Journal of Applied Finance & Banking, Vol. 15, No. 6, 2025, 1-12
ISSN: 1792-6580 (print version), 1792-6599(online)
https://doi.org/10.47260/jafb/1561
Scientific Press International Limited

Recreational Activities and Tourism Expenditure in Taiwan: An Online Buzz Perspective

Chien-Jung Ting[1] and Yu-Jung Chen[1]

Abstract

This study examines the influence of online search behavior—specifically keywords related to recreational activities popular among Taiwanese domestic travelers—on total tourism expenditure. Utilizing high-frequency data from Google Trends, the analysis investigates how public interest in specific leisure activities correlates with tourism spending. Principal Component Analysis (PCA) is used to extract key indicators representing aggregated search trends, which are then incorporated into a Vector Autoregression (VAR) model to assess their dynamic relationship with tourism expenditure. Results indicate that among four major categories—gourmet, nature sightseeing, other leisure, and cultural experiences—six keywords ("drinking coffee," "whale watching," "shopping," "farm," "indigenous culture," and "boating") significantly affect domestic tourism expenditure. These activities contribute to broader consumption in areas such as lodging, transportation, and dining. The study contributes by (1) demonstrating the predictive utility of online search data for tourism economics, and (2) highlighting the growing significance of specific recreational activities, consistent with official statistics.

JEL classification numbers: C60, O11, R11.
Keywords: Principal Component Analysis (PCA), Recreational Activities, Tourism Expenditure, Vector Autoregression (VAR).

---

[1] Department of Leisure, Recreation, and Tourism Management, Southern Taiwan University of Science and Technology, Taiwan.

Article Info: *Received:* September 21, 2025. Revised: October 19, 2025.
Published online: October 23, 2025.

1. Introduction

Taiwan's diverse natural landscapes, rich cultural assets, and well-developed transportation infrastructure have positioned it as a leading domestic tourism destination. As living standards improve and consumer preferences diversify, domestic tourism has shifted from conventional sightseeing to including eco-tourism, cultural participation, rural travel, and outdoor recreation. This shift, accelerated by the post-COVID-19 "revenge travel" trend, has further revitalized the domestic tourism market.

According to the 2023 Taiwan Tourism Survey, domestic trips increased by 22.66% year-on-year, reaching 200 million trips. The data indicate a weekend-oriented travel pattern, with 55.3% of trips occurring on weekends, and highlight that 82.9% of travelers view tourism primarily as a form of relaxation—underscoring the integration of leisure travel into everyday life.

The Tourism Administration classifies recreational activities into seven categories: nature sightseeing, cultural experience, gourmet, sports, theme parks, other leisure activities, and visits to friends or relatives. While nature sightseeing remains dominant, interest in cultural and culinary tourism is growing. This classification provides a valuable framework for examining how various leisure preferences shape tourism expenditure.

With the increasing accessibility of high-frequency online data, recent research has adopted Google Trends as a proxy for real-time public interest. Building on this approach, the present study analyzes the online search behavior of Taiwanese travelers to identify which recreational activity types most strongly predict domestic tourism expenditure.

Using government classifications and prior literature, keyword groups were constructed and Google Trends data retrieved. Principal Component Analysis (PCA) was employed to extract key indicators, which were then incorporated into a Vector Autoregression (VAR) model. Following the framework of Matsumoto et al. (2013), 40 recreational subcategories were grouped into six broader categories to construct diffusion indices for empirical analysis.

The study has two main objectives: (1) to improve forecasts of domestic tourism expenditure using real-time online data; and (2) to assess how interest in specific recreational activities influences tourism-related consumption patterns. Addressing a gap in the literature, this research moves beyond destination-level analyses to explore macro-level dynamics between leisure preferences and tourism spending, offering insights that can inform targeted tourism policies and support sustainable development in Taiwan's domestic tourism sector.

2. Literature Review

The increasing availability of high-frequency data, particularly from online search behavior, has significantly enhanced the timeliness and granularity of economic forecasting. Compared to traditional low-frequency government-published statistical data, such data provide dynamic insights into public sentiment and behavioral trends.

This study builds on a growing body of literature employing online search volume as a proxy for economic activity. The relevant research can be categorized into three main strands:

(1) Forecasting Economic Indicators with High-Frequency Online Data

Initial studies, such as Klein and Park (1993,1995), demonstrated the potential of using high-frequency indicators to nowcast low-frequency macroeconomic variables. Subsequent research employed dynamic factor models and principal component techniques to improve forecasting accuracy (e.g., Evans, 2005; Marcellino et al., 2003; Boivin and Ng, 2005; Barhoumi et al., 2010). Vosen and Schmidt (2011, 2012), and Kholodilin et al. (2010) showed that search volume data could outperform traditional models in predicting private consumption.

(2) Applications in Labor Markets and Financial Behavior

Askitas and Zimmermann (2009) and D'amuri and Marcucci (2010) demonstrated a strong relationship between search trends and unemployment rates, while Takeda and Wakao (2014) linked Google Trends data to investor behavior in Japan, finding stronger correlations with trading volume than with price movements.

(3) Tourism-Related Applications

Most relevant to the present study, Matsumoto et al. (2013) examined how tourism-related keyword searches influenced private consumption in post-disaster Japan. Choi and Varian (2012) further showed that incorporating search data improved the performance of tourism forecasting models.

In the Taiwanese context, the 2023 Taiwan Tourism Survey reported 206.75 million domestic trips and NT$495.4 billion in total expenditure—marking a 26.15% increase from pre-pandemic levels, largely attributed to the resurgence of domestic travel demand.

Building on prior studies and government classification systems, this research categorizes recreational search behavior into six major types and 40 subcategories (see Table 1), encompassing activities such as nature sightseeing, cultural experience, gourmet, sports, theme parks, and other leisure activities.[2]

[2] The Tourism Administration under Taiwan's Ministry of Transportation and Communications categorizes recreational activities into seven major types: nature sightseeing, cultural experience, gourmet, sports, theme parks, other leisure activities, and visits to friends or relatives. However, as the category " visits to friends or relatives" yielded no available data in the Google Trends database, this study excludes it from the analysis.

Table 1: Data Sources and Descriptions

<table>
<tr><th>Category</th><th colspan="2">Name</th><th>Data Source</th><th>Data Period</th><th>Frequency</th></tr>
<tr><td>Target Variable</td><td colspan="2">Total Domestic Tourism Expenditure</td><td>Tourism Administration, Ministry of Transportation and Communications</td><td>2004-2024</td><td>Monthly</td></tr>
<tr><td rowspan="7">Keywords for Main Recreational Activities Engaged in During Domestic Travel</td><td>Six Major Activity Categories</td><td>40 Subcategories of Activities</td><td rowspan="7">Google trends</td><td rowspan="7">2004-2024</td><td rowspan="7">Annual</td></tr>
<tr><td>Nature Sightseeing Activities</td><td>Visiting Waterfalls, Hiking, Mountain Climbing, Camping, Watching Pandas, Whale Watching, Bird Watching, Flower Viewing, Cherry Blossom Viewing, Watching the Sunrise</td></tr>
<tr><td>Cultural Experience Activities</td><td>Historic Sites, Festivals, Performances, Exhibitions, Traditional Arts, Indigenous Culture, Farm</td></tr>
<tr><td>Gourmet Activities</td><td>Specialty Cuisine, Night Market Snacks, Fine Tea Tasting, Drinking Coffee, Afternoon Tea, Culinary Classes</td></tr>
<tr><td>Sports Activities</td><td>Swimming, Fishing, Baseball, Rock Climbing, Cycling, Ball Games, Road Running</td></tr>
<tr><td>Theme Park Activities</td><td>Roller Coaster</td></tr>
<tr><td>Other Leisure Activities</td><td>Driving Around, Soaking in Hot Springs, Strolling, Shopping,Watching Movies, Boating, Cable Car, Tourist Factories, Hot Air Ballooning</td></tr>
</table>

3. Methodological Framework

This study investigates whether high-frequency online search data—specifically, keyword search volumes related to popular recreational activities—can effectively predict low-frequency indicators such as Taiwan's domestic tourism expenditure. By integrating digital behavioral data with time-series econometric modeling, this research aims to assess the forecasting potential of real-time online signals in the tourism context.

3.1 Step 1: Dimensionality Reduction via Diffusion Index Construction

To address the high dimensionality of keyword data while preserving its informational content, Principal Component Analysis (PCA) is employed to extract latent common factors—termed Diffusion Indices—from the keyword time series. These indices capture the co-movement among search trends and serve as key inputs for subsequent modeling.

Let $\mathbf{X}_{\mathbf{T\times N}}$ be the observed data matrix and $\mathbf{F}_{\mathbf{T\times k}}$ the diffusion index matrix. The model is defined as:

$$X = F\beta + \varepsilon \tag{1}$$

where β is the factor loading matrix and ε is a white noise error term. Since F and β are both unobserved, a two-step estimation procedure (Stock and Watson, 1998a, 1998b, 2002a, 2002b) is applied: (1) Step 1: Estimate F by extracting the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of X; and (2) Step 2: Estimate β using Ordinary Least Squares (OLS).

The optimization objective is to minimize the residual sum of squares:

$$\min_F \sum_{i=1}^{N} (X_i - F\beta_i)' (X_i - F\beta_i) \tag{2}$$

This is equivalent to:

$$\min_F \{trace[X'X - X'F(F'F)^{-1}F'X]\} \tag{3}$$

or, alternatively, the maximization problem:

$$\max_{\mathrm{F}} trace(X'F(F'F)^{-1}F'X) \tag{4}$$

The resulting F comprises the leading k eigenvectors of the sample covariance matrix of X, representing the most significant dimensions of variation in search activity.

3.1.1 Data Preprocessing

All keyword time series are tested for stationarity using the Augmented Dickey-Fuller (ADF) test. Non-stationary series integrated of order one, I(1), are log-transformed and differenced. The resulting stationary matrix X is used for diffusion index extraction. The derived indices, denoted F_1 、F_2 、 ... F_n, reflect underlying recreational activity trends and serve as potential predictors of tourism expenditure. Factor loadings help identify the most influential keywords in each index.

3.2 Step 2: Vector Autoregressive (VAR) Modeling

To assess the dynamic relationships between the diffusion indices and tourism expenditure, a Vector Autoregressive (VAR) model is applied. This framework captures feedback and lead-lag effects among multiple time series.
The VAR(p) model is specified as:

$$Z_t = AZ_{t-1} + \varepsilon_t \tag{5}$$

where Z_t is a vector comprising tourism expenditure and the diffusion indices, A is the coefficient matrix, and ε_t is the innovation term.
Forecasts are generated via:

$$E_t(Z_{t+j}) = A^j Z_t E_t(y_{t+j}) = e'_2 A^j Z_t \tag{6}$$

where e'_2 selects the tourism expenditure component. The reduced-form solution is expressed as:

$$x_t = e'(I - \beta A)^{-1} \beta A Z_t \tag{7}$$

Significance of coefficients is evaluated using the Wald test to identify which diffusion indices serve as effective leading indicators.

3.2.1 Model Specification

The final VAR model includes four variables: total domestic tourism expenditure (Y_t) and three diffusion indices ($F1_t$, $F2_t$, and $F3_t$). The system can be represented as[3]:

[3] According to the principle of selecting the model with smaller AIC, BIC, and HQ values, we choose a lag order of four for the Vector Autoregression (VAR) model, denoted as VAR (4). The results of the VAR model are presented with its four equations labeled sequentially as Model 1, Model 2, Model 3, and Model 4. The following are the specifications for Models 1 through 4:

Model 1: $Y_t = \hat{\alpha}_1 + \hat{\beta}_1 Y_{t-1} + \hat{\beta}_2 Y_{t-2} + \hat{\beta}_3 Y_{t-3} + \hat{\beta}_4 Y_{t-4} + \hat{\beta}_5 DF1_{t-1} + \hat{\beta}_6 F1_{t-2} + \hat{\beta}_7 F1_{t-3} + \hat{\beta}_8 F1_{t-4} + \hat{\beta}_9 F2_{t-1} + \hat{\beta}_{10} F2_{t-2} + \hat{\beta}_{11} F2_{t-3} + \hat{\beta}_{12} F2_{t-4} + \hat{\beta}_{13} F3_{t-1} + \hat{\beta}_{14} F3_{t-2} + \hat{\beta}_{15} F3_{t-3} + \hat{\beta}_{16} F3_{t-4}$

Model 2: $F1_t = \hat{\alpha}_2 + \hat{\beta}_{17} Y_{t-1} + \hat{\beta}_{18} Y_{t-2} + \hat{\beta}_{19} Y_{t-3} + \hat{\beta}_{20} Y_{t-4} + \hat{\beta}_{21} F1_{t-1} + \hat{\beta}_{22} F1_{t-2} + \hat{\beta}_{23} F1_{t-3} + \hat{\beta}_{24} F1_{t-4} + \hat{\beta}_{25} F2_{t-1} + \hat{\beta}_{26} F2_{t-2} + \hat{\beta}_{27} F2_{t-3} + \hat{\beta}_{28} F2_{t-4} + \hat{\beta}_{29} F3_{t-1} + \hat{\beta}_{30} F3_{t-2} + \hat{\beta}_{31} F3_{t-3} + \hat{\beta}_{32} F3_{t-4}$

Model 3: $F2_t = \hat{\alpha}_3 + \hat{\beta}_{33} Y_{t-1} + \hat{\beta}_{34} Y_{t-2} + \hat{\beta}_{35} Y_{t-3} + \hat{\beta}_{36} Y_{t-4} + \hat{\beta}_{37} F1_{t-1} + \hat{\beta}_{38} F1_{t-2} + \hat{\beta}_{39} F1_{t-3} + \hat{\beta}_{40} F1_{t-4} + \hat{\beta}_{41} F2_{t-1} + \hat{\beta}_{42} F2_{t-2} + \hat{\beta}_{43} F2_{t-3} + \hat{\beta}_{44} F2_{t-4} + \hat{\beta}_{45} F3_{t-1} + \hat{\beta}_{46} F3_{t-2} + \hat{\beta}_{47} F3_{t-3} + \hat{\beta}_{48} F3_{t-4}$

Model 4: $F3_t = \hat{\alpha}_4 + \hat{\beta}_{49} Y_{t-1} + \hat{\beta}_{50} Y_{t-2} + \hat{\beta}_{51} Y_{t-3} + \hat{\beta}_{52} Y_{t-4} + \hat{\beta}_{53} F1_{t-1} + \hat{\beta}_{54} F1_{t-2} + \hat{\beta}_{55} F1_{t-3} + \hat{\beta}_{56} F1_{t-4} + \hat{\beta}_{57} F2_{t-1} + \hat{\beta}_{58} F2_{t-2} + \hat{\beta}_{59} F2_{t-3} + \hat{\beta}_{60} F2_{t-4} + \hat{\beta}_{61} F3_{t-1} + \hat{\beta}_{62} F3_{t-2} + \hat{\beta}_{63} F3_{t-3} + \hat{\beta}_{64} F3_{t-4}$

$$\begin{bmatrix} Y_t \\ F1_t \\ F2_t \\ F3_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ Y_{t-3} \\ Y_{t-4} \\ F1_{t-1} \\ F1_{t-2} \\ F1_{t-3} \\ F1_{t-4} \\ F2_{t-1} \\ F2_{t-2} \\ F2_{t-3} \\ F2_{t-4} \\ F3_{t-1} \\ F3_{t-2} \\ F3_{t-3} \\ F3_{t-4} \end{bmatrix} + \begin{bmatrix} \varepsilon_{Y_t} \\ \varepsilon_{F1_t} \\ \varepsilon_{F2_t} \\ \varepsilon_{F3_t} \end{bmatrix} \tag{8}$$

where B is the matrix of estimated coefficients and ε is the error vector. This structure enables empirical assessment of how online search-based indicators influence tourism expenditure, offering a novel approach for real-time forecasting in tourism economics.

4. Data Description

This study employs high-frequency search volume data from Google Trends, spanning January 2004 to December 2024. The dataset includes keywords associated with recreational activities commonly pursued by Taiwanese domestic tourists. The corresponding low-frequency data on total domestic tourism expenditure is sourced from the Tourism Administration, Ministry of Transportation and Communications.

5. Empirical Findings

5.1 Factor Loadings and Thematic Interpretation

Principal Component Analysis (PCA) results are summarized in Table 2. Using the maximum eigenvalue criterion, F2—derived from keywords—demonstrates the strongest predictive relevance for tourism expenditure. The six keywords with the highest absolute factor loadings on F_2 are:

Table 2: The factor loadings of diffusion indices

Keywords for Main Recreational Activities Engaged in During Domestic Travel	Diffusion indices		
	F1	F2	F3
Components of Diffusion index(keywords)	**Ranking of components (by weight)**		
Visiting Waterfalls			–
Hiking			–
Mountain Climbing			–
Camping	3		–
Watching Pandas			–
Whale Watching		2	–
Bird Watching			–
Flower Viewing			–
Cherry Blossom Viewing			–
Watching the Sunrise	7		–
Historic Sites	6		–
Festivals			–
Performances			–
Exhibitions			–
Traditional Arts	4		–
Indigenous Culture		4	–
Farm		3	–
Specialty Cuisine			–
Night Market Snacks			–
Fine Tea Tasting			–
Drinking Coffee		1	–
Afternoon Tea			–
Culinary Classes			–
Swimming			–
Fishing	5		–
Baseball			–
Rock Climbing			–
Cycling			–
Ball Games			–
Road Running			–
Roller Coaster			–
Driving Around			–
Soaking in Hot Springs			–
Strolling			–
***Shopping* (2)**	1	2	–
Watching Movies	2		–
Boating		4	–
Cable Car	7		–
Tourist Factories			–
Hot Air Ballooning			–
First r eigenvalues of the correlation matrix	4.29	24.52	3.89
Variability explained	0.82		

Note: 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 2, we choose those diffusion indices having significant effects on dependent variables in VAR tests in Table 3, that's F1 and F2. 3. Based on the ranking of components, we conclude the higher-weight variables in bold italics, which are "drinking coffee, whale watching, shopping, farm, indigenous culture, and boating."

Table 3: VAR results

Model 1		Model 2		Model 3		Model 4	
Coefficient estimates							
$\hat{\alpha}_1$	2.227*(1.275)	$\hat{\alpha}_2$	0.006(0.013)	$\hat{\alpha}_3$	-0.097*(0.052)	$\hat{\alpha}_4$	-0.004(0.081)
$\hat{\beta}_1$	0.694(0.262)	$\hat{\beta}_{17}$	0.145(0.304)	$\hat{\beta}_{33}$	-0.019(0.287)	$\hat{\beta}_{49}$	-0.040(0.287)
$\hat{\beta}_2$	18.866(21.341)	$\hat{\beta}_{18}$	-44.695 (38.422)	$\hat{\beta}_{34}$	24.636(37.138)	$\hat{\beta}_{50}$	-2.812(22.481)
$\hat{\beta}_3$	-1.529***(6.189)	$\hat{\beta}_{19}$	5.750(8.879)	$\hat{\beta}_{35}$	-9.575(8.358)	$\hat{\beta}_{51}$	2.294(5.655)
$\hat{\beta}_4$	-0.046(3.652)	$\hat{\beta}_{20}$	-0.160 (4.093)	$\hat{\beta}_{36}$	1.514(4.018)	$\hat{\beta}_{52}$	1.029(3.439)
$\hat{\beta}_5$	0.004(0.003)	$\hat{\beta}_{21}$	-0.002(0.003)	$\hat{\beta}_{37}$	-0.003(0.003)	$\hat{\beta}_{53}$	0.00002(0.003)
$\hat{\beta}_6$	1.429***(0.219)	$\hat{\beta}_{22}$	-0.355(0.394)	$\hat{\beta}_{38}$	0.331(0.381)	$\hat{\beta}_{54}$	-0.449*(0.231)
$\hat{\beta}_7$	0.058(0.064)	$\hat{\beta}_{23}$	-0.019(0.091)	$\hat{\beta}_{39}$	0.064(0.086)	$\hat{\beta}_{55}$	-0.0007(0.058)
$\hat{\beta}_8$	-0.013(0.038)	$\hat{\beta}_{24}$	0.065(0.042)	$\hat{\beta}_{40}$	-0.077*(0.041)	$\hat{\beta}_{56}$	0.035(0.054)
$\hat{\beta}_9$	-0.0003(0.011)	$\hat{\beta}_{25}$	-0.0005(0.013)	$\hat{\beta}_{41}$	0.013(0.012)	$\hat{\beta}_{57}$	0.003(0.012)
$\hat{\beta}_{10}$	0.347(0.877)	$\hat{\beta}_{26}$	-1.781(1.58)	$\hat{\beta}_{42}$	1.685(1.527)	$\hat{\beta}_{58}$	-0.077(0.924)
$\hat{\beta}_{11}$	1.091(0.254)	$\hat{\beta}_{27}$	-0.677*(0.365)	$\hat{\beta}_{43}$	0.188(0.344)	$\hat{\beta}_{59}$	-0.147(0.232)
$\hat{\beta}_{12}$	-0.153***(0.262)	$\hat{\beta}_{28}$	0.115(0.168)	$\hat{\beta}_{44}$	-0.090(0.165)	$\hat{\beta}_{60}$	0.054(0.141)
$\hat{\beta}_{13}$	-0.007(0.017)	$\hat{\beta}_{29}$	0.009(0.019)	$\hat{\beta}_{45}$	0.015(0.018)	$\hat{\beta}_{61}$	-0.015(0.018)
$\hat{\beta}_{14}$	-0.465(1.35)	$\hat{\beta}_{30}$	0.539(2.432)	$\hat{\beta}_{46}$	1.22(2.349)	$\hat{\beta}_{62}$	-1.349(1.422)
$\hat{\beta}_{15}$	0.091(0.262)	$\hat{\beta}_{31}$	-0.318(0.562)	$\hat{\beta}_{47}$	0.373(0.529)	$\hat{\beta}_{63}$	-0.450(0.358)
$\hat{\beta}_{16}$	0.323(0.231)	$\hat{\beta}_{32}$	-0.468*(0.259)	$\hat{\beta}_{48}$	-0.130(0.254)	$\hat{\beta}_{64}$	-0.384*(0.218)
det(SSE)	0.00000014						
AIC	-12.558						
BIC	-9.856						
HQ	-11.581						

Note: 1. Here, the results of the VAR model are presented with its four equations labelled sequentially as Model 1, Model 2, Model 3, and Model 4. The following are the specifications for Models 1 through 4:

Model 1: $Y_t = \hat{\alpha}_1 + \hat{\beta}_1 Y_{t-1} + \hat{\beta}_2 Y_{t-2} + \hat{\beta}_3 Y_{t-3} + \hat{\beta}_4 Y_{t-4} + \hat{\beta}_5 DF1_{t-1} + \hat{\beta}_6 F1_{t-2} + \hat{\beta}_7 F1_{t-3} + \hat{\beta}_8 F1_{t-4} + \hat{\beta}_9 F2_{t-1} + \hat{\beta}_{10} F2_{t-2} + \hat{\beta}_{11} F2_{t-3} + \hat{\beta}_{12} F2_{t-4} + \hat{\beta}_{13} F3_{t-1} + \hat{\beta}_{14} F3_{t-2} + \hat{\beta}_{15} F3_{t-3} + \hat{\beta}_{16} F3_{t-4}$

Model 2: $F1_t = \hat{\alpha}_2 + \hat{\beta}_{17} Y_{t-1} + \hat{\beta}_{18} Y_{t-2} + \hat{\beta}_{19} Y_{t-3} + \hat{\beta}_{20} Y_{t-4} + \hat{\beta}_{21} F1_{t-1} + \hat{\beta}_{22} F1_{t-2} + \hat{\beta}_{23} F1_{t-3} + \hat{\beta}_{24} F1_{t-4} + \hat{\beta}_{25} F2_{t-1} + \hat{\beta}_{26} F2_{t-2} + \hat{\beta}_{27} F2_{t-3} + \hat{\beta}_{28} F2_{t-4} + \hat{\beta}_{29} F3_{t-1} + \hat{\beta}_{30} F3_{t-2} + \hat{\beta}_{31} F3_{t-3} + \hat{\beta}_{32} F3_{t-4}$

Model 3: $F2_t = \hat{\alpha}_3 + \hat{\beta}_{33} Y_{t-1} + \hat{\beta}_{34} Y_{t-2} + \hat{\beta}_{35} Y_{t-3} + \hat{\beta}_{36} Y_{t-4} + \hat{\beta}_{37} F1_{t-1} + \hat{\beta}_{38} F1_{t-2} + \hat{\beta}_{39} F1_{t-3} + \hat{\beta}_{40} F1_{t-4} + \hat{\beta}_{41} F2_{t-1} + \hat{\beta}_{42} F2_{t-2} + \hat{\beta}_{43} F2_{t-3} + \hat{\beta}_{44} F2_{t-4} + \hat{\beta}_{45} F3_{t-1} + \hat{\beta}_{46} F3_{t-2} + \hat{\beta}_{47} F3_{t-3} + \hat{\beta}_{48} F3_{t-4}$

Model 4: $F3_t = \hat{\alpha}_4 + \hat{\beta}_{49} Y_{t-1} + \hat{\beta}_{50} Y_{t-2} + \hat{\beta}_{51} Y_{t-3} + \hat{\beta}_{52} Y_{t-4} + \hat{\beta}_{53} F1_{t-1} + \hat{\beta}_{54} F1_{t-2} + \hat{\beta}_{55} F1_{t-3} + \hat{\beta}_{56} F1_{t-4} + \hat{\beta}_{57} F2_{t-1} + \hat{\beta}_{58} F2_{t-2} + \hat{\beta}_{59} F2_{t-3} + \hat{\beta}_{60} F2_{t-4} + \hat{\beta}_{61} F3_{t-1} + \hat{\beta}_{62} F3_{t-2} + \hat{\beta}_{63} F3_{t-3} + \hat{\beta}_{64} F3_{t-4}$

2. The values in parentheses represent standard deviations. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

- *Drinking coffee* (gourmet),
- *Whale watching* (nature sightseeing),
- *Shopping* and *farm* (leisure),
- *Indigenous culture* and *boating* (cultural experience).

These keywords span four core recreational categories, suggesting that multidimensional experiences—often involving transportation, accommodation, and local services—are more likely to drive substantial tourism expenditure.

5.2 VAR Model Results

Table 3 reports the results from the estimated VAR(4) model. Key findings include "*lagged tourism expenditure* (Y_{t-3}) has a significant influence on current expenditure (Y_t)," and "*diffusion index 1 at lag 2* ($F1_{t-2}$) and *diffusion index 2 at lag 4* ($F2_{t-4}$) are significant predictors of Y_t."

These results confirm that keyword-based diffusion indices contain valuable leading information for forecasting domestic tourism expenditure. While no evidence of reverse causality is found, the predictive direction—from online behavior to economic outcome—is robust and statistically significant.

Interestingly, despite high search frequency, keywords such as "*performances*," "*cycling*," "*road running*," and "*shopping*" show limited impact on overall expenditure[4]. This is likely due to their association with low-cost, short-duration, or local activities. For instance: (1) *performances* and *exhibitions* often generate local foot traffic without triggering overnight travel; (2) *cycling* and *road running* typically incur minimal costs and are not tied to tourism-specific services; and (3) *shopping* has increasingly shifted to e-commerce, diminishing its travel-related expenditure impact.

6. Conclusion and Policy Implications

This study identifies six high-impact recreational keywords— *"drinking coffee," "whale watching," "shopping," "farm," "indigenous culture,"* and *"boating"*— as key predictors of Taiwan's domestic tourism expenditure. These findings underscore the importance of experiential tourism activities spanning gourmet, nature sightseeing, other leisure, and cultural experience.

The results align with official surveys conducted by the Tourism Administration, offering empirical support for the economic significance of demand-side behavior captured via online search trends.

To strengthen the value-added potential of Taiwan's domestic tourism sector, the following strategies are proposed: (1) developing sports tourism to attract both

---

[4] Descriptive statistics indicate that the most frequently searched keywords include "*shopping*," "*performances*," "*cycling*," "*exhibitions*," and "*road running*." These keywords reflect prevailing public interests but vary in their economic impact. Due to space limitations, descriptive statistics of the keyword data are not included in this paper. Interested readers may contact the author for further information.

domestic and international visitors, leveraging Taiwan's natural and urban environments; (2) enhancing tourism infrastructure—such as sport-friendly accommodations and facilities—to extend visitor stays and associated spending; and (3) integrating theme parks with surrounding attractions to diversify offerings, lengthen visit duration, and stimulate local economies. These recommendations aim to foster a *diversified, high-value, and experience-driven tourism model*, enhancing Taiwan's competitiveness in the post-pandemic travel economy.

References

[1] Askitas, N. and Zimmermann, K.F. (2009). Google Econometrics and Unemployment Forecasting, German Council for Social and Economic Data (RatSWD) Research Notes No. 41.

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

[3] Boivin, J. and Ng, S. (2005). Understanding and Comparing Factor-Based Forecasts, International Journal of Central Banking, 1(3), pp. 117 - 151.

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

[5] D'Amuri, F. and Marcucci, J. (2010). "Googleit!" Forecasting the US Unemployment Rate with a Google Job Search index, FEEM Working Paper No. 31.2010.

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

[7] Kholodilin, K. A., Podstawski, M. and Siliverstovs, B. (2010). 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.

[8] Klein, L.R. and Park, J.Y. (1993). Economic Forecasting at High-Frequency Intervals, Journal of Forecasting, 12, pp. 301 - 319.

[9] Klein, L.R. and Park, J.Y. (1995). The University of Pennsylvania Model for HighFrequency Economic Forecasting, Economic and Financial Modeling, pp. 95 - 146.

[10] Marcellino, M., Stock, J. and Watson, M. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. European Economic Review, 47(1), pp. 1 - 18

[11] Matsumoto, A., Matsumura, K. and Shiraki, N. (2013). Potential of Search Data in Assessment of Current Economic Conditions, No. 2013-04-18. Bank of Japan.

[12] Stock, J.H. and Watson, M.W. (1998a). Diffusion Indexes, NBER Working Paper No. W6702.

[13] Stock, J.H. and Watson, M.W. (1998b). Forecasting inflation, Journal of Monetary Economics 44, pp. 293 - 335.

[14] Stock, J.H. and Watson, M.W. (2002a). Macroeconomic Forecasting Using Diffusion Indexes, Journal of Business and Economic Statistics, 20(2), pp. 147 - 162.
[15] 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), pp. 1167 - 1179.
[16] 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, pp. 1 - 18.
[17] Taiwan's Ministry of Transportation and Communications (MOTC) (2023). 2023 Taiwan Tourism Survey, 2023 survey from Tourism statistics database of the tourism administration (MOTC).
[18] Vosen, S. and Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends, Journal of Forecasting, 30.6: pp. 565 - 578.
[19] Vosen, S. and Schmidt, T. (2012). A monthly consumption indicator for Germany based on Internet search query data, Applied Economics Letters19.7: pp. 683 - 687.