Global Review of Tourism and Social Science

Global Review of Tourism and Social Sciences

Journal Homepage: https://journal.gpp.or.id/index.php/grtss/index

Is Google Trends a Leading Indicator for Tourism Demand? Evidence from Malaysia, 2014–2024

Md. Akramul Bari[1]*, Ali Abdi Hassan[2]
[1]Faculty of Business, Multimedia University, Malaysia
[2]Seattle University, Washington, United States of America

ARTICLE INFO

Submitted: April 8, 2026
Revised: May 22, 2026
Accepted: June 15, 2026
Section: Economics and Finance

Corresponding Author: Md. Akramul Bari

Corresponding Author Email:
akramul.bari@mmu.edu.my

ABSTRACT: Google Trends search-interest data are widely reported to improve tourism-demand forecasts, but most evidence relies on monthly or weekly series. This study examines whether an annual composite Google Trends Index (GTI), constructed from four Malaysia-related English-language queries, is associated with and adds incremental predictive information for Malaysian international tourist arrivals from 2014 to 2024. The full-sample contemporaneous correlation is very high ($r = 0.952$; $r^2 \approx 0.906$), but the association reverses in the small pre-pandemic window (2015–2019; $r = -0.902$), while the one-year-lagged correlation is near zero ($r = -0.127$, $N = 4$). In nested annual regressions, lagged GTI is not statistically significant and is affected by substantial multicollinearity. A leave-one-year-out sensitivity analysis also yields higher RMSE and MAE for the GTI-augmented model than for the AR(1) benchmark. Because this validation is not temporally ordered and the sample is extremely small, it should not be interpreted as definitive prospective forecasting evidence. Instead, the combined findings indicate that the strong pandemic-inclusive correlation is unstable and likely reflects shared collapse-and-recovery dynamics. Claims about routine forecasting value for Malaysia therefore require replication with a longer monthly series, source-market-sensitive search measures, and rolling-origin validation.

Keywords: Google Trends; tourism demand forecasting; Malaysia; structural break; COVID-19; forecast validation

INTRODUCTION

Since Google made its search-interest index publicly available, researchers have used it to forecast a wide range of economic activity ahead of official statistics, including automobile sales, unemployment claims, and travel demand (Choi & Varian, 2012). Tourism has become one of the most active application areas for this idea, on the logic that a trip typically begins with an online search: a traveller planning a holiday searches for destinations, flights, and hotels weeks or months before arriving, so search volume should rise ahead of — and therefore help predict — realised visitor arrivals (Bangwayo-Skeete & Skeete, 2015; Yang et al., 2015). A substantial literature reports that adding Google Trends data to conventional time-series or econometric forecasting models improves accuracy relative to autoregressive benchmarks, across destinations ranging from the Caribbean to China to South Africa (Botha & Saayman, 2024; Gunter & Önder, 2016; Rashad, 2022).

A smaller but growing body of work urges caution. Search-based indices are sensitive to query selection and aggregation, platform and language coverage, and structural breaks. This concern is especially salient for tourism studies that span the COVID-19 pandemic, when border restrictions sharply reduced both realised travel and the incentive to search for trips. Two series that collapse and recover in the same period can correlate strongly even when neither provides a stable lead over the other. A forecasting model built on such a sample may therefore mistake shared-shock co-movement for an exploitable relationship.

This paper examines that concern using Malaysia as a case study. Malaysia is a useful setting because it is a major Southeast Asian inbound market, its arrivals fell from more than 27 million in 2014 to 134,728 in 2021 and recovered to just over 25 million in 2024, and a national arrivals series can be compared with a consistently constructed search index. The study combines a composite Google Trends Index based on four Malaysia-related queries with annual arrivals for 2014–2024. It addresses three linked questions: how strongly the series are associated over the full sample, whether that association is stable in the pre-pandemic window, and whether lagged GTI adds predictive information beyond an AR(1) benchmark in the available annual data.

The question addressed here also has practical significance beyond Malaysia. National tourism boards and destination marketing organisations across Southeast Asia have increasingly incorporated Google Trends dashboards into routine demand monitoring, often informally and without the kind of subperiod or out-of-sample validation applied in this paper, on the reasonable but untested assumption that a positive correlation observed in recent, pandemic-inclusive data will continue to hold in ordinary planning conditions. If the pattern documented here for Malaysia — a strong pandemic-inclusive correlation that is largely or entirely a shared-shock artefact — generalises to other ASEAN destinations with similarly severe COVID-19-era arrivals collapses, then Google Trends dashboards built or validated primarily on 2020–2024 data may be systematically overstating the tool's value for routine, non-crisis demand planning, even where the same tool may genuinely assist with crisis-recovery nowcasting specifically, as Botha and Saayman (2024) find for South Africa. This distinction between crisis-recovery value and routine planning value is one that the existing literature has not, to the authors' knowledge, tested explicitly for any Southeast Asian destination, and is a second, more applied motivation for the design used in this paper.

The contribution is threefold. First, the study provides a Malaysia-specific examination of the Google Trends-arrivals relationship that explicitly contrasts the pandemic-inclusive sample with a pre-pandemic window. Second, it complements in-sample association measures with a model-comparison sensitivity analysis, while acknowledging that the available annual series is too short for strong prospective forecasting claims. Third, it provides a concrete illustration of the warning by Bokelmann and Lessmann (2019): a very strong full-sample correlation ($r = 0.952$; $r^2 \approx 0.906$) can coexist with unstable subperiod relationships and no demonstrated incremental predictive value in a small annual sample.

LITERATURE REVIEW

Google Trends as an Economic Nowcasting Tool

The use of internet search data to forecast economic activity was popularised by Choi and Varian (2012), who showed that adding Google Trends search categories to simple autoregressive models improved near-term forecasts of automobile sales, home sales, and unemployment claims. Since then, the approach has been extended to a wide range of macroeconomic indicators, including unemployment rate nowcasting using daily and weekly search data, and business-cycle indicators such as purchasing manager indices. The underlying logic in each case is the same: because search activity typically precedes the corresponding real-world action — searching for a house before buying it, searching for unemployment benefits before filing a claim — search volume should carry forward-looking information that official statistics, released with a lag, do not yet reflect.

Beyond unemployment and vacation-destination planning, Google search data have been applied to nowcast a wide range of other economic indicators with generally positive results: D'Amuri and Marcucci (2017) show that an internet job-search index outperforms traditional leading indicators, including the Survey of Professional Forecasters, in predicting the US unemployment rate across multiple business-cycle phases, while Carrière-Swallow and Labbé (2013) demonstrate that a Google Trends automotive-purchase index improves nowcasts of automobile sales in Chile, an emerging-market setting with comparatively low internet penetration, suggesting that search-based nowcasting value is not confined to high-income, high-connectivity economies. These applications share a common structure with the tourism-forecasting literature reviewed below: a search index is used to augment, rather than replace, a conventional autoregressive or econometric benchmark, and the reported gains are evaluated primarily through in-sample fit improvement rather than through the kind of subperiod stability test applied in this paper.

Tourism was one of the earliest and remains one of the most active application domains for this idea, since trip planning is a search-intensive activity spanning destination research, accommodation booking, and flight comparison. Bangwayo-Skeete and Skeete (2015) constructed a composite "hotels and flights" search index for five Caribbean destinations and found that autoregressive mixed-data sampling models incorporating the index outperformed SARIMA and simple autoregressive benchmarks in most out-of-sample forecasting experiments. Yang et al. (2015) compared Google and Baidu search data in forecasting Chinese tourist volumes to a single destination and found that the locally dominant search engine, Baidu, outperformed Google, underscoring that search-engine market share in the origin market matters for forecasting performance. Gunter and Önder (2016) found that Google Analytics website-traffic data improved city-level tourism arrivals forecasts for Vienna beyond what Google Trends search data alone achieved. More recently, Botha and Saayman (2024) extended this literature to the post-pandemic recovery context specifically, using mixed-data sampling models to show that Google Trends data improved South African tourism recovery forecasts one to two quarters ahead, with the benefit growing larger, not smaller, after COVID-19. Rashad (2022) similarly found that a Google-developed tourism recovery tracking product improved forecasts of Dubai's tourism demand in the same recovery period. Taken together, this literature has generally reported positive findings, and recent systematic reviews confirm that search-based indices are now a well-established input to tourism forecasting models across a wide range of destinations and modelling approaches (Song et al., 2019; Li et al., 2021).

Biases, Structural Breaks, and Validation

A parallel strand of literature identifies important limitations in Google Trends-based forecasting. Dergiades et al. (2018) show that language coverage and search-engine market share can bias destination-level indices when visitors originate from multiple countries. Their corrected Cyprus index outperforms a naive aggregate, implying that keyword language, source-market composition, and platform coverage should be treated as modelling decisions rather than neutral data choices.

More directly relevant to the present design, Bokelmann and Lessmann (2019) demonstrate that Google Trends variables can appear predictive in-sample while contributing little to out-of-sample accuracy when common trends, seasonality, or structural breaks drive both the search index and tourism demand. A recent critical assessment reaches a similar conclusion: reported gains are highly sensitive to the sample period, keyword construction, and model specification (Mikulić & Baumgärtner, 2025). These findings motivate explicit subperiod checks and temporally appropriate forecast validation.

Methodologically, the literature has evolved considerably beyond the simple linear regressions used in early Google Trends tourism studies. Mixed-data sampling models that combine high-frequency search data with lower-frequency arrivals series, machine learning approaches such as least absolute shrinkage and selection operator (lasso) variable selection and long short-term memory networks, and hybrid models that combine search indices with macroeconomic covariates have all been applied to improve on simple autoregressive or search-augmented benchmarks (Song et al., 2019; Li et al., 2021). A recurring theme across this more recent, methodologically sophisticated literature is that the incremental value of search data is highly context-dependent: it varies with forecast horizon, data frequency, destination characteristics, and — as Yang et al. (2022) demonstrate directly — with structural features of the destination economy itself, rather than being a fixed, generalisable property of Google Trends data as a predictor class. This context-dependence is precisely why a single-destination, subperiod-comparison design of the kind used in this paper, rather than a pooled cross-country average effect, is informative: it tests whether the relationship holds for one specific, real destination under one specific, real set of economic and search-behaviour conditions, rather than assuming that a positive average effect estimated across many heterogeneous destinations necessarily applies to Malaysia.

Tourism Demand Forecasting in Malaysia

Malaysia-specific forecasting research has generally used conventional time-series or macroeconomic approaches rather than search-based indicators. Kadir and Karim (2012) examined tourism and economic growth using arrivals from ASEAN-5 source markets, while Mohd Lip et al. (2020) compared SARIMA and Holt-Winters models for arrivals from Singapore, Korea, and the United Kingdom, and Arokiasamy et al. (2021) applied an auto-regressive neural network approach to model whether foreign visitor arrivals from ten major source markets would revert to their pre-pandemic trend following COVID-19 border closures, finding substantial heterogeneity in projected recovery speed across source markets. These studies underline source-market heterogeneity and the importance of model choice, but they do not test whether online search intensity adds predictive information to aggregate national arrivals.

Destination-level studies reinforce this picture. Mustafa et al. (2021) forecast international tourist arrivals to Langkawi conditional on the GDP growth of three major source markets — China, Saudi Arabia, and the United Kingdom — using Holt linear trend models, finding that source-market income growth, rather than any destination-side indicator, was the dominant driver of projected post-COVID-19 recovery paths, with substantial heterogeneity across source markets. Together with Kadir and Karim (2012) and Mohd Lip et al. (2020), this body of work establishes that Malaysian tourism forecasting has relied almost entirely on arrivals persistence, macroeconomic source-market conditions, and conventional time-series decomposition. None of these studies incorporates a digital search-interest indicator, and none tests whether an apparent association between a leading indicator and arrivals survives a pre-pandemic or otherwise shock-free subsample — the specific gap this paper addresses. This omission matters because Malaysia's largest source markets, particularly Singapore, Indonesia, and China, are demonstrably heterogeneous in income growth, currency regime, and pandemic recovery trajectory, which is precisely the kind of heterogeneity that a single, English-language, worldwide-aggregated Google Trends index cannot distinguish, and which the source-market-specific forecasting literature suggests may matter more than aggregate search interest for explaining year-to-year movements in national arrivals.

Malaysia's tourism sector is also embedded within broader patterns of ASEAN economic and financial integration that shape year-to-year fluctuations in visitor demand independently of any search-based indicator. Regional financial integration initiatives among the ASEAN-5 economies have been shown to strengthen cross-border capital and, indirectly, travel linkages (Robiyanto et al., 2023), while exchange-rate volatility among ASEAN-5 trading and tourism partners has measurable effects on cross-border commodity and, by extension, travel-related flows (Handoyo et al., 2023). Evidence from tourism-led growth research outside Southeast Asia similarly shows that the tourism-economy relationship can shift materially between ordinary and crisis periods (Gričar et al., 2021; Haller et al., 2021), reinforcing the general methodological point made throughout this paper: relationships estimated over a window that includes a large regional or global shock should not be assumed to hold in steady-state, non-crisis conditions, whether the relationship in question concerns tourism and exchange rates, tourism and economic growth, or — as tested here — tourism and online search interest.

More broadly, digital and behavioural indicators of visitor interest are an increasingly active research theme across Southeast Asian tourism scholarship, even where, as in the present study, aggregate national statistics remain the primary outcome of policy interest. Community-based tourism research from Indonesia, for example, documents how shared-value and participatory models are increasingly used to embed digital engagement and destination marketing into local tourism development (Aminullah & Wusko, 2025), while qualitative case-study evidence on the tourist experience shows that visitor perceptions and destination sustainability are closely intertwined with the same kinds of pre-trip information-seeking behaviour that search-based indices such as the Google Trends Index used in this paper are intended to capture (Yuli, 2024). This literature is a useful complement to the quantitative, national-level test conducted here: where the present paper asks whether aggregate search interest predicts aggregate arrivals, this destination-level literature suggests that the underlying behavioural link between search, interest, and travel may be more visible and more stable at the level of individual destinations, campaigns, or communities than in a single, undifferentiated national time series.

Synthesis of the Literature

The literature therefore presents two competing expectations. Search activity may contain timely information because travel planning precedes realised arrivals, as shown in several destination studies (Bangwayo-Skeete & Skeete, 2015; Yang et al., 2015; Botha & Saayman, 2024). At the same time, search indices can be distorted by language and platform coverage, model overfitting, and common shocks (Dergiades et al., 2018; Bokelmann & Lessmann, 2019). The unresolved question for Malaysia is not merely whether the two series correlate, but whether the association is stable and whether lagged search information adds anything beyond arrivals persistence in the available annual sample.

The study addresses three exploratory research questions:

RQ1: How strongly is the composite Google Trends Index associated with Malaysian international tourist arrivals over 2014–2024?
RQ2: How stable is that association in the pre-pandemic 2015–2019 window?
RQ3: Does lagged Google Trends information add predictive information beyond an AR(1) benchmark in the available annual sample?

METHODS

This exploratory study uses annual data for Malaysia from 2014 to 2024 (N = 11), combining a composite Google Trends search-interest index with international tourist-arrival figures. The single-country design and annual frequency follow the available data rather than an optimal forecasting design. Accordingly, the analysis is intended to diagnose association stability and incremental information in this sample, not to establish a definitive operational forecasting model.

Google Trends data were retrieved from the Google Trends web interface for four Malaysia-related search queries—"Malaysia holiday," "visit Malaysia," "Kuala Lumpur hotel," and "flights to Malaysia"—compared jointly so that the exported series shared the same 0–100 normalisation base, worldwide geographic scope, and monthly resolution from January 2014 through December 2024. A planned fifth query, "Malaysia travel," did not appear as a distinct column in the archived export and was therefore excluded. The exact retrieval date, query type, category setting, and original export should accompany the final article as supplementary documentation because Google Trends values can vary across downloads.

The four monthly series were aggregated to annual averages for 2014–2024, standardised individually, and averaged to form a composite Google Trends Index (GTI), consistent with composite-index applications in the literature (Bangwayo-Skeete & Skeete, 2015; Yang et al., 2015). Standardisation reduces mechanical domination by queries with higher index levels. Because only English-language worldwide queries are used, the GTI should be interpreted as a broad search-interest proxy rather than a complete measure of travel intention across Malaysia's multilingual source markets.

International tourist arrivals were compiled from Tourism Malaysia releases and cross-checked against Malaysia's open arrivals data, Department of Statistics Malaysia tourism publications, and the 2024 parliamentary figure reported by Reuters. The resulting annual series retains the precision reported by the underlying sources. A source-by-year reconciliation table should be included in the supplementary material so that each annual observation can be independently reproduced.

The decision to work at annual rather than monthly frequency reflects a specific and disclosed data-access constraint rather than a methodological preference. A complete, independently verifiable monthly series of Malaysian international tourist arrivals spanning the full 2014–2024 window was not available through the sources accessible to the authors at the time of writing; Tourism Malaysia's monthly bulletins are not uniformly archived in a single public location covering the entire period, and third-party aggregators that do report monthly figures typically require paid subscriptions for historical series or provide only the most recent one to two months without a downloadable historical table. Annual totals, by contrast, are independently reported and cross-checkable across multiple public sources — government releases, wire-service reporting, and parliamentary records — for every year in the sample, which is why the annual series, despite its coarser temporal resolution, was judged more reliable and more fully verifiable than a monthly series that could not be independently reconstructed and checked line by line. Readers seeking to replicate or extend this analysis at monthly frequency would need to obtain the underlying monthly bulletins directly from Tourism Malaysia or the Department of Statistics Malaysia, which the authors were unable to compile into a complete verified series for the full sample period.

Table 1. Variable Definitions and Data Sources

Variable	Measurement / Source	Expected sign
International tourist arrivals (dependent)	Annual international tourist arrivals to Malaysia, compiled from Tourism Malaysia releases and cross-checked against data.gov.my, DOSM, and Reuters.	n/a

Google Trends Index (GTI)	Composite z-score average of four Malaysia-related search terms, annualised from monthly Google Trends data (worldwide, 2014–2024).	+/− (tested)
Arrivals (t−1)	One-year lagged international tourist arrivals, used as the autoregressive benchmark predictor.	+
GTI (t−1)	One-year lagged Google Trends Index, used to test leading-indicator value.	+/− (tested)

Three analytical steps address the research questions. First, Pearson correlation coefficients are calculated between the GTI and arrivals, contemporaneously and at a one-year lag, for the full 2014–2024 sample and for the pre-pandemic 2015–2019 window. Because the latter contains only five contemporaneous and four lagged observations, these coefficients are descriptive indicators of instability rather than precise population estimates.

As an additional robustness check on the Pearson correlations reported above, Spearman rank correlations are also computed for both the full sample and the pre-pandemic subsample. Rank-based correlation is less sensitive than the Pearson coefficient to the small number of extreme observations that dominate an eleven-year series spanning a pandemic collapse and recovery, and provides a useful check on whether the reported pattern of reversal is an artefact of a few large outlying values (specifically, 2020 and 2021) rather than a more general feature of the ranking of years by arrivals and search interest.

Second, two nested ordinary least squares models are estimated. Model A is an AR(1) benchmark, regressing current-year arrivals on one-year-lagged arrivals. Model B adds the one-year-lagged GTI. This structure follows the common test of whether search data add information beyond arrivals persistence (Choi & Varian, 2012; Bangwayo-Skeete & Skeete, 2015). With only ten usable observations after lagging, coefficient significance, information criteria, and variance inflation factors are reported descriptively and should not carry the inferential weight they would have in a longer series.

The nested model comparison in Table 4, and the use of information criteria (AIC, BIC) alongside coefficient significance to adjudicate between the AR(1) benchmark and the GTI-augmented model, follows the general logic of formal model-specification testing established in the econometrics literature (Hausman, 1978), applied here in its simplest form given the very small sample rather than through a formal Hausman-type test statistic. Because the analysis in this paper is a single-country annual time series rather than a cross-sectional panel, standard panel-data robust covariance estimators are not directly applicable; nonetheless, the general principle that inference in small, serially dependent samples requires caution about standard-error reliability is well established in both the time-series literature on heteroskedasticity- and autocorrelation-consistent covariance estimation (Newey & West, 1987) and the panel-data literature on comparing alternative standard-error correction approaches (Petersen, 2009; Baltagi, 2021). This is one further reason the coefficient p-values reported in Table 4 are treated as descriptive rather than strongly inferential, consistent with the treatment of all significance tests throughout this paper.

Third, model sensitivity is examined using leave-one-year-out cross-validation (LOYOCV). Each of the ten usable observations is omitted in turn, the model is re-estimated on the remaining nine, and an error is calculated for the omitted year. RMSE and MAE summarise the resulting errors (Hyndman & Koehler, 2006). LOYOCV is useful here as a small-sample influence diagnostic, but it is not a prospective time-series forecast design because later observations can enter the training set for earlier holdout years. The comparison therefore addresses whether the GTI-augmented specification is more stable under observation deletion; a rolling-origin or expanding-window evaluation is required for a stronger operational forecasting test.

RESULTS

Descriptive Overview

Table 2 summarises the annual series. Malaysian international tourist arrivals ranged from a low of 134,728 in 2021, at the depth of pandemic-related border closures, to a high of 27.4 million in 2014. The composite Google Trends Index also reached its sample minimum in 2021 and recovered thereafter. The descriptive pattern therefore suggests that the full-sample co-movement is strongly influenced by the collapse-and-recovery period, a possibility examined in the subperiod and model-sensitivity analyses below.

Table 2. Descriptive Statistics, Malaysia 2014–2024 (N = 11 years)

Variable	Mean	Std. Dev.	Min	Max
International tourist arrivals (millions)	18.95	9.83	0.13	27.44
Google Trends Index (composite, z-score)	0.09	0.98	−2.00	1.18
"Malaysia holiday" (raw index, 0–100)	44.18	9.99	22.92	55.33
"visit Malaysia" (raw index, 0–100)	13.10	4.04	4.67	17.17
"Kuala Lumpur hotel" (raw index, 0–100)	56.81	18.20	21.75	83.92
"flights to Malaysia" (raw index, 0–100)	10.92	3.34	4.42	16.42

Correlation Analysis: Full Sample vs. Pre-Pandemic Subsample

Table 3 reports the study's central descriptive result. Across 2014–2024, the composite GTI is very strongly associated with contemporaneous arrivals ($r = 0.952$; $r^2 \approx 0.906$), while the one-year-lagged correlation is lower ($r = 0.649$). In the pre-pandemic window, the contemporaneous coefficient reverses sign ($r = -0.902$) and the lagged coefficient is near zero ($r = -0.127$). The sharp contrast answers RQ1 and RQ2 by showing that the pandemic-inclusive association is large but not stable across the examined windows.

Table 3. Correlation Between Google Trends Index and Tourist Arrivals: Full Sample vs. Pre-Pandemic Subsample

Correlation	Full sample (2014–2024)	N	Pre-pandemic (2015–2019)	N
Contemporaneous (GTI_t, $Arrivals_t$)	0.952	N = 11	−0.902	N = 5
One-year lag (GTI_{t-1}, $Arrivals_t$)	0.649	N = 10	−0.127	N = 4

The pre-pandemic estimates are based on only five contemporaneous and four lagged observations. They cannot establish the magnitude or sign of the underlying pre-pandemic relationship with statistical precision. Their appropriate use is narrower: they show that the full-sample coefficient is highly sensitive to the selected period and is therefore insufficient, by itself, to demonstrate a stable leading-indicator mechanism.

Table 4 reports the two nested OLS models. In Model A, lagged arrivals are positive and marginally significant ($\beta = 0.613$, $p = 0.054$), with $R^2 = 0.389$. In Model B, adding lagged GTI raises R^2 only to 0.421; neither lagged-arrivals ($p = 0.962$) nor lagged-GTI ($p = 0.552$) is significant, and both AIC and BIC increase. The VIF of 11.28 for each predictor indicates substantial multicollinearity. Given N = 10, these results should be interpreted as descriptive evidence that the augmented annual specification is unstable, not as a definitive parameter test.

Table 4. AR(1) Benchmark vs. AR(1) + Google Trends Index (Dependent variable: $Arrivals_t$)

Variable	Model A: AR(1) only	Model B: AR(1) + GTI
Arrivals (t−1)	0.613* (0.271)	0.047 (0.948)
GTI (t−1)	—	6,163,000 (9,860,000)
Constant	7,212,000 (5,870,000)	18,260,000 (18,700,000)
Observations	10	10
R^2	0.389	0.421
AIC	349.18	350.64
BIC	349.79	351.55
VIF (lagged predictors)	n/a	11.28

Note. Standard errors in parentheses. * $p < 0.10$. No coefficient in Model B reaches conventional significance. AIC and BIC are lower (better) for Model A in both cases.

Model Sensitivity Under Leave-One-Year-Out Cross-Validation

Table 5 reports the leave-one-year-out sensitivity analysis. Model B's RMSE is 9.98 million arrivals, 5.8% higher than Model A's 9.43 million; its MAE is 8.16 million, 12.8% higher than Model A's 7.23 million. Thus, deleting one annual observation at a time does not reveal a stability advantage from adding lagged GTI. This answers RQ3 cautiously: the available annual sample does not demonstrate incremental predictive information from GTI. Because the folds are not temporally ordered, the result must not be described as definitive prospective out-of-sample forecast performance.

Individual Search-Term Contributions and Rank-Based Robustness

The composite Google Trends Index aggregates four search terms that differ conceptually: "visit Malaysia" and "Malaysia holiday" most directly capture demand-side travel intent, while "Kuala Lumpur hotel" and "flights to Malaysia" capture booking-stage search behaviour closer to the point of an actual trip. Table 6 decomposes the full-sample and pre-pandemic correlations by individual term, to test whether the reversal documented for the composite index in Table 3 is uniform across terms or concentrated in particular ones.

Table 6. Individual Search-Term Correlations: Full Sample vs. Pre-Pandemic Subsample

Search term	Full sample r (2014–2024)	Pre-pandemic r (2015–2019)	Change
"Malaysia holiday"	0.922	−0.846	−1.768
"visit Malaysia"	0.903	−0.615	−1.518
"Kuala Lumpur hotel"	0.928	−0.085	−1.013
"flights to Malaysia"	0.873	−0.461	−1.334
Composite GTI	0.952	−0.902	−1.854

Every individual term shows the same qualitative pattern documented for the composite index: a strong positive full-sample correlation that weakens or reverses in the pre-pandemic subsample. The magnitude of the reversal varies, however. "Kuala Lumpur hotel", the most infrastructure- and supply-oriented of the four terms, has the largest full-sample correlation ($r = 0.928$) but the smallest pre-pandemic correlation in absolute terms ($r = -0.085$, essentially zero), suggesting that its full-sample relationship with arrivals is almost entirely a pandemic-collapse artefact with no discernible steady-state association at all. "Malaysia holiday", by contrast, shows the strongest pre-pandemic reversal ($r = -0.846$), indicating that even a demand-intent search term does not behave as a stable positive leading indicator once the pandemic years are removed. No individual term in this composite retains a positive pre-pandemic correlation with arrivals, which strengthens the paper's central conclusion: the instability documented for the composite GTI is not an artefact of how the four terms were combined, but is present at the level of every underlying query.

Table 7. Spearman Rank Correlation Robustness Check (Composite GTI)

Correlation	Full sample ϱ	p-value	Pre-pandemic ϱ	p-value
Contemporaneous (GTI_t, $Arrivals_t$)	0.782	0.004	−0.900	0.037
One-year lag (GTI_{t-1}, $Arrivals_t$)	0.673	0.033	0.400	0.600

Table 7 confirms that the reversal documented using Pearson correlation is not driven by the small number of extreme pandemic-year observations. The Spearman rank correlation, which depends only on the ordering of years rather than the magnitude of the collapse and recovery, shows the same qualitative pattern: a strong, statistically significant positive rank correlation over the full sample ($\varrho = 0.782$, $p = 0.004$) that reverses to a strong negative rank correlation in the pre-pandemic subsample ($\varrho = -0.900$, $p = 0.037$), a reversal that remains statistically significant despite the small pre-pandemic sample size. At the one-year lag, the rank correlation falls from a modest but significant full-sample value ($\varrho = 0.673$, $p = 0.033$) to a small and non-significant pre-pandemic value ($\varrho = 0.400$, $p = 0.600$, $N = 4$), mirroring the Pearson result in Table 3. Because this pattern holds under both a correlation measure sensitive to the magnitude of the pandemic collapse (Pearson) and one that is not (Spearman), the instability documented in this paper cannot be attributed simply to the extreme scale of the 2020–2021 outliers; it reflects a genuine reordering of years between the full sample and the pre-pandemic subsample.

Table 5. Leave-One-Year-Out Cross-Validation: Model-Sensitivity Errors

Metric	Model A: AR(1) only	Model B: AR(1) + GTI	Error change (B relative to A)
RMSE (arrivals, millions)	9.43	9.98	+5.8% (higher)
MAE (arrivals, millions)	7.23	8.16	+12.8% (higher)

DISCUSSION

The central result is a divergence between a very strong pandemic-inclusive association ($r = 0.952$; $r^2 \approx 0.906$) and an absence of demonstrated incremental information from lagged GTI in the small annual models. Taken alone, the full-sample correlation resembles the positive evidence often reported in search-based tourism forecasting (Bangwayo-Skeete & Skeete, 2015; Botha & Saayman, 2024; Rashad, 2022). The subperiod reversal, high multicollinearity, and higher LOYOCV errors show, however, that the association is highly sample-dependent and cannot be treated as evidence of a stable annual leading indicator.

This pattern is consistent with the general warning by Bokelmann and Lessmann (2019) that shared trends and structural breaks can create apparently predictive Google Trends relationships. It also complements the bias concerns identified by Dergiades et al. (2018). In Malaysia, both arrivals and travel-related search interest fell sharply during border restrictions and then recovered. That common collapse-and-recovery path is a plausible explanation for much of the full-sample co-movement, although the present annual design cannot identify a causal mechanism.

The result does not necessarily contradict Botha and Saayman (2024), who find gains from Google Analytics in South African tourism-recovery forecasts. Their quarterly mixed-data design uses higher-frequency search information and a richer covariate set, whereas the present study uses eleven annual observations and a sparse AR benchmark. Search behaviour may be useful for short-horizon recovery nowcasting while remaining too aggregated to improve routine annual forecasts. Frequency, forecast horizon, data construction, and evaluation design are therefore central to reconciling apparently different findings.

The null result reported here for Malaysia is not an isolated finding. In the most directly comparable large-sample test available, Yang, Fan, Jiang, and Liu (2022) applied lasso variable selection to daily tourism-demand forecasting across 74 countries during 2020 and found that, in general, online search queries did not generate more accurate forecasts than benchmark models; where search data did help, the benefit was systematically concentrated in countries with more severe pandemic exposure, a higher dependence on inbound tourism, and island geography. Malaysia is a large, income-diversified, multi-source-market economy for which inbound tourism, while economically important, is not the singular economic dependency it is for small island states, and its 2021 collapse, while severe, was not uniquely extreme relative to the 74-country sample Yang et al. examine. Read against their cross-country evidence, the absence of incremental forecasting value found here is not an anomaly specific to this paper's annual Malaysian data; it is consistent with a broader pattern in which search-based indicators add the least value for precisely the kind of large, diversified, multi-market destination that Malaysia represents.

The instability documented here also has a direct parallel in recent work on the pandemic-era relationship between tourism demand and its determinants more broadly. Liu, Wen, Liu, and Song (2024) show, using time-varying-parameter mixed-data-sampling models, that the relationship between tourism demand and its usual determinants was itself unstable throughout the COVID-19 period, changing systematically in strength and even sign as the crisis evolved from initial shock through prolonged restriction to recovery. Their solution is not to abandon determinant-based forecasting but to let the estimated relationship vary over time rather than assuming a single fixed coefficient across the whole sample — an approach this paper's simple two-period comparison (full sample versus pre-pandemic subsample) only crudely approximates. This suggests a constructive reframing of the present findings: rather than concluding that Google Trends has no relationship whatsoever to Malaysian tourism demand, the more precise conclusion is that any such relationship is time-varying rather than fixed, and that a single pooled annual coefficient, whether estimated over eleven pandemic-inclusive years or over five pre-pandemic years, is the wrong object to estimate in the first place. A time-varying-parameter or regime-switching specification, applied to a longer or higher-frequency Malaysian series, would be a natural next step and would test directly whether the Google Trends-arrivals relationship strengthens, weakens, or reverses across identifiable phases of the pandemic and recovery rather than assuming it is constant within each subsample.

The term-level results in Table 6 add a further layer of caution beyond the composite-index finding. That "Kuala Lumpur hotel" — arguably the search term most directly tied to an imminent, bookable trip — shows almost no pre-pandemic association with arrivals ($r = -0.085$) is difficult to reconcile with the theoretical mechanism underlying most of the Google Trends tourism-forecasting literature, in which search interest is presumed to rise ahead of a trip precisely because travellers search for accommodation as part of concrete trip planning (Bangwayo-Skeete & Skeete, 2015; Yang et al., 2015). If booking-stage search interest does not track arrivals even loosely in ordinary, non-crisis years for a destination as large and well-established as Malaysia, this raises a further question, beyond the scope of the annual data used here, about whether the search terms most theoretically suited to capturing near-term travel intent are in practice the ones for which Malaysia's aggregate worldwide Google Trends index is least informative — possibly because a large share of actual bookings for Malaysia are conducted through the search engines, travel agents, or local-language platforms of specific major source markets such as Singapore, Indonesia, and China, none of which is well captured by an English-language, worldwide-aggregated query.

The limitations are substantial. Annual aggregation removes seasonality and short planning lags that motivate search-based forecasting; the pre-pandemic window contains only four or five observations; English-language worldwide queries may not represent Malaysia's major source markets; and LOYOCV is not temporally ordered. A stronger replication should use verified monthly arrivals, multilingual and source-market-specific search measures, archived query settings, and rolling-origin evaluation. Mixed-data and richer demand models used in the wider literature (Bangwayo-Skeete & Skeete, 2015) would then provide a more credible test of operational forecast value.

CONCLUSION

This study examined whether a composite Google Trends Index is associated with and adds predictive information for Malaysian international tourist arrivals from 2014 to 2024. The full-sample contemporaneous association is very strong, but it reverses in the small pre-pandemic window, while the lagged association is near zero in that window. In the nested annual models, lagged GTI is not significant, is highly collinear with lagged arrivals, and does not reduce leave-one-year-out sensitivity errors. The evidence therefore does not establish a stable annual leading-indicator relationship for Malaysia; it is more consistent with a pandemic-inclusive association that is highly dependent on the chosen sample period.

The paper's main contribution is methodological caution. It provides a Malaysia-specific illustration of how a large contemporaneous correlation can coexist with unstable subperiod results and no demonstrated incremental information in a small lagged model, reinforcing the concerns raised by Bokelmann and Lessmann (2019). For practitioners, the implication is not that Google Trends is inherently unusable, but that pandemic-inclusive correlations should not be converted into routine planning tools without higher-frequency data, source-market validation, and temporally ordered forecast testing.

Future research should first reproduce the annual series from an archived source-by-year data file and reconcile all descriptive statistics with the processing code. It should then repeat the analysis with monthly arrivals and rolling-origin forecasts, test multilingual and source-market-specific queries, and compare Malaysia with other ASEAN destinations. Until those steps are completed, the present findings should be read as an exploratory diagnosis of instability rather than a definitive rejection of search-based tourism forecasting.

Beyond the monthly-data and source-market-specific replication already noted, three further extensions would sharpen the conclusions reached here. First, applying a time-varying-parameter or regime-switching model, in the spirit of Liu et al. (2024), to a longer Malaysian series would test directly whether the Google Trends-arrivals relationship strengthens, weakens, or reverses across distinct phases of the pandemic and recovery, rather than assuming a single fixed relationship within each of the two subsamples compared here. Second, decomposing the composite index into its individual search terms, as Table 6 begins to do for this annual series, and testing each term's relationship separately at monthly frequency would help identify whether particular categories of search behaviour — destination research, accommodation booking, or flight search — carry more genuine forecasting information than others, rather than treating "Google Trends" as a single undifferentiated predictor. Third, given that the individual-term results in Table 6 raise the possibility that Malaysia's major source markets are poorly represented by an English-language, worldwide-aggregated query, future work should test source-market-specific and multilingual search indices — for example, Chinese-language queries for the Chinese market, Bahasa Indonesia queries for the Indonesian market — following the language- and platform-bias corrections proposed by Dergiades et al. (2018) for exactly this kind of multi-source-market destination. Until these extensions are carried out, the present study's contribution should be read as a documented instability and a methodological caution specific to the annual, composite, worldwide-aggregated Google Trends index tested here, rather than as a comprehensive verdict on whether any form of online search data could ever add forecasting value for Malaysian tourism.

Funding. The authors declare that this research received no external funding.
Conflict of Interest. The authors declare that they have no known financial or personal relationships that could have influenced the work reported in this article.
Data Availability. The country-year data used in this study were compiled from publicly available UN Tourism and Our World in Data sources described in the Methods section. The processed dataset and analysis code are available from the corresponding author upon reasonable request.
Author Contributions. M.A.B.: Conceptualisation, methodology, formal analysis, data curation, writing—original draft, and project administration. A.A.H.: Conceptualisation, validation, and writing—review and editing.

REFERENCES

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. Tourism Management, 46, 454–464. https://doi.org/10.1016/j.tourman.2014.07.014

Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google Trends data: An analysis of the effects on tourism demand forecasting in Germany. Tourism Management, 75, 1–12. https://doi.org/10.1016/j.tourman.2019.04.015

Botha, I., & Saayman, A. (2024). Does Google Analytics improve the prediction of tourism demand recovery? Forecasting, 6(4), 908–924. https://doi.org/10.3390/forecast6040045

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88(s1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. Tourism Management, 66, 108–120. https://doi.org/10.1016/j.tourman.2017.10.014

Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. Annals of Tourism Research, 61, 199–212. https://doi.org/10.1016/j.annals.2016.10.007

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Kadir, N., & Karim, M. Z. A. (2012). Tourism and economic growth in Malaysia: Evidence from tourist arrivals from ASEAN-5 countries. Ekonomska Istraživanja, 25(4), 1089–1100. https://doi.org/10.1080/1331677X.2012.11517550

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. Tourism Management, 83, 104245. https://doi.org/10.1016/j.tourman.2020.104245

Mikulić, J., & Baumgärtner, R. M. (2025). Google Trends and Baidu Index data in tourism demand forecasting: A critical assessment of recent applications. Tourism Management, 110, 105164. https://doi.org/10.1016/j.tourman.2025.105164

Mohd Lip, N., Jumery, N. S., Ahmad Termizi, F. A., Mulyadi, N. A., Anuar, N., & Ithnin, H. (2020). Forecasting international tourist arrivals in Malaysia using SARIMA and Holt-Winters model. Journal of

Tourism, Hospitality and Environment Management, 5(18), 41–51. https://doi.org/10.35631/JTHEM.518004

Rashad, A. S. (2022). The power of travel search data in forecasting the tourism demand in Dubai. Forecasting, 4(3), 674–684. https://doi.org/10.3390/forecast4030036

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting. Annals of Tourism Research, 75, 338–362. https://doi.org/10.1016/j.annals.2018.12.001

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. Tourism Management, 46, 386–397. https://doi.org/10.1016/j.tourman.2014.07.019

Mustafa, H., Ahmed, F., Zainol, W. W., & Enh, A. M. (2021). Forecasting the impact of gross domestic product (GDP) on international tourist arrivals to Langkawi, Malaysia: A post-COVID-19 future. Sustainability, 13(23), 13372. https://doi.org/10.3390/su132313372

Yang, Y., Fan, Y., Jiang, L., & Liu, X. (2022). Search query and tourism forecasting during the pandemic: When and where can digital footprints be helpful as predictors? Annals of Tourism Research, 93, 103365. https://doi.org/10.1016/j.annals.2022.103365

Liu, Y., Wen, L., Liu, H., & Song, H. (2024). Predicting tourism recovery from COVID-19: A time-varying perspective. Economic Modelling, 135, 106706. https://doi.org/10.1016/j.econmod.2024.106706

Aminullah, A., & Wusko, A. U. (2025). Pentahelix model to create shared value in empowering the people of Winong Village, Gempol District, Pasuruan Regency. Global Review of Tourism and Social Sciences, 1(2), 59–67. https://doi.org/10.53893/grtss.v1i2.348

Arokiasamy, A. R. A., Smith, P. M. R., & Kijbumrung, T. (2021). Conceptualizing post-COVID-19 Malaysia's tourism recovery: An auto-regressive neural network analysis. Emerging Science Journal, 5, 119–129. https://doi.org/10.28991/esj-2021-SPER-10

Baltagi, B. H. (2021). Econometric analysis of panel data (6th ed.). Springer. https://doi.org/10.1007/978-3-030-53953-5

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. Journal of Forecasting, 32(4), 289–298. https://doi.org/10.1002/for.1252

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. International Journal of Forecasting, 33(4), 801–816. https://doi.org/10.1016/j.ijforecast.2017.03.004

Gričar, S., Bojnec, Š., Karadžić, V., & Backović Vulić, T. (2021). Tourism-led economic growth in Montenegro and Slovenia. Economic Research-Ekonomska Istraživanja, 34(1), 3401–3420. https://doi.org/10.1080/1331677X.2021.1875858

Haller, A. P., Butnaru, G. I., Hârşan, G. D. T., & Ştefănică, M. (2021). The relationship between tourism and economic growth in the EU-28. Is there a tendency towards convergence? Economic Research-Ekonomska Istraživanja, 34(1), 1121–1145. https://doi.org/10.1080/1331677X.2020.1819852

Handoyo, R. D., Alfani, S. P., Ibrahim, K. H., Sarmidi, T., & Haryanto, T. (2023). Exchange rate volatility and manufacturing commodity exports in ASEAN-5: A symmetric and asymmetric approach. Heliyon, 9(2), e13067. https://doi.org/10.1016/j.heliyon.2023.e13067

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251–1271. https://doi.org/10.2307/1913827

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708. https://doi.org/10.2307/1913610

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. The Review of Financial Studies, 22(1), 435–480. https://doi.org/10.1093/rfs/hhn053

Robiyanto, R., Nugroho, B. A., Handriani, E., & Frensidy, B. (2023). Measuring the effectiveness of ASEAN-5 initiatives from emerging market portfolio's perspective. Cogent Business & Management, 10(1), 2167292. https://doi.org/10.1080/23311975.2023.2167292

Yuli, S. B. C. (2024). Understanding the dynamics of tourist experience through a qualitative lens: A case study approach in Indonesia. Global Review of Tourism and Social Sciences, 1(1), 30–40. https://doi.org/10.53893/grtss.v1i1.323