

Mueller, Rosanna; Sobreira, Nuno

Article

Tourism forecasts after COVID-19: Evidence of Portugal

Annals of Tourism Research Empirical Insights

Provided in Cooperation with:

Elsevier

Suggested Citation: Mueller, Rosanna; Sobreira, Nuno (2024) : Tourism forecasts after COVID-19: Evidence of Portugal, Annals of Tourism Research Empirical Insights, ISSN 2666-9579, Elsevier, Amsterdam, Vol. 5, Iss. 1, pp. 1-14,
<https://doi.org/10.1016/j.annale.2024.100127>

This Version is available at:

<https://hdl.handle.net/10419/337785>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Tourism forecasts after COVID-19: Evidence of Portugal[☆]

Rosanna Mueller^{a,1}, Nuno Sobreira^{b,*,2}

^a ISEG Lisbon School of Economics & Management, Universidade de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal

^b Research in Economics and Mathematics (REM – CEMAPRE), ISEG Lisbon School of Economics & Management, Universidade de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal

ARTICLE INFO

Handling Editor: Masiero Lorenzo

Keywords:

Tourism forecasting competition
Portuguese regions
Pandemic impact
Forecast breakdown
Rankings change
Data adjustments

ABSTRACT

Based on a comprehensive tourism forecasting competition using Portugal's regional data, we study the impact of COVID-19 on the ability of time series models to forecast tourism demand. We find that the stable seasonal patterns observed before the pandemic did not persist in 2020, but regions with higher weights of domestic tourism showed much lower tourism declines and seasonal shifts. Although this change was temporary, it caused significant forecast breakdowns in all methods. However, the intensity of the break differed across models leading to important changes in model rankings, especially in the most affected regions. We discuss the effectiveness and implications of applying straightforward data adjustments and how they can attenuate the pandemic impact on *ex-post* assessment of tourism forecasts.

1. Introduction

The tourism sector plays a key role in the worldwide economy. According to the World Travel & Tourism Council (WTTC), the contribution of travel and tourism to worldwide GDP was approximately 7.6% in 2022 and accounts for 295 million jobs (one in 11 jobs worldwide) (WTTC, 2023). Moreover, the WTTC expectation is that these numbers continue to grow at a fast pace so that the year 2024 may already exceed the 2019 peak. The importance of tourism has also been increasing in many countries and it is, in fact, one of the main income sources in many developing economies (UNWTO, 2023b). Hence, obtaining accurate forecasts for tourism becomes quite relevant as shocks in this sector have important effects on the performance of national economies and, even more broadly, on the worldwide economy (see Liu, Kim, & Song, 2022, for a recent literature review on tourism economic impact). Having accurate methods to forecast tourism may also be decisive for the various actors involved in this industry. For example, tourism-related businesses may need precise forecasts for better planning and management of the resources needed to satisfy demand regarding transport,

accommodations, staff, amenities, etc.

Recently in 2020, strict travel restrictions and regulations were imposed as a response to the COVID-19 pandemic which led to a sudden and sharp decline in tourism demand in 2020. Even in regions with less stringent restrictions on people's movement, it was observed a significant decline in the number of trips and changes in travel behavior (Park, Kim, & Ho, 2022).

Conventional forecasting methods are based on historical patterns observed in-sample, such as trend, cycle, and seasonality, and make forecasts based on the assumption that these patterns will somewhat persist in the future. Hence, even if a given method has proven to be a powerful tool for forecasting tourism demand until February 2020, its accuracy will inevitably decrease in a relevant manner with data from the COVID-19 pandemic due to the strong and sudden changes in tourism patterns during this period. This may pose important challenges to *ex-post* evaluation and comparison of different forecasting methods because the results of the competition may be biased by the observations influenced by the pandemic period if the structural break caused by COVID-19 is unmodelled. This fact raises two important and interrelated

[☆] We wish to thank CEMAPRE and the Department of Mathematics from ISEG. Financial support from FCT -Fundação para a Ciência e Tecnologia through national funds (CEMAPRE/REM - UIDB/05069/2020) is gratefully acknowledged.

* Corresponding author.

E-mail address: nsobreira@iseg.ulisboa.pt (N. Sobreira).

¹ Rosanna Mueller holds a Master's Degree in Data Analytics for Business at ISEG, has relevant working experience as Project Manager and her main interests are machine learning, Big Data, and economic forecasting.

² Nuno Sobreira is Assistant Professor and a Program Director of the Masters in Applied Econometrics and Forecasting and the PhD in Applied Mathematics for Economics and Management at ISEG. His main interests are issues on inference for non-stationary time series and economic forecasting.

<https://doi.org/10.1016/j.annale.2024.100127>

Received 14 August 2023; Received in revised form 19 February 2024; Accepted 22 February 2024

Available online 27 February 2024

2666-9579/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

research questions which we try to answer and contribute with this work. First, is the comparison of the forecasting performance among different methods and corresponding rankings significantly affected by including the observations from COVID-19 in the empirical analysis (research question 1)? Second, if so, which strategies can be designed to attenuate the biases caused by COVID-19 on *ex-post* assessment of tourism forecasts and what are the key challenges (research question 2)?

We address the first question by running a forecasting competition among various methods, comparing the results that exclude the observations from the COVID-19 period to those that include this period. A vast array of modeling approaches is included in the competition such as Naive methods (NAIVE), Autoregressive Integrated Moving Average (ARIMA), Dynamic Regression (DREG) and Neural Network (NNET) models, different developments from the Theta (THETA) and Exponential Smoothing (ETS) frameworks, and the combination of models included in the competition (COMB). We focus on univariate time series models that produce forecasts based only on the history of the variable being forecasted. Due to their ease of implementation, competitive performance, and reliance on readily available data, these models remain popular in the literature on tourism demand forecasting (Song, Qiu, & Park, 2019).

Our empirical results illustrate a significant deterioration in the forecast accuracy of all methods following the COVID-19 outbreak. However, we find that the structural break caused by the pandemic affected the forecasting performance of each class of models differently leading to relevant variations in their ranking positions, especially in the regions whose tourism was mostly affected by the pandemic crisis. Hence, a first implication from our empirical results is that practitioners need to acknowledge the importance of COVID-19 on the results of forecasting competitions involving tourism data.

This motivates and shows how important it is to address our second research question mentioned above, specifically, how to deal with the impact of the outlying COVID-19 observations on the results of *ex-post* evaluation and comparison of competing forecasting methodologies.

We propose and compare different approaches that are easy to understand and implement regardless of the number of methods included in the tourism forecasting competition which consist of either discarding the observations mostly affected by COVID-19 or replacing them with counterfactual values representing what would have occurred if there was no COVID-19. Despite their *ad hoc* nature and the unprecedented and complex changes caused by COVID-19 in tourism data patterns, we show that some of these strategies have been effective in stabilizing the dynamics of the forecast errors and model ranking positions, making them consistent with the patterns before COVID-19.

The analysis is done using regional-level data from Portugal for the number of overnight stays in tourist accommodations. In recent decades, Portugal has experienced extraordinary growth in tourism, making it one of the country's most important economic sectors accounting for 11.8% of Portugal's 2019 GDP and >320 thousand jobs (Jornal de Negócios, 2023; Turismo de Portugal, 2023b). Furthermore, Portugal has been consistently attracting more attention in the global tourism context, as evidenced by the numerous awards and distinctions granted in the last few years (Turismo de Portugal, 2023a).

In summary, this work makes the following contributions to the tourism forecasting literature. First, despite numerous studies about forecasting tourism demand in multiple geographies, there are only a limited number of articles directed towards the Portuguese case (see, for example, Andraz, Gouveia, & Rodrigues, 2009, Rodrigues & Gouveia, 2004, Teixeira & Fernandes, 2014). Hence, we conduct an empirical analysis and examine the results of a large forecasting competition at a relatively unexplored but thriving and increasingly popular tourism destination. Second, we contribute to the recent and fast-growing literature about the impact of COVID-19 observations for empirical analysis on tourism demand. Specifically, this work compares tourism dynamics in Portugal across regions before and following the COVID-19 outbreak and studies the main differences in *ex-post* evaluation of

several forecasting methods when the dataset includes observations that cover the COVID-19 pandemic period. Third, we show how the results of tourism forecasting competitions can be influenced by the inclusion of data following the COVID-19 outbreak and we suggest and compare simple strategies that can account for the deterioration of forecast accuracy and the biases in model ranking positions caused by the COVID-19 data.

2. Related literature

This work is broadly related to the large amount of literature about tourism demand forecasting. In their review of this literature, Song et al. (2019) organizes the most common pre-COVID-19 approaches into four main categories: time series models (Naive, Exponential Smoothing, ARIMA, ...), econometric models (Static and Dynamic Regressions, ARIMA with exogenous variables, Autoregressive Distributed Lag-Error Correction Models (ARDL-ECM), Vector Autoregressions, ...), AI-based models (Artificial Neural Networks, Support Vector Regressions, Fuzzy time series, combinations of AI and traditional time series models, ...), and judgmental methods (Scenario analysis, Delphi surveys, ...).

The COVID-19 pandemic posed a major challenge for empirical analysis and forecasting with tourism demand data as it caused an exceptionally large and sudden drop in the time series values as we will see in Section 3. However, outliers and structural breaks have already been found in many macroeconomic and financial time series. Consequently, the implications of these phenomena for forecasting have already attracted substantial interest in the econometric literature. We may find works that study the negative consequences for forecast performance of not appropriately modeling the breaks and propose adjustment strategies in the presence of breaks. Chen and Liu (1993) studied the impact of four types of outliers, namely innovation outliers, additive outliers, level shifts, and temporary changes, in the forecasts and corresponding accuracies under an ARIMA framework and suggested combining forecasts of the adjusted series (*i.e.*, not influenced by the outliers) with estimates of the outlier effects. Similarly, Koehler, Snyder, Ord, and Beaumont (2012) studied the impact of outliers in ETS forecasts and proposed an augmented innovation state space model with terms that specifically account for these outliers, developing a methodology for estimation and forecasting. Pesaran, Pick, and Pranovich (2013) considered breaks in the slope coefficients of the linear regression model and suggested estimating the parameters based on a weighted-type least squares procedure, with the weights designed to minimize the 1-step ahead forecast errors. Another common practice is to estimate the models using only the most recent observations, such as the post-break observations. A relevant question is how many observations should be included, and both Pesaran and Timmermann (2007) and Inoue, Jin, and Rossi (2017) have established methods for selecting the estimation window size.

Finally, this research is related with several recent studies that have emerged during the early stages of the pandemic about tourism forecasting within the framework of the crisis caused by the COVID-19 outbreak. In July 2020, the Curated Collection of Annals of Tourism Research organized a tourism forecasting competition where, in the first stage, the competing teams produced baseline forecasts based on the best-performing methods from tourism forecasting competitions with pre-COVID-19 data and, in the second stage, had to predict the future path of tourism for 2020 and 2021 (Kourentzes et al., 2021; Liu, Vici, Ramos, Giannoni, & Blake, 2021; Qiu et al., 2021). The teams differentiated themselves not only by the forecasting methods employed but also by the way they characterized different scenarios for tourism recovery (mild, medium, severe) and the strategies adopted to define judgmental adjustments to the baseline forecasts for each origin-destination pair and scenario. Kourentzes et al. (2021) collected data about the expectations of experts regarding travel disruptions in the forecasted years. Qiu et al. (2021) implemented a 7-step methodology supported by information about the evolution of the pandemic and

travel restrictions, while Liu et al. (2021) combined a judgmental approach with the COVID-19 Risk Exposure (CORE) index.

More efforts can be found in the literature regarding the modeling and forecasting of tourism variables following the COVID-19 pandemic. Zhang, Song, Wen, and Liu (2021) forecasted tourist arrivals to Hong Kong from 16 source markets under three pandemic scenarios using baseline forecasts obtained from the ARDL-ECM methodology, adjusted by a qualified Delphi panel considering the impact of the pandemic crisis. Zhang and Lu (2022) combined baseline forecasts for Hong Kong's hotel room demand based on the ARDL-ECM framework with a two-step scenario strategy that characterizes the recovery path of each source market according to the COVID-19 Travelable Index and the pandemic evolution. Fotiadis, Polyzos, and Huan (2021) trained models belonging to the Long Short Term Memory neural network and the Generalized Additive frameworks on previous crises to establish different scenarios in forecasts for international tourism demand following the pandemic outbreak. Provenzano and Volo (2022) compared different univariate models with pre-COVID-19 data and then augmented the Seasonal ARIMA model with a lockdown dummy variable and a regressor for the number of COVID-19 cases. This regressor was then used to establish scenarios for the spread of the pandemic and to obtain forecasts for tourism demand in Lombardia. Chen, Sun, and Liao (2023) explored the effectiveness of using a novel algorithm called BB crogging and augmenting the information set used to build the forecasts with a new COVID-19 impact indicator. Li, Yang, Guo, Sun, and Wang (2023) introduced a new decomposition-ensemble framework for forecasting daily tourism data and proved the superiority of this methodology before and during COVID-19.

Our work differs from the existing literature in several ways. First, as in the aforementioned works, the data adjustment strategies that we explore can be used to improve the forecasting accuracy in tourism datasets that encompass the COVID-19 period. However, while it may be challenging for practitioners to identify in the literature and apply the adjustment tailored to each specific forecasting method, the strategies under analysis in this work are designed to be user-friendly and can be applied easily, regardless of the number of methods being compared. Second, the COVID-19 pandemic has resulted in enormous movements in tourism data without parallel to any other historical period (Assaf, Kock, & Tsionas, 2022; Chen et al., 2023; Li et al., 2023; Song, Qiu, & Park, 2023). This unique context recommends caution in the use of the strategies proposed for previous extraordinary events. Most of the recent studies related to COVID-19 proposed judgmental forecasts and either could not evaluate the accuracy of the proposed forecasts due to non-existent post-COVID-19 data or only used a very limited number of data points. Hence, an additional advantage of our approach is that we empirically show that some of the proposed strategies can be useful for *ex-post* assessment of tourism forecasts when observations affected by COVID-19 are included in the dataset. Third, we provide evidence that the structural break caused by COVID-19 can significantly affect the results of *ex-post* evaluation and comparison of competing forecasting methods. This highlights the need for researchers to account for these consequences in the context of tourism forecasting competitions. The strategies introduced in this work aim to provide an initial proposal for addressing this problem.

3. Data description and exploratory analysis

The data used in this work is the monthly number of overnight stays in tourist accommodations from January 2000 to December 2022 (276 observations) of the seven regions of Portugal defined by NUTS2 ("Nomenclatura das Unidades Territoriais para Fins Estatísticos"): Azores, Alentejo, Algarve, Central, Lisbon Metropolitan Area (Lisbon M. A.), Madeira and North. These names are based on the "Inquérito à Permanência de Hóspedes na Hotelaria e outros Alojamentos" (Survey of Guest Stays in Hotels and Other Accommodations) conducted by Instituto Nacional de Estatística (INE), which is the official public institution

responsible for Portugal's national statistics. This survey must be responded to and sent to INE every month by Hotels, rural tourism accommodations, and local accommodations ("Alojamento Local") with tourist use.

Fig. 1 shows the time series plots for all seven regions. It reveals that the time series are characterized by a strong seasonality and an upward trend until the end of 2019. For all regions, August is usually the month with the highest number of overnight stays, while December or January are the months reporting the lowest values. Seasonality seems to be more pronounced in Algarve and Azores while we find smaller differences between the high and the low season of tourism in the regions of Lisbon M.A. and Madeira. For example, in 2019, the number of overnight stays increased by around 84% from January to August in Algarve, but only by 47% in Lisbon M.A.

However, Fig. 1, and in more detail the seasonal plot of the most recent years of data in Fig. 2 shows a major change in 2020, characterized by a sharp decline in overnight stays, followed by a slow recovery of tourism in the subsequent years. In March 2020, all regions recorded a year-over-year percentage decrease in overnight stays of >50%, rising to >94% in April 2020. This was due to the travel restrictions imposed worldwide as a response to the spread of COVID-19.

From August 2021 onwards, the number of monthly overnight stays in most regions recovered, approximately, to the pre-pandemic levels reaching only slightly lower values than in 2019, the year before the COVID-19 outbreak. In fact, in some regions such as Alentejo, Azores, Madeira, and North, the number of overnight stays in 2022 already surpassed the values recorded in most months of 2019.

Interestingly, the change in Portuguese tourism behavior in 2020 was not uniform across regions being more significant where domestic tourism had historically less weight. The decline in the number of overnight stays after February 2020, measured by the year-over-year percentage change from 03/2020 to 12/2020, was on average highest for Lisbon M.A. (-81.6%), Madeira (-77.7%), and Azores (-76.4%). On the other hand, the regular and logical seasonal patterns observed until 2019, were hardly observed anywhere in 2020 but the changes in seasonality were more pronounced in these regions. For example, Fig. 2 shows that February 2020 became the peak month of the year in Madeira and Lisbon M.A.

4. Methodology

This section describes the methods employed in this work to forecast tourism demand and details the methodology used to compare their forecasting performance.

We begin by obtaining, for each method and region, the series of h -steps-ahead forecasts for the overnight stays using both recursive and rolling window schemes and varying the forecast horizon h . The recursive window scheme is useful as it uses the data effectively available to the forecaster and hence replicates what he can do in real-time. The rolling window scheme always uses the same sample size to produce forecasts which helps to control potential biases in the results caused by older observations.

Each forecasting method's initial training step consists of using an initial window of 138 observations to select the best model specification and estimate its parameters. Specifically, for each forecast technique, the first training step consists of selecting the best specification and estimating its parameters, whenever any of these two steps are needed. For example, when applying the ARIMA approach represented by eq. (1) below, the number of regular and seasonal first differences (d and D), the AR lags (p and P), and the MA lags (q and Q) are specified. Additionally, the unknown parameters in (L) , $\Phi(L^{12})$, $\theta(L)$, and $\Theta(L^{12})$ are estimated. Next, we obtain the 1-, 3- and 6- months-ahead predictions based on the outcome of this training step. The initial sample of 138 observations is then either rolled forward (rolling window) or expanded to next period's observation (recursive window). This new sample is used for another training phase and for obtaining another round of 1-, 3- and 6-

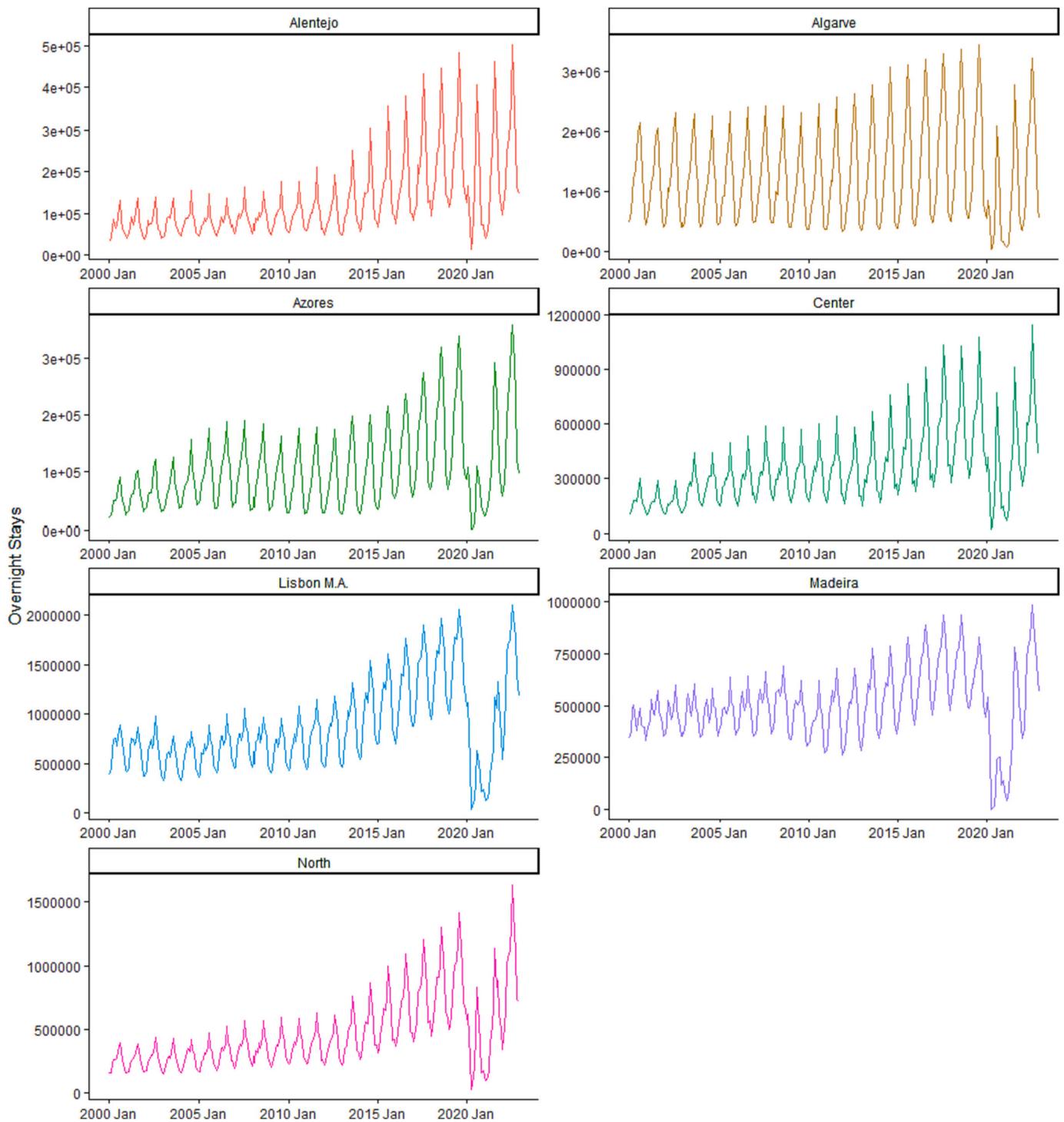


Fig. 1. Monthly number of overnight stays in touristic accommodations per region in Portugal.

– months-ahead predictions. This process is repeated until the end of the sample. Hence, parameter estimates, and model specifications can change as a different training sample is used in each iteration. We have chosen a window size of 138 as it represents about one-half of the sample observations.

To assess the performance of the different methods, the resulting h -steps-ahead forecasts need to be compared to the actual number of overnight stays in each month and region. The evaluation of the forecast accuracy is based on the Root Mean Squared Error (RMSE) which averages the squared forecast errors. Several studies discuss the use of various forecasting accuracy measures, their merits, caveats, and the

circumstances where one metric is more appropriate than another (see [Koutsandreas, Spiliotis, Petropoulos, & Assimakopoulos, 2022](#), for a recent and interesting discussion). We use the RMSE because, despite its limitations, it is among the most popular forecast performance measures and is an appropriate error statistic for the comparisons being made in this forecasting competition. Hence, RMSE values are obtained for all combinations of regions, forecast methods, cross-validation schemes, and forecast horizons.

As regards to choosing the forecast methods to be included in the competition, we can opt for one of two approaches: one approach involves analyzing plots and statistics from each series and choosing one

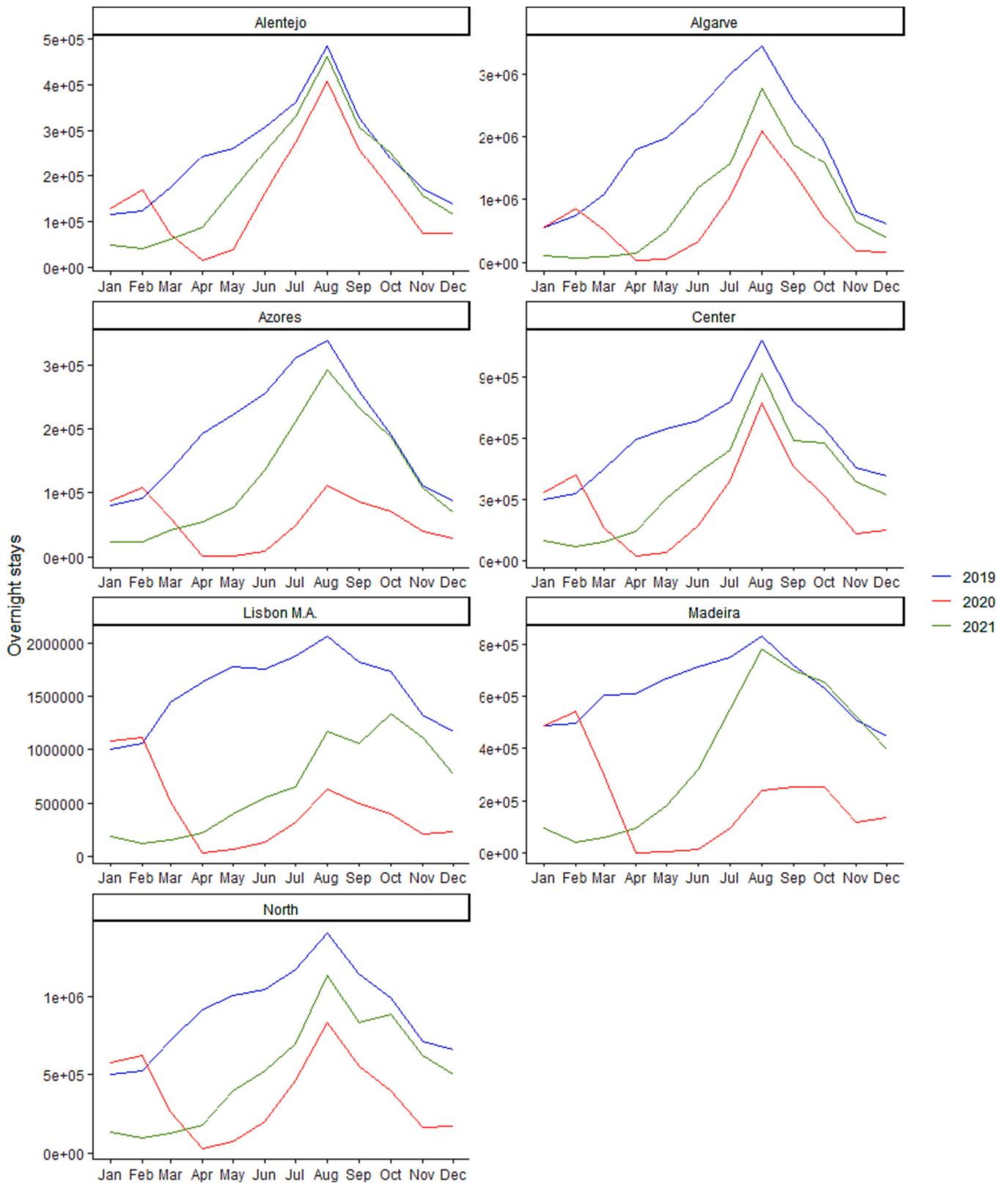


Fig. 2. Seasonal plot of the overnight stays in touristic accommodations per region in Portugal in 2020–2022.

or a small set of candidate models within a given class of methods that should better fit the features identified in the data; another approach is more agnostic and includes a more expanded set of the most widely used models within each class of methods. We opted for the second approach for two reasons: first, it allows us to analyze the results over a broader

range of forecasting techniques. Second, we observe that the characteristics of tourism data exhibited important changes during the COVID-19 period as highlighted in section 3, making it less advisable to rely only on one or two model specifications determined beforehand.

Hence, we have included several techniques to forecast the number

of overnight stays which can be organized into seven model classes: NAIVE, ETS, THETA, ARIMA, DREG, NNET, and COMB. Another issue worth highlighting is the use of variance stabilizer transformations. Whenever it makes sense, each method may appear until three times in the forecasting competition: with forecasts based on models fitted with the original, log-transformed (following, for example, Lütkepohl & Xu, 2012), and Box-Cox transformed data (with the power transformation selected according to Guerrero, 1993). Sections 4.1 to 4.7 provide further details about the forecast methods used for each modeling category.

4.1. NAIVE

We obtain forecasts based on the random walk, random walk with drift and seasonal random walk. Given the seasonal patterns observed in the data, we also consider an interesting alternative, which is to apply an STL decomposition method (Cleveland, Cleveland, McRae, & Terpening, 1990) to obtain seasonally adjusted time series and produce forecasts based on the random walk and the random walk with drift. These forecasts are then reseasonalised so that forecasts for the original data are obtained. For implementation purposes, we used the `rwf()`, `snaive()`, and `stlf()` functions available in the R programming language.

This type of methods is often used to forecast tourism due to the ease of implementation and reasonable prediction accuracy in many cases (Song et al., 2019). In fact, naive techniques were found to produce competitive forecasts for tourism data (Athanasopoulos, Hyndman, Song, & Wu, 2011; Witt & Witt, 1995) or even outperform more complex models in some studies (see Goh & Law, 2011, and references therein).

4.2. ETS

Another set of forecasting methods used in this work is based on the state space approach applied to Exponential Smoothing. This technique is very popular mainly because of its transparency and relatively good performance (see Athanasopoulos et al., 2011, Kourentzes & Athanasopoulos, 2019, and references therein, for examples with tourism data).

We follow Hyndman, Koehler, Snyder, and Grose (2002) and Hyndman, Koehler, Ord, and Snyder (2008) and apply the automatic modeling approach that selects the ETS model minimizing the Corrected Akaike Information Criterion (AICc) to forecast the overnight stays of each region. Additionally, our forecast competition includes the most well-known ETS methods: simple exponential smoothing, Holt's linear trend, Holt's damped trend, Holt-Winters with linear and damped trend, and with additive and multiplicative seasonality. It also includes a decomposition method like the one described in section 4.1 but where the forecasts for the seasonally adjusted series are derived from a non-seasonal ETS model selected according to the minimization of the AICc criterion.

Finally, we also consider the BATS and TBATS models introduced by De Livera, Hyndman, and Snyder (2011), and Complex Exponential Smoothing (CES) proposed by Svetunkov and Kourentzes (2018). To estimate the unknown parameters and select the specification that best fits the data according to each method, we use the functions `ses()`, `holt()`, `hw()`, `ets()`, `bats()`, `tbats()`, and `auto.ces()` that implement the algorithms proposed by the authors of these methods in R.

4.3. THETA

We also obtained forecasts for the overnight stays in each region based on the Theta model (Assimakopoulos & Nikolopoulos, 2000) and on recent algorithms developed to improve the performance of the baseline method. The Theta model gained popularity given its simplicity, ease of use and impressive performance in the M3 competition, outperforming several more sophisticated methods across different series, and forecasting horizons (Makridakis & Hibon, 2000).

Several developments have been made to this method, namely, the

dynamic optimized Theta model (DOTM) proposed by Fiorucci, Pellegrini, Louzada, Petropoulos, and Koehler (2016) that optimizes the Theta coefficient of the short-run component (instead of setting $\Theta = 2$) and allows the level and slope of the long-run trend to be time-varying. Another improvement named AutoTheta was proposed by Spiliotis, Assimakopoulos, and Makridakis (2020) that generalized the trend, and seasonal components from the classical Theta model and allowed both an additive and a multiplicative form such that multiple models become available as in the ETS framework. To choose the most appropriate forecasting model they suggest minimizing the in-sample Mean Absolute Error.

The baseline Theta method was implemented with the `thetaf()` function in R. We applied the DOTM along with the simpler Dynamic Standard Theta (DSTM), Optimized Theta (OTM) and Standard Theta (STM) models through the functions from the `forecTheta` package named with the abbreviation of each method. The forecasts produced by the AutoTheta algorithm were based on the code publicly available for the `AutoTheta()` function. Finally, we also included predictions based on an STL decomposition method where each of these methods is used to forecast the seasonally adjusted data as explained in section 4.1.

4.4. ARIMA

The class of ARIMA models corresponds to one of the most popular approaches for time series forecasting. ARIMA models have also been widely used to forecast tourism demand, and the Seasonal ARIMA model often produces the most accurate forecasts taking advantage of the seasonality in this type of data (Claveria & Torra, 2014).

The most general form for monthly data is the Seasonal ARIMA model which assumes that the dynamics of the series of overnight stays, y_t , are well represented by:

$$\phi(L)\Phi(L^{12})\Delta^d\Delta_{12}^D y_t = c + \theta(L)\Theta(L^{12})\varepsilon_t \quad (1)$$

where L is the lag operator such that $Ly_t = y_{t-1}$, the autoregressive and moving average polynomials are defined as $\phi(L) = 1 - \sum_{i=1}^p \phi_i L^i$, $\Phi(L^{12}) = 1 - \sum_{i=1}^P \Phi_i L^{12i}$, $\theta(L) = 1 - \sum_{i=1}^q \theta_i L^i$, $\Theta(L^{12}) = 1 - \sum_{i=1}^Q \Theta_i L^{12i}$, with roots outside the unit circle, $\Delta^d = (1 - L)^d$, $\Delta_{12}^D = (1 - L^{12})^D$, and ε_t is the error term which is assumed to be white noise.

As can be seen from Eq. (1), we need to decide about the number of regular and seasonal first differences (d and D), the AR lags (p and P) and the MA lags (q and Q). After deciding the proper differencing and lag orders, the unknown parameters in Eq. (1) need to be estimated. We include several possibilities in our tourism forecasting competition. We apply the automatic algorithm proposed by Hyndman and Khandakar (Hyndman & Khandakar, 2008) using the ADF test (Dickey & Fuller, 1979) to select d . Moreover, we considered the forecasts of ARIMA models chosen by the Hyndman and Khandakar algorithm but assuming the following restrictions: $d = 1$, $D = 1$, search limited to non-seasonal models, and all possible combinations of these three restrictions.

Finally, we also considered the decomposition method described in section 4.1 but where the forecasts for the seasonally adjusted series are obtained from a non-seasonal ARIMA model selected by the Hyndman and Khandakar algorithm.

We used the `auto.arima()` function in R to produce the forecasts according to these different ARIMA specifications.

4.5. DREG

Another approach is to forecast the series of overnight stays based on linear regression models with errors that may follow an ARIMA process:

$$y_t = \beta_0 + \sum_{j=1}^k \beta_j x_{jt} + u_t, t = 1, \dots, T$$

where β_0 is the intercept, $\beta_j, j = 1, \dots, k$ are the slope coefficients, $x_{jt}, j = 1, \dots, k$ are the regressors, and u_t follows, in the most general case, a Seasonal ARIMA process as described by Eq. (1).

The idea is to choose regressors representing the trend and seasonal components such that the short-term dynamics are captured by a more parsimonious ARIMA model than the one selected in section 4.4. To implement this approach, we considered regression models with a linear time trend only, seasonal dummy variables only, and with both linear trend and seasonal dummies. We also included in the competition regression models that replace the linear trend or the seasonal dummies (or both) by piecewise linear trends with a knot in 2010 and Fourier regressors varying from 1 to 5 terms, respectively. The Hyndman and Khandakar algorithm implemented by the `auto.arima()` function is then used to select the ARIMA model that best fits the residuals of each specification. The search for the best fitting ARIMA model to u_t is restricted to models without stochastic trends, stochastic seasonality, or both, depending on the deterministic components.

4.6. NNET

Another class of methods used in this work belongs to the framework of Neural Networks, which is a very popular alternative both for modeling and forecasting economic and financial time series. Artificial neural networks have also emerged in the tourism forecasting literature and outperformed more traditional time series models in some studies (see, for example, Peng, Song, & Crouch, 2014, Saayman & Botha, 2017).

In this paper, we will consider methods for a feedforward neural network with a single hidden layer that receives lagged values of the variable to be predicted as inputs. We focus on this model as it corresponds to the most popular type of neural network architecture for univariate time series forecasting. To determine the NNET model structure and estimate the unknown parameters, specifically, the lag order, number of nodes in the hidden layer, the activation function for the hidden layer and learning method, we implemented the algorithms defined by the `mlp()` and `elm()` functions from the `nnfor` R package and the `nnetar()` function from the `forecast` package using the inputs set by default and compared the results. We also implemented the KNN autoregressive model based on the `tsfknn` package. In addition to these forecasts, we also obtained the forecasts based on the STL decomposition method as in section 4.1, where each of these algorithms is used to predict the seasonally adjusted component.

4.7. COMB

Finally, we consider combinations of the forecasts obtained from the methods described in sections 4.1 to 4.6. Forecast combinations can be a competitive strategy and many times outperform most or all individual methods (Clemen, 1989). However, only in more recent years we can find works that explore the benefits of combining in the context of forecasting tourism demand and study the performance of such an approach under different weighting schemes (Claveria, Monte, & Torra, 2016; Wu & Blake, 2023).

To implement this approach in our work, we need to decide how to combine the forecasts as there is no consensus in the literature about this problem. We propose a two-step procedure: first, we find and select the model within each class that had the lowest RMSE in the last 12 months; second, following the same reasoning used by, for example, Petropoulos and Svetunkov (2020), we choose the median operator to combine the forecasts of the models selected in the first step.

5. Empirical results

5.1. Research question 1: what is the influence of COVID-19 on the assessment of competing tourism demand forecasts?

This section assesses the ability of the different methods to forecast Portugal's tourism demand. Specifically, our main interest is to study whether there were significant changes in the relative forecasting performance and corresponding ranking positions among different methods following the COVID-19 outbreak.

We describe first the results from the pre-COVID-19 era. First, we obtain rolling and recursive window 1-, 3- and 6-months ahead forecasts for the overnight stays of the 7 Portuguese regions based on all the forecasting methods described in section 4. Then, to summarize the forecast accuracy of each model class, we first find the forecasting method that minimizes the RMSE within each class in each region and average the RMSE over all regions. The column "Ratio" within "2013–2019" of Table 1 reports the Ratio of the average RMSE obtained in this manner of each model class against the best-performing class between 2013 and 2019 based on the recursive window scheme.³

As for column "# Best", it informs about the number of regions where a model for a given class was the top-ranked method. To make Table 1 and other similar tables easier to read, we use a color palette where the columns "Ratio" and "# Best" are printed with a darker color as the values get closer to 1 (model class with the lowest average RMSE) and 7 (total number of regions), respectively.

The column "Ratio" within "2013–2019" from Table 1 shows that the NNET class had, on average, the best 1-step ahead forecasting performance in the period between 2013 and 2019 with a 6% advantage over the COMB scheme and the ARIMA methodology, and 12% advantage over the ETS framework which are placed in the 2nd, 3rd and 4th places, respectively (see rows corresponding to "h = 1"). The remaining methods exhibit a more sizeable disadvantage in terms of forecast accuracy. Moreover, the column "# Best" reveals that the NNET or the ARIMA best models outperform the other model classes in 6 out of 7 regions in coherence with the RMSE ratio results.

Focusing now on the same columns but with "h = 3" and "h = 6", we find that, before COVID-19, the top-3 places were occupied by the same forecast approaches at longer forecast horizons (3 and 6 months ahead). Indeed, the best NNET models continue to be, on average, the most accurate methods and even become top-ranked in more regions than in the 1-month horizon.

To have a more precise idea about the methods that are contributing to these rankings, the columns headed "2013–2019" from Table 2 list the top-10 methods in the pre-COVID-19 period ranked according to the average 1-step ahead RMSE over all regions and it also informs about the number of regions where they achieved the top-3 positions. We see that the methods which stood more often on a podium position are the median combination scheme, the MLP neural network either fitted to the levels or first differences, and the ARIMA based on the Hyndman and Khandakar algorithm applied to the seasonally differenced and log-transformed data.

To understand if the same model rankings hold across the years before COVID-19 or are the outcome of some subsample period, we also compare the performance of the same models within each class of methods in shorter periods. Specifically, Table 1 reports the "Ratio" and "# Best" results of the same models within each class in the periods 2013–2014, 2015–2016, and 2017–2019. Apart from some exceptions, the results continue to highlight the performance of NNET, ARIMA, and COMB.

³ We focus most of our attention on the results obtained from the recursive window forecasts, but we highlight that only small differences have been found when the rolling window scheme was used as documented in the supplementary appendix.

Table 1

Comparing forecast accuracy before COVID-19 based on the best performing models within each class: recursive window forecasts.

Model class	h	2013–2014		2015–2016		2017–2019		2013–2019	
		Ratio	# Best						
NNET	1	1.00	3	1.06	1	1.00	2	1.00	3
ARIMA	1	1.09	1	1.00	5	1.09	1	1.06	3
COMB	1	1.00	2	1.12	0	1.12	2	1.06	1
ETS	1	1.19	0	1.11	1	1.10	1	1.12	0
NAIVE	1	1.08	1	1.26	0	1.36	0	1.23	0
THETA	1	1.08	0	1.21	0	1.42	1	1.24	0
DREG	1	1.19	0	1.39	0	1.52	0	1.35	0
NNET	3	1.02	0	1.03	3	1.00	3	1.00	6
COMB	3	1.00	2	1.02	1	1.13	1	1.04	0
ARIMA	3	1.01	3	1.00	3	1.21	1	1.07	1
ETS	3	1.24	0	1.17	0	1.08	2	1.16	0
NAIVE	3	1.04	1	1.24	0	1.45	0	1.22	0
THETA	3	1.25	0	1.30	0	1.59	0	1.37	0
DREG	3	1.27	1	1.70	0	1.89	0	1.58	0
NNET	6	1.00	2	1.00	2	1.00	5	1.00	5
COMB	6	1.08	1	1.09	2	1.15	0	1.10	0
ARIMA	6	1.11	1	1.06	2	1.24	0	1.14	2
ETS	6	1.08	1	1.25	1	1.33	1	1.20	0
NAIVE	6	1.10	1	1.39	0	1.47	0	1.28	0
THETA	6	1.41	0	1.52	0	1.64	1	1.50	0
DREG	6	1.30	1	1.69	0	1.73	0	1.52	0

Note: **Ratio** – Ratio between the average RMSEs across regions of the best methods within each class relative to the top-performing class overall; **# Best** – number of regions where a model from each class was the top-ranked method.

Table 2

Top-10 best performing methods and corresponding model families based on recursive window 1-step ahead forecasts for the periods 2013–2019 and 2020–2022.

2013–2019				2020–2022			
Model class	Method	RMSE	# Best	Model class	Method	RMSE	# Best
NNET	mlp	36,077	2	DREG	dregtrendseasonal	139,347	3
NNET	mlp_d1	36,389	2	DREG	dregseasonal	144,916	1
COMB	comb_median	36,581	2	DREG	dregtrendfourier5	145,341	3
ARIMA	arima_ds1	36,595	1	NAIVE	rwstlf	145,472	3
NNET	mlp_log	37,614	2	NAIVE	rwdstlf	145,727	2
NNET	mlp_d1_log	37,903	1	NNET	elmstl	146,010	0
ARIMA	arima_d1	38,156	3	ETS	bats	150,913	0
ARIMA	arima_log	38,283	2	NNET	elm_d1	151,171	0
ARIMA	arima	39,162	1	NNET	elmstl_d1	151,836	2
ARIMA	arima_d1_ds1_log	39,411	2	DREG	dregtrendbrseasonal	152,864	1

Note: the meaning of the **Method** abbreviations is available in the supplemental appendix; **RMSE** – average RMSE across regions of each method; **# Best** – number of regions where each method achieved one of the top-3 positions.

The columns headed “2013–2014”, “2015–2016”, and “2017–2019” from [Table 1](#) use the most accurate models during 2013–2019 to make comparisons in the considered subperiods. Another possibility is to obtain the “Ratio” and “# Best” results based on the forecasts of the models within each class that minimize the RMSE in each subperiod per region. Overall, we find that this analysis produces similar rankings as [Table 1](#) because either the best-performing models coincide or the relative RMSE differences are marginal at best.⁴

We now assess the performance of the different methods to forecast the post-COVID-19 tourism demand observations and discuss the main differences relative to the results before COVID-19.

[Fig. 3](#) illustrates, for each of the 7 Portuguese regions, how the 1-step ahead RMSE from the best performing method within each class during the pre-COVID-19 period developed with the inclusion of more observations of the overnight stays. We find that the 1-step RMSE of most models maintained relatively stable levels before the COVID-19 outbreak with the lines going almost in parallel in such a way that the rankings remained largely unchanged between 2013 and 2019 following what we already found in [Table 1](#). However, the inclusion of the observations following the COVID-19 outbreak led to a strong and sudden

upward-level shift in the RMSE for all methods and regions. Specifically, [Fig. 3](#) shows that the highest increase in the RMSE occurs when we add the values of overnight stays from March 2020 and April 2020. Furthermore, the inclusion of the observations from the remaining months of 2020, 2021, and 2022 did not, in general, lead to another shift in the RMSEs.⁵

A deeper analysis of the results reveals that the deterioration in forecast accuracy differs across models and regions. If, for each model class, we average the RMSE values used in [Fig. 3](#) over all regions, we find that the performance of the NNET, COMB, and ARIMA methods dropped by >80% during March and April 2020 whereas the NAIVE and DREG methods depreciated, on average, by “only” around 60%.

To understand whether these differences in the intensity of the forecast breakdown significantly changed the rankings, we consider the methods within each class that performed better before COVID-19 per region and assess their relative accuracy when predicting tourism demand in the post-pandemic period. Hence, we average the RMSE of the forecasts for 2020, 2021, and 2022 and the overall period 2020–2022 across regions, based on the same models which were used to build

⁵ We would find a very similar picture if the RMSEs for longer horizons were plotted instead. These figures are available upon request.

⁴ These results are available upon request.

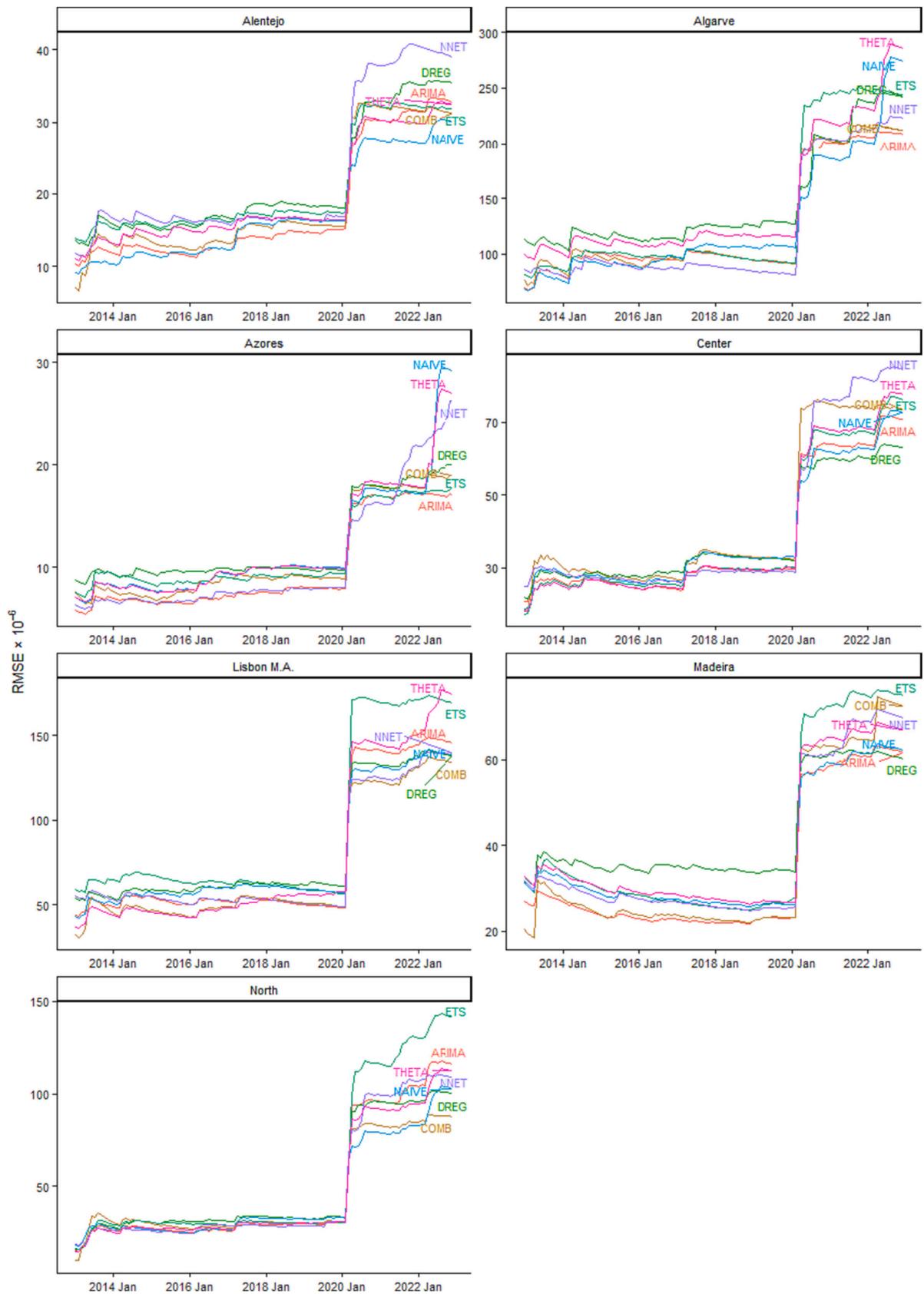


Fig. 3. Evolution of the 1-step ahead RMSE from the Pre-COVID-19 best performing methods within each model family on the 7 Portuguese regions: recursive window forecasts.

Table 1.

Using this strategy, the columns labeled “Ratio” in Table 3 report the average RMSE ratios of the different classes of models against the most accurate methods. As in Table 1, we also report the number of regions where each model class was the best-ranked forecasting approach in column “# Best”.

The comparison between Tables 1 and 3 reveals interesting changes in the rankings because of considering the COVID-19 period. Specifically, the NAIVE methods were competing for the best 1-month ahead forecast performance in 2020 and 2021 (rows corresponding to $h = 1$) although their ranking dropped substantially with the tourism recovery in 2022. Similarly, the selected DREG models improved significantly in the post-pandemic period and are now among the top-performing methods, moving up from the last to the top-3 positions in most cases. Conversely, the ETS method, which performed well in the pre-COVID-19 period, became one of the worst-ranked methods. The ARIMA and NNET also moved down in their rankings for shorter-term forecasts, although their average RMSE differences relative to the top-ranked methods are not substantial. Finally, the COMB scheme is either ranked first or competes closely with the top-performing methods. This performance was already observed in the pre-COVID-19 period, showing that this method is more robust against the sudden and strong changes that followed the COVID-19 outbreak.

Instead of comparing model classes based on the pre-COVID-19 most accurate models, we may find the methods that better forecast the observations from the post-pandemic period. The columns headed “2020–2022” from Table 2 inform about the top-10 methods, ranked according to the average RMSE across regions and the number of regions where each method stood on the podium in the period 2020–2022. Comparing the columns headed “2013–2019” with “2020–2022” from Table 2, we find that the types of models listed in the top-10 places are now very different. The top-10 list now includes NAIVE techniques with an emphasis on the random walk method (with or without drift) combined with an STL decomposition, linear regressions with deterministic terms accounting for trend and seasonality, and the BATS model belonging to the DREG and ETS approaches, respectively. Moreover, as before COVID-19, a method from the NNET family is often on the podium but, this time, it is the ELM (and not the MLP) that is propelling the NNET class to the top of the rankings.

In summary, it seems that the rankings of the forecasting methods

under analysis have been relatively stable before the COVID-19 period both in the overall global ranking and within each specific model class. Specifically, the most successful methods in forecasting tourism demand from 2013 to 2019 include more sophisticated univariate methods for modeling trend and seasonality, such as ARIMA and MLP with seasonal inputs and using statistical criteria for model selection. However, including in the empirical analysis the observations following the COVID-19 outbreak not only generated significant upward shifts in the RMSEs but also important changes in model rankings that the forecaster needs to be aware of when assessing the forecast performance of different methods. This seems to be happening whether we compare only the best-performing models from the pre-COVID era or if we find the best performers based on data that already includes the observations affected by the COVID-19 pandemic. Particularly, basic approaches for accounting for trend and seasonality such as using linear trends, dummy variables or Fourier terms, and simpler forecasting methods such as the random walk (with or without drift) or linear regression models with more parsimonious ARMA errors, seem to be more successful in forecasting the post-COVID-19 period.

Such substantial differences in the relative forecasting performance of conventional time series methods between the pre-pandemic and post-pandemic periods may create uncertainty for the forecaster about the best way to proceed. However, these differences are justified as these forecasting techniques did not account for the enormous and unprecedented movements in tourism data following COVID-19. Consequently, the results that include the post-COVID-19 observations have limited practical value. First, the estimates of the model parameters and, consequently, the forecasts become biased as the observations mostly affected by COVID-19 are included in the training sample. Second, including these observations in the test data generates outlying forecast errors that significantly influence the values of the forecasting accuracy measures. This reinforces the importance of discussing different adjustments to the data mostly affected by COVID-19 with the goal of either accounting for or, at least, diminishing the influence of these outlying observations in the *ex-post* evaluation and comparison of competing forecasting methods. This is what is done in the next section.

Table 3

Comparing forecast accuracy after COVID-19 using the Pre-COVID-19 top performing models within each class: recursive window forecasts.

Model class	h	2020		2021		2022		2020–2022	
		Ratio	# Best	Ratio	# Best	Ratio	# Best	Ratio	# Best
COMB	1	1.05	2	1.00	0	1.00	3	1.00	3
DREG	1	1.05	2	1.20	3	1.11	2	1.05	2
ARIMA	1	1.13	0	1.03	2	1.23	1	1.08	1
NNET	1	1.16	0	1.27	0	1.31	1	1.16	0
NAIVE	1	1.00	2	1.01	1	2.04	0	1.17	1
ETS	1	1.36	1	1.17	0	1.20	0	1.27	0
THETA	1	1.16	0	1.07	1	2.22	0	1.29	0
COMB	3	1.00	3	1.01	2	1.00	0	1.00	5
ARIMA	3	1.09	0	1.00	1	1.28	0	1.10	0
ETS	3	1.14	0	1.22	0	1.20	2	1.17	0
DREG	3	1.18	2	1.12	2	1.04	2	1.17	0
NNET	3	1.12	2	1.12	1	1.38	1	1.17	0
NAIVE	3	1.30	0	1.05	0	2.08	1	1.40	1
THETA	3	1.11	0	1.09	1	2.64	1	1.40	1
COMB	6	1.00	2	1.10	3	1.00	3	1.00	3
DREG	6	1.07	1	1.00	1	1.16	1	1.05	1
ETS	6	1.05	2	1.22	1	1.01	3	1.06	1
ARIMA	6	1.07	0	1.08	1	1.20	0	1.07	1
NNET	6	1.03	2	1.25	0	1.20	0	1.09	0
THETA	6	1.05	0	1.25	1	1.66	0	1.15	1
NAIVE	6	1.13	0	1.25	0	1.89	0	1.24	0

Note: **Ratio** – Ratio between the average RMSEs across regions of the best methods within each class relative to the top-performing class overall; **# Best** – number of regions where a model from each class was the top-ranked method.

5.2. Research question 2: how can we attenuate the COVID-19 impact on ex-post assessment of competing forecasts?

In this section, we consider simple and common-sense strategies to mitigate the influence of the extreme COVID-19 observations on the ex-post assessment of competing forecasts and we evaluate if these strategies have been effective in achieving their intended goal. The strategies that we compare are:

- (a) discard the 2020 observations, $z_t^{(a)}$
- (b) discard the 2020 and 2021 observations, $z_t^{(b)}$
- (c) replace the 2020 observations with the mean per month of 2019 and 2021, $z_t^{(c)}$
- (d) replace the 2020 and 2021 observations with weighted means per month of 2019 and 2022, $z_t^{(d)}$

Since our series of regional overnight stays cover the period from January 2000 to December 2022 (276 observations), and considering that y_{240} , y_{252} , and y_{264} correspond to December values for the years 2019, 2020, and 2021, respectively, the adjusted time series per region are obtained as follows:

$$z_t^{(a)} = \begin{cases} y_t, & 1 \leq t \leq 240 \\ y_{t+12}, & 241 \leq t \leq 264 \end{cases} \quad z_t^{(c)} = \begin{cases} y_t, & 1 \leq t \leq 240 \vee 253 \leq t \leq 276 \\ (y_{t-12} + y_{t+12})/2, & 241 \leq t \leq 252 \end{cases}$$

$$z_t^{(b)} = \begin{cases} y_t, & 1 \leq t \leq 240 \\ y_{t+24}, & 241 \leq t \leq 252 \end{cases} \quad z_t^{(d)} = \begin{cases} y_t, & 1 \leq t \leq 240 \vee 265 \leq t \leq 276 \\ (2y_{t-12} + y_{t+24})/3, & 241 \leq t \leq 252 \\ (y_{t-24} + 2y_{t+12})/3, & 253 \leq t \leq 264 \end{cases}$$

Specifically, we start by generating the adjusted time series for overnight stays in the seven regions using each of these strategies, resulting in time series denoted as $z_t^{(j)}$, $j \in \{a, b, c, d\}$ for each region. Then, we calculate recursive window 1-, 3-, and 6-steps ahead forecasts for each adjusted time series according to each forecasting method.⁶ Finally, we compute the RMSEs for the year 2022 according to each case and rely on these results to make performance comparisons. We selected the year 2022 to ensure that the evaluation of the adjustment strategies is based on forecast errors from the same period.

Strategies (a) until (d) are commonly used when dealing with missing observations, although they have already been discussed in the context of extreme observations related to COVID-19 (Lenza & Primiceri, 2022). The idea of strategies (a) and (b) is to drop potentially contaminated outlier observations, while strategies (c) and (d) replace these contaminated observations with values obtained through linear interpolation from recent years, which were not or were less affected by the COVID-19 crisis. The advantages of these strategies are that they are straightforward to understand and implement and serve as a starting point for discussion and comparison of more complex schemes. Moreover, as shown below, some of these strategies have contributed to bringing the dynamics of forecast errors and rankings closer to the results obtained until 2019. However, these strategies also come with limitations. Indeed, dropping observations may lead to the loss of information that could be useful for the forecasting exercise. On the other hand, replacing observations may bias the results if the counterfactual exercise uses values substantially different from what would have occurred without COVID-19.

Table 4 presents the relative accuracy of the forecasts for the data-adjusted year 2022 obtained from the models used to produce Table 1, i.e., the best performers within each class before COVID-19. The results in Table 4 provide useful insights on which strategies are more effective in reducing the impact of COVID-19 on ex-post assessment of tourism forecasts and how the results are sensitive to the model family

⁶ Results for the rolling window case are provided in the supplementary appendix.

and forecast horizon.

Indeed, it seems that approach (d) which replaces the 2020 and 2021 observations by linear interpolation is the one that better preserves the model rankings observed before COVID-19. After comparing the columns headed “2013–2019” and “replace 2020–2021” in Tables 1 and 4, respectively, we find that this is the only strategy where the top-3 positions remain occupied by the same class of models, NNET, COMB, and ARIMA, regardless of the forecast horizon. Moreover, we only observe slight moves in rankings as, for example, for $h = 1$, both ARIMA and COMB exchanged their relative positions and the same occurred to THETA and NAIVE. We also find higher values in the ratios between the top-3 methodologies and the others as, for example, THETA, NAIVE and DREG display average RMSEs values that are more than the double of the best performing NNET models. This is natural as these values are based on the comparison of actual and forecasted values from less data points.

Fig. 4 replicates the same information from Fig. 3 but now applied to the 1-step ahead forecasts of the overnight stays adjusted according to strategy (d). The line plots reinforce the idea that this data adjustment approach was able to attenuate the impact of the COVID-19 shock on the performance of most forecasting methodologies. In fact, the upward level shift in the RMSE values in March and April 2020 is not observed anymore and the accuracy values display relatively stable patterns throughout the sample period in most of the regions.

Another strategy that deserves attention is (b) which, besides the NNET methods, reveals rankings and RMSE ratios like those observed during the period 2013–2019. Particularly, when comparing the columns headed “2013–2019” with “omit 2020–2021” in Tables 1 and 4, respectively, we observe that both COMB and ARIMA consistently maintain the top-3 positions while methods such as NAIVE and DREG occupy the bottom positions for most cases. Moreover, despite some fluctuations in the average RMSE values, the methodologies remain roughly in the same relative positions over longer forecast horizons as it was already observed in pre-COVID-19 period.

A different picture can be drawn by looking at the results from strategies (a) and (c) in columns headed “omit 2020” and “replace 2020”, respectively, of Table 4. Under these approaches, the model positions seem to be more in line with the rankings presented in Table 3, thereby deviating from the objective we have in mind. For example, the NNET methodology drops out from the top-3 positions while DREG rises to the top-3, even becoming the top-performing methodology for longer forecast horizons. These changes in rankings were already observed when the post-COVID-19 observations were incorporated in the forecast analysis in Table 3. Under these adjustment strategies, we also observe less consistent rankings across forecast horizons, and diverging from the pre-COVID-19 rankings from Table 1. Hence, it seems that the model rankings are more consistent with those before the pandemic when adjustments are applied to both the observations from 2020 and 2021.

6. Discussion and conclusions

Time series models have proven to be powerful tools for forecasting tourism demand. However, the COVID-19 pandemic led to an unprecedented drop in tourism levels in 2020 with 1.1 billion fewer international tourists (overnight visitors) worldwide, which represented a 72% decrease in international traveling (UNWTO, 2023a). A change of this order of magnitude naturally poses challenges and doubts when modeling and forecasting tourism demand as models which provide reasonably accurate forecasts may no longer display similar performance.

This work aims to further analyze this situation by running a competition to forecast tourism demand between several methods using regional data from Portugal and analyzing the most relevant changes in forecast performance after including the pandemic period in the dataset. We find that the COVID-19 pandemic had severe effects on Portugal’s tourism but not at the same level for all regions, with regions having

Table 4

Comparing the accuracy of Pre-COVID-19 top performing models to forecast 2022 adjusted tourism demand: recursive window forecasts.

Model class	h	(a) omit 2020		(b) omit 2020–2021		(c) replace 2020		(d) replace 2020–2021	
		Ratio	# Best	Ratio	# Best	Ratio	# Best	Ratio	# Best
NNET	1	1.47	1	1.29	0	2.28	0	1.00	4
COMB	1	1.00	3	1.07	0	1.00	4	1.08	1
ARIMA	1	1.16	0	1.00	5	1.61	0	1.22	2
ETS	1	1.27	0	1.09	0	1.42	0	1.63	0
THETA	1	1.41	1	1.05	0	1.34	0	2.18	0
NAIVE	1	1.33	1	1.12	1	1.24	2	2.31	0
DREG	1	1.14	1	1.18	1	1.09	1	2.87	0
NNET	3	1.86	0	1.35	0	2.88	0	1.00	3
COMB	3	1.00	3	1.07	1	1.48	0	1.26	3
ARIMA	3	1.21	1	1.00	3	2.16	0	1.53	0
ETS	3	1.17	2	1.07	1	1.42	1	1.89	0
THETA	3	1.41	0	1.05	0	1.32	0	2.63	0
NAIVE	3	1.65	1	1.21	1	1.48	2	2.66	1
DREG	3	1.18	0	1.47	1	1.00	4	3.82	0
COMB	6	1.23	0	1.06	0	1.62	0	1.00	3
NNET	6	2.20	0	1.16	2	2.06	0	1.06	0
ARIMA	6	1.49	1	1.04	1	2.12	0	1.20	0
NAIVE	6	1.84	1	1.00	3	1.60	1	1.21	4
ETS	6	1.22	2	1.17	0	1.46	0	1.78	0
THETA	6	1.20	1	1.06	1	1.18	2	2.06	0
DREG	6	1.00	2	1.35	0	1.00	4	2.64	0

Note: **Ratio** – Ratio between the average RMSEs across regions of the best methods within each class relative to the top-performing class overall; **# Best** – number of regions where a model from each class was the top-ranked method.

higher weights of domestic tourism showing much lower declines in demand and fewer changes in seasonality in 2020.

Although this change in tourism dynamics was temporary, we find evidence that this was enough to cause significant forecast breakdowns in all considered time series methods. Moreover, the intensity of the break in forecast performance was found to differ across models leading to important changes in model rankings both in general and within specific model classes, especially in the regions most affected during the COVID-19 period.

Finally, we explored the effectiveness of some simple and common-sense strategies to adjust the data that was most influenced by the pandemic, like omitting certain observations or applying linear interpolation techniques, in attenuating the pandemic impact on *ex-post* assessment of tourism forecasts.

The results from this research have relevant theoretical and practical implications for both the forecasting literature and the tourism sector. Previous studies have shown that ignoring structural breaks can tremendously impact the accuracy of forecasts, particularly when the forecast's origin is close to the break date. Our work reinforces the importance of this finding by highlighting the substantial effect of the recent COVID-19 outbreak on point forecasts for tourism demand produced from 2020 to the present day. However, most of this literature has primarily focused on classical time series models such as Box-Jenkins-type ARIMA or ETS. This research documents losses in forecasting accuracy across a much wider range of methods and emphasizes that the impact of the COVID-19 break may vary substantially across different model families. Consequently, the rankings of models can change substantially when structural breaks are not considered. In fact, our findings confirm that models that consistently performed well in the pre-COVID period are now apparently performing worse, while others have become more competitive due to the inclusion of observations affected by COVID-19. This seems to be happening both when comparing forecasting methods in general and within the specific model classes.

From a practitioner's point of view, these results may create uncertainty about the best way to proceed when the pandemic period is included in the dataset. This could lead to suboptimal decisions such as, for example, choosing less accurate forecasting methods and tourism forecasts. This may result in misleading scenarios about the performance of national economies and tourism-related businesses, potentially leading to wrong planning and management decisions.

The existing forecasting literature proposes various strategies to address the effects of structural breaks in point forecasts. However, practical challenges, such as time and workload constraints, can make it difficult for practitioners to apply the adjustments tailored to each specific forecasting method. Alternatively, we take the practitioner's point of view and consider intuitive and straightforward data adjustments. Our empirical findings reveal the need to adjust not only 2020 but also 2021 observations. Moreover, among the considered strategies, linear interpolation appears to be more effective in stabilizing the dynamics of the forecast errors and model rankings, making them more consistent with the patterns observed until 2019. Finally, it should be acknowledged that while some of these strategies seem to be useful, they have some limitations, as outlined in section 5.4. An additional limitation of this work is that we can only conjecture that this linear interpolation strategy will reduce the bias in forecasting more recent tourism data, such as for the years 2023 and 2024. However, we have not empirically validated this conjecture. Hence, designing more robust adjustment systems and proving that they effectively account for the effects of the breaks caused by COVID-19 when forecasting more recent tourism data is an interesting avenue of research.

It would also be of interest to expand the forecasting competition to econometric models that incorporate explanatory variables. After performing a comprehensive forecasting competition, Athanasopoulos et al. (2011) found that univariate approaches forecast tourism demand more accurately than methods that use explanatory variables in line with previous studies. Nevertheless, these results were based on time series with regular patterns and could change with observations “contaminated” by the COVID-19 outbreak.

CRedit authorship contribution statement

Rosanna Mueller: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nuno Sobreira:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

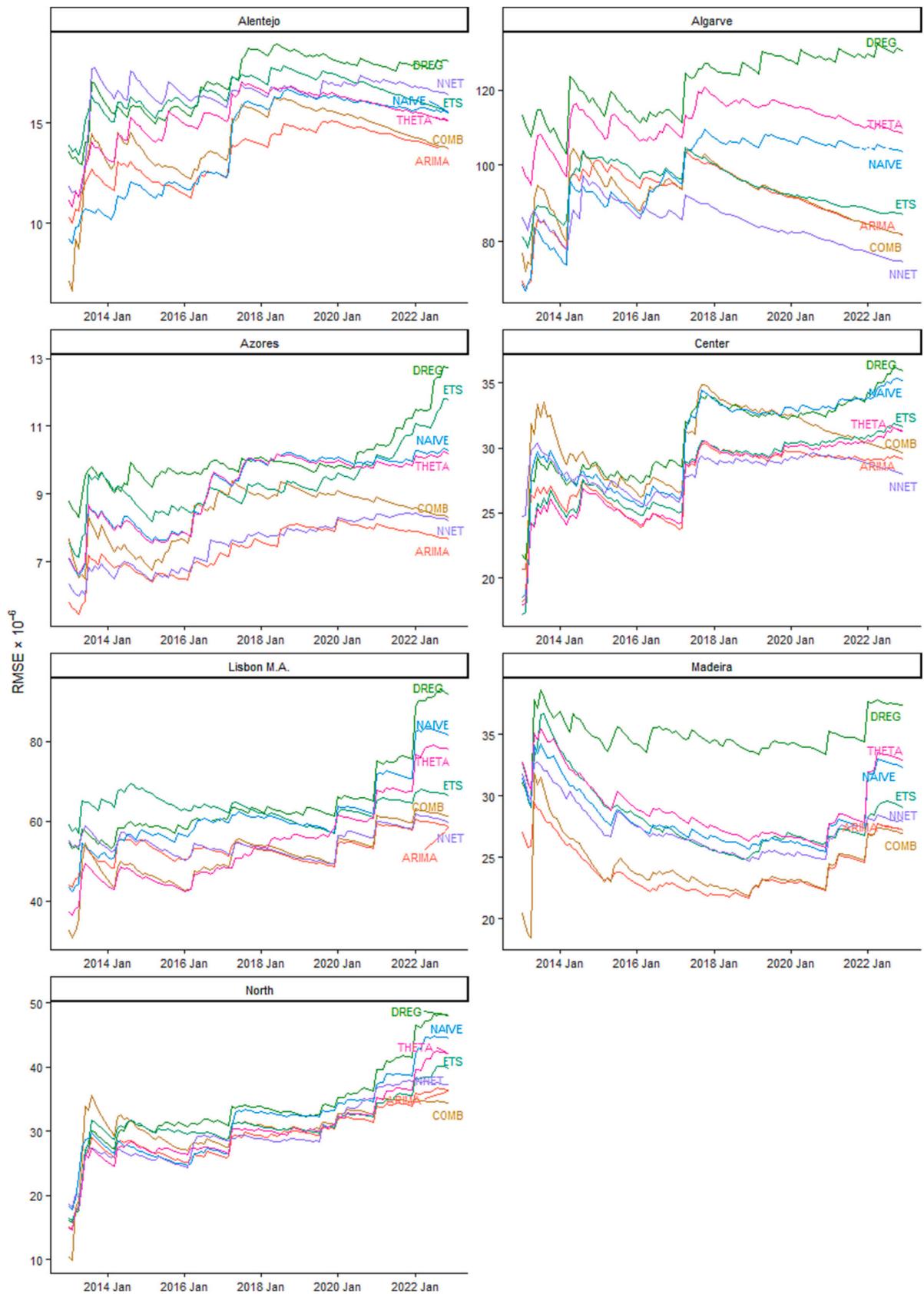


Fig. 4. Evolution of the 1-step ahead RMSE from the Pre-COVID-19 best performing methods within each model family under data adjustment strategy (d) (replace the 2020 and 2021 observations by linear interpolation): recursive window forecasts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to three anonymous reviewers for their very helpful comments. We wish to thank CEMAPRE and the Department of Mathematics from ISEG. Financial support from FCT-Fundação para a Ciência e Tecnologia through national funds (CEMAPRE/REM - UIDB/05069/2020) is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.annale.2024.100127>.

References

- Andraz, J. L., Gouveia, P. M., & Rodrigues, P. M. (2009). Modelling and forecasting the UK tourism growth cycle in Algarve. *Tourism Economics*, 15(2), 323–338.
- Assaf, A. G., Kock, F., & Tsionas, M. (2022). Tourism during and after COVID-19: An expert-informed agenda for future research. *Journal of Travel Research*, 61(2), 454–457.
- Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting*, 16(4), 521–530.
- Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3), 822–844.
- Chen, C., & Liu, L.-M. (1993). Forecasting time series with outliers. *Journal of Forecasting*, 12(1), 13–35.
- Chen, D., Sun, F., & Liao, Z. (2023). Forecasting tourism demand of tourist attractions during the COVID-19 pandemic. *Current Issues in Tourism*, 1–19.
- Claveria, O., Monte, E., & Torra, S. (2016). Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*, 23(6), 428–431.
- Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220–228.
- 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. *Journal of Official Statistics*, 6(1), 3–73.
- 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.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.
- Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., & Koehler, A. B. (2016). Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, 32(4), 1151–1161.
- Fotiadis, A., Polyzos, S., & Huan, T.-C. T. (2021). The good, the bad and the ugly on COVID-19 tourism recovery. *Annals of Tourism Research*, 87, Article 103117.
- Goh, C., & Law, R. (2011). The methodological progress of tourism demand forecasting: A review of related literature. *Journal of Travel & Tourism Marketing*, 28(3), 296–317.
- Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1), 37–48.
- Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer Science & Business Media.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27, 1–22.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.
- Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1), 55–67.
- Jornal de Negócios. (2023) (accessed 2023a). Turismo pesa 8% do PIB <https://www.jornaldenegocios.pt/empresas/turismo-lazer/detalhe/turismo-pesa-8-do-pib>.
- Koehler, A. B., Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). A study of outliers in the exponential smoothing approach to forecasting. *International Journal of Forecasting*, 28(2), 477–484.
- Kourentzes, N., & Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, 75, 393–409.
- Kourentzes, N., Saayman, A., Jean-Pierre, P., Provenzano, D., Sahli, M., Seetaram, N., & Volo, S. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Africa team. *Annals of Tourism Research*, 88, Article 103197.
- Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2022). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, 73(5), 937–954.
- Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*, 37(4), 688–699.
- Li, Y., Yang, D., Guo, J., Sun, S., & Wang, S. (2023). Daily tourism demand forecasting before and during COVID-19: Data predictivity and an improved decomposition-ensemble framework. *Current Issues in Tourism*, 1–21.
- Liu, A., Kim, Y. R., & Song, H. (2022). Toward an accurate assessment of tourism economic impact: A systematic literature review. *Annals of Tourism Research Empirical Insights*, 3(2), Article 100054.
- 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, Article 103182.
- Lütkepohl, H., & Xu, F. (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics*, 42, 619–638.
- Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451–476.
- Park, S., Kim, Y. R., & Ho, C. S. T. (2022). Analysis of travel mobility under Covid-19: Application of network science. *Journal of Travel & Tourism Marketing*, 39(3), 335–352.
- Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, 45, 181–193.
- Pesaran, M. H., Pick, A., & Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2), 134–152.
- Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1), 134–161.
- Petropoulos, F., & Svetunkov, I. (2020). A simple combination of univariate models. *International Journal of Forecasting*, 36(1), 110–115.
- Provenzano, D., & Volo, S. (2022). Tourism recovery amid COVID-19: The case of Lombardy, Italy. *Tourism Economics*, 28(1), 110–130.
- 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, Article 103155.
- Rodrigues, P. M., & Gouveia, P. M. (2004). An application of PAR models for tourism forecasting. *Tourism Economics*, 10(3), 281–303.
- Saayman, A., & Botha, I. (2017). Non-linear models for tourism demand forecasting. *Tourism Economics*, 23(3), 594–613.
- Song, H., Qiu, R. T., & 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., Qiu, R. T., & Park, J. (2023). Progress in tourism demand research: Theory and empirics. *Tourism Management*, 94, Article 104655.
- Spiliotis, E., Assimakopoulos, V., & Makridakis, S. (2020). Generalizing the theta method for automatic forecasting. *European Journal of Operational Research*, 284(2), 550–558.
- Svetunkov, I., & Kourentzes, N. (2018). *Complex exponential smoothing for seasonal time series*. Technical report. Department of Management Science, Lancaster University.
- Teixeira, J. P., & Fernandes, P. O. (2014). Tourism time series forecast with artificial neural networks. *Têkhne*, 12(1–2), 26–36.
- Turismo de Portugal. (2023a). Prémios e distinções (accessed 2023) https://www.turismodeportugal.pt/en/quem_somos/Organizacao/Premios_Distincoes/Pages/default.aspx.
- Turismo de Portugal. (2023b). TravelBI - Turismo de Portugal. População Empregada - 2022 (accessed 2023b) <https://travelbi.turismodeportugal.pt/emprego/populacao-empregada-2022/>.
- UNWTO. (2023a). Impact assessment of the COVID-19 outbreak on international tourism (Accessed in 2023a) <https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism>.
- UNWTO. (2023b). Why tourism? (accessed 2023b) <https://www.unwto.org/why-tourism>.
- Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, 11(3), 447–475.
- WTTC. (2023). Global travel and tourism catapults into 2023, says wtcc (accessed 2023) <https://wtcc.org/news-article/global-travel-and-tourism-catapults-into-2023-says-wtcc>.
- Wu, X., & Blake, A. (2023). Does the combination of models with different explanatory variables improve tourism demand forecasting performance? *Tourism Economics*, 29(8), 2032–2056.
- Zhang, H., & Lu, J. (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, Article 103149.