Forecasting daily visits in Shanghai with Model Combination and Telco Big Data *

Nicola Camatti† Giulia Carallo†

Roberto Casarin† Xiang Feng‡

† Ca' Foscari University of Venice

‡ Shanghai Normal University

Abstract

Analyzing and forecasting visitor flows has significant potential to enhance policy decisions, facilitate long- and short-term planning of local resources, and optimize tourism offerings. Leveraging Big Data in this context provides distinct advantages, particularly regarding real-time predictions. This paper proposes an econometric approach to account for uncertainty in the prediction model, a key source of forecasting errors. A description and prediction of daily visits in Shanghai City are provided using models commonly employed in tourist flows analysis. We show that model combination and selection procedures can serve as valuable tools for policymakers, equipping them with accurate forecasts to support tourist flow management.

Keywords: Tourism flows; Mobile positioning; Smart Cities; Shanghai; Forecasting

1 Introduction

Managing the volatility of tourism dynamics has become increasingly important as destinations experience greater variability in visitor numbers across regions and over time. These oscillations, whether a spike that exceeds a destination's capacity or a rapid decline

*Corresponding author: Nicola Camatti. Address: Department of Economics, Ca' Foscari University of Venice, San Giobbe Campus, Cannaregio 873, 30121 Venice, Italy, nicola.camatti@unive.it (e-mail).
NC acknowledges financial support from the European Union, Seventh Framework Program HORIZON 2020. RC acknowledges the support from the European Union - Next Generation EU - Project 'GRINS - Growing Resilient, INclusive and Sustainable'; the National Recovery and Resilience Plan (NRRP). The computation has been performed using Matlab through the HPC cluster at the Venice Centre for Risk Analytics at Ca' Foscari University of Venice.

in visitor numbers, can trigger multidimensional socio-economic and environmental impacts. For example, when visitation exceeds acceptable sustainability limits, overtourism can detract from the tourism experience, overstress infrastructure and public services, and threaten environmental sustainability (Lin [2024], Mussoni and Vici [2025]). However, prolonged low visitation can lead to weakened local economies, diminished corporate profit, and inefficient use of public resources (Bisht et al. [2025], Blázquez-Salom et al. [2023]). Both extremes expose the structural fragility of tourism systems, exacerbate regional disparities, and threaten balanced and equitable development (Song et al. [2019], Zheng et al. [2024]).

In this context, accurate demand forecasting is essential for anticipating demand fluctuations and supporting evidence-based policy-making. Reliable projections provide government officials and key tourism stakeholders with a basis for balancing the economic advantages of tourism against the destination's environmental and infrastructural conditions, thereby minimizing risk from either excessive or insufficient demand for tourism space. Forecasting insights also support strategic planning by helping decision-makers and policymakers determine how and when to allocate limited resources, adjust operational capacities, and, if needed, use adaptive management processes to enhance resilience and sustainability in the long term (Duan et al. [2022], Fan et al. [2025], Prayag et al. [2024]).

The integration of real-time big data into forecasting is a new area of research in tourism and hospitality. Using Big Data in policy-making offers numerous advantages, particularly in providing real-time predictive insights. Access to real-time data enables the identification of daily patterns in urban systems and facilitates optimal data processing to formulate fundamental responses for policymakers, as highlighted by Kandt and Batty [2021]. The extensive collection of big data enables the analysis of tourists' preferences and characteristics, empowering decision-makers in tourism planning. The advantages of leveraging big data can help mitigate risks associated with extreme scenarios, as discussed in studies by Ardito et al. [2019] and Li et al. [2018]. In this paper, we use a Big Data dataset provided by Shanghai Unicom, the second-largest mobile phone provider in Shanghai. This dataset contains real-time tourists' positions in the Shanghai metropolis from 1 December 2020 to 17 January 2022. It has been successfully used in Camatti et al. [2024] to study tourism spillover effects within the Shanghai districts. We consider a period during which a significant event, namely the Chinese New Year, occurs and study the forecasting ability of standard models in periods with an extremely large number of visits.

Some strategies for addressing forecasting errors and model uncertainty are proposed.

Although the growing availability of high-frequency and behaviorally rich data has significantly enhanced the empirical basis of tourism forecasting, improved data alone do not automatically translate into more reliable predictions. Even with detailed real-time information on visitor movements, researchers and policymakers must still decide how to model these dynamics, which variables to include, how to specify temporal structures, and which assumptions to impose on the data-generating process. In this sense, a fundamental methodological challenge persists in the form of model uncertainty. Tourism demand forecasting has evolved from classical time-series approaches – such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA), designed to capture seasonality and persistence – to more flexible systems that integrate statistical and computational techniques (Chu [2009], Önder and Gunter [2016], Song and Li [2008], Song and Zhang [2025]). However, empirical evidence consistently shows that no single model systematically outperforms others across destinations, time horizons, or data environments. Forecast accuracy in tourism analysis depends not only on the richness of the available data, but also on model specification, parameterization, and the structural assumptions underlying the forecasting framework. Tourism systems are inherently dynamic and volatile, and forecasting errors may arise as much from specification risk as from stochastic variability in the data itself. This challenge has become increasingly relevant as the diversity of forecasting methodologies available to researchers and practitioners has grown. Traditional econometric approaches now coexist alongside machine learning techniques and hybrid frameworks that combine statistical rigor with computational flexibility. Each approach captures different dimensions of tourism dynamics, and is characterized by distinct assumptions, strengths, and limitations. Consequently, no single model can be expected to consistently outperform others across all destinations, time periods, or forecasting horizons. These considerations underscore the importance of forecasting frameworks that explicitly recognize and manage model uncertainty rather than implicitly ignoring it. By acknowledging that different models may provide complementary insights, researchers can develop more adaptive and robust forecasting systems that better capture the evolving nature of tourism demand.

An original contribution of this paper is to address model uncertainty through forecast combination. Instead of selecting a single preferred specification, we implement a Bayesian Model Averaging (BMA) framework that combines predictions from alternative models estimated on the same dataset. In the seminal work of Bates and Granger [1969] and its subsequent extensions (Billio et al. [2013], Fawcett et al. [2015], Geweke and Amisano [2011], Gneiting and Ranjan [2011, 2013], McAlinn and West [2019]), BMA assigns dynamic weights to competing models based on predictive performance. In our application, SARIMA-type models serve as structured and interpretable core specifications, and their forecasts are combined probabilistically to account for time variation in model performance. Following Billio et al. [2013] and Costantini et al. [2016], we use common forecast performance metrics to guide model dynamics and also incorporate economic evaluation to illustrate how model selection can be aligned with specific policy objectives. By integrating high-frequency mobile positioning data on Shanghai's tourist flows into a coherent probabilistic averaging framework, our study demonstrates how forecast robustness can be enhanced, specification uncertainty explicitly quantified, and predictive performance improved in periods of extreme demand. The resulting approach provides a scalable, decision-oriented tool for managing tourism volatility in complex urban environments such as Shanghai.

The remainder of the paper is organized as follows. Section 3 introduces the forecasting models and the model combination strategy. Section 4 presents the empirical application to forecasting visits to Shanghai. Section 5 concludes and outlines future research directions.

2 Literature Review

Tourism has become a key driver of economic growth in many countries and regions, stimulating the development of a wide range of related sectors such as transportation, hospitality, and culture (Lin [2024]). While tourism generates significant development opportunities, its strong interdependence with other industries increases the complexity of its management and planning processes (Chen et al. [2025], Duan et al. [2022], Mussoni and Vici [2025]). These structural characteristics require forward-looking strategies to ensure balanced and sustainable growth.

In recent years, evolving behavioral changes and market conditions in response to major events, such as the COVID-19 pandemic, naturally called for flexible, resilient, and sustainable tourism management and governance (Fan et al. [2025], Prayag et al. [2024]). Rather than focusing solely on recovery from shocks, emerging paradigms feature structural strengthening of tourism systems and their ability to adapt to demand changes and long-term sustainability challenges. Within this framework, reliable demand forecasts are crucial for effective data-driven policymaking and tourism governance. Accurate forecasting enables public authorities to anticipate fluctuations in visitor flows and align medium- and long-term investments with expected demand, thereby strengthening transport systems, optimizing environmental carrying capacity, and improving service provision (Ghalehkhondabi et al. [2019], Guizzardi et al. [2021], Wang et al. [2019]). Linking infrastructure planning and resource allocation to tourism dynamics predictions enhances policy coherence and reduces the risks of over- and under-investment at the aggregate and regional levels. Forecasting provides a critical foundation for informed decision-making in budgeting, staffing, and destination development (Jakovlev and Kitanov [2024], Zhou [2021]), while also guiding marketing and promotional strategies through alignment with anticipated demand trends (He and Qian [2025]). Beyond these operational applications, demand forecasting serves a dual role as both a management tool and a strategic instrument for driving tourism toward a more sustainable path (Liang et al. [2025]). Accurate projections favor balancing economic gains with environmental and operational limits, risk mitigation, efficient resource use, and sustainable development (Song et al. [2019], Zheng et al. [2024]). They further support the identification of ecological carrying capacities and the design interventions to safeguard fragile ecosystems (Lu and Chen [2023], Sejati et al. [2025]). Finally, by anticipating variations in visitor flows, destination managers can improve efficiency, resilience, and competitiveness (Hewapathirana [2025], Zheng et al. [2024]).

Many forecasting approaches have been developed over the last few decades, ranging from traditional time-series and econometric models to state-of-the-art artificial intelligence (AI) and hybrid approaches, with the aim of improving predictive accuracy (Bi et al. [2022], Song et al. [2019]). Although many of these methods perform well in specific empirical contexts, no single forecasting approach consistently demonstrates superior accuracy across all settings (Höpken et al. [2021], Pan et al. [2025]). Consequently, enhancing the generalizability, robustness, and adaptability of forecasting models has become a central focus of contemporary research. The historical evolution of tourism demand forecasting reflects a broader methodological and epistemological shift: from static, theory-driven econometric models to adaptive, data-integrated systems. Advances in computational power and the increasing availability of high-frequency and behaviorally rich data have fostered the convergence of statistical inference, economic reasoning, and machine learning techniques, fundamentally reshaping how researchers analyze the complexity of tourism dynamics (Song and Li [2008], Song and Zhang [2025]).

Time-series models, particularly within the Box–Jenkins framework (Box and Jenkins [1976], Box et al. [2015]), have long dominated tourism demand forecasting. Basic time-series approaches include autoregressive (AR), moving average (MA), single exponential smoothing (ES), and historical average (HA) models (Wan and Song [2018]). These models have been widely employed over the past five decades due to their straightforward implementation and their capacity to capture historical patterns effectively. Among them, ARIMA models are frequently applied (Lim and McAleer [2002]), and their seasonal extension, SARIMA, is extensively used in tourism time-series analysis (Chu [2009]). Early forecasting studies relied heavily on ARIMA and SARIMA models to capture persistence, cyclicality, and temporal dependence in tourism demand, and the empirical evidence has consistently confirmed their predictive reliability (Chu [2009], Önder and Gunter [2016]). Nevertheless, the linear structure of ARIMA-type models limits their ability to account for structural variability, nonlinear patterns, and volatility clustering that increasingly characterize tourism demand. Despite these limitations, ARIMA and SARIMA remain essential and frequently serve as interpretable cores within more flexible hybrid frameworks (Gil-Alana et al. [2004], Koc and Altinay [2007]). They have been extended along several directions. For example, the ARFIMA model introduces fractional integration (Chu [2009]), while ARIMAX and SARIMAX incorporate exogenous variables to increase explanatory power. For example, Li et al. [2018] used an ARX model to explore the impact of relative climate variability on tourism demand. Beyond ARIMA-type specifications, additional econometric frameworks have been developed to capture inter-temporal relationships and structural dynamics in tourism demand. These include the Autoregressive Distributed Lag Model (ADLM) (Huang et al. [2017]), the Error Correction Model (ECM) (Vanegas Sr [2013]), the Vector Autoregressive (VAR) model (Gunter and Önder [2016]), and Time-Varying Parameter (TVP) models (Song et al. [2011]). Moreover, time-varying volatility has been incorporated into tourism forecasting by assuming Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models for the error variance. ARIMA-GARCH and SARIMA-GARCH specifications are particularly appropriate when tourism demand exhibits conditional heteroskedasticity, allowing for seasonality and time-varying variance (Araya et al. [2024], Tsui and Balli [2017]), and have been shown to be well-suited for forecasting high-frequency flows (Chan et al. [2005]). Divino and McAleer

[2010] applied ARIMA-GARCH to model and forecast daily tourist arrivals in Peru, while Liang [2014] employed SARIMA-GARCH to forecast tourism demand in Taiwan. In this paper, we build on this literature and provide a suitable framework for incorporating uncertainty in SARIMA-GARCH specifications into tourism forecasts.

In parallel with these developments, the extensive use of large-scale, high-frequency, and behaviorally rich datasets has accelerated the need for adaptive forecasting techniques in tourism studies that combine econometric rigor with machine-learning flexibility [e.g., see Camatti et al., 2024]. Machine learning techniques such as support vector machines (Pai et al. [2014]), random forests (He and Qian [2025]), and hybrid systems that incorporate search engine data (Sun et al. [2019]) have shown strong potential for capturing nonlinear relationships and complex dependencies (Li et al. [2018]). Ensemble and hybrid frameworks, including ARIMAX artificial neural networks (NN) (Wen et al. [2019]), gradient boosting (Liu et al. [2019]), and recurrent NN (Bi et al. [2020]), combine linear and nonlinear components to improve predictive performance. However, empirical evidence shows that these approaches do not consistently outperform traditional models across all settings (Volchek et al. [2019]), and their increasing complexity raises concerns about interpretability, overfitting, and parameter sensitivity (He and Qian [2025], Wei and Mukherjee [2024]). Thus, standard ARIMA and SARIMA models continue to play a central role, often providing an interpretable baseline for hybrid and ensemble mechanisms (Coshall and Charlesworth [2011], Hassani et al. [2017], Shen et al. [2011]).

Model uncertainty arises from the coexistence of competing prediction approaches and from the recognition that the forecasting performance depends not only on data quality but also on model specification and selection, an aspect often neglected in tourism economics (Peng et al. [2014]). No model can be expected to consistently outperform all others across different time periods, destinations, or datasets (Song et al. [2019], Yang et al. [2015]). Forecast Errors may arise from structural uncertainty in specification choices, parameter selection, or omitted variables (Armstrong et al. [2015]). This awareness has redirected methodological research toward forecast combination strategies. In tourism econometrics, it has been shown that forecast combination outperforms individual models, producing more robust and accurate predictions (Claveria et al. [2015], Clements and Hendry [2004], Wang et al. [2023]). In this study, we adopt a BMA strategy instead of relying on a single selected model. The seminal contribution of Bates and Granger [1969] demonstrated that combining forecasts from different models can reduce forecast variance and enhance reliability. BMA has become standard in many fields of economics, including regional and urban economics, for addressing uncertainty in specification and variable selection (Crespo Cuaresma and Feldkircher [2013], Fernández et al. [2001], LeSage and Parent [2007], Moral-Benito [2015]). In tourism econometrics, BMA has been applied by Oh and Morzuch [2005] via an equally weighted combination of the best-performing time-series models. An application to travelers' arrivals to Singapore between 1977 and 1990 shows that BMA forecasts always outperform the poorest individual model, and perform better than the best model. Wong et al. [2007] applied BMA to Hong Kong inbound tourism forecasting and found that combined forecasts can outperform the worst individual models. Compared to the two studies, we extend the analysis from point to density forecasts, from static to dynamic weighting, and from mean squared error and mean absolute error measures to general performance measures. See also Song and Li [2008], Song et al. [2019], Song and Zhang [2025] for a review. Outside the BMA, a frequentist approach based on Weighted-Average Least Squares (WALS) has been applied by Wan and Zhang [2009] to combine time-series models for visitor arrivals in Barbados between 1956 and 1992. The authors found that WALS forecasts outperform the worst model in the combination. The advantage of WALS over BMA is its lower computational cost; however, its optimality is not guaranteed when used with ARIMA models, which are standard in tourism econometrics. A summary of the main approaches to tourism forecasting (Time Series Models, Time Series with Stochastic Volatility, Machine Learning, Hybrid Machine Learning, Ensemble Machine Learning, Forecast Combination), including the references discussed in this paper, is given in Table 1. Rather than selecting a single "best" model, BMA assigns posterior probabilities to models based on their predictive performance and computes a weighted average of individual model predictions. By combining predictive distributions, BMA explicitly quantifies uncertainty and produces forecasts that are statistically consistent and operationally robust (Billio et al. [2013], Geweke and Amisano [2011], Gneiting and Ranjan [2013]). In tourism forecasting models, uncertainty has been recognized as a central challenge, motivating the use of forecast combination (Coshall and Charlesworth [2011], Peng et al. [2014], Song and Li [2008]).

Among existing approaches to BMA, a modeling framework known as Bayesian agent and expert opinion analysis was introduced by Lindley et al. [1979] and subsequently discussed in later works (Genest and Schervish [1985], West [1984, 1988, 1992]). Recently, this approach has been further extended through Bayesian predictive synthesis (John-

Approach	Model/Algorithm	References	Description
TSM	AR, MA, ES, HA	Wan and Song [2018]	Basic models used to capture historical patterns; simple and widely applied.
	ARIMA	Chu [2009], Lim and McAleer [2002], Önder and Gunter [2016]	Standard model for capturing autocorrelation and temporal dynamics; strong empirical reliability.
	SARIMA	Chu [2009], Önder and Gunter [2016]	Seasonal extension of ARIMA, suitable for tourism data with seasonality.
	ARFIMA	Chu [2009]	Introduces fractional integration to capture long-memory processes.
	ARIMAX / SARIMAX / ARX	Li et al. [2018]	Incorporates exogenous variables to enhance explanatory power.
	ADLM	Huang et al. [2017]	Dynamic model with distributed lags for intertemporal relationships.
	ECM	Vanegas Sr [2013]	Error correction model capturing long-run equilibrium relationships.
	VAR	Gunter and Önder [2016]	Multivariate framework capturing interdependencies among variables.
	TVP models	Song et al. [2011]	Time-varying parameters to account for structural changes.
TSM-V	ARIMA-GARCH	Araya et al. [2024], Divino and McAleer [2010], Tsui and Balli [2017]	Combines mean dynamics with conditional volatility modeling.
	SARIMA-GARCH	Araya et al. [2024], Liang [2014], Tsui and Balli [2017]	Incorporates both seasonality and heteroskedasticity.
ML	Support Vector Machines	Pai et al. [2014]	Captures complex nonlinear relationships.
	Random Forest	He and Qian [2025]	Ensemble method robust to nonlinearities and interactions.
	Search data-based models	Sun et al. [2019]	Integrates search engine data to improve forecasting performance.
H-ML, E-ML	ARIMAX + Neural Networks	Wen et al. [2019]	Combines linear and nonlinear components.
	Gradient Boosting	Liu et al. [2019]	Boosting ensemble approach enhancing predictive accuracy.
	Recurrent Neural Networks	Bi et al. [2020]	Suitable for sequential data and dynamic temporal dependencies.
FC	Bayesian Model Averaging	Oh and Morzuch [2005], Wong et al. [2007]	Applied classical BMA to linear models, including ARIMA and ADL, with static and standard combination weights.
	Frequentist Model Averaging	Wan and Zhang [2009]	Applied Weighted Average Least Square to linear regression and ARIMA models and provide combination confidence bounds.

Table 1: Summary of the literature review on tourism demand forecasting models. Approaches: Time Series Models (TSM), TSM with Dynamic Volatility (TSM-SV), Machine Learning (ML), Hybrid ML (H-ML), Ensemble ML (E-ML), Forecast Combination (FC).

Method	References	Description
FC	Bates and Granger [1969], Claveria et al. [2015], Clements and Hendry [2004], Wang et al. [2023]	Combines multiple models to reduce variance and improve robustness.
BMA	Crespo Cuaresma and Feldkircher [2013], Fernández et al. [2001], LeSage and Parent [2007], Moral-Benito [2015]	Weighted averaging of models based on posterior probabilities.
EOM	Genest and Schervish [1985], Lindley et al. [1979], West [1984, 1988, 1992]	Modeling framework combining judgmental probability assessments.
BPS	Aastveit et al. [2023], Johnson and West [2025], McAlinn [2021], McAlinn et al. [2020], McAlinn and West [2019], Tallman and West [2024]	Probabilistic framework combining predictive distributions.
DMA	Billio et al. [2013], Geweke and Amisano [2011], Raftery et al. [1997]	Model weights evolve over time based on performance.
BMA-E	Liu et al. [2025], Llewellyn et al. [2023], Sun et al. [2023]	Advanced approaches with adaptive weights and alternative data.

Table 2: Summary of the literature review on model and forecast combination. Methods: Forecast Combination (FC), Bayesian Model Averaging (BMA), Bayesian Predictive Synthesis (BPS), Expert Opinion Models (EOM), Dynamic Model Averaging (DMA), BMA Extensions (BMA-E)

son and West [2025], McAlinn et al. [2020], McAlinn and West [2019]), which provides a probabilistic and coherent framework. Recent advances extend BMA through time- and feature-varying model averaging (Sun et al. [2023]), hybrid Bayesian–bootstrap approaches (Liu et al. [2025]), and adaptive weighting schemes incorporating alternative data sources (Llewellyn et al. [2023]), while broader Bayesian Model Combination frameworks such as Dynamic Model Averaging allow model weights to evolve with predictive performance (Billio et al. [2013], Geweke and Amisano [2011], Raftery et al. [1997]). A summary of the main approaches to forecast combination (Forecast Combination, BMA, Bayesian Predictive Synthesis, Expert Opinion Models, Dynamic Model Averaging, BMA Extensions), including the references discussed in this paper, is given in Table 2.

In this paper, we apply BMA to a set of alternative SARIMA specifications with and without GARCH errors. Within our BMA framework, we assume that the combination weights depend on the model's predictive performance. Standard predictive performance metrics are considered (e.g., see Costantini et al. [2016]) together with cost functions in order to account for the economic evaluation of the forecast and to align statistical relevance and policy and tourism management objectives (Geweke and Amisano [2011], Gneiting and Ranjan [2011, 2013], McAlinn and West [2019]).

Expanding on this methodological rationale, the empirical phase of this research utilizes mobile positioning data to implement the proposed forecasting framework. These data, obtained from mobile network capabilities, provide detailed, real-time data on tourist presence and mobility behaviors (Chen et al. [2025], Metulini and Carpita [2024]). Mobile data provide dynamic measures of how people move through space and time, mitigating some of the limitations of conventional surveys and administrative data sources. In addition, mobile data have been used to understand the spatial and temporal impacts of social and cultural events, thereby further expanding their relevance to tourism research. More recently, the development of big data—high-frequency, complex information from mobile devices, social media, and digital transactions—has transformed our understanding of tourist behavior (Wu et al. [2025]). These data sources can identify travel pathways, peak engagement periods, and popular attractions, yielding useful insights for destination and infrastructure management. Predictive analytics will further extend this capacity to understand mobility patterns and data-informed tourism policymaking (Ramos et al. [2021]).

Nevertheless, predicting mobility is inherently challenging due to nonlinear interactions among behavioral, environmental, and demographic variables. Classic models (ARIMA models) rely on linearity and stationarity assumptions, while machine learning models (methods) can better handle (model nonlinearity) nonlinear functions but frequently do not capture temporal dependence. To address these issues, recent studies have employed ensemble frameworks to combine existing models, leveraging the strengths of both traditional and machine-learning approaches, thereby improving flexibility and predictive performance while remaining cautious about overfitting and parameter selection (Punia and Shankar [2022], Wei and Mukherjee [2024]). The current study put forward the combination of mobile positioning data within a BMA framework – a probabilistic approach that weights multiple model specifications dynamically based on predictive performance. This controls for uncertainty while also capturing the temporal and structural variability of tourist mobility. With the addition of mobile data, the new approach improves forecasting accuracy and resilience while leveraging BMA's flexibility to produce predictive outcomes, offering a scalable, generalizable tool for adaptive and sustainable tourism management.

3 Methodology

The temporal evolution of visitation rates across statistical areas reflects the inherently dynamic nature of tourism demand. The temporal patterns in observed visitation rates across statistical areas exhibit both periodicity and trend. Understanding these temporal dynamics is essential for tourism planning, destination management, and the efficient allocation of resources and services. Rather than fluctuating randomly, visitation counts exhibit clear periodic behavior alongside persistent changes over time, highlighting the need for modeling approaches that capture both regular cycles and stochastic variation. To capture these complex patterns, this study adopts dynamic stochastic models that represent the persistent, cyclical, and irregular components embedded in tourism visitation counts. Such models are particularly valuable in tourism analytics because they not only provide accurate forecasts of future demand but also help identify underlying temporal structures critical to policy design, infrastructure planning, and market responsiveness. Their flexibility and robustness have made them widely applicable in tourism forecasting research, and they can be readily implemented using modern analytical environments such as Python, MATLAB, and R. In this study, the empirical analysis was conducted in MATLAB using the Econometrics Toolbox. Specifically, we used the `arima`, `estimate`, `forecast`, `garch`, `lbqtest`, `mae`, `mape`, and `rmse` libraries within the econometric toolbox in Matlab, for inference and model diagnostics.

3.1 Bayesian predictive synthesis

Bayesian predictive synthesis (BPS) and various generalizations of the framework have been introduced in particular for forecasting in economics and finance by Aastveit et al. [2023], Bassetti et al. [2018], McAlinn [2021], McAlinn et al. [2020]. Johnson and West [2025] and McAlinn and West [2019] introduced the mixture-BPS, which includes Bayesian Model Averaging as a special case.

Let $\mathcal{M}_j$, with $j = 1, \ldots, J$, be a set of J models, where each $\mathcal{M}_j$ defines a predictive density function $p_j(\mathbf{y}|\mathcal{M}_j)$, which set is defined as $\mathcal{P} = \{p_1(\mathbf{y}|\mathcal{M}_1), \ldots, p_J(\mathbf{y}|\mathcal{M}_J)\}$. Let $\pi_j = P(\mathcal{M}_j)$ be the model probabilities. Under mixture-BPS , the predictive distribution, conditional on learning $\mathcal{P}$, can be written as

$$f(\mathbf{y}|\mathcal{P}) \propto \sum_{j=0}^{J} \pi_j \alpha_j(\mathbf{y}) p_j(\mathbf{y}|\mathcal{M}_j), \tag{1}$$

where $\alpha_j(\cdot)$ $j = 0, 1, \ldots, J$ are non-negative weight functions and π_j $j = 0, 1, \ldots, J$ are the initial model probabilities. The probability density function of the baseline model $\mathcal{M}_0$ is denoted with $p_0(\mathbf{y}|\mathcal{M}_0)$. Equation 1 can be rewritten so that models can be explored for various choices of $\alpha(\mathbf{y})$. The predictive distribution $f(\mathbf{y}|\mathcal{P})$ becomes:

$$f(\mathbf{y}|\mathcal{P}) = \sum_{j=0}^{J} \tilde{\pi}_j f_j(\mathbf{y}|\mathcal{M}_j),\ f_j(\mathbf{y}|\mathcal{M}_j) = \frac{\alpha_j(\mathbf{y}) p_j(\mathbf{y}|\mathcal{M}_j)}{\alpha_j},\ j = 0, 1, \ldots, J \tag{2}$$

where

$$\alpha_j = \int_y \alpha_j(\mathbf{y}) p_j(\mathbf{y}|\mathcal{M}_j) d\mathbf{y},\ \tilde{\pi}_j = k\pi_j\alpha_j,\ k^{-1} = \sum_{j=0}^{J} \pi_j\alpha_j. \tag{3}$$

Bayesian Model Averaging is a special case of the mixture BPS when $\pi_0 = 0$ and $\alpha_j(\mathbf{y}) = 1$ for $j = 1, \ldots, J$ (e.g., see Tallman and West [2024]).

3.2 Model classes

The stochastic model, coupled with an inference procedure, facilitates generating daily predictions of the variable of interest over a specified time horizon. In the subsequent section, we illustrate a 60-day forecast for Shanghai City, using the aggregated statistical area. This methodology is versatile and can be extended to provide disaggregated forecasting analyses, including area-specific predictions for various statistical areas of interest. Furthermore, it enables predictions of visitor origin and type, as well as forecasts of flows between and within statistical areas.

Empirical tourism analysis must explicitly model trend and seasonal behaviour and ignoring these components can lead to biased conclusions. Beyond weather conditions – particularly extremes such as humid summers that shape leisure demand – seasonal patterns in Shanghai visits are also influenced by pronounced domestic tourism peaks and recurring business travel cycles. We employ SARIMA models, initially proposed by Box and Jenkins [1976], which are particularly well suited to our dataset. Differencing is used to address the trend, while seasonal differencing captures seasonal dynamics. Let us introduce the lag polynomials $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\theta(L) = 1 - \theta_1 L - \ldots - \theta_q L^q$ for the non-seasonal dynamics and $\Phi(L^s) = 1 - \Phi_1 L^s - \ldots - \Phi_P L^{sP}$ and $\Theta(L^s) = 1 - \Theta_1 L^s - \ldots - \Theta_Q L^{sQ}$ for the seasonal dynamics, with $(1-L)^d z_t$ the dth order differencing operator, $(1-L^s)^D z_t$ the Dth order seasonal differencing operator, p the autoregressive order, q the moving average order, d the number of differencing operations, P, D and Q the seasonal orders, and s the seasonality. The $SARIMA(p, d, q) \times (P, D, Q)_s$ model takes the following form:

$$\phi(L)\Phi(L^s)(1 - L^s)^D (1 - L)^d z_t = \theta(L)\Theta(L^s) u_t. \tag{4}$$

See also Box et al. [2015] for further details.

Time–varying volatility, that is heteroschedasticity, in tourist visits arises from the persistence or clustering of exogenous shocks, such as pandemics, exchange rate movements, or global income fluctuations. Other factors, such as mega-events, expos, and sudden travel booms, can generate bursts of variability that are not constant over time. These features are well–captured by GARCH models. In the GARCH(p, q) specification Bollerslev [1987], it is assumed that $\mathbb{E}(u_t|\mathcal{F}_{t-1}) = 0$ and $\mathbb{V}(u_t|\mathcal{F}_{t-1}) = \sigma_t^2$, where:

$$\sigma_t^2 = \alpha(L)u_t^2 + \beta(L)\sigma_t^2, \tag{5}$$

where $\mathcal{F}_{t-1}$ is the information set including past observations $\mathcal{F}_{t-1} = \sigma(\{z_s\}_{s\leq t-1})$ with lag polynomials $\alpha(L) = \alpha_0 + \alpha_1 L + \ldots, \alpha_q L^q$ for the ARCH component and $\beta(L) = \beta_1 L + \ldots + \beta_p L^p$ for the GARCH component. The SARIMA and SARIMA-GARCH models are estimated by maximum likelihood (Ling and McAleer [2003a,b]).

3.3 Combination weight specification

We consider Bayesian model averaging (e.g., see Bates and Granger [1969]) and assume time-varying combination weights. Following Costantini et al. [2016], we consider some of the most used forecasting performance measures to build the combination weights. The weights are specified based on a rolling estimate of the model's performance measures (e.g., see Billio et al. [2013]). Three groups of measures are considered: i) point forecast ability, ii) density forecast ability, and iii) economic evaluation. In the following, $\hat{y}_{\ell,s}$ denote the forecast made at time $s-1$ for time s using model $\mathcal{M}_\ell$, $\ell = 1, \ldots, N$ with N the number of models, and τ denotes the window size.

The first group of measures includes: Mean Squared Forecast Error (MSFE), Exponential Mean Squared Forecast Error (EMSFE), Mean Squared Root Forecast Error (MSRFE), Discount Mean Squared Forecast Error (DMSFE), Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), Hit/Success Rates (HR), and Exponential of Hit/Success Rates (EHR). The MSFE and EMSFE weights are:

$$w_{j,t}^{(MSFE)} = \frac{MSFE_{j,t}^{-1}}{\sum_{\ell=1}^{N} MSFE_{\ell,t}^{-1}}, \quad w_{j,t}^{(EMSFE)} = \frac{\exp(-MSFE_{j,t})}{\sum_{\ell=1}^{N} \exp(-MSFE_{\ell,t})} \tag{6}$$

$$\text{MSFE}_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} (\hat{y}_{\ell,s} - y_s)^2, \tag{7}$$

where k_j denotes the number of parameters of model $\mathcal{M}_j$. The exponential transform is commonly used in this literature to ensure numerical stability and smoothness of the weights when the MSFE is close to zero. The MSRFE and DMSFE weights are defined as:

$$w_{j,t}^{(MSRFE)} = \frac{MSRFE_{j,t}^{-1}}{\sum_{\ell=1}^{N} MSRFE_{\ell,t}^{-1}}, \quad MSRFE_{\ell,t} = \sqrt{\frac{1}{\tau} \sum_{s=t-\tau+1}^{t} (\hat{y}_{\ell,s} - y_s)^2} \tag{8}$$

$$w_{j,t}^{(DMSFE)} = \frac{DMSFE_{j,t}^{-1}}{\sum_{\ell=1}^{N} DMSFE_{\ell,t}^{-1}}, \quad DMSFE_{\ell,t} = \sum_{s=t-\tau+1}^{t} \theta^{t-s} (\hat{y}_{\ell,s} - y_s)^2 \tag{9}$$

where $0 \leq w_{j,t} \leq 1$ are the combination weights, such that $w_{1,t} + \ldots + w_{N,t} = 1$. Similarly, MAE and MAPE weights are defined as

$$w_{j,t}^{(MAE)} = \frac{MAE_{j,t}^{-1}}{\sum_{\ell=1}^{N} MAE_{\ell,t}^{-1}}, \quad \text{MAE}_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} |\hat{y}_{\ell,s} - y_s|, \tag{10}$$

$$w_{j,t}^{(MAPE)} = \frac{MAPE_{j,t}^{-1}}{\sum_{\ell=1}^{N} MAPE_{\ell,t}^{-1}}, \quad \text{MAPE}_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} \frac{|\hat{y}_{\ell,s} - y_s|}{y_s}. \tag{11}$$

HR and EHR are built using the DA measure:

$$w_{j,t}^{(HR)} = \frac{DA_{j,t}}{\sum_{\ell=1}^{N} DA_{\ell,t}}, \quad w_{j,t}^{(EHR)} = \frac{\exp(DA_{j,t})}{\sum_{\ell=1}^{N} \exp(DA_{\ell,t})}, \tag{12}$$

$$\text{DA}_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} \mathbb{I}(\text{sgn}(\hat{y}_{\ell,s} - \hat{y}_{\ell,s-1}) = \text{sgn}(y_s - y_{s-1})), \tag{13}$$

where $\mathbb{I}(x = y)$ is the indicator function which take values 1 if $x = y$ and 0 otherwise and $\text{sgn}(x)$ returns -1 if x is negative, 0 if x is null, and 1 if x is positive.

The second group comprises measures of density-forecast ability, such as Log-Score

(LS), Continuous Ranked Probability Score (CRPS), and Frequentist Model Averaging (FMA). Let $IC_{\ell,s}$ be the predictive information criterion of model $\mathcal{M}_\ell$ at time s evaluated on the point forecast $\hat{y}_{\ell,s}$, $F_{\ell,s}(\cdot) = F(\cdot|\mathcal{M}_\ell, y_1, \ldots, y_{s-1})$ the forecasting distribution at time $s-1$ for time s, and $f_{\ell,t}(\cdot)$ its probability density function. The CRPS, LS, and FMA weights are defined as:

$$w_{j,t}^{(LS)} = \frac{LS_{j,t}}{\sum\limits_{\ell=1}^{N} LS_{\ell,t}}, \quad \mathrm{LS}_{\ell,t} = -\frac{1}{\tau} \sum_{s=t-\tau+1}^{t} \log f_{\ell,s}(y_s), \tag{14}$$

$$w_{j,t}^{(CRPS)} = \frac{CRPS_{j,t}}{\sum\limits_{\ell=1}^{N} CRPS_{\ell,t}}, \quad \mathrm{CRPS}_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} \int_{-\infty}^{\infty} \left(F_{\ell,s}(\hat{y}_s) - \mathbb{I}\{\hat{y}_s \geq y_s\}\right)^2 d\hat{y}_s, \tag{15}$$

$$w_{j,t}^{(FMA)} = \frac{ICj,t}{\sum\limits_{\ell=1}^{N} IC\ell,t}, \quad \mathrm{IC}_{\ell,t} = -2 \sum_{s=t-\tau+1}^{t} \log(f_{\ell,s}(\hat{y}_{\ell,t})) + w(\tau,k), \tag{16}$$

where $w(\tau,k) = 2k$ for the Akaike Information Criterion (AIC), $w(\tau,k) = \log(\tau)k$ for the Bayesian Information Criterion (BIC) and $w(\tau,k) = 2\log(\log(\tau))k$ for the Hannan and Quinn Criterion (HQ), with k the number of parameters of a given model.

The third group comprises two measures of the economic impact of forecast errors: the Economic Evaluation of Directional Forecasts (EEDF) and the Weighted MSFE (WMSFE). The EEDF is built on the Directional Value (DV) measure

$$w_{j,t}^{(EEDF)} = \frac{DV_{j,t}}{\sum\limits_{\ell=1}^{N} DV_{\ell,t}}, \quad \mathrm{DV}_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} |y_s - y_{s-1}| \mathbb{I}(\mathrm{sgn}(\hat{y}_{\ell,s} - \hat{y}_{\ell,s-1}) = \mathrm{sgn}(y_s - y_{s-1})) \tag{17}$$

which indicates the model's ability to capture the correct change in direction of the predicted variable, accounting for the magnitude of the realized changes. The WMSFE exploits the economic relevance of forecasting, incorporating the costs incurred by the municipality related to tourism. The idea is to measure how prediction errors vary across days of the year, accounting for cost asymmetries in over- and underestimating flows across holiday and non-holiday periods. WMSFE is computed for each model as follows:

$$\tilde{w}_{j,t} = \frac{WMSFE_{j,t}^{-1}}{\sum_{\ell=1}^{N} WMSFE_{\ell,t}^{-1}}, \quad WMSFE_{\ell,t} = \frac{1}{\tau} \sum_{s=t-\tau+1}^{t} (\hat{y}_{\ell,s} - y_s)^2 c_s \tag{18}$$

such that $0 \leq \tilde{w}_{j,t} \leq 1$ and $\tilde{w}_{1,t} + \ldots + \tilde{w}_{N,t} = 1$. The weights c_s are chosen based on policy analysis, such as the extra costs associated with overtourism during certain periods of the year.

4 Empirical Analysis

4.1 Data description

The dataset comprises daily tourist visits to Shanghai from 1 December 2020 to 17 January 2022, covering all 16 administrative districts—Huangpu, Xuhui, Changning, Jing'an, Putuo, Hongkou, Yangpu, Minhang, Baoshan, Jiading, Pudong, Jinshan, Songjiang, Qingpu, Fengxian, and Chongming. It is derived from real-time positional data of mobile phone signals provided by Shanghai Unicom, the city's second-largest mobile service provider. The dataset encompasses users from diverse telephone companies in proportions reflecting their market shares, and it is relied upon by the Shanghai municipal government for tourism flow prediction, underscoring its accuracy, inclusiveness, and institutional relevance. The final panel contains 409 daily observations per district [e.g., see Camatti et al., 2024]. The target population comprises tourists visiting Shanghai's districts during the study period. Operationally, a "tourist" is defined as a mobile phone user who records an overnight stay in a district either on the night prior to or following the visit, and whose activity is not attributable to other categories of users. China Unicom provided pre-filtered data to exclude residents, hikers, workers, students, and commuters. Under the intra-destination flow framework, a tourist is counted once per district per day, although multiple districts may be visited on the same day. Based on SIM card activation records, tourists are classified as either Shanghai locals or tourists from outside Shanghai. These clearly defined behavioral criteria ensure that the analytical population closely corresponds to visitors engaging with districts across Shanghai, thereby supporting data representativeness and comprehensive coverage of the defined tourist population.

The top plot in Figure 1 represents the trends in visits to Shanghai during the period from December 1, 2020, to January 16, 2022. We conducted a preliminary analysis of the dataset by examining the temporal dependence structure. The autocorrelation functions in Panel (b) allow for identifying persistence and seasonal effects. In addition, Phillips–Perron and augmented Dickey–Fuller tests reject the null hypothesis of a unit root at the 5% level for all series. Together, these results provide a compelling rationale for employing seasonal ARMA models to capture the dynamic structure of tourism flows

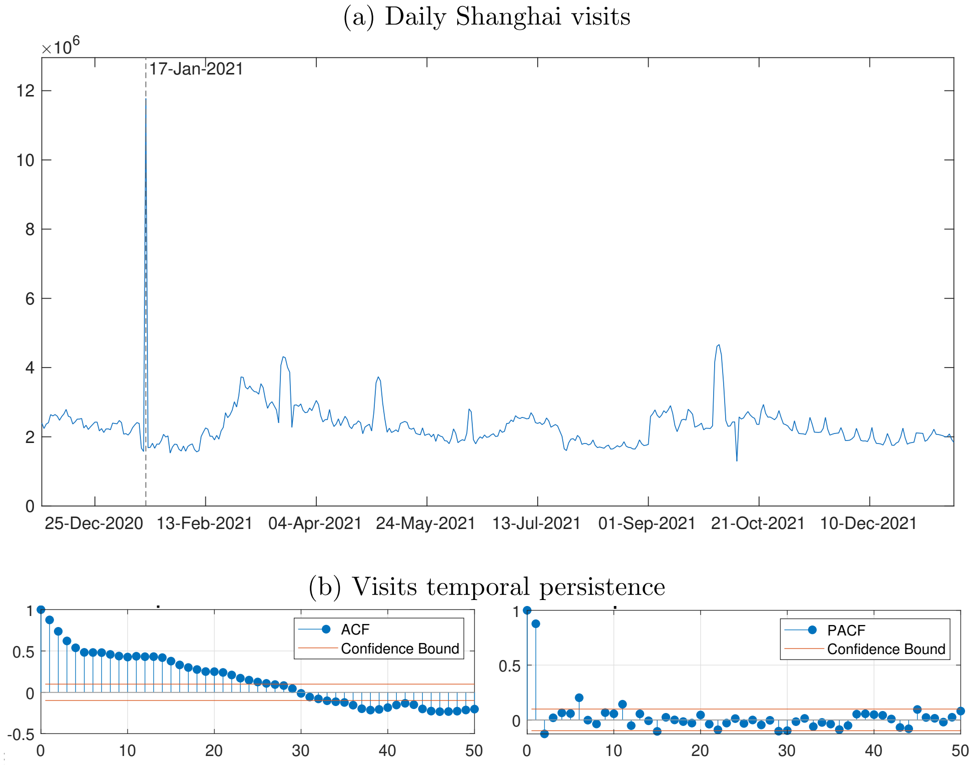

Figure 1: Shanghai visits data. Top: Millions of visits to Shanghai in the period from 1^{st} December 2020 to 16^{th} January 2022. The vertical dashed line highlights an outlier. Bottom: Autocorrelation (left) and partial autocorrelation functions (right), which are two measures of dependence of the current number of visits on the numbers in the days before (from 1 to 50).

across Shanghai's districts.

4.2 Model selection

We estimate the following ten different SARIMA specifications. Following the literature, low-order AR and MA terms are adopted for the non-seasonal component, with first-order integration [e.g., Song et al., 2019, Wan and Zhang, 2009], while the seasonal specification is selected after a preliminary AIC-based analysis. Of the 20 models initially specified, we report results for the eight with the lowest AIC. GARCH components with lags of 1 and 2 are specified in line with the standard tourism time-series literature [e.g., Chan et al., 2005, Divino and McAleer, 2010, Tsui and Balli, 2017]. Of the 7 GARCH models initially specified, we report results for the 2 with the lowest AIC.

The first three models in our set are two SARIMA$(2,0,2)(24,0,0)_5$ without and with restrictions and a SARIMA$(2,0,2)(24,0,0)_5$ without intercept:

- $\mathcal{M}_1$: $(1-\phi_2 L^2)(1-\Phi_5 L^5-\Phi_{10}L^{10}-\Phi_{50}L^{50}-\Phi_{80}L^{80}-\Phi_{120}L^{120})y_t = \mu+(1+\theta_2 L^2)u_t$,
- $\mathcal{M}_2$: $(1-\phi_1 L-\phi_2 L^2)(1-\Phi_5 L^5-\Phi_{10}L^{10}-\Phi_{50}L^{50}-\Phi_{80}L^{80}-\Phi_{120}L^{120})y_t = \mu+(1+\theta L+\theta_2 L^2)u_t$,
- $\mathcal{M}_3$: $(1-\phi_1 L-\phi_2 L^2)(1-\Phi_5 L^5-\Phi_{10}L^{10}-\Phi_{50}L^{50}-\Phi_{80}L^{80}-\Phi_{120}L^{120})y_t = (1+\theta L+\theta_2 L^2)u_t$.

The second block of models includes a SARIMA$(2,1,0)(16,0,0)_5$ and a SARIMA$(0,1,2)(16,0,0)_5$:

- $\mathcal{M}_4$: $(1-\phi_1 L-\phi_2 L^2)(1-L)(1-\Theta_5 L^5-\Theta_{40}L^{40}-\Theta_{80}L^{80})y_t = \mu+u_t$,
- $\mathcal{M}_5$: $(1-L)(1-\Theta_5 L^5-\Theta_{40}L^{40}-\Theta_{80}L^{80})y_t = \mu+(1+\theta L+\theta_2 L^2)u_t$.

The third block includes three models without MA components, that are a SARIMA$(1,1,0)(16,0,0)_5$, a SARIMA$(2,0,0)(16,0,0)_5$ and a SARIMA$(2,0,0)(24,0,0)_5$:

- $\mathcal{M}_6$: $(1-\phi_1 L)(1-L)(1-\Theta_5 L^5-\Theta_{40}L^{40}-\Theta_{80}L^{80})y_t = \mu+u_t$,
- $\mathcal{M}_7$: $(1-\phi_1 L-\phi_2 L^2)(1-\Theta_5 L^5-\Theta_{40}L^{40}-\Theta_{80}L^{80})y_t = \mu+u_t$,
- $\mathcal{M}_8$: $(1-\phi_1 L-\phi_2 L^2)(1-\Theta_5 L^5-\Theta_{40}L^{40}-\Theta_{80}L^{80}-\Theta_{120}L^{120})y_t = \mu+u_t$.

The last block includes heteroskedasticity in the different model specifications. Namely, a SARIMA$(2,0,2)(24,0,0)_5$ with GARCH(2,2) errors model and a SARIMA$(2,0,0)(24,0,0)_5$ with GARCH(2,2) errors model:

- $\mathcal{M}_9$: $(1-\phi_1 L-\phi_2 L^2)(1-\Phi_5 L^5-\Phi_{10}L^{10}-\Phi_{50}L^{50}-\Phi_{80}L^{80}-\Phi_{120}L^{120})y_t = \mu+(1+\theta L+\theta_2 L^2)u_t,\ u_t=\sigma_t z_t$,
- $\mathcal{M}_{10}$: $(1-\phi_1 L-\phi_2 L^2)(1-\Theta_5 L^5-\Theta_{40}L^{40}-\Theta_{80}L^{80}-\Theta_{120}L^{120})y_t = \mu+u_t,\ u_t=\sigma_t z_t$,

where for both models the GARCH(2,2) component writes as $\sigma_t^2 = \gamma_0+\alpha_1\sigma_{t-1}^2+\alpha_1\sigma_{t-2}^2+\beta_1 z_{t-1}^2+\beta_2 z_{t-2}^2$. We conducted various diagnostic checks on the residuals for each model specification, employing autocorrelation and periodogram analyses. We compared the residuals' quantiles to theoretical quantiles of a normal distribution and applied the Ljung-Box Q-test to assess residual autocorrelation. Figures A.5-A.14 in the Appendix present the results of these diagnostic checks for each model.

The optimal model was selected by minimizing the absolute values of both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The values for AIC and BIC are presented in Table 3. Diagnostic results indicate that the preferred

Model	AIC	BIC	MAE	MAPE	RMSFE	LS	CRPS	DA	DV
$\mathcal{M}_1$	123.56	158.26	2.35	101.52	2.52	-8.44	2.01	0.51	1.10
$\mathcal{M}_2$	-102.60	-60.19	2.37	103.42	2.49	-16.00	2.13	0.47	1.02
$\mathcal{M}_3$	-104.51	-65.96	2.36	102.47	2.51	-16.44	2.11	0.48	1.01
$\mathcal{M}_4$	102.48	129.47	9.84	455.98	11.14	-42.38	9.51	0.48	1.04
$\mathcal{M}_5$	103.06	130.05	1.81	86.98	2.54	-7.59	1.52	0.51	1.11
$\mathcal{M}_6$	103.53	126.66	**1.06**	**45.99**	**1.28**	**-2.40**	**0.79**	0.49	1.06
$\mathcal{M}_7$	84.10	111.09	2.14	92.96	2.34	-8.06	1.82	0.48	1.04
$\mathcal{M}_8$	-105.23	-70.53	2.21	95.21	2.33	-13.86	1.96	**0.53**	**1.12**
$\mathcal{M}_9$	-491.97	-434.14	12.13	549.91	14.60	-37.72	11.43	0.45	0.97
$\mathcal{M}_{10}$	**-1690**	**-1650**	2.89	126.85	2.95	-50.20	2.87	0.49	1.06

Table 3: In-sample performance for each model $\mathcal{M}_\ell$, $\ell = 1, \ldots, 10$. Mean values over 60 days of the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Forecasting Error (RMSFE), Log-Score value (LS), Continuous Ranked Probability Score (CRPS), Directional Accuracy (DA), and Directional Value (DV).

model is $\mathcal{M}_{10}$ – that is the SARIMA$(2,0,0)(24,0,0)_5$ with GARCH(2,2) errors, followed by $\mathcal{M}_9$ – that is a SARIMA$(2,0,2)(24,0,0)_5$ with GARCH(2,2) errors. In conclusion, the diagnostics favor models with a seasonal component, more lags than models $\mathcal{M}_4$ to $\mathcal{M}_7$, and heteroscedastic effects. In the next section, we show that out-of-sample forecasting performance differs from in-sample performance and varies over time, which naturally calls for either sequential selection of the best model or dynamic combination techniques.

4.3 Prediction and combination

Panel (a) of Figure 2 shows the number of visits for the statistical area of Shanghai City from 01-Dec-2020 to 17-Jan-2022 (black solid line). In the same picture, the red dashed line shows the predicted daily number of visits (in Millions) from 18-Jan-2021 to 18-Mar-2022. The top plot shows forecasts for the model with the best in-sample and out-of-sample performance, that is, SARIMA$(2,0,0)(24,0,0)_5$-GARCH(2,2) ($\mathcal{M}_{10}$). The bottom plot in Panel (a) shows the forecasts for the BMA-MSFE model combination. When integrating predicted values into a city management process, careful consideration of various scenarios and acknowledgment of prediction uncertainty are paramount. Recognizing the dynamic nature of visits, it is imperative to address potential variations in anticipated outcomes. To address this, our approach goes beyond mere point prediction based on a single model. We propose two strategies to deal with model uncertainty. The first one is based on using density prediction, where the density incorporates the uncertainty around the point prediction. The second strategy is based on dynamically combining predictions from

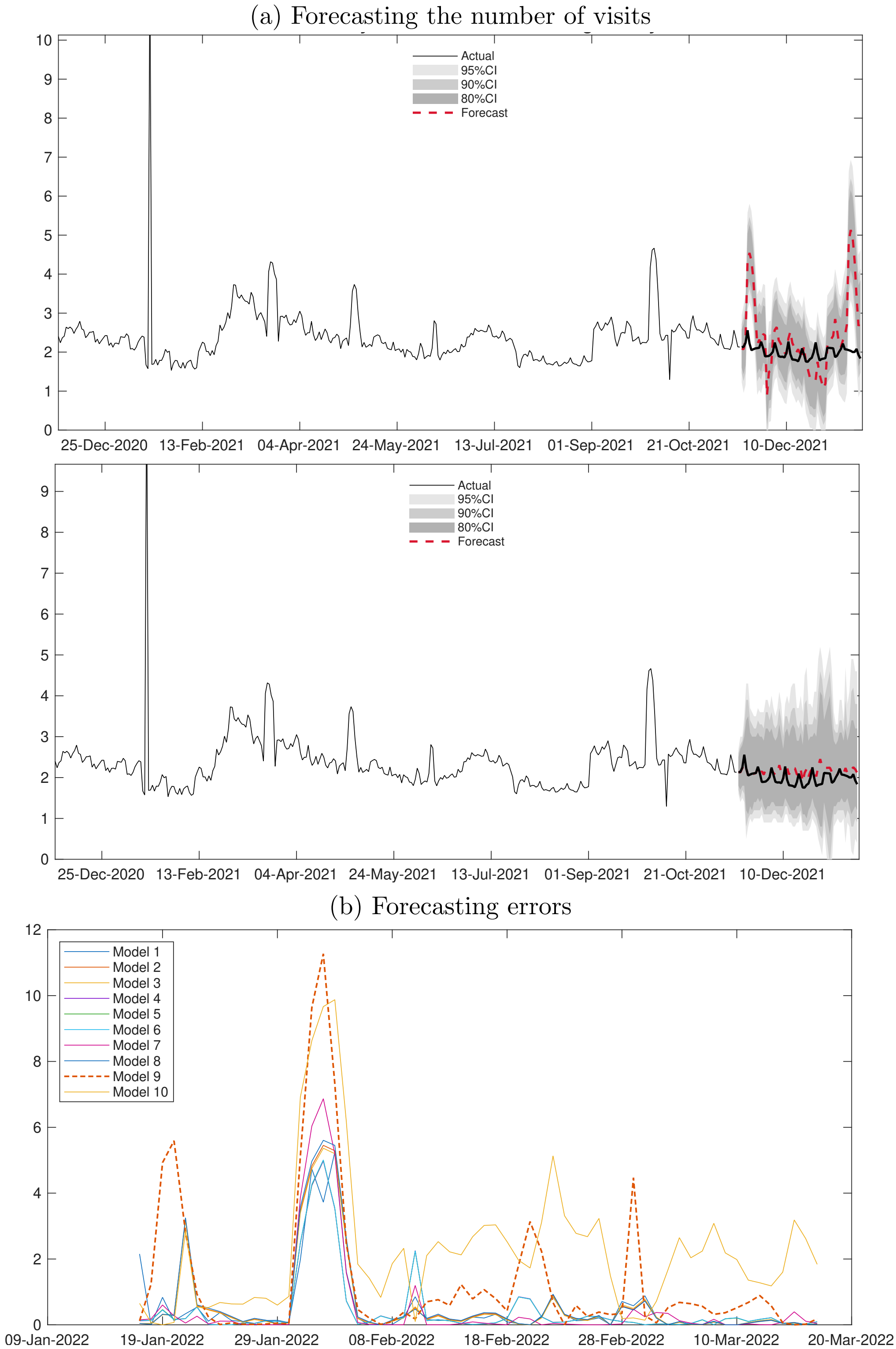

Figure 2: Panel (a): Forecasted visits (dashed red) using the best model $\mathcal{M}_{10}$ (top) and the BMA-MSFE combination (bottom). Actual number of visits from 01-Dec-2020 to 17-Jan-2022 in the Shanghai City area (black solid), and forecast confidence intervals (CI) at the 80%, 90%, and 95% levels (gray areas). Panel (b): Mean Square Forecasting Errors (top), $\text{MSFE}_{\ell,t}$ for each model $\mathcal{M}_\ell$, $\ell = 1, \dots, 10$.

different models, thus accounting for both possible model specifications and temporal variation in the data-generating process.

Regarding forecast uncertainty, we provide a more comprehensive depiction by including confidence intervals at different levels. For each model $\mathcal{M}_j$, $F_{j,s+1|s}(y)$ is the cumulative distribution of the number of visits at time $s+1$, forecasted at time s associated with the predictive density $f_{j,s+1|s}(y|\mathcal{M}_j)$. For each model it is a normal distribution with mean and variance given by conditional mean $\mu_{j,s+1|s}$ and conditional variance $\sigma^2_{j,s+1|s}$ of the SARIMAs and SARIMA-GARCHs, whereas for the BMAs is a mixture of normal distributions, $F_{s+1|s}(y) = w_{1,s}F_{1,s+1|s}(y) + \ldots + w_{J,s}F_{J,s+1|s}(y)$. The conditional mean is the point forecast and the forecast confidence intervals (CIs) at time s are the number of visits $\underline{y}_{j,s+1|s}$, and $\bar{y}_{j,s+1|s}$, such that the probability that $\underline{y}_{j,s+1|s} \leq y_{s+1} \leq \bar{y}_{j,s+1|s}$ is equal to $\alpha\%$ with $\alpha = 80\%, 90\%, 95\%$, where $\underline{y}_{j,s+1|s}$ and $\bar{y}_{j,s+1|s}$ are computed using the inverse cumulative distribution. For normal distribution $\underline{y}_{j,s+1|s}$ and $\bar{y}_{j,s+1|s}$ can be evaluated using standard normal quantile function, while for the mixture $\underline{y}_{j,s+1|s}$ is the smallest y such that $F_{j,s+1|s}(y) \geq \alpha/2$ and $\bar{y}_{j,s+1|s}$ is the smallest y such that $F_{j,s+1|s}(y) \geq 1-\alpha/2$, which can be computed numerically using bisection or a fine grid. These intervals provide insights into expected values and indicate the potential deviation of future observations from the point forecast, contributing to a nuanced understanding of the model's predictive uncertainty. The grey areas represent the uncertainty around the point prediction, measured by CIs at 80% (dark grey), 90% (grey), and 95% (light grey). If future prediction errors are stable, unbiased, and normally distributed, approximately 80% of future outcomes are expected to fall within the 80% uncertainty bands shown in dark grey, with the remaining outcomes falling symmetrically above and below the bands. In formulas, the coverage probability is computed as $\tau^{-1}\sum_{s=t-\tau+1}^{t} \mathbb{I}(\mu_{j,s+1|s} \in (\underline{y}_{j,s+1|s}, \bar{y}_{j,s+1|s}))$. This comprehensive perspective ensures decision-makers a more robust and informed response to potential scenarios for a better city management. The point and density forecasts obtained with the ten models are in Fig. B.15 of the Appendix.

Regarding model uncertainty, we evaluated the out-of-sample predictive performance of each model described in Section 3. Point-forecast measures such as MAE, MAPE, RMSFE, and density-forecast measures such as Log-Score (LS), and the Continuous Ranked Probability Score (CRPS), indicate that models without the GARCH effects outperform the ones with GARCH effects (see Tab. 4). The results reported in the top plot of Fig. 3 show the greater predictive ability (larger weights on average) of Model 7, the SARIMA$(2,0,0)(16,0,0)_5$

Model	AIC	BIC	MAE	MAPE	RMSFE	LS	CRPS	DA	DV
$\mathcal{M}_1$	76.83	112.96	0.44	22.33	0.60	-1.39	0.31	0.52	0.97
$\mathcal{M}_2$	-188.91	-144.76	0.36	18.14	0.49	-1.34	0.30	0.42	0.79
$\mathcal{M}_3$	-190.80	-150.66	0.37	18.72	0.50	-1.35	0.31	0.42	0.79
$\mathcal{M}_4$	52.81	80.90	0.67	34.97	0.77	-2.15	0.57	0.52	0.98
$\mathcal{M}_5$	53.49	81.58	0.67	34.73	0.76	-2.12	0.55	0.52	0.98
$\mathcal{M}_6$	54.38	78.46	0.67	34.84	0.77	-2.23	**0.65**	0.5	0.95
$\mathcal{M}_7$	31.27	59.36	**0.33**	**17.43**	**0.44**	-1.41	0.25	0.5	0.94
$\mathcal{M}_8$	-191.30	-155.18	0.35	17.87	0.48	**-1.34**	0.27	0.43	0.83
$\mathcal{M}_9$	-1200	-1140	0.59	29.85	0.66	-1.37	0.43	**0.57**	**1.08**
$\mathcal{M}_{10}$	**-2480**	**-2440**	0.67	33.22	1.01	-1.56	0.45	0.43	0.82

Table 4: Out-of-sample forecast accuracy evaluated sequentially for each model $\mathcal{M}_\ell$, $\ell = 1, \dots, 10$. Mean values over 60 days, respectively of the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Forecasting Error (RMSFE), Log-Score value (LS), Continuous Ranked Probability Score (CRPS), Directional Accuracy (DA), and Directional Value (DV).

without the MA component, followed by Model 8, the SARIMA$(2,0,0)(24,0,0)_5$ without restrictions and Model 2, the SARIMA$(2,0,2)(24,0,0)_5$ with restrictions (see Fig. B.15 of the Appendix). Nevertheless, when model parsimony is taken into consideration, as indicated by the AIC and BIC performance measures, GARCH models $\mathcal{M}_9$ and $\mathcal{M}_{10}$ exhibit better performance. Compared to the other models, they are able to capture the reduced variability in that last part of the sample (see last row in Fig. B.15 of the Appendix) and the decrease in the flows as confirmed by the directional accuracy measures, DA and DV, favor models in the GARCH family.

We apply BMA and use time-varying combination weights based on rolling estimates of the model's forecasting performance measures. Panel (b) of Fig. 2 shows the model MSFE over time, and Fig. 3 shows BMA weights based on the MSFEs. There are periods, e.g., the beginning of January and February, when Models 1 and 4 outperformed, and other periods when Models 6 and 7 performed better. Models without GARCH effects outperform GARCH models, except during periods of higher series variability. The time variations in predictive ability are confirmed by other measures, which induce time variations in the BMA weights (see Figs. C.18-C.19 in the Appendix). The time-varying nature of the problem naturally calls for either the sequential selection of the best model or the combination of models. We apply the different BMA procedures presented in the previous section and obtain the out-of-sample point and density predictions reported in Figs. C.16-C.17 of the Appendix. The BMA specifications exhibit different levels of

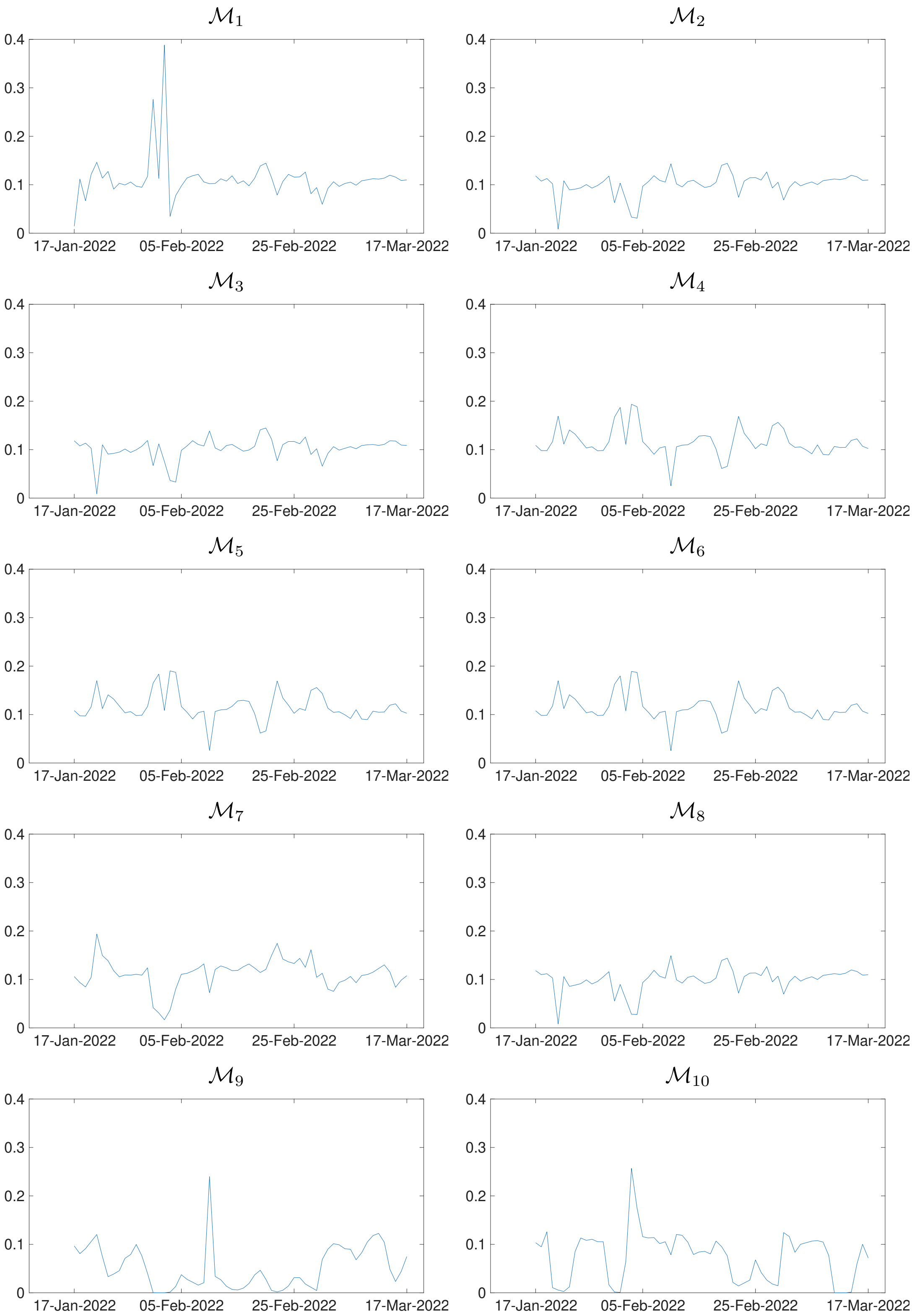

Figure 3: Combination weights $w_{\ell,t}^{MSFE}$ for each model $\mathcal{M}_\ell$, $\ell = 1, \ldots, 10$ for $t = 1, \ldots, 60$, that is the period from 17^{th} January 2022 to 17^{th} March 2022.

forecasting errors (see Panel (a) of Fig. 4). In our results, BMA based on MAE, MAPE, and RMSFE (top left) yields nearly identical errors; similarly, FMA based on AIC, BIC, and HQ (bottom right) yields similar errors. BMA-LS and BMA-CRPS exhibit different performance, with BMA-CRPS usually performing better (bottom right). The BMA based on DA, performs worse than its exponential version, that is, the BMA-EHR (middle-right). On average, the point-forecast performance of BMA-MSFE is superior to that of the other weighting strategies. In particular, it also outperforms the best individual model, as shown by comparing the black lines in the left and right plots of Panel (b) in Figure 4. This indicates that the combination based on MSFE weights can exploit the information across the different models more effectively than relying on a single specification. The out-of-sample point and density forecast of the BMA-MSFE is reported for comparison purposes in the second plot of Panel (a) of Fig. 4. The best model combination provides a better fit to seasonality and variability than the best individual model (compare with the first plot in the same panel). Regarding the forecast CIs coverage, there are three groups of models. The predictive intervals of the models with the best density forecasts (large CRPS), $\mathcal{M}_4$, $\mathcal{M}_5$, and $\mathcal{M}_6$, are wider and include all realizations, indicating over-coverage. Models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, $\mathcal{M}_7$, and $\mathcal{M}_8$ have better coverage and point forecast performances, but the lowest density forecast performances (e.g, low CRPS). The third group includes the SARIMA-GARCHs $\mathcal{M}_9$ and $\mathcal{M}_{10}$, which have narrower intervals, good coverage, and intermediate density- and point-forecasting abilities. The combined forecasts provide a compromise among the three model classes and better capture the dynamics and variability of the tourist flows.

Among the various urban dimensions affected by tourism, waste management is among the most significantly affected, as increased visitor flows often strain existing collection and disposal infrastructure [e.g., see Xiao et al., 2020, Yuxi et al., 2023]. Thus, one can expect that the economic implications of a forecast error differ depending on whether the cleaning and waste management system is operating under high or low pressure. Indeed, large forecasting errors tend to occur during holiday periods when visit volumes spike. Cost-based weighting schemes reduce the influence of these large errors and tend to favor models that perform more reliably during non-holiday periods. To exploit the economic relevance of forecasting error, we use the average daily and per-visit cleaning costs. We assumed an average cost of 178 (in 10,000 Yuan/Day) and 2.3 million visits per day during the non-holiday periods and 440 (in 10,000 Yuan/Day) during holidays, with 12 million

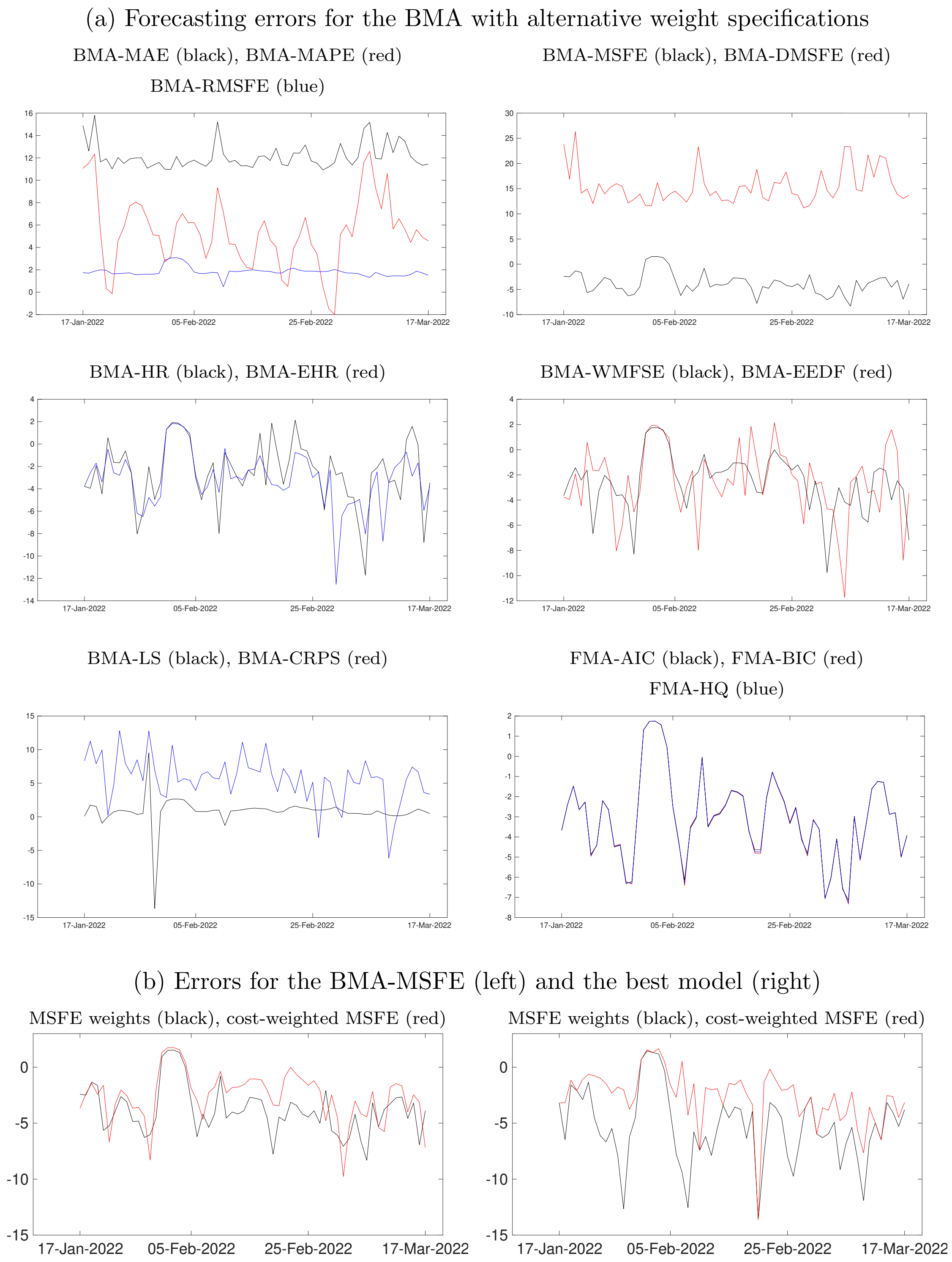

Figure 4: Forecasting errors (in logarithms). Panel (a) Alternative BMA specifications. Panel (b) BMA with MSFE weights (left) and the best model (right). Model weights $w_{\ell,t}$ are based on MSFEs (black) and $\tilde{w}_{\ell,t}$ are based on cost–weighted MSFEs (red), $\ell = 1,\dots,10$, $t = 1,\dots,60$, that is the period from 17^{th} January 2022 to 17^{th} March 2022. Tourism costs are the ones incurred by the municipality of Shanghai for cleaning services.

trips (Culture and Bureau [2024]). The relative cost per day and visit, c_s, is equal to 0.7 during non-holiday periods and 0.3 during holiday periods. With these weights, the WMSFE is computed for each model as in equation 18.

The two plots in Panel (b) of Figure 4 show the errors (in logarithms) of forecasts based on equally weighted errors (black line) and on cost-weighted errors (red line). The forecasting errors for prediction averaging and the best model are shown in the top and bottom plots, respectively. In the Appendix, Fig. C.20 shows the MSFE-based and cost-based weights for each model. Comparing the results in the two plots shows that a model combination strategy yields smaller forecasting errors than a prediction strategy that selects the best model. In addition, comparing the red and black lines shows that using a cost-weighted combination or selection strategy can reduce forecasting errors. Finally, among measures that incorporate an economic value component, BMA-WMFSE performs better than BMA specifications based on more standard default measures, such as EEDF (see the middle–right plot of Panel (a)). This result suggests that the proposed cost-weighted measure is more effective at capturing the relevant economic trade-offs embedded in the forecasting exercise and leads to improved forecast-combination performance.

5 Conclusion

Using Big Data and predictive models is crucial when making timely policy decisions for tourism management. Usually, when predicting tourist flows, multiple models are applied, which adds uncertainty and poses challenges for their use in supporting policy-making. We propose a flexible yet straightforward procedure based on forecast combination models to address model uncertainty in decision-making. First, using Telcom Big Data to forecast daily visits to Shanghai, we show that model uncertainty arises when using prediction models from classes commonly used in practice, such as ARIMA, ARIMA with seasonal and GARCH effects.

The empirical analysis identifies SARIMA with seasonal dynamics and heteroskedastic errors (GARCH) as the most suitable model class for forecasting tourism flows. This model class performs consistently well both in-sample and out-of-sample, demonstrating robustness in estimation and forecasting. However, forecasting accuracy varies over time, suggesting that reliance on a single model may not always be optimal. For this reason, forecast combination strategies prove particularly valuable. By integrating information from multiple models rather than relying exclusively on a single best specification, combined forecasts achieve lower prediction errors and greater stability over time. This improvement becomes even more pronounced when cost-weighted combination methods are employed, as they allow the forecasting framework to place greater emphasis on models that perform better under specific conditions. These findings underscore the value of flexible and adaptive forecasting frameworks in tourism analysis, where uncertainty and rapid changes in demand patterns are intrinsic features of the sector.

From a policy perspective, these findings carry important implications for tourism destinations. The results provide practical evidence that more accurate forecasting tools can significantly support public decision-making. Improved predictions of tourism flows allow policymakers and destination management organizations to anticipate demand fluctuations, mitigate congestion during peak periods, and better plan infrastructure, mobility systems, and visitor services. More reliable forecasts also help optimize the allocation of public resources, enabling authorities to design and implement policies that balance tourism growth with sustainability objectives and the well-being of local communities. Moreover, the results highlight the strategic importance of integrating diverse data sources and analytical approaches into tourism governance. The increasing availability of Big Data—particularly high-frequency mobility data—creates opportunities to expand operational management tools and develop more flexible analytical frameworks. In this context, the ability to combine different modelling approaches becomes particularly valuable, as it allows policymakers to continuously incorporate new data streams and methodological advances. Such an adaptive analytical environment supports more responsive, evidence-based tourism management and strengthens destinations' capacity to address complex challenges related to demand volatility, resource management, and sustainable development.

Despite these encouraging results, some limitations remain. The modelling framework relies on traditional time-series specifications and could be further improved by exploring hybrid approaches that combine statistical and machine learning methods. Moreover, the analysis focuses on a specific dataset and context; future research should test the proposed models across different destinations and time horizons to assess their generalizability and to support the development of more adaptive forecasting tools. Future research should also focus on improving data availability at the individual point-of-interest level, with the aim of testing the applicability of the models developed in this study at a finer spatial scale and analysing tourism dynamics across individual attractions within the same city. Such developments could provide policymakers with more detailed information to support the planning and management of tourism services.

References

Aastveit, K. A., Cross, J. L., and van Dijk, H. K. (2023). Quantifying time-varying forecast uncertainty and risk for the real price of oil. *Journal of Business & Economic Statistics*, 41(2):523–537.

Araya, H. T., Aduda, J., and Berhane, T. (2024). A hybrid GARCH and deep learning method for volatility prediction. *Journal of Applied Mathematics*, 2024(1):6305525.

Ardito, L., Cerchione, R., Del Vecchio, P., and Raguseo, E. (2019). Big data in smart tourism: challenges, issues and opportunities. *Current Issues in Tourism*, 22(15):1805–1809.

Armstrong, J. S., Green, K. C., and Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, 68(8):1717–1731.

Bassetti, F., Casarin, R., and Ravazzolo, F. (2018). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*, 113(522):675–685.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Bi, J.-W., Han, T.-Y., and Li, H. (2022). International tourism demand forecasting with machine learning models: The power of the number of lagged inputs. *Tourism Economics*, 28(3):621–645.

Bi, J.-W., Liu, Y., and Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research*, 83:102923.

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2):213–232.

Bisht, A., Kumar, H., Nag, A., and Kumar, V. (2025). Managing seasonality and overtourism for sustainable tourism development: Challenges and strategies. *International Journal of Research and Review*, 12(3):443–453.

Blázquez-Salom, M., Cladera, M., and Sard, M. (2023). Identifying the sustainability indicators of overtourism and undertourism in majorca. *Journal of Sustainable Tourism*, 31(7):1694–1718.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

Box, G. and Jenkins, G. (1976). Analysis: Forecasting and control. *San Francisco*, 10.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Camatti, N., Carallo, G., Casarin, R., and Feng, X. (2024). Measuring daily tourism mobility spillover at the intra-metropolis level with mobile positioning data. *Regional Studies & Regional Science*, 11(1):701–723.

Chan, F., Lim, C., and McAleer, M. (2005). Modelling multivariate international tourism demand and volatility. *Tourism Management*, 26(3):459–471.

Chen, Y., Hu, T., and Law, R. (2025). Tourism demand forecasting: A novel multi-channel imaging model. *Journal of Tourism Futures*.

Chu, F.-L. (2009). Forecasting tourism demand with arma-based methods. *Tourism Management*, 30(5):740–751.

Claveria, O., Monte, E., and Torra, S. (2015). Tourism demand forecasting with neural network models: Different ways of treating information. *International Journal of Tourism Research*, 17(5):492–500.

Clements, M. and Hendry, D. (2004). Pooling of forecasts. *Econometrics Journal*, 7(1):1–31.

Coshall, J. T. and Charlesworth, R. (2011). A management orientated approach to combination forecasting of tourism demand. *Tourism Management*, 32(4):759–769.

Costantini, M., Cuaresma, J. C., and Hlouskova, J. (2016). Forecasting errors, directional accuracy and profitability of currency trading: The case of eur/usd exchange rate. *Journal of Forecasting*, 35(7):652–668.

Crespo Cuaresma, J. and Feldkircher, M. (2013). Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics*, 28(4):720–741.

Culture, S. and Bureau, T. (2024). Shanghai monthly tourism data report. Shanghai, China. https://whlyj.sh.gov.cn/tjzl/20250527/4468038bff1a454497366fcc99e91941.html.

Divino, J. A. and McAleer, M. (2010). Modelling and forecasting daily international mass tourism to Peru. *Tourism Management*, 31(6):846–854.

Duan, J., Xie, C., and Morrison, A. M. (2022). Tourism crises and impacts on destinations: A systematic review of the tourism and hospitality literature. *Journal of Hospitality & Tourism Research*, 46(4):667–695.

Fan, D. X., Uzunboylu, N., Efthymiou, L., Thrassou, A., and Vrontis, D. (2025). Innovation and value creation amidst global challenges and crises in tourism and hospitality industry. In *Global Challenges and Uncertainty in Tourism and Hospitality, Volume II: Marketing Advancements and Value Creation*, pages 1–25. Springer.

Fawcett, N., Kapetanios, G., Mitchell, J., and Price, S. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188:150–165.

Fernández, C., Ley, E., and Steel, M. F. J. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16(5):563–576.

Genest, C. and Schervish, M. J. (1985). Modeling expert judgments for Bayesian updating. *The Annals of Statistics*, pages 1198–1212.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.

Ghalehkhondabi, I., Ardjmand, E., Young, W. A., and Weckman, G. R. (2019). A review of demand forecasting models and methodological developments within tourism and passenger transportation industry. *Journal of Tourism Futures*, 5(1):75–93.

Gil-Alana, L. A., Gracia, F. P. D., and Cunado, J. (2004). Seasonal fractional integration in the Spanish tourism quarterly time series. *Journal of Travel Research*, 42(4):408–414.

Gneiting, T. and Ranjan, R. (2011). Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules. *Journal of Business and Economic Statistics*, 29:411–422.

Gneiting, T. and Ranjan, R. (2013). Combining Predictive Distributions. *Electronic Journal of Statistics*, 7:1747–1782.

Guizzardi, A., Pons, F. M. E., Angelini, G., and Ranieri, E. (2021). Big data from dynamic pricing: A smart approach to tourism demand forecasting. *International Journal of Forecasting*, 37(3):1049–1060.

Gunter, U. and Önder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of Tourism Research*, 61:199–212.

Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., and Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, 63:112–127.

He, M. and Qian, X. (2025). Forecasting tourist arrivals using stl-xgboost method. *Tourism Economics*, 32(2):408–436.

Hewapathirana, I. U. (2025). Advancing tourism demand forecasting in Sri Lanka: evaluating the performance of machine learning models and the impact of social media data integration. *Journal of Tourism Futures*, 11(2):261–285.

Höpken, W., Eberle, T., Fuchs, M., and Lexhagen, M. (2021). Improving tourist arrival prediction: a big data and artificial neural network approach. *Journal of Travel Research*, 60(5):998–1017.

Huang, X., Zhang, L., and Ding, Y. (2017). The baidu index: Uses in predicting tourism flows–a case study of the forbidden city. *Tourism Management*, 58:301–306.

Jakovlev, Z. and Kitanov, V. (2024). The importance of the planning function in the business financial decision-making process in tourism enterprises. *International Journal of Economics, Management and Tourism*, 4(2):109–116.

Johnson, M. C. and West, M. (2025). Bayesian predictive synthesis with outcome-dependent pools. *Statistical Science*, 40(1):109–127.

Kandt, J. and Batty, M. (2021). Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, 109:102992.

Koc, E. and Altinay, G. (2007). An analysis of seasonality in monthly per-person tourist spending in Turkish inbound tourism from a market segmentation perspective. *Tourism Management*, 28(1):227–237.

LeSage, J. P. and Parent, O. (2007). Bayesian model averaging for spatial econometric models. *Geographical Analysis*, 39(3):241–267.

Li, H., Goh, C., Hung, K., and Chen, J. L. (2018). Relative climate index and its effect on seasonal tourism demand. *Journal of Travel Research*, 57(2):178–192.

Liang, X., Li, X., Shu, L., Wang, X., and Luo, P. (2025). Tourism demand forecasting using graph neural network. *Current Issues in Tourism*, 28(6):982–1001.

Liang, Y.-H. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Computers & Industrial Engineering*, 74:111–119.

Lim, C. and McAleer, M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management*, 23(4):389–396.

Lin, W. L. (2024). Tourism and economic growth: Assessing the significance of sustainable competitiveness using a dynamic panel data approach. *International Journal of Tourism Research*, 26(5):e2760.

Lindley, D. V., Tversky, A., and Brown, R. V. (1979). On the reconciliation of probability assessments. *Journal of the Royal Statistical Society. Series A (General)*, 142(2):146–180.

Ling, S. and McAleer, M. (2003a). Asymptotic theory for a vector arma-garch model. *Econometric Theory*, 19(2):280–310.

Ling, S. and McAleer, M. (2003b). On adaptive estimation in nonstationary arma models with garch errors. *The Annals of Statistics*, 31(2):642–674.

Liu, X., Chen, Y.-n., Qiu, Z., and Chen, M.-r. (2019). Forecast of the tourist volume of Sanya city by XGBoost model and GM model. In *2019 International Conference on*

Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pages 166–173.

Liu, X., Liu, A., Chen, J. L., Li, G., and Song, H. (2025). Tourism demand forecasting in normal and crisis times: Combining bootstrap-aggregating and Bayesian approaches. *Journal of Hospitality & Tourism Research*, 50(3):323–340.

Llewellyn, M., Ross, G., and Ryan-Saha, J. (2023). Covid-era forecasting: Google trends and window and model averaging. *Annals of Tourism Research*, 103:103660.

Lu, J. and Chen, H. (2023). Dynamic evaluation and forecasting analysis of touristic ecological carrying capacity of forest parks in China. *Forests*, 15(1):38.

McAlinn, K. (2021). Mixed-frequency Bayesian predictive synthesis for economic nowcasting. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 70(5):1143–1163.

McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020). Multivariate Bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, 115(531):1092–1110.

McAlinn, K. and West, M. (2019). Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of econometrics*, 210(1):155–169.

Metulini, R. and Carpita, M. (2024). Modeling and forecasting traffic flows with mobile phone big data in flooding risk areas to support a data-driven decision making. *Annals of Operations Research*, 342(3):1629–1654.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29(1):46–75.

Mussoni, M. and Vici, L. (2025). Exogenous shocks and resilience in tourism. a regional science perspective. *Scienze Regionali*, 24(3):297–309.

Oh, C.-O. and Morzuch, B. J. (2005). Evaluating time-series models to forecast the demand for tourism in Singapore: Comparing within-sample and postsample results. *Journal of Travel Research*, 43(4):404–413.

Önder, I. and Gunter, U. (2016). Forecasting tourism demand with Google Trends for a major European city destination. *Tourism Analysis*, 21(2-3):203–220.

Pai, P.-F., Hung, K.-C., and Lin, K.-P. (2014). Tourism demand forecasting using novel hybrid system. *Expert Systems with Applications*, 41(8):3691–3702.

Pan, M., Liao, Z., Wang, Z., Ren, C., Xing, Z., and Li, W. (2025). Tourism forecasting: A dynamic spatiotemporal model. *Annals of Tourism Research*, 110:103871.

Peng, B., Song, H., and Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, 45:181–193.

Prayag, G., Jiang, Y., Chowdhury, M., Hossain, M. I., and Akter, N. (2024). Building dynamic capabilities and organizational resilience in tourism firms during covid-19: A staged approach. *Journal of Travel Research*, 63(3):713–740.

Punia, S. and Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems*, 258:109956.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191.

Ramos, V., Yamaka, W., Alorda, B., and Sriboonchitta, S. (2021). High-frequency forecasting from mobile devices' bigdata: An application to tourism destinations' crowdedness. *International Journal of Contemporary Hospitality Management*, 33(6):1977–2000.

Sejati, A., Putri, S., Tyas, W., Buchori, I., Handayani, W., Basuki, Y., Barbarossa, G., and Husna, I. (2025). Predicting urban carrying capacity to support sustainable tourism using gis. *Journal of Policy Research in Tourism, Leisure and Events*, 17(3):632–655.

Shen, S., Li, G., and Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, 38(1):72–89.

Song, H. and Li, G. (2008). Tourism demand modelling and forecasting—a review of recent research. *Tourism Management*, 29(2):203–220.

Song, H., Li, G., Witt, S. F., and Athanasopoulos, G. (2011). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27(3):855–869.

Song, H., Qiu, R. T., and 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. and Zhang, H. (2025). Tourism demand modelling and forecasting: A horizon 2050 paper. *Tourism Review*, 80(1):8–27.

Sun, S., Wei, Y., Tsui, K.-L., and Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, 70:1–10.

Sun, Y., Zhang, J., Li, X., and Wang, S. (2023). Forecasting tourism demand with a new time-varying forecast averaging approach. *Journal of Travel Research*, 62(2):305–323.

Tallman, E. and West, M. (2024). Bayesian predictive decision synthesis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(2):340–363.

Tsui, W. H. K. and Balli, F. (2017). International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure. *Tourism Economics*, 23(2):403–428.

Vanegas Sr, M. (2013). Co-integration and error correction estimation to forecast tourism in El Salvador. *Journal of Travel & Tourism Marketing*, 30(6):523–537.

Volchek, K., Liu, A., Song, H., and Buhalis, D. (2019). Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tourism Economics*, 25(3):425–447.

Wan, A. T. and Zhang, X. (2009). On the use of model averaging in tourism research. *Annals of Tourism Research*, 36(3):525–532.

Wan, S. K. and Song, H. (2018). Forecasting turning points in tourism growth. *Annals of Tourism Research*, 72:156–167.

Wang, X., Hyndman, R. J., Li, F., and Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547.

Wang, Y., Xie, T., and Jie, X. (2019). A mathematical analysis for the forecast research on tourism carrying capacity to promote the effective and sustainable development of tourism. *Discrete & Continuous Dynamical Systems-Series S*, 12.

Wei, Z. and Mukherjee, S. (2024). Analyzing and forecasting service demands using human mobility data: A two-stage predictive framework with decomposition and multivariate analysis. *Expert Systems with Applications*, 238:121698.

Wen, L., Liu, C., and Song, H. (2019). Forecasting tourism demand using search query data: A hybrid modelling approach. *Tourism Economics*, 25(3):309–329.

West, M. (1984). Bayesian aggregation. *Journal of the Royal Statistical Society: Series A (General)*, 147(4):600–607.

West, M. (1988). Modelling expert opinion (with discussion). *In Bayesian Statistics 3*, pages 493—508.

West, M. (1992). Modelling agent forecast distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 54(2):553–567.

Wong, K. K., Song, H., Witt, S. F., and Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, 28(4):1068–1078.

Wu, D. C., Zhong, S., Wu, J., and Song, H. (2025). Tourism and hospitality forecasting with big data: A systematic review of the literature. *Journal of Hospitality & Tourism Research*, 49(3):615–634.

Xiao, S., Dong, H., Geng, Y., Francisco, M.-J., Pan, H., and Wu, F. (2020). An overview of the municipal solid waste management modes and innovations in Shanghai, China. *Environmental Science and Pollution Research*, 27(24):29943–29953.

Yang, X., Pan, B., Evans, J. A., and Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46:386–397.

Yuxi, Z., Filimonau, V., Ling-en, W., and Linsheng, Z. (2023). The impact of tourism on municipal solid waste generation in China. *Journal of Cleaner Production*, 427:139255.

Zheng, W., Li, J., and Li, C. (2024). Tourism demand forecasting using complex network theory. *Asia Pacific Journal of Tourism Research*, 29(3):302–320.

Zhou, W. (2021). Prediction of urban and rural tourism economic forecast based on machine learning. *Scientific Programming*, 2021(1):4072499.

A Further model diagnostics

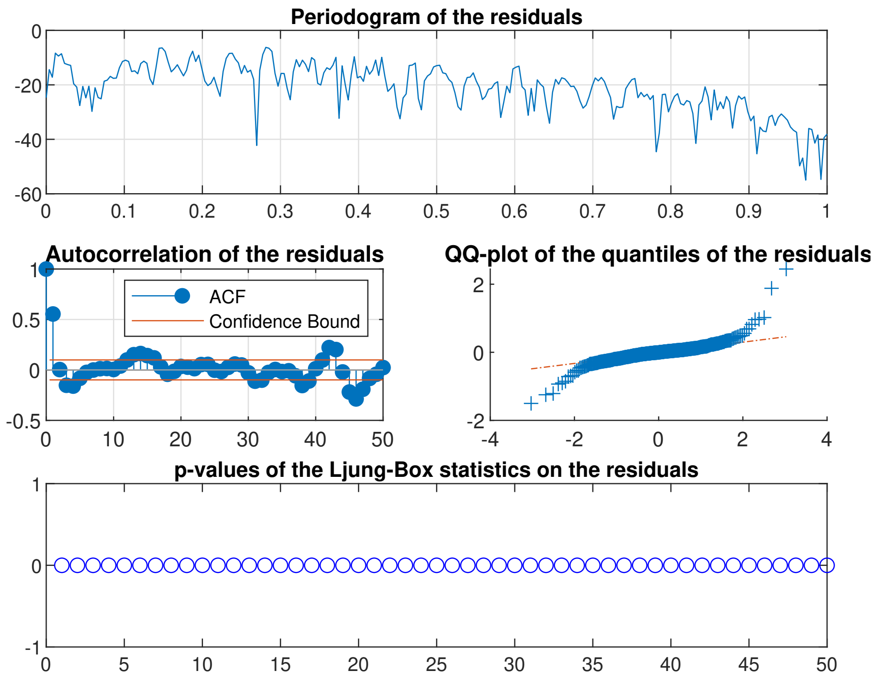

Figure A.5: Diagnostics for model $\mathcal{M}_1$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p-values of the Ljung–Box Q-test (lower panel).

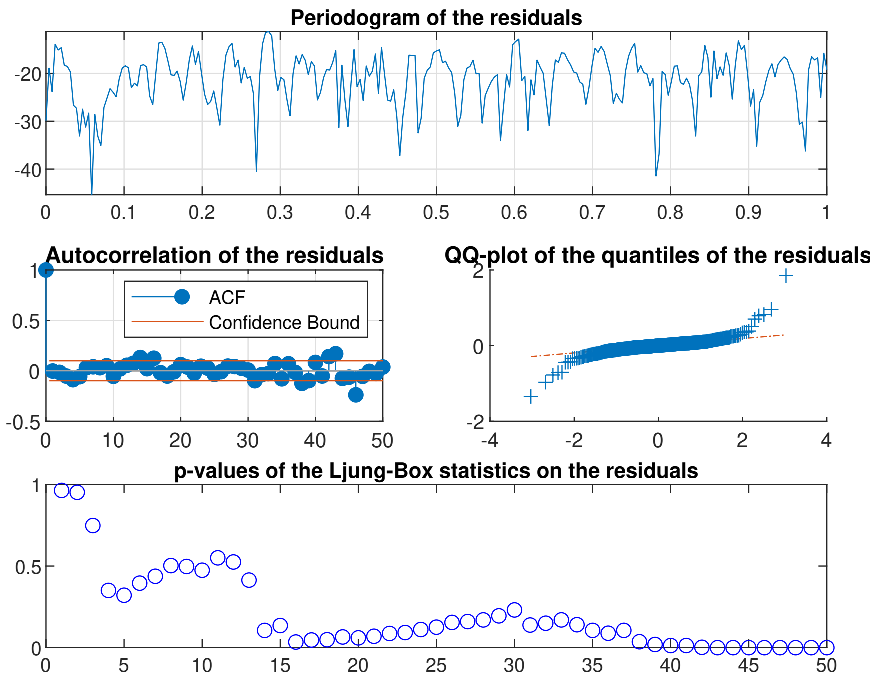

Figure A.6: Diagnostics for model $\mathcal{M}_2$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p-values of the Ljung–Box Q-test (lower panel).

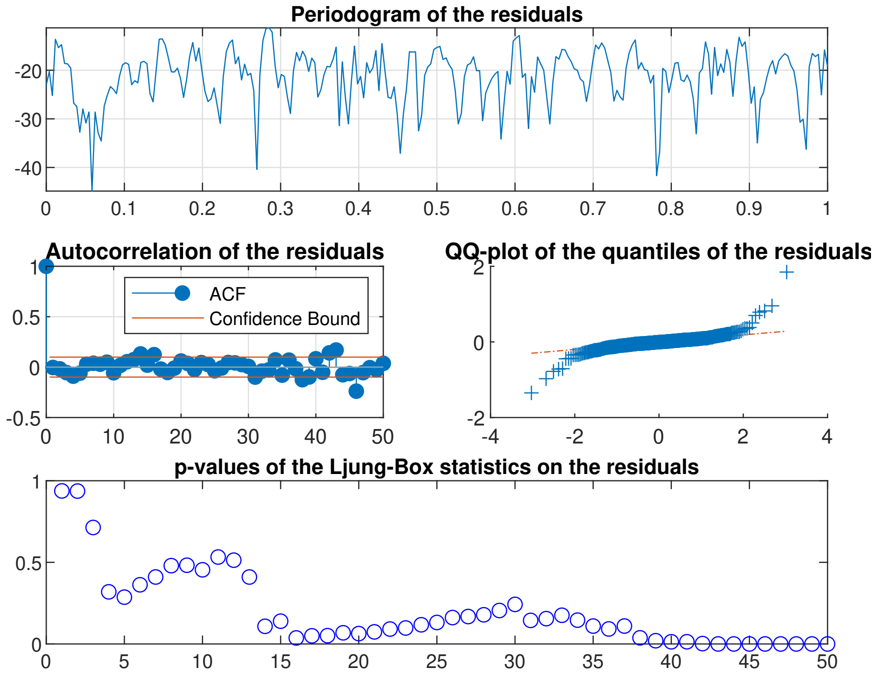

Figure A.7: Diagnostics for model $\mathcal{M}_3$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p–values of the Ljung–Box Q–test (lower panel).

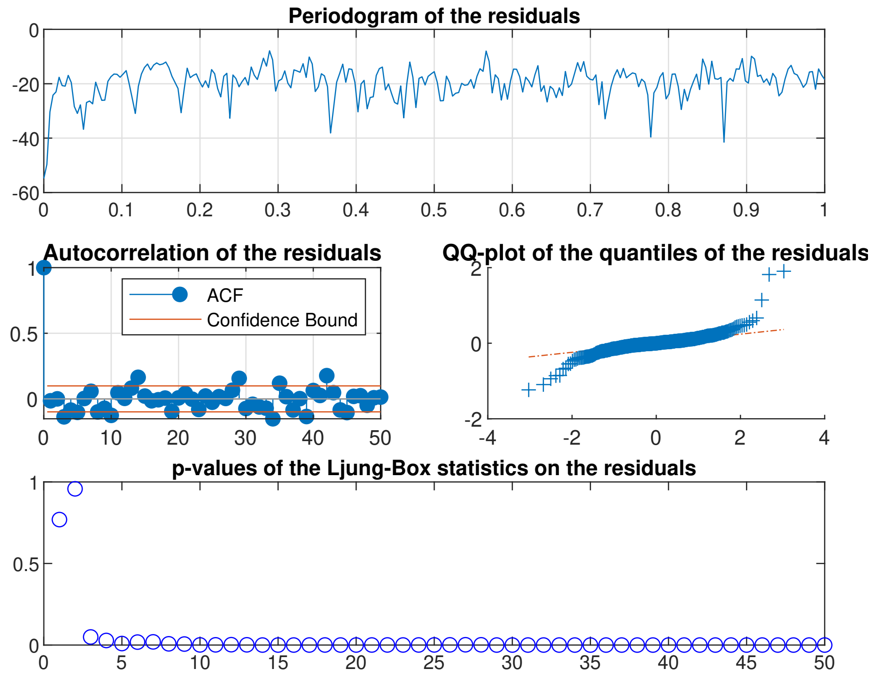

Figure A.8: Diagnostics for model $\mathcal{M}_4$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p–values of the Ljung–Box Q–test (lower panel).

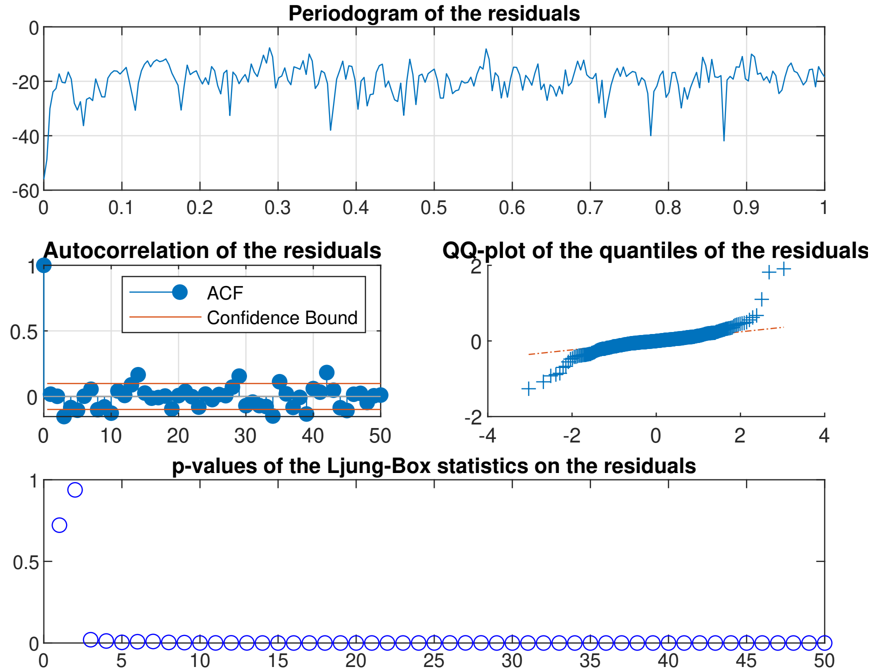

Figure A.9: Diagnostics for model $\mathcal{M}_5$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p-values of the Ljung–Box Q–test (lower panel).

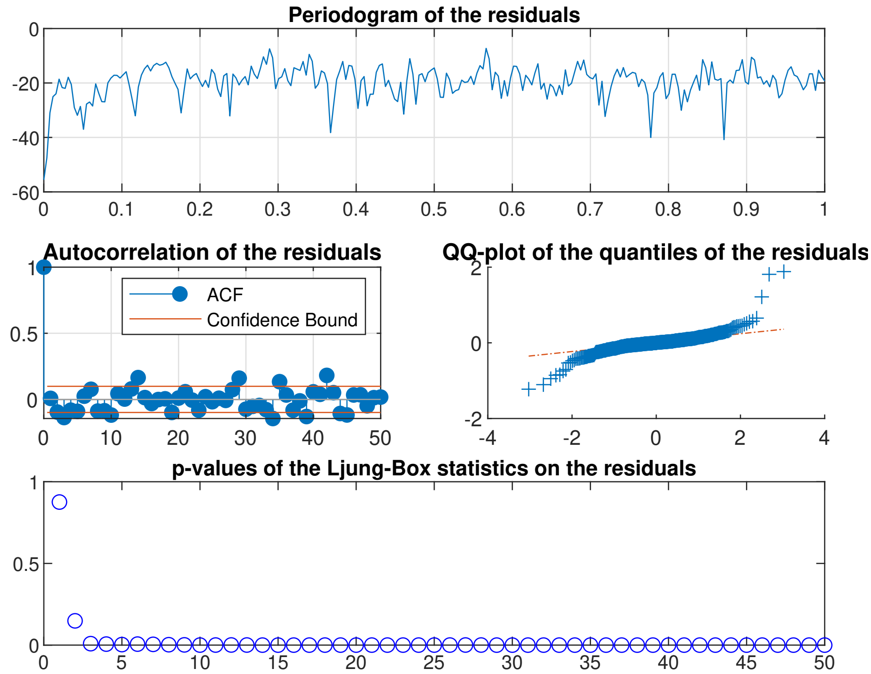

Figure A.10: Diagnostics for model $\mathcal{M}_6$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p-values of the Ljung–Box Q–test (lower panel).

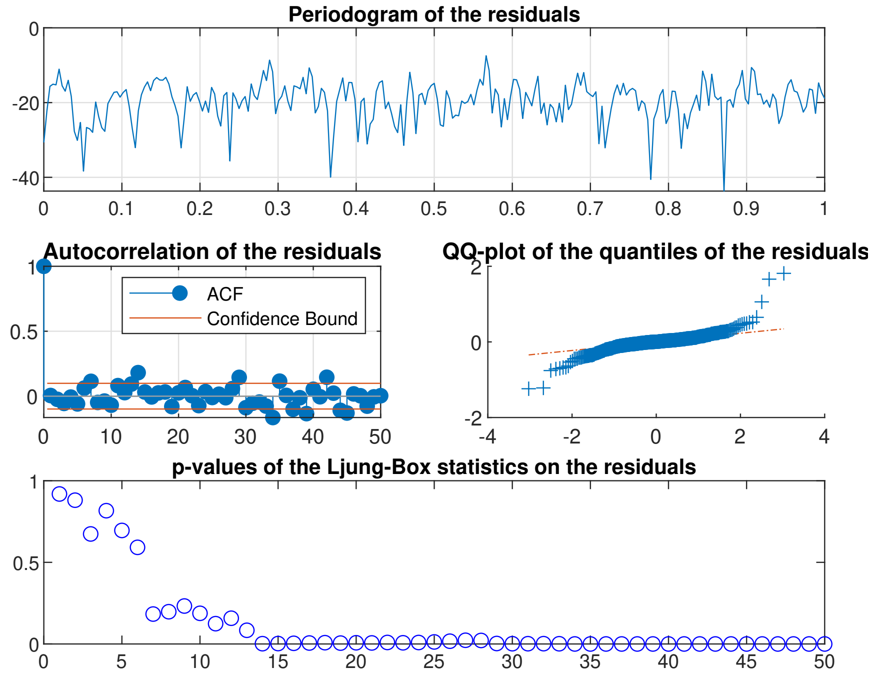

Figure A.11: Diagnostics for model $\mathcal{M}_7$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p–values of the Ljung–Box Q–test (lower panel).

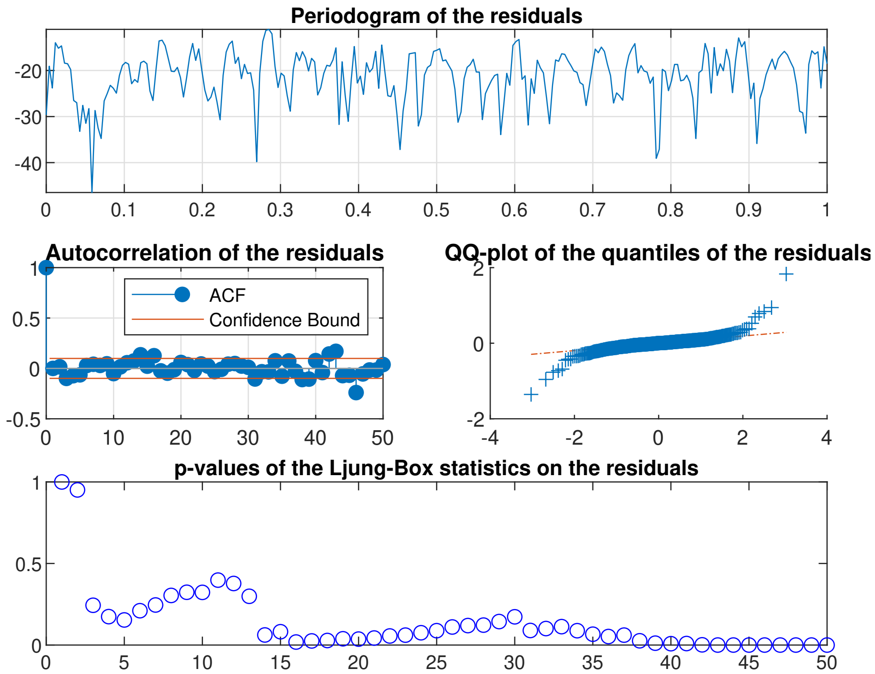

Figure A.12: Diagnostics for model $\mathcal{M}_8$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p–values of the Ljung–Box Q–test (lower panel).

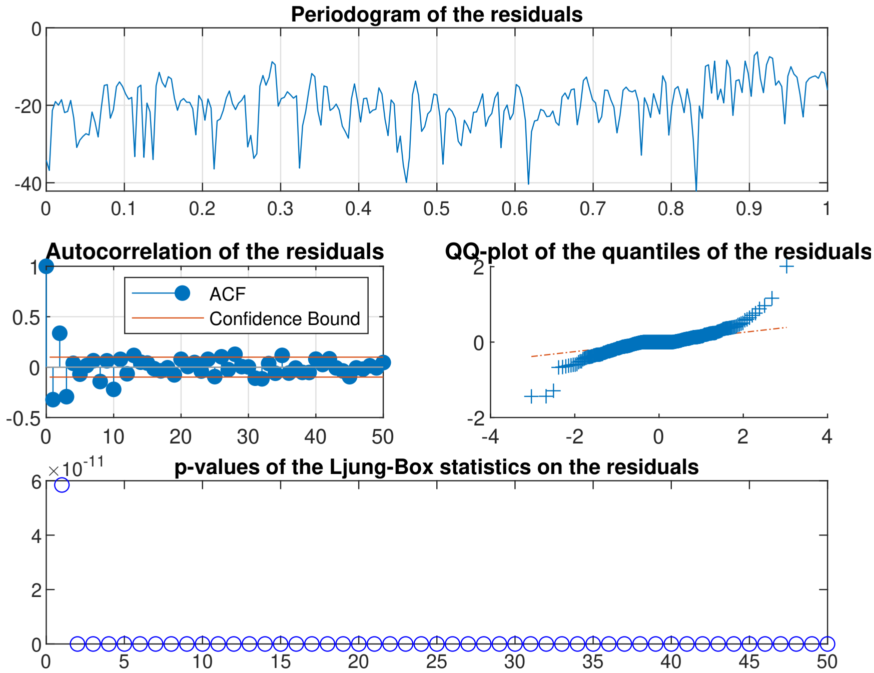

Figure A.13: Diagnostics for model $\mathcal{M}_9$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p–values of the Ljung–Box Q–test (lower panel).

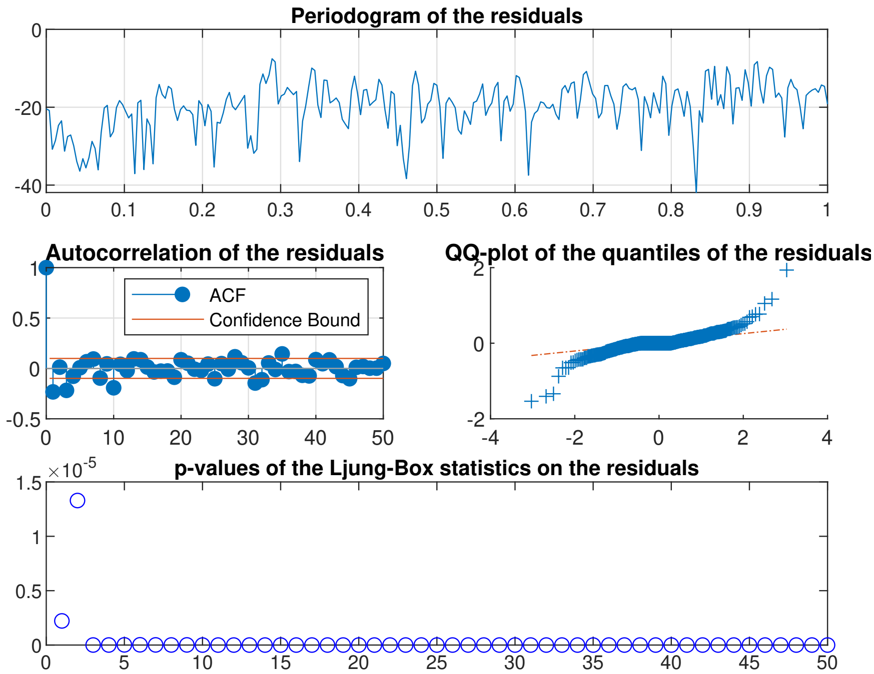

Figure A.14: Diagnostics for model $\mathcal{M}_{10}$: periodogram (upper panel), ACF (middle left panel), Q–Q plot (middle right panel), and p–values of the Ljung–Box Q–test (lower panel).

B Out–of–sample model forecasts

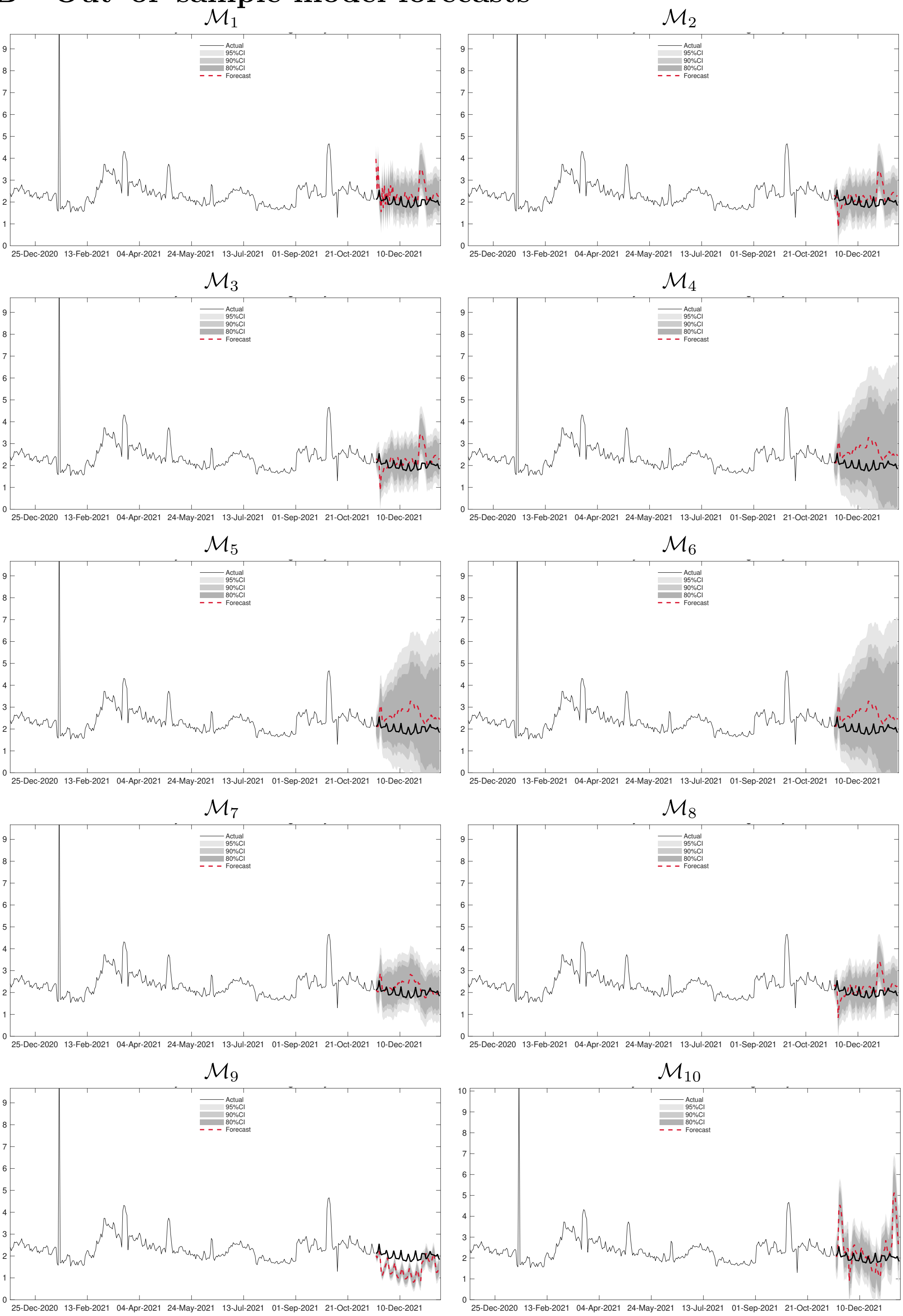

Figure B.15: Actual number of visits from 01-Dec-2020 to 17-Jan-2022 in Shanghai City area (black solid) and the forecasted visits (dashed red) using different models $\mathcal{M}_j$ $j = 1, \ldots, 10$. The gray areas represent confidence intervals (CI) at different levels.

C Out–of–sample BMA forecasts

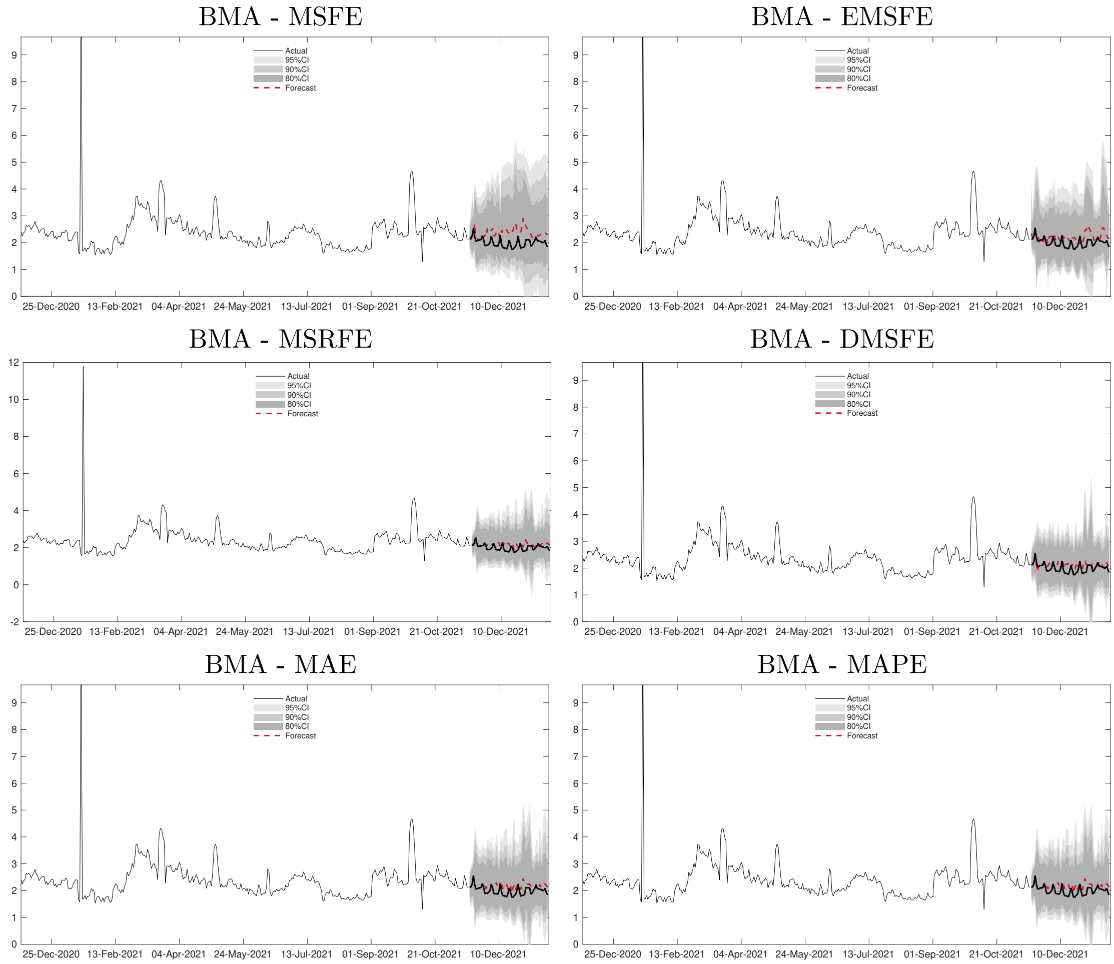

Figure C.16: Actual number of visits from 01-Dec-2020 to 17-Jan-2022 in Shanghai City area (black solid) and the forecasted visits (dashed red) using different BMA specifications. The gray areas represent confidence intervals (CI) at different levels.

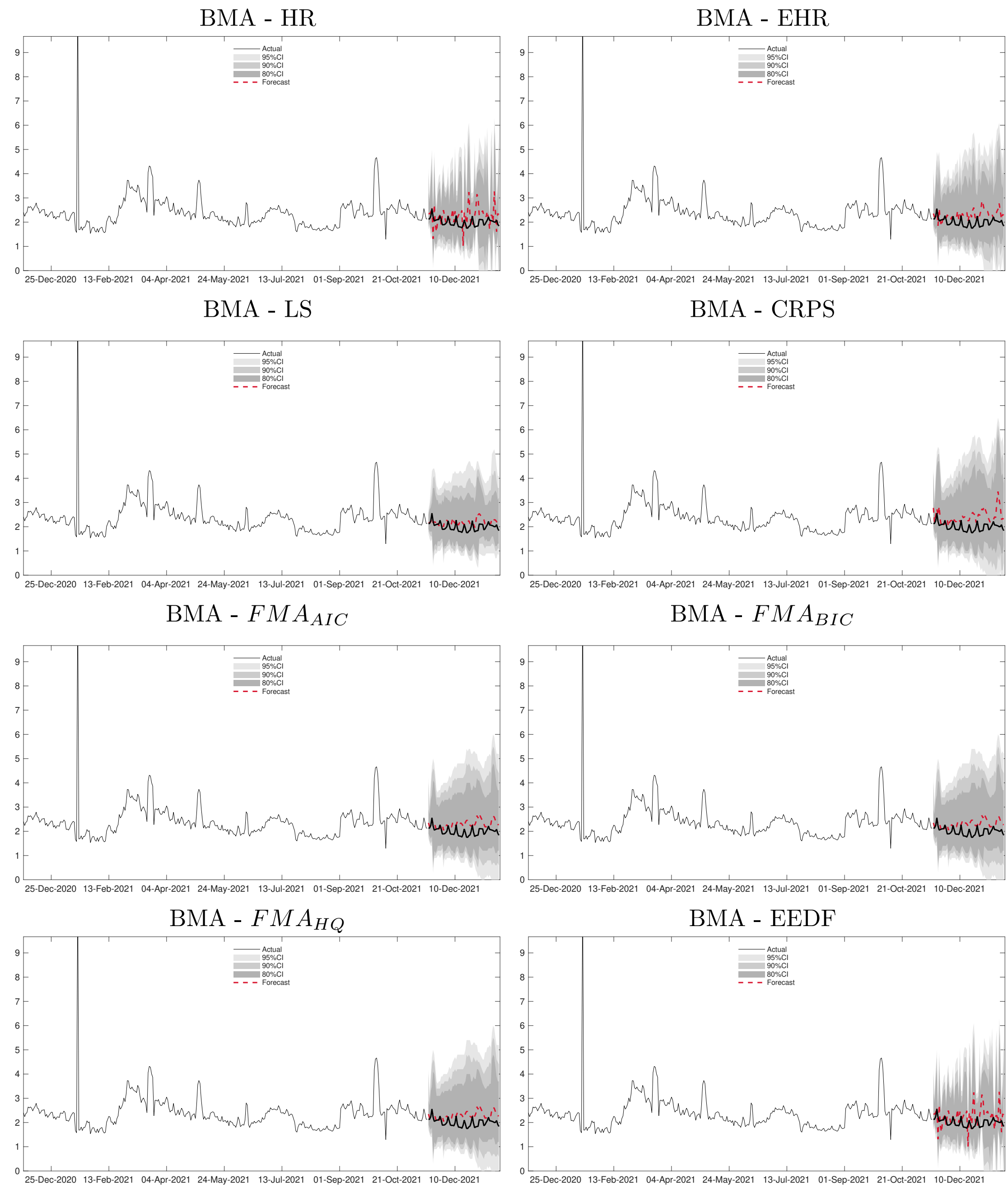

Figure C.17: Actual number of visits from 01-Dec-2020 to 17-Jan-2022 in Shanghai City area (black solid) and the forecasted visits (dashed red) using different BMA specifications. The gray areas represent confidence intervals (CI) at different levels.

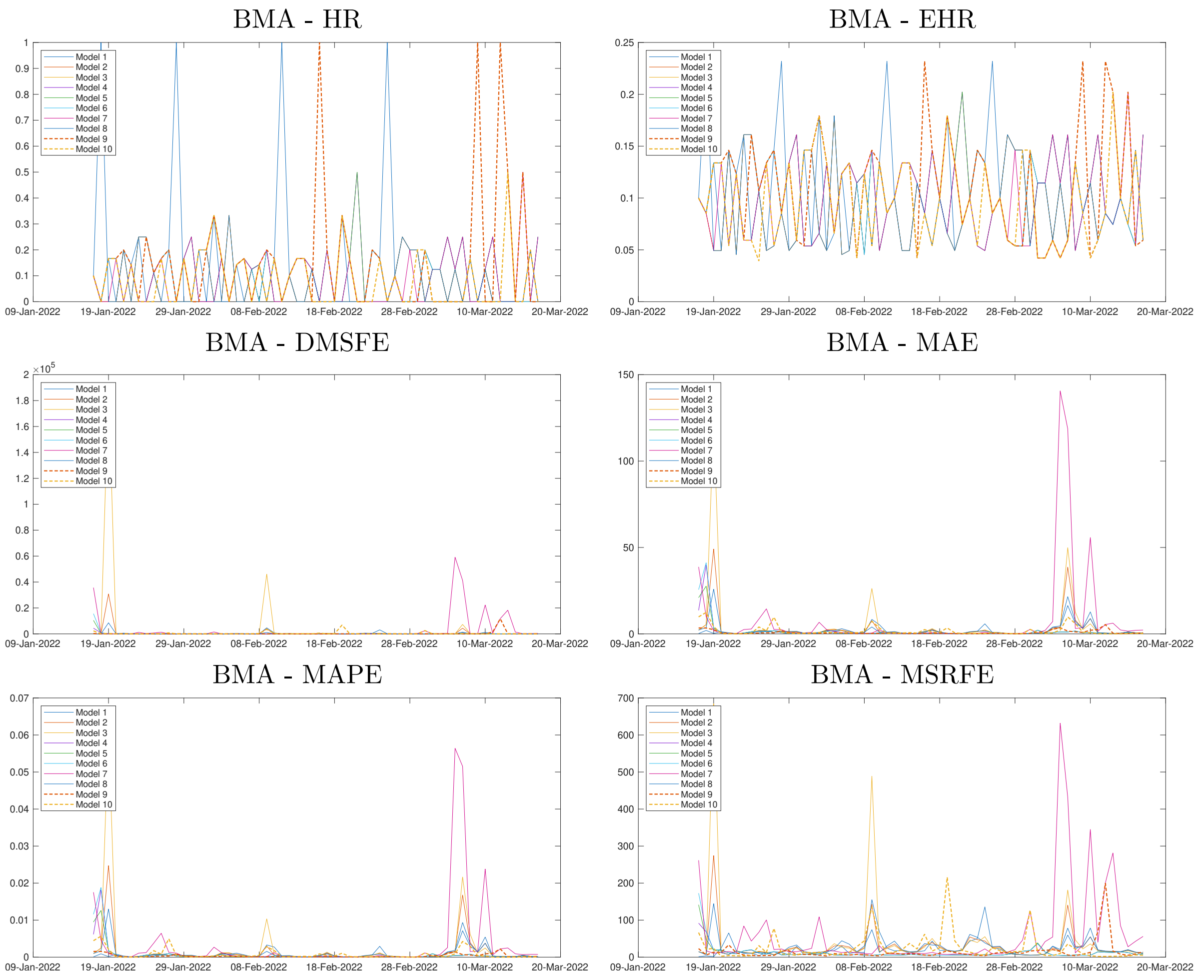

Figure C.18: Model weights for different BMA methods.

Figure C.19: Model weights for different BMA methods.

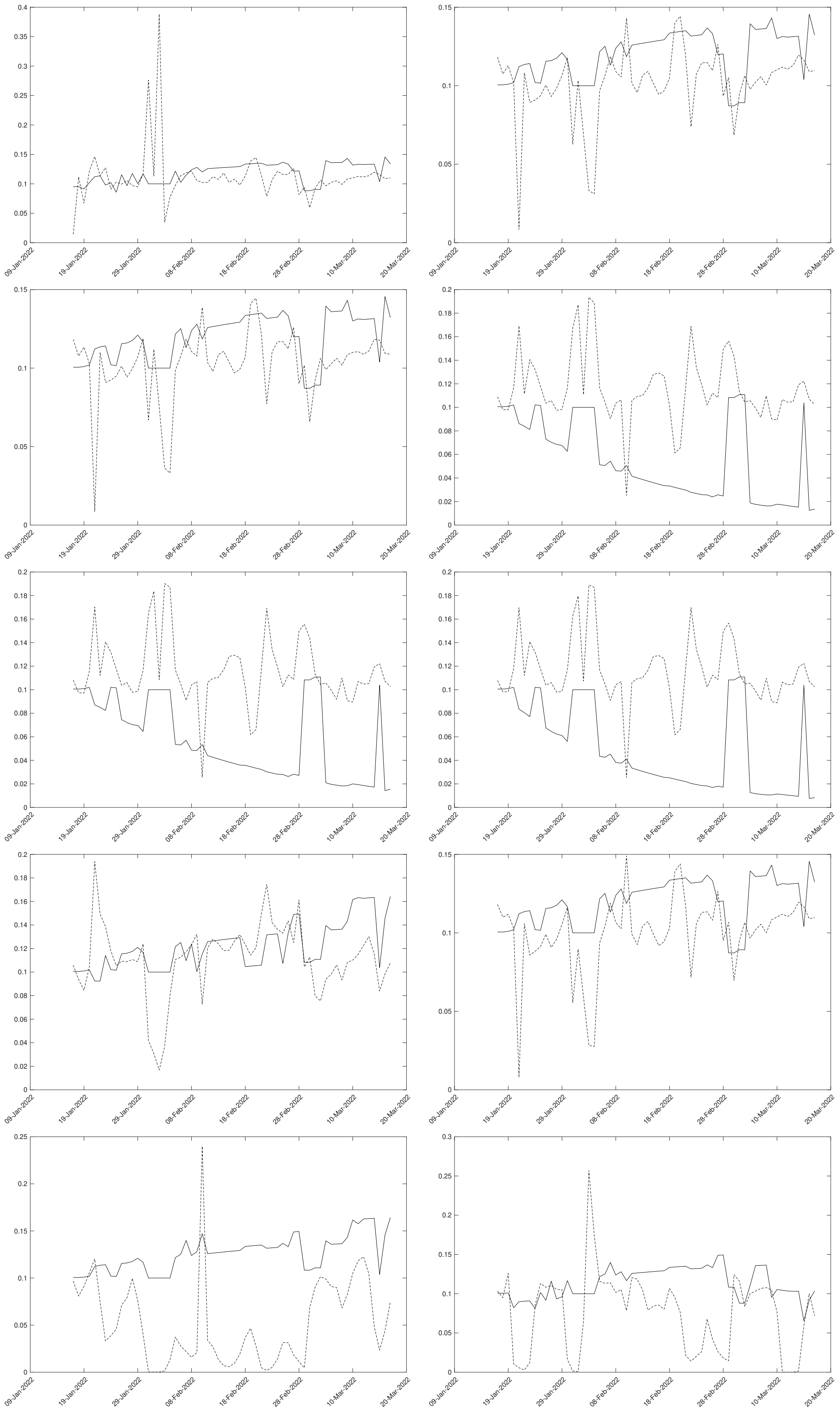

Figure C.20: Model weights based on MSFEs, $w_{\ell,t}$ $\ell = 1, \ldots, 10$ (solid line) and on cost–weighted MSFEs, $\tilde{w}_{\ell,t}$ $\ell = 1, \ldots, 10$ (dashed line).

D Reproducibility Materials

The code and data supporting the results presented in this article have been made publicly available through Code Ocean, a cloud-based computational reproducibility platform. The complete reproducibility package can be accessed at: `https://doi.org/10.24433/CO.3148292.v1`. This repository contains the source code, data, and computational environment required to reproduce the analyses and results reported in the paper. Code Ocean provides a citable and executable research capsule, facilitating transparency and reproducibility.