Japan.

[9] Choi, Hyunyoung, and Hal Varian. (2012). "Predicting the present with google trends." Economic Record, 88(s1):2-9.

[10] Connor, G. and Korajczyk, R. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. Journal of Financial Economics, 15(3), 373-394

[11] Connor, G. and Korajczyk, R. (1993). A Test for the Number of Factors in an Approximate Factor Model. Journal of Finance, 48(4), 1263-91.

[12] D' Amuri, F., and Marcucci, J. (2010)."Googleit!"Forecasting the US Unemployment Rate with a Google Job Search index. Fondazione Eni Enrico Mattei Working Papers.

[13] Evans, Martin D. D. (2005). "Where Are We Now? Real-Time Estimates of the Macroeconomy, " International Journal of Central Banking, 1(2), September.

[14] Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia Reichlin (2005): The Generalized Dynamic Factor Model, Journal of the American Statistical Association, 100:471, 830-840

[15] Giannone, D. and Reichlin, L. & Sala, L. (2005). "Monetary Policy in Real Time," NBER Chapters, in: NBER Macroeconomics Annual 2004, 19, 161-224

[16] Giannone, D. and Reichlin, L. & Small, D. (2008). "Nowcasting: The real-time informational content of macroeconomic data," Journal of Monetary Economics, 55(4), 665-676.

[17] Giannone, D. and Reichlin, L. &Bańbura, M. (2010). "Nowcasting," Working Paper Series 1275, European Central Bank.
[18] Kabundi, A., Nel, E. and Ruch, F., 2016. "Nowcasting Real GDP growth in South Africa," Working Papers 581, Economic Research Southern Africa.
[19] Klein, L. R., and Park, J. Y.(1994). Current Quarterly Models of the United States Economy Forecast Summary. Weekly Report.
[20] Kholodilin, Konstantin A., Maximilian Podstawski, and BorissSiliverstovs. (2010)"Do Google searches help in nowcasting private consumption? A real-time evidence for the US." KOF Swiss Economic Institute working Paper256.
[21] Lahiri, K., and Monokroussos, G. (2013). Nowcasting US GDP: The role of ISM business surveys. International Journal of Forecasting, 29(4), 644-658.
[22] Liebermann, J. (2012). "Short-term forecasting of quarterly gross domestic product growth," Central Bank of Ireland, Working Paper, 74-84.
[23] Luciani, M. and L. Ricci, (2013), "Nowcasting Norway," ECARES, Working Papers.
[24] Luciani, M., Pundit, M., Ramayandi, A., and Veronese, G., 2018. "Nowcasting Indonesia," Empirical Economics, Springer, vol. 55(2), pages 597-619, September.
[25] Marcellino, M. and Stock, J. and Watson, M. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. European Economic Review, 47(1), 1-18
[26] Mazzi, G. L., and Montana, G. (2009). A system of rapid estimates to improve real-time monitoring of the economic situation the case of the Euro Area. National Institute Economic Review, 210(1), 63-70.
[27] Matsumoto, Azusa, Kohei Matsumura, and Noriyuki Shiraki (2013). "Potential of Search Data in Assessment of Current Economic Conditions." No. 2013-04-18. Bank of Japan.
[28] Notini, H. H. and J. V. Issler& C. Rodrigues & S. Matos &Bonelli, R. (2012), "Nowcasting Brazil monthly GDP: A State-Space approach," Economics, Working Papers.
[29] Nikolaos Askitas and Klaus F. Zimmermann (2009). Google Econometrics and Unemployment Forecasting. Discussion Paper No. 4201.
[30] Stock, J.H. and Watson, M.W. (1998a) Diffusion Indexes. NBER Working Paper, No. W6702.
[31] Stock, J.H. and Watson, M.W. (1998b), Forecasting inflation. Journal of Monetary Economics 44, 293-335.
[32] Stock, J.H. and Watson, M.W. (2002a) "Macroeconomic Forecasting Using Diffusion Indexes," Journal of Business and Economic Statistics, 20(2), 147-162
[33] Stock, J.H. and Watson, M.W. (2002b). Forecasting Using Principal Components from a Large Number of Predictors. Journal of the American Statistical Association, 97(460), 1167-1179

[34] Takeda, F., and Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pacific-Basin Finance Journal, 27, 1-18.
[35] Yiu, Matthew S. and Chow, Kenneth K. (2011). Nowcasting Chinese GDP: Information Content of Economic and Financial Data. HKIMR Working Paper No. 04/2011.
[36] Vosen, Simeon, and Torsten Schmidt. (2011) "Forecasting private consumption: survey-based indicators vs. Google trends." Journal of Forecasting,30.6: 565-578.
[37] Vosen, Simeon and Torsten Schmidt. (2012). "A monthly consumption indicator for Germany based on Internet search query data. "Applied Economics Letters19.7: 683-687.