

Forecasting Chinese Outbound Tourism Recovery:

A Triple-layer Forecast Combination Framework

Abstract

Forecast combinations became particularly significant in the post-pandemic era due to heightened uncertainty. This study introduces a Triple-layer Forecast Combination Framework to predict Chinese outbound tourism recovery from August 2023 to July 2024 across 20 destinations. The framework integrates baseline quantitative models, expert-based model selection, and real-time judgmental adjustments to enhance forecast accuracy in post-crisis contexts. Results show Chinese visitor arrivals rebounding, on average, to 80% of July 2019 levels by mid-2024, with East and Southeast Asia—particularly Hong Kong SAR, Macao SAR, and Thailand—recovering faster than long-haul markets such as Hawaii, Canada, and the Czech Republic. By combining statistical rigor with contextual insight, the framework supports replicable, adaptive forecasting under uncertainty for tourism recovery planning.

Keywords: Forecast combination; Delphi method; Judgmental adjustments; Chinese outbound tourism; Recovery pattern

1. Introduction

The COVID-19 pandemic caused an unprecedented disruption to global tourism, and China's outbound tourism sector was among the most severely affected. Prior to the pandemic, China was the world's largest source of outbound tourists, with about 155 million international trips in 2019. Chinese travelers spent over USD 254.6 billion abroad that year, making them the top spenders worldwide (UN Tourism, 2024). This dominance sharply reversed in 2020 when the pandemic caused outbound travel to plummet by over 80% (China Tourism Academy, 2021). The pandemic effectively halted outbound travel for nearly three years, and international arrivals from China dropped to near zero.

Since early 2023, China's full border reopening has triggered a pronounced resurgence in outbound travel, with significant implications for global destination markets. Chinese outbound trips soared to 101 million, and international tourism expenditure hit US\$196.5 billion (UN Tourism, 2024). This re-establishes China as the world's leading nation in international tourism spending, unequivocally demonstrating the robust recovery momentum and immense scale of its outbound tourism market. Such protracted and complex recovery trajectory poses a critical challenge for tourism planners and businesses worldwide, who require reliable forecasts to inform recovery strategies, resource allocation, and marketing efforts in the post-pandemic era.

Several international destinations, particularly in the Asia-Pacific region, remain highly dependent on Chinese visitors; Chinese tourists comprised 25%–30% of all international arrivals in Japan, Korea (ROK), and Thailand before the pandemic and over 60% in the Maldives (Oxford Economics, 2023). The economic contributions of Chinese visitors have been substantial, accounting for as much as 2.7% of Thailand's gross domestic product (GDP) and 0.8% of Japan's before the pandemic (Bank of Thailand, 2019; Japan National Tourism Organization, 2019). Accurate forecasting of Chinese outbound tourism recovery is therefore essential for policymakers, destination-management organizations, and tourism businesses to implement proactive arrangements; allocate resources efficiently; support capacity planning, marketing, and policy formulation (Song & Zhang, 2025).

Yet, forecasting tourism demand following a global crisis remains challenging. The COVID-19 pandemic significantly disrupted tourism demand, introducing heightened volatility and structural changes and making many traditional forecasting models ineffective (Gössling et al., 2021; Liu et al., 2021). Conventional statistical methods, built on assumptions of stationarity, struggle to address “structural change” from sudden policy-induced shocks and rapidly changing tourist preferences (Gössling et al., 2021; Korinth, 2022). Throughout the pandemic, models relying on stable parameters struggled to capture the sudden decline in global travel and the subsequent unstable rebound (Wu et al., 2017). As Korinth (2022) demonstrated in Poland, even established models like the autoregressive integrated moving average model saw their accuracy deteriorated due to drastic, policy-driven demand shifts. These challenges highlight the growing consensus that post-crisis tourism forecasting necessitates more robust and adaptive methodologies capable of addressing macroeconomic volatility, geopolitical uncertainty, and rapidly evolving market structures (Liu et al., 2025).

Recent literature has responded to these challenges with various methodological innovations.

Judgmental forecasting methods, such as Delphi surveys and scenario analysis, are especially effective in managing heightened uncertainty (Kourentzes et al., 2021; Zhang et al., 2021), drawing on experts' insights for *ex-ante* forecasts valuable during crises (Lin & Song, 2015). Other innovations include pandemic-related indices that track variables such as infection rates, vaccine coverage, travel restrictions, and policy interventions to inform nuanced, scenario-based models (Liu et al., 2021; Zhang & Lu, 2022). Furthermore, forecast combination techniques have become widespread, demonstrating improved accuracy, especially during crises (Song et al., 2009; Zhang et al., 2021). The persistent effects of crises have necessitated a methodological shift towards more dynamic frameworks, leveraging historical recovery rates from past crises (e.g., SARS, the Global Financial Crisis, and Middle East respiratory syndrome) (Fotiadis et al., 2021; Polyzos et al., 2020), high-frequency data (Qiu et al., 2021), and integration of quantitative and qualitative forecasts (i.e., hybrid approaches).

Despite these advancements, current methodologies, when used in isolation, still face critical limitations in addressing the complex and dynamic recovery of markets like China's outbound tourism. Judgmental forecasting offers valuable expert insight but may lack consistency. Pandemic indices present useful macro-level indicators but often fail to capture micro-level behavioral shifts (Liu et al., 2025). Artificial intelligence (AI) models, while powerful, typically require extensive, stable high-frequency data often unavailable post-crisis. Simultaneously, traditional statistical models, though sometimes perceived as less agile, retain significant value for interpretability, structure, and data efficiency (Liu et al., 2022). The key, therefore, lies not in replacing traditional or judgmental models but in strategically combining diverse forecasting methods to create a more resilient and context-aware framework.

Against this backdrop, this study seeks to answer the research question: How can a robust and adaptive forecasting methodology be developed to accurately predict post-crisis Chinese outbound tourism recovery by systematically integrating quantitative and qualitative insights? This leads to two primary objectives: (i) to develop a more robust and adaptive forecasting methodology specifically tailored to post-crisis tourism recovery and (ii) to provide actionable recovery insights for tourism industry practitioners and destination-management organizations, specifically in the Chinese outbound market.

This study makes a unique methodological contribution by introducing a structured, multilayered forecasting framework—the Triple-layer Forecast Combination Framework—that systematically and transparently integrates statistical, judgmental, and contextual components. Unlike previous approaches that rely on either data-driven models or expert input in isolation, this framework formalizes the integration process across three stages: baseline quantitative forecasting, expert-driven model selection, and real-time judgmental adjustments. This phased structure surpasses *ad hoc* combination methods by offering a replicable and theoretically grounded approach to forecast construction in volatile and uncertain conditions. Furthermore, it emphasizes the continuing relevance of traditional models—not as standalone tools but as essential building blocks when used within a thoughtfully designed ensemble. By reconciling methodological rigor with contextual adaptability, this framework advances the methodological discourse on crisis-era tourism forecasting and offers a roadmap for future research on post-disruption demand modeling.

The remainder of this paper is structured as follows. Section 2 reviews relevant literature on tourism demand forecasting, combined forecasting, model selection, and the impacts of crises, identifying gaps this study addresses. Section 3 introduces the proposed methodology and outlines the Triple-layer Forecast Combination Framework's implementation. Section 4 presents and discusses the empirical results. Section 5 concludes with key findings, methodological contributions, managerial implications, and directions for future research.

2. Literature Review

Foundational forecasting models and their limitations in crises

Tourism demand forecasting traditionally uses three main families of models. Time series models, such as Exponential Smoothing and Seasonal Autoregressive Integrated Moving Average, are well-suited to capturing historical patterns and seasonality in stable environments, owing to their simplicity and transparency (De Livera et al., 2011; Hyndman & Khandakar, 2008). Econometric models incorporate causal variables to explain demand fluctuations, making them particularly valuable for policy analysis and scenario evaluation (Lin et al., 2015). More recently, AI-based techniques have been adopted for their ability to model complex, nonlinear relationships in large datasets with fewer statistical assumptions (Polyzos et al., 2020). Despite the emergence of advanced methods, traditional models remain foundational in the field due to their interpretability and demonstrated effectiveness under normal conditions.

However, the core assumption of these classical models—that future patterns will resemble the past—is fundamentally violated during and after major crises. The unprecedented shocks, structural breaks, and extreme volatility associated with events like the COVID-19 pandemic render historical data an unreliable guide to future behavior (Gössling et al., 2021; Liu et al., 2021; Korinth, 2022). Consequently, in such conditions, reliance on a single model or purely quantitative forecasts may compromise forecasting robustness and accuracy. Meanwhile, purely quantitative forecasts are limited by the historical trend and pattern that causes failure if left without human judgmental input. This critical limitation underscores the need to shift toward more resilient, adaptive forecast-combination and hybrid approaches that are explicitly designed to perform under the turbulent conditions of post-crisis recovery.

Enhancing recovery forecasts with combination and hybrid approaches

Given that no single forecasting model is universally optimal or capable of addressing all requirements (Zhang, 2003), forecast combination methods provide a robust strategy to integrate accurate predictions while mitigating model-selection risks (Bates & Granger, 1969). Approaches to forecast combination can be broadly classified into three categories: combinations of multiple quantitative models, combinations of multiple judgmental forecasts, and hybrid approaches integrating both quantitative and judgmental components (Flores & White, 1988). Each category addresses unique forecasting challenges and contributes to enhanced accuracy and robustness in different contexts.

A primary strategy involves the weighting scheme used for combination. Over time, forecast combination techniques have evolved from simple averages to more sophisticated approaches featuring time-varying weights, nonlinear integration, and cross-model learning (Wang et al.,

2023). Common combination techniques in the tourism domain include average-based combinations (Coshall & Charlesworth, 2011; Wong et al., 2007), forecast-error-based weighting (Coshall, 2009), and regression-based integration methods (Cang, 2014; Shen et al., 2011). Simple averaging provides a robust baseline, but its static nature limits its ability to react to changing market dynamics. In contrast, time-varying weighting schemes directly enhance adaptability under volatility. By assigning more weight to models that have performed better in the recent past (e.g., using forecast-error-based weighting), these methods can dynamically adjust to shifting data patterns. This mechanism also accelerates responsiveness to emerging recovery signals, as models that first capture the nascent rebound are automatically given greater influence in the combined forecast. However, these approaches have predominantly focused on mechanical combination of multiple quantitative forecasts.

The forecast combination approach has gained considerable prominence in tourism forecasting due to repeated empirical demonstrations that no single model consistently achieves superior performance. Forecast combination involves integrating two or more model forecasts to achieve greater accuracy or stability than any model alone (Flores & White, 1988). For example, early work by Fritz et al. (1984) highlighted how combining competing forecasts reduced errors and improved overall predictive accuracy. Subsequent studies support this view, showing that combined forecasts not only enhance accuracy but also reduce forecast variance (Hibon & Evgeniou, 2005). A comprehensive review by Song et al. (2019) reported a marked increase in combined- or hybrid-model use over the past two decades, with the vast majority of such studies finding combined forecasts to be more accurate than individual models. Their observations further indicate that combined forecasts generally outperformed even the single best model's forecast in accuracy. The underlying rationale for this consistent superiority is twofold: Combined methods can capture complementary insights from different models and offer robustness against model misspecification and structural breaks, which are common during periods of instability (Timmermann, 2006).

Furthermore, forecast combination is especially valuable during crises, when data patterns can change abruptly, and traditional models often lose reliability (Song et al., 2009; Zhang et al., 2021). For instance, during the 2007–2008 global financial crisis, Song et al. (2010) combined an autoregressive distributed lag model with scenario analysis to evaluate its impact on Hong Kong's tourism demand. Page et al. (2012) similarly used time-varying parameter models and scenario analysis to explore the effects of both economic shocks and the swine flu pandemic on inbound tourism in the United Kingdom.

COVID-19 has accelerated innovation in tourism forecasting, prompting researchers to use forecast combination methods to address rapidly changing market conditions and unprecedented disruptions in data patterns. To accelerate responsiveness to emerging recovery signals, recent studies have focused on hybrid models that incorporate high-frequency, non-traditional data sources, often pandemic-scenario indices or other leading indicators. For example, Liu et al. (2021) proposed a scenario-based judgmental forecasting method using a COVID-19 Risk Exposure Index and expert judgment to develop *ex-ante* scenarios. Qiu et al. (2021) introduced a two-stage, three-scenario forecasting framework employing stacked models and recovery scenarios to forecast tourism demand across 20 destinations. Wu et al.

(2023) proposed an integrated seasonal autoregressive integrated moving average-mixed data-sampling approach and incorporated heterogeneous data sources, including monthly economic indices, daily search queries, and weekly online news, to improve forecasting accuracy during the pandemic. More recently, Li et al. (2025) introduced a novel combination strategy employing a multi-objective swarm intelligence optimization algorithm, further improving accuracy under pandemic conditions. Such approaches directly bolster recovery forecasting by grounding predictions in real-time indicators of policy changes and behavioral shifts rather than relying solely on lagged official tourism statistics.

While dynamic weighting improves adaptability, it cannot account for events or information entirely absent from the historical data. Here, hybrid methods integrating expert judgment become indispensable. Hybrid approaches are particularly valuable in volatile contexts where traditional statistical models may struggle to adapt to structural breaks or unprecedented events, and expert knowledge can provide essential contextual calibration (Liu et al., 2022). These methods include judgmental adjustments to model-generated forecasts, algorithmic integration of expert opinions with statistical outputs, and expert-led selection among competing forecasts (Petropoulos & Siemsen, 2023; Song et al., 2013). The Delphi method, for instance, systematically consolidates expert opinion to generate forecasts, providing crucial robustness to structural breaks. When historical data on post-pandemic travel behavior are nonexistent, expert knowledge on travel restrictions, consumer sentiment, and policy changes can guide forecasts away from model-induced absurdities. For example, Zhang et al. (2021) developed a scenario-based Delphi framework that combined econometric forecasting with expert-driven scenarios to forecast tourism recovery in Hong Kong. Kourentzes et al. (2021) demonstrated that univariate models combined with cross-sectional hierarchical forecasting techniques, when paired with expert adjustments, outperformed multivariate models in generating *ex-ante* tourist arrival forecasts. This indicates how expert input can correct model misspecification in a crisis. The integration of judgment ensures that forecasts remain plausible and contextually relevant. Collectively, these studies underscore the growing relevance of hybrid forecasting approaches as adaptive tools for managing uncertainty in tourism demand forecasting during crisis and post-crisis periods.

The critical role of model selection in volatile environments

The effectiveness of any combination strategy hinges on the quality of the models in its pool. Hence, model selection is fundamental to constructing effective forecast combinations. Traditional selection algorithms typically rely on historical data and are guided by various criteria, including information criteria, assessments of past forecasting performance, and diagnostics based on time series characteristics (Petropoulos & Siemsen, 2023). Statistical approaches to model selection include information criteria (e.g., Akaike Information Criterion, corrected Akaike Information Criterion, Bayesian Information Criterion), validation and cross-validation, descriptive statistics, discriminant analysis, and rules grounded in variance or domain expertise. These methods aim to balance model complexity with predictive fit (Gardner, 2006) and are particularly well-suited to short or stable time series where holdout validation may be infeasible (Petropoulos & Siemsen, 2023). Validation and cross-validation are also prevalent for model selection based on historical performance, especially in machine learning contexts.

In stable periods, model selection can rely on retrospective statistical criteria. However, in a post-crisis context where the past is a poor predictor of the future, these methods can be misleading. This highlights the growing importance of judgmental model selection. Research shows that combining a carefully selected subset of forecasts often yields better accuracy than aggregating all available models (Wang et al., 2023). Judgmental selection can help exclude poorly performing models, especially when statistical criteria are inconclusive. For example, Petropoulos et al. (2018) found that judgmental model selection identified the best- and worst-performing models more effectively than purely algorithmic approaches. Furthermore, combining statistical and judgmental selections yielded superior performance relative to either method alone (Petropoulos et al., 2018).

While information criteria and cross-validation are widely utilized for forecasting model selection, primarily by comparing forecasts with observed actual values (Petropoulos & Siemsen, 2023), relatively few studies have focused on evaluating *ex-ante* forecast performance prior to data availability. In such contexts, judgmental model selection outperforms purely statistical procedures, particularly in its ability to discern and select representative models before outcomes are realized. Petropoulos and Siemsen (2023) introduced a new selection criterion that integrates past model fit with human judgment to evaluate the representativeness of out-of-sample forecasts, surpassing sole reliance on traditional comparisons between actual and predicted values. Their method allows for asynchronous comparison and *ex-ante* evaluation, demonstrating that judgmental selection based on representativeness can improve forecasting performance.

Synthesis and research gaps

The literature overwhelmingly demonstrates that forecast combination is indispensable in navigating post-crisis uncertainty. Despite this, the tourism forecasting literature continues to exhibit notable limitations in model-selection practices. Most studies either adopt retrospective statistical evaluation, such as information criteria and forecast error analysis, or rely on *ad hoc* expert judgment, lacking a systematic mechanism to combine the strengths of both. This dichotomy restricts adaptability, particularly in volatile post-crisis contexts where historical data may no longer be representative and expert insights are critical. Moreover, although combination methods are widely used, few studies have established a structured, multi-stage framework that coherently integrates model selection, forecast combination, and judgmental adjustment. These gaps highlight the need for more integrated and adaptive forecasting frameworks capable of effectively responding to post-crisis volatility, especially in the context of China's outbound tourism recovery. This recovery involves destination-specific dynamics, policy-driven fluctuations, and behavioral shifts difficult to capture using standard models alone.

The motivations for combining forecasts stem from the recognized limitations of single-type models in volatile contexts. Combinations of quantitative models alone often fail to capture structural breaks and the dynamic nature of crises, while purely judgmental methods lack statistical robustness and replicability (Bunn & Wright, 1991). Effectively integrating quantitative and qualitative approaches can bridge this gap by leveraging the strengths of both, producing forecasts that are statistically sound and better adapted to the uncertainties of

post-crisis recovery (see Table 1).

Table 1 Categories of Forecast Combination

Category	Key Principle	Base Forecasts	Combination Technique
Quantitative	Combining the outputs of multiple statistical or machine learning models to improve accuracy and reduce error variance.	Simple extrapolative methods (e.g. Naive methods, exponential smoothing methods), time-series methods, econometric methods, artificial intelligence methods.	Systematic: Simple average, forecast-error-based weighting, regression-based integration Intuitive: Ad-hoc selection or subjective weighting of model outputs based on analyst judgment
Judgmental	Aggregating forecasts generated independently by multiple human experts to leverage collective wisdom and diverse perspectives.	Delphi method, scenario writing, nominal group technique, other judgmental forecasting methods.	Systematic: Simple average, consensus rules (e.g., median forecast, trimmed mean). Intuitive: Informal discussion, negotiation, or dominant-expert influence
Hybrid	Integrating quantitative, model-based forecasts with qualitative expert knowledge or judgmental adjustments.	Quantitative model outputs combined with expert judgmental methods (e.g., Delphi, scenario analysis).	Systematic (mechanical integration): simple average of judgmental and quantitative forecasts, correction for bias, and bootstrapping Intuitive (voluntary integration): Judgmental adjustments to quantitative model outputs

Source: Adapted from Flores and White (1988).

3. The Triple-layer Forecast Combination Framework

This study proposes the Triple-layer Forecast Combination Framework, a systematic approach designed to generate accurate and adaptive forecasts, as illustrated in Figure 1. This framework operationalizes the hybrid forecasting concept by sequentially integrating quantitative and qualitative methods across three distinct but interconnected stages. A broad pool of quantitative baseline forecasts (Stage 1) is refined through an expert-driven model-selection and -combination process (Stage 2). A final layer of judgmental adjustments based on real-time contextual information (Stage 3) ensures the final output reflects a robust synthesis of statistical evidence and expert insight.

Stage 1: Quantitative modeling and forecasting

In the initial stage, quantitative techniques such as time series analysis, econometric models, and AI tools were used to generate baseline forecasts. Specifically, 10 individual models were utilized in R from August 2023 to July 2024 to predict Chinese outbound visitor arrivals to 20 destinations: Canada, Chile, Mexico, Chinese Taipei, Hong Kong SAR, Japan, Korea (ROK),

Macao SAR, Maldives, Cambodia, Indonesia, Singapore, New Zealand, the United States, Thailand, Turkey, Australia, Hawaii, Austria, and the Czech Republic.

The forecasting models in this study were intentionally selected to produce a comprehensive set of forecasts incorporating a wide range of assumptions and structural forms. In particular, the decision to employ 10 distinct models in Stage 1 was guided by the fundamental principle of assembling a diverse and representative portfolio, which is essential for achieving robustness in the highly volatile and uncertain conditions typical of post-crisis environments. Rather than seeking to identify a single superior model, the aim was to generate a broad array of plausible forecast trajectories. These trajectories are generated from multiple model families, each uniquely equipped to capture different features and patterns inherent in the data. The performance of each model was rigorously evaluated using the Mean Absolute Scaled Error, which helps identify predictions that may require further refinement.

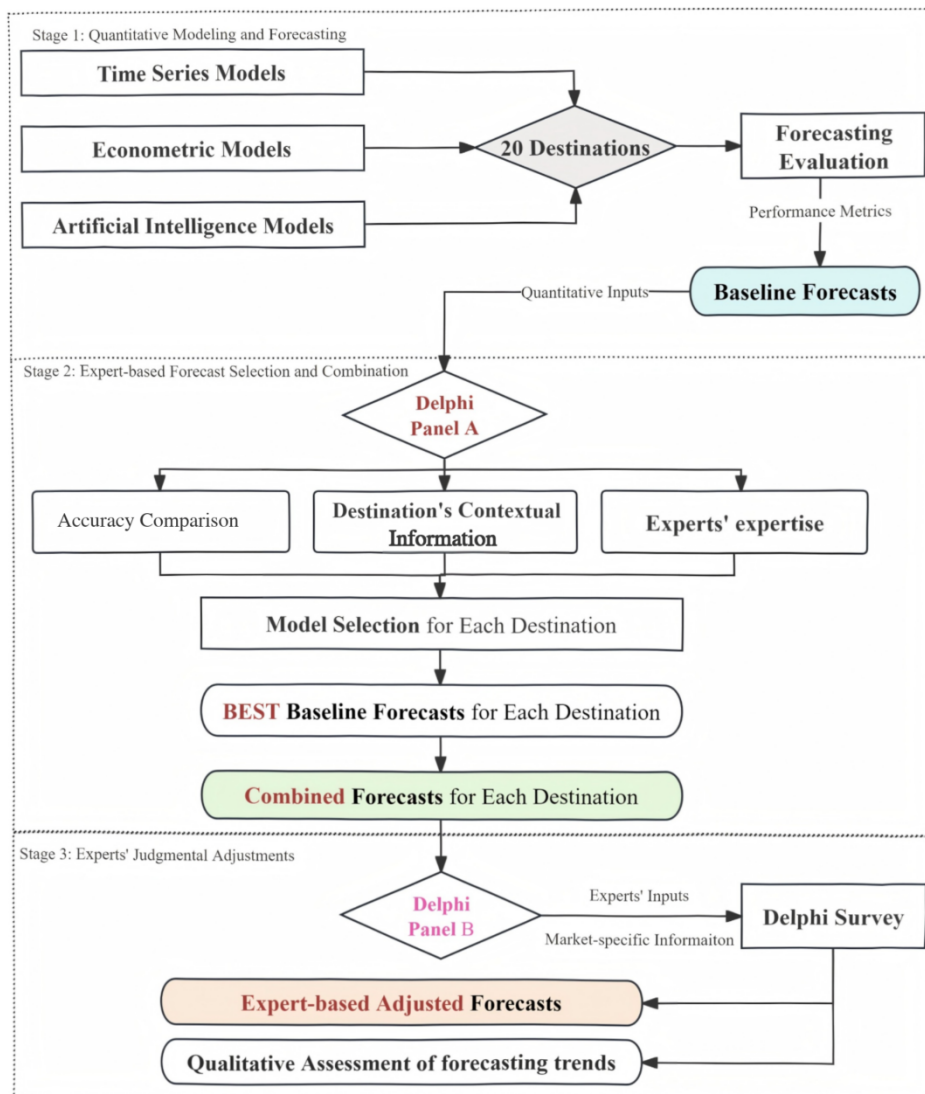


Figure 1 The Triple-layer Forecast Combination Framework

This study aggregated monthly visitor arrival data from January 2000 to February 2023 at quarterly observations for 20 destinations. Following demand theory, independent variables

included GDP, exchange rates, and consumer price indices; 2010 served as the base year (values set at 100). Data was sourced from the International Monetary Fund and the Census and Economic Information Centre. Econometric models assessed economic factors affecting tourism demand by incorporating Chinese visitor income (GDP) and relative prices adjusted by the China–destination exchange rates. Dummy variables captured seasonality and specific events like the COVID-19 pandemic. Before model estimation, visitor arrivals and explanatory variables were preprocessed. To address missing values, Kalman Smoothing was applied to structural time series models (Harvey, 1990) using the “na_kalman” function from the “imputeTS” package in R (Moritz & Bartz-Beielstein, 2017). This ensured time-series continuity while preserving temporal correlations. Finally, a log-plus-one transformation was used to handle zero values and stabilize the variance.

Time series methods

To analyze historical tourism demand patterns, six distinct time series models were employed: seasonal naïve (Model_{SNAIVE}); exponential smoothing state-space (Model_{ETS}); seasonal autoregressive integrated moving average (Model_{SARIMA}); seasonal autoregressive integrated moving average with exogenous inputs (Model_{SARIMAX}); seasonal and trend decomposition using Loess (Model_{STL}); and trigonometric seasonality, Box-Cox transformation, auto-regression moving average errors, trend, and seasonal components (Model_{BCT}). Model_{SNAIVE} forecasts seasonal data by replicating values from the corresponding period of the previous year. Model_{ETS} applies exponential smoothing with automatic selection of trend and seasonality components (Hyndman & Khandakar, 2008). These two models consistently serve as benchmarks in tourism demand forecasting studies.

Model_{SARIMA} and Model_{SARIMAX} were built using the Box-Jenkins approach, capturing both trends and seasonal patterns in tourism demand. The “auto.arima” algorithm automated model selection and estimation. To account for external factors influencing tourism demand, Model_{SARIMAX} included exogenous variables, such as GDP and exchange rates, reflecting economic influences on visitor arrivals. Model_{STL} decomposed visitor arrival data into trend, seasonal, and remainder components (Cleveland et al., 1990), then applied either Model_{ETS} or the auto-regression integrated moving average model to forecast the remainder. These components were then resynthesized to produce the final forecasts. Model_{BCT} was implemented to address complex seasonal patterns, particularly those with multiple and non-integer seasonality structures, owing to the utility of the Box-Cox transformation and trigonometric terms in facilitating the flexible modeling of cyclical behavior (De Livera et al., 2011).

Econometric methods

Two econometric models produced baseline forecasts: the autoregressive distribution lagged model (Model_{ARDL}) and static regression with a time-varying parameter model (Model_{SRTVP}). Model_{ARDL} was well-suited to capturing both short-term dynamics and long-term equilibrium relationships with explanatory variables, including China’s GDP, relative prices, and event-specific dummies. Lag structures were selected automatically using the “auto_ardl” function, which also performs a bounds test for cointegration to facilitate the estimation of long-run elasticities (Pesaran et al., 2001). The explanatory variables were projected onto the forecast

horizon using univariate models like Model_{ETS}. Given the dynamic nature of these variables, Model_{SRTVP} was also employed using the “tvLM” function from the “tvReg” package. Forecasts for visitor arrivals were generated through a custom routine (forecast.tvlm), which offers the advantage of recursively updating parameters as new data become available.

AI methods

In this study, our application of “AI methods” is specifically focused on a well-established subclass: feedforward neural networks. Two feedforward neural networks were employed to generate baseline forecasts: the univariate multilayer perception (Model_{MP}) and the univariate extreme learning machine (Model_{MLP}). Model_{MP} uses lagged visitor arrivals values as input predictors to capture nonlinear temporal dependencies. To ensure replicability, a fixed random seed is applied. The training of Model_{MP} was conducted through backpropagation, with validation procedures in place to minimize the risk of overfitting. Model_{MLP} shares a similar network structure but differs in its training approach. Following Huang et al. (2006), this model randomly assigns the majority of weights in the hidden layer, focusing learning exclusively on the output layer coefficients (Huang et al., 2006). This simplification substantially accelerates the training process while maintaining a high level of forecasting accuracy.

Stage 2: Expert-based forecast selection and combination

In Stage 2, a modified Delphi method was employed to refine the baseline forecasts generated in Stage 1. Given the significant variability among the 20 destinations, a one-size-fits-all model is inappropriate. Therefore, it was crucial to select and combine models that align most effectively with each destination’s contextual specifications. The experts assessed both quantitative metrics and destination-specific factors to identify the most suitable forecasts, drawing on their knowledge of origin markets, policy environments, and tourism market trends. To ensure transparency and rigor, they were also required to justify their choice.

The judgmental selection was conducted by a dedicated panel of five senior experts selected based on rigorous criteria: (i) advanced academic credentials (all hold PhDs in relevant fields); (ii) extensive professional experience in tourism analytics and consultancy; and (iii) a proven research track record, including publications in leading journals and major industry reports on tourism demand. This specific blend of qualifications was deemed essential for the demanding nature of Stage 2, which required the technical acumen to interpret quantitative metrics like the Mean Absolute Scaled Error alongside the deep domain knowledge needed to assess the contextual plausibility of each forecast for the Chinese outbound market.

Expert forecasters selected the models using a structured, multi-criteria framework designed to integrate statistical performance with qualitative insight (Figure 2). The selection process unfolded in three sequential stages. First, an initial quantitative assessment ranked candidate models based on forecasting performance metrics, primarily the Mean Absolute Scaled Error, a robust metric for comparing point forecast accuracy (Franses, 2016). This error measurement is particularly advantageous as it remains reliable even with zero or near-zero actual values, substantial data scale variation across time series, and pronounced seasonal fluctuations (Hyndman & Koehler, 2006). Second, experts evaluated these statistically-ranked

models for contextual plausibility by analyzing destination-specific information, such as historical data patterns, geographic factors, and policy contexts. Experts received information on historical time series, seasonal structures, and projected recovery paths generated by each of the 10 baseline models from Stage 1 (see Figure 3). In the final synthesis stage, experts applied their domain knowledge of market dynamics to integrate the quantitative and contextual evidence, culminating in a justified model selection for each destination. From this process, a subset of “plausible models” was identified from the initial pool of 10 candidates; a model was included if selected by at least one expert. The forecast was then calculated as the simple arithmetic average of all unique models in the plausible subset, where each model contributed equally. This approach advances the traditional Delphi technique by systematically embedding expert-driven model selection within a structured forecast combination procedure, ensuring the resulting forecasts are data-driven, statistically robust, and contextually validated.

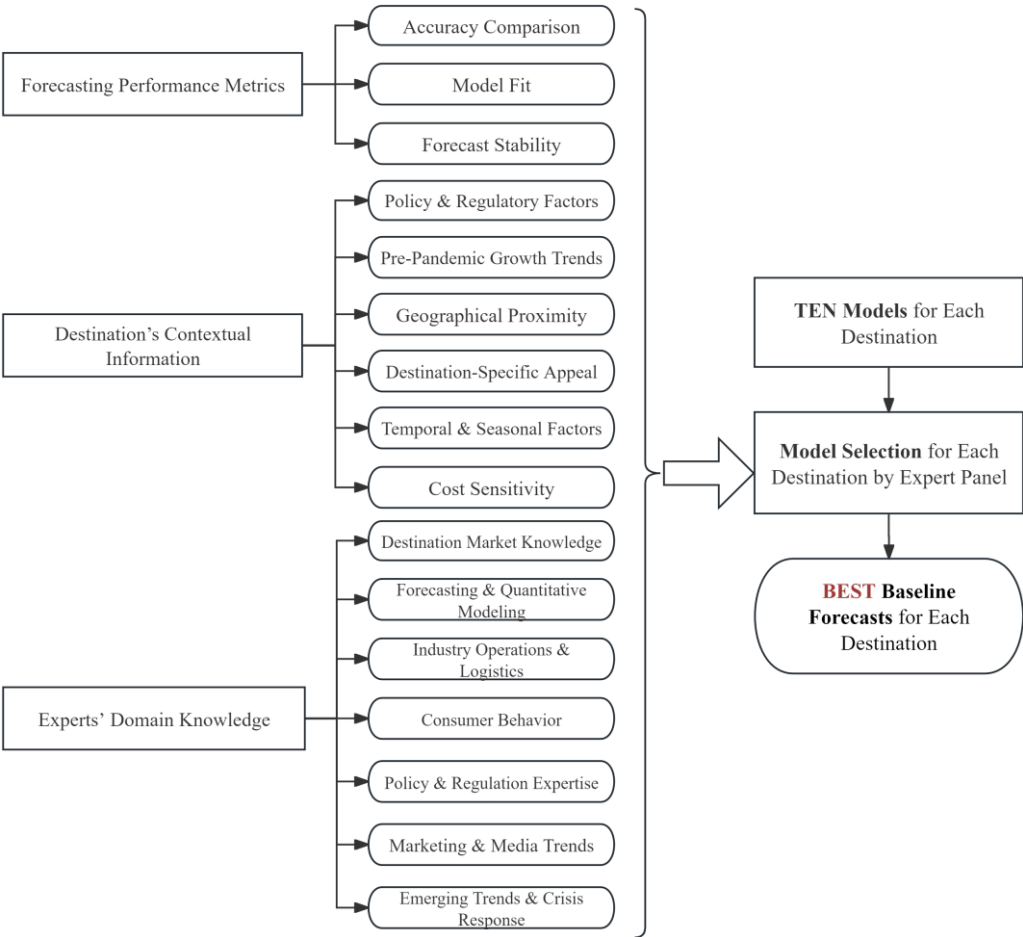


Figure 2 Framework of the Judgmental Selection Process

Figure 4 and Table 2 collectively illustrate the selection frequencies and geographic patterns of optimal forecasting models across destinations. Model_{SRTVP} was most frequently selected, chosen for 12 destinations (60%), underscoring experts’ preference for its flexibility in capturing structural changes over time. Model_{MLP} and Model_{SARIMAX} followed closely, being optimal for 10 (50%) and 9 (45%) destinations, respectively, likely due to their robustness and adaptability to destination-specific contexts. Model_{SARIMA}, although a more traditional

approach, was selected for 8 destinations (40%), indicating its continued relevance, particularly in stable, seasonal markets. Experts thus demonstrated a clear preference for models capable of capturing nonlinearity (Model_{MLP}), external variables (Model_{SARIMAX}), or time-varying dynamics (Model_{SRTVP}), highlighting the importance of flexible and adaptive tools in post-pandemic tourism demand forecasting. Notably, Model_{ARDL} and Model_{BCT} were not chosen for any destination, as their outputs were deemed inconsistent with observed trends. Geographic patterns in model selection were also apparent: Model_{MLP} was favored for North American destinations, particularly where complex, multidimensional factors contributed to heightened uncertainty, while Model_{SARIMA} and Model_{SARIMAX} were preferred for Australian destinations with stable, seasonal demand. For East and Southeast Asian destinations, Model_{SRTVP} and Model_{SARIMAX} were commonly selected, reflecting expectations of rapid recovery and consistent seasonal patterns.

Experts were asked to provide justifications for their model selections. Figure 5 summarizes the relative importance of key factors identified across the panel. This hierarchy of factors reflects experts' emphasis on integrating structural market characteristics, external risks, and demand-side behavioral drivers into model selection. Among the most frequently cited was pre-pandemic growth, with historical trends serving as a reference point for anticipating recovery trajectories. Geographical proximity to China was also emphasized, as destinations closer to the origin market were generally expected to recover more rapidly than long-haul counterparts. Seasonality, particularly the strength and regularity of seasonal patterns, was another common consideration. Experts also highlighted the influence of external factors, such as political tensions and diplomatic relations, which were seen to significantly affect demand uncertainty. Administrative barriers, including travel restrictions and visa policies, were consistently recognized as major determinants of recovery speed. Additional emphasis was placed on destination-specific appeal, including cultural attractiveness, quality of tourism infrastructure, and perceived safety, which are particularly salient for Southeast Asian destinations. Lastly, travel costs and price sensitivity were considered especially important for long-haul destinations. In essence, these varied factors, encompassing both historical data and real-world dynamics, underpinned the experts' nuanced choices in model selection.

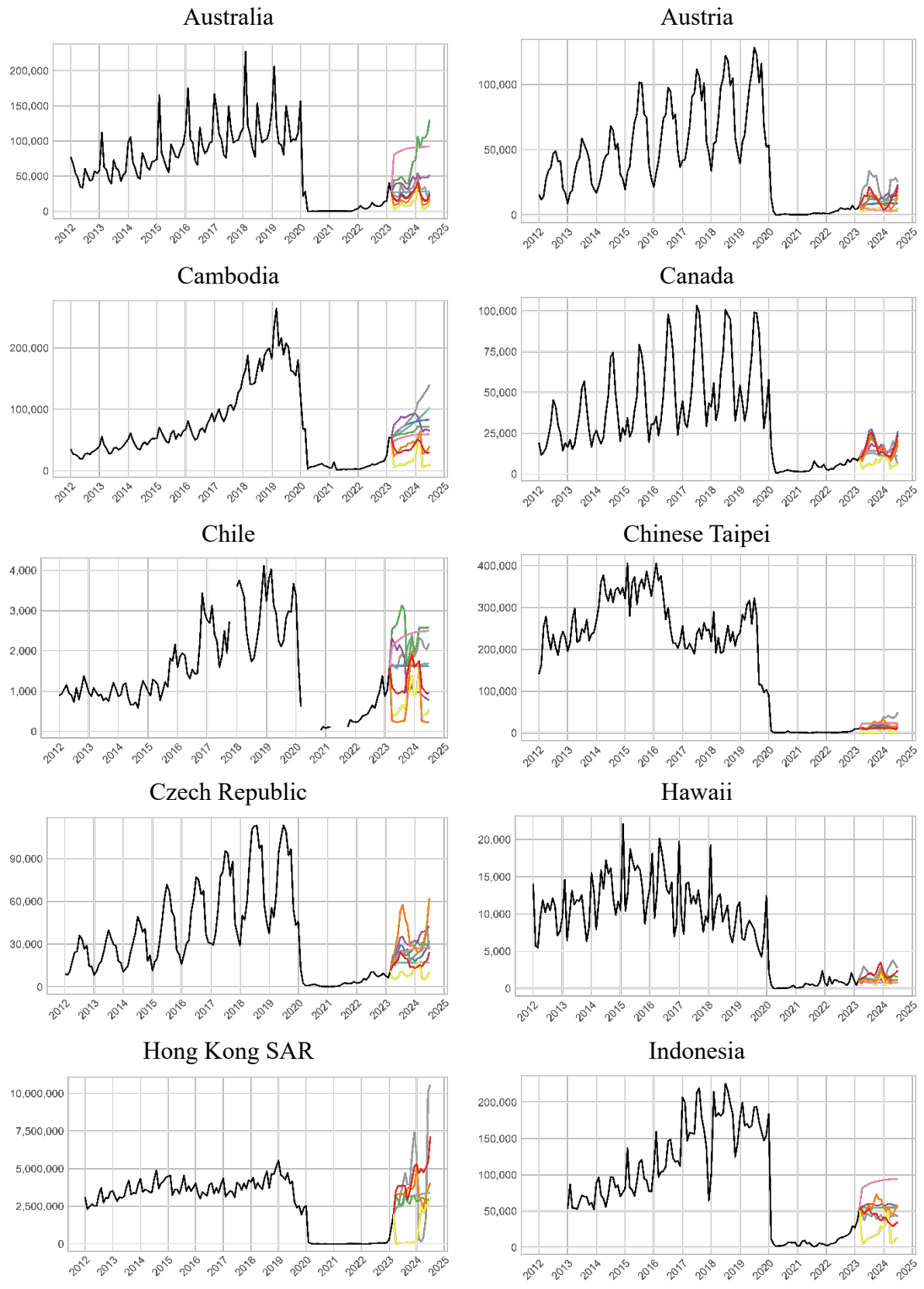


Figure 3 Supporting Information Provided to Experts

Note: Lines represent baseline forecasts for Chinese outbound tourism to a destination: Historical data (black), Model_{SNAIVE} (yellow), Model_{ETS} (blue), Model_{SARIMA} (green), Model_{SARIMAX} (purple), Model_{STL} (orange), Model_{BCT} (red), Model_{ARDL} (brown), Model_{SRTVP} (pink), Model_{MP} (gray), Model_{MLP} (cyan).

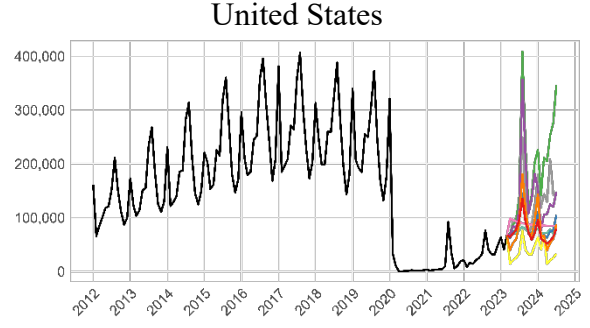
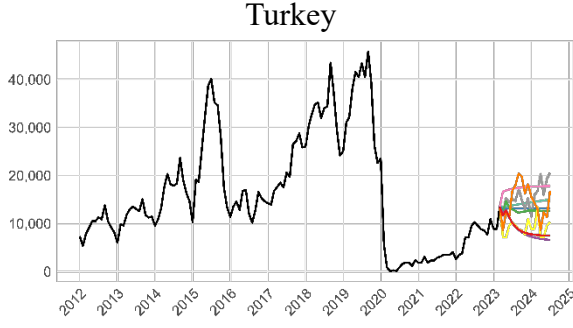
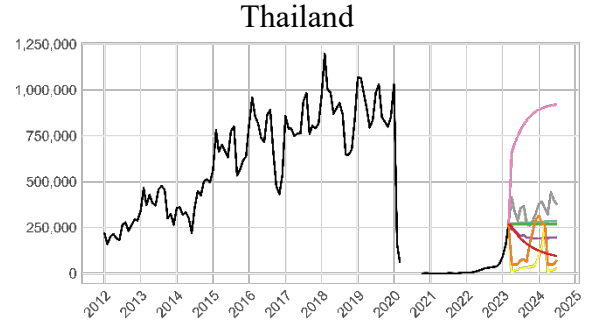
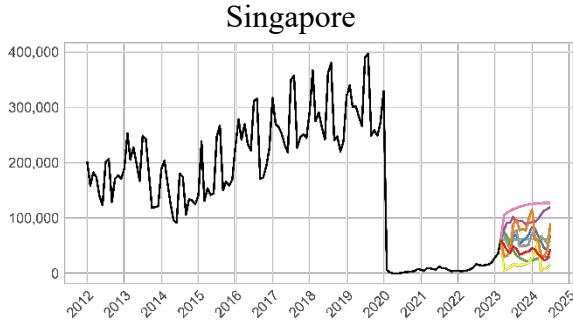
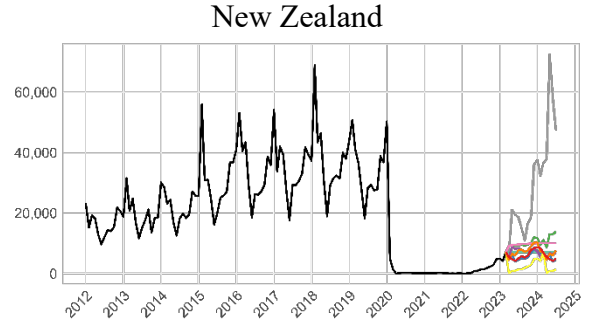
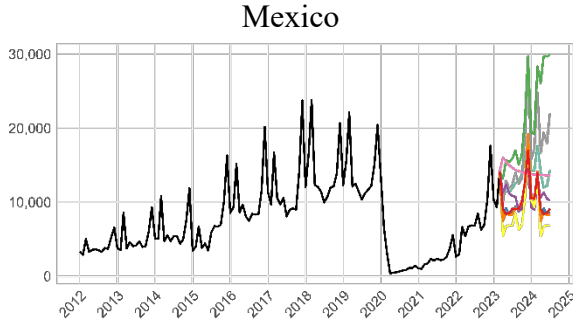
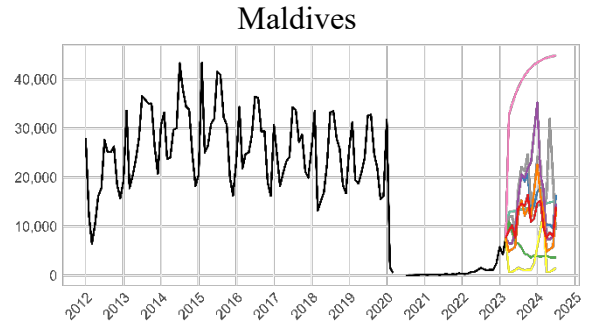
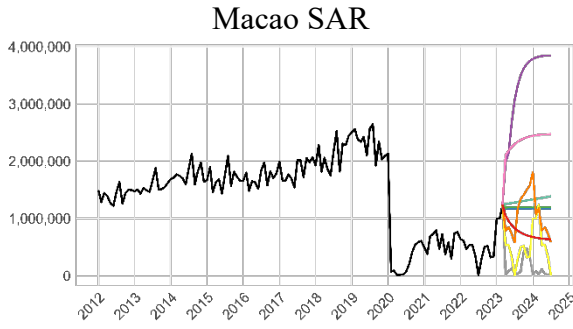
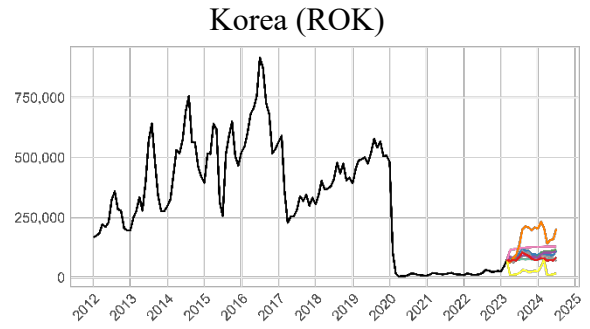
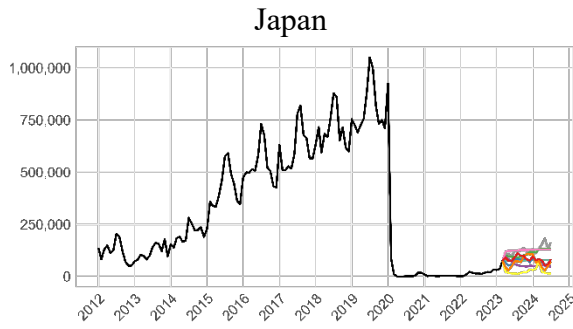


Figure 3 Supporting Information Provided to Experts (Cont.)

Table 2 Final Model Selection by the Expert Panel

Destination	Model Selection
Canada	Model _{ETS} +Model _{SARIMAX}
Chile	Model _{SARIMA} +Model _{MLP}
Mexico	Model _{SARIMAX} +Model _{SRTVP} +Model _{MLP}
Chinese Taipei	Model _{SARIMAX} +Model _{STL} +Model _{MLP}
Hong Kong SAR	Model _{SNAIVE} +Model _{SARIMA}
Japan	Model _{SRTVP} +Model _{MLP}
Korea (ROK)	Model _{STL} +Model _{SRTVP}
Macao SAR	Model _{SNAIVE} +Model _{SARIMAX} +Model _{SRTVP}
Maldives	Model _{SARIMAX} +Model _{MLP}
Cambodia	Model _{SARIMA} +Model _{SRTVP}
Indonesia	Model _{STL} +Model _{SRTVP}
Singapore	Model _{SARIMAX} +Model _{SRTVP}
New Zealand	Model _{SARIMA} +Model _{SRTVP}
United States	Model _{ETS} +Model _{MLP}
Thailand	Model _{SRTVP} +Model _{MLP}
Turkey	Model _{SRTVP} +Model _{MLP}
Australia	Model _{SARIMA} +Model _{SARIMAX} +Model _{SRTVP}
Hawaii	Model _{SARIMA} +Model _{MLP}
Austria	Model _{SARIMA} +Model _{SARIMAX}
Czech Republic	Model _{SARIMA} +Model _{SARIMAX} +Model _{SRTVP} +Model _{MLP}

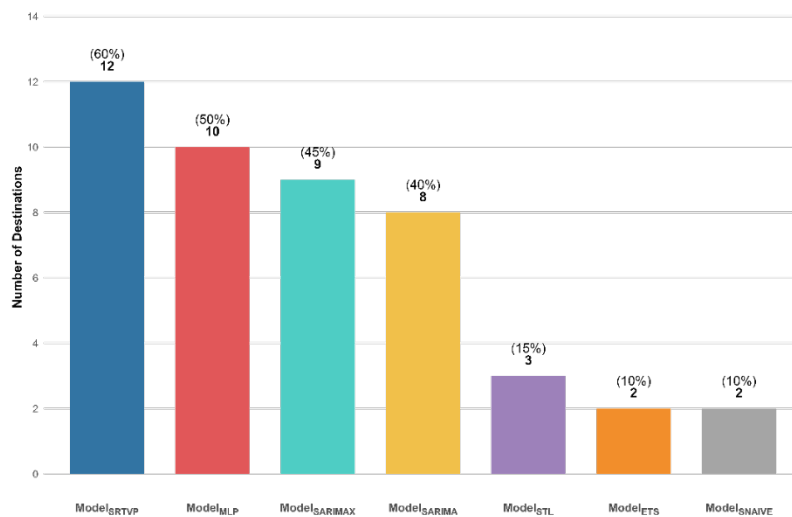


Figure 4 Frequency of Selected Models by the Delphi Panel

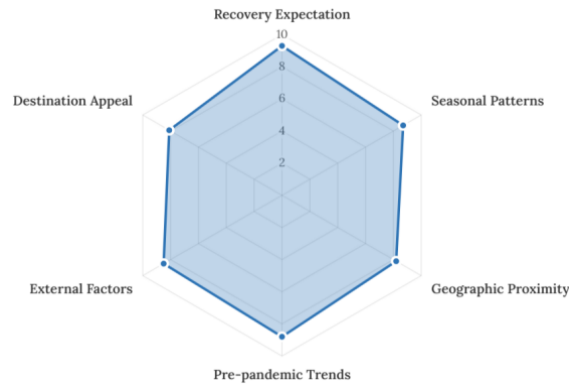


Figure 5 Relative Importance of Factors in Expert Model Selection

Once individual experts selected the most appropriate forecasting models, the next step involved combining the selected forecasts. A key consideration in this process was the choice of a weighting scheme. Among various options, the simple average method was adopted due to its operational simplicity, widespread use as a benchmark, and well-documented effectiveness in the literature (Clemen, 1989; Song et al., 2009). Previous research has consistently shown that simple averaging performs comparably to, or better than, more complex weighting strategies (Armstrong, 2001; Shen et al., 2011; Stock & Watson, 2004; Wang et al., 2023). Its advantages include eliminating the need to estimate optimal weights, decreasing variance and bias by averaging forecasting errors, and mitigating risks associated with weight misspecification (Palm & Zellner, 1992). As the selected forecasts demonstrated similar levels of accuracy, there was no empirical justification for assigning differential weights. Equal weighting was thus applied to ensure an unbiased, stable composite forecast that leverages the strengths of all models while diluting individual weaknesses.

Stage 3: Experts' judgmental adjustments

The third stage involved applying additional judgmental adjustments to further enhance accuracy based on a comprehensive Delphi survey conducted by tourism and hospitality professionals. By integrating expert judgment with evolving market conditions and emerging trends, the forecasting process is enhanced to achieve an optimal balance between quantitative and qualitative evidence. This is accomplished through the synthesis of statistical outputs with real-time market dynamics and expert insights. Thus, the judgmental selection processes in Stage 2 and final judgmental adjustments in Stage 3 together constitute the second layer of the forecast combination framework.

Experts calibrated model-generated forecasts using concurrent indicators including travel mobility trends, flight capacity data, and evolving visa policies. A key component of this stage was a Delphi survey focused on Chinese outbound departures to the 20 selected destinations. Invitations were distributed via email and WeChat to 400 professionals representing academia, government, travel agencies, accommodation providers, and attraction operators. A total of 129 valid responses were collected and analyzed. The composition of the expert panel is summarized in Table 3. Panelists were asked to provide adjustments for each destination.

To facilitate informed judgments, participants received an Excel template featuring tabular

and graphical summaries of current forecasts along with updated information on visa policies and destination-specific tourism statistics. The experts were encouraged to justify their estimates, elaborating on the contextual factors informing their assessments. Although the experts were not asked to directly use search engine indices, which are recognized as effective short-run predictors of tourism demand (Qiu et al., 2021), indicators such as flight capacity, mobility trends, and visa policy changes capture comparable dimensions of real-time travel intentions and thus serve a similar role in refining short-term forecasts. In this sense, the expert adjustments approximate the demand signals that search data provide, offering a different but parallel lens on evolving recovery dynamics. This process revealed a significant surge in visitor arrivals following China’s 2023 travel restriction lift, underscoring the necessity of real-time data integration to improve forecast validity.

Table 3 Expert Panel Composition

Sector	Number	%
Academic institutions	28	21.7%
Government	15	11.6%
Travel agents	48	37.2%
Other tourism sectors	38	29.5%
Grand total	129	100.0%

Findings from the Delphi survey supported more optimistic adjustments for short-haul destinations—particularly the two SARs (Macao and Hong Kong) and Korea (ROK)—driven by cultural proximity, robust transport connectivity, and relaxed visa regimes. While Thailand and the Maldives were also expected to recover strongly due to their appeal as beach destinations, some experts raised concerns over safety issues that could moderate the pace of recovery. In contrast, long-haul destinations were projected for more gradual, constrained recovery trajectories, reflecting cost considerations, travel hesitancy, and policy uncertainties. In sum, this final stage integrated diverse real-time data and expert consensus to produce context-sensitive forecasts that are both empirically grounded and adaptable to post-pandemic volatility.

4. Results and Discussion

Forecasting performance against benchmark models

To assess forecasting performance for the competition period (August 2023–July 2024), the proposed specification ($\text{Model}_{\text{FINAL}}$) was compared with four benchmark models: $\text{Model}_{\text{NAIVE}}$, $\text{Model}_{\text{SNAIVE}}$, $\text{Model}_{\text{SARIMA}}$, and $\text{Model}_{\text{ETS}}$. Mean Absolute Scaled Error was used as the primary accuracy measure, and its mean and standard deviation across destinations were examined to capture average performance and stability (Table 4). $\text{Model}_{\text{FINAL}}$ achieved the lowest mean error (1.248) and the smallest standard deviation (0.795), indicating the highest overall accuracy and most consistent performance. All benchmark models produced larger and more variable errors. The seasonal naïve model ($\text{Model}_{\text{SNAIVE}}$) performed worst (3.178), while the seasonal autoregressive integrated moving average model ($\text{Model}_{\text{SARIMA}}$) and the naïve model ($\text{Model}_{\text{NAIVE}}$) showed similar, intermediate accuracy. The exponential smoothing model ($\text{Model}_{\text{ETS}}$) also underperformed relative to $\text{Model}_{\text{FINAL}}$, with higher mean error and

greater dispersion.

Table 4 Descriptive Statistics of Mean Absolute Scaled Errors for Forecasting Models

Model	Mean	Standard Deviation	Minimum	Maximum
Model _{FINAL}	1.248	0.795	0.182	3.173
Model _{SARIMA}	2.424	1.279	0.279	4.688
Model _{NAIVE}	2.494	1.281	0.353	4.894
Model _{ETS}	2.819	1.716	0.353	7.169
Model _{SNAIVE}	3.178	1.597	0.208	5.471

Figure 6 shows the distribution of mean absolute scaled error for each model across all destinations. Consistent with the descriptive statistics, Model_{FINAL} has the lowest median error (approximately 1.5) and the narrowest interquartile range, indicating superior central accuracy and lower forecast variance. The benchmark models display higher medians (Model_{SARIMA} \approx 2.5, Model_{NAIVE} \approx 2.5, Model_{ETS} \approx 3.0, Model_{SNAIVE} \approx 3.5) and greater dispersion. The wider upper tail of Model_{ETS} (upper whisker above 6.0) and the high variability of Model_{NAIVE} highlight their instability. Overall, these results confirm that the proposed framework delivers more robust and reliable forecasts across diverse destinations.

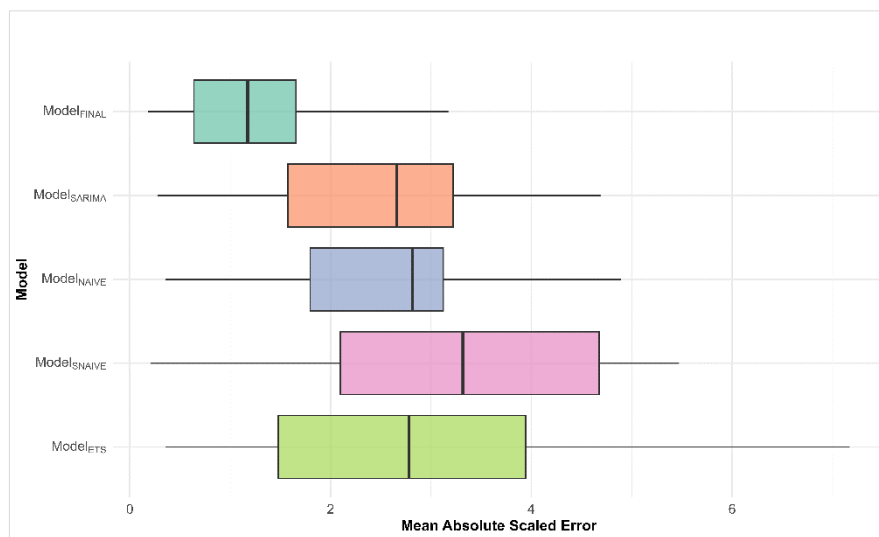


Figure 6 Forecasting Accuracy Comparison of Models

Forecasting performance by destination

Table 5 reports ex-ante forecast accuracy, measured by the Mean Absolute Scaled Error, for 20 destinations from August 2023 to July 2024. The proposed method outperformed all benchmark models in 16 destinations (80%), with particularly low errors for Chinese Taipei (0.182), Indonesia (0.306), Hawaii (0.454), Austria (0.543), and Canada (0.642). Benchmark models rarely led: Model_{SARIMA} ranked first in only two destinations (10%). Performance varied across destinations, reflecting differences in historical patterns and structural characteristics. In Chinese Taipei, the proposed method (0.182) outperformed the second-best model, Model_{SARIMA} (0.327), by 44.3%. For Indonesia, the improvement reached 78.2%, with

the proposed method achieving an error of 0.306 compared with 1.403 for the best-performing benchmark. These results suggest that destinations with relatively stable demand and strong regional linkages may exhibit more predictable recovery paths that the proposed framework can exploit effectively.

Table 5 Forecasting Accuracy by Destination and Forecast Horizons

Destination	Model_{NAIVE}	Model_{SNAIVE}	Model_{SARIMA}	Model_{ETS}	Model_{FINAL}
Australia	2.687	5.471	3.779	4.537	1.634
Austria	2.126	2.162	1.996	1.819	0.543
Cambodia	0.525	1.905	3.135	0.807	0.628
Canada	1.927	2.253	1.573	1.440	0.642
Chile	2.065	2.526	1.676	2.065	0.665
Chinese Taipei	0.353	0.458	0.327	0.353	0.182
Czech Republic	0.788	0.999	0.567	1.492	0.649
Hawaii	0.546	0.208	0.279	0.535	0.454
Hong Kong SAR	3.091	4.911	2.562	7.169	1.988
Indonesia	1.403	1.839	1.403	1.403	0.306
Japan	4.894	5.075	4.610	4.866	3.173
Korea (ROK)	3.214	3.479	3.242	3.160	1.280
Macao SAR	3.053	4.733	3.061	3.223	1.634
Maldives	2.989	4.827	1.829	4.225	2.511
Mexico	2.890	3.229	3.221	2.418	0.936
New Zealand	3.016	3.405	1.573	3.142	1.495
Singapore	3.788	4.207	2.991	3.848	2.170
Thailand	3.222	4.333	3.222	3.222	1.134
Turkey	4.557	4.653	4.688	4.556	1.722
United States	2.745	2.877	2.754	2.105	1.213

The model also performed consistently well in East Asia, ranking first for Hong Kong SAR (1.988), Macao SAR (1.634) and Korea (ROK) (1.280). Relative to Model_{SARIMA}, error reductions were 22.4% for Hong Kong SAR (2.562), 46.6% for Macao SAR (3.061) and 60.5% for Korea (ROK) (3.242). These findings point to the role of geographic proximity, dense pre-existing tourism flows and supportive policy environments in generating more regular recovery dynamics that can be captured by the proposed framework.

Japan was the most difficult destination to forecast, with the highest error (3.173). Nevertheless, the proposed model still outperformed all benchmarks, improving on Model_{SARIMA} (4.610) by 31.2%, Model_{NAIVE} (4.894) by 35.2%, Model_{SNAIVE} (5.075) by 37.5% and Model_{ETS} (4.866) by 34.8%. The elevated errors likely reflect volatility driven by factors not fully captured in historical data, including post-pandemic structural shifts, changing visa policies and heightened geopolitical tensions.

The proposed method did not dominate in only three cases. For Maldives (2.511), it ranked second after Model_{SARIMA} (1.829); for Cambodia (0.628), it ranked second after Model_{NAIVE} (0.525); and for Hawaii (0.454), it ranked third after Model_{SNAIVE} (0.208) and Model_{SARIMA}

(0.279), while still outperforming Model_{NAIVE} (0.546) and Model_{ETS} (0.535). Hawaii's third-place ranking, despite a relatively low error, reflects the exceptional performance of seasonal benchmarks model in a market with pronounced seasonality.

Forecasting Chinese Outbound Tourism Recovery (August 2023–July 2024)

Figure 7 presents the adjusted forecasting results for Chinese outbound tourism to 20 destinations from August 2023 to July 2024, incorporating insights from 129 Delphi experts. The results highlight substantial heterogeneity in recovery patterns across destinations. By July 2024, outbound travel from China was projected to reach approximately 80% of its July 2019 levels across all 20 destinations (Figure 8), reflecting a strong overall rebound alongside persistent regional disparities. Generally, destinations with favorable entry policies, strong marketing, and well-connected flight networks—particularly in East and Southeast Asia—were expected to lead the recovery. In contrast, long-haul markets in Europe and North America, while showing positive trends, face slower growth due to policy barriers, cost pressures, and shifting traveler preferences.

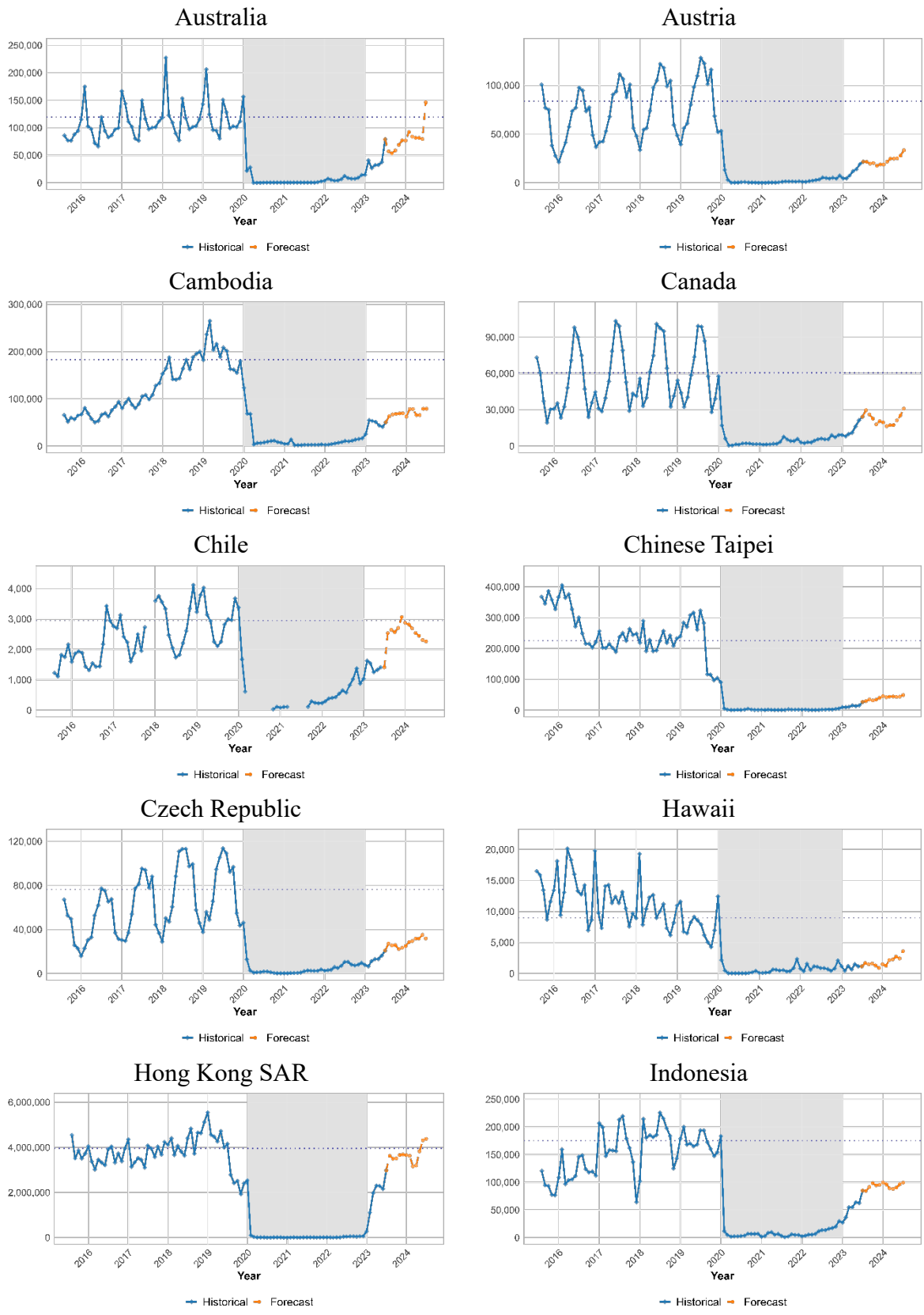


Figure 7 Final Forecasts for Chinese Outbound Tourism to 20 Destinations

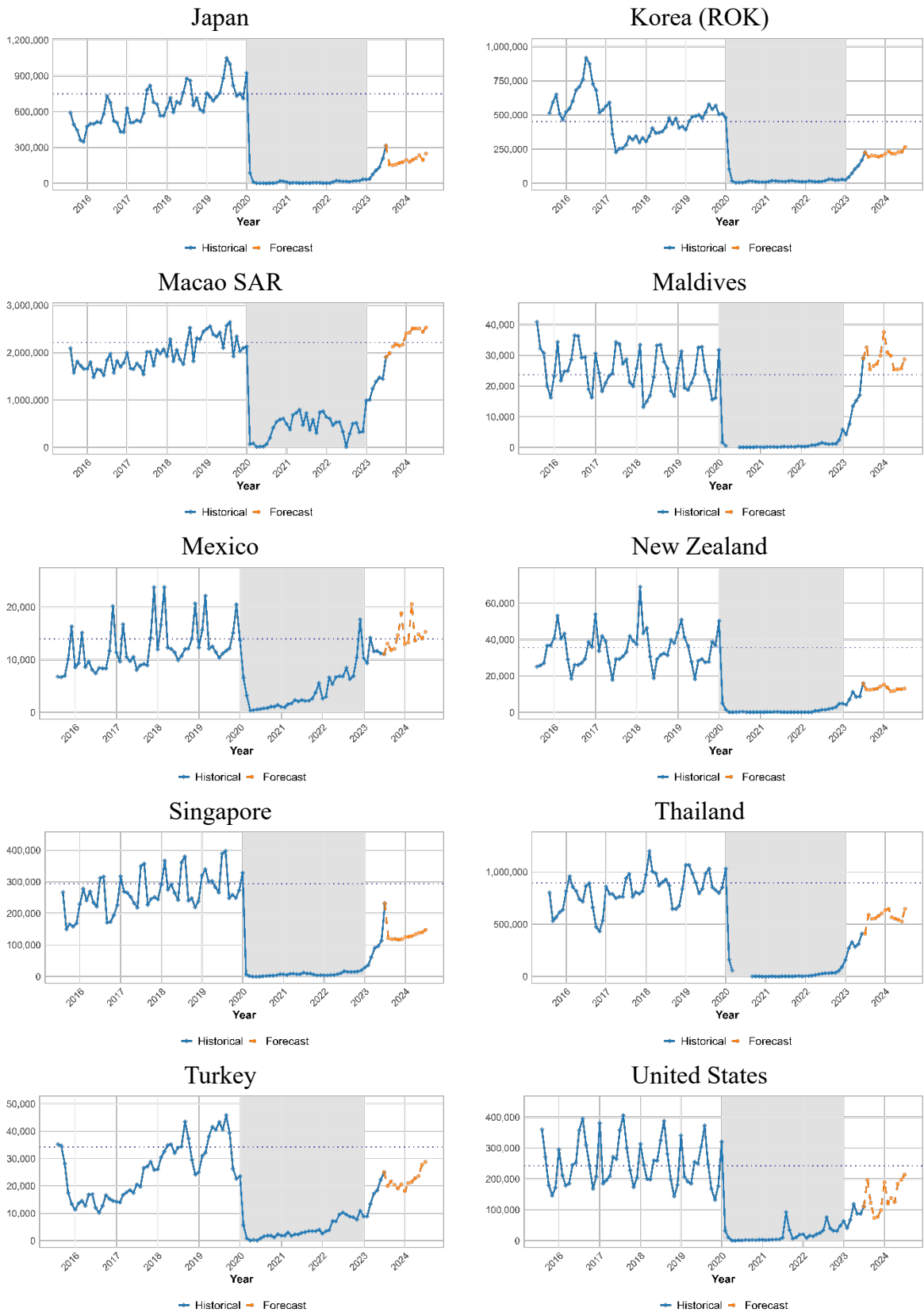


Figure 7 Final Forecasts for Chinese Outbound Tourism to 20 Destinations (Cont.)

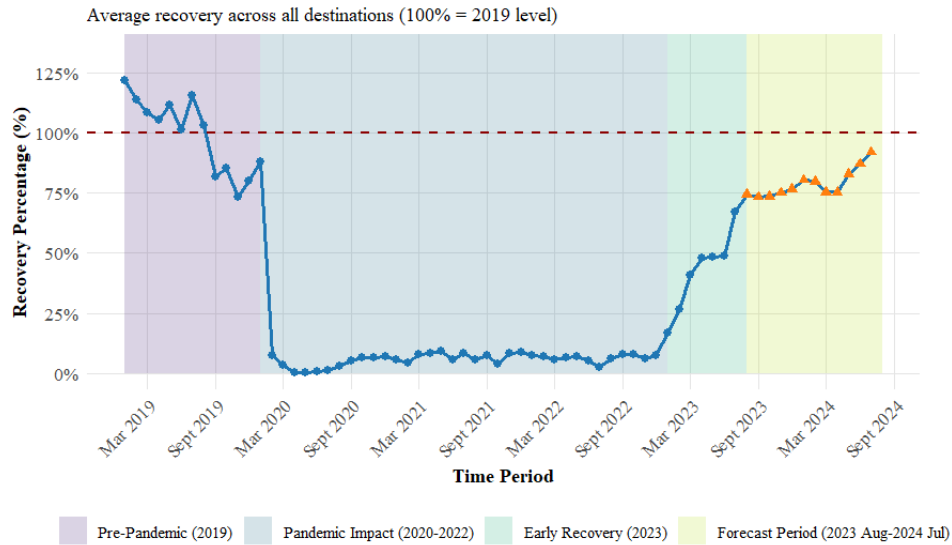


Figure 8 Chinese Outbound Tourism Recovery Trend (2019–2024)

East Asia and the Pacific

Hong Kong SAR, historically a top destination, was forecast to see arrivals increase from 3.52 million in August 2023 to 4.37 million by July 2024. This strong recovery was driven by simplified visa protocols, efficient cross-border transport (including high-speed rail), and concerted marketing campaigns. Macao SAR’s visitor numbers were likewise projected to rise from 1.98 million to over 2.53 million, fueled by visa-free short stays, coordinated border policies with Mainland China, and ongoing entertainment diversification beyond gaming.

Chinese Taipei’s predicted increase (from under 29,000 in August 2023 to nearly 49,000 arrivals in July 2024) was slower, reflecting the ongoing cross-strait geopolitical complexities. Japan and Korea (ROK) were projected to grow steadily, supported by cultural appeal, favorable exchange rates, and lenient visa policies. Departures to Japan were expected to climb from 157,100 in August 2023 to 248,600 by July 2024, while departures to Korea (ROK) are also projected to increase due to the popularity of cosmetics, shopping, and cultural festivals.

Southeast Asia

Thailand led this region, with Chinese departures rising from 588,900 to over 647,700, supported by visa-on-arrival policies and affordability. Safety concerns, however, pose a constraint. Singapore saw Chinese arrivals rising from 121,300 in August 2023 to 148,300 by July 2024, owing to increased flight connectivity, strong healthcare standards, and digital promotions that highlight its culinary and family-oriented attractions.

Indonesia benefited from Bali’s popularity, “Work from Bali” promotions, and digital visa initiatives, though growth was tempered by regional competition and residual pandemic cautions. Cambodia, featuring Angkor Wat and coastal areas, similarly experiences steady gains, though constrained by infrastructure and limited airline capacity. The Maldives saw an obvious fluctuation in tourist arrivals, possibly due to seasonality, constrained flight

availability, and intensifying competition from other island destinations.

Oceania

The number of Chinese traveling to Australia showed a marked upswing, from approximately 57,100 (August 2023) to over 146,200 by July 2024, reflecting resumed flight services, appeal of diverse natural attractions, and strong educational linkages. New Zealand followed a slower trajectory due to greater distance, fewer direct flights, and generally higher travel expenses.

Europe

Austria and the Czech Republic were projected to regain Chinese visitors at a moderate but consistent rate. Austria's totals increased from 21,500 to 33,300 over the forecast period, supported by cultural tours and winter sports, but dampened by flight and visa challenges. The Czech Republic's recovery centered on Prague but was limited by air capacity. Turkey posted modest gains, driven by heritage tourism and favorable currency rates, though political and economic instability tempered growth.

Americas

North and South American destinations exhibited the slowest recoveries, challenged by long distances, higher travel costs, and visa complexities. The United States, a key market for business, education, and leisure, eased many health restrictions but remained cautious in visa issuance; consumer confidence and airline capacity have yet to fully rebound. Canada displayed similar trends, competing with the United States for leisure and student travelers. Meanwhile, Mexico demonstrated relatively strong recovery growth due to improved air connectivity and diversified source markets but faces competition from closer tropical destinations.

Chinese departures to Hawaii nearly doubled, from about 1,700 to 3,700, supported by its premium appeal despite flight constraints and high travel costs. Chile, however, showed a slight decline—from roughly 2,534 to 2,262—reflecting distance, cost, and global economic uncertainty.

5. Conclusion

The enduring aftermath of the COVID-19 pandemic continues to hinder the complete resurgence of international tourism, particularly impacting China as a pivotal outbound market. Recognizing the inadequacy of conventional models in such volatile contexts, this study introduces a novel Triple-layer Forecast Combination Framework to forecast Chinese outbound visitor arrivals to 20 international destinations from August 2023 to July 2024. The framework's primary aim is to improve forecasting accuracy and reliability during a period of profound uncertainty by systematically integrating quantitative modeling with expert judgment.

The empirical results reveal several insights. First, the proposed framework consistently demonstrates superior predictive accuracy compared to both individual statistical models and traditional quantitative combination approaches. This highlights that combining diverse

quantitative methods and incorporating expert-driven adjustments are crucial for generating robust forecasts in dynamic environments. Second, despite the increased availability of recovery-period data partially restoring the utility of traditional statistical models, judgmental interventions—particularly when structured through Delphi-based mechanisms—remain indispensable for refining post-pandemic forecasting. Third, the recovery trajectories across destinations are highly heterogeneous, driven by a combination of factors including geographical proximity, policy responses, consumer preferences, destination appeal, and flight capacity. Notably, long-haul destinations faced a more protracted recovery than closer markets—a trend aligning with prior research (e.g., Liu et al., 2021) and highlighting geographical distance’s enduring influence.

On average, Chinese outbound departure to the 20 destinations was projected to recover to 80% of pre-pandemic (July 2019) levels by July 2024. A swifter recovery was observed among East Asian destinations (e.g., Macao SAR and Hong Kong SAR exhibiting the highest recovery rates), while long-haul regions such as Europe and North America were expected to lag. Short-haul destinations such as Thailand, Korea (ROK), and Singapore also showed relatively strong recovery. In contrast, destinations like Austria, New Zealand, and Canada were expected to have the slowest rebound. These disparities are attributable to factors like streamlined entry protocols, geographical proximity, and targeted marketing campaigns in East and Southeast Asian markets, versus lingering travel barriers, limited flight capacity, and elevated travel costs in long-haul markets. Crucially, the study emphasizes that future forecasting should not rely solely on historical patterns but also incorporate real-time, often qualitative, information to capture rapidly evolving conditions.

This study provides significant methodological contributions by developing a structured, multiphase forecasting framework that addresses critical limitations in tourism demand forecasting, especially during periods of post-crisis uncertainty. The proposed framework enhances current knowledge by introducing a formal process for integrating quantitative model selection with expert evaluations, thereby transcending purely statistical or *ad hoc* judgmental approaches and providing a methodological blueprint for combining objective performance metrics with subjective, contextual knowledge. Moreover, it incorporates judgmental adjustments as an essential, structured layer within the forecasting process, allowing systematic incorporation of real-time market signals, emerging trends, and nuanced destination-specific knowledge, which significantly enhances forecast adaptability and external validity. Finally, it provides a replicable design for combining diverse forecasting models in a coherent, multi-stage sequence, thereby improving transparency and adaptability and shifting theoretical focus from isolated model performance to designing resilient, context-aware, and methodologically integrative forecasting systems. Thus, this study contributes to the evolving discourse on forecasting under uncertainty and offers a blueprint for future methodological development in tourism demand prediction.

In addition to its methodological innovations, this study offers two theoretical contributions to the field of tourism forecasting. First, it advances the theoretical understanding of model selection in volatile environments by formalizing the role of expert judgment as a legitimate and necessary input when historical data is unreliable. This positions model selection not just as a technical procedure but as a theoretically grounded process involving representativeness,

contextual plausibility, and expert interpretation. Second, our findings contribute to theoretical debates about the comparative utility of traditional versus AI-based forecasting approaches. The frequent expert preference for models such as SARIMA, SARIMAX, and ETS—even over advanced neural network models—suggests that transparent, interpretable, and statistically grounded models retain significant theoretical and practical value, particularly under post-crisis conditions where data limitations and structural breaks persist.

The practical implications of this study are substantial, providing actionable guidance for tourism policymakers, destination-management organizations, and industry stakeholders in navigating post-pandemic recovery. The framework empowers government agencies with an adaptive, evidence-based tool for informing policy timing, border reopening strategies, and visa regulation adjustments. For instance, the observed faster recovery in East Asian destinations (e.g., Hong Kong SAR, Macao SAR) due to simplified protocols highlights the direct impact of policy on recovery speed, enabling costly over- or under-preparation in infrastructure, customs processing, and public health measures to be avoided. Destination-management organizations can further utilize the detailed, destination-specific forecasts (e.g., the strong projected growth for Japan and Korea (ROK) attributed to marketing and cultural appeal) to implement dynamic marketing campaigns and optimize promotional budgeting, which facilitates precise targeting of high-potential source markets and informed prioritization of partnerships with airlines and travel intermediaries, maximizing return on investment.

For local tourism businesses (e.g., hotels, attractions, transport operators, and retailers), these forecasts offer a crucial planning horizon for capacity management, staffing, pricing, and product development; understanding the expected pace and composition of demand recovery (e.g., the slower rebound for long-haul destinations like Hawaii) allows businesses to strategically align supply with demand, mitigate labor shortages, and optimize revenue management, enhancing operational efficiency and profitability. Broadly, this study highlights the importance of combining expert insights with quantitative rigor to inform tourism recovery strategies, encouraging governments and destination marketing organizations to institutionalize expert panels and scenario-based forecasting as routine planning components, particularly for volatile or rapidly evolving situations. The forecasting system presented here offers a replicable structure adaptable to specific market needs and generalizable for preparedness for future tourism crises.

This study has several limitations that offer avenues for further research. First, the constrained availability of highly reliable data during the early recovery period may limit model precision and the stability of long-term projections. Future research could explore advanced data-imputation techniques or leverage alternative, real-time data sources to address this. Second, while the incorporation of expert judgment via the Delphi method significantly enhances forecast contextualization, it inherently introduces potential subjectivity stemming from expert selection and interpretation. Developing standardized protocols for expert elicitation and methods for quantifying judgmental uncertainty is crucial for future research. Third, while the framework is designed to be adaptable, its current application to China's outbound tourism market means applying it to other contexts—such as inbound or domestic tourism—would necessitate recalibration to reflect different behavioral, structural, and policy dynamics. Fourth, the expert panel's rejection of the forecasts generated by the autoregressive distributed

lag model and trigonometric seasonality, Box-Cox transformation, auto-regression moving average errors, trend, and seasonal components demonstrates that post-crisis forecasting requires structural flexibility over historical fit. Future research should focus on adaptive methods like regime-switching or dynamic models capable of incorporating structural breaks and real-time data. Additionally, while the Delphi survey in Stage 3 provides crucial contextual depth, it is constrained by time lags and subjectivity. Future research could address this by integrating high-frequency indicators like search data and booking volumes to improve short-term forecast accuracy. Such data could be incorporated as an exogenous variable in quantitative models or, more pragmatically, provided to the expert panel via a real-time dashboard to inform their adjustments in Stages 2 and 3. Finally, it is important to acknowledge that this framework is time-intensive due to the considerable effort required for quantitative modeling and expert elicitation. A valuable direction for future research is the development of a Decision Support System to streamline its application. Such a system could automate the Stage 1 modeling and provide an efficient interface for experts to submit their judgments in Stages 2 and 3. This would significantly increase forecasting efficiency not by replacing human insight but by empowering experts to focus on the critical judgmental tasks where they add the most value.

References

- Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Springer. <https://doi.org/10.1007/978-0-306-47630-3>.
- Bank of Thailand. (2019). The economic impact of Chinese tourism in Thailand. Retrieved from <https://www.bot.or.th/>. Accessed July 17, 2023.
- Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451–468.
- Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: Issues and analysis. *Management Science*, 37 (5), 501–518.
- Cang, S. (2014). A comparative analysis of three types of tourism demand forecasting models: Individual, linear combination and non-linear combination. *International Journal of Tourism Research*, 16(6), 596–607.
- China Tourism Academy. (2021). *Annual report of China outbound tourism development 2021*.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559–583.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–73.
- Coshall, J. T. (2009). Combining volatility and smoothing forecasts of UK demand for international tourism. *Tourism Management*, 30(4), 495–511.
- Coshall, J. T., & Charlesworth, R. (2011). A management orientated approach to combination forecasting of tourism demand. *Tourism Management*, 32(4), 759–769.
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527.
- Flores, B. E., & White, E. M. (1988). A framework for the combination of forecasts. *Journal of the Academy of Marketing Science*, 16, 95–102.
- Fotiadis, A., Polyzos, S., & Huan, T. C. (2021). The good, the bad and the ugly on COVID-19 tourism recovery. *Annals of Tourism Research*, 87, 103117.
- Franses, P. H. (2016). A note on the mean absolute scaled error. *International Journal of Forecasting*, 32(1), 20–22.
- Fritz, R. G., Brandon, C., & Xander, J. (1984). Combining time-series and econometric forecast of tourism activity. *Annals of Tourism Research*, 11(2), 219–229.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22(4), 637–666.
- Gössling, S., Scott, D., & Hall, C. M. (2021). Pandemics, tourism and global change: A rapid assessment of COVID-19. *Journal of Sustainable Tourism*, 29(1), 1–20.
- Harvey, A. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
- Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1), 15–24.
- Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70, 489–501.

- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 26(3), 1–22.
<https://doi.org/10.18637/jss.v027.i03>.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Japan National Tourism Organization. (2019). Annual report on inbound tourism statistics. Retrieved from <https://www.jnto.go.jp/>. Accessed July 10, 2023.
- Korinth, B. (2022). Implications of Covid-19 on tourism sector in Poland: Its current state and perspectives for the future. *Tourism Recreation Research*, 47(5–6), 636–640.
- Kourentzes, N., Saayman, A., Pierre, P. J., Provenzano, D., Sahli, M., Seetaramf, N., & Volo, S. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Africa team. *Annals of Tourism Research*, 88, 103197.
- Li, G., Wu, D. C., Zhou, M., & Liu, A. (2019). The combination of interval forecasts in tourism. *Annals of Tourism Research*, 75, 363–378.
- Li, H. Y., Guo, H. G., Wang, J. Z., Wang, Y., Wu, C. Y. (2025). Tourism combination forecasting with swarm intelligence. *Annals of Tourism Research*, 111, 103932.
- Lin, V.S., Liu, A., & Song, H. (2015). Modeling and forecasting Chinese outbound tourism: an econometric approach. *Journal of Travel & Tourism Marketing*, 32(1/2), 34–49.
- Liu, A., Vici, L., Ramos, V., Giannoni, S., & Blake, A. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Europe team. *Annals of Tourism Research*, 88, 103182.
- Liu, A., Lin, V. S., Li, G., & Song, H. (2022). Ex ante tourism forecasting assessment. *Journal of Travel Research*, 61(1), 64–75.
- Liu, X., Liu, A., Chen, L., Li, G., & Song, H. (2025). Tourism demand forecasting in both normal and crisis times: Combining the bootstrap-aggregating method and the Bayesian approach. *Journal of Hospitality and Tourism Research*.
- Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *The R Journal*, 9(1), 207–218.
- Oxford Economics. (2023). *Tourism economics: Chinese outbound travel recovery forecasts*.
- Page, S., Yeoman, I., Munro, C., Connell, J., & Walker, L. (2006). A case study of best practice – Visit Scotland’s prepared response to an influenza pandemic. *Tourism Management*, 27(3), 361–393.
- Palm, F. C., & Zellner, A. (1992). To combine or not to combine? Issues of combining forecasts. *Journal of Forecasting*, 11(8), 687–701.
- Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16(3), 289–326.
- Petropoulos, F., & Siemsen, E. (2023). Forecast selection and representativeness. *Management Science*, 69(5), 2672–2690.
- Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. (2018). Judgmental selection of forecasting models. *Journal of Operations Management*, 60(1), 34–46.
- Polyzos, S., Samitas, A., & Spyridou, A. (2020). Tourism demand and the COVID-19 pandemic: An LSTM approach. *Tourism Recreation Research*, 46(2), 175–187.

- Qiu, R. T., Liu, A., Stienmetz, J. L., & Yu, Y. (2021). Timing matters: Crisis severity and occupancy rate forecasts in social unrest periods. *International Journal of Contemporary Hospitality Management*, 33(6), 2044–2064.
- Qiu, R. T., Wu, D. C., Dropsy, V., Petit, S., Pratt, S., & Ohe, Y. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and Pacific team. *Annals of Tourism Research*, 88, 103155.
- Shen, S., Li, G., & Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, 38(1), 72–89.
- Song, H., & Zhang, H. (2025). Tourism demand modelling and forecasting: A Horizon 2050 paper. *Tourism Review*, 80(1), 8–27.
- Song, H., Gao, B. Z., & Lin, V. S. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, 29(2), 295–310.
- Song, H., Lin, S., Zhang, X., & Gao, Z. (2010). Global financial/economic crisis and tourist arrival forecasts for Hong Kong. *Asia Pacific Journal of Tourism Research*, 15(2), 223–242.
- Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338–362.
- Song, H., Witt, S. F., Wong, K. F., & Wu, C. (2009). An empirical study of tourism demand forecast combination. *Journal of Hospitality and Tourism Research*, 33(1), 3–29.
- Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405–430.
- Timmermann, A. (2006). Forecast combinations. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting (1st ed.)* (pp. 135–196). Elsevier.
- UN Tourism. (2024). *World tourism barometer and statistical annex, November 2024*.
- Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4), 1518–1547.
- Wong, K. K., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, 28(4), 1068–1078.
- Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*, 29(1), 507–529.
- Wu, J., Li, M., Zhao, E., Sun, S., & Wang, S. (2023). Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tourism Management*, 98, 104759.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175.
- Zhang, H. Y., & Lu, J. P. (2022). Forecasting hotel room demand amid COVID-19. *Tourism Economics*, 28(1), 200–221.
- Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research*, 87, 103149.