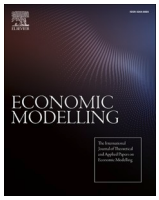
Economic Modelling 135 (2024) 106706

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling



Predicting tourism recovery from COVID-19: A time-varying perspective☆

Ying Liu [a,b], Long Wen [c,*], Han Liu [a,d], Haiyan Song [b]

[a] School of Business and Management, Jilin University, Changchun, 130012, China
[b] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong SAR, China
[c] School of Economics, University of Nottingham Ningbo China, Ningbo, 330200, China
[d] Center for Quantitative Economics, Jilin University, Changchun, 130012, China

ARTICLE INFO

Handling editor: Angus Chu

Original content: Predicting tourism recovery from COVID-9: A time-varying perspective (Original data)

JEL classification:
C32
C53
Z3

Keywords:
Tourism recovery
Nowcasting
Time-varying
Mixed-frequency
COVID-19

ABSTRACT

The uncertainties associated with the coronavirus disease 2019 (COVID-19) pandemic significantly reduced the accuracy of traditional econometric models in forecasting tourism demand, as the relationship between tourism demand and its determinants during the crisis changes over time. To address these inaccuracies, we apply three Factor mixed data sampling (MIDAS) models with different time-varying parameter (TVP) settings: Factor TVP-MIDAS, Factor MIDAS with stochastic volatility (Factor MIDAS-SV), and Factor TVP-MIDAS-SV. We examine the dynamic relationship between tourism demand and its influencing factors, capture the uncertainty and volatility in the data, and provide short-term forecasting and nowcasting. We expose the Factor MIDAS models with TVP specifications to different combinations of determinants to examine their performance. The empirical results show that the Factor MIDAS models with TVP settings performed better than the Factor MIDAS model in the short-term forecasting and nowcasting of tourism demand during COVID-19. The results also suggest that high-frequency data complement these Factor MIDAS models with TVP settings in improving the forecasting and nowcasting accuracy during crises.

1. Introduction

The unexpected coronavirus disease (COVID-19) outbreak in late 2019 adversely affected human health and global economy. The economy has now largely recovered from the initial shocks caused by travel restrictions and other public health measures taken to deal with the pandemic. The World Bank (2022) reported that global gross domestic product (GDP) fell by 5.7% in 2020 compared to 2019 due to the impact of the pandemic, but rose again in 2021. However, recurring outbreaks of COVID-19 continue to adversely affect the tourism industry, which is highly sensitive to crises. The United Nations World Tourism Organization (UNWTO, 2022) reported that international tourist arrivals fell by 73% in 2020 compared to 2019 and recovered only by 4% in 2021 compared to 2020. Furthermore, the number of international tourist arrivals in 2022 was 63% of that in 2019 (UNWTO, 2023). This partial recovery led to more challenges, such as labor shortages and supply constraints, introducing further uncertainty. To sustain this recovery, policy and business decisions must be made based on timely short-term tourism forecasting and nowcasting (Foroni et al., 2022).

The high degree of uncertainty associated with crises makes it difficult to accurate predicting their impact. Simple time-series models work well during normal periods but fail to capture the effects of unexpected crises as they require rich and readily available time-series data without drastic fluctuations to produce reliable forecasts and nowcasts (Larson Sinclair, 2022). Scholars have developed various econometric models with nonlinear specifications to attempt to capture the structural changes caused by unprecedented events, including Markov-switching (Guérin and Marcellino, 2013), threshold (Ferrara et al., 2015), and time-varying parameter (TVP; Page et al., 2012) models. To address the issue of parameter instability in the context of the COVID-19 crisis, vector autoregressive (VAR) frameworks have been proposed that enable models with TVPs (Götz and Hauzenberger, 2021) or relax the standard distribution of the error variance (Carriero et al., 2022). These specifications can improve parameter stability and forecast

☆ We gratefully acknowledge the constructive comments and suggestions from the editor and two anonymous reviews.
* Corresponding author. School of Economics, University of Nottingham Ningbo, 330200, No.199 Taikang East Road, Ningbo, China.
E-mail addresses: curley.liu@polyu.edu.hk (Y. Liu), long.wen@nottingham.edu.cn (L. Wen), hanliu@jlu.edu.cn (H. Liu), haiyan.song@polyu.edu.hk (H. Song).

https://doi.org/10.1016/j.econmod.2024.106706
Received 16 December 2022; Received in revised form 27 February 2024; Accepted 28 February 2024
Available online 5 March 2024

accuracy.

Long delays in the release of official statistics are common, so the availability of current and complete data for real-time forecasting and nowcasting presents another challenge (Ferrara and Sheng, 2022). The issue of low frequency of official statistics has been addressed through short-term forecasting exercises using weekly labor market statistics, daily online search queries, and even hourly electricity consumption data (Jardet and Meunier, 2022; Schaer et al., 2019). Studies have shown that analyzing these and other types of high-frequency data can help in tracking economic changes during crises (Chen et al., 2020; Coibion et al., 2020; Jardet and Meunier, 2022). Yang et al. (2022) showed that daily search query data can help improve the accuracy of tourism demand forecasting for countries with relatively high numbers of confirmed COVID-19 cases, are highly dependent on inbound tourism, and/or have no land borders.

Recent studies that have estimated and forecasted the impact of the COVID-19 pandemic have focused on key macroeconomic indicators such as GDP (Foroni et al., 2022), employment rate (Coibion et al., 2020), and inflation (Bobeica and Hartwig, 2023), along with tourism demand indicators such as tourist arrivals (Liu et al., 2021) and hotel room demand (Zhang and Lu, 2022). However, few studies have examined real-time tourism demand forecasting and nowcasting during the pandemic, and the selected models do not consider the structural changes caused by the pandemic (Wu et al., 2022; Yang et al., 2022). The impact of the COVID-19 pandemic on tourism demand has been assessed through scenario analyses, which does not allow for accurate real-time estimates (Liu et al., 2021; Qiu et al., 2021; Zhang et al., 2021).

To address this research gap, we investigate potential methods for providing accurate short-term forecasting and real-time nowcasting of tourism recovery while considering the impact of the COVID-19 crisis. We combine low-frequency statistics from official sources with daily online search query data as explanatory variables in our forecasting models. We consider three aspects in the process of model specification: (1) it is necessary to reduce the dimensionality of the data to minimize information loss when adding large quantities of high-frequency data econometric models (Nakajima and Sueishi, 2022); (2) we should consider the mismatch in data frequency when using both monthly and daily data to forecast monthly tourism demand; and (3) the specified model should be able to capture the structural changes induced by the COVID-19 pandemic.

We use a two-step modeling approach to address these requirements. First, we use the factor model proposed by Stock and Watson (2002) to reduce dimensionality by extracting common factors from the high-frequency data. Second, we propose three mixed data sampling (MIDAS) models using different TVP settings, namely, TVP-MIDAS, MIDAS with stochastic volatility (MIDAS-SV), and TVP-MIDAS-SV, to examine the short-term forecasting and nowcasting performance of tourism demand during COVID-19. The proposed models can simultaneously handle different frequencies of the dependent and independent variables and capture possible structural breaks and uncertainties by allowing parameters and the error variance to change over time. Baumeister and Guérin (2021) and Foroni et al. (2022) used a basic MIDAS-type model to generate nowcasts of macroeconomic variables during the COVID-19 pandemic and adjust forecasts based on the patterns of identified impact of past crises (e.g., financial crises). However, these basic models cannot monitor pandemic-induced structural changes, and no study in the COVID-19 context has considered the nowcasting performance of the MIDAS model involving TVPs. Our study is, therefore, the first to apply Factor MIDAS models with different TVP specifications to predict post-COVID-19 pandemic tourism demand. The different time-varying specifications allow for variation in the intercept and parameters, as well as error variance over time. These features are expected to minimize the potential loss of forecasting accuracy during crises.

Schumacher (2014) used the particle filtering algorithm to estimate the TVP-MIDAS model. However, as the particle filtering procedure only occurs in one iteration, the convergence of the model is vulnerable to initial values. In contrast, our application of the improved iterated filtering (IF2) algorithm proposed by Ionides et al. (2015) to estimate the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models represents a novel approach and has a strong theoretical justification for model convergence to maximum likelihood estimation. In addition to generating short-term forecasts and nowcasts of tourism recovery, we assess whether (1) our proposed Factor MIDAS models with TVP settings can improve the accuracy of short-term forecasting and nowcasting of tourism recovery amid the ongoing destabilization resulting from the COVID-19 pandemic-related mobility restrictions and public health measures; (2) the information sets generate additional marginal gains to improve nowcasting accuracy (and if so, when); and (3) the model can capture the sudden pandemic-induced decline, and if so, whether the TVP channel or the time-varying error variance channel contributes more to capturing this impact.

We make the following contributions to the literature. First, we overcome the limitations of traditional constant-parameter econometric models using three Factor MIDAS models with distinct TVP settings. These new models use high-frequency search query data and capture the dynamic relationships between dependent and independent variables and uncertainty in the data. Second, we evaluate whether these models can provide marginal advantages in forecasting and nowcasting tourism recovery in the context of the COVID-19 pandemic. We find a complementary relationship between the Factor TVP-MIDAS model and high-frequency data, as these data improve both the forecasting and nowcasting performance of the model. Third, we examine the potential reasons for the probable superiority of the Factor MIDAS models with TVP over the Factor MIDAS model in nowcasting. We find that the Factor MIDAS-SV and Factor TVP-MIDAS-SV models can rapidly and accurately capture the impact of reoccurring outbreaks of the COVID-19 on monthly tourist arrivals using only the daily indexes included as the determining factor. This provides useful information for destination stakeholders to develop the most effective policies for reducing potential economic losses during the crisis.

The rest of this study is presented as follows. In Section 2, we review the literature on tourism demand forecasting and nowcasting during crises. Section 3 describes the methods and data used in this study. Section 4 discusses the forecasting and nowcasting strategies and empirical results. In Section 5, we discuss the limitations of the study and potential areas for future research.

2. Literature review

2.1. Forecasting and nowcasting during crises

Econometric models dominate the tourism forecasting literature and are typically specified as multiple-equation or single-equation models (Bańbura et al., 2013). Multiple-equation models, such as vector autoregression (VAR) models (Kuzin et al., 2011) and simultaneous equations models (Li et al., 2006), forecast the co-movements of the dependent and independent variables. Single-equation models, such as the bridge equation model (Andreini et al., 2023) and MIDAS (Foroni and Marcellino, 2014), are used to analyze dependent variables. As both multiple- and single-equation models depend on historical data for consistent model estimation (Huber et al., 2023), external shocks such as epidemics and natural disasters can lead to structural changes, resulting in accurate forecasts by models. Specifying a reliable model that can capture real-time unprecedented downturns in a crisis context is challenging because the relationship between the dependent and independent variables, the size of the shocks, and the drivers of the crisis can differ and change over time (Foroni et al., 2022).

Many recent studies have focused on developing new frameworks of VAR-type models with multiple equations to improve the reliability of short-term forecasts and nowcasts during the COVID-19 pandemic. For example, Schorfheide and Song (2021) and Lenza and Primiceri (2020)

suggested that one can improve the nowcasting performance of VAR models by considering the extreme data variations in the first half of 2020 as possible outliers and mixed distribution of the model error variance or models with stochastic volatility to reduce the impact of structural instability on nowcasting accuracy. Similarly, Bobeica and Hartwig (2023) and Carriero et al. (2022) found that the VAR models' forecasting properties could be enhanced by relaxing the assumption of the standard Gaussian distribution of model errors to Student's *t*-distribution during the pandemic. Götz and Hauzenberger (2021) used mixed-frequency VAR models involving TVPs to predict the unprecedented decline in macroeconomic variables caused by the COVID-19 pandemic. They found that including a few independent variables with similarly large deviations, the forecasting models could accurately capture the pandemic's impact.

MIDAS-type models have also recently been used to generate forecasts and nowcasts during the COVID-19 pandemic. These models can use high-frequency independent variables to explain low-frequency dependent variables without aggregation procedure and are therefore applicable to short-term forecasting and nowcasting research in the pandemic context. For example, Foroni et al. (2022) used MIDAS and unrestricted MIDAS (UMIDAS) models to generate forecasts and nowcasts of the GDP of the US and other G7 countries during the pandemic and the recovery period. To account for the lack of historical data during the pandemic, which forecasting models require, they adjusted the forecasts and nowcasts based on the recovery patterns identified during the 2007–2008 global financial crisis. Similarly, Baumeister and Guérin (2021) used MIDAS and UMIDAS models to assess whether an indicator based on a set of global economic variables could help improve global GDP growth forecasts and nowcasts during normal times and the pandemic. However, no studies have tested the forecasting and nowcasting performance of a TVP-MIDAS model in this context.

2.2. Tourism demand forecasting amid the COVID-19 pandemic

Tourism demand is particularly susceptible to the effects of crises such as natural disasters, terrorism, political turmoil, and epidemics because tourists are risk-averse, and any perceived threat to their health and safety can affect their travel decisions. The resulting decline in tourist arrivals can have economic consequences for destinations and the tourism industry in general (Speakman and Sharpley, 2012). The COVID-19 pandemic has had a major impact on the tourism industry since late 2019. The pandemic led to a severe economic and social crisis, and a series of travel bans and lockdown measures were introduced in various countries to prevent the spread of COVID-19 (Qiu et al., 2021). The pandemic's impact on tourism demand has been specifically assessed in the literature. For example, Gössling et al. (2021) compared the pandemic to previous crises, such as the outbreak of severe acute respiratory syndrome, after examining its effects on global tourism through the end of March 2020. They argued that COVID-19-related travel restrictions have caused the greatest damage to global tourism since World War II. Hao et al. (2020) developed a COVID-19 management framework based on their analysis of the pandemic's impact on the Chinese hotel industry. They suggested that various aspects of the industry were permanently affected, such as investment preferences and product design. Liew (2022) examined the extent of the pandemic's effects on the share prices of Booking.com, Expedia Group, and Trip.com, the three largest online tourism companies, and found that the overall tourism industry performance declined rapidly during the pandemic. Polyzos et al. (2021) assessed the impact of the COVID-19 outbreak on tourist arrivals from China to the US and Australia and identified its adverse effects and suggested it could take nearly a year for arrivals to return to prepandemic levels.

The increased levels of uncertainty that arises during protracted crises such as the COVID-19 pandemic can drastically influence future behavior in the tourism industry. Therefore, forecasting tourism demand is even more important for policymakers during crises than in normal times (Liu et al., 2021). However, few studies have focused on forecasting tourism demand recovery during the COVID-19 pandemic. Korinth (2022) used the autoregressive integrated moving average (ARIMA) model to forecast passenger air traffic and the share of Poland's accommodation options used in the first quarter of 2020. The results showed that the pandemic-related travel restrictions and public health measures negatively and substantially affected Poland's tourism industry. Using traditional time-series models such as autoregressive moving average (ARMA) and ARMA with exogenous variable models, Wickramasinghe and Ratnasiri (2021) demonstrated that forecast accuracy was improved by incorporating Google search data into forecasts of guest night stay by international tourist arrivals to Sri Lanka in the first quarter of 2020. Yang et al. (2022) used the least absolute shrinkage and selection operator method with a series of regressors to forecast daily tourism demand across 75 countries during the pandemic. They showed that Google search data could improve the accuracy of these forecasts in certain situations, such as for countries that depend heavily on inbound tourism or those with no land borders.

There is scarce information on tourism during the pandemic; therefore, recent studies have adopted judgmental two-stage adjustments (Kourentzes et al., 2021; Liu et al., 2021; Polyzos et al., 2021; Qiu et al., 2021; Zhang et al., 2021) to assess the future improvement in tourism demand. First, *ex ante* baseline forecasts are generated based on appropriate models estimated based on precrisis data and judgmental approaches, such as scenario analyses and Delphi methods, which are then used to adjust these forecasts by considering different pandemic scenarios. However, judgmental adjustments can be problematic because of subjective assessments, which may introduce bias, human error, and lead to a lack of replicability. Although the aforementioned studies have considered several scenarios, most have focused on the long-term implications of the pandemic rather than short-term forecasting and nowcasting, which are also crucial for the timely implementation of crisis management plans.

3. Methodology and data

3.1. Model setup

We apply three Factor MIDAS models with different TVP settings to address the issues of high dimensionality, frequency mismatch, and possible structural changes. Using high-frequency data to capture the impact of the COVID-19 pandemic, the models are expected to improve the accuracy of forecasting and nowcasting tourist arrivals during such crises. We take a two-step modeling approach to construct our Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models. First, the factor model extracts common factors from high-frequency data. We then introduce different TVP specifications into the Factor MIDAS model to build the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models and include the extracted factors as independent variables into the models.

We sequentially introduce the Factor MIDAS, Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models for a better understanding of the model construction process. These models are built using a two-step method, in which we first introduce the factor model, which reduces the dimensionality of huge high-frequency time-series data by extracting a few common factors and is commonly used to forecast and construct leading factors (Forni et al., 2004). In specifying the factor model, $\{x_{i,t_m}^{(m)}, i = 1, \ldots n, t_m = 1, \ldots, T_m\}$ denotes a large set of high-frequency time series that correspond to the low-frequency dependent variable $\{y_t, t = 1, \ldots, T\}$, which is only observable every m period, where $T_m = mT$. For example, $m = 3$ if y_t is a quarterly variable and $x_{i,t_m}^{(m)}$ is a monthly variable. The specification of the factor model can be expressed as follows:

$$x_{i,t_m}^{(m)} = \lambda_{i1} f_{1,t_m}^{(m)} + \lambda_{i2} f_{2,t_m}^{(m)} + \ldots + \lambda_{ir} f_{r,t_m}^{(m)} + e_{i,t_m} \quad (1)$$

Table 1
IF2 algorithm pseudocode.

Input
Simulator for $p(\mathbf{S}_0\|\theta)$
Simulator for $p(\mathbf{S}_t\|\mathbf{S}_{t-1},\theta)$, t in $1:T$
Evaluator for $p(\mathbf{D}_t\|\mathbf{S}_t,\theta)$, t in $1:T$
Data, $\mathbf{D}^*_{1:T}$
Number of iterations, M
Number of particles, J
Initial parameter sample, $\{\Phi^0_j, j \text{ in } 1:J\}$
Perturbation sequence, $\sigma_{1:M}$
Procedure
For m in $1:M$.
$\Phi^{F,m}_{0,j} \sim N(\Phi^{m-1}_j, \sigma_m)$ for j in $1:J$.
$\mathbf{S}^{F,m}_{0,j} \sim p(\mathbf{S}_0 \mid \Phi^{F,m}_{0,j})$ for j in $1:J$.
For t in $1:T$.
$\Phi^{P,m}_{t,j} \sim N(\Phi^{F,m}_{t-1}, \sigma_m)$ for j in $1:J$.
$\mathbf{S}^{P,m}_{t,j} \sim p(\mathbf{S}_t \mid \mathbf{S}^{F,m}_{t-1,j}, \Phi^{P,m}_{t,j})$ for j in $1:J$.
$w^m_{t,j} = p(\mathbf{D}^*_t \mid \mathbf{S}^{P,m}_{t,j}, \Phi^{P,m}_{t,j})$ for j in $1:J$.
Draw $k_{1:J}$ with $p(k_j = i) = w^m_{t,i} / \sum_{u=1}^{J} w^m_{t,u}$.
$\Phi^{F,m}_{t,j} = \Phi^{P,m}_{t,k_j}$ and $\mathbf{S}^{F,m}_{t,j} = \mathbf{S}^{P,m}_{t,k_j}$ for j in $1:J$.
End For.
Set $\Phi^m_j = \Phi^{F,m}_{Tj}$ for j in $1:J$.
End For.

Notes: (1) The superscript F in $\Phi^{F,m}_{t,j}$ and $\mathbf{S}^{F,m}_{t,j}$ represents solutions to the filtering problem, and the superscript P in $\Phi^{P,m}_{t,j}$ and $\mathbf{S}^{P,m}_{t,j}$ represents solutions to the prediction problem; (2) the weights $w^m_{t,j}$ yield the likelihood of the designated data $\mathbf{D}^*_t$.

where λ_{ij} is the factor loading and $f^{(m)}_{j,t_m}$ is a common factor, with $j = 1,\ldots, r$. The idiosyncratic component e_{i,t_m} is the part of $x^{(m)}_{i,t_m}$ not explained by the factors. The number of factors is determined by the information criterion IC_{p2} proposed by Bai and Ng (2002), and we use the static principal component analysis model introduced by Stock and Watson (2002) to estimate the factor model.

(1) Factor MIDAS model

The Factor MIDAS model proposed by Marcellino and Schumacher (2010) uses the MIDAS model to regress low-frequency dependent variables on extracted high-frequency common factors $\{f^{(m)}_{j,t_m}, j = 1,\ldots, r\}$. The specification of the Factor MIDAS model for forecasting y_{t+1} can be written as follows:

$$y_{t+1} = c + \sum_{j=1}^{r} \beta_j b_j(L_m, \boldsymbol{\theta}_j) f^{(m)}_{j,t} + \varepsilon_t \tag{2}$$

$$b_j(L_m, \boldsymbol{\theta}_j) = \sum_{k=0}^{K_j} c(k, \boldsymbol{\theta}_j) L^m_k, c(k, \boldsymbol{\theta}_j) = \frac{\exp(\theta_{1,j}k + \theta_{2,j}k^2)}{\sum_{k=0}^{K_j} \exp(\theta_{1,j}k + \theta_{2,j}k^2)} \tag{3}$$

where the polynomial $b_j(L_m, \boldsymbol{\theta}_j) = \sum_{k=0}^{K_j} c(k, \boldsymbol{\theta}_j) L^m_k$ assigns weights to lags 0 to K of the j-th high-frequency factor based on the exponential Almon specification, which is the most commonly used weighting scheme and highly flexible in generating different weight shapes (Ghysels et al., 2007). L^m is the high-frequency lag operator, and $L^m_k f^{(m)}_{j,t} = f^{(m)}_{j,t-k/m}$. K_j is the maximum lag on the j-th high-frequency variable. ε_t represents the independent white noise processes with a mean of zero and variance denoted by σ^2_ε. The nonlinear least squares (NLS) method is used to estimate the Factor MIDAS model.

The Factor MIDAS model can be further extended to include autoregressive terms and variables measured at a frequency identical to that of the independent variable, which can be expressed as follows:

$$y_{t+1} = c + \gamma y_t + \sum_{p=1}^{P} \alpha_p x_{p,t} + \sum_{j=1}^{r} \beta_j b_j(L_m, \boldsymbol{\theta}_j) f^{(m)}_{j,t} + \varepsilon_t \tag{4}$$

The AR coefficient γ and the coefficients α_p of the low-frequency variables $x_{p,t}$ can be estimated alongside other coefficients using the NLS method.

(2) Factor TVP-MIDAS model

We include the extracted common factors $\{f^{(m)}_{i,t}, j = 1,\ldots, r\}$ into the TVP-MIDAS model proposed by Schumacher (2014) to specify the Factor TVP-MIDAS model. The specification of this model with one lag of y_{t+1} and low-frequency independent variables $\{x_{p,t}, p = 1,\ldots, P\}$ for forecasting y_{t+1} can be written as follows:

$$y_{t+1} = c_t + \gamma_t y_t + \sum_{p=1}^{P} \alpha_{p,t} x_{p,t} + \sum_{j=1}^{r} \beta_{j,t} b_{j,t}(L_m, \boldsymbol{\theta}_{j,t}) f^{(m)}_{j,t} + \varepsilon_t \tag{5}$$

$$b_{j,t}(L_m, \boldsymbol{\theta}_{j,t}) = \sum_{k=0}^{K_j} c(k, \boldsymbol{\theta}_{j,t}) L^m_k, c(k, \boldsymbol{\theta}_{j,t}) = \frac{\exp(\theta_{1,j,t}k + \theta_{2,j,t}k^2)}{\sum_{k=0}^{K_j} \exp(\theta_{1,j,t}k + \theta_{2,j,t}k^2)} \tag{6}$$

$$c_t = c_{t-1} + \varepsilon_{c,t} \tag{7}$$

$$\gamma_t = \gamma_{t-1} + \varepsilon_{\gamma,t} \tag{8}$$

$$\alpha_{p,t} = \alpha_{p,t-1} + \varepsilon_{\alpha_p,t} \tag{9}$$

$$\beta_{j,t} = \beta_{j,t-1} + \varepsilon_{\beta_j,t} \tag{10}$$

$$\theta_{1,j,t} = \theta_{1,j,t-1} + \varepsilon_{1,j,t}, \theta_{2,j,t} = \theta_{2,j,t-1} + \varepsilon_{2,j,t} \tag{11}$$

Unlike their counterparts in the Factor MIDAS model, parameters c_t, γ_t, $\alpha_{p,t}$, $\beta_{j,t}$, $\theta_{1,j,t}$, and $\theta_{2,j,t}$ of the Factor TVP-MIDAS model follow the random-walk process, which is commonly used in the TVP literature, especially when multiple parameters are allowed to change over time (Primiceri, 2005). ε_t, $\varepsilon_{\gamma,t}$, $\varepsilon_{\alpha_p,t}$, $\varepsilon_{\beta_j,t}$, $\varepsilon_{1,j,t}$, and $\varepsilon_{2,j,t}$ are independent white noise processes, which have a mean of zero and variances of σ^2_ε, σ^2_c, σ^2_γ, $\sigma^2_{\alpha_p}$, $\sigma^2_{\beta_j}$, $\sigma^2_{\varepsilon_{1,j}}$, and $\sigma^2_{\varepsilon_{2,j}}$, respectively. To ensure the stability of the weighting scheme, we impose some restrictions on $\theta_{2,j,t}$ (Lütkepohl, 1981) and create a newly defined variable $\theta^*_{2,j,t}$ by setting $\theta_{2,j,t} = -\exp(\theta^*_{2,j,t})$ rather than allowing $\theta_{2,j,t}$ to evolve freely as a random-walk process. Equations (6) and (11) then become Equations (12) and (13), respectively.

$$b_{j,t}(L_m, \boldsymbol{\theta}_{j,t}) = \sum_{k=0}^{K_j} c(k, \boldsymbol{\theta}_{j,t}) L^m_k, c(k, \boldsymbol{\theta}_{j,t}) = \frac{\exp\left(\theta_{1,j,t}k - \exp\left(\theta^*_{2,j,t}\right)k^2\right)}{\sum_{k=0}^{K_j} \exp\left(\theta_{1,j,t}k - \exp\left(\theta^*_{2,j,t}\right)k^2\right)} \tag{12}$$

$$\theta_{1,j,t} = \theta_{1,j,t-1} + \varepsilon_{1,j,t}, \theta^*_{2,j,t} = \theta^*_{2,j,t-1} + \varepsilon_{2,j,t} \tag{13}$$

The Factor TVP-MIDAS model can be represented in a nonlinear state–space form for estimation:

$$y_{t+1} = \Phi_t \mathbf{F}_t(\boldsymbol{\lambda}_t) + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2_\varepsilon) \tag{14}$$

$$\begin{bmatrix} \boldsymbol{\lambda}_t \\ \Phi_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}_{t-1} \\ \Phi_{t-1} \end{bmatrix} + \mathbf{w}_t, \mathbf{w}_t \sim N(0, \mathbf{W}) \tag{15}$$

where Equation (14) is the measurement equation and Equation (15) is the transition equation. All states are classified into Φ_t and $\boldsymbol{\lambda}_t$ based on the nonlinearity of the system to convert the Factor TVP-MIDAS model

into a state–space form. Specifically, the vector $\lambda_t = (\theta_{1,1,t}, \ldots, \theta_{1,r,t}, \theta^*_{2,1,t}, \ldots, \theta^*_{2,r,t})$ includes all nonlinear states not amenable to Kalman filtering, while the vector $\Phi_t = (c_t, \gamma_t, \alpha_{1,t}, \ldots, \alpha_{P,t}, \beta_{1,t}, \ldots, \beta_{r,t})$ contains all linear states.

For Equations (14) and (15), $\mathbf{F}_t(\lambda_t)$, $\mathbf{w}_t$ and $\mathbf{W}$ are defined as:

$$\mathbf{F}_t(\lambda_t)' = \left(1 \; y_t \; x_{1,t} \; \ldots \; x_{P,t} \; b_{1,t}(L_m, \boldsymbol{\theta}_{1,t}) f^{(m)}_{1,t} \; \ldots \; b_{r,t}(L_m, \boldsymbol{\theta}_{r,t}) f^{(m)}_{r,t}\right) \tag{16}$$

$$\mathbf{w}_t = \begin{pmatrix} \varepsilon_{1,1,t} \\ \vdots \\ \varepsilon_{1,r,t} \\ \varepsilon_{2,1,t} \\ \vdots \\ \varepsilon_{2,r,t} \\ c_t \\ \varepsilon_{\gamma_t} \\ \varepsilon_{\alpha_{1,t}} \\ \vdots \\ \varepsilon_{\alpha_{P,t}} \\ \varepsilon_{\beta_{1,t}} \\ \vdots \\ \varepsilon_{\beta_{r,t}} \end{pmatrix}, \mathbf{W} = \begin{pmatrix} \sigma^2_{\varepsilon_{1,1,t}} & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \vdots & \sigma^2_{\varepsilon_{1,r,t}} & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma^2_{\varepsilon_{2,1,t}} & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & \sigma^2_{\varepsilon_{2,r,t}} & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & \sigma^2_{c_t} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & \sigma^2_{\gamma_t} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & \sigma^2_{\alpha_{1,t}} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & \vdots & \sigma^2_{\alpha_{P,t}} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & \vdots & 0 & \sigma^2_{\beta_{1,t}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma^2_{\beta r,t} \end{pmatrix} \tag{17}$$

(3) Factor MIDAS-SV model

We incorporate stochastic volatility into the MIDAS model to develop the MIDAS-SV model. We allow the error variance of the MIDAS model to change over time using an AR(1) process (Götz and Hauzenberger, 2021). Subsequently, we integrate the derived common factors $\{f^{(m)}_{j,t_m}, j = 1, \ldots, r\}$ into the MIDAS-SV model to define the Factor MIDAS-SV model. The model specification for the Factor MIDAS-SV model, which includes one lag of y_{t+1} and low-frequency independent variables $\{x_{p,t}, p = 1, \ldots, P\}$ to forecast y_{t+1}, can be expressed as follows:

$$y_{t+1} = c + \gamma y_t + \sum_{p=1}^{P} \alpha_p x_{p,t} + \sum_{j=1}^{r} \beta_j b_j(L_m, \boldsymbol{\theta}_j) f^{(m)}_{j,t} + \sqrt{\exp(h_t)}\varepsilon_t \tag{18}$$

$$b_j(L_m, \boldsymbol{\theta}_j) = \sum_{k=0}^{K_j} c(k, \boldsymbol{\theta}_j) L^m_k, c(k, \boldsymbol{\theta}_j) = \frac{\exp\left(\theta_{1,j} k - \exp\left(\theta^*_{2,j}\right) k^2\right)}{\sum_{k=0}^{K_j} \exp\left(\theta_{1,j} k - \exp\left(\theta^*_{2,j}\right) k^2\right)} \tag{19}$$

$$h_t = \rho h_{t-1} + \varepsilon_{h,t}, |\rho| < 1 \tag{20}$$

where εt represents a white noise process, which have a mean of zero and a variance of one. h_t represents the stochastic volatility estimate and follows a stationary AR(1) specification. $\varepsilon_{h,t}$ is an independent white noise process whose associated variance σ^2_h follows an inverse gamma distribution, denoted as $\sigma^2_h \sim IG(v_h, V_h)$. The initial values of the parameters associated with stochastic volatility are selected based on Götz and Hauzenberger (2021).

The following state–space form can be used to estimate the Factor MIDAS-SV model:

$$y_{t+1} = \boldsymbol{\Phi}\mathbf{F}_t(\lambda) + v_t, v_t \sim N(0, V_t(h_t)) \tag{21}$$

$$h_t = \rho h_{t-1} + \varepsilon_{h,t}, |\rho| < 1 \tag{22}$$

where $\lambda = (\theta_{1,1}, \ldots, \theta_{1,r}, \theta^*_{2,1}, \ldots, \theta^*_{2,r})'$, $\mathbf{F}_t(\lambda)' = (1 \; y_t \; x_{1,t} \cdots x_{P,t} \; b_1(L_m, \boldsymbol{\theta}_1) f^{(m)}_{1,t} \cdots b_r(L_m, \boldsymbol{\theta}_r) f^{(m)}_{r,t})$, $\Phi = (\gamma, \alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_r)'$, $v_t = \sqrt{\exp(h_t)_{\varepsilon_t}}$, and $V_t(h_t) = \exp(h_t)\sigma^2_\varepsilon$.

(4) Factor TVP-MIDAS-SV model

We develop the Factor TVP-MIDAS-SV model by introducing stochastic volatility into the TVP-MIDAS model proposed by Schumacher (2014) and incorporating the obtained high-frequency common factors $\{f^{(m)}_{j,t_m}, j = 1, \ldots, r\}$ as independent variables. We consider two distinct channels of time variation to capture potential structural breaks in tourism demand and the uncertainties evolving from the COVID-19 pandemic. The first channel allows parameters to vary over time, including intercepts, coefficients, and the shape parameters of the nonlinear lag polynomial of the high-frequency common factors. The second channel allows for variation in error variance. The specification of the Factor TVP-MIDAS-SV model for forecasting y_{t+1} can be written as follows:

$$y_{t+1} = c_t + \gamma_t y_t + \sum_{p=1}^{P} \alpha_{p,t} x_{p,t} + \sum_{j=1}^{r} \beta_{j,t} b_{j,t}(L_m, \boldsymbol{\theta}_{j,t}) f^{(m)}_{j,t} + \sqrt{\exp(h_t)}\varepsilon_t \tag{23}$$

$$b_{j,t}(L_m, \boldsymbol{\theta}_{j,t}) = \sum_{k=0}^{K_j} c(k, \boldsymbol{\theta}_{j,t}) L^m_k, c(k, \boldsymbol{\theta}_{j,t}) = \frac{\exp\left(\theta_{1,j,t} k - \exp\left(\theta^*_{2,j,t}\right) k^2\right)}{\sum_{k=0}^{K_j} \exp\left(\theta_{1,j,t} k - \exp\left(\theta^*_{2,j,t}\right) k^2\right)} \tag{24}$$

$$c_t = c_{t-1} + \varepsilon_{c,t} \tag{25}$$

$$\gamma_t = \gamma_{t-1} + \varepsilon_{\gamma,t} \tag{26}$$

$$\alpha_{p,t} = \alpha_{p,t-1} + \varepsilon_{\alpha_p,t} \tag{27}$$

$$\beta_{j,t} = \beta_{j,t-1} + \varepsilon_{\beta_j,t} \tag{28}$$

$$\theta_{1,j,t} = \theta_{1,j,t-1} + \varepsilon_{1,j,t}, \theta^*_{2,j,t} = \theta^*_{2,j,t-1} + \varepsilon_{2,j,t} \tag{29}$$

$$h_t = \rho h_{t-1} + \varepsilon_{h,t}, |\rho| < 1 \tag{30}$$

The Factor TVP-MIDAS-SV model uses the same TVP setting as the Factor TVP-MIDAS model, while the stochastic volatility setting follows that of the MIDAS-SV model. The Factor TVP-MIDAS-SV model can be converted to the following state–space form:

$$y_{t+1} = \Phi_t \mathbf{F}_t(\lambda_t) + v_t, v_t \sim N(0, V_t(h_t)) \tag{31}$$

$$\begin{bmatrix} \lambda_t \\ \boldsymbol{\Phi}_t \\ h_t \end{bmatrix} = \begin{bmatrix} \lambda_{t-1} \\ \boldsymbol{\Phi}_{t-1} \\ \rho h_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_t \\ \varepsilon_h \end{bmatrix}, \mathbf{w}_t \sim N\left(0, \mathbf{W}\right), \varepsilon_h \sim N(0, \sigma^2_h) \tag{32}$$

where the states λ_t and Φ_t are defined as $\lambda_t = (\theta_{1,1,t}, \ldots, \theta_{1,r,t}, \theta^*_{2,1,t}, \ldots, \theta^*_{2,r,t})'$ and $\Phi_t = (c_t, \gamma_t, \alpha_{1,t}, \ldots, \alpha_{P,t}, \beta_{1,t}, \ldots, \beta_{r,t})'$, $\mathbf{F}_t(\lambda_t)' = (1 \;\; y_t \;\; x_{1,t} \cdots x_{P_s,t} \;\; b_{1,t}(L_m, \boldsymbol{\theta}_t) f^{(m)}_{1,t} \cdots b_{r,t}(L_m, \boldsymbol{\theta}_t) f^{(m)}_{r,t})$, $v_t = \sqrt{\exp(h_t)}\varepsilon_t$, $V_t(h_t) = \exp(h_t)\sigma^2_\varepsilon$, and $\mathbf{w}_t$ and $\mathbf{W}$ are defined as

$$\mathbf{w}_t = \begin{pmatrix} \varepsilon_{1,1,t} \\ \vdots \\ \varepsilon_{1,r,t} \\ \varepsilon_{2,1,t} \\ \vdots \\ \varepsilon_{2,r,t} \\ c_t \\ \varepsilon_{\gamma_t} \\ \varepsilon_{\alpha_{1,t}} \\ \vdots \\ \varepsilon_{\alpha_{P,t}} \\ \varepsilon_{\beta_{1,t}} \\ \vdots \\ \varepsilon_{\beta_{r,t}} \end{pmatrix}, \mathbf{W}$$

$$= \begin{pmatrix} \sigma^2_{\varepsilon_{1,1,t}} & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \vdots & \sigma^2_{\varepsilon_{1,r,t}} & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma^2_{\varepsilon_{2,1,t}} & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & \sigma^2_{\varepsilon_{2,r,t}} & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & \sigma^2_{c_t} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & \sigma^2_{\gamma_t} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & \sigma^2_{\alpha_{1,t}} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & \vdots & \sigma^2_{\alpha_{P,t}} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & \vdots & 0 & \sigma^2_{\beta_{1,t}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma^2_{\beta_{r,t}} \end{pmatrix} \tag{33}$$

3.2. Model estimation

As the states $\theta_{1,j,t}$, $\theta^*_{2,j,t}$, and h_t are nonlinear in the transition equation of the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models, the standard Kalman filter cannot be used as it can only estimate linear models. Therefore, we use a particle filter, a sequential Monte Carlo method that simulates and mutates state (particle) samples according to the state–space dynamics in the model, to estimate the MIDAS-type models with nonlinear TVPs (Schumacher, 2014).

We use the Factor TVP-MIDAS-SV model as a case study to illustrate the estimation process. If we denote all states c_t, γ_t, $\alpha_{p,t}$, $\beta_{j,t}$, $\theta_{1,j,t}$, $\theta^*_{2,j,t}$, and h_t as $\mathbf{S}_t$, designate data up to t as $\mathbf{D}_t = \{\mathbf{D}_{t-1}, y_t, x_t\}$, and denote the variances of the white noise processes ε_t, and unknown initial states $\mathbf{S}_0$ as θ, then the particle filter sequentially produces new particles based on the transition densities $p(\mathbf{S}_t|\mathbf{S}_{t-1}, \theta)$ in the prediction step. The particle filter also adjusts the particles' weights using measurement density $p(\mathbf{D}_t|\mathbf{S}_t, \theta)$, and the sample is updated in the filtering step. However, the time-invariant parameters represented by θ are often unknown and must be estimated. The particle filter method includes an additional artificial random-walk process for θ with small variances. With this added noise, the particles and their associate θ_t can be produced sequentially during each period. The resulting samples can then be used to infer the properties of θ. However, this method involves only one iteration of the filtering procedure and its convergence property is not guaranteed.

In contrast, the convergence of iterated filtering to the maximum likelihood estimation has been theoretically justified (Ionides et al., 2011, 2015). Such filtering involves multiple repetitions of the particle filtering procedure, during which the intensity of the random perturbations of θ decreases toward zero and the parameters converge to maximum likelihood estimates. The IF2 proposed by Ionides et al. (2015) is considerably more effective than the single-iteration particle filter. We, therefore, use the IF2 algorithm in this study to estimate the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models. Table 1 shows the algorithm pseudocode for the model estimation.

The T loop, spanning lines 4 through 10 in Table 1, constitutes a fundamental particle filter used to first obtain the maximum likelihood estimates of the TVPs θ and to address model parameter stochasticity. The particle filter is then repeated several times with decreasing random-walk intensities ($\boldsymbol{\sigma}_{1:M}$) through the M loop. The random-walk intensity, often called the temperature, decreases with successive filtering iterations according to a cooling schedule. Thus, σ_m becomes smaller as m increases. We focus on the initial value parameters, which represent the initial states in this study and are a subset of the time-invariant parameters of the Factor TVP-MIDAS model. Only the early time points contain information about the initial states; therefore, these parameters are usually estimated inconsistently with increasing length of the time series. Therefore, the IVPs are only perturbed at time zero. More details related to the iterated filtering method are given by Ionides et al. (2015).

3.3. Data

Fig. 1 shows the data obtained on monthly overnight tourist arrivals into Hainan province from the Wind database to measure demand for tourism (Yang et al., 2015; Zhang et al., 2017). Daily data are only available on the Baidu search engine from January 2011; therefore, the sample period is from January 2011 to December 2021. We convert the number of tourist arrivals into a natural logarithm using the model specifications used in earlier studies (Wen et al., 2021). Hainan province is a popular tourist destination in China, recording more than 87 million tourist arrivals in 2019, representing an annual growth rate of 9% (Hainan Tourism Bureau, 2020). The initial COVID-19 outbreak in December 2019 was detrimental to tourism demand. There were 64 million tourist visits in 2020, down 23% drop from the previous year. The numbers gradually began to recover in February 2020; however, it was one of the first provinces to reopen for tourism during the pandemic. Tourist arrivals exceeded 81 million in 2021, almost returning to pre-pandemic levels. Although the unexpected reemergence of COVID-19 later in the year resulted in another substantial drop in arrivals (Fig. 1).

We use two types of data as the independent variables to model and forecast tourism demand recovery: monthly macroeconomic variables and daily search query data. First, economic theory suggests that a destination's tourism demand is affected by its specific prices, prices in competing destinations (substitute prices), and the country of origin, or by the region's tourism income (Song and Li, 2008; Song and Lin, 2010; Song et al., 2003, 2009). Therefore, we include the relative tourism price, P_{Hainan}, of Hainan province and the tourism income of source regions in mainland China (excluding Hainan), Y, as influencing factors in the model. We calculate P_{Hainan} based on the consumer price index in Hainan compared to that of China as a whole and express it as CPI_{Hainan}/CPI_{China}. GDP, which is often used to measure tourism income, is released quarterly with considerable time lags; therefore, we select monthly industrial value added as a proxy for GDP (Chatziantoniou et al., 2016). The substitute prices should be determined from the top five tourist destinations in mainland China. However, Hainan province's particular climate and natural landscape offer tourists a unique tropical island experience in mainland China. Thus, it is difficult for Hainan province to identify substitute tourist destinations, so we omit substitute

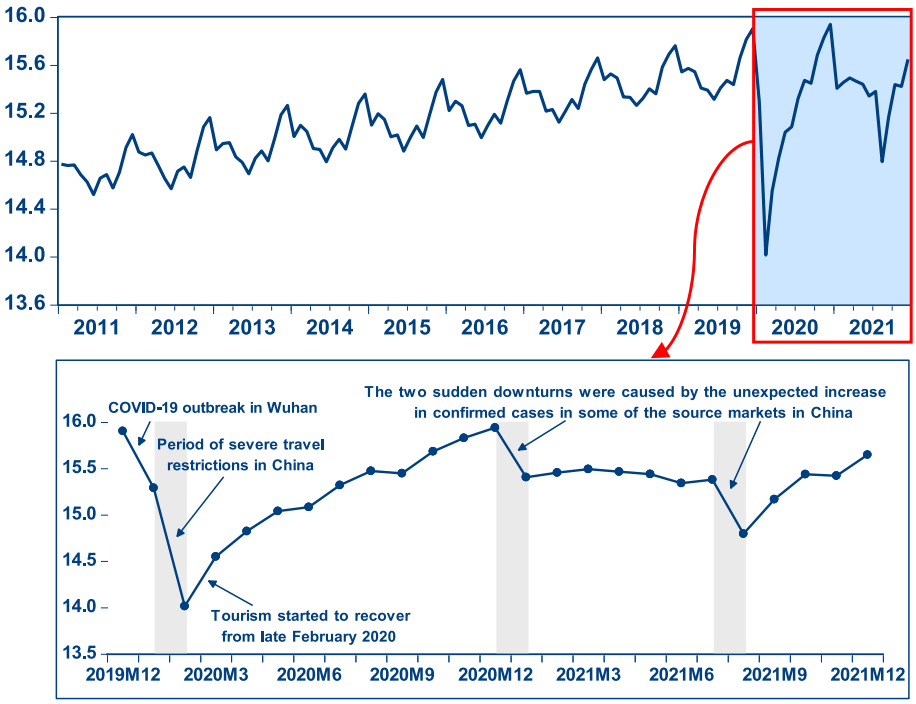
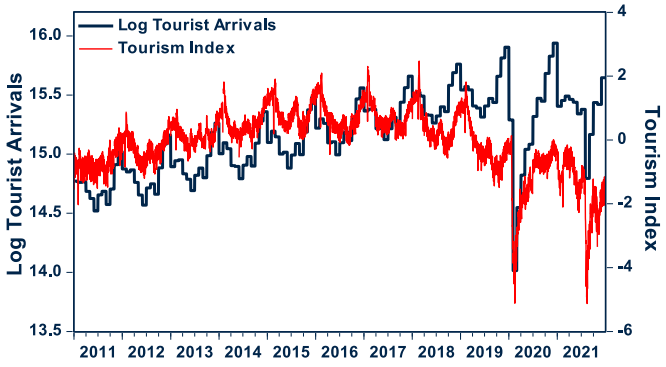
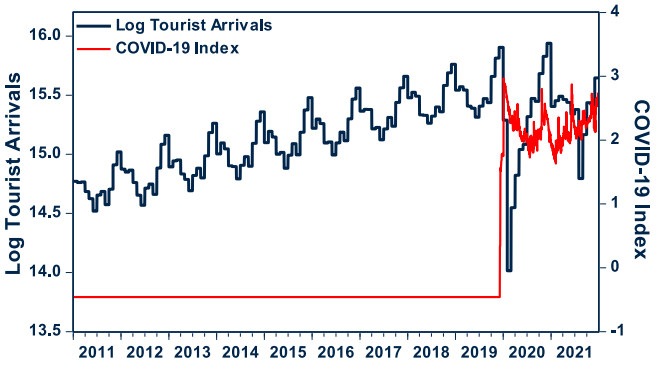


Fig. 1. Tourist arrivals (in logarithm form).



(a) Tourism Index



(b) COVID-19 Index

Fig. 2. Daily indexes and log of tourist arrivals.

prices (Li et al., 2017a). We include monthly time-varying seasonal dummies to capture seasonality (Shen et al., 2009; Zhang et al., 2021).

Second, research has shown that relevant real-time data can be useful in modeling and forecasting tourism demand, particularly in a crisis context (Kourentzes et al., 2021). We therefore collect daily search query data related to tourism and the COVID-19 pandemic from Baidu, which we use in the modeling process to determine the most effective method for measuring the impact of the pandemic on tourism demand. Search queries are typically derived through intuition and domain knowledge (Li et al., 2017b; Yang et al., 2015). We follow other tourism demand forecasting studies to initially consider search queries related to lodging, dining, attractions, transportation, tours, and shopping as aspects of tourism planning (Wen et al., 2021). We also consider pandemic-related queries, which reflect tourists' perceived risk about a destination. We develop our search queries in Baidu as follows.

(i) We define the categories of search queries related to aspects of tourism planning and the COVID-19 pandemic. Each category initially contains various selected search queries.
(ii) We further expand the pool of queries by adding other queries strongly correlated with the initial queries using functions in the Baidu Index. We repeat this step until we reach convergence.
(iii) The numbers of queries below a certain threshold are not reported on Baidu; therefore, we manually check the availability of each search.
(iv) Finally, we select 90 Baidu search queries. Table A in the Appendix presents the details.

The large number of relevant search queries results in high dimensionality, hence daily data cannot be directly analyzed. Therefore, we use the factor model to reduce the dimensionality and extract information from the large volume of daily data. We finally construct two indexes: (1) a tourism index constructed from the search queries related to the six aspects of tourism planning mentioned earlier and (2) a COVID-19 Index constructed using information from two sources: the search queries related to the pandemic and China's stringency index (collected from Our World Data (https://ourworldindata.org/coronavirus)), which represents the overall severity of China's pandemic-related measures. Fig. 2 shows the relationships between the daily search query indexes and the log of monthly tourist arrivals.

4. Empirical results

Our empirical study focuses on forecasting and nowcasting tourism recovery during COVID-19 pandemic. We aim to test the feasibility of capturing the tourism recovery process after demand suddenly due to the pandemic using a time-varying, mixed-frequency model with various information sets. We chose Hainan province, one of China's most popular tourist destinations, as the research context because domestic tourism resumed in this province soon after the pandemic, although COVID-19 continued to reemerge intermittently.

4.1. Forecasting and nowcasting strategies

We monitor the impact of the pandemic on real-time tourism demand during COVID-19 resurgence. We assess the one-step-ahead forecasting and nowcasting performance of the Factor TVP-MIDAS model. One-step-ahead short-term forecasting involves analyzing data for the independent variables available at the end of the month to predict the dependent variable for the following month (Baffigi et al., 2004). In contrast, nowcasting involves predicting the current month's dependent variable using the real-time high-frequency variables of the same month. The nowcasts are continuously updated with data on the high-frequency independent variables as they become available (Bańbura et al., 2011).

We use data from January 2011 (2011M1) to December 2020 (2020M12) for the model estimates and that from January 2021 (2021M1) to December 2021 (2021M12) to evaluate the one-step-ahead forecasting and nowcasting performance of the selected models. Using an expanding window, we create one-step-ahead forecasts and nowcasts. In the forecasting process, we use the independent variables up to month $t-1$ and the dependent variables up to month t for model estimation. Using data on the dependent variables at time t, we generate the one-step-ahead forecasts for month $t+1$. In the nowcasting process, we use the information available at the beginning of month $t+1$ and then re-estimate the model to update the same month's nowcasts when new daily data become available. As the search query data are updated every day, we repeat this process for each day of the month.

We use three common forecast error measures to evaluate the accuracy of model forecasts and nowcasts: the mean absolute percentage error (MAPE); the root mean square error (RMSE), based on the growth rate; and the mean absolute scaled error (MASE). These can be expressed as follows:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\widehat{y}_t - y_t}{y_t}\right| \tag{34}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\Delta\widehat{y}_t - \Delta y_t\right)^2} \tag{35}$$

$$MASE = \frac{\sum_{t=1}^{N}\left|\widehat{y}_t - y_t\right|/N}{\sum_{t=2}^{T}\left|y_t - y_{t-1}\right|/(T-1)} \tag{36}$$

where $\widehat{y}_t$ is the forecast or nowcast at time t, y_t is the actual value, $\Delta\widehat{y}_t = (\widehat{y}_t - y_{t-1})/y_{t-1} \times 100$, $\Delta y_t = (y_t - y_{t-1})/y_{t-1} \times 100$, N is the length of the forecast period, and in this study $N = 12$. The use of multiple measures can help in better comprehension of the forecasting accuracy of the models being evaluated and assess whether their forecasting performance is stable during the out-of-sample period.

For comparison, we select three commonly used models as benchmarks to test the performance of tourism demand forecasting methods: the seasonal naïve (SNaïve), seasonal ARIMA (SARIMA), and single exponential smoothing (ETS) models (Hyndman, 2018). The SNaïve model assumes that the future values of a time series will correspond to the most recent value from the previous seasonal period. The SARIMA model, which combines an autoregressive (AR) component, a moving average component, and seasonal patterns into a single model, has become increasingly popular in recent years. The ETS model uses a weighted average of past observations with exponentially decreasing weights (Athanasopoulos et al., 2011). We then add the Factor MIDAS, Factor TVP-MIDAS, and Factor MIDAS-SV models to determine whether the different specifications of TVPs can improve forecasting and nowcasting accuracy. We set the number of lags of daily indexes K_j to 30 (Wen et al., 2021). The initialization of parameters to be estimated in these models is given by the estimation results of the Factor MIDAS model. The residual variance of this model is used to initialize the variance of the particle sets, while the initial variances of the shocks in the TVPs are determined based on those by Schumacher (2014). Given the data, we then use the iterated filtering algorithm to estimate the posterior distribution of the parameters. We specify several particles (e.g., 2000) for the iterated filtering algorithm, allowing us to explore various possible parameter values.

To evaluate the marginal gains of different information sets, we consider three model specifications in the Factor MIDAS and TVP-MIDAS models: (1) we use the full set of information, including the

Table 2
One-step-ahead forecasting results.

A. Model with macro variables, lagged tourist arrivals, and daily indexes							
	SNaïve	**SARIMA**	**ETS**	**Factor MIDAS**	**Factor TVP-MIDAS**	**Factor MIDAS-SV**	**Factor TVP-MIDAS-SV**
MAPE	41.59%	22.03%	19.69%	18.37%	18.22%	18.30%	**18.06%**
RMSE	45.52%	29.31%	23.87%	22.57%	**21.26%**	22.42%	22.15%
MASE	3.04	1.44	1.22	1.15	1.13	1.14	**1.12**
B. Model with lagged tourist arrivals and daily indexes							
	SNaïve	**SARIMA**	**ETS**	**Factor MIDAS**	**Factor TVP-MIDAS**	**Factor MIDAS-SV**	**Factor TVP-MIDAS-SV**
MAPE	41.59%	22.03%	**19.69%**	20.85%	20.95%	20.88%	20.84%
RMSE	45.52%	29.31%	**23.87%**	26.46%	25.95%	26.49%	26.11%
MASE	3.04	1.44	**1.22**	1.33	1.36	1.33	1.32
C. Model with daily indexes							
	SNaïve	**SARIMA**	**ETS**	**Factor MIDAS**	**Factor TVP-MIDAS**	**Factor MIDAS-SV**	**Factor TVP-MIDAS-SV**
MAPE	41.59%	22.03%	**19.69%**	23.49%	23.04%	23.49%	23.21%
RMSE	45.52%	29.31%	**23.87%**	29.25%	26.92%	29.25%	28.64%
MASE	3.04	1.44	**1.22**	1.64	1.50	1.64	1.61

Note: The figures displayed in bold represent the best forecasting performance for each measurement error.

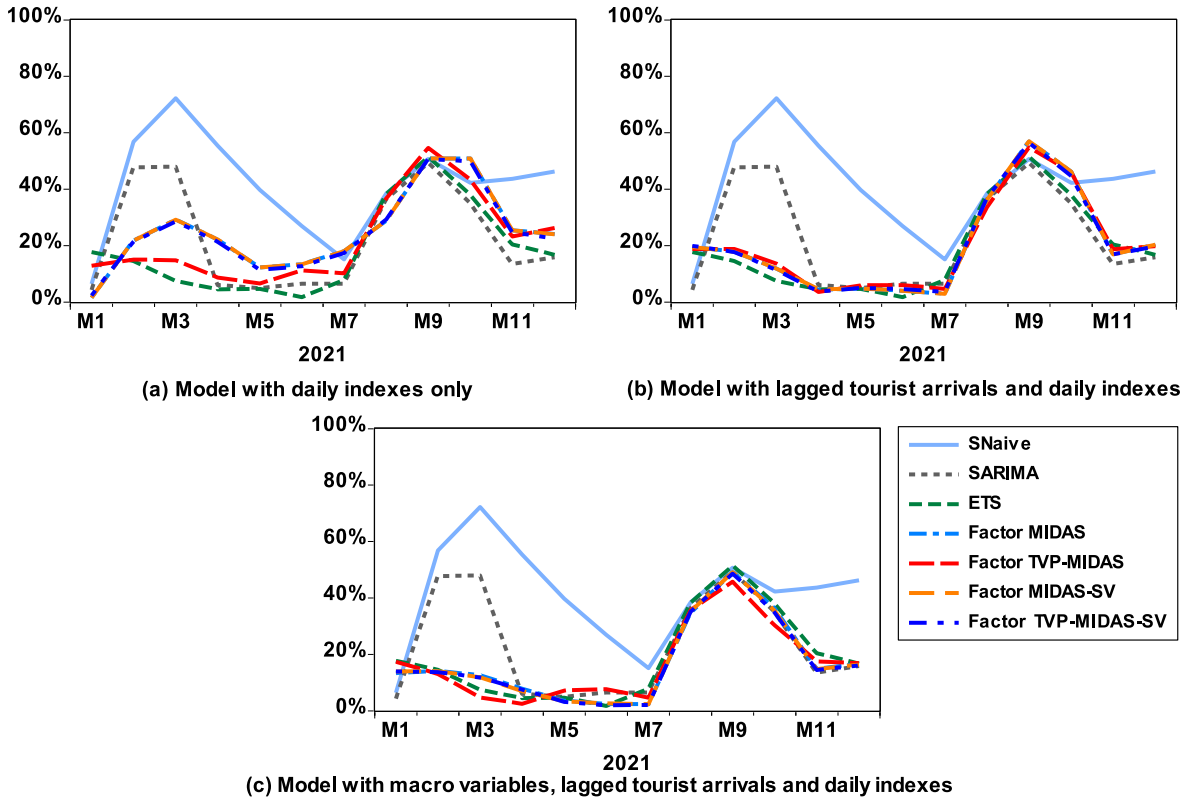


Fig. 3. One-step-ahead forecasting rolling RMSEs.

monthly macroeconomic variables, monthly lagged dependent variables, and daily indexes for modeling and forecasting; (2) we construct models by combining the lagged dependent variables and daily indexes to assess the marginal gains of integrating the macroeconomic variables; and (3) we use only the daily indexes for modeling and forecasting to evaluate whether such data can be used solely to generate reliable forecasts and nowcasts during crises.

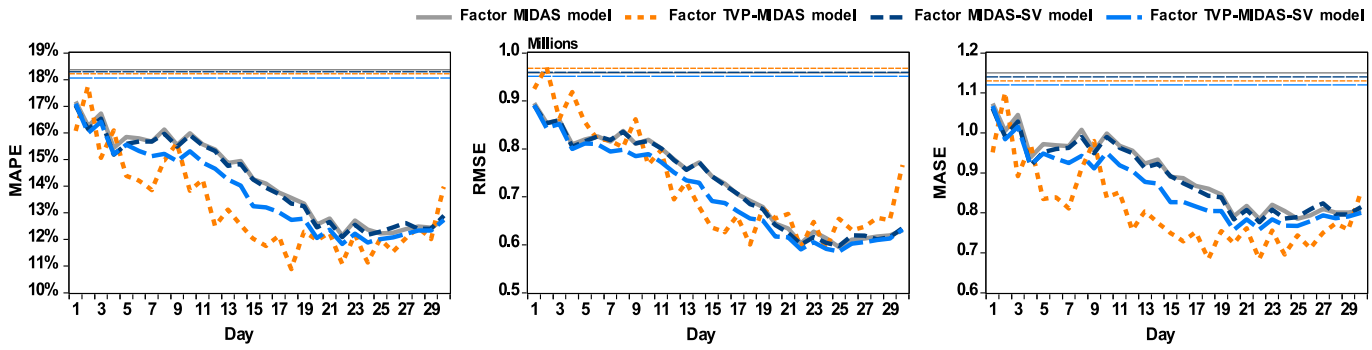
4.2. Short-term forecasting results

This subsection presents the one-step-ahead forecasting results of tourism recovery during the COVID-19 pandemic using various information sets. Table 2 shows the one-step-ahead forecast accuracy results for each model, with Panels A to C corresponding to the forecasting results of the mixed-frequency models with three information sets. The error measures indicate that the SNaïve model demonstrates worst performance as the COVID-19 pandemic led to a substantial decline in tourist arrivals in early 2020. The SARIMA model performs better than the SNaïve model, indicating that including lagged dependent variables, current and lagged random shocks, and degrees of integration improves forecasting performance (Song et al., 2019). The ETS model outperforms the other time-series models, suggesting that it is suitable for considering the different characteristics of time-series data by incorporating level, trend, and seasonal components (Kourentzes et al., 2014). The Factor MIDAS-type models with daily indexes, or with lagged tourist arrivals and daily indexes, have lower forecasting performance than the ETS model when compared with the time-series models. However, the Factor MIDAS-type models that include macroeconomic variables, lagged tourist arrivals, and daily indexes have better forecasting accuracy. The results show that adding macroeconomic variables can provide more comprehensive information about the overall economic environment than just adding the lagged tourist arrivals and daily indexes. This is consistent with our expectation that using additional explanatory variables can improve the accuracy of short-term forecasting. Furthermore, the Factor TVP-MIDAS-SV model with the full information set outperforms other models, according to the MAPE and MASE metrics, as shown in Panel A of Table 2. These results indicate that incorporating the SV model helps capture the impact of COVID-19 more accurately when using a dataset that includes macroeconomic variables and daily indexes for the short-term forecasting of tourism demand.

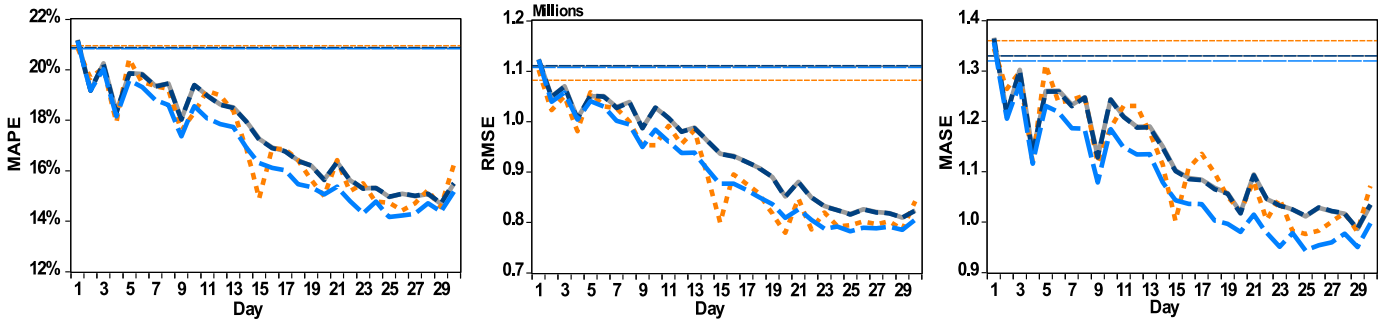
We also obtain the rolling RMSEs of the one-step-ahead forecasting results shown in Table 2 to examine the time-varying forecasting performance of the models (Escribano and Wang, 2021). We calculate the rolling RMSEs based on growth rates over a two-month moving window to capture possible short-term fluctuations in the forecast errors. Fig. 3 shows the rolling RMSEs for the one-step-ahead forecasting results of models with three sets of information sets. The SNaïve model is the least robust across all forecast horizons and has the highest RMSEs, followed by the SARIMA model. In contrast, the ETS and MIDAS-type models show relatively stable performance across all forecast periods, with significant RMSE spikes observed only in August 2021 owing to the impact of COVID-19. These results suggest that the ETS and MIDAS models are more robust and reliable for one-step-ahead forecasting, except under exceptional circumstances such as the pandemic.

4.3. Nowcasting results

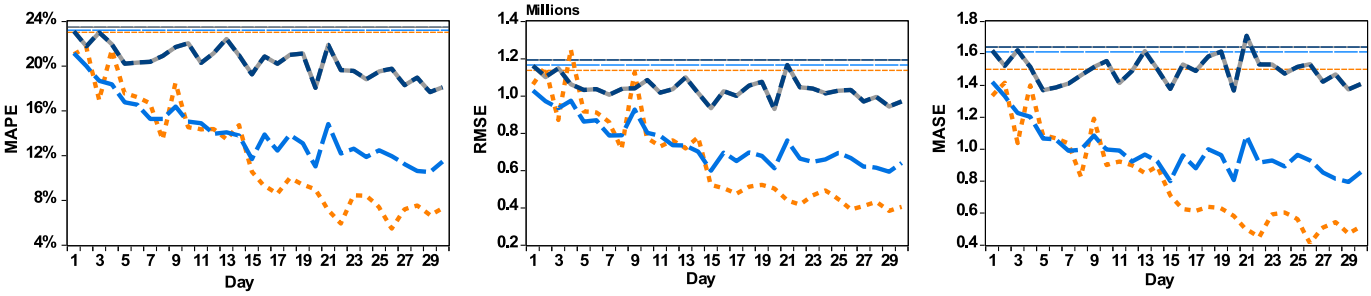
Given the COVID-19 pandemic, timely nowcasts of tourism demand are especially crucial for policymakers and stakeholders developing flexible crisis management plans. However, it is challenging to create accurate nowcasts during unexpected events (Larson and Sinclair, 2022). The daily updated nowcasts of tourist arrivals during the COVID-19 pandemic generated by MIDAS-type models with high-frequency search query data are presented in this subsection. We also examine whether it is possible to capture the unexpected effects of the pandemic using these models during the nowcasting process. MIDAS-type models can generate daily updated nowcasts upon availability of new search query data, unlike the benchmark models of SNaïve



(a) Nowcasts of MIDAS-type models with macro variables, lagged tourist arrivals, and daily indexes



(b) Nowcasts of MIDAS-type models with lagged tourist arrivals and daily indexes



(c) Nowcasts of MIDAS-type models with daily indexes

Fig. 4. Comparison of nowcasting performance by the Factor MIDAS-type models.

and SARIMA, which cannot incorporate independent variables to generate nowcasts. Nowcasting performance can also be improved by the status of tourists' travel intentions and the pandemic's severity, as shown by the relevant daily search queries.

Fig. 4 shows a comparison of nowcasting performance between the Factor MIDAS, Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models. This allows us to assess whether diverse TVP settings can improve the nowcasting performance in scenarios with varied information sets. The horizontal lines on each panel of Fig. 4 represent the one-step-ahead forecasting performance of the corresponding MIDAS-type models. First, the comparison of the nowcasts with their corresponding forecasts in Fig. 4 shows that most nowcasts for the same mixed-frequency model are below the corresponding forecasting results and the gaps widen as new daily indexes become available, suggesting that the accuracy of nowcasting for the same mixed-frequency model is better than that of forecasting. Second, even with different information sets, the evolution of the nowcasts for MIDAS-type models slopes downward over the month in most cases, indicating that integrating more daily updated data can help improve nowcasting accuracy. This result is consistent with that of Wen et al. (2021). Third, we observe that the nowcasts for the Factor MIDAS-SV model are similar to those for the Factor MIDAS model, regardless of the information sets used. This suggests that including the SV model may not substantially improve the nowcast accuracy without adding real-time data for the current month. However, the Factor TVP-MIDAS-SV model reflects the improved nowcasting performance of the Factor TVP-MIDAS model and outperforms

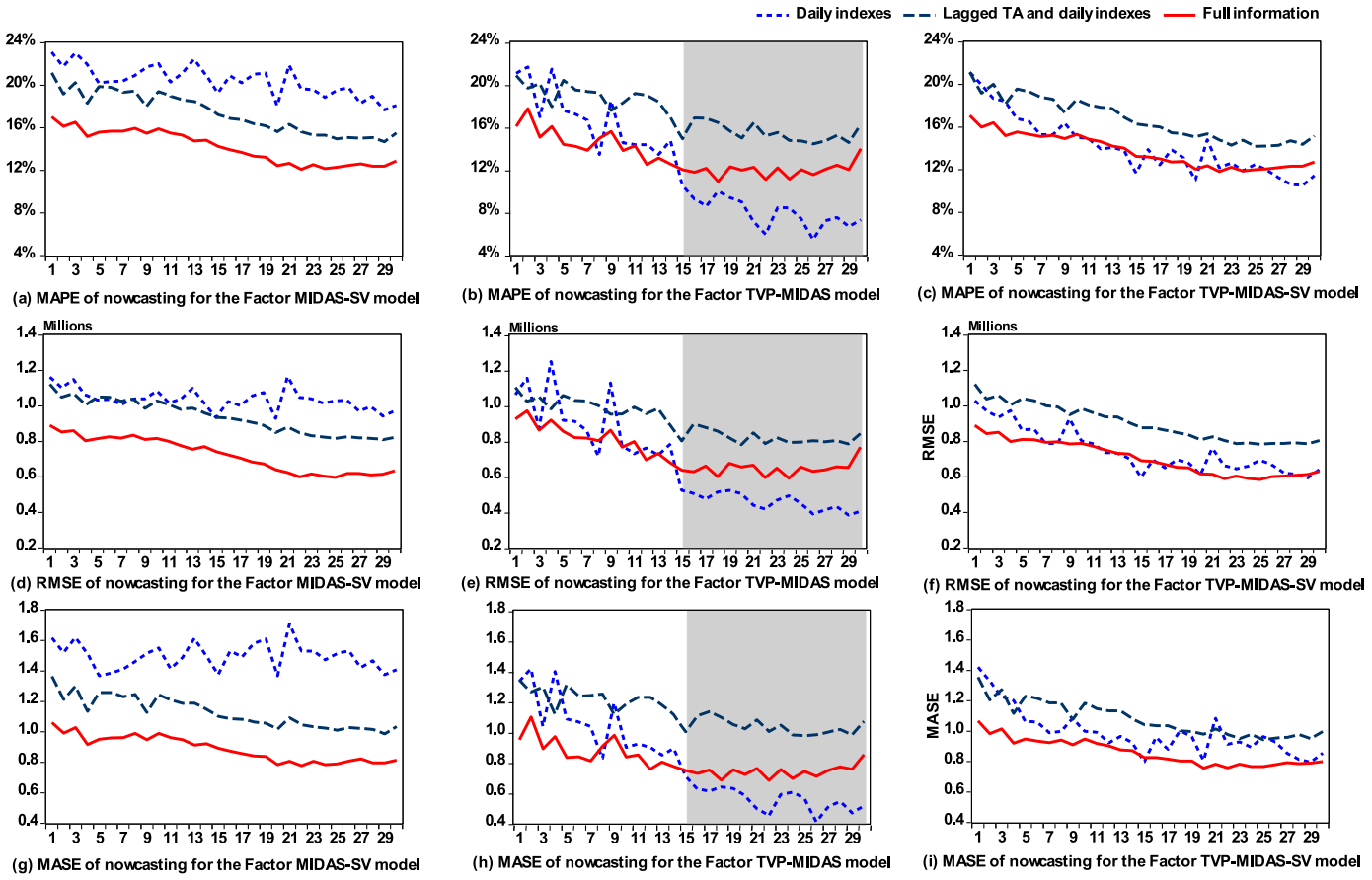


Fig. 5. Comparison of nowcasting performance with different information sets.

the benchmark Factor MIDAS and Factor MIDAS-SV models. These findings suggest that incorporating the TVP method, which can capture changes in the relationships between variables, into the Factor MIDAS model is more effective in improving nowcasting accuracy than depending on the SV model alone.

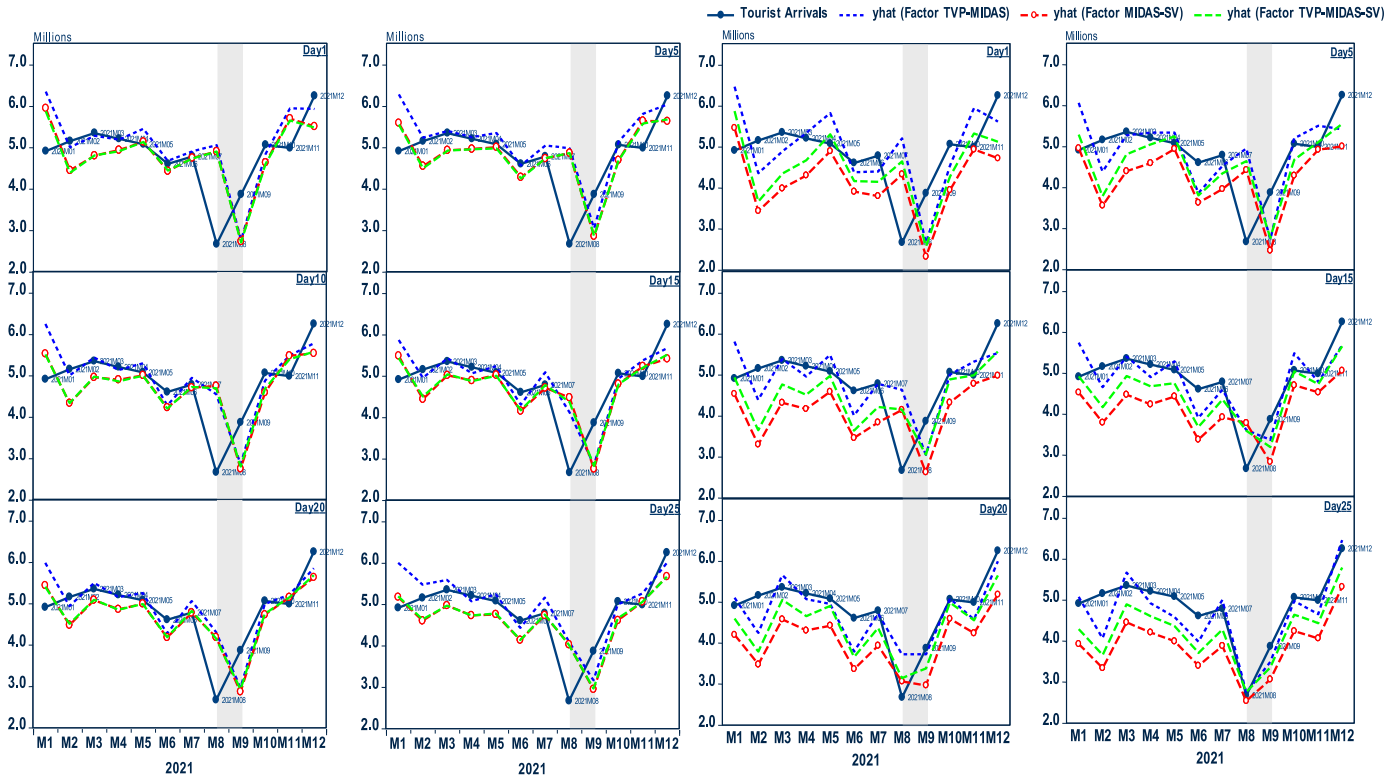
Fig. 5 presents a comparison of the nowcasting performance of the same model using different information sets, which helps examine whether adding the macroeconomic variables, lagged dependent variable, and high-frequency indexes can lead to marginal gains in nowcasting accuracy. We first examine the Factor MIDAS model's nowcasting performance. Fig. 5(a), (d), and (g) show that the Factor MIDAS model with daily indexes (bright blue dashed line) represents the worst performance in terms of all error measures. Adding lagged dependent variables improves nowcasting performance and adding the macroeconomic variables results in remarkable nowcasting accuracy gains. Similarly, Fig. 5(c), (f), and (i) show that the Factor TVP-MIDAS-SV model with the full information set outperforms the other datasets.

Regarding the nowcasting results of the Factor TVP-MIDAS model, Fig. 5(b), (e), and (h) shows that this model performs worst with lagged tourist arrivals and daily indexes. However, the accuracy of the updated nowcasts improves on the availability of more daily data. The Factor TVP-MIDAS model with full information (red solid lines in Fig. 5(b), (e), and (h)) performs the best in the first half of the month. Surprisingly, the Factor TVP-MIDAS model that includes only daily indexes outperforms all competing models in the second half of the month, and the error measures for this model are relatively small. For example, MAPE values remain below 10% and MASE values below 0.8 in the second half of the month. This indicates that the Factor TVP-MIDAS model outperforms the benchmark models in terms of the nowcasting performance and better utilizes the valuable information in high-frequency data.

As shown earlier, the TVP-MIDAS model with full information performs best in the first half of the month, while the model with daily indexes performs best only in the second half of the month. To further understand this change in the nowcasting performance for the model that includes only daily indexes, we compare the actual nowcasts generated by the MIDAS-type models with the full information set and those with daily indexes only. Fig. 6 plots the actual nowcasts generated on the 1st, 5th, 10th, 15th, 20th, and 25th days of the month from the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models with full information and those with daily indexes only. As Fig. 6 shows, we also use tourist arrivals in Hainan province to compare the actual numbers and nowcasts. The resurgence of the COVID-19 pandemic in China in 2021M8 led to a significant drop in tourist arrivals, making it difficult to capture the unexpected impact of the COVID-19 pandemic on tourism demand.

Fig. 6(a) shows the actual nowcasts generated by the MIDAS-type model with full information. The Factor TVP-MIDAS model outperforms the Factor MIDAS-SV and Factor TVP-MIDAS-SV models in the normal period, when there is no reemergence of the COVID-19 infections (e.g., 2021M1 to 2021M7 and 2021M10 to 2021M12). Any volatility or uncertainty in the data is typically absorbed by MIDAS-type models with stochastic volatility, generating nowcast values that are often lower than the actual values during the normal period. However, none of the MIDAS-type models with TVP settings accurately nowcast the sudden drop in tourist arrivals caused by the recurrence of the COVID-19 pandemic in 2021M8.

The nowcasts of the MIDAS-type models integrating only the daily indexes are shown in Fig. 6(b). Our findings indicate that, first, consistent with the results in Fig. 6(a), the Factor TVP-MIDAS model outperforms the Factor MIDAS-SV and Factor TVP-MIDAS-SV models in the normal period; however, its nowcasts during this period show greater fluctuations than those generated by the Factor TVP-MIDAS model with



(a). MIDAS-type models with full information

(b). MIDAS-type models with daily indexes only

Fig. 6. Evolution of the actual nowcasts of MIDAS-type models with different information sets.

the full information set. Second, the nowcast values of the MIDAS-type models remain below the actual tourist arrivals in the normal period, particularly for the Factor MIDAS-SV and Factor TVP-MIDAS-SV models, indicating that the MIDAS models with stochastic volatility overestimate the negative impact of the pandemic when there are no COVID-19 cases. Third, in Fig. 6(b), the nowcasts obtained on the first day of the month show that none of the MIDAS-type models can capture the sudden drop in tourist arrivals in 2021M8 quickly enough and that the actual nowcasts generated on the 5th and 10th days of the month do not change substantially as more data become available. The nowcasts generated by the MIDAS models with stochastic volatility on the 20th day of the month, however, show a remarkable decrease in tourist arrivals in 2021M8 than the nowcasts generated by the Factor TVP-MIDAS model. By the end of the month, the nowcasts generated by all models are similar to the actual numbers.

In summary, daily index data provide timely signals of sudden declines arising from unexpected crises. Compared with the Factor TVP-MIDAS model, combining the SV approach with the Factor MIDAS or Factor TVP-MIDAS models can rapidly and accurately nowcast the unexpected impact of the COVID-19 pandemic on tourist arrivals on the availability of more daily data. However, all models can predict the impact of COVID-19 with the availability of sufficient accessible daily data, such as after 20 days in each month. Adding the SV model, which can capture the uncertainty or volatility in the data, to the Factor MIDAS and Factor TVP-MIDAS models yield nowcasts that consistently underestimate the actual values during the normal period. This is shown by the nowcasting performance of the Factor TVP-MIDAS model with daily indexes, which improves substantially in the middle of the month. Furthermore, there is no evidence of the relative nowcasting power of daily indexes after we add macroeconomic variables and lagged tourist arrivals, as the sudden drop in tourist arrivals in 2021M8 cannot be captured by MIDAS-type models with a full information set. However, such statistical data can enhance the nowcasting accuracy during normal times by providing more information on past trends and cyclical changes of the dependent variable, whereas daily data can better reflect the current situation. The information lag of macroeconomic variables, especially during rapidly changing economic situations (such as that caused by the COVID-19 pandemic), may thus limit their usefulness in nowcasting.

The major differences between the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models using the same daily indexes arise because of the distinct configurations of the TVPs. In the Factor TVP-MIDAS and Factor TVP-MIDAS-SV models, all parameters, including the intercept, coefficients, and shape parameters of the nonlinear lag polynomial of the explanatory variables, follow random walks, while the error variance in the Factor MIDAS-SV and Factor TVP-MIDAS-SV models vary over time. These characteristics may help in determining the relationship between variables and possible structural breaks in case of unexpected events (Schumacher, 2014). Therefore, to identify the source of the differences in the results between MIDAS-type models with daily indexes, we further analyze the changes in the TVPs. We first examine the changes in the weights of the tourism and COVID-19 indexes calculated using the shape parameters of the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models. As we can obtain 360 group weights (12 months, 30 nowcasts per month) for each daily index in each model when generating nowcasts from 2021M1–2021M12, we use 2021M8—the period when the nowcasts demonstrate the maximum change—as an example to analyze the differences in the MIDAS-type models. Fig. 7 shows the evolution of the weights of the tourism index and the COVID-19 Index obtained using the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models on the 5th, 15th, and 25th days of the month. The weights of each index are shown as a function of the lags, with the maximum lag increasing as new daily data become available.

The evolution of the weights of the tourism index in Fig. 7 shows that the maximum weight obtained with the Factor TVP-MIDAS-SV model

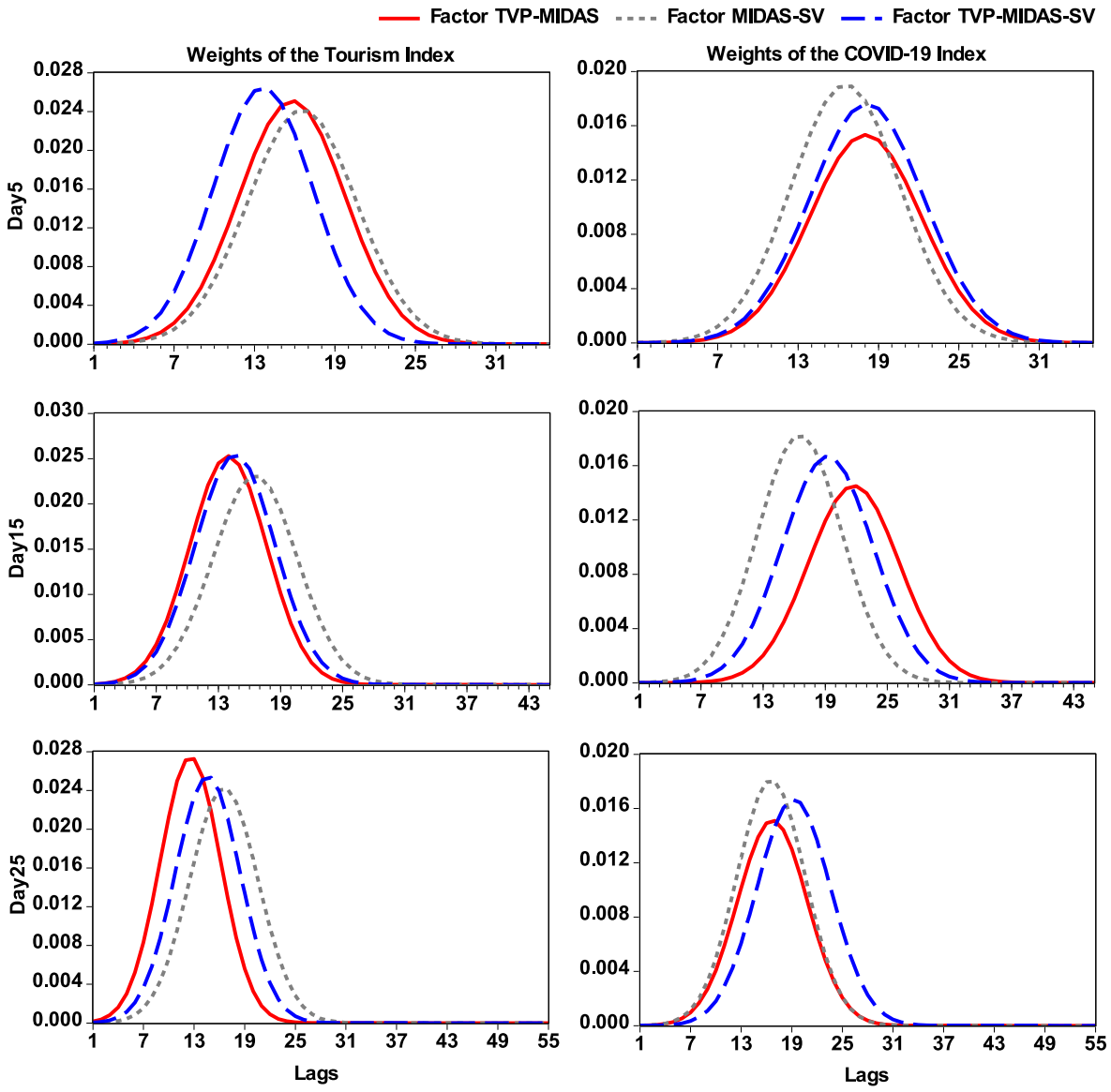


Fig. 7. Evolution of the exponent Almon lag weights over time in 2021M8.

consistently occurs at a smaller lag than using the Factor MIDAS-SV model. For example, when nowcasting on the 5th day of the month, the maximum weight of the tourism index obtained with the Factor TVP-MIDAS-SV model occurs at approximately lag 13, whereas that obtained by the Factor MIDAS-SV model occurs at about lag 16. Moreover, as more daily data become available, the Factor TVP-MIDAS model assigns greater importance to the updated data than other models. Regarding the weights of the COVID-19 Index, we observe that on the 5th day of the month, the Factor MIDAS-SV model attaches more importance to the latest data than the Factor TVP-MIDAS and Factor TVP-MIDAS-SV models.

However, no substantial difference is found in the evolution of the weights of the COVID-19 Index generated by the TVP-MIDAS and Factor MIDAS-SV models, indicating that on availability of more daily data, the daily indexes on tourism arrivals obtained using the Factor TVP-MIDAS model have a greater impact on the latest data than the Factor MIDAS-type models with stochastic volatility. Furthermore, in the nowcasting process, the weights of lags greater than 31 days tend toward zero, implying that more recent data are important for monitoring shifting tourism demand dynamics and that earlier data should therefore be discounted. This agrees with the findings of Schorfheide and Song (2021).

Finally, we analyze the difference in the TVPs of the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models. Fig. 8 shows the evolutions of the intercepts and the time-varying coefficients of the daily indexes and stochastic volatilities generated by the Factor TVP-MIDAS, Factor MIDAS-SV, and Factor TVP-MIDAS-SV models that incorporate only daily indexes for nowcasting on the 5th, 15th, and 25th days of 2021M8. The stochastic volatilities are plotted in terms of the standard deviations $\exp(h_t/2)$.

First, we observe that the changes in the intercept can be estimated using the changes in the long-term growth rate of tourist arrivals. Fig. 8 (a) shows that the intercepts of the Factor TVP-MIDAS and Factor TVP-MIDAS-SV models increased by approximately 0.5 on a logarithmic scale from 2011 to 2019, which correspond to more than 3 million additional tourist arrivals. However, the outbreak of COVID-19 in 2020 led to a significant decline in arrivals. Second, we observe that the tourism index and tourist arrivals in Hainan province are positively correlated, indicating a corresponding increase in the number of tourist arrivals as tourism-related Baidu searches increase. Third, the coefficients of the COVID-19 Index in Fig. 8(c) show a substantial downward trend since 2020. A negative relationship is expected between the COVID-19 Index

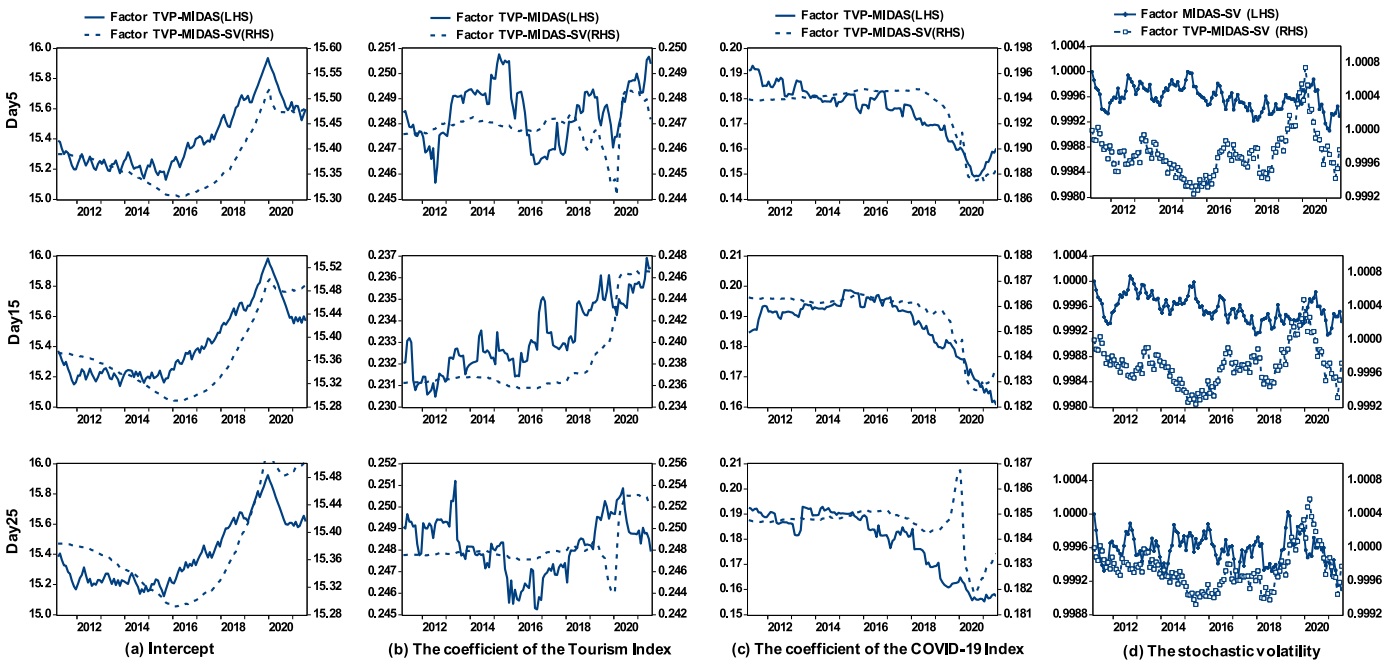


Fig. 8. Evolution of time-varying coefficients from 2011M1–2021M8.

and tourist arrivals as an increase in the COVID-19 Index (indicating greater severity and stricter prevention measures) may negatively impact tourist arrivals. However, the overall analysis reveals a positive correlation between these factors. This unexpected outcome can be a result of the index reflecting both the severity of the pandemic in Hainan province and the strictness of China's prevention and control measures. The slight increase in the COVID-19 Index when the situation in Hainan was relatively stable may have been the result of the severity of the epidemic in other parts of China; thus, the index may not have shown a negative correlation with the tourist arrivals, and thereby indicating that the impact of COVID-19 Index on tourist arrivals is influenced by factors outside Hainan province. Fourth, Fig. 8(d) shows the evolution of stochastic volatility for both the Factor MIDAS-SV and Factor TVP-MIDAS-SV models, showing a significant spike at the beginning of 2020 with some minor fluctuations. This indicates that the COVID-19-related uncertainty significantly affected the tourism industry in Hainan province. We also see that the time-varying coefficients of the Factor TVP-MIDAS-SV model evolve more smoothly than those of the Factor TVP-MIDAS model, as the SV model introduced into the former accounts for fluctuations and uncertainties in the data. Overall, we find that the feasibility and utility of the TVPs increase the likelihood of detecting potential instability when unexpected crises arise.

5. Conclusions

In this study, we generate short-term forecasts and real-time nowcasts of tourism recovery in China's Hainan province, a destination affected by the intermittent measures implemented to minimize the spread of COVID-19. It is difficult to determine a reliable method for generating accurate forecasts and nowcasts from a large amount of data during crises. We accurately capture the impact of the pandemic on tourism demand using Factor MIDAS models with different TVP settings. Such methods can simultaneously deal with the problems of high dimensionality, frequency mismatch resulting from high-frequency data, and any structural changes. We also consider different information sets and examine their marginal improvements in forecasting and nowcasting performance during the COVID-19 pandemic.

The empirical results show that the Factor TVP-MIDAS-SV model that includes all collected data outperforms the benchmark models in short-term forecasting. This indicates that short-term forecasting accuracy can be improved by combining as much available information as possible. A comparison of the forecasts and nowcasts obtained using MIDAS-type models with different information sets shows that the nowcasting performance of MIDAS-type models is consistently better than their forecasting performance. The evolution of the nowcasts shows that the Factor TVP-MIDAS model that integrates only daily indexes demonstrates outstanding nowcasting performance in the second half of the month as the impact of the COVID-19 pandemic on tourism demand can be more accurately captured as daily data become available. Although the integration of the SV approach into the Factor MIDAS or Factor TVP-MIDAS models can provide faster and more accurate nowcasting of the unexpected impact of the COVID-19 pandemic on tourist arrivals than the Factor TVP-MIDAS model, it may lead to nowcasts that consistently underestimate the actual values during the normal period. The addition of monthly macroeconomic variables and lagged dependent variables also weakens the predictive power of the daily indexes. Thus, macroeconomic variables appear to be less useful for increasing nowcasting accuracy in times of crises than daily indexes that are updated in real time.

In conclusion, we find that high-frequency data and the Factor MIDAS models with TVPs share a mutually reinforcing relationship. Owing to the lack of regularly updated high-frequency data, these models cannot accurately capture the negative impact of the COVID-19 pandemic. Similarly, the Factor MIDAS model is unable to use the valuable information gained from high-frequency data without TVP specifications. This validates the proposition by Foroni et al. (2022) that finding a reliable TVP model with explanatory variables that change over time is the best solution to improve the accuracy of forecasting and nowcasting during a crisis.

The findings of this study have significant policy implications for the management and recovery of the tourism industry during crises, such as the COVID-19 pandemic. Our study highlights the importance of using reliable forecasting and nowcasting methods that can accurately capture the impact of such crises on tourism demand. Combining a large amount of available and preferably high-frequency data can improve the accuracy of short-term forecasting. Factor MIDAS models with TVPs accurately capture the impact of the pandemic on tourism demand. However, it is essential to update these models with real-time data to ensure their

effectiveness. Our study also highlights that the macroeconomic variables have limited utility in improving nowcasting accuracy during crises compared with daily indexes. Therefore, policymakers should prioritize the integration of real-time data and adjust their forecasting and nowcasting approaches accordingly. The study also reveals the significance of using a TVP model with explanatory variables that change over time, as this can enhance forecasting and nowcasting accuracy during crises. Overall, this study can serve as an alternative benchmark for tourism demand nowcasting during crises, as the baseline forecasts or nowcasts produced by traditional time series or econometric models will rapidly become outdated.

Our study has some limitations that can be addressed in future research. We focus on domestic tourism demand in a single destination in China, but with the lifting of travel restrictions, the patterns of domestic and international tourism demand will differ greatly from the case of Hainan. Therefore, the generalizability of the empirical results to other countries is likely to be limited. In addition, the high-frequency daily data used in our study are obtained exclusively from the Baidu search engine. Future studies may consider other types of high-frequency data, such as those of reviews and hotel bookings. Finally, we limited our study to generating and evaluating point forecasts and nowcasts. Interval and density forecasts can provide information about the uncertainties associated with forecasts and nowcasts, which can be valuable to decision makers in crisis and postcrisis periods. Therefore, these approaches could also be considered in future studies.

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.

Data availability

Data will be made available on request.

Predicting tourism recovery from COVID-9: A time-varying perspective (Original data) (Mendeley Data)

Acknowledgments

This work was supported by the Hong Kong Research Grant Committee [grant number PolyU–15502120] and the National Natural Science Foundation of China [grant number 72004106].

Appendix

Table A
Basic search queries.

No.	Search query	No.	Search query	No.	Search query	No.	Search query
	Dining		***Transportation***	47	Wenchang tourism	71	Yanoda Rainforest Cultural Tourism Zone
1	**Hainan cuisine**	24	**Hainan map**	48	Wanning tourism		***Lodging***
2	Hainanese chicken rice	25	Haikou map	49	Sanya tourism	72	**Hainan accommodation**
3	Coconut chicken	26	Sanya map	50	Sanya tourism tips	73	**Hainan hotels**
4	Wenchang chicken	27	Sanya tourist map	51	Sanya travel agencies	74	5-star hotels in Hainan
5	Hainan specialty	28	Hainan tourist map	52	Sanya travel cost	75	Resorts in Hainan
6	Hainan specialty snacks	29	**Hainan road trip**	53	**Hainan self-guided tour**	76	Sheraton Sanya Resort
7	Hainan specialty fruits	30	Sanya road trip	54	Sanya self-guided tour	77	Haikou hotels
8	Haikou specialty	31	Hainan car rental	55	Sanya self-guided tour tips	78	Sanya hotels
9	Sanya specialty	32	Sanya car rental	56	Wuzhizhou island tips	79	Atlantis Sanya
10	Sanya seafood	33	Haikou car rental	57	**Hainan weather**	80	Resort Intime Sanya
11	Wenchang cuisine	34	**Hainan airlines**	58	Sanya weather	81	Sanya hotel booking
12	Hainan fruit	35	China Eastern Airlines	59	Haikou weather	82	Best hotels in Sanya
13	Sanya cuisine	36	China Southern Airlines	60	Wenchang weather	83	Seaview hotels in Sanya
14	Haikou cuisine	37	**Airline tickets to Hainan**	61	Wanning weather	84	Haitang Bay hotels
	Shopping	38	Airline tickets to Sanya		***Attractions***	85	Yalong Bay hotels
15	**Hainan duty-free shops**	39	Cheap flights to Sanya	62	**Hainan tourist attractions**	86	The Ritz-Carlton Sanya
16	Hainan duty-free shop official website	40	Airline tickets to Haikou	63	Nanwan Monkey Island	87	Mandarin Oriental, Sanya
17	Haikou duty-free shops	41	**Haikou airport**	64	Haikou tourist attractions	88	InterContinental Sanya Resort
18	Haikou duty-free shop official website	42	Sanya airport	65	Holiday Beach		***COVID-19-related***
19	**Sanya shopping**		***Tours***	66	Yalong Bay Underwater World	89	**Coronavirus pandemic latest news in Hainan**
20	Sanya duty-free shop tips	43	**Hainan tourism**	67	Haikou Arcade Streets	90	**Coronavirus pandemic latest news in Sanya**
21	Sanya duty-free shops	44	Haikou tourism	68	Dongjiao Coconut Forest		
22	Sanya duty-free shop address	45	Hainan tourism tips	69	Sanya tourist attractions		
23	Sanya duty-free shop opening hours	46	Sanya one-day trip	70	Sanya Forest Park		

Note: Search queries in bold represent the initial keywords.

References

Andreini, P., Hasenzagl, T., Reichlin, L., Senftleben-König, C., Strohsal, T., 2023. Nowcasting German GDP: foreign factors, financial markets, and model averaging. Int. J. Forecast. 39, 298–313.

Athanasopoulos, G., Hyndman, R.J., Song, H., Wu, D.C., 2011. The tourism forecasting competition. Int. J. Forecast. 27, 822–844.

Baffigi, A., Golinelli, R., Parigi, G., 2004. Bridge models to forecast the euro area GDP. Int. J. Forecast. 20, 447–460.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. In: Timmermann (Ed.), Handbook of Economic Forecasting G. Elliott & A. Elsevier, Amsterdam, pp. 195–237.

Bańbura, M., Giannone, D., Reichlin, L., 2011. Nowcasting. In: Clements, M.P., Hendry, D.F. (Eds.), The Oxford Handbook on Economic Forecasting. Oxford University Press, Oxford, pp. 193–224.

Baumeister, C., Guérin, P., 2021. A comparison of monthly global indicators for forecasting growth. Int. J. Forecast. 37, 1276–1295.

Bobeica, E., Hartwig, B., 2023. The COVID-19 shock and challenges for inflation modeling. Int. J. Forecast. 39, 519–539.

Carriero, A., Clark, T.E., Marcellino, M., Mertens, E., 2022. Addressing COVID-19 outliers in BVARs with stochastic volatility. Rev. Econ. Stat. 1–38 https://doi.org/10.1162/rest_a_01213.
Chatziantoniou, I., Degiannakis, S., Eeckels, B., Filis, G., 2016. Forecasting tourist arrivals using origin country macroeconomics. Appl. Econ. 48, 2571–2585.
Chen, S., Igan, D.O., Pierri, N., Presbitero, A.F., 2020. Tracking the economic impact of COVID-19 and mitigation policies in Europe and the United States. International monetary fund working paper No. 20/125. IMF Working Papers 20.
Coibion, O., Gorodnichenko, Y., Weber, M., 2020. Labor Markets during the COVID-19 Crisis: A Preliminary View. NBER Working Paper No, 27017.
Escribano, Á., Wang, D., 2021. Mixed random forest, cointegration, and forecasting gasoline prices. Int. J. Forecast. 37, 1442–1462.
Ferrara, L., Marcellino, M., Mogliani, M., 2015. Macroeconomic forecasting during the great recession: the return of nonlinearity? Int. J. Forecast. 31, 664–679.
Ferrara, L., Sheng, X.S., 2022. Guest editorial: economic forecasting in times of COVID-19. Int. J. Forecast. 38, 527–528.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model consistency and rates. J. Econom. 119, 231–255.
Foroni, C., Marcellino, M., 2014. A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. Int. J. Forecast. 30, 554–568.
Foroni, C., Marcellino, M., Stevanovic, D., 2022. Forecasting the COVID-19 recession and recovery: lessons from the financial crisis. Int. J. Forecast. 38, 596–612.
Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS regressions: further results and new directions. Econom. Rev. 26, 53–90.
Gössling, S., Scott, D., Hall, C.M., 2021. Pandemics, tourism and global change: a rapid assessment of COVID-19. J. Sustain. Tourism 29, 1–20.
Götz, T.B., Hauzenberger, K., 2021. Large mixed-frequency VARs with a parsimonious time-varying parameter structure. Econom. J. 24, 442–461.
Guérin, P., Marcellino, M., 2013. Markov-switching MIDAS models. J. Bus. Econ. Stat. 31, 45–56.
Hainan Tourism Bureau, 2020. Tourist Volume and Income of Hainan Province. http://lwt.hainan.gov.cn/xxgk_55333/lytj/2019data/202001/t20200115_2735863.html. (Accessed 22 August 2022).
Hao, F., Xiao, Q., Chon, K., 2020. COVID-19 and China's hotel industry: impacts, a disaster management framework, and post-pandemic agenda. Int. J. Hospit. Manag. 90, 102636.
Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., Schreiner, J., 2023. Nowcasting in a pandemic using nonparametric mixed frequency VARs. J. Econom. 232, 52–69.
Hyndman, R.J., 2018. A Forecast Ensemble Benchmark. https://robjhyndman.com/hyndsight/benchmark-combination/. (Accessed 29 August 2022).
Ionides, E.L., Bhadra, A., Atchadé, Y., King, A., 2011. Iterated filtering. Ann. Stat. 39, 1776–1802.
Ionides, E.L., Nguyen, D., Atchadé, Y., Stoev, S., King, A.A., 2015. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 112, pp. 719–724.
Jardet, C., Meunier, B., 2022. Nowcasting world GDP growth with high-frequency data. J. Forecast. 41, 1181–1200.
Korinth, B., 2022. Implications of COVID-19 on tourism sector in Poland: its current state and perspectives for the future. Tour. Recreat. Res. 47, 636–640.
Kourentzes, N., Petropoulos, F., Trapero, J.R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. Int. J. Forecast. 30, 291–302.
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. Ann. Tourism Res. 88, 103197.
Kuzin, V., Marcellino, M., Schumacher, C., 2011. MIDAS vs. mixed-frequency VAR: nowcasting GDP in the euro area. Int. J. Forecast. 27, 529–542.
Larson, W.D., Sinclair, T.M., 2022. Nowcasting unemployment insurance claims in the time of COVID-19. Int. J. Forecast. 38, 635–647.
Lenza, M., Primiceri, G.E., 2020. How to Estimate a VAR after March 2020. NBER Working Paper No. 27771.
Li, G., Song, H.Y., Witt, S.F., 2006. Time varying parameter and fixed parameter linear AIDS: an application to tourism demand forecasting. Int. J. Forecast. 22, 57–71.
Li, H., Song, H., Li, L., 2017a. A dynamic panel data analysis of climate and tourism demand: additional evidence. J. Trav. Res. 56, 158–171.
Li, X., Pan, B., Law, R., Huang, X.K., 2017b. Forecasting tourism demand with composite search index. Tourism Manag. 59, 57–66.
Liew, V.K.S., 2022. The effect of novel coronavirus pandemic on tourism share prices. J. Tourism Futur. 8, 109–124.
Liu, A.Y., Vici, L., Ramos, V., Giannoni, S., Blake, A., 2021. Visitor arrivals forecasts amid COVID-19: a perspective from the Europe team. Ann. Tourism Res. 88, 103182.
Liu, H., Liu, Y., Li, G., Wen, L., 2021. Tourism demand nowcasting using a LASSO-MIDAS model. Int. J. Contemp. Hospit. Manag. 33, 1922–1949.
Lütkepohl, H., 1981. A model for non-negative and nonpositive distributed lag functions. J. Econom. 16, 211–219.
Marcellino, M., Schumacher, C., 2010. Factor MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP. Oxf. Bull. Econ. Stat. 72, 518–550.
Nakajima, Y., Sueishi, N., 2022. Forecasting the Japanese macroeconomy using high-dimensional data. Jpn. Econ. Rev. 73, 299–324.
Page, S., Song, H., Wu, D.C., 2012. Assessing the impacts of the global economic crisis and swine flu on inbound tourism demand in the United Kingdom. J. Trav. Res. 51, 142–153.
Polyzos, S., Samitas, A., Spyridou, A.E., 2021. Tourism demand and the COVID-19 pandemic: an LSTM approach. Tour. Recreat. Res. 46, 175–187.
Primiceri, G.E., 2005. Time varying structural vector autoregressions and monetary policy. Rev. Econ. Stud. 72, 821–852.
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. Ann. Tourism Res. 88, 103155.
Schaer, O., Kourentzes, N., Fildes, R., 2019. Demand forecasting with user-generated online information. Int. J. Forecast. 35, 197–212.
Schorfheide, F., Song, D., 2021. Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. NBER Working Paper No. 29535. SSRN Electron. J.
Schumacher, C., 2014. MIDAS regressions with time-varying parameters: an application to corporate bond spreads and GDP in the euro area: ZBW-Deutsche Zentralbibliothek für Wirtschaftswissenschaften, leibniz-Informationszentrum Wirtschaft. In: Und Hamburg, K. (Ed.), Annual Conference 2014 (Hamburg): Evidence-Based Economic Policy 100289. German Economic Association.
Shen, S.J., Li, G., Song, H.Y., 2009. Effect of seasonality treatment on the forecasting performance of tourism demand models. Tourism Econ. 15, 693–708.
Song, H., Li, G., 2008. Tourism demand modeling and forecasting—a review of recent research. Tourism Manag. 29, 203–220.
Song, H., Lin, S., 2010. Impacts of the financial and economic crisis on tourism in Asia. J. Trav. Res. 49, 16–30.
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. Ann. Tourism Res. 75, 338–362.
Song, H., Witt, S.F., Li, G., 2009. The Advanced Econometrics of Tourism Demand. Routledge, New York.
Song, H., Wong, K.K.F., Chon, K.K.S., 2003. Modeling and forecasting the demand for Hong Kong tourism. Int. J. Hospit. Manag. 22, 435–451.
Speakman, M., Sharpley, R., 2012. A chaos theory perspective on destination crisis management: evidence from Mexico. J. Destin. Market. Manag. 1, 67–77.
Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econ. Stat. 20, 147–162.
United Nations World Tourism Organization, 2022. UNWTO World Tourism Barometer. https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2022-01/220118-Barometersmall.pdf?VersionId=_PBIQdr4u_qM0w56.l0NpfGPzylGu6Md. (Accessed 22 August 2022).
United Nations World Tourism Organization, 2023. UNWTO World Tourism Barometer. https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2023-01/UNWTO_Barom23_01_January_EXCERPT.pdf?VersionId=_2bbK5GIwk5KrBGJZt5iNPAGnrWoH8NB. (Accessed 1 August 2023).
Wen, L., Liu, C., Song, H.Y., Liu, H., 2021. Forecasting tourism demand with an improved mixed data sampling model. J. Trav. Res. 60, 336–353.
Wickramasinghe, K., Ratnasiri, S., 2021. The role of disaggregated search data in improving tourism forecasts: evidence from Sri Lanka. Curr. Issues Tourism 24, 2740–2754.
World Bank, 2022. Global Economic Prospects. https://www.worldbank.org/en/publication/global-economic-prospects. (Accessed 28 September 2022).
Wu, E.H.C., Hu, J., Chen, R., 2022. Monitoring and forecasting COVID-19 impacts on hotel occupancy rates with daily visitor arrivals and search queries. Curr. Issues Tourism 25, 490–507.
Yang, X., Pan, B., Evans, J.A., Lv, B., 2015. Forecasting Chinese tourist volume with search engine data. Tourism Manag. 46, 386–397.
Yang, Y., Fan, Y., Jiang, L., Liu, X., 2022. Search query and tourism forecasting during the pandemic: when and where can digital footprints be helpful as predictors? Ann. Tourism Res. 93, 103365.
Zhang, B.R., Huang, X.K., Li, N., Law, R., 2017. A novel hybrid model for tourist volume forecasting incorporating search engine data. Asia Pac. J. Tourism Res. 22, 245–254.
Zhang, H., Lu, J., 2022. Forecasting hotel room demand amid COVID-19. Tourism Econ. 28, 200–221.
Zhang, H., Song, H., Wen, L., Liu, C., 2021. Forecasting tourism recovery amid COVID-19. Ann. Tourism Res. 87, 103149.