ASIA PACIFIC JOURNAL OF TOURISM RESEARCH
2021, VOL. 26, NO. 12, 1350–1363
https://doi.org/10.1080/10941665.2021.1983623

Routledge
Taylor & Francis Group

Forecasting the demand for tourism using combinations of forecasts by neural network-based interval grey prediction models

Yi-Chung Hu

Department of Business Administration, Chung Yuan Christian University, Taoyuan City, Taiwan

ABSTRACT
In contrast to point forecasting, interval forecasting provides the degree of variation associated with forecasts. Accurate forecasting can help governments formulate policies for tourism, but little attention has been paid to interval forecasting of tourism demand. This study contributes to apply neural networks to develop interval models for tourism demand forecasting. Since combined forecasts are likely to improve the accuracy of point forecasting, forecast combinations are used to construct the proposed models. Besides, grey prediction models without requiring that data follow any statistical assumption serve as constituent models. Empirical results show that the proposed models outperform other considered interval models.

KEYWORDS
Tourism demand; interval forecasting; forecast combination; grey theory; neural network; artificial intelligence

1. Introduction

According to the World Travel and Tourism Council (2019), the travel and tourism sector was one of the world's largest in 2018, and generated 10.4% of the world's GDP. Accurate forecasts of the demand for tourism are important for improving development and investment strategies for travel and tourism, (Li et al., 2019; Sun et al., 2016). The appropriate policies can significantly increase investment in tourism by the private sector (Wu et al., 2017). Forecasts of the demand for tourism help industry practitioners, such as hotels and airlines, formulate financial and marketing strategies as well. In light of the increasing importance of forecasting the demand for tourism, research has focused on applying quantitative methods to generate accurate forecasts.

A majority of studies on forecasting the demand for tourism have preferred to focus on point forecasting over interval forecasting (Wu et al., 2017). In contrast to point forecasting, interval forecasting estimates a range of forecasts consisting of upper and lower bounds (UBs and LBs) (Chatfield, 2016), and can provide more information on the variation associated with point forecasts (Li et al., 2019). Even though interval forecasting provides more useful information to policymakers and industry practitioners that can help them formulate the relevant policies and strategies, little attention has been paid to it in the context of the demand for tourism (Li et al., 2019; Wu et al., 2017). To fill this gap in research, this study applies neural networks (NNs) to develop the interval prediction models on the basis of point forecasting. In light of the facilitation of improving accuracy of point forecasting by forecast combinations (Gunter & Önder, 2016; Osyczka, 2003), combinations of point forecasts are incorporated into interval models here. As the available data on tourism are not always compliant with the assumption of a specific statistical distribution, we use grey prediction models, which do not require that the data follow any statistical assumption (Dang et al., 2016; Liu et al., 2017) to generate individual point forecasts. We examine the forecasting performance of the proposed interval prediction models, with the mechanism of combinations of point forecasts, in terms of the inbound demand for tourism in Taiwan and Mainland China.

The remainder of this paper is organized as follows: Section 2 reviews the literature on point and interval

CONTACT Yi-Chung Hu ychu@cycu.edu.tw

forecasting, grey prediction, combinations of point forecasts, and measurements of forecasting accuracy. Section 3 focuses on the methodology by introducing four grey prediction models, the means to assess accuracy, the proposed interval grey prediction models, and data for empirical analysis. Section 4 presents the empirical results, and Section 5 offers conclusions of this study as well as directions for future work.

2. Literature review

2.1. Tourism point forecasting

Time series models, econometric approaches, and Artificial Intelligence (AI)-based methods are the most commonly used techniques to generate point forecasts (Song et al., 2019; Wu et al., 2017). Examples include the autoregressive integrated moving-average model (ARIMA) (Li et al., 2020; Tsui et al., 2014) for time series analyses, the autoregressive distributed lag model (Onafowora & Owoye, 2012), and models formulated using neural networks (NNs) (Claveria et al., 2015; Hu et al., 2019; Yao & Cao, 2020), deep learning (Law et al., 2019), and pattern recognition (Hu et al., 2021) in AI-based research.

Grey prediction has drawn significant research attention in the last decade because it can use a limited number of samples to generate details about unknown systems (Dang et al., 2016; Liu et al., 2017). Compared with statistical time series models, the conformation of data to any statistical assumptions is not required for grey prediction (Liu et al., 2017). Liou (2013), and Peng and Tzeng (2013) have noted that the statistical assumptions made in economics and statistics are often unrealistic. Despite the advantages of grey prediction, its application to forecasting the demand for tourism is limited, with the following exceptions: the grey-Markov model developed by Sun et al. (2016), the hybrid Markov modification model proposed by Wang (2004), the NN-based grey-Markov model developed by Hu et al. (2019), fractional grey models by Hu (2021), fourier residual modification by Huang and Lee (2011), predictions of tourism flow by Liu et al. (2014), and interval forecasts obtained using a nonlinear grey Bernoulli model (NGBM(1,1)), by Chen et al. (2019). Thus, we use grey prediction models to generate individual forecasts for combinations of forecasts, instead of statistical time series models.

2.2. Tourism interval forecasting

Although point and interval forecasting have their own specific functions for decision-making, there is limited research on interval forecasting of tourism demand. On the basis of point forecasting, the related studies for interval forecasting include a tourism forecasting competition using coverage probabilities by Athanasopoulos et al. (2011); the use of the bias-corrected bootstrap by Kim et al. (2010); the evaluation of alternative prediction intervals by Kim et al. (2011); and tourism demand elasticity by Song et al. (2010). The establishment of an interval model for these studies is dependent on the outcomes from single prediction models. Recently, Li et al. (2019) derived interval forecasts from the combined density in which a combined point forecast is used as the expected value. Their results reflect that not only combined interval models outperformed individual models on average, but it is feasible to develop interval forecasting models with point forecast combinations as well.

2.3. Tourism forecast combinations

Point forecast combinations based on certain weighting schemes are likely to improve the accuracy of forecasting the demand for tourism (Andrawis et al., 2011; Qiu et al., 2021; Shen et al., 2011; Song et al., 2019). For instance, Andrawis et al. (2011) proposed combinations of short- and long-term forecasts, and Cang (2014) combined NNs and support vector regression to this end. Combined econometric models were examined by Shen et al. (2011) as well.

Among the various weighting schemes for combinations of point forecasts, Wallis (2005) and Genre et al. (2013) showed that the simple average outperforms more complicated schemes. Because the historical performance of individual models should be taken into account in weighting schemes (Wu et al., 2017), the weighted average can be used to aggregate point forecasts instead of the simple average (Gunter & Önder, 2016). Song et al. (2009) have recommended combined forecasts for tourism by using statistical analysis. Empirical evidence has shown that none of single models outperforms others in all situations (Song & Li, 2008; Volchek et al., 2019; Wu et al., 2017). Forecast combinations can prevent the selection of a single, inappropriate forecasting model (Kascha & Ravazzolo, 2012). The above findings inspire us to apply combinations of point forecasts to develop interval forecasting models to predict the demand for tourism.

2.4. Forecasting performance measurement

To measure the performance of interval forecasting, both the rate of coverage rate and the width of the interval are usually considered (Li et al., 2019). The latter denotes the mean width of intervals at a certain confidence level, and the former denotes the rate of actual demand within the intervals. An ideal prediction interval is narrow, with a rate of coverage that approximates to the rate of nominal coverage. However, the wider is the prediction interval, the higher is the corresponding coverage rate.

To jointly access the interval width and coverage rate, Khosravi et al. (2010, 2011a) developed the coverage-length-based criterion (CLC) by multiplying the normalized mean length (NMPIL) by the coverage probability (PICP) of the intervals. The characteristic of the CLC is that the intervals of prediction with an unsatisfactorily high probability of coverage are heavily penalized. To avoid the unreasonableness arising from a zero width index resulting in a zero CLC, Quan et al. (2014) proposed the coverage width-based criterion (CWC) to better integrate the coverage rate with the width of the interval. The smaller the CWC is, the better is the quality of the constructed intervals.

Using an NN model with two output nodes, Khosravi et al. (2011b) further applied the CLC to develop a lower upper-bound estimation (LUBE) without assumptions about data or error distributions to generate prediction intervals. Quan et al. (2014) applied the LUBE with the CWC to forecast electrical load by optimizing the NN structure as well. Although we use the NN with two output nodes and the CWC to set up the proposed interval model, our method for training an NN is different from that of the LUBE.

3. Methodology

3.1. Individual forecasting models

In response to the literature on grey prediction models that are primarily used for forecasting the demand for tourism, the GM(1,1), NGBM(1,1), fractional GM(1,1), and fractional NGBM(1,1) models are used to generate individual point forecasts for the proposed interval models.

3.1.1. GM(1,1) model

Consider an original sequence $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$. Using the one-order accumulated generating operation (1-AGO), a new sequence, $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$, is expressed as

$$x_k^{(1)} = \sum_{j=1}^{k} x_j^{(0)},\ k = 1, 2, \ldots, n \tag{1}$$

For the demand for tourism with temporal fluctuations, the AGO helps determine developmental trends from historical data (Sun et al., 2016). The GM (1,1) is approximated by

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{2}$$

Its solution is expressed as

$$\hat{x}_k^{(1)} = \left(x_1^{(0)} - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a} \tag{3}$$

in which a and b are usually estimated by the ordinary least squares (OLS). The point forecast of $x_k^{(0)}$ ($k = 2, 3, \ldots, n$) is $\hat{x}_k^{(1)} - \hat{x}_{k-1}^{(1)}$ as

$$\hat{x}_k^{(0)} = (1 - e^{a})\left(x_1^{(0)} - \frac{b}{a}\right)e^{-a(k-1)} \tag{4}$$

3.1.2. NGBM(1,1) model

A one-order sequence in the NGBM(1,1) is approximated by the Bernoulli equation (Chen, 2008; Chen et al., 2008) as

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b(x^{(1)})^{b} \tag{5}$$

$\hat{x}_k^{(1)}$ (k = 1, 2, …, n) is

$$\hat{x}_k^{(1)} = \left[\left((x_1^{(0)})^{1-}b - \frac{b}{a}\right)e^{-a(k-1)(1-}b) + \frac{b}{a}\right]^{1/(1-}b) \tag{6}$$

where $\beta \neq 1$, and a and b can be measured by the OLS.

3.1.3. Fractional GM(1,1) model

The 1-AGO views each sample with the same degree of importance (Fang, 2020; Zeng & Meng, 2015). A fractional-order accumulation can relieve this restriction by applying a fractional parameter p ($0 < p < 1$) to fit the developing trends of a sequence (Wu et al., 2013, 2015). Let $GM^{p}(1,1)$ denote a fractional GM(1,1) with the p-AGO. A p-order sequence, $x^{(p)} = (x_1^{(p)}, x_2^{(p)},$

$\ldots, x_n^{(p)}$), is expressed by the following:

$$x_k^{(p)} = \sum_{i-1}^{k} \binom{k-i+p-1}{k-i} x^{(0)}(i),\ k = 1, 2, \ldots, n \quad (7)$$

where

$$\binom{k-i+p-1}{k-i} = \frac{(k-i+p-1)(k-i+p-2)\ldots(p+1)p}{(k-i)!} \quad (8)$$

$x^{(p)}$ is approximated by a p-order differential equation:

$$\frac{dx^{(p)}}{dt} + ax^{(p)} = b \quad (9)$$

$\hat{x}_k^{(p)}$ turns out to be

$$\hat{x}_k^{(p)} = \left(x_1^{(0)} - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a} \quad (10)$$

To obtain $\hat{x}_k^{(0)}$, the (1−p)-AGO is applied to ($\hat{x}_1^{(p)}, \hat{x}_2^{(p)}, \ldots, \hat{x}_n^{(p)}$), and the 1-IAGO is then applied to the resultant sequence.

3.1.4. Fractional NGBM(1,1) model

A fractional NGBM(1,1) (NGBM^p(1,1)) is extended from the GM^p(1,1). The aforementioned p-order sequence is approximated by

$$\frac{dx^{(p)}}{dt} + ax^{(p)} = b(x^{(p)})^{\beta} \quad (11)$$

where its solution is expressed as

$$\hat{x}_k^{(p)} = \left[\left((x_1^{(0)})^{1-}b - \frac{b}{a}\right)e^{-a(k-1)(1-}b) + \frac{b}{a}\right]^{1/(1-}b) \quad (12)$$

where $\beta \neq 1$, and both a and b are estimated by OLS.

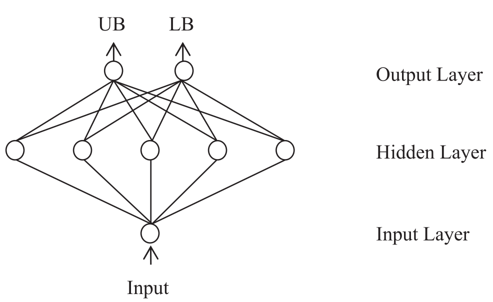

Figure 1. A schematic depiction of an NN for generating intervals.

3.2. Assessment of forecasting accuracy

In point forecasting, the linear interactive and general optimizer (LINGO) can be applied to construct optimized versions of individual prediction models, by minimizing the commonly-used mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\hat{x}_k^{(0)} - x_k^{(0)}}{x_k^{(0)}}\right| \times 100\% \quad (13)$$

In interval forecasting, the NMPIL and PICP are defined as follows:

$$\text{NMPIL} = \frac{1}{nR}\sum_{i=1}^{n}(u_i - l_i) \quad (14)$$

$$\text{PICP} = \frac{1}{n}\sum_{i=1}^{n} c_i \quad (15)$$

where u_i and l_i are UBs and LBs of $x_i^{(0)}$, respectively, and R is equal to the difference between the maximum and minimum real demand. In the PICP, $c_i = 1$ if $l_i \le x_i^{(0)} \le u_i$; otherwise, $c_i = 0$.

To impose higher penalties on wider intervals, Quan et al. (2014) provided the prediction interval normalized root-mean-square width (PINRW) to set up an interval model as

$$\text{PINRW} = \frac{1}{R}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(u_i - l_i)^2} \quad (16)$$

A comprehensive measure, namely the CWC, fuses informativeness (PINRW or NMPIL) with validity (PICP), and is formulated as follows (Quan et al., 2014):

$$\text{CWC} = \text{PIW} + e^{-\eta(\text{PICP}-\mu)} \quad (17)$$

where PIW is equal to the PINRW and NMPIL for training and testing, respectively. η and μ are two preassigned parameters that determine how sharply and where the function rises, and μ is usually considered to be the nominal coverage rate that is equivalent to the given confidence level, $(1-\alpha)$%. The CWC exponentially increases as PICP < μ regardless of the width of the interval.

An NN with two output nodes and one hidden layer is shown schematically in Figure 1. It is used to construct the proposed interval models. Given a data sequence, when a time point is presented to an NN, we can obtain its corresponding UB and LB in the output layer. The corresponding CWC can be obtained after all time points have been presented to an NN.

3.3. The proposed NN-based interval grey prediction models

3.3.1. Establishment of single interval models

For a grey point forecasting model, say M_j ($j = 1, 2, 3, 4$), we assume that n_1 training samples are used to construct a grey prediction model. Then, n point forecasts $\hat{x}^{(1)}_{j,1}, \hat{x}^{(1)}_{j,2}, \ldots, \hat{x}^{(0)}_{j,n_1}, \hat{x}^{(0)}_{j,n_1+1}, \ldots, \hat{x}^{(0)}_{j,n_1+n_2}$ ($j = 1, 2, 3, 4$) are obtained from the constructed M_j, where $n_1 + n_2 = n$. Because $\hat{x}^{(1)}_{j,1}, \hat{x}^{(1)}_{j,2}, \ldots, \hat{x}^{(0)}_{j,n_1}$ is not the main concern of decision-makers in tourism, a new sequence $\tilde{x}^{(0)}_j = (x^{(0)}_1, x^{(0)}_2, \ldots, x^{(0)}_{n_1}, \hat{x}^{(0)}_{j,n_1+1}, \ldots, \hat{x}^{(0)}_{j,n_1+n_2})$, comprising n_1 original samples and n_2 single-model point forecasts, is used to build an NN for M_j. We can easily compute the CWC with respect to $(x^{(0)}_1, x^{(0)}_2, \ldots, x^{(0)}_{n_1}, \hat{x}^{(0)}_{j,n_1+1}, \ldots, \hat{x}^{(0)}_{j,n_1+n_2})$ once the n required UBs and LBs have been obtained. An interval model with the minimum CWC is constructed by using the genetic algorithm (GA) to find the optimal weights of the connection of the NN.

3.3.2. Establishment of interval models with forecast combinations

The NN-based interval prediction model with combinations of point forecast is constructed using the following three primary steps:

(1) Generation of point forecasts: Build four optimized grey models by using LINGO to generate individual sequences of $\tilde{x}^{(0)}_1$, $\tilde{x}^{(0)}_2$, $\tilde{x}^{(0)}_3$, and $\tilde{x}^{(0)}_4$.
(2) Combination of point forecasts: Generate combined sequences through the weighted average.
(3) Construction of interval models: Minimize the CWC with respect to a combined sequence by using the GA to determine the optimal weights of the connection of the NN.

3.3.2.1. Combinations of point forecasts. Excluding the integration of four single-model point forecasts from four individual grey prediction models (four-model), six and four combinations of point forecasts are obtained from any given two (two-model) and three (three-model) individual models, respectively. Each combination has its own interval prediction model.

The reciprocal of the MAPEs are used to express the relative weights of the individual models. The larger the MAPE of a prediction model is, the less it contributes to the combined forecast. Assume that the relative weight of M_j is w_j ($j = 1, 2, 3, 4$). Then, for instance, a two-model combination with the weighted average at time t is obtained by combining the sequences of $\tilde{x}^{(0)}_1$ and $\tilde{x}^{(0)}_2$ as

$$\hat{x}^{(0)}_{1,2,t} = \frac{w_1}{w_1 + w_2}\hat{x}^{(0)}_{1,t} + \frac{w_2}{w_1 + w_2}\hat{x}^{(0)}_{2,t} \tag{18}$$

where $n_1+1 \le t \le n_1+n_2$. Then, an NN can be constructed for the combined sequence $\tilde{x}^{(0)}_{1,2} = (x^{(0)}_1, x^{(0)}_2, \ldots, x^{(0)}_{n_1}, \hat{x}^{(0)}_{1,2,n_1+1}, \ldots, \hat{x}^{(0)}_{1,2,n_1+n_2})$ with n_2 additional combined forecasts.

Likewise, a three-model combination at time t is obtained by combining the sequences of $\tilde{x}^{(0)}_1$, $\tilde{x}^{(0)}_2$, and $\tilde{x}^{(0)}_3$ as

$$\hat{x}^{(0)}_{1,2,3,t} = \frac{w_1}{w_1 + w_2 + w_3}\hat{x}^{(0)}_{1,t} + \frac{w_2}{w_1 + w_2 + w_3}\hat{x}^{(0)}_{2,t} + \frac{w_3}{w_1 + w_2 + w_3}\hat{x}^{(0)}_{3,t} \tag{19}$$

As a result, a combined sequence $\tilde{x}^{(0)}_{1,2,3} = (x^{(0)}_1, x^{(0)}_2, \ldots, x^{(0)}_{n_1}, \hat{x}^{(0)}_{1,2,3,n_1+1}, \ldots, \hat{x}^{(0)}_{1,2,3,n_1+n_2})$ is generated for such a combination. It is easy to derive $\hat{x}^{(0)}_{1,2,3,4,t}$ by combining the sequences of $\tilde{x}^{(0)}_1$, $\tilde{x}^{(0)}_2$, $\tilde{x}^{(0)}_3$, and $\tilde{x}^{(0)}_4$ at time t as

$$\hat{x}^{(0)}_{1,2,3,4,t} = \frac{1}{w_1 + w_2 + w_3 + w_4}\sum_{j=1}^{4} w_j \hat{x}^{(0)}_{j,t} \tag{20}$$

The CWC corresponding to a combined sequence can be computed once the n required UBs and LBs have been obtained. In the same way, we use the GA to construct an interval model with the minimum CWC.

3.3.2.2. Determination of UBs and LBs. The objective of the problem we address here is to minimize the CWC with respect to a given data sequence. The GA toolbox in MATLAB was used to discover the optimal weights of the connection of an NN with five hidden nodes in a hidden layer. With the CWC as the fitness function, the interval grey prediction model we want to determine is represented by a chromosome. The lower the fitness value is, the better the performance of the chromosome is. As a result, an optimal, combined model can be obtained from the best chromosomes with the minimum fitness value in successive generations. Although several parameters, such as population size, and the crossover and mutation rates may affect the performance of the GA, none of the optimal set of parameter specifications is available. Following the principles recommended by Osyczka (2003) and Ishibuchi et al. (2004), the total number of generations, population size, crossover rate, and mutation rate were set to 500, 100, 0.95, and 0.01, respectively.

3.4. Data description and modelling

We selected Mainland China as the tourism destination to verify the proposed method. Modelling international arrivals is becoming increasingly important for Mainland China, as it was the fourth most visited country by international tourists in 2019. Taiwan was chosen as well because in 2020, the US magazine *Global Traveler* listed Taiwan as the best leisure destination and the second-best adventure destination in Asia. This indicates that the future of the tourism market in Taiwan is promising.

The primary source markets for tourists to Taiwan are Japan, Hong Kong/Macau, Mainland China, South Korea, the US, and Southeast Asia. As listed in Table 1, annual data on the demand for inbound tourism to Taiwan from 2003 to 2017, collected from the tourism statistics database of the Taiwan Tourism Bureau (2021), were used to examine forecasting performance as well. Data from 2014 to 2017 were used for ex-post testing (i.e. $n_2 = 4$).

Japan, South Korea, Russia, the US, Malaysia, Philippines, Mongolia, and Singapore were the primary source markets for tourists to Mainland China. Annual data from 2003 to 2017 were collected from the China Economic and Social Data Platform, implemented by the China National Knowledge Infrastructure (2021), to examine forecasting performance, and are listed in Table 2. Data from 2014 to 2017 were used for ex-post testing.

We examined the forecasting performance of the proposed interval grey prediction models with different combinations of point forecasts. With $\eta = 5$, the average CWC for each NN-based prediction model was obtained from 10 independent trials. In addition to the combined density, several distinctive interval forecasting models that did not use combinations of point forecasts, including the LUBE, the grey number grey modification model (GGMM(1,1)) (Shih et al., 2011), and interval GM(1,1) (IGM(1,1)) (Chen et al., 2019), were considered.

(1) GGMM(1,1): Given $x_m^{(0)} = (x_m^{(0)}, x_{m+1}^{(0)}, \ldots, x_n^{(0)}) = (x_{m,1}^{(0)}, x_{m,2}^{(0)}, \ldots, x_{m,n-m+1}^{(0)})$ $(1 \le m \le n-3)$, $\hat{x}_{m,k}^{(0)}$ is expressed as

$$\hat{x}_{m,k}^{(0)} = (1 - e^{a_m})(x_n^{(1)} - \frac{b_m}{a_m})e^{-a_m(k-n)}, k = 2, 3, \ldots, n-m+1 \quad (21)$$

Because there are m sequences, $\hat{x}_{u,k}^{(0)}$ and $\hat{x}_{l,k}^{(0)}$ are $\max_{1..r}\{\hat{x}_{m,k}^{(0)}\}$ and $\min_{1..r}\{\hat{x}_{m,k}^{(0)}\}$, respectively, where $r = \min\{k, m\}$.

(2) IGM(1,1): $x_1^{(0)}, x_2^{(0)}, \ldots,$ and $x_n^{(0)}$ were used to create a linear regression line, and the upper wrapping sequence $x_{u'}^{(0)} = (x_{u',1}^{(0)}, x_{u',2}^{(0)}, \ldots, x_{u',n_1}^{(0)})$ and the lower wrapping sequence $x_{l'}^{(0)} = (x_{l',1}^{(0)}, x_{l',2}^{(0)}, \ldots, x_{l',n_2}^{(0)})$ could then be generated according to the signs of individual residuals, where $n_1 + n_2 = n$. To set-up the interval GM(1,1), the GM(1,1) used $x_{u'}^{(0)}$ and $x_{l'}^{(0)}$ to generate the upper and lower bounds, respectively.

(3) Combined density: The density functions of four individual grey prediction models were assumed to conform to a normal distribution centred on one associated point forecast, with the variance estimated by past forecasting errors. For instance, the residuals of $\hat{x}_{i,1}^{(0)}, \hat{x}_{i,2}^{(0)}, \ldots, \hat{x}_{i,t-1}^{(0)}$ contributed to estimate the variance in the density forecast at time t.

Because there were four individual density forecasts of a random variable, denoted by y_t, a combined density forecast at time t $(n_1 + 1 \le t \le n_2)$ is expressed as

$$f_t(y_t) = \sum_{i=1}^{4} w_i f_{it}(y_t) \quad (22)$$

where $f_{it}(y_t)$ is a density forecast from M_i $(i = 1, 2, 3, 4)$, $w_i \ge 0$, and sums to one. Then, the mean and the variance of $f_t(y_t)$ is formulated as following:

$$E(y_t) = \hat{y}_t = \frac{1}{4}\sum_{i=1}^{4} \hat{x}_{i,t}^{(0)} \quad (23)$$

$$Var(y_t) = \frac{1}{4}\sum_{i=1}^{4} \sigma_{i,t}^2 + \frac{1}{4}\sum_{i=1}^{4} (\hat{x}_{i,t}^{(0)} - \hat{y}_t)^2 \quad (24)$$

where $\hat{x}_{i,t}^{(0)}$ is the point forecast that represents the mean value of $f_{it}(y_t)$, and $\sigma_{i,t}^2$ denotes the variance of $\hat{x}_{i,1}^{(0)}, \hat{x}_{i,2}^{(0)}, \ldots, \hat{x}_{i,t-1}^{(0)}$. Let Q be the quantile function such that $(Q(q/2), Q(1-q/2))$ denotes a combined interval with confidence level p, in which $q = 1-p$. The confidence interval is specified as 80% $(\mu = 0.8)$, following Li et al. (2019).

4. Empirical results

4.1. Inbound demand for tourism in Taiwan (Case I)

The results of the interval widths and coverage rates for the individual interval prediction models are shown in Figure 2. For the sake of simplicity, each

Table 1. Historical annual foreign tourists from six major markets to Taiwan (unit: persons).

Year	Japan	Hong Kong /Macau	Korea	Mainland China	USA	Southeast Asia
2003	657053	323178	92893		272858	457103
2004	887311	417087	148095		382822	568269
2005	1124334	432718	182517		390929	636925
2006	1161489	431884	196260		394802	643338
2007	1166380	491437	225814		397965	700287
2008	1086691	618667	252266	329204	387197	725751
2009	1000661	718806	167641	972123	369258	689027
2010	1080153	794362	216901	1630735	395729	911174
2011	1294758	817944	242902	1784185	412617	1071975
2012	1432315	1016356	259089	2586428	411416	1132592
2013	1421550	1183341	351301	2874702	414060	1261596
2014	1634790	1375770	527684	3987152	458691	1388305
2015	1627229	1513597	658757	4184102	479452	1425485
2016	1895702	1614803	884397	3511734	523888	1653908
2017	1898854	1692063	1054708	2732549	561365	2137138

interval width was set to a unit of 10^4 people. The coverage rates of the proposed interval models that use combined forecasts were higher than the average coverage rates of single interval grey models. Furthermore, there was a trade-off between the rate of coverage and interval width. For instance, the combined density had a higher coverage rate than that of the proposed interval models, but it had a wider interval. This is why a reasonable and comprehensive measure —namely, the CWC—was applied to evaluate forecasting performance. A lower CWC represented a better interval forecast.

As is shown in Table 3, no single interval grey prediction model outperformed the others in any source market. The LUBE, GGMM(1,1), and IGM(1,1) performed worse than the NN-based interval models with forecast combinations. In most cases, the accuracy of the individual NN-based interval models with forecast combinations was lower than that of single interval grey models across the six source markets. The average CWCs of the forecasting performance of the NN-based single interval grey models could be improved by fusing several point forecasts. Altogether, the forecasting accuracy of the proposed models with combined forecasts was superior to those of the other interval models considered.

4.2. Inbound demand for tourism in Mainland China (Case II)

The coverage rates and interval widths of the individual models are shown in Figure 3. The average coverage rates of the proposed models that used combined forecasts were higher than those of single interval prediction models. Table 4 shows that no single interval grey model outperformed the others in any source market. As in Case I, the accuracy of a majority of the individual NN-based interval models with forecast combinations was lower than that of NN-based single interval grey models across all eight source markets.

To measure overall prediction performance, the average CWCs of individual interval prediction models were considered across 14 data sequences from Taiwan and Mainland China, as shown in Figure 4. The average accuracy of the individual NN-based interval models with point forecast combinations were lower than that of the other interval models considered. The superiority of the proposed interval models over LUBE indicates that a supplement of the future central tendency can help improve the forecasting performance of LUBE.

Table 5 further shows the worst forecasts (i.e. the worst CWC) for each combination of the proposed interval models. A majority of the proposed interval models with combined forecasts were better than the single interval grey prediction models across the 14 source markets. Therefore, the proposed models can help decision-makers without knowledge or experience of the forecasting capability of individual interval prediction models. This result is consistent with that reported by Li et al. (2019), who studied the inbound the demand for tourism for Hong Kong.

4.3. Assessment of point forecasts

To assess the forecasting performance of the proposed interval models with combinations of point forecast, several widely used time series models, including the ARIMA, exponential smoothing (ES), structural time series (STS), and multi-layer perceptron (MLP), were used. Tables 6 and 7 list the results as

Table 2. Historical annual foreign tourists from eight major markets to Mainland China (unit: ten thousand persons).

Year	Korea	Japan	Russia	USA	Malaysia	Mongolia	Philippines	Singapore
2003	194.55	225.48	138.07	82.25	43.01	41.83	45.77	37.81
2004	284.49	333.48	179.22	130.86	74.19	55.38	54.94	63.68
2005	354.53	339.00	222.39	155.55	89.96	64.20	65.40	75.59
2006	392.40	374.59	240.51	171.03	91.06	63.12	70.42	82.79
2007	477.71	397.75	300.39	190.12	106.20	68.20	83.30	92.20
2008	396.04	344.61	312.34	178.64	104.05	70.53	79.53	87.58
2009	319.75	331.75	174.30	170.98	105.90	57.67	74.89	88.95
2010	407.64	373.12	237.03	200.96	124.52	79.44	82.83	100.37
2011	418.54	365.82	253.63	211.61	124.51	99.42	89.43	106.30
2012	406.99	351.82	242.61	211.81	123.55	101.05	96.20	102.77
2013	396.90	287.80	218.60	208.50	120.70	105.00	99.70	96.70
2014	418.20	271.80	204.60	209.30	113.00	108.30	96.80	97.10
2015	444.44	249.77	158.23	208.58	107.55	101.41	100.40	90.53
2016	477.53	258.99	197.66	224.96	116.54	158.12	113.51	92.46
2017	386.38	268.30	235.68	231.29	123.32	186.45	116.85	94.12

evaluated by the MAPE and root mean-squared error (RMSE) for four interval and four point forecasting models for the source markets considered, respectively.

Both the average MAPEs and the RMSEs of the combined density models and the proposed interval models with forecast combinations were lower than those of the ARIMA, ES, STS, and MLP. The average MAPEs of the proposed model with two- and three-model combinations of forecasts were 15.077% and 14.954%, respectively, lower than those of the point forecasting models considered. The development of interval models with combinations of point forecast reflected the superiority in the provision of the central tendency in terms of accuracy of forecasting over the point forecasting models considered. Of course, interval and point forecasting cannot be used as substitutes for each other; instead, the use of both can provide comprehensive and useful information for decision-making (Li et al., 2019).

4.4. Statistical analysis

The forecasting performance of the different interval prediction models was compared on the 14 datasets by using the Friedman test. This led us to reject the null hypothesis, whereby the considered interval models had equal average ranks, because the Friedman statistic of 10.257 exceeded the critical value of $F(7, 91)$ at a 10% significance *level*.

We then used the post-hoc Nemenyi test to compare the differences in average rank between

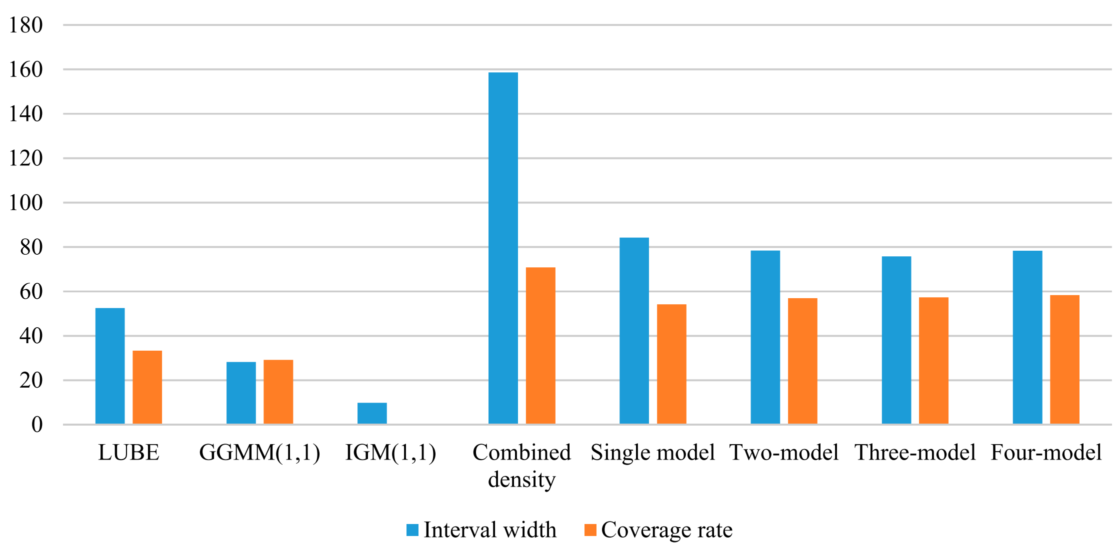

Figure 2. The interval widths and the coverage rates for individual interval prediction models (Case I).

Table 3. CWC for individual interval prediction models (Case I).

	Japan	HK/Macau	Korea	Mainland China	USA	Southeast Asia	Average
LUBE	4.864	4.692	54.914	16.473	16.504	4.923	17.062
GGMM(1,1)	4.861	1.734	55.305	54.849	54.900	4.663	29.385
IGM(1,1)	55.069	54.624	54.748	54.624	54.742	54.693	54.750
Combined density	1.117	1.423	55.208	1.751	16.421	5.646	13.594
NN-based single-model							
GM(1,1)	5.133	1.137	55.170	2.290	5.541	2.112	
NGBM(1,1)	5.144	1.118	55.332	5.417	55.557	2.013	
$GM^P(1,1)$	5.171	1.208	55.167	5.426	5.883	2.064	
$NGBM^P(1,1)$	5.087	1.034	55.222	5.285	5.860	2.111	
Average	5.134	1.125	55.223	4.605	18.210	2.075	14.395
NN-based two-model	4.135	1.110	55.189	4.116	7.316	1.677	12.257
NN-based three-model	4.375	1.108	55.207	4.585	10.648	2.036	12.993
NN-based four-model	1.990	1.292	55.294	5.436	5.522	2.033	11.928

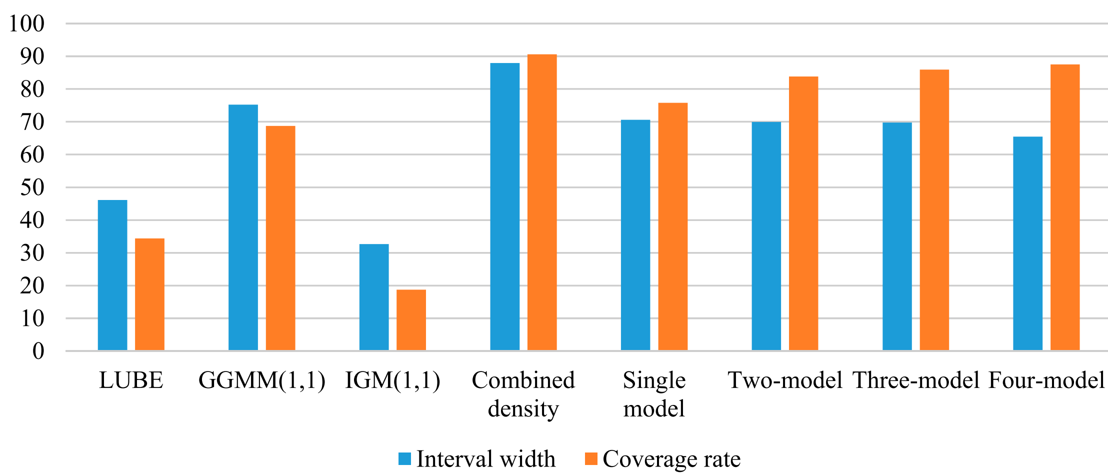

Figure 3. The interval widths and the coveragike rates for individual interval prediction models (Case II).

Table 4. CWCs for individual interval prediction models (Case II).

	Korea	Japan	Russia	USA	Malaysia	Mongolia	Philippines	Singapore	Average
LUBE	4.717	55.202	4.870	15.938	54.838	5.003	0.797	54.840	24.526
GGMM(1,1)	4.805	1.813	1.053	0.678	4.862	54.993	4.618	0.932	9.219
IGM(1,1)	54.623	16.249	54.936	1.501	54.895	15.844	54.599	16.021	33.584
Combined density	1.865	1.924	1.867	1.159	1.157	2.205	1.191	1.106	1.559
NN-based single-model									
GM(1,1)	1.673	16.235	1.896	1.088	16.286	5.324	0.674	16.192	
NGBM(1,1)	4.938	0.958	0.940	1.011	0.899	5.105	0.616	1.051	
$GM^P(1,1)$	1.727	0.969	1.002	0.938	0.939	5.344	0.586	0.938	
$NGBM^P(1,1)$	15.852	1.910	1.779	0.963	0.960	5.359	0.638	0.930	
Average	6.047	5.018	1.404	1.000	4.771	5.283	0.629	4.778	3.616
NN-based two-model	3.278	1.457	1.387	0.954	1.965	5.215	0.613	0.926	1.974
NN-based three-model	1.688	1.647	1.610	0.925	1.045	5.199	0.628	0.926	1.709
NN-based four-model	1.867	1.042	1.799	0.969	0.889	5.083	0.830	0.867	1.668

the interval models with the critical difference at a 10% significance *level*. Table 8 shows the results of the test as well as the differences between the proposed models and the other interval models considered. We can see that, for instance, the difference in average rank between the proposed NN-based interval model with three-model forecast combinations and the IGM (1,1) was 4.464. The results can be summarized as follows:

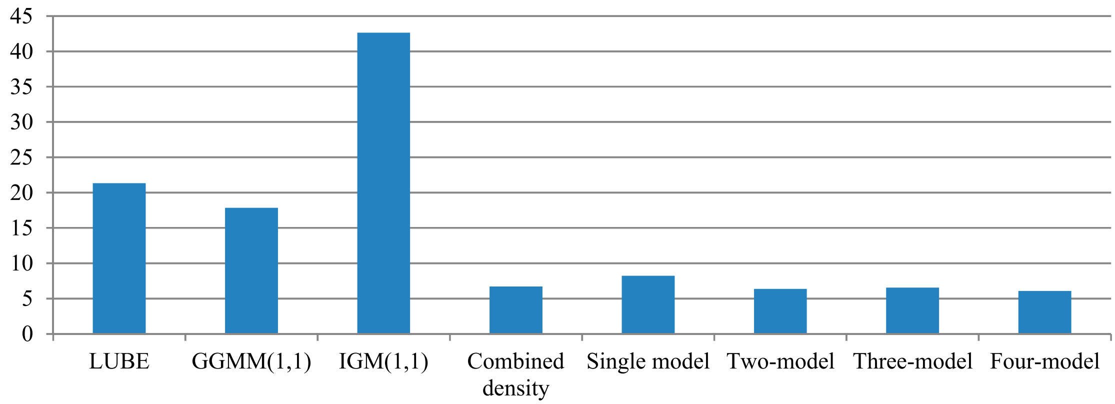

Figure 4. The average CWCs for individual interval prediction models.

Table 5. Highest CWCs for the proposed interval prediction models.

Case I	Japan	HK/ Macau	Korea	Mainland China	USA	Southeast Asia	Average		
NN-based single-model	5.171	1.137	55.332	5.426	55.557	2.112	20.789		
NN-based two-model	5.260	1.126	55.275	16.658	5.592	1.991	14.317		
NN-based three-model	5.168	1.175	55.272	16.581	5.443	2.099	14.292		
NN-based four-model	1.990	1.292	55.294	5.436	5.522	2.033	11.928		
Case II	**Korea**	**Japan**	**Russia**	**USA**	**Malaysia**	**Mongolia**	**Philippines**	**Singapore**	**Average**
NN-based single-model	15.852	16.235	1.896	1.088	16.286	5.359	0.674	16.192	9.198
NN-based two-model	4.896	1.980	1.817	1.044	5.127	5.299	0.652	0.998	2.727
NN-based three-model	1.697	1.916	1.863	1.042	1.162	5.239	0.677	0.972	1.821
NN-based four-model	1.867	1.042	1.799	0.969	0.889	5.083	0.830	0.867	1.668

Table 6. MAPEs for point forecasting results.

	Combined density	NN-based two-model	NN-based three-model	NN-based four-model	ARIMA	ES	STS	MLP
Case I								
Japan	10.933	11.140	10.865	10.856	19.527	22.245	20.446	17.601
HK/Macau	8.826	9.016	9.079	9.110	3.952	32.658	2.730	23.852
Korea	54.562	54.562	54.563	54.563	38.539	59.269	50.306	59.636
Mainland China	29.074	26.521	25.706	25.422	31.792	30.155	30.656	17.579
USA	16.740	16.618	16.598	16.589	16.374	18.477	11.984	20.797
SA	5.625	5.767	5.755	5.703	9.064	28.410	8.234	18.830
Case II								
Korea	8.480	5.846	6.744	8.680	8.482	8.462	7.830	9.578
Japan	14.510	14.132	14.024	13.972	51.229	23.109	5.376	35.282
Russia	15.377	15.606	15.087	14.722	17.818	19.390	25.840	18.064
USA	3.106	3.648	3.391	3.416	8.121	5.072	5.747	4.404
Malaysia	16.077	15.469	15.309	15.234	6.589	6.326	19.591	19.258
Mongolia	16.420	16.410	16.407	16.406	18.310	23.765	18.117	19.706
Philippines	3.965	4.325	4.079	4.017	4.625	10.566	4.357	12.358
Singapore	13.177	12.012	11.746	11.628	12.856	5.946	14.720	7.080
Average	15.491	15.077	14.954	15.023	17.663	20.989	16.138	20.288

(1) The average CWC of the individual interval models with combinations of point forecasts were significantly lower than the average CWC of the four single-interval grey models.

(2) The proposed interval models with forecast combinations significantly outperformed the LUBE, GGMM(1,1), and IGM(1,1). The average CWC of the proposed single interval models

Table 7. RMSEs for point forecasting results.

	Combined density	NN-based two-model	NN-based three-model	NN-based four-model	ARIMA	ES	STS	MLP
Case I								
Japan	21.061	20.980	20.942	20.922	23.379	26.902	23.863	18.757
HK/Macau	19.589	20.031	20.144	20.202	8.781	52.566	4.480	38.245
Korea	47.823	47.823	47.824	47.824	34.947	52.549	45.266	52.555
Mainland China	108.757	102.542	96.475	93.584	122.347	128.268	117.507	69.359
USA	9.064	9.187	9.187	9.189	9.281	10.390	6.812	11.476
SA	13.816	13.792	13.630	13.551	27.308	58.501	24.781	44.399
Case II								
Korea	44.985	50.967	53.234	50.183	41.372	45.673	40.164	54.938
Japan	40.128	39.647	39.159	38.914	152.503	60.854	15.513	92.530
Russia	33.595	35.421	33.894	33.165	38.642	42.091	52.811	38.436
USA	7.172	10.212	8.392	7.816	22.674	15.146	13.025	10.903
Malaysia	18.701	18.131	17.892	17.778	8.586	8.222	22.870	22.291
Mongolia	26.502	26.515	26.512	26.510	39.197	52.950	37.397	42.811
Philippines	4.814	5.074	4.876	4.777	5.038	14.603	5.069	13.252
Singapore	12.625	11.560	11.283	11.159	13.634	6.003	14.899	9.153
Average	29.188	29.420	28.817	28.255	39.121	41.051	30.318	37.079

Table 8. Results of the nonparametric Friedman test with a post-hoc Nemenyi test.

Interval model	Average rank	Friedman statistic	Comparison group	Nemenyi test
LUBE	5.714	10.257 > $F(7, 91)$	LUBE vs. NN1	0.643
GGMM(1,1)	5.5		LUBE vs. NN2	3.107 *
IGM(1,1)	7.143		LUBE vs. NN3	3.036 *
NN1	5.071		LUBE vs. NN4	2.5 *
NN2	2.607		GGMM(1,1) vs. NN1	0.429
NN3	2.679		GGMM(1,1) vs. NN2	2.893 *
NN4	3.214		GGMM(1,1) vs. NN3	2.821 *
Combined density	4.071		GGMM(1,1) vs. NN4	2.286 *
			IGM(1,1) vs. NN1	2.071 *
			IGM(1,1) vs. NN2	4.536 *
			IGM(1,1) vs. NN3	4.464 *
			IGM(1,1) vs. NN4	3.929 *
			Combined density vs. NN1	−1
			Combined density vs. NN2	1.464
			Combined density vs. NN3	1.393
			Combined density vs. NN4	0.857
			NN1 vs. NN2	2.464 *
			NN1 vs. NN3	2.392 *
			NN1 vs. NN4	1.857 *

Note: NN1, NN2, NN3, and NN4 denote NN-based single-, two-, three-, and four models, respectively.
*A significant statistical difference with the critical difference of 1.784 at a 10% significance *level*.

significantly outperform that of the IGM(1,1) as well.

(3) No significant difference in average rank was noted between the proposed interval models with combinations of point forecasts and the combined density models. However, the proposed interval models outperformed the combined density models in terms of average rank.

(4) No significant difference in average ranks was noted among the two-, three-, and four-model combinations. Therefore, we cannot conclude whether forecast combinations with more models can outperform those with fewer models as well. The empirical results that used four prediction models reported by Song et al. (2009) support this finding.Note: NN1, NN2, NN3, and NN4 denote NN-based single-, two-, three-, and four models, respectively.* A significant statistical difference with the critical difference of 1.784 at a 10% significance *level*.

5. Discussion and conclusions

This study developed NN-based interval grey prediction models for predicting the demand for tourism.

We also investigated whether the proposed interval models with combined point forecasts could outperform single interval grey prediction models. Four grey point forecasting models were used to generate point forecasts and interval prediction models: the GM(1,1), NGBM(1,1), fractional GM(1,1), and fractional NGBM(1,1). We examined all possible combinations of two to four grey prediction models. The forecasting performance, as measured by the CWC, was examined based on the demand for tourism in Taiwan and China from main source markets.

No single-interval model outperformed any other in the two real cases considered. The proposed interval models with combinations of point forecasts were significantly better than those interval models that did not use combinations of point forecasts—the LUBE, GGMM(1,1), and IGM(1,1). The average forecasting performance of the four single-interval grey models can be significantly improved by incorporating combinations of point forecasts into them. Besides, the individual grey prediction models selected to be the constituents belong to the univariate time series models. Such models do not consider independent variables that affect demand such that they can avoid the adverse impact of incomplete information associated with explanatory factors (Cang, 2014). Thus, the proposed interval models did not require any statistical assumption regarding the statistical time series models, and avoid the difficulty of determining relevant variables that affect the demand for tourism.

The LUBE and the proposed interval models were common in that they used the NN structures with two output nodes. However, the main difference was that the LUBE simply used $(x_1^{(0)}, x_2^{(0)}, \ldots, x_{n_1}^{(0)})$ with n_1 historical samples as target values to set-up the NN, and then generated variations of n_2 samples. The proposed interval models used not only $(x_1^{(0)}, x_2^{(0)}, \ldots, x_{n_1}^{(0)})$, but also n_2 additional point or combined forecasts that reflected the central tendencies in the future. Since the proposed interval models with forecast combination significantly outperformed the LUBE, we can conclude that the addition of combined forecasts to the sequence of n_1 historical samples effectively improved the forecasting performance of the LUBE. The addition of combined forecasts may help the proposed model catch more estimation of future variation.

For decision-makers, especially those who have no knowledge of interval models, this study provides useful findings that can help avoid mistakes when choosing individual models. Wu et al. (2017) also claimed that the use of combined forecasting techniques is less risky than any single model. Here, the Mann–Whitney U test was used to determine whether the average CWCs of the individual NN-based interval models with different combinations for Case II were lower than those for Case I. This led us to reject the null hypothesis at a 10% significance *level* because the statistics for two-, three-, and four-model combinations were 0.0063, 0.0147, and 0.026, respectively, all lower than 0.1. Even though only two tourism destinations were considered, the results reflect the usefulness of the proposed interval models for forecasting the demand for tourism to Mainland China.

In spite of the above, this study has a few limitations that should be addressed in future work. First, although only grey prediction models were considered for point forecasting, other models, such as the ARIMA, vector autoregressive model, and back-propagation NN, should be examined in this context as well. Second, future research should explore weighting schemes other than the weighted average, such as the variance–covariance method, the multiple forecast encompassing *F*-test recommended by Gunter and Önder (2016) and Shen et al. (2011), as well as their effects on performance in terms of interval forecasting. Finally, we addressed the development of prediction models with a single variable such that exogenous variables, such as the GDP and the price of oil, were not considered in gauging the demand for tourism. Because multi-variate models are expected to yield better forecasting performance than single-variable models (Hu, 2020; Ma et al., 2019; Wu et al., 2018; Xie et al., 2020), multi-variate grey prediction models should be considered as constituents of the proposed interval models.

Acknowledgements

This research is supported by the Ministry of Science and Technology, Taiwan, under grant MOST 108-2410-H-033 -038-MY2.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Ministry of Science and Technology, Taiwan: [Grant Number MOST 108-2410-H-033 -038-MY2].

ORCID

Yi-Chung Hu http://orcid.org/0000-0003-3090-1515

References

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, *27*(3), 870–886. https://doi.org/10.1016/j.ijforecast.2010.05.019

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844. https://doi.org/10.1016/j.ijforecast.2010.04.009

Cang, S. (2014). A comparative analysis of three types of tourism demand forecasting models: Individual, linear combination and non-linear combination. *International Journal of Tourism Research*, *16*(6), 596–607. https://doi.org/10.1002/jtr.1953

Chatfield, C. (2016). *The analysis of time series: An introduction*. CRC Press.

Chen, C. I. (2008). Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate. *Chaos, Solitons and Fractals*, *37*(1), 278–287. https://doi.org/10.1016/j.chaos.2006.08.024

Chen, C. I., Chen, H. L., & Chen, S. P. (2008). Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear grey Bernoulli model NGBM(1,1). *Communications in Nonlinear Science and Numerical Simulations*, *13*(6), 1194–1204. https://doi.org/10.1016/j.cnsns.2006.08.008

Chen, Y. Y., Liu, H. T., & Hsieh, H. L. (2019). Time series interval forecast using GM(1,1) and NGBM(1, 1) models. *Soft Computing*, *23*(5), 1541–1555. https://doi.org/10.1007/s00500-017-2876-0

China National Knowledge Infrastructure. (2021). China Economic and Social Data Platform. Retrieved May 1, 2021, from http://Data.Oversea.Cnki.Net/Chn

Claveria, O., Monte, E., & Torra, S. (2015). Tourism demand forecasting with neural network models: Different ways of treating information. *International Journal of Tourism Research*, *17*(5), 492–500. https://doi.org/10.1002/jtr.2016

Dang, Y. G., Wang, Z. X., Qian, W. Y., & Xiong, P. P. (2016). *Grey prediction techniques and methods*. Science Press.

Fang, J. (2020). Prediction and analysis of regional economic income multiplication capability based on fractional accumulation and integral model. *Chaos, Solitons & Fractals*, *130*, 109441. https://doi.org/10.1016/j.chaos.2019.109441

Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, *29*(1), 108–121. https://doi.org/10.1016/j.ijforecast.2012.06.004

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

Hu, M., Qiu, R. T. R., Wu, D. C., & Song, H. (2021). Hierarchical pattern recognition for tourism demand forecasting. *Tourism Management*, *84*, 104263. https://doi.org/10.1016/j.tourman.2020.104263

Hu, Y. C. (2020). A multivariate grey prediction model with grey relational analysis for bankruptcy prediction problems. *Soft Computing*, *24*(6), 4259–4268. https://doi.org/10.1007/s00500-019-04191-0

Hu, Y. C. (2021). Forecasting tourism demand using fractional grey prediction models with Fourier series. *Annals of Operations Research*, *300*(2), 467–491. https://doi.org/10.1007/s10479-020-03670-0

Hu, Y. C., Jiang, P., & Lee, P. C. (2019). Forecasting tourism demand by incorporating neural networks into grey-Markov models. *Journal of the Operational Research Society*, *70*(1), 12–20. https://doi.org/10.1080/01605682.2017.1418150

Huang, Y. L., & Lee, Y. H. (2011). Accurately forecasting model for the stochastic volatility data in tourism demand. *Modern Economy*, *2*(5), 823–829. https://doi.org/10.4236/me.2011.25091

Ishibuchi, H., Nakashima, T., & Nii, M. (2004). *Classification and modeling with linguistic information granules: Advanced approaches to linguistic data mining*. Springer.

Kascha, C., & Ravazzolo, F. (2012). Combining inflation density forecasts. *Journal of Forecasting*, *29*(1-2), 231–250. https://doi.org/10.1002/for.1147

Khosravi, A., Nahavandi, S., & Creighton, D. (2010). Construction of optimal prediction intervals for load forecasting problems. *IEEE Transactions on Power Systems*, *25*(3), 1496–1503. https://doi.org/10.1109/TPWRS.2010.2042309

Khosravi, A., Nahavandi, S., & Creighton, D. (2011a). Prediction interval construction and optimization for adaptive neuro-fuzzy inference systems. *IEEE Transactions on Fuzzy Systems*, *19*(5), 983–988. https://doi.org/10.1109/TFUZZ.2011.2130529

Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011b). Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Transactions on Neural Networks*, *22*(3), 337–346. https://doi.org/10.1109/TNN.2010.2096824

Kim, J. H., Song, H., & Wong, K. K. F. (2010). Bias-corrected bootstrap prediction intervals for autoregressive model: New alternatives with applications to tourism forecasting. *Journal of Forecasting*, *29*(7), 655–672. https://doi.org/10.1002/for.1150

Kim, J. H., Wong, K., Athanasopoulos, G., & Liu, S. (2011). Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals. *International Journal of Forecasting*, *27*(3), 887–901. https://doi.org/10.1016/j.ijforecast.2010.02.014

Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, *75*, 410–423. https://doi.org/10.1016/j.annals.2019.01.014

Li, G., Wu, C., Zhou, M., & Liu, A. (2019). The combination of interval forecasts in tourism. *Annals of Tourism Research*, *75*, 363–378. https://doi.org/10.1016/j.annals.2019.01.010

Li, H., Hu, M., & Li, G. (2020). Forecasting tourism demand with multisource big data. *Annals of Tourism Research*, *83*, 102912. https://doi.org/10.1016/j.annals.2020.102912

Liou, J. J. H. (2013). New concepts and trends of MCDM for tomorrow – in honor of professor Gwo-Hshiung Tzeng on the occasion of his 70th birthday. *Technological and Economic Development of Economy*, *19*(2), 367–375. https://doi.org/10.3846/20294913.2013.811037

Liu, S., Yang, Y., & Forrest, J. (2017). *Grey data analysis: Methods, models and applications*. Springer.

Liu, X., Peng, H., Bai, Y., Zhu, Y., & Liao, L. (2014). Tourism flows prediction based on an improved grey GM(1,1) model. *Procedia-Social and Behavioral Sciences*, *138*, 767–775. https://doi.org/10.1016/j.sbspro.2014.07.256

Ma, X., Liu, Z., & Wang, Y. (2019). Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China. *Journal of Computational and Applied Mathematics*, *347*, 84–94. https://doi.org/10.1016/j.cam.2018.07.044

Onafowora, O. A., & Owoye, O. (2012). Modelling international tourism demand for the Caribbean. *Tourism Economics*, *18*(1), 159–180. https://doi.org/10.5367/te.2012.0102

Osyczka, A. (2003). *Evolutionary algorithms for single and multicriteria design optimization*. Physica-Verlag.

Peng, K. H., & Tzeng, G. H. (2013). A hybrid dynamic MADM model for problems-improvement in economics and business. *Technological and Economic Development of Economy*, *19*(4), 638–660. https://doi.org/10.3846/20294913.2013.837114

Qiu, R. T. R., Wu, D. C., Dropsy, V., Petit, S., Pratt, S., & Ohe, Y. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and pacific team. *Annals of Tourism Research*, *88*, 103155. https://doi.org/10.1016/j.annals.2021.103155

Quan, H., Srinivasan, D., & Khosravi, A. (2014). Uncertainty handling using neural network-based prediction intervals for electrical load forecasting. *Energy*, *73*, 916–925. https://doi.org/10.1016/j.energy.2014.06.104

Shen, S., Li, G., & Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, *38*(1), 72–89. https://doi.org/10.1016/j.annals.2010.05.003

Shih, C. S., Hsu, Y. T., Yeh, J., & Lee, Y. P. (2011). Grey number prediction using the grey modification model with progression technique. *Applied Mathematical Modelling*, *35*(3), 1314–1321. https://doi.org/10.1016/j.apm.2010.09.008

Song, H., Kim, J. H., & Yang, S. (2010). Confidence intervals for tourism and hospitality demand elasticity. *Annals of Tourism Research*, *37*(2), 377–396. https://doi.org/10.1016/j.annals.2009.10.002

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting – A review of recent research. *Tourism Management*, *29*(2), 203–220. https://doi.org/10.1016/j.tourman.2007.07.016

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

Song, H., Witt, S. F., Wong, K. F., & Wu, D. C. (2009). An empirical study of forecast combination in tourism. *Journal of Hotel & Tourism Research*, *33*(1), 3–29. https://doi.org/10.1177/1096348008321366

Sun, X., Sun, W., Wang, J., & Gao, Y. (2016). Using a grey-Markov model optimized by cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. *Tourism Management*, *52*, 369–379. https://doi.org/10.1016/j.tourman.2015.07.005

Taiwan Tourism Bureau. (2021). Tourism Statistics Database of the Taiwan Tourism Bureau. Retrieved May 1, 2021, from https://stat.taiwan.net.tw/inboundSearch

Tsui, W. H. K., Balli, H. O., Gilbey, A., & Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, *42*, 62–76. https://doi.org/10.1016/j.tourman.2013.10.008

Volchek, K., Liu, A., Song, H., & Buhalis, D. (2019). Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tourism Economics*, *25*(3), 425–447. https://doi.org/10.1177/1354816618811558

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, *67*(s1), 983–994. https://doi.org/10.1111/j.1468-0084.2005.00148.x

Wang, C. H. (2004). Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management*, *25*(3), 367–374. https://doi.org/10.1016/S0261-5177(03)00132-8

World Travel and Tourism Council. (2019). The Economic Impact of Travel of Travel & Tourism, World Travel and Tourism Council, London.

Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*, *29*(1), 507–529. https://doi.org/10.1108/IJCHM-05-2015-0249

Wu, L., Gao, X., Xiao, Y., Yang, Y., & Chen, X. (2018). Using a novel multi-variable grey model to forecast the electricity consumption of shandong province in China. *Energy*, *157*, 327–335. https://doi.org/10.1016/j.energy.2018.05.147

Wu, L., Liu, S., Fang, Z., & Xu, H. (2015). Properties of the GM(1,1) with fractional order accumulation. *Applied Mathematics and Computation*, *252*, 287–293. https://doi.org/10.1016/j.amc.2014.12.014

Wu, L., Liu, S., Yao, L., & Yan, S. (2013). The effect of sample size on the grey system model. *Applied Mathematical Modelling*, *37*(9), 6577–6583. https://doi.org/10.1016/j.apm.2013.01.018

Xie, M., Wu, L., Li, B., & Li, Z. (2020). A novel hybrid multivariate nonlinear grey model for forecasting the traffic-related emissions. *Applied Mathematical Modelling*, *77*, 1242–1254. https://doi.org/10.1016/j.apm.2019.09.013

Yao, Y., & Cao, Y. (2020). A neural network enhanced hidden Markov model for tourism demand forecasting. *Applied Soft Computing*, *94*, 106465. https://doi.org/10.1016/j.asoc.2020.106465

Zeng, B., & Meng, W. (2015). *Research on fractional accumulating generation operators and grey prediction models*. Scientific Press.