

Katz, Harrison; Savage, Erica; Coles, Peter A.

Article

Lead times in flux: Analyzing Airbnb booking dynamics during global Upheavals (2018-2022)

Annals of Tourism Research Empirical Insights

Provided in Cooperation with:

Elsevier

Suggested Citation: Katz, Harrison; Savage, Erica; Coles, Peter A. (2025) : Lead times in flux: Analyzing Airbnb booking dynamics during global Upheavals (2018-2022), Annals of Tourism Research Empirical Insights, ISSN 2666-9579, Elsevier, Amsterdam, Vol. 6, Iss. 2, pp. 1-16, <https://doi.org/10.1016/j.annale.2025.100185>

This Version is available at:

<https://hdl.handle.net/10419/337843>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



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



Lead times in flux: Analyzing Airbnb booking dynamics during global Upheavals (2018–2022)

Harrison Katz^{a,*}, Erica Savage^{b,2}, Peter Coles^{c,3}

^a Data Science, Forecasting, Airbnb, United States

^b Finance, Forecasting, Airbnb, United States

^c Data Science, Airbnb, United States

ARTICLE INFO

Editor: Dr. Lorenzo Masiero

Keywords:

Forecasting
Lead time distribution
Normalized L1 distance
COVID-19 pandemic
Airbnb
Tourism analytics
Consumer behavior
Hospitality
Revenue management

ABSTRACT

Short-term changes in booking behaviors can significantly undermine naive forecasting methods in the travel and hospitality industry, especially during periods of global upheaval. Traditional metrics like average or median lead times capture only broad trends, often missing subtle yet impactful distributional shifts. In this study, we introduce a normalized L1 (Manhattan) distance to measure the full distributional divergence in Airbnb booking lead times from 2018 to 2022, with particular emphasis on the COVID-19 pandemic. Using data from four major U.S. cities, we find a two-phase pattern of disruption: a sharp initial change at the pandemic's onset, followed by partial recovery but persistent divergences from pre-2018 norms. Our approach reveals shifts in travelers' planning horizons that remain undetected by conventional summary statistics. These findings highlight the importance of examining the *entire* lead-time distribution when forecasting demand and setting pricing strategies. By capturing nuanced changes in booking behaviors, the normalized L1 metric enhances both demand forecasting and the broader strategic toolkit for tourism stakeholders, from revenue management and marketing to operational planning, amid continued market volatility.

1. Introduction

Booking lead time—the interval between the booking date and the actual stay date—is a key metric in the travel and hospitality industry. It influences operational and strategic decisions, including pricing strategies, inventory management, marketing campaigns, and customer relationship management. A thorough understanding of booking lead times enables businesses to optimize resource allocation, anticipate demand fluctuations, and enhance competitiveness in a dynamic market environment.

For Airbnb's own analytics teams and for professional property-management firms that oversee dozens or even hundreds of listings, anticipating *when* bookings will arrive is as important as forecasting *how many*. Accurate lead-time forecasts drive automated pricing rules, trigger targeted promotions, schedule cleaning crews, and help

managers decide when to onboard additional units. Sudden shifts in booking windows therefore have direct financial and operational consequences at the platform and portfolio level, even if a single-listing host may feel them less acutely.

Prior research has underscored the importance of lead times in tourism management. [Lim and McAleer \(2002\)](#) showed how lead times affect revenue management strategies in the hospitality sector. [Tussyadiah and Pesonen \(2016\)](#) examined consumer decision-making processes in tourism, finding that booking lead times are influenced by a complex interplay of factors, including perceived risk, travel purpose, and technological advancements. Lead times are frequently incorporated into forecasts of aggregate demand ([Fiori & Foroni, 2019](#); [Fiori & Foroni, 2020](#); [Pereira, 2016](#); [Webb, Schwartz, Xiang, & Altin, 2022](#); [Zaki, 2022](#)), while [Katz, Brusch, and Weiss \(2024\)](#) and [de Oliveira et al. \(2021\)](#) presented models for forecasting lead times directly. Moreover,

* Corresponding author.

E-mail addresses: harrison.katz@airbnb.com (H. Katz), erica.savage@airbnb.com (E. Savage), peter.coles@airbnb.com (P. Coles).

¹ Bio: Harrison Katz, PhD in Statistics from UCLA, is a Tech Lead Data Scientist on Airbnb's forecasting team. Previously with the Federal Reserve Board of Governors in risk analysis, his research focuses on financial markets, Bayesian forecasting, and compositional time series.

² Bio: Erica Savage is an Associate Principal focused on forecasting and financial modeling at Airbnb.

³ Bio: Peter Coles is Airbnb's Head Economist and Director of Data Science. Formerly Head Economist at eBay and a Harvard professor, his research covers marketplace design, friction reduction, data strategy, and public policy. He holds degrees from Princeton and Stanford.

Song and Li (2008) and Huang and Zheng (2023) have reviewed the broader context of tourism demand forecasting, emphasizing the growing importance of incorporating booking windows into predictive frameworks.

Global events, particularly crises such as pandemics, have profound effects on travel behaviors and lead times. The COVID-19 pandemic, in particular, has led to unprecedented disruptions in the tourism industry (Mueller & Sobreira, 2024; Okafor, Khalid, & Gopalan, 2022; Qiu, Park, Li, & Song, 2020; Sainaghi & Chica-Olmo, 2022; Sharma, Shin, Santa-Maria, & Nicolau, 2021). According to Gössling, Scott, and Hall (2021), the pandemic resulted in massive declines in international and domestic travel, altering consumer risk perceptions and booking behaviors. Sigala (2020a) emphasized that COVID-19 not only impacted demand but also accelerated digital transformation and the adoption of flexible booking policies in the industry. Research suggests that travelers' perceived risk and uncertainty lead to significant changes in booking patterns. For example, Lo, Cheung, and Law (2011) found that during health crises, travelers tend to delay bookings or opt for destinations perceived as safer. Fuchs and Reichel (2011) showed that risk reduction strategies, such as flexible cancellation policies and enhanced safety measures, become more critical in influencing booking lead times.

Additionally, recent research in the airline industry underscores similar shifts in booking lead-time patterns under COVID-19. Zhang, Yang, Wang, Bian, and Zhang (2021) investigated the Chinese aviation market and revealed how pandemic-related uncertainties drove travelers to shorten planning horizons and switch travel modes. Their moment-based analysis demonstrates that variance and higher-order distributional features of lead times can provide actionable insights into passenger behavior, a perspective that aligns well with the comprehensive approach advocated in this paper.

Despite recognition of these shifts, much of the existing literature relies on traditional statistical measures—such as mean and median lead times—to analyze booking behaviors. While these measures are useful, they may not fully capture the distributional changes and increased variability introduced by external shocks like pandemics. As Assaf and Tsonas (2019a, 2019b, 2019d) argued, there is a need for more advanced analytical approaches in tourism research to better understand complex consumer behaviors. Likewise, Assaf and Tsonas (2019c) underscored the importance of distribution-aware metrics in examining tourism demand, especially during times of heightened uncertainty.

This study addresses this gap by introducing the normalized Manhattan (L1) distance metric as a novel approach to analyzing variations in booking lead time distributions. Unlike traditional metrics, the normalized L1 distance captures changes across the entire distribution, providing a more sensitive and comprehensive measure of divergence in booking behaviors over time. This method allows for the detection of shifts that might be overlooked when relying solely on averages or standard deviations.

The primary objective of this research is to assess the effectiveness of the normalized L1 distance metric in detecting and quantifying changes in Airbnb booking lead times during periods of global upheaval, specifically the COVID-19 pandemic. By analyzing data from Airbnb properties across four major U.S. urban centers—Austin, Boston, Miami, and San Francisco—from 2018 to 2022, we aim to provide a comprehensive perspective on the evolving dynamics of urban tourism. Our work also resonates with the emerging body of research on digital platforms' responses to crises; as noted by Tussyadiah and Pesonen (2016) and Gyódi (2021), peer-to-peer accommodation providers (like Airbnb) can exhibit different booking behaviors compared to traditional hotels, a gap that demands further empirical and methodological attention.

Furthermore, this study contributes to the literature by offering empirical evidence on how external shocks impact booking behaviors in the sharing economy context. By situating our findings within the broader theoretical frameworks of risk perception and consumer behavior during crises, we enhance the understanding of how travelers

adjust their planning in response to heightened uncertainty. Our focus on Airbnb properties answers recent calls (e.g., Sigala, 2020) to examine how platform-specific features—such as flexible listing types, decentralized inventory, and digital-first user experiences—shape travel demand under extreme market disruptions.

The normalized L1 distance is used to quantify the divergence in lead time distributions across different years, offering a non-parametric measure that adjusts for scale differences by normalizing the traditional L1 (Manhattan) distance. This approach gives a global summary of lead time distributional changes, allowing a more nuanced understanding of how booking patterns have shifted in response to significant events. This framework provides practical insights for industry stakeholders and also advances methodological approaches in tourism research, as it addresses the need for enhanced analytics that capture full distributional variations rather than focusing on singular statistical descriptors alone.

2. Methodology

This section describes our approach to studying changes in Airbnb booking lead-time distributions from 2018 to 2022. In particular, we define the key variable of interest (§2.1), introduce the normalized L1 (Manhattan) distance for quantifying distributional divergences (§2.3), and outline our time-series decomposition and interpretative strategy (§2.4).

2.1. Data and variable of interest

We analyze Airbnb booking data for four major U.S. cities (Austin, Boston, Miami, and San Francisco) from January 2018 through December 2022. For each city, we consider both *destination* bookings (nights stayed in the city) and *origin* bookings (nights booked from the city), distinguishing between domestic and international travel corridors. In total, the dataset comprises monthly aggregates of booking records, where each record indicates how many nights were booked in a given city (or from a given city) for specific calendar months of stay.

A key variable in our study is the *lead time*, defined as the number of days between the booking date and the check-in date. We confine our analysis to lead times from 0 to 365 days before check-in. Although travelers *can* book more than one year in advance, these cases represent under 0.5 % of total bookings in our dataset and are thus negligible. By restricting the lead-time domain to 0–365 days, we ensure that more than 99 % of all nights are captured, while also simplifying the analysis (e.g., a uniform bin structure for the lead-time distribution).

2.2. Monthly lead-time distributions

Let B_t be the set of all bookings whose check-in date (i.e., first night) falls in month t . For each booking $i \in B_t$,

- $\Delta_i \in \{0, 1, \dots, 365\}$ denotes the lead time (in days) between its booking date and its check-in date,
- ℓ_i denotes the number of nights in booking i .

For each lead time $\Delta \in \{0, \dots, 365\}$, define

$$w_t(\Delta) = \sum_{\substack{i \in B_t \\ \Delta_i = \Delta}} \ell_i,$$

that is, the total number of *nights* contributed by all bookings in B_t whose lead time is exactly Δ .⁴

Since Δ ranges from 0 to 365, we have 366 possible lead-time cate-

⁴ For example, if two guests each book 5-night stays with $\Delta_i = 3$, then they add $5 + 5 = 10$ to $w_t(3)$.

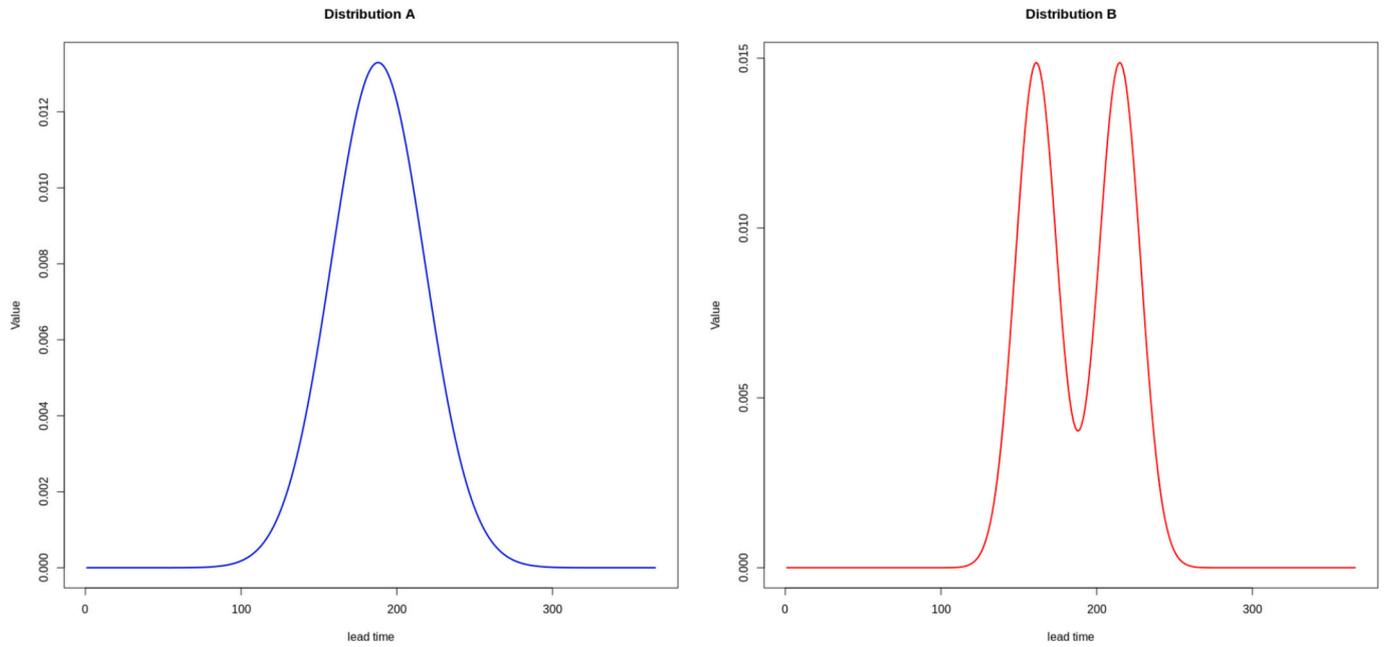


Fig. 1. Example of two lead-time distributions, A (left) and B (right), that share the same mean ($\mu = 188$) and standard deviation ($\sigma = 30$), but exhibit notably different shapes. Distribution A is roughly unimodal, while Distribution B is bimodal. Despite having identical mean and SD, the two distributions have an L1 distance of 0.25, highlighting how summary statistics alone may fail to capture important shifts.

gories. To convert these totals into a *probability distribution*, we normalize:

$$L_t(\Delta) = \frac{w_t(\Delta)}{\sum_{\delta=0}^{365} w_t(\delta)}, \Delta = 0, 1, \dots, 365.$$

Hence, $L_t(\Delta)$ represents the *proportion* of all nights (among bookings that start in month t) that were booked Δ days in advance. By construction,

$$\sum_{\Delta=0}^{365} L_t(\Delta) = 1.$$

2.3. Descriptive analysis

A descriptive analysis is performed for each city, examining each year separately. We look at:

- **Mean trip lead times:** The mean trip lead time provides the average number of days between the booking date and the trip date, weighted by the trip length. It gives an overall sense of how far in advance travelers are planning their trips. The mean is calculated using the formula:

$$\text{Mean} = \frac{1}{n} \sum_{i=1}^n x_i$$

where n is the total number of trips and x_i is the lead time for each trip.

- **Median trip lead times:** The median trip lead time indicates the central tendency of trip lead times, offering a measure less sensitive to extreme outliers compared to the mean. This is particularly useful for understanding typical trip planning behavior in the presence of significant variability. It is calculated as follows:
 - If n (the number of observations) is odd, the median is the middle value:

$$\text{Median} = x_{\left(\frac{n+1}{2}\right)}$$

- If n is even, the median is the average of the two middle values:

$$\text{Median} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+1}{2}\right)}}{2}$$

where $x_{(i)}$ represents the i -th ordered lead time.

- **Standard deviation of trip lead times:** The standard deviation measures the variability in trip lead times, indicating how spread out the planning patterns are around the mean. This statistic helps to assess the consistency of trip planning behaviors over time. The standard deviation is computed as:

$$\text{Standard Deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \text{Mean})^2}$$

where x_i is the lead time for each trip, and the mean is as defined above.

2.4. Divergence analysis using normalized L1 distance

While basic summary statistics (mean, median, and standard deviation) can highlight overall trends in lead times, they may overlook *distributional shifts* that arise during periods of upheaval. To address this, we employ the *normalized L1 distance* (a.k.a. the Manhattan distance).

2.4.1. Definition

The L1 distance between two probability distributions p and q on the same domain (in our case, $\Delta = 0, 1, \dots, 365$) essentially sums the absolute differences of their probabilities at each point:

$$D_{L1}(p, q) = \sum_{\Delta=0}^{365} |p(\Delta) - q(\Delta)|.$$

If the distributions do not overlap at all (one is 1 in places where the other is 0, and vice versa), the L1 distance is maximal. Because we want a measure in $[0, 1]$, we divide by 2:

$$D(p, q) = \frac{1}{2} \sum_{\Delta=0}^{365} |p(\Delta) - q(\Delta)|.$$

Interpretation: $D(p, q)$ is the fraction of the total probability mass that has “shifted.” A value of 0 means no difference (identical distributions), and 1 implies total divergence.

Fig. 1 illustrates why L1 distance can be more informative. Distributions A and B both have mean 188 and standard deviation 30, yet differ substantially in shape (unimodal vs. bimodal). Commonly used summary statistics do not reveal these differences, whereas the normalized L1 distance reveals a notable shift: here, $D(A, B) = 0.25$, indicating that 25 % of the probability mass has effectively “moved.”

2.4.2. Year-over-year vs. baseline comparisons

We calculate the normalized L1 distance between lead-time distributions in two ways:

1. Year-over-Year (YoY) comparison

$$l_{YoY,t}^i = \frac{1}{2} \sum_{\Delta=0}^{\Delta_{max}} |L_t(\Delta) - L_{t-12}(\Delta)|.$$

- This measures how much the distribution in month t diverges from the same calendar month in the previous year.
- Baseline-2018 Comparison:

$$l_{YoY2018,t}^i = \frac{1}{2} \sum_{\Delta=0}^{\Delta_{max}} |L_t(\Delta) - L_{2018,t}(\Delta)|,$$

where $L_{2018,t}(\Delta)$ denotes the distribution for the corresponding month of 2018. This approach captures how far month t is from a pre-pandemic benchmark.

These two perspectives reveal distinct patterns. The YoY metric highlights sequential shifts (e.g., from 2020 to 2021), while the Baseline-2018 metric clarifies whether distributions are returning to pre-crisis norms or have settled into new equilibria.

2.5. Seasonal-Trend Decomposition (STL)

After computing the L1 distances over time, we apply a Seasonal-Trend decomposition using Loess (STL) (Cleveland, Cleveland, McRae, Terpenning, et al., 1990) to each L1 time series. The STL method splits each series y_t into three components:

$$y_t = T_t + S_t + R_t,$$

where T_t is a slowly varying trend, S_t is a seasonal component capturing recurring monthly or weekly patterns, and R_t is a remainder (irregular fluctuations). We prefer STL for its:

- Robustness to Outliers: It uses iterative Loess smoothing to handle sudden spikes or dips.
- Flexibility in Seasonal Patterns: Each city may show distinct seasonalities (e.g., academic schedules, major events).

Ease of Interpretation: We can pinpoint the long-term shift in lead-

time divergences by examining T_t .

Using STL on $l_{YoY,t}^i$ and $l_{YoY2018,t}^i$ helps identify whether divergence spikes are driven mainly by structural (long-term) changes or by seasonal and event-driven fluctuations.

2.6. Forward-looking extensions of the L1 distance

Booking curves are often needed *before* they are complete. We therefore define a “partial” normalized L1 distance that compares the in-progress lead-time distribution to its historical analogue at the same horizon H days before arrival. Concretely:

- Partial Distribution Construction: For each arrival date, identify what fraction of total bookings (so far) accrued at lead times from 0 to H days out, where H is a “checkpoint” (e.g., 30 days before arrival).
- Comparative Baseline: Compute an average partial distribution at H days out using data from prior years or seasons. This yields a historical reference $L_{hist}^{(H)}$ to which the current partial distribution $L_t^{(H)}$ can be compared.
- Partial L1 Metric:

$$D_H(t) = \frac{1}{2} \sum_{\delta=0}^H |L_t^{(H)}(\delta) - L_{hist}^{(H)}(\delta)|.$$

- A higher $D_H(t)$ signals that current booking behavior is unfolding quite differently from the norm.
- Early Warning: When $D_H(t)$ exceeds a chosen threshold, managers can anticipate a final distribution that may deviate from historical patterns. They might respond by adjusting prices, experimenting with targeted promotions, or revising inventory allocations.

Because these “forward-looking” metrics capture divergences *before* the final booking curve is observed, they could serve as an *early-warning tool* in times of heightened volatility. While we have not implemented this approach in the current analysis, it represents a natural follow-up for practitioners seeking near-real-time anomaly detection.

3. Toy example: impact of lead time changes on naive forecast methods

3.1. Setup

To illustrate the sensitivity of the normalized L1 metric to shifts in lead time distributions and to demonstrate how changes in lead times can affect commonly used forecast methods, we present a simulated example using a fictional town called B-Ville. In this example, we assume that in the year 2019, B-Ville had 1000 total bookings for a specific trip date, and in 2020, it had 1200 total bookings for the same trip date.

We created a lead time distribution—a compositional vector of size 30—representing the proportion of bookings made at each lead time from 1 to 30 days before the trip date. To simulate changes in booking behaviors from 2019 to 2020, we introduced a small perturbation to the 2019 lead time distribution by adding random noise drawn from a half-normal distribution with scale parameter $\sigma = 0.05$ to each component and then re-normalizing the vector to ensure it sums to one. This resulted in the 2020 lead time distribution, reflecting a shift in booking patterns. The lead time distributions for both years are shown in Fig. 2.

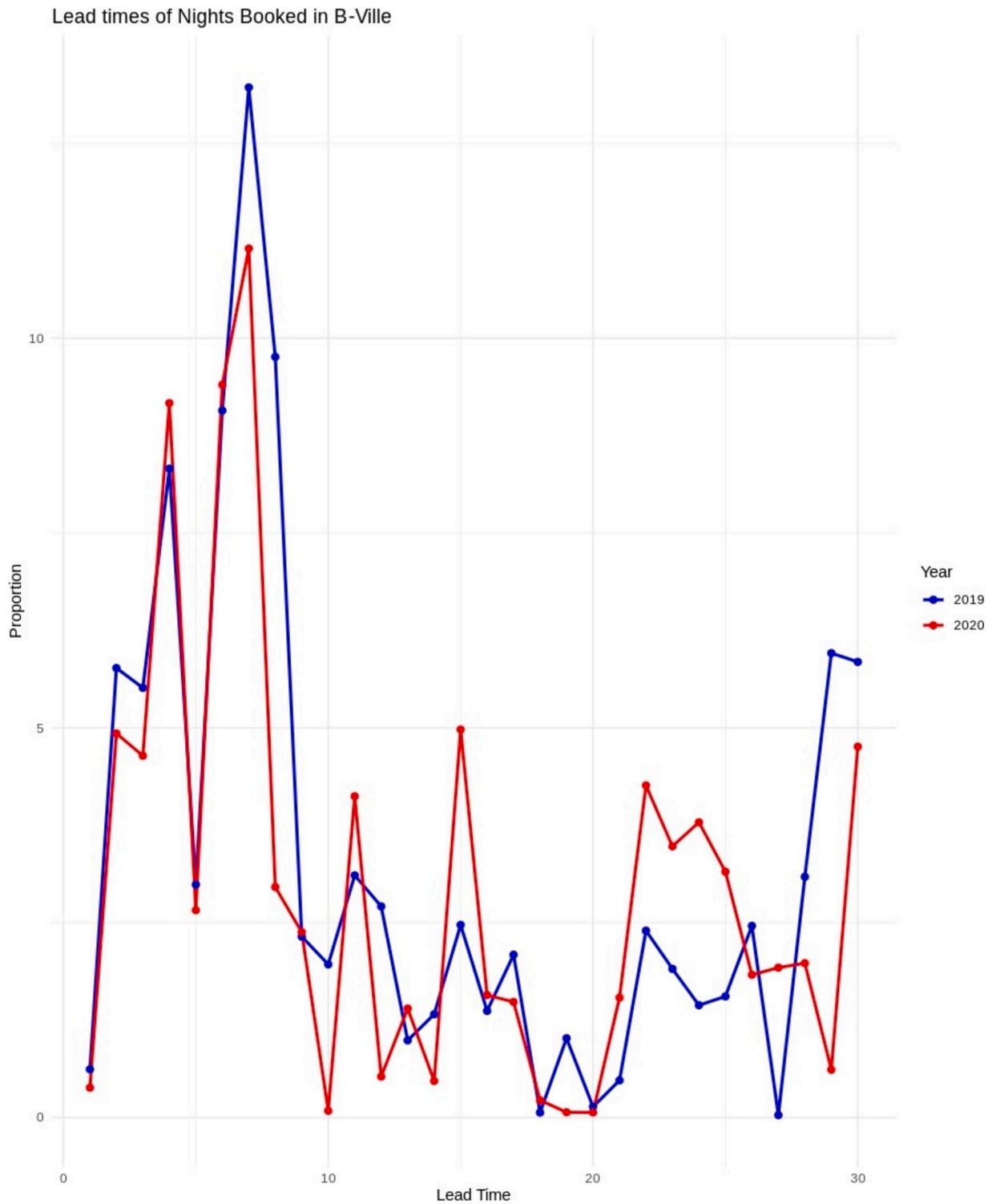


Fig. 2. The distribution of lead times for the fictional town B-Ville, illustrating the changes in booking behaviors across two consecutive years. This figure shows how the normalized L1 distance metric captures small perturbations in lead times better than traditional summary statistics like mean, median, or standard deviation, thereby emphasizing its sensitivity to distributional changes in booking behaviors.

3.2. Results and effects on the pickup method

The pickup method⁵ is a common forecasting technique (Pereira, 2016) in the hospitality industry, where future bookings are predicted based on historical booking patterns and current observed bookings. To estimate the pickup curves, we calculated the cumulative sum of the lead time distributions for each year. The pickup curves represent the accumulation of bookings as the trip date approaches, showing the cumulative percentage of total bookings made up to each lead time.

For instance, suppose we are 17 days before the trip date in 2020 and have recorded 723 bookings so far. In 2019, the pickup curve indicates that by 17 days before the trip date, 69 % of bookings had been made. Applying the multiplicative pickup method, we would predict the total number of bookings for 2020 to be

$$\text{Forecasted Bookings} = \frac{\text{Current Bookings}}{\text{Cumulative\%att} = 17(\text{from 2019 data})} = \frac{723}{0.69} \approx 1,047.$$

However, the actual total bookings in 2020 are 1200. This means that the forecasted bookings underestimate the actual bookings by approximately 150 bookings, or 12.5 %. This discrepancy arises because the lead time distribution shifted in 2020, with more bookings occurring closer to the trip date—a change captured by the normalized L1 distance between the 2019 and 2020 lead time distributions, which is 0.2156. This is illustrated in Fig. 3, which displays both the historical pickup curves and the forecasted total bookings at each time point.

Traditional summary statistics fail to capture this change effectively. As shown in Table 1, the mean lead time increased by only 3.58 %, the median remained unchanged, and the standard deviation decreased by 4.30 %. These minor changes in conventional metrics mask the significant distributional shift affecting the pickup method’s forecast.

This example illustrates how changes in lead time distributions, as quantified by the normalized L1 metric, can impact the accuracy of forecasts made using historical lead times. When lead time distributions shift, relying solely on historical booking patterns may lead to underestimation or overestimation of future bookings.

3.3. Bounding the forecast error of naive methods using the normalized L1 distance

We aim to derive a bound on the relative forecast error ϵ of the pickup method that incorporates both the normalized L1 distance D between the historical and actual lead time distributions and the time remaining until the trip date.

Let:

- Δ_{max} is the maximum lead time considered (e.g., 365 days).
- $L_{hist}(k)$ and $L_{actual}(k)$ denote the historical and actual lead time distributions, respectively, where k ranges from 1 to Δ_{max} .
- $C_{hist}(\Delta) = \sum_{k=1}^{\Delta} L_{hist}(k)$ and $C_{actual}(\Delta) = \sum_{k=1}^{\Delta} L_{actual}(k)$ are the cumulative distributions up to lead time Δ .
- $D = \frac{1}{2} \sum_{k=1}^{\Delta_{max}} |L_{hist}(k) - L_{actual}(k)|$ is the normalized L1 distance between the distributions.

The relative forecast error ϵ of the pickup method is defined as:

$$\epsilon = \frac{\hat{B}_{total} - B_{total}}{B_{total}}$$

⁵ At the city-level and monthly granularity we study, Airbnb supply is effectively elastic—hosts can add or release nights in response to price signals—so modeling strict capacity limits would not change our focus on shifts in booking timing; explicit constraints would, of course, be required in settings with fixed inventory such as hotels.

where:

- \hat{B}_{total} is the forecasted total number of bookings using the pickup method.
- B_{total} is the actual total number of bookings.

Using the pickup method, the forecasted total bookings are calculated as:

$$\hat{B}_{total} = \frac{B_{obs}}{C_{hist}(\Delta)}$$

where:

- B_{obs} is the number of bookings observed up to lead time Δ .
- $C_{hist}(\Delta)$ is the cumulative historical booking proportion up to lead time Δ .

At time Δ , the actual cumulative booking proportion is:

$$C_{actual}(\Delta) = \frac{B_{obs}}{B_{total}}$$

Substituting these expressions into Equation [eq:relative forecast error], we derive (see Appendix 5.5) the following bound on the relative forecast error ϵ :

$$|\epsilon| \leq \frac{2D \left(1 - \frac{\Delta}{\Delta_{max}}\right)}{C_{hist}(\Delta)}$$

This bound decreases as Δ increases (i.e., as we approach the trip date) and approaches zero when $\Delta = \Delta_{max}$.

3.3.1. Estimating the normalized L1 distance D

It is important to note that D represents the divergence between the historical and actual lead time distributions, which is unknown in real time because the actual distribution $L_{actual}(k)$ is not fully observed until after the trip date. However, practitioners can estimate D by:

- Historical Analysis: Examining historical patterns of divergence during similar periods or events to estimate a typical or maximum value of D .
- Scenario Analysis: Considering specific values of D based on past maximum observed divergences to model potential forecast errors under different scenarios.
- Real-Time Monitoring: Tracking early indicators of changes in booking behaviors (e.g., sudden shifts in booking pace) to adjust the estimated D .

By using these methods, forecasters can approximate D and apply the bound in Equation [eq:error bound improved] to assess the potential forecast error under various scenarios.

3.3.2. Example

Revisiting the toy example of B-Ville, suppose we estimate D based on historical divergences:

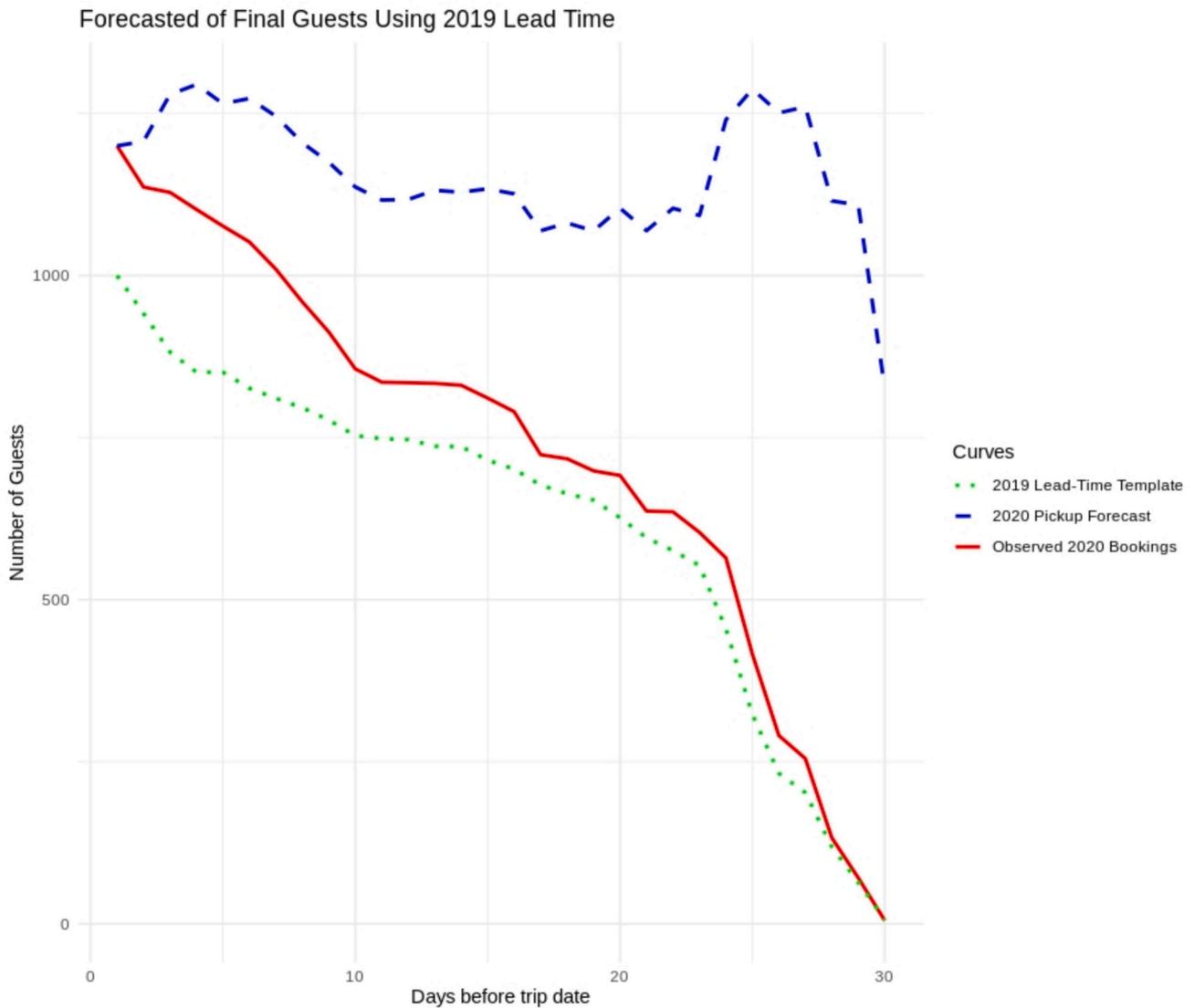


Fig. 3. Pickup curves—the cumulative share of bookings by days before arrival for B-Ville in 2019 and 2020. The 2019 Lead-Time Template (green, dotted) is the 2019 pickup curve, and the pickup method assumes this pattern will hold for 2020. The 2020 Pickup Forecast (blue, dashed) is produced by rescaling the 2019 template each day to the bookings already on the books. The Observed 2020 Bookings curve (red, solid) shows the realized 2020 trajectory. Because that trajectory departs unevenly from the 2019 shape, the forecast alternately over- and under-predicts final demand across the booking window. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Results from simulated lead times data in B-Ville. The L1 distance change better captures the distributional shifts compared to mean, median, or standard deviation alone.

	Year 1	Year 2	Change (%)
Mean	12.38066	12.82447	3.58
Median	8	8	0
Standard deviation	9.417028	9.012016	-4.30
L1 Distance change		0.2156462	

- Estimated Normalized L1 Distance: $D = 0.2156$ (based on a previously observed maximum divergence).

Maximum Lead Time: $\Delta_{\max} = 30$ days.

- At $\Delta = 17$ days (13 days before the trip date):

$$1 - \frac{\Delta}{\Delta_{\max}} = 1 - \frac{17}{30} = \frac{13}{30}$$

$$C_{\text{hist}}(17) = 0.69$$

$$|\epsilon| \leq \frac{0.2156 \times 2 \times \frac{13}{30}}{0.69} \approx 0.27$$

- At $\Delta = 25$ days (5 days before the trip date):

$$1 - \frac{\Delta}{\Delta_{\max}} = 1 - \frac{25}{30} = \frac{5}{30} = \frac{1}{6}$$

$$C_{\text{hist}}(25) = 0.90$$

$$|\epsilon| \leq \frac{0.215 \times 2 \times \frac{1}{6}}{0.90} \approx 0.08$$

This example demonstrates how D allows practitioners to use the bound to assess potential forecast errors at different lead times.

Note.

The detailed derivation of this bound is provided in Appendix 5.5, where we explain how it arises from the properties of cumulative distributions and the normalized L1 distance.

4. Airbnb results

In this section, we present both conventional summary metrics (mean, median, standard deviation) and the normalized L1 divergence for booking lead times across four U.S. cities (Austin, Boston, Miami, and San Francisco) from 2018 to 2022. Our aim is to show how these two approaches complement each other in revealing shifts in booking behavior—while mean or median changes capture overall trends in travelers’ planning horizons, the L1 metric uncovers subtle distributional movements that may not be reflected in central tendencies.

4.1. Evolution of Lead times

The analysis of Airbnb booking data from 2018 to 2022 shows significant variations in lead times across the cities of Austin, Boston, Miami, and San Francisco. We examined the mean, median, and standard deviation (SD) of lead times, segmented by domestic and international travel, and categorized by origin (bookings from the city) and destination (bookings to the city). Figs. 4 and 5 show the trends in mean and median lead times, while Supplementary Tables 1 and 2 provide detailed yearly summaries.

4.1.1. 2018–2019: pre-pandemic period

During the pre-pandemic period, Miami consistently had the shortest median lead times for both inbound and outbound domestic travel, suggesting a higher propensity for last-minute bookings among travelers to and from Miami. For international travel, bookings originating from Miami and those destined for Austin had the shortest median lead times. In contrast, San Francisco had the longest median lead times for domestic and international destination bookings, indicating that travelers to San Francisco tended to plan their trips further in advance. On an origin basis, Boston had the longest median lead times for both domestic and international bookings, suggesting that residents of Boston planned their trips earlier than those from other cities.

4.1.2. 2020: impact of the COVID-19 pandemic

The onset of the COVID-19 pandemic in 2020 caused significant disruptions in travel patterns, particularly in international travel. International destination median lead times surged by at least 20 % in every city, with Boston and San Francisco experiencing the most substantial increases of 64 % and 42 %, respectively. The SD of lead times also reached its highest levels for international travel in 2020, reflecting heightened variability in booking behaviors.

Domestic travel had mixed responses. While Boston’s domestic destination median lead times increased by 17 %, both San Francisco and Miami experienced decreases of 11 % and 8 %, respectively. This suggests that travelers to certain cities reacted differently to the pandemic’s onset. The SD for domestic travel lead times increased in several cities, indicating greater variability amid uncertainty.

4.1.3. 2021: recovery and shifts in booking behavior

In 2021, as travel began to rebound, domestic lead times increased across all cities. Notably, Miami’s origin median lead time surged by 35 %, and Austin’s destination median lead time increased by 21 %. This suggests a return to more advanced planning for domestic trips. For international travel, lead times generally decreased in most cities, possibly reflecting pent-up demand and travelers’ eagerness to book trips as restrictions eased. San Francisco and Miami saw significant reductions in destination median lead times by 48 % and 52 %, respectively. Boston’s origin median lead time for international travel decreased by 29 %.

4.1.4. 2022: stabilization and continued variability

By 2022, changes in mean and median lead times for domestic travel were generally less pronounced, suggesting a stabilization of booking behaviors. However, San Francisco experienced a notable 28 % increase in destination median lead times, indicating that travelers to San Francisco were planning trips even further in advance than before. The SD reached its highest level for all origin domestic travel in 2022, implying persistent variability.

For international travel, mean and median lead times increased again for all cities, reflecting a cautious return to international travel with longer planning horizons. San Francisco experienced the largest year-over-year increases in origin mean and median lead times at 27 % and 31 %, respectively. Similarly, bookings to Miami saw median lead times increase by 23 % and mean lead times by 22 %.

4.2. Divergence analysis using normalized L1 distance

To capture shifts across the entire lead-time distribution, we use the normalized L1 distance, calculated both relative to the baseline year 2018 ($l_{YoY2018,t}^1$) and relative to the same month in the previous year ($l_{YoY,t}^1$). Figs. 6–7 depict the baseline comparisons, while Figs. 8–9 illustrate the year-over-year divergences. The associated numeric summaries appear in Supplementary Tables 7–10.

4.2.1. Baseline comparison divergence ($l_{YoY2018,t}^1$)

Comparisons with 2018 underscore how dramatically the pandemic reshaped booking behaviors. For international travel, $l_{YoY2018,t}^1$ peaked in 2020 across all four cities, indicating substantial distributional disruption. In Boston, for instance, the median divergence for international destination bookings rose from 0.05 in 2019 to 0.34 in 2020, while in Austin it climbed from 0.05 to 0.17. Although these divergences eased somewhat in 2021 and 2022, they generally remained elevated compared to pre-pandemic levels, suggesting that international booking patterns never fully reverted to their 2018 norms.

A similar story emerges for domestic bookings: origin San Francisco’s median $l_{YoY2018,t}^1$ increased from 0.03 in 2019 to 0.11 in 2020 and stabilized around 0.11–0.12 thereafter, while destination Boston’s

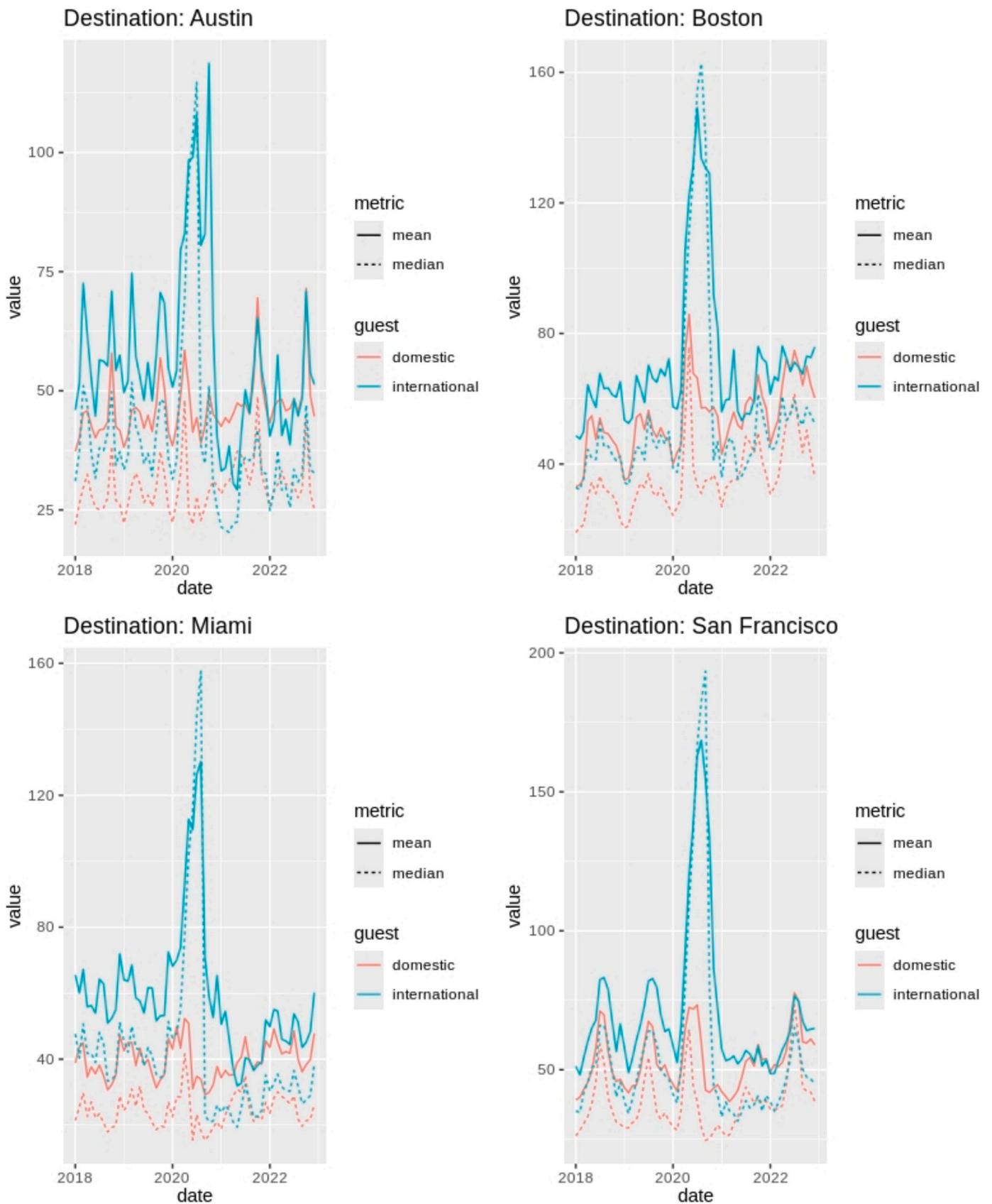


Fig. 4. Trends in mean (solid line) and median (dashed line) lead times for Airbnb bookings to Austin, Boston, Miami, and San Francisco from 2018 to 2022. The figure differentiates domestic (red) and international (blue) travel, showing how booking behaviors evolved over time, particularly during significant events such as the COVID-19 pandemic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

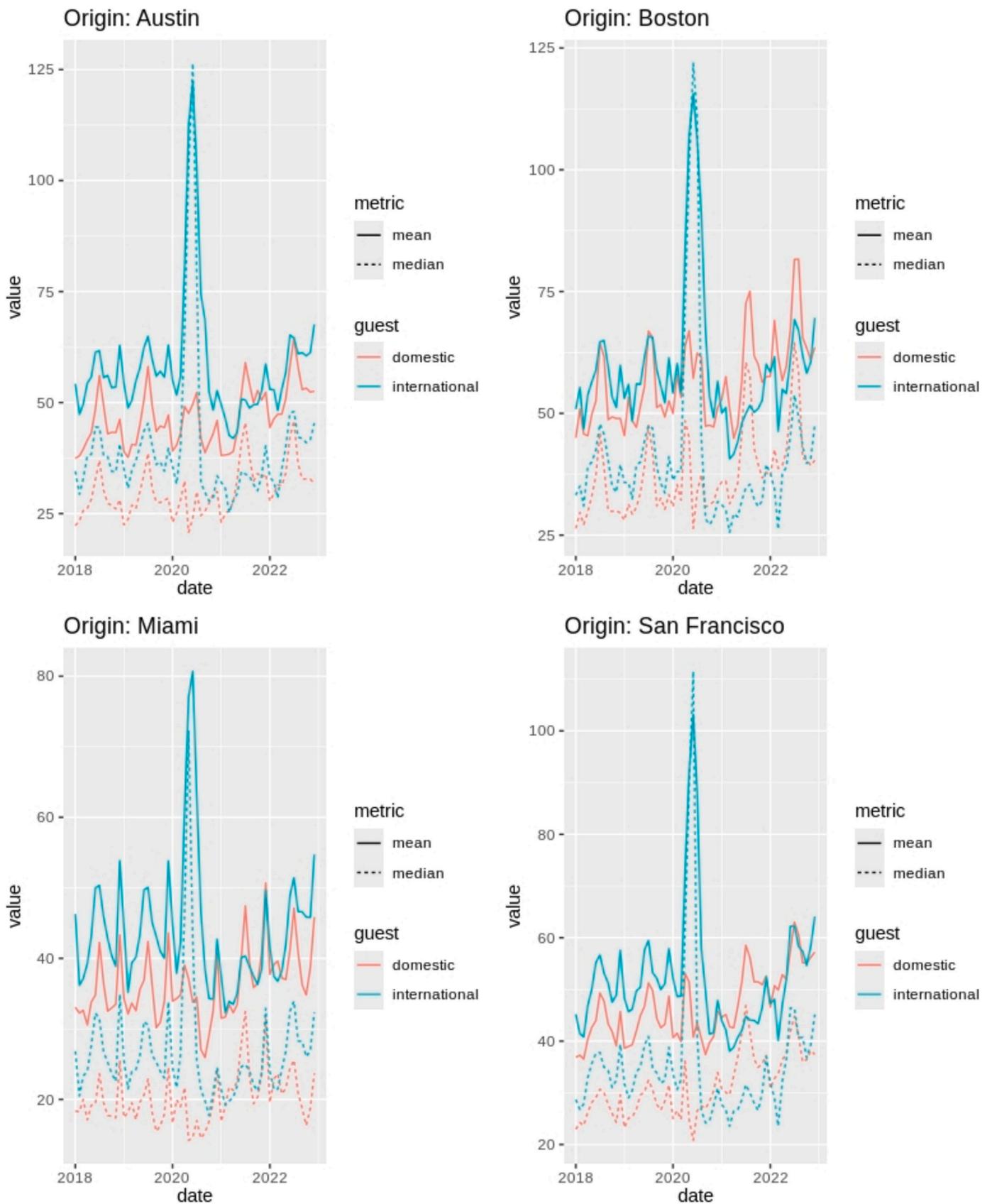


Fig. 5. Trends in mean (solid line) and median (dashed line) lead times for Airbnb bookings from Austin, Boston, Miami, and San Francisco from 2018 to 2022. The figure differentiates domestic (red) and international (blue) travel, showing how booking behaviors evolved over time, particularly during significant events such as the COVID-19 pandemic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

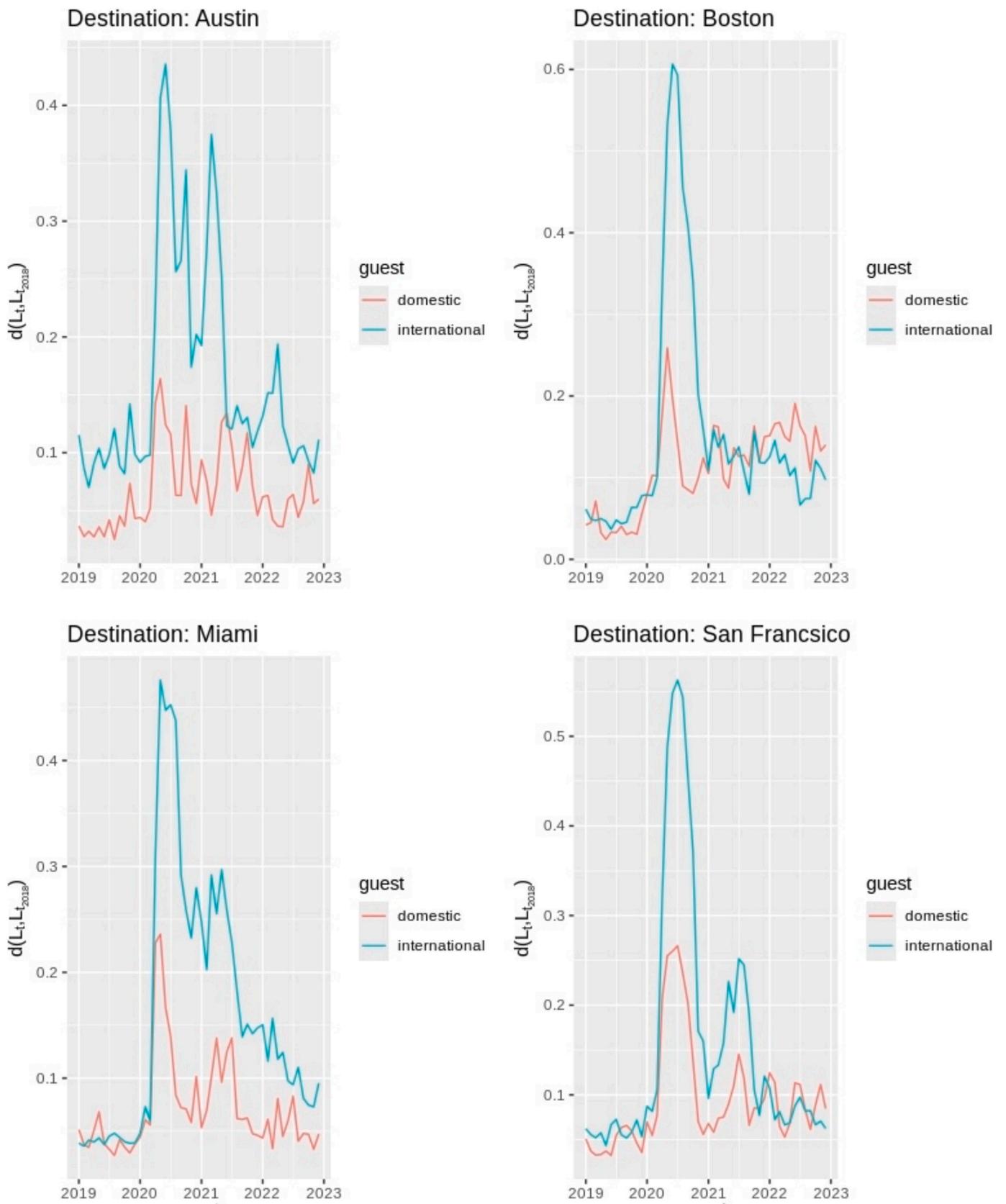


Fig. 6. Normalized L1 divergence in destination lead times for Airbnb bookings, taking 2018 as the baseline year.

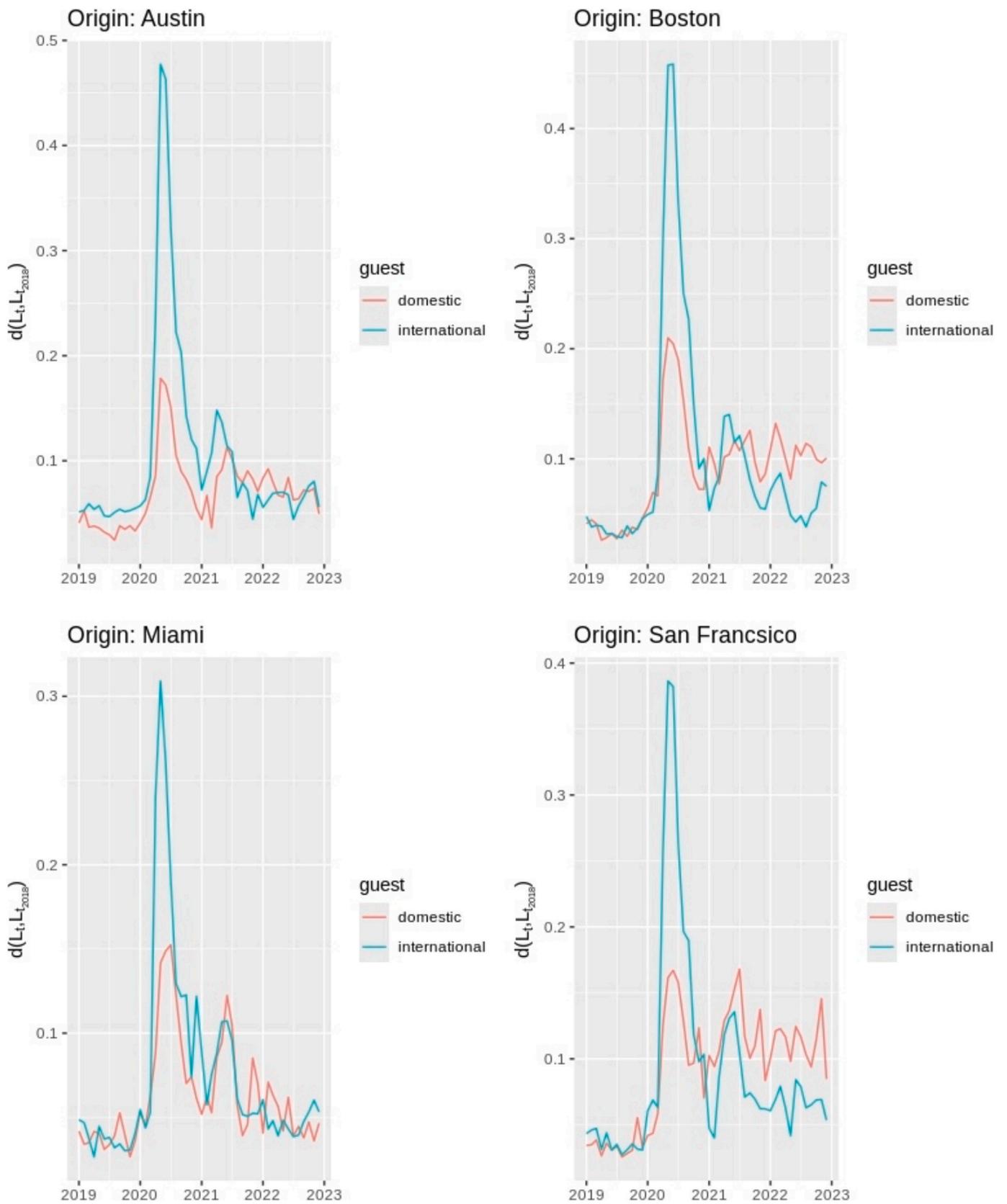


Fig. 7. Normalized L1 divergence in origin lead times for Airbnb bookings, taking 2018 as the baseline year.

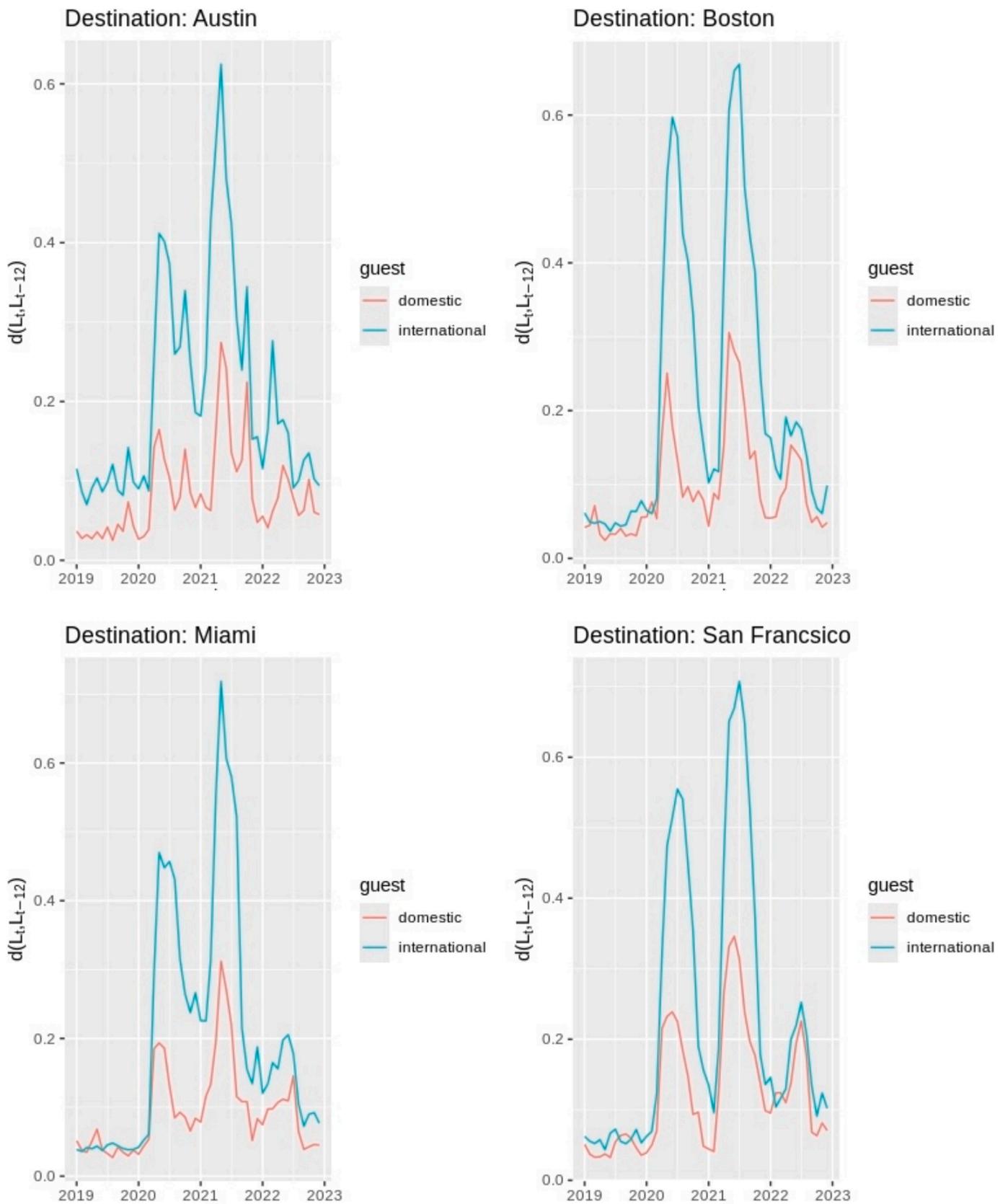


Fig. 8. Normalized L1 divergence in destination lead times for Airbnb bookings, taking the previous year as the baseline year.

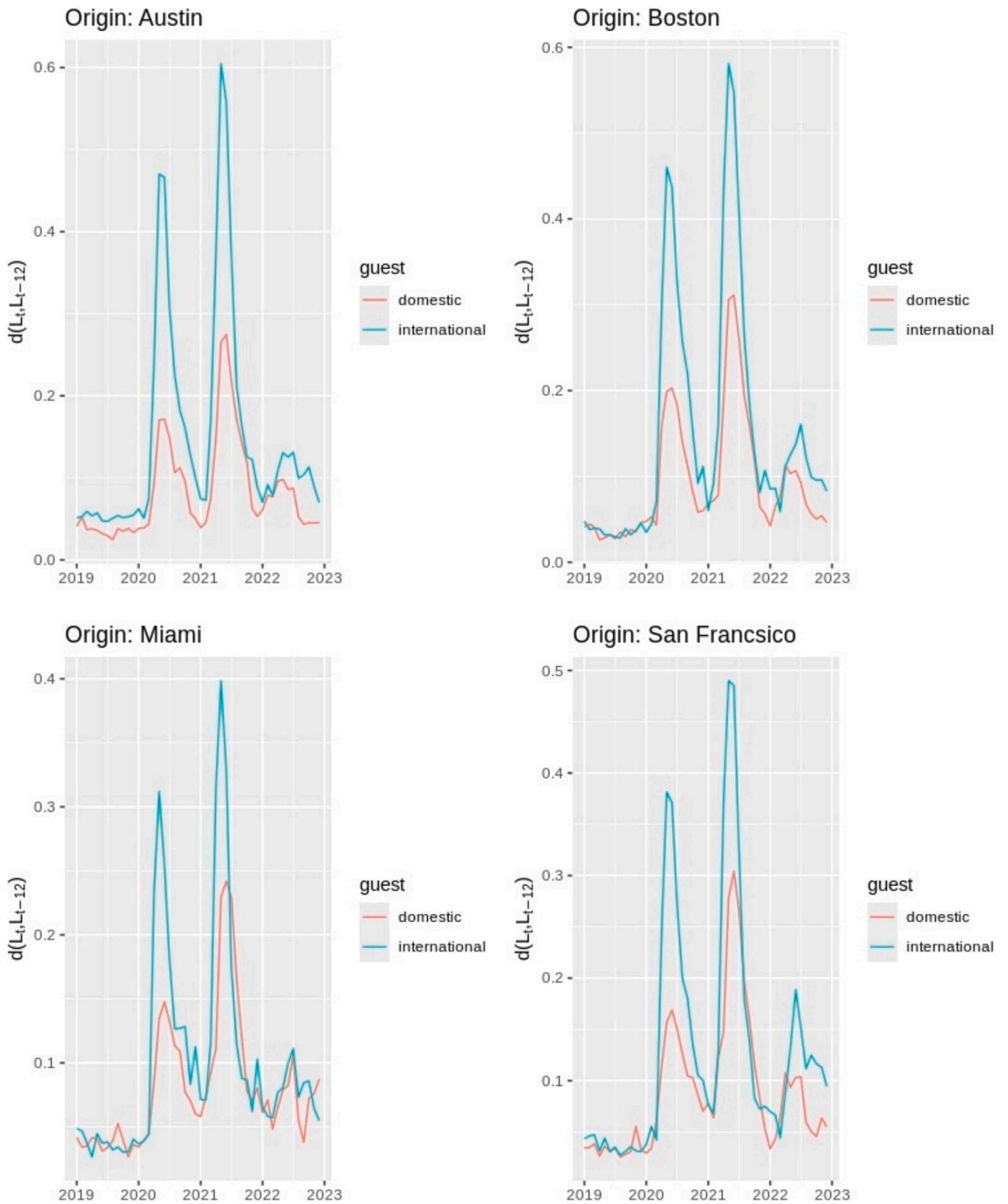


Fig. 9. Normalized L1 divergence in origin lead times for Airbnb bookings, taking the previous year as the baseline year.

divergence rose from 0.03 to 0.10 and ultimately reached 0.15 by 2022. These sustained gaps indicate that, even when the mean or median domestic lead time did not exhibit a massive year-to-year swing, *parts* of the distribution migrated toward notably shorter or longer booking windows. As a result, the overall distribution in 2022 still looked markedly different from its 2018 baseline.

4.2.2. Year-over-year divergence ($l_{YoY,t}^1$)

Year-over-year divergences (Figs. 8–9) highlight temporal shifts *within* the pandemic period. For domestic travel (red lines), $l_{YoY,t}^1$ often surged in 2021 as the industry began recovering, then tapered in 2022. In San Francisco, the median $l_{YoY,t}^1$ for domestic destination bookings jumped from 0.04 in 2019 to 0.12 in 2020 and 0.19 in 2021, mirroring the transitions seen in the simpler metrics but adding clarity on *how much* of the distribution had changed. Meanwhile, in Miami, the domestic origin year-over-year divergence rose from 0.04 in 2019 to 0.08 in 2020 and 0.10 in 2021, underscoring ongoing flux.

For international bookings (blue lines), peaks occurred in either 2020 or 2021. The median $l_{YoY,t}^1$ for bookings from Miami, for instance, reached 0.13 in 2020 before easing slightly; in Boston, the international destination divergence swelled from 0.05 in 2019 to 0.33 in 2020 and 0.39 in 2021. Although these values generally drifted downward in 2022, they remained well above pre-pandemic norms, reflecting prolonged turbulence in international travel. Indeed, the “two-phase” anomaly—where 2020 itself becomes a new, unconventional baseline—left many city pairs experiencing persistently high year-over-year divergences in 2021.

4.2.3. Correlation analysis of lead time divergences

Supplementary Figs. 1 and 2 present correlations between domestic and international l^1 divergences, separating same-month effects from those that emerge after a one-month lag. High correlations within the same travel type and direction suggest consistent behavioral shifts, while lower correlations between different travel types indicate distinct factors influencing domestic and international booking patterns.

In Miami, there was a strong correlation between domestic and international divergences, suggesting that local conditions similarly affected both travel segments. For instance, Miami’s domestic origin $l_{YoY2018,t}^1$ (bottom-left panel of Supplementary Fig. 1) had a correlation of 0.83 with its international destination divergence. In contrast, San Francisco (bottom-right panel of Supplementary Fig. 1) exhibited lower correlations between domestic and international travel, implying that different factors influenced booking behaviors in these segments.

Lag effects were also observed, particularly in Austin, where domestic origin trends influenced international origin trends with a delay (correlation of 0.76 in $l_{YoY2018,t}^1$). Consistently high correlations within categories across cities indicate a degree of interconnectedness in travel patterns, with domestic travel patterns significantly influencing international travel in some cases.

4.2.4. Seasonal-Trend Decomposition (STL) analysis

The Seasonal-Trend decomposition using LOESS (STL) on the normalized L1 distance time series for both the baseline year 2018 ($l_{YoY2018,t}^1$) and the year-over-year ($l_{YoY,t}^1$) divergences are shown in Supplementary Figs. 3–6. The trend components from the STL give insights into the persistent changes in booking lead time distributions over time.

For the $l_{YoY2018,t}^1$ divergences (Supplementary Figs. 3–4), the trend components for both domestic and international travel peaked around mid-2020 to mid-2021, corresponding with the peak impact of the COVID-19 pandemic on global travel. These trend components do not return to zero by 2022, indicating that booking lead time distributions have not reverted to their pre-pandemic (2018) patterns. For example,

the trend components for domestic origin divergences in Boston and San Francisco remain elevated going into 2023, suggesting sustained shifts in booking behaviors in these markets. This persistent divergence implies that travelers continue to book accommodations with lead times significantly different from those in 2018, reflecting lasting changes in planning behavior possibly due to altered risk perceptions, remote work arrangements, or new travel preferences.

In contrast, the $l_{YoY,t}^1$ divergences (Supplementary Figs. 5–6) show a different pattern. For domestic travel, the trend components generally decrease toward pre-pandemic levels by 2022, indicating that year-over-year changes in booking behaviors are stabilizing. However, the divergences remain higher than zero, suggesting that while the rate of change is normalizing, booking patterns have not fully returned to previous norms. For international origin travel, the trend components show an acceleration of divergence going into 2023. This indicates that international booking behaviors continue to change significantly compared to the immediate past year.

5. Discussion

5.1. Key insights on lead-time distributions

Our analysis shows that the COVID-19 pandemic triggered marked changes in Airbnb booking lead times, often in ways that simple mean or median metrics would fail to capture. By studying the *entire* distribution of lead times through the normalized L1 distance, we revealed substantial shifts, including cases where domestic travel in certain cities (e.g., Austin, Miami) rebounded more quickly, whereas others (e.g., Boston, San Francisco) continued to diverge from 2018 baselines. While these differences could reflect variations in local economic composition, traveler preferences, or policy approaches (Sigala, 2020), our data do not directly measure the causal impact of local regulations. Rather, the persistent L1 divergences remind us that forecasting models relying heavily on historical norms may underperform when external shocks alter traveler behavior in fundamental ways.

Several city-year examples illustrate how L1 can capture internal shifts that remain obscured by stable or modest changes in averages. For instance, in San Francisco’s domestic destination travel during 2020, the mean lead time rose by only about 2 % (see Supplementary Table 5), yet the mean $l_{YoY2018,t}^1$ exceeded 0.16 (Supplementary Table 9), implying that a notable fraction of bookings migrated to the extremes (very short or very long lead times). Likewise, destination Miami’s domestic bookings in 2020 showed a mean that was nearly unchanged from 2018, yet the L1 distance spiked to around 0.11, indicating tail-driven reallocation. Boston’s 2020 domestic origin market is even more instructive—its median lead time decreased by 3 % over 2019, while the median $l_{YoY,t}^1$ was 0.10 (Supplementary Table 7). Thus, although many travelers booked closer to the travel date, a sizable subset shifted to much earlier or much later bookings, preserving wide dispersion. In each case, L1 reveals the depth of these distributional shifts, underscoring how stable or even partially “recovered” means and medians can mask significant reconfigurations of booking horizons.

5.2. Implications for forecasting and revenue management

When the L1 distance rises, the historical booking curve becomes an unreliable reference and naïve pickup rules are prone to systematic error—under-predicting if reservations shift toward last-minute booking or over-predicting if guests start planning further ahead. Revenue managers can respond in three complementary ways. First, they should regularly recalibrate their forecasting models with distribution-aware metrics that capture recent shifts in booking timelines. Second, treating the L1 distance as an early warning signal enables “what-if” scenario analysis that brackets best- and worst-case demand before launching

new prices or campaigns. Finally, because international lead-time patterns often diverge more sharply than domestic ones, managers should track L1 distances separately by market segment to refine pricing and inventory decisions accordingly.

5.3. Conclusion and future directions

By emphasizing the *full* distribution of lead times, this study demonstrates how the COVID-19 pandemic produced lasting changes that are not fully captured by traditional statistics such as the mean or median. In revealing reallocation among short-, medium-, and long-lead segments, the normalized L1 distance emerges as a valuable diagnostic tool for detecting and quantifying booking-behavior shifts. These insights are particularly relevant to industry stakeholders looking to adapt forecasting and operational strategies to volatile travel markets.

Despite this contribution, our analysis is subject to several limitations. First, we focused on four major U.S. cities using data from a single accommodation platform. Results may not generalize to all markets or lodging segments, particularly those with smaller or more heterogeneous demand bases. Second, we did not isolate other potential drivers of lead-time divergence, such as flight capacity, local pandemic severity, or remote-work policies. Incorporating these factors might clarify whether certain market segments or traveler demographics are more prone to last-minute shifts.

Building on these insights, future research could broaden the dataset to include additional destinations or compare multiple lodging types. Methodologically, integrating the L1 distance into machine learning or advanced time-series models could improve forecast robustness, especially under fluctuating conditions. Another promising direction involves piloting a *forward-looking* or *partial-L1* approach in real-world revenue management systems, where even small early warnings about distributional shifts can yield significant benefits in pricing, marketing, and inventory controls.

CRedit authorship contribution statement

Harrison Katz: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Erica Savage:** Writing – review & editing, Writing – original draft. **Peter Coles:** Writing – review & editing, Methodology.

Declaration of competing interest

Harrison Katz reports a relationship with Airbnb that includes: employment.

Erica Savage reports a relationship with Airbnb that includes: employment.

Peter Coles reports a relationship with Airbnb that includes: employment.

Acknowledgements

The authors thank Sean Wilson, Jess Needleman, Liz Medina, and Jenny Cheng for helpful discussions and Ellie Mertz for championing our research. The authors would also like to thank the Editor, the Associate Editor, and three anonymous reviewers for their careful reading, comments, and suggestions to an earlier version of the paper.

Data availability

The primary dataset used in our study, *Lead Times in Flux: Analyzing Airbnb Dynamics During Global Upheavals (2018–2022)*, is not publicly available due to confidentiality constraints.

References

- Assaf, A. G., & Tsionas, M. (2019a). Non-parametric regression for hypothesis testing in hospitality and tourism research. *International Journal of Hospitality Management*, 76, 43–47.
- Assaf, A. G., & Tsionas, M. G. (2019b). Bayesian dynamic panel models for tourism research. *Tourism Management*, 75, 582–594.
- Assaf, A. G., & Tsionas, M. G. (2019c). Forecasting occupancy rate with Bayesian compression methods. *Annals of Tourism Research*, 75, 439–449.
- Assaf, A. G., & Tsionas, M. G. (2019d). Quantitative research in tourism and hospitality: An agenda for best-practice recommendations. *International Journal of Contemporary Hospitality Management*, 31(7), 2776–2787.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I., et al. (1990). StI: A seasonal-trend decomposition. *Journal of Official Statistics*, 6(1), 3–73.
- Fiori, A. M., & Foroni, I. (2019). Reservation forecasting models for hospitality smes with a view to enhance their economic sustainability. *Sustainability*, 11(5), 1274.
- Fiori, A. M., & Foroni, I. (2020). Prediction accuracy for reservation-based forecasting methods applied in revenue management. *International Journal of Hospitality Management*, 84, Article 102332.
- Fuchs, G., & Reichel, A. (2011). An exploratory inquiry into destination risk perceptions and risk reduction strategies of first-time vs repeat visitors to a highly volatile destination. *Tourism Management*, 32(2), 266–276.
- Gössling, S., Scott, D., & Hall, C. M. (2021). Pandemics, tourism and global change: A rapid assessment of covid-19. *Journal of Sustainable Tourism*, 29(1), 1–20.
- Gyódi, K. (2021). *Airbnb and hotels during COVID-19: Different strategies to survive*. 16 (1) pp. 168–192. *International Journal of Culture, Tourism and Hospitality Research*.
- Huang, L., & Zheng, W. (2023). Hotel demand forecasting: A comprehensive literature review. *Tourism Review*, 78(1), 218–244.
- Katz, H., Bruschi, K. T., & Weiss, R. E. (2024). A Bayesian Dirichlet auto-regressive moving average model for forecasting lead times. *International Journal of Forecasting*, 40(4), 1556–1567.
- Lim, C., & McAleer, M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management*, 23(4), 389–396.
- Lo, A. S., Cheung, C., & Law, R. (2011). Hong Kong residents' adoption of risk reduction strategies in leisure travel. *Journal of Travel & Tourism Marketing*, 28(3), 240–260.
- Mueller, R., & Sobreira, N. (2024). *Tourism forecasts after covid-19: Evidence of Portugal*. *Annals of Tourism Research Empirical Insights*, 5 (1) p. 100127.
- Okafor, L., Khalid, U., & Gopalan, S. (2022). Covid-19 economic policy response, resilience and tourism recovery. *Annals of Tourism Research Empirical Insights*, 3(2), Article 100073.
- de Oliveira, M. B., Zucchi, G., Lippi, M., Cordeiro, D. F., da Silva, N. R., Iori, M., et al. (2021). Lead time forecasting with machine learning techniques for a pharmaceutical supply chain. *In ICEIS*, 1, 634–641.
- Pereira, L. N. (2016). An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management*, 58, 13–23.
- Qiu, R. T., Park, J., Li, S., & Song, H. (2020). Social costs of tourism during the covid-19 pandemic. *Annals of Tourism Research*, 84, Article 102994.
- Sainaghi, R., & Chica-Olmo, J. (2022). The effects of location before and during covid-19: Impacts on revenue of airbnb listings in Milan (Italy). *Annals of Tourism Research*, 96, Article 103464.
- Sharma, A., Shin, H., Santa-Maria, M. J., & Nicolau, J. L. (2021). Hotels' covid-19 innovation and performance. *Annals of Tourism Research*, 88, Article 103180.
- Sigala, M. (2020). Tourism and covid-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, 117, 312–321.
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting: A review of recent research. *Tourism Management*, 29(2), 203–220.
- Tussyadiah, I. P., & Pesonen, J. (2016). Impacts of peer-to-peer accommodation use on travel patterns. *Journal of Travel Research*, 55(8), 1022–1040.
- Webb, T., Schwartz, Z., Xiang, Z., & Altin, M. (2022). Hotel revenue management forecasting accuracy: The hidden impact of booking windows. *Journal of Hospitality and Tourism Insights*, 5(5), 950–965.
- Zaki, K. (2022). Implementing dynamic revenue management in hotels during covid-19: Value stream and wavelet coherence perspectives. *International Journal of Contemporary Hospitality Management*, 34(5), 1768–1795.
- Zhang, L., Yang, H., Wang, K., Bian, L., & Zhang, X. (2021). The impact of COVID-19 on airline passenger travel behavior: An exploratory analysis on the Chinese aviation market. *Journal of Air Transport Management*, 95, Article 102084.