Show-Up Profiles for Scheduled Services: Estimation and Applications

Fernando Bernstein[1] N. Bora Keskin[1] Adam Mersereau[2] Morgan Wood[3] Serhan Ziya[3]

[1]Fuqua School of Business, Duke University
[2]Kenan-Flagler School of Business, University of North Carolina at Chapel Hill
[3]Department of Statistics & Operations Research, University of North Carolina at Chapel Hill

Problem definition: Motivated by passenger arrivals at the security checkpoint of the Raleigh-Durham International Airport, we develop methods to study arrivals to a system in which arrivals are tied to scheduled events, such as flights. A key concept for modeling arrivals in such systems is the "show-up profile," a parametric probability distribution describing how far in advance passengers arrive for their flights. These profiles can be combined based on a known flight schedule to yield an effective and interpretable aggregate passenger arrival forecast. Existing industry practice and academic work estimate show-up profiles using customer surveys or other data sources matching arriving passengers with flights, which are typically not available to U.S. airports. This motivates our study of an easy to implement and dynamic method for estimating show-up profiles. **Methodology/results:** We introduce an innovative solution for estimating show-up profiles using infrared-beam people-counting sensors and a structural estimation approach that does not require a mapping of passengers to flights. A direct maximum likelihood approach is intractable, but we propose a tractable approximation and prove that it yields consistent estimates of the underlying show-up profile parameters. Our approach produces forecasting results comparable to pure machine learning methods in our airport context, yields significantly improved adaptive forecasts when combined with machine learning methods, and reveals empirical insights about passenger behavior variations across different times of day and flight destinations. **Managerial implications:** Our work presents a novel application of Internet of Things (IoT) technology to service operations with incomplete data and demonstrates the value of integrating known operational structure with black-box forecasting approaches. The methods we develop and test can be readily applied at U.S. airports and other transportation hubs, and they can be adapted to other event-driven service environments such as theaters, healthcare facilities, and museums.

1. Introduction

A key challenge in running a complex operation like an airport is managing the passenger experience, which is itself a multifaceted task. Passengers are unhappy when they have difficulty finding a parking spot, when they encounter expensive or limited food options, and when they find neglected bathroom facilities. However, a factor that is most associated with passenger dissatisfaction is crowding at the airport. Crowding and delays not only color the passengers' experience while they are physically at the airport, but delays may also lead to missed flights which have more serious consequences for passengers. Therefore, airport administrators and designers are deeply interested in managing passenger flows and crowding through reactive measures such as deploying staff for crowd management, short-term controls such as staff scheduling, and long-term decisions such as designing terminals with sufficient capacity. A prerequisite for managing passenger flows is a model that can predict passenger arrivals (to check-in counters, the security checkpoint line, gates, etc.) and can facilitate sensitivity analysis of different arrival scenarios.

To address this challenge, the aviation industry commonly utilizes *show-up profiles*—also known as *arrival profiles* and *earliness arrival functions* (IATA 1995, Ashford et al. 2011, Isarsoft 2024). A show-up

profile specifies the probability distribution for how early (relative to the flight's departure time) passengers of a flight arrive at a particular location (e.g., the security checkpoint queue) in the airport. To illustrate the notion and use of a show-up profile, consider the left panel of Figure 1, which plots a show-up profile used by a planner at the Raleigh-Durham International Airport (RDU) in the U.S. state of North Carolina, which we partnered with for this research. The show-up profile for a particular flight allows airlines and airport administrators to anticipate flight-specific crowds at the check-in, security checkpoint, and boarding areas. More importantly, by stacking the show-up profiles for all flights in a day, as shown on the right panel in Figure 1, one can naturally derive aggregate forecasts of arrivals at commonly used areas of the airport.

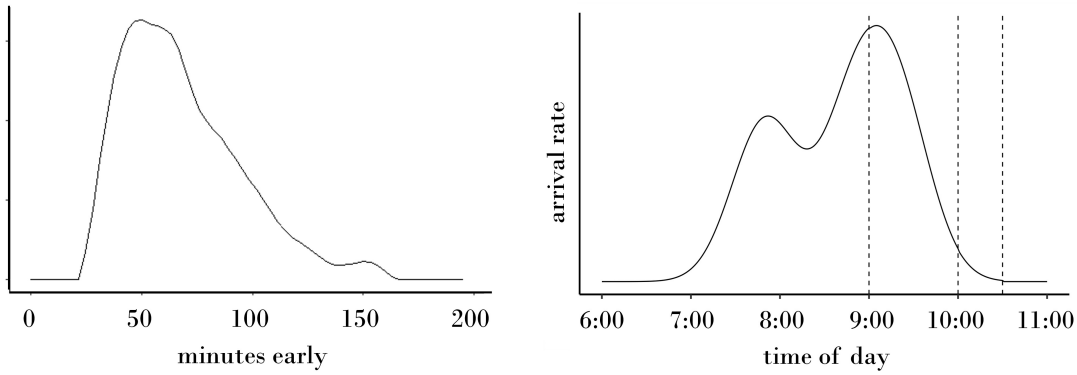

Figure 1 Example of a possible show-up profile—in this figure, an example used by a planner at the RDU airport—and a resulting arrival distribution derived from "stacking" the show-up profile according to scheduled flight departure times which are displayed using dotted vertical lines in the right panel.

Despite the crucial role show-up profiles play in some of the important operational decisions at airports, to the best of our knowledge, no prior work has provided a road map for how they can be estimated in practice from readily obtainable data and how the estimated show-up profiles can potentially be used for planning purposes. The objective of this paper is to help fill this gap.

Estimating show-up profiles requires data on passenger flows. Historically, airport planners have used standard show-up profiles—e.g., a prominent reference manual on airport management (IATA) provides a stock show-up profile that is commonly employed both in practice and academic studies (Ahyudanari and Vandebona 2005, Cheng 2014)—or they have estimated show-up profiles by surveying arriving customers in airports (Park and Ahn 2003, Rauch and Kljajić 2006, Bruno et al. 2019). Using standard show-up profiles has obvious shortcomings, as prior work has found that passengers' show-up patterns vary by departure airport and with various factors including destination, time of day, day of week, carrier, and passenger fare class (Chun and Mak 1999, Postorino et al. 2019). We posit that show-up profiles are also likely to shift seasonally and over time. While surveys can be helpful in estimating show-up profiles depending on passenger or flight characteristics, it is difficult to scale this method to obtain reliable estimates or to update the estimates with changing conditions.

An alternative approach to estimate show-up profiles is to use identifiable data on passenger arrivals, collected through baggage handling information systems, airline-specific departure control systems, and government security agencies (Postorino et al. 2019). However, baggage handling information systems only reflect passengers checking bags, which is just a (potentially biased) sample of the overall passenger population. Departure control systems can capture passenger check-ins, but the data are owned by airlines and do not capture operationally useful arrival timestamps from passengers checking in online. Finally, government security agencies such as the Transportation Security Agency (TSA) in the United States do not routinely share information on passenger flows with airports. This is the case for our airport partner, RDU. Therefore, it is valuable to identify an inexpensive and automatic way to collect data and estimate show-up profiles dynamically from available data. In this paper, we propose an approach to estimate the show-up profiles by collecting data on passenger flows from infrared-beam people counter sensors installed at the airport. These counters are easy to install and maintain, they are inexpensive, and they report data on passenger flows at the airport in real time.

The TSA checkpoint area at RDU experiences congestion frequently over the course of a day, serves as a bottleneck in passenger flows, and causes crowding in adjacent areas. This congestion is likely to worsen in the future. RDU, like many airports around the U.S. and globally, has seen record-breaking passenger arrivals in recent months and years (Bergin 2024, Josephs 2024). Moreover, the airport is moving ahead with expansion plans, which include a longer runway to accommodate larger aircraft, and it expects increases in passenger traffic in the coming years. In view of the challenges of managing congestion around the security checkpoint, we installed infrared-beam people counters at the entrances to the TSA checkpoint area in Terminal 1 of RDU. These sensors report in real time in 15 minute increments the number of passengers arriving at the checkpoint. An example of a day of sensor data can be seen in Figure 2 along with scheduled flight departure times overlaid as vertical lines.

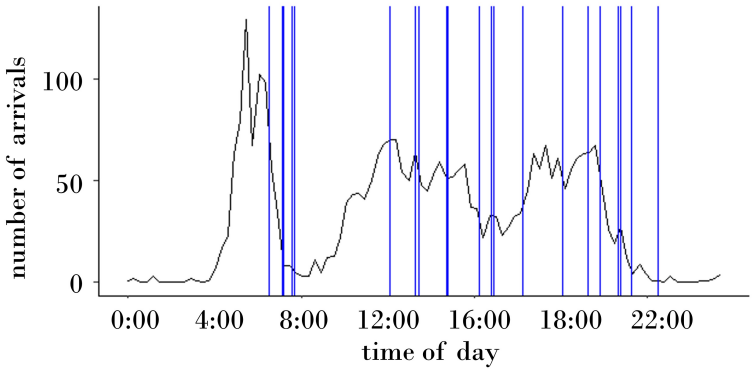

Figure 2 Passenger arrival and scheduled flight departure data from April 15, 2022, at RDU Terminal 1. Sensor counts of the number of arrivals in each 15 minute interval are depicted in black and the scheduled flight departure times are overlaid as vertical lines.

While collecting data using people counters is straightforward and inexpensive, using that data to estimate show-up profiles for individual flights poses a significant challenge as it is not possible to match passenger

arrivals to specific flights. (For example, a passenger arriving to the TSA checkpoint area at 9:15 a.m. may be 45 minutes early for a 10:00 a.m. flight or 75 minutes early for a 10:30 a.m. flight.) This makes estimation a non-trivial task. If one knew to which flight each passenger tracked by the sensors was assigned, then estimating a show-up profile would be a straightforward density estimation problem; see, e.g., Postorino et al. (2019). The main challenge we tackle in this work is the estimation of show-up profiles with no information on the matching of passengers to flights.

In view of these limitations, we propose a method for structurally estimating show-up profiles from data collected using the people counters. We consider a parametric model that assumes the show-up profile belongs to a known family of distributions. We assume the number of passengers assigned to each flight is known, but the matching of passengers to flights is not. The problem can be reframed as estimating the parameters of the show-up profile's distribution, which we accomplish through maximum likelihood estimation. The true likelihood function is intractable to work with, but we solve a much simpler approximate likelihood function. In Section 3, we provide theoretical justification for this method by proving that it results in consistent parameter estimates under mild assumptions.

Figure 3 presents one such show-up profile estimate (for a single flight) obtained by assuming a truncated normal distribution and using maximum likelihood. This results in a mean earliness of 66 minutes with a standard deviation of 37 minutes. Compared with the industry show-up profile presented in Figure 1, forecasting arrivals using our method (which we refer to as *structural forecasting*) results in improved forecasting performance. Specifically, considering the root mean forecasting squared error, we find that the expected error across 15-minute intervals decreases by 13% using our method (from a root-mean squared error of 15.8 individuals to 13.7 individuals). We provide further details and analysis in Section 4.

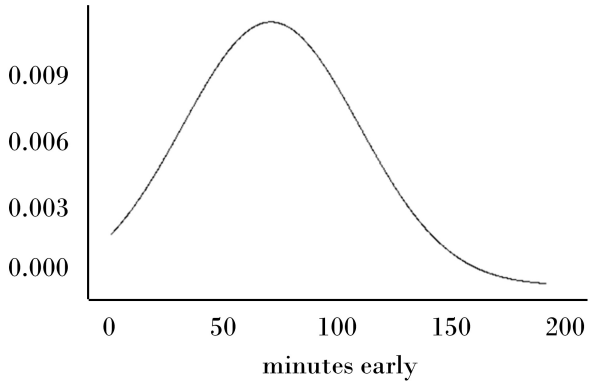

Figure 3 Example of a show-up profile trained using maximum likelihood estimation.

We further compare the performance of the structural forecasting method with black-box machine learning methods that use the time of day, day of week, multiple lagged passenger arrival counts, and multiple measures of the number of upcoming flights (for example, number of flights in the next 30 minutes, next 1 hour, etc.) and previous flights to predict future passenger arrivals. We present three main sets of performance results in Section 4. First, we show that the basic structural method performs similarly to machine

learning methods, although in some cases slightly worse, and discuss reasons for this difference. Second, we demonstrate how the structural method can be improved by training multiple show-up profiles that differ based on flight characteristics such as departure time and destination. Finally, we show that improved forecasting performance can be obtained by combining machine learning methods, which can take into account many variables including recent lagged arrivals, and structural methods, which take advantage of the special problem structure albeit in a static way. We accomplish this combination by using the estimate generated by the structural method as an additional feature variable for the machine learning methods. We find that we can sometimes improve machine learning methods by carefully reducing the set of feature variables to guide the methods to rely on relevant structure.

Our show-up profile estimation approach can easily be extended to allow the show-up profiles to depend on flight features and can therefore support, in addition to forecasting, empirical investigations of passenger behavior based on aggregate arrival counts. To illustrate this point, we focus on two questions: whether passengers' show-up behavior changes depending on the time of the flight and whether the flight destination characteristics play any role. Our analysis in Section 5 reveals that the answer to both questions is affirmative. In particular, we find that passengers arrive closer to departure times when their flights are early in the morning and when the flight has a destination that serves as a hub, but we find the opposite for flights to leisure destinations.

Airports are service environments in which customers experience services at scheduled times (i.e., flight departure times) and thus are similar to other service environments such as entertainment venues like movie multiplexes, where customer arrivals are driven by movie start times, or large healthcare facilities, which house multiple clinics and use block scheduling. Therefore, even though this paper is directly motivated by airports, the proposed methodology and results are relevant to a broader class of service environments.

The structure of the paper is as follows. We begin with a review of the relevant literature in Section 2. This section is followed by the formal problem formulation, development of our estimation method, and proof of the method's consistency, all provided in Section 3. Section 4 compares the performance of our method with that of a benchmark developed using the industry show-up profile estimate, as well as those under various machine learning methods. In Section 5, we present the results of our empirical analysis and provide our concluding remarks in Section 6. The appendices contain mathematical derivations, a discussion on consistency, and further robustness checks of numerical results.

2. Literature Review

Show-up profiles (Isarsoft 2024), also known as passenger arrival patterns (Gao 2021), lead time distributions (Kennon et al. 2013), and earliness distributions (TransSolutions 2011), are often used in airport settings to predict throughput and congestion for resource planning as well as to better understand passenger behaviour. However, unlike in our work where we do not know the earliness of arriving passengers,

the show-up profiles that are estimated in the literature assume access to direct measurements of passenger earliness, possibly by utilizing passenger surveys or recording boarding pass information at checkpoints. Within the United States, boarding pass information is not recorded or stored (U.S. Department of Homeland Security 2007) or shared with airports.

Alodhaibi et al. (2019) investigate the impact that different arrival patterns have on congestion. Based on a model of the Brisbane Airport, the authors simulate the congestion at check-in counters, security, and immigration under various passenger show-up profiles. They find that the show-up profile's mean and maximum drastically affect the average and maximum amount of time spent waiting in each queue, especially when looking at times of high congestion. This implies that using the correct show-up profile is pivotal to having an accurate estimation of future arrivals.

Postorino et al. (2019) train show-up profiles based on earliness measurements collected using Bar Coded Boarding Pass (BCBP) technologies to better understand passenger behavior. Similarly to our work, they train a parametric model of earliness and, as we also do in Section 5, they allow this model to differ based on various flight characteristics. The authors find that both the time of day and type of carrier (low-cost carrier or full carrier) affect the show-up profile. Unlike our study, Postorino et al. (2019) have access to the exact earliness values of each individual passenger through the BCBP technology, which greatly simplifies the estimation step. This technology is not used in the United States.

A common use of show-up profiles is in the estimation of passenger arrivals to determine optimal allocation of check-in counters; see, e.g., Chun and Mak (1999), Al-Sultan (2018), and the survey by Lalita and Murthy (2022). Often, these show-up profiles are determined using passenger surveys (see Ashford et al. 1976, Park and Ahn 2003, Rauch and Kljajić 2006, Bruno et al. 2019). As these studies assume direct access to observations of passenger earliness, the estimation of a show-up profile amounts to a straightforward application of density estimation. Unlike the approaches in this literature, our work does not have direct earliness measurements and thus cannot match arrivals to specific flights, which introduces unique challenges in estimation. However, these studies motivate that show-up profiles vary with a number of factors, including departure time, type of haulage (first- and business-class versus economy-class passengers), and departure distance (domestic versus international flights), some of which we consider in Section 5.

Other studies, such as the work by Ahyudanari and Vandebona (2005), use generic show-up profiles such as the distribution found in the Airport Development References Manual by the International Air Transport Association (IATA 1995) to estimate future arrivals. However, the IATA manual recommends that practitioners do their own survey to generate show-up profiles tailored to the airport being modeled. Our work deviates from this approach as we do not rely on surveys or previously existing profiles but instead estimate profiles directly from aggregate arrival data.

Forecasting future arrivals in airport settings goes beyond using show-up profiles. A stream of literature also considers using machine learning models and other models to forecast arrivals. However, as in the

show-up profile models in the literature referenced above, these papers assume access to the historical earliness of each passenger. Sayın et al. (2023), for example, forecast arrivals using linear models and various types of recurrent neural networks that make predictions based on upcoming and past number of flight departures, distinguishing by type of aircraft (top large body, medium large body, etc). Their study relies on access to the flight info of each passenger through boarding pass scans at the check-in counters, whereas we use aggregate arrival data without flight-specific identifiers.

Machine learning models are also used to estimate distributions of other elements of passenger flow such as connection times. Both Guo et al. (2020) and Guo et al. (2022) focus on modeling connecting passenger connection times with real-time data to improve multiple systems in the airport. These papers use regression trees to estimate the distribution of connection times, which is possible as they have access to information on every passenger and the flights (inbound and outbound) that they are associated with. Connection times, the region, time of day, day of week, and planned connection time are all considered as factors to forecast connection times.

Instead of using machine learning models or show-up profiles to forecast arrivals for check-in allocation, Parlar and Sharafali (2008) train a "show up process." They view passengers showing up for a flight as a death process. An arrival of a passenger from a fixed population, which encompasses all of the passengers that will be arriving, is viewed as an occurrence, i.e., a "death." For each passenger, the time until their arrival is assumed to follow an exponential distribution. The authors use a maximum likelihood technique to determine the rate of this death process. Our methodology similarly uses maximum likelihood estimation, but unlike Parlar and Sharafali (2008), we do not assume knowledge of specific passenger-to-flight mappings, instead we leverage aggregated arrival data only.

Forecasting arrivals is important in many settings other than airports, such as call centers (see Ibrahim and L'Ecuyer 2013) and hospitals and clinics (see Kim et al. 2015, among others). In particular, distributions of patient arrivals, similar to show-up profiles, are commonly used in healthcare settings to estimate how individuals will arrive for appointments or events. A common behavior to model is the incidence of no-shows or cancellations (see Liu et al. 2010, Kim et al. 2018). Dantas et al. (2018) provide a comprehensive literature review on the topic of no-shows. Unlike in a healthcare environment, the uncertainty in passenger arrivals at airports is around the time of their arrival, rather than on the possibility of having no-shows. There are other papers that consider non-punctual customers, such as the works by Jouini et al. (2022) and Koeleman and Koole (2012). While this literature provides valuable insights into predicting individual arrival behavior, it typically assumes knowledge of the specific event individuals are attending, which differs from our context in airport arrival forecasting, where we are not aware of which flight a passenger is arriving for.

Our work is part of a larger movement to apply IoT (Internet of Things) sensors to measure flows of people in various service operations contexts. Bernstein et al. (2024) use people-counting sensors to monitor

passenger congestion in an airport setting like ours. Martínez-de Albéniz and Valdivia (2019) use sensors to measure the effects of displayed exhibits on museum traffic. In the context of retail, Caro and Sadr (2019) survey opportunities for the use of IoT in brick-and-mortar retail stores. Perdikaki et al. (2012), Lu et al. (2013), Lee et al. (2021), Martínez-de Albéniz and Belkaid (2021), and Musalem et al. (2021) all use people-counting sensors to gather data from retail settings for empirical analysis. IoT has also been used to gather data to improve healthcare operations; see, for example, Laudanski et al. (2021).

3. Estimating a Show-up Profile

In this section, we develop a method to estimate show-up profiles of passenger arrivals. In Section 3.1, we introduce a model of passenger arrivals and provide a formal definition of a show-up profile. In Section 3.2, we construct a maximum likelihood estimate of the show-up profile, and in Section 3.3, we show that this estimate is consistent under mild assumptions.

3.1. A Show-up Profile of Passenger Arrivals

Suppose that we have passenger arrival data collected over M days, and on each day $m \in \{1, \ldots, M\}$, we make observations over a bounded time range, $[0, T]$. Specifically, there are K scheduled flights on each day, where flight $k \in \{1, \ldots, K\}$ is scheduled to depart at time $\tau_k \in [0, T]$. For expositional purposes, we make the assumption that flight schedules are identical across days. We relax this assumption in Sections 4 and 5, and we formally describe the relaxation in the appendices.

We assume that N_k, the number of passengers scheduled for flight k, is known and has the same value for all days. (In practice, airport planners can accurately estimate N_k by considering seat capacities on the corresponding flight and the expected load factor.) The total number of passengers to arrive on any given day is then equal to $\sum_{k=1}^{K} N_k$.

The arrival time of each passenger depends only on the passenger's scheduled flight. Conditional on the passenger's scheduled flight time, the passenger's arrival time is independent of other flight times and the arrival times of other passengers. Furthermore, all passengers' earliness to their respective flights are stochastically identical. (In subsequent sections, we allow the show-up profiles to vary based on scheduled flight time and destination.) Specifically, each passenger arrives for their flight according to an identical show-up profile D_θ, where $D_\theta(t)$ represents the probability that a given passenger arrives more than t time units before their scheduled flight departure. The show-up profile D_θ follows a known parametric family characterized by a parameter θ which is contained in some subspace Θ of $\mathbb{R}^n$, where $n \in \{1, 2, \ldots\}$. For a fixed θ, $D_\theta(t)$ is a survival function, or equivalently, $1 - D_\theta(t)$ is a cumulative distribution function (c.d.f.). Passengers do not arrive at the TSA line after their respective scheduled flight time and arrive at most a time units before, where $a \in (0, \infty)$; for example, one can consider that passengers arrive no earlier than, say, $a = 360$ minutes prior to the scheduled departure of their flight. Therefore, $D_\theta(t) = 1$ for $t < 0$, and $D_\theta(t) = 0$ for $t > a$.

Our goal is to estimate θ and thus obtain an estimate of the show-up profile D_θ, using the arrival data aggregated over all flights. That is, we estimate θ even though we do not know which specific flight each arrival is associated with.

Each day is divided into I non-overlapping time intervals each of length r, for some fixed $r \in (0, \infty)$. These time intervals, which we call *sensor intervals*, correspond to data collection intervals of the people counters we installed at the airport. Sensor interval $i \in \{1, \ldots, I\}$ is defined as $(t_{i-1}, t_i]$, where $t_0 = 0$, $t_i = t_{i-1} + r$ for $i \in \{1, \ldots, I-1\}$, and $t_I = T \geq \tau_K$, i.e., the last flight departure occurs prior to the end of the last sensor interval. Let X_i^m denote the total number of arrivals (for any flight) in sensor interval i, i.e., in the time interval $(t_{i-1}, t_i]$, on day m. Our objective is to estimate θ given x_i^m, the realization of random variables X_i^m for $i \in \{1, \ldots, I\}$ and $m \in \{1, \ldots, M\}$.

For expositional simplicity, we assume that flights depart on interval boundaries such that for each $k \in \{1, \ldots, K\}$, $\tau_k = t_i$ for some $i \in \{1, \ldots, I\}$. This assumption is also relaxed in our numerical experiments, and we again leave the modeling details to the appendices. We also assume that the first sensor interval occurs early enough that all arrivals are observed; that is, the first flight of the day occurs at time a or later.

For any $i \in \{1, \ldots, I\}$, let

$$B_\theta(i) = D_\theta((i-1)r) - D_\theta(ir)$$

be the probability that a random passenger arrives in the ith sensor interval before their flight departure time; equivalently, $B_\theta(i)$ is the probability that the passenger arrives between $(i-1)r$ and ir minutes early relative to their flight departure time. We call the probability mass function $\{B_\theta(i),\, i = 1, \ldots, I\}$ the *binned distribution* or *binned show-up profile*. An example of a binned distribution for a passenger on flight 1 with a departure time of 7:30 is given in Figure 4. In the figure, the sensor interval size, r, is 15 minutes.

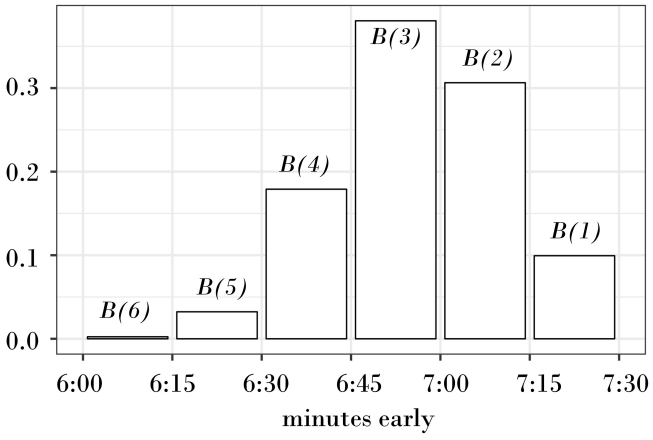

Figure 4 Example of a binned distribution with interval size of $r = 15$ minutes and flight departure time τ_1 of 7:30, where the x-axis represents the time of day. The binned distribution's dependence on θ has been suppressed for clarity of presentation.

3.2. Estimation of the Show-up Profile

Let $\text{int}(k)$ denote the sensor interval in which the kth flight departs; i.e., $\text{int}(k) = i$ if $\tau_k = t_i$. On any day m, the probability that a passenger on the kth flight arrives in interval $i \leq \text{int}(k)$, i.e., the $(\text{int}(k) - i + 1)$st interval before the departure time of the passenger's flight, is equal to $B_\theta(\text{int}(k) - i + 1)$. Also, the probability that a random passenger arriving on day m is on flight k is $N_k / \sum_{s=1}^K N_s$. Thus, by the law of total probability, the probability that a passenger on a fixed day m arrives in interval i, which we refer to as $P_\theta(i)$, is given by

$$P_\theta(i) = \frac{\sum_{k=1}^K B_\theta(\text{int}(k) - i + 1) N_k}{\sum_{k=1}^K N_k}.$$

For any fixed day m, we observe a collection of arrivals $X^m = \{X_i^m,\, i = 1, \ldots, I\}$. Note that the arrival observations $\{X^m,\, m = 1, \ldots, M\}$ are independent across days, as arrivals on each day are naturally independent. However, within a day, individual passenger arrival times are not independent. To see why there is a dependence across arrivals within a day, suppose we have observed N_K passengers arriving between times τ_{K-1} and τ_K. That is, we have seen as many passengers during the period of time between flight $K-1$ (the next-to-last flight) and flight K (the last flight) as there are on flight K. Because we assume that all passengers arrive before their scheduled flights, it follows that these passenger arrivals are the full set of arrivals for flight K. All other prior passenger arrivals must therefore belong to flights $1, \ldots, K-1$, which results in a modified arrival time distribution. More generally, there is a complex dependence between passenger arrivals induced by the flight assignments. Explicitly expressing the true distribution of arrival vectors, much less calculating it, is quite complicated. Hence, our goal of finding the parameter θ that drives these observations becomes intractable. For this reason, we now introduce a simpler *surrogate problem* based on a mixture distribution.

In the surrogate problem, we do not enforce that each flight has a fixed number of passengers. Rather, we allow passengers to arrive for each flight k with probability $N_k / \sum_{s=1}^K N_s$, independently of each other. An alternative interpretation is that each passenger arrives according to a mixture distribution of the show-up profiles for all flights weighted by $N_k / \sum_{s=1}^K N_s$. Hence, as before, the probability a passenger arrives in interval i is equal to $P_\theta(i)$. In this surrogate problem, we let $\tilde{X}^m = \{\tilde{X}_i^m,\, i = 1, \ldots, I\}$ denote the counts of arrivals across all intervals on day m.

As each arrival can arrive independently in any of I potential sensor intervals, each with corresponding probability $P_\theta(i)$, $\tilde{X}^m = \{\tilde{X}_i^m,\, i = 1, \ldots, I\}$ follows a multinomial distribution with parameters $\sum_{k=1}^K N_k$, I, and $\{P_\theta(i),\, i = 1, \ldots, I\}$, corresponding to the number of trials, number of mutually exclusive events, and event probabilities, respectively. Suppose that we had M observations $\{\tilde{x}^m\}_{m=1}^M$ from the surrogate problem. Then, we could express the likelihood of the show-up profile parameter ϑ as

$$L\left(\vartheta; \{\tilde{x}^m\}_{m=1}^M\right) = \prod_{m=1}^M \left(\frac{(\sum_{k=1}^K N_k)!}{\prod_{i=1}^I \tilde{x}_i^m!} \prod_{i=1}^I P_\vartheta(i)^{\tilde{x}_i^m} \right).$$

We can thus infer θ in the surrogate problem via maximum likelihood estimation as follows:

$$\tilde{\theta}_M = \arg\max_{\vartheta \in \Theta} \left\{ \sum_{m=1}^{M} \sum_{i=1}^{I} \tilde{x}_i^m \log P_\vartheta(i) \right\}. \tag{1}$$

Equation (1) is a conceptual construct because the real system does not produce observations $\{\tilde{x}^m\}_{m=1}^M$. Substituting the true observations $\{x^m\}_{m=1}^M$, we arrive at the following estimate:

$$\hat{\theta}_M = \arg\max_{\vartheta \in \Theta} \left\{ \sum_{m=1}^{M} \sum_{i=1}^{I} x_i^m \log P_\vartheta(i) \right\}. \tag{2}$$

Given the estimate $\hat{\theta}_M$ of θ, we also have an estimate of the binned distribution B_θ, given by $B_{\hat{\theta}_M}$, and an estimate of the show-up profile D_θ, given by $D_{\hat{\theta}_M}$. This estimate is practical to compute, and in the following section, we prove that it is a consistent estimator of θ.

3.3. Consistency of the Estimated Show-up Profile

The main result of this section, Theorem 1, shows that the parameter estimate $\hat{\theta}_M$ in (2) is weakly consistent under the following set of assumptions.

ASSUMPTION 1.

1.1 *The true parameter θ lies within a compact subspace Θ of $\mathbb{R}^n$, where for each $\vartheta \in \Theta$, $B_\vartheta(i) = 0$ if and only if $B_\theta(i) = 0$.*

1.2 *The binned distribution $B_\vartheta(i)$ is continuous in ϑ for all i and $\vartheta \in \Theta$.*

1.3 *The parameter θ can be uniquely identified from the binned distribution B_θ.*

Assumptions 1.1 and 1.2 are mild regularity conditions and are used to guarantee that the log-likelihood is bounded and continuous. However, in general, the identifiability of the parameter θ from the binned distribution (Assumption 1.3) cannot be guaranteed under all scenarios. Certain parameter and interval configurations may lead to multiple estimates based on the observed data. Even though one can construct examples of non-identifiability (see Appendix EC.2), the underlying scenarios in these examples appear to be unrealistic, and we have not faced the problem of non-identifiability in any of our numerical experiments. One simple scenario in which the parameters are identifiable is when the show-up profile follows a discrete distribution with bin width equal to the length of the sensor interval, i.e. $r = s$. In this scenario, the show-up profile is equivalent to the binned distribution. Alternatively, if the show-up profile follows a normal distribution, and the interval size is small enough such that $I \geq 3$, then the mean and standard deviation are identifiable. Likewise, if the show-up profile follows a triangular distribution with unknown mode, then the mode is identifiable as well. In Appendix EC.2, we provide further discussion of the identifiability of the parameter θ from the binned distribution.

We first present the following proposition, which indicates that the maximum likelihood estimate of the surrogate problem converges almost surely to the true parameter θ.

PROPOSITION 1. *Under Assumption 1, the maximum likelihood estimate $\tilde{\theta}_M$ for the surrogate problem in (1) is strongly consistent.*

PROOF. We complete this proof in two steps. First, we derive a strongly consistent estimate $\hat{P}_M(i)$ of $P_\theta(i)$, namely, the probability of a random passenger arriving in a given sensor interval i. Second, we show that B_θ can be expressed as a linear function of P_θ; say, $B_\theta(i) = g_i(P_\theta)$ for $i \in \{1,\dots,I\}$. Thus, given the consistent estimate $\hat{P}_M$ of P_θ, the estimate $\hat{B}_M(i) = g_i(\hat{P}_M)$ is a consistent estimate of $B_\theta(i)$. Finally, by Assumption 1.3, θ is uniquely identified from B_θ, which implies that the estimate $\tilde{\theta}_M$ such that $B_{\tilde{\theta}_M}(i) = \hat{B}_M(i) = g_i(\hat{P}_M)$ for all i is a consistent estimate of θ.

Step 1: Derive a consistent estimate of P_θ. Let $Y_i^M = \sum_{m=1}^M X_i^m$ be the total number of passengers across all M days to arrive in sensor interval i. Recall that the observations on day m, $\tilde{X}^m$, follow a multinomial distribution with parameters $\sum_{k=1}^K N_k$, I, and $\{P_\theta(i),\, i=1,\dots,I\}$, corresponding to the number of trials, number of mutually exclusive events, and event probabilities, respectively. Analogously, across multiple days, the sum of observations in the same period $Y^M = \{Y_i^M,\, i=1,\dots,I\}$ follows a multinomial distribution with parameters $M\sum_{k=1}^K N_k$, I, and $\{P_\theta(i),\, i=1,\dots,I\}$, corresponding to the number of trials, number of mutually exclusive events, and event probabilities, respectively.

As the maximum likelihood estimate of a multinomial distribution is strongly consistent, the estimate

$$\hat{P}_M = \arg\max_P \left\{ \sum_{i=1}^I \sum_{m=1}^M \tilde{x}_i^m \log P(i) : \sum_{i=1}^I P(i) = 1,\ P(i) > 0 \text{ for } i = 1,\dots,I \right\}$$

is a strongly consistent estimate of P_θ. Based on this, we deduce that if the system of equations

$$P_\theta(i) = \frac{\sum_{k=1}^K B_\theta(\text{int}(k) - i + 1) N_j}{\sum_{k=1}^K N_k},\ i = 1,\dots,I, \tag{3}$$

has a unique solution $\{B_\theta(j),\, j=1,\dots,I\}$, then the solution $\{\hat{B}(j),\, j=1,\dots,I\}$ to

$$\hat{P}_M(i) = \frac{\sum_{k=1}^K \hat{B}(\text{int}(k) - i + 1) N_j}{\sum_{k=1}^K N_k},\ i = 1,\dots,I,$$

is a strongly consistent estimate of $\{B_\theta(j),\, j=1,\dots,I\}$.

Step 2: Express $B_\theta(i)$ as a linear function of $P_\theta(i)$ for each $i = 1,\dots,I$. For notational simplicity, we set the end time of the final sensor interval equal to the time of the final flight $T = t_I = \tau_K$. We also assume that every period has exactly one departing flight. This can be done by allowing the total number of passengers on a flight to equal zero when there is no actual flight departing in that period, and expressing the total number of passengers as a sum if there are multiple flights within that period. Thus, N_i corresponds to the number of passengers with a flight departing in sensor interval i. This simplifies the notation such that $I = K$ and causes the flight departure at the end of the day to be non-empty, i.e., $N_I > 0$.

To show that equation (3) has a unique solution, we solve backwards recursively, starting in the final sensor interval T, i.e., the interval of the final flight. We have the following probability of an arrival in the final interval I:

$$P_\theta(I) = \frac{B_\theta(1)N_I}{\sum_{k=1}^{I} N_k}.$$

Thus, we have the following solution for $B_\theta(1)$:

$$B_\theta(1) = \frac{P_\theta(I)\sum_{k=1}^{I} N_k}{N_I}.$$

Considering the prior interval $I-1$, we have that

$$P_\theta(I-1) = \frac{B_\theta(1)N_{I-1} + B_\theta(2)N_I}{\sum_{k=1}^{I} N_k}.$$

Solving for $B_\theta(2)$, we deduce that

$$B_\theta(2) = \frac{P_\theta(I-1)\sum_{k=1}^{I} N_k - B_\theta(1)N_{I-1}}{N_I}.$$

More generally, for any $i \in \{0, \ldots, I-1\}$,

$$P_\theta(I-i) = \frac{\sum_{j=0}^{i} B_\theta(i-j+1)N_{I-j}}{\sum_{k=1}^{I} N_k}.$$

Hence, setting $N_i = 0$ for any $i \leq 0$, we obtain the following recursive solution:

$$B_\theta(i+1) = \frac{P_\theta(I-i)\sum_{k=1}^{I} N_k - \sum_{j=1}^{i} B_\theta(i-j+1)N_{I-j}}{N_I}. \qquad \square$$

In the proof of Proposition 1, the identifiability of θ from the binned distribution B_θ (Assumption 1.3) ensures that the parameter θ can be uniquely identified from P_θ as well. Proposition 1 follows from this result and the strong consistency of the maximum likelihood estimate for a multinomial distribution.

We now turn our attention to the true observations and the estimate of primary interest, $\hat{\theta}_M$. We prove the main result, Theorem 1, by first considering the consistency of the maximum value, rather than the consistency of the maximizer. This consistency is obtained by using laws of large numbers. Then identifiability, as established in the proof of Proposition 1, as well as an argmax consistency theorem is used to show the desired consistency of the maximizer.

THEOREM 1. *Under Assumption 1, the estimate $\hat{\theta}_M \to \theta$ in probability as $M \to \infty$.*

PROOF. For any possible parameter $\vartheta \in \Theta$ and vector $x \in \mathbb{R}^I$, let

$$f(\vartheta, x) = \sum_{i=1}^{I} x_i \log P_\vartheta(i),$$

and

$$F(\vartheta) = \sum_{k=1}^{K} N_k \sum_{i=1}^{I} P_\theta(i) \log P_\vartheta(i).$$

As the arrival of passengers on each day is independent, for each fixed i, X_i^m are independent across m. Moreover, because $\mathbb{E}[X_i^m] = N_k P_\theta(i)$ for each sensor interval i and day m, we deduce from the strong law of large numbers that

$$\lim_{M\to\infty}\left\{\frac{1}{M}\sum_{m=1}^{M} f(\vartheta, X^m)\right\} = \mathbb{E}\left[f(\vartheta, X^1)\right] = F(\vartheta) \text{ almost surely.}$$

To prove that $\hat{\theta}_M \to \theta$ in probability, we use the argmax continuous mapping theorem. Specifically, we invoke Theorem 3.2.2 of van der Vaart and Wellner (2023) by choosing the index n as M, the function $\mathbb{M}$ as F, and the indexed functions $\mathbb{M}_n$ as $\frac{1}{M}\sum_{m=1}^{M} f(\cdot, X^m)$, each maximized over the space $H = \Theta$ with maximizers $\hat{h} = \theta$ and $\hat{h}_n = \hat{\theta}_M$, respectively. For clarity, we restate the result as follows.

Theorem 3.2.2 of van der Vaart and Wellner (2023) (argmax continuous mapping). *Let $\mathbb{M}_n$, $\mathbb{M}$ be stochastic processes indexed by a metric space H such that $\mathbb{M}_n \to \mathbb{M}$ in probability in $l^\infty(K)$ for every compact $K \subset H$. Suppose that almost all sample paths $h \mapsto \mathbb{M}(h)$ are upper semicontinuous and possess a unique maximum at a (random) point $\hat{h}$, which as a random map in H is tight. If the sequence $\hat{h}_n$ is uniformly tight and satisfies $\mathbb{M}_n(\hat{h}_n) \geq \sup_h \mathbb{M}_n(h) - o_P(1)$, then $\hat{h}_n \to \hat{h}$ in probability in H.*

We begin by verifying the first set of conditions in Theorem 3.2.2 of van der Vaart and Wellner (2023). To this end, we aim to show that $\max_{\vartheta\in\Theta}\left|\frac{1}{M}\sum_{m=1}^{M} f(\vartheta, X^m) - F(\vartheta)\right|$ converges to zero in probability by the uniform law of large numbers (ULLN). In particular, we claim that under Assumption 1, the ULLN presented in Lemma 2.4 of Newey and McFadden (1994) holds in our setting. (Shi 2022 notes that Lemma 2.4 of Newey and McFadden 1994 is missing certain required measurability conditions. However, because of Assumption 1.2, our problem satisfies these additional criteria.) To establish this claim, we verify that (a) the space Θ is compact, (b) at all $\vartheta \in \Theta$, the mapping $\vartheta \mapsto f(\vartheta, X^m)$ is continuous almost surely, and (c) for some measurable function $G(X^m)$ with finite expectation, $|f(\vartheta, X^m)| \leq G(X^m)$. Because condition (a) holds under Assumption 1.1, we proceed by verifying conditions (b) and (c). For all i and $\vartheta \in \Theta$, we deduce from Assumptions 1.1 and 1.2 that the binned distribution value $B_\vartheta(i)$ is bounded and continuous in ϑ. Moreover, for all j, $P_\vartheta(j)$ is a weighted average of the binned distribution values; thus, for all j, $P_\vartheta(j)$ is also bounded and continuous in ϑ. By Assumption 1.1, $B_\vartheta(i) = 0$ if and only if $B_\theta(i) = 0$. Therefore, for all j, $P_\vartheta(j)$ is bounded and continuous in ϑ, with $P_\vartheta(j) = 0$ if and only if $P_\theta(j) = 0$. As $X_j^m = 0$ for all m if $P_\theta(j) = 0$, it also holds that the mapping $\vartheta \mapsto f(\vartheta, X^m)$ is bounded and continuous on Θ almost surely. Similarly, the mapping $\vartheta \mapsto F(\vartheta)$ is bounded and continuous on Θ. This verifies the first set of conditions in Theorem 3.2.2 of van der Vaart and Wellner (2023) as $\max_{\vartheta\in\Theta}\left|\frac{1}{M}\sum_{m=1}^{M} f(\vartheta, X^m) - F(\vartheta)\right|$ = $\sup_{\vartheta\in\Theta}\left|\frac{1}{M}\sum_{m=1}^{M} f(\vartheta, X^m) - F(\vartheta)\right|$ because both $\frac{1}{M}\sum_{m=1}^{M} f(\vartheta, X^m)$ and $F(\vartheta)$ are bounded, and Θ is a bounded metric space.

Next, we verify the second set of conditions of Theorem 3.2.2 of van der Vaart and Wellner (2023). Recall that F is continuous (and thus upper semi-continuous). To prove that F has a unique maximum

at θ, note that a parameter $\vartheta \in \Theta$ is identifiable from P_ϑ as shown in the proof of Proposition 1. Since $F(\vartheta) = \sum_{k=1}^{K} N_k \sum_{i=1}^{I} P_\theta(i) \log P_\vartheta(i)$ and the function $g(p) = \sum_i p_i \log(q_i)$, subject to each p_i and q_i being positive and the sums $\sum_i p_i$ and $\sum_i q_i$ equaling one, has the unique maximum $p = q$, the function F has a unique maximum θ by the identifiability established in the proof of Proposition 1.

We next consider the tightness of θ and of the random variables $\{\hat{h}_M,\, M = 1, 2, \ldots\}$. By definition, a random variable Y is considered tight if for all $\epsilon > 0$, there exists a finite N such that $\mathbb{P}(|Y| \geq N) \leq \epsilon$. Since θ is a deterministic value, this holds. Similarly, a sequence of random variables $\{Y_\alpha,\, \alpha \in A\}$ is considered tight if for all $\epsilon > 0$, there exists a finite N such that $\sup_{\alpha \in A} \mathbb{P}(|Y_\alpha| \geq N) \leq \epsilon$. As the space Θ is bounded, the estimates $\hat{h}_M$ are also bounded for all M. Taking any strict upper bound of the absolute value of Θ, e.g., $\max_{\vartheta \in \Theta} |\vartheta| + 1$, it follows that $\sup_{M \geq 1} \mathbb{P}(|\hat{h}_M| \geq \max_{\vartheta \in \Theta} |\vartheta| + 1) = 0$.

Finally, by the definition of $\hat{h}_M$, it holds that $\frac{1}{M} \sum_{m=1}^{M} f(\hat{h}_M, X^m) = \sup_{\vartheta \in \Theta} \left\{ \frac{1}{M} \sum_{m=1}^{M} f(\vartheta, X^m) \right\}$. We have therefore verified all the conditions in Theorem 3.2.2 of van der Vaart and Wellner (2023), concluding the proof of the result. □

Having established a consistent estimator of show-up profiles, we next turn our attention to forecasting using show-up profiles in Section 4 and to an investigation of how various factors correlate with passenger earliness in Section 5.

4. Forecasting

In this section, we use the show-up profile introduced in Section 3 to forecast future passenger arrivals to the TSA checkpoint area given the schedule of flight departures. Because this forecasting method leverages the structure inherent to the arrival process, we refer to it as the *structural forecasting method*, or simply, the *structural method*. To estimate show-up profiles, we employ the maximum likelihood estimation procedure described in Section 3.2, generalized to allow for differing flight schedules across days. We provide a detailed derivation of the generalized approach in Appendix EC.1.

Given a parameter estimate $\hat{\theta}$ and an estimated binned show-up profile $B_{\hat{\theta}}$, we can forecast arrivals across a day by "stacking" the binned show-up profile in front of each scheduled flight departure time. That is, to forecast the number of passengers that will arrive in sensor interval j, we sum the probabilities that a passenger will arrive in sensor interval j, over all flights departing after the interval. To make this idea more precise, consider a future schedule of K^F flights departing at times $\{\tau_k^F,\, k = 1, \ldots, K^F\}$ with N_k^F passengers on the kth flight. As before, we assume that we observe all arrivals within a day (i.e., the first flight of the day occurs after time a and the final flight occurs before time T). Thus, for each sensor interval i, we obtain the forecasted number of arrivals as

$$\hat{A}(i) = \sum_{k=1}^{K^F} N_k^F B_{\hat{\theta}}(\text{int}^F(k) - i), \tag{4}$$

where the kth flight in the future flight schedule departs in time interval $\text{int}^F(k)$.

In practice, the number of passengers on each flight N_k^F needs to be estimated. In the airport terminal that is the focus of our study all flights were on Boeing 737 airplanes (with a few exceptions on similarly-sized Airbus airplanes), so we estimate N_k^F by dividing the total number of passenger arrivals that day by the total number of flights. In reality, the airport could fine-tune this estimate using information available from the airlines on the number of passengers per flight or by simply multiplying the known capacity of each flight by an expected load factor (for example, 95%). Load factors themselves are relatively stable and predictable.

The remainder of this section investigates the predictive power of forecasts generated using the structural method. We seek to answer several questions through this analysis: (a) How does the predictive performance of structural forecasting compare with that of benchmark methods? (b) How can we extend structural forecasting to improve its performance? (c) How can we combine the benefits of structural forecasting with those of other approaches?

4.1. Benchmark Forecasting Methods

As benchmarks for structural forecasting, we consider several time-series forecasting methods based on autoregressive and machine learning models. These include the autoregressive model ARIMAX and several machine learning-inspired approaches including lasso and ridge regression, regression trees, bagging trees, and random forests. We train all benchmark models using a broad set of covariates, including time-of-day and day-of-week indicators, and counts of future flights, past flights, and past passenger arrivals. Table 1 includes a list of covariates employed by our benchmark forecasting methods. Specifically, the variables $\mathrm{next\_}*$ count the number of flights scheduled to depart in an upcoming time period; e.g., $\mathrm{next\_60min}$ counts the number of flights scheduled to depart in the next 60 minutes after the focal time period. The variables $\mathrm{prev\_}*$ count the number of flights scheduled to depart in a trailing period of time; e.g., $\mathrm{prev\_60min}$ counts the number of flights scheduled to depart in the 60 minutes prior to the focal time period. The variables $\mathrm{lag\_}*$ count the number of passengers who arrived in a past 15-minute interval; e.g., $\mathrm{lag\_15min}$ counts the number of arrivals in the 15-minute interval ending 15 minutes prior to the focal interval, whereas $\mathrm{lag\_1day\_30min}$ counts the number of passengers who arrived in the 15-minute interval that ended 1 day and 30 minutes prior to the focal time period.

description	variables
temporal variables	hour, weekday, month indicators
numbers of upcoming flights	next_30min, next_60min, next_90min, next_120min, next_150min, next_180min
numbers of past flights	prev_30min, prev_60min, prev_120min
lagged arrival counts	lag_15min, lag_30min, lag_45min, lag_60min, …, lag_120min, lag_1day_0min, lag_1day_15min, lag_1day_30min, …, lag_1day_120min, lag_1week_0min, lag_1week_15min, lag_1week_30min, …, lag_1week_120min, lag_2week_0min, lag_2week_15min, lag_2week_30min, …, lag_2week_120min, lag_3week_0min, lag_3week_15min, lag_3week_30min, …, lag_3week_120min, lag_4week_0min, lag_4week_15min, lag_4week_30min, …, lag_4week_120min

Table 1 List of all variables available to the benchmark forecasting models.

We note that the structural forecast does not rely on lagged variables and therefore is not adaptive. That is, the structural point forecast is the same regardless of the forecast lead time. This is an advantage for implementation, as the structural approach can be calculated well in advance with a consistent set of covariates, but it is a potential drawback in terms of performance. We present one way to add adaptivity to the structural forecast in Section 4.5.

For the benchmark methods, we consider forecast lead times ranging from 15 minutes (equivalent to one period in advance, given our sensors provide counts at 15-minute intervals) to four weeks (reflecting the planning horizon for certain staffing decisions). Longer lead times constrain the set of lagged covariates available to the benchmark forecasting methods. For example, for a 60-minute forecast lead time, the benchmark approaches only have access to lags greater than or equal to 60 minutes. All lagged arrivals are available to the benchmark methods when considering a 15-minute lead time; no lagged arrivals are available to the benchmark methods when considering a 4-week lead time.

In the results that follow, we measure the accuracy of our point forecasts using root-mean squared error (RMSE) where, given true arrival observations $A(1),\ldots,A(T)$ and arrival forecasts $\hat{A}(1),\ldots,\hat{A}(T)$,

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(A(i)-\hat{A}(i)\right)^2}$$

for forecasting in sensor intervals $1,\ldots,T$. We omit the performance between 10:00 p.m. and 2:00 a.m. when the TSA checkpoint area is idle. Results in this section are based on a training set ranging from January 1, 2022 until April 14, 2022 and a test set ranging from April 15, 2022 until May 9, 2022. We consider other training and test sets in Appendix EC.3, where we find that our main qualitative findings are robust.

4.2. Configuring and Tuning

We seek to narrow the set of benchmark methods to simplify our later analysis. Table 2 reports on the performances of several of the methods.

	forecast lead time							
method	15 minutes	1 hour	1 day	1 week	2 weeks	3 weeks	4 weeks	no lagged arrival data
lasso regression	12.03	12.77	12.87	13.12	13.92	14.43	14.80	15.68
ridge regression	12.21	12.78	12.85	13.18	13.98	14.49	14.83	15.61
bagging tree	12.34	12.37	12.45	12.52	12.81	13.15	13.20	13.62
random forest, $m=\sqrt{p}$	11.95	12.91	12.86	12.84	13.27	13.32	12.91	12.66
random forest, $m=p/3$	11.96	12.37	12.41	12.39	12.82	12.86	12.76	12.57
ARIMAX	12.35	17.00	21.04	23.61	23.61	23.61	23.61	23.61

Table 2 Performance of machine learning methods (RMSE) across various lead times.

The two regularized regression models, lasso and ridge regression, show similar performances across all lead times. In each model, the weight of the regularization is chosen using 5-fold cross validation. For clarity of presentation in our analysis going forward, we use ridge regression to represent these models.

In each random forest model, we build 100 trees and randomly sample m variables (out of the total number of variables p which depends on the lead time) as candidates at each split. We use both $m = \sqrt{p}$ and $m = p/3$, which are typical choices in the literature (see, e.g., Breiman 2002, James et al. 2013). Bagging trees are random forests with m equal to the total number of variables p, or equivalently, random forest models that allow all variables to be candidates at each split. The three tree-based methods exhibit similar performances to each other across various lead times, and we select the random forest with $m = p/3$ to highlight in our remaining results.

The random forest models perform similarly to the regularized regression models when the lead time is small. However, as the lead time increases, the random forest model outperforms the regularized regression models. To understand why, we note that the relationship between the recent lagged arrival variables and the upcoming number of arrivals is plausibly approximately linear, whereas the way the upcoming number of arrivals relates to flight and temporal variables is likely non-linear. It is reasonable that a weighted average of the recent lagged arrival data is a good predictor of the upcoming number of arrivals, thus both sets of models do well when there is a short lead time. Nevertheless, when recent lagged arrival data is not available, the models have to rely on flight and temporal data instead. As random forest models allow for non-linear predictions, they have an advantage in this context, achieving better performance when the lead time is large.

Finally, we also consider an ARIMAX model, i.e., an ARIMA model with regressors. For ARIMAX, rather than feeding the model all the lagged arrival counts displayed in Table 1, we choose the set of past arrival counts to consider based on cross validation. This model compares similarly to others when predicting the number of arrivals in the next 15 minutes. However, when forecasting with longer lead times, the ARIMAX model's performance degrades quickly. We note that as the lead time increases, we expect the ARIMAX model to behave similarly to an unregularized linear regression model. We observe that the ARIMAX model is dominated by the regularized regression models in this case. For this reason, we choose not to highlight ARIMAX as a benchmark in the results to follow.

For each of the structural forecasting methods, the show-up profile is assumed to follow a truncated normal distribution (truncated between 0 and 5 hours). We also consider other distributions and show their performances in Table 3. As all performances were similar, we default to the truncated normal distribution.

distribution of show-up profile	RMSE
truncated normal distribution (min = 0, max = 5)	13.74
beta distribution (min = 0, max = 5)	13.83
triangular distribution	13.87

Table 3 RMSE when assuming show-up profiles that follow different parametric distributions.

4.3. Evaluating Predictive Performance

Our first insight is that the structural forecasting method performs similarly to the benchmark machine learning models, though slightly worse than the random forest model. Figure 5 plots the RMSE of the structural method, where the show-up profile is estimated through maximum likelihood estimation, and of well-performing machine learning methods, ridge regression and random forest, across various lead times. The RMSE values furthest to the right on the x-axis in Figure 5 correspond to forecasts computed with no lagged arrival counts. Each error can be viewed as the approximate expected error of forecasting arrivals in any 15 minute period between 2:00 a.m. and 10:00 p.m.

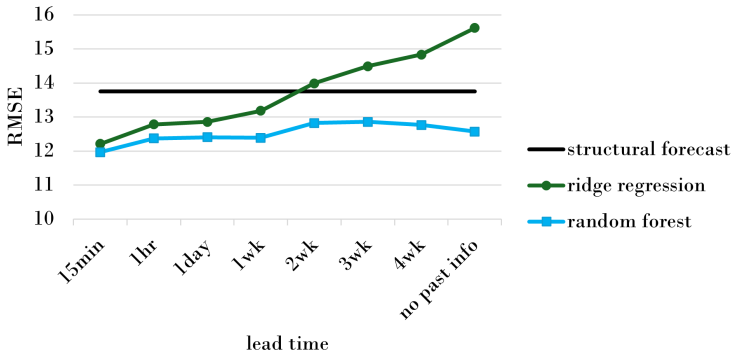

Figure 5 Performance of (i) structural forecasting with a truncated normal show-up profile, (ii) ridge regression, and (iii) random forest.

We note that the RMSE of the structural forecast appears as a straight horizontal line in Figure 5. The structural method is invariant to lead time because it estimates the number of arrivals based only on the show-up profile, an estimate of passenger counts on each flight, and the flight schedule within the test set. In particular, the structural method does not take lagged arrival counts into account.

The similarity in performance is surprising given how few variables the structural method uses versus the machine learning methods. While the structural method is only given the flight schedule and an estimated number of passengers on each flight, the machine learning methods are given the flight schedule, lagged arrival data, and temporal information. Even in the case of a long lead time or no lagged arrival data, the ability of the machine learning methods to adjust the forecast based on temporal information such as time of day and day of week give the machine learning methods an advantage over the structural methods.

Looking at the performance of the random forest model in Figure 5, we notice that it levels off for long lead times and even slightly improves when there is no lagged data relative to 2- and 3-week forecast lead times. That is, the random forest model can actually benefit from having access to less data. Our interpretation of this observation is that for long lead-times the random forest approach is better off learning the relationship between passenger arrivals and flight departures than relying on outdated temporal passenger arrival patterns. Eliminating lagged passenger arrivals from the training set forces the random forest to learn

the relationship between passenger arrivals and flight departures. This demonstrates that more data is not always better and that there can be advantages to using only a few variables which are known to be highly relevant according to a known structure.

4.4. Improving the Structural Methods

A simple approach to improving the structural methods is to train two distinct show-up profiles: one for passengers with flights departing before 9:00 a.m. (referred to as "pre-9am") and another for passengers with flights departing after 9:00 a.m. ("post-9am"). This can be achieved by allowing the mean of each distribution to differ. The resulting show-up profiles are displayed in Figure 6. Here, the standard deviation for each distribution is assumed to be equal and estimated to be 37 minutes. The fitted distributions have means of 54 and 75 minutes for the pre-9am and post-9am flight departure times, respectively. That is, passengers arriving for early flights tend to arrive later, closer to their departure times, while passengers arriving for later flights tend to arrive farther in advance of their flights.

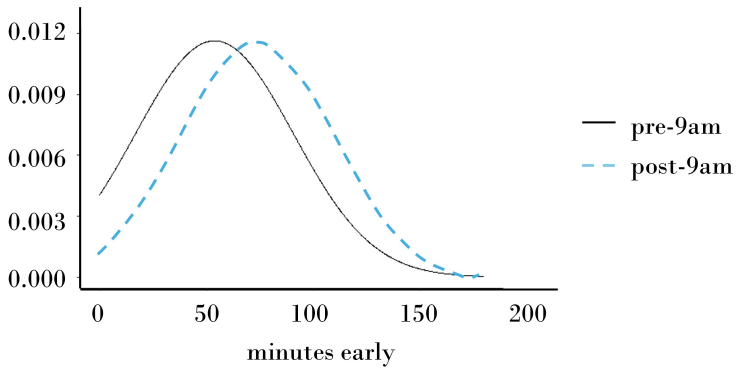

Figure 6 Two show-up profiles fitted for passengers arriving for flights scheduled to depart in the morning before 9:00 a.m. (solid curve) and flights scheduled after 9:00 a.m. (dashed curve). The pre-9am show-up profile shows passengers that arrive closer to their flight's departure time.

The difference between the two show-up profile means in Figure 6 is over 20 minutes, suggesting that we may achieve a substantially better model fit using two show-up profiles: one for pre-9am flights and another for post-9am flights. We find that this approach indeed yields an improvement in performance as depicted in Figure 7. We may try estimating different show-up profiles according to other characteristics. For example, Figure 7 shows that fitting different show-up profiles by both time of day (pre-9am versus post-9am) and destination yields a further improvement in out-of-sample forecast performance. There are 14 destination airports represented in our data, and we incorporate this information in our analysis using indicator variables.

Figures 6 and 7 suggest that the show-up profile of a passenger depends on their flight's characteristics. We further explore this idea in Section 5.

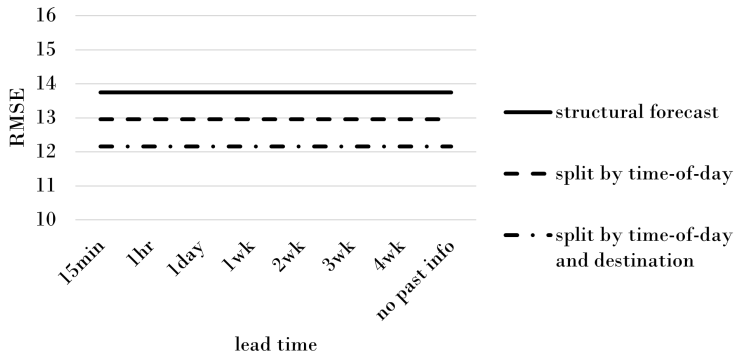

Figure 7 Performance of (i) basic structural forecasting, (ii) structural forecasting with pre-9am and post-9am profiles, and (iii) structural forecasting with profiles varying by time of day and destination.

4.5. Combining Structural Methods and Machine Learning for Best Performance

We next develop a method to leverage each technique's advantage. Namely, we employ the specific structure captured by the structural method and also the enhanced information and variable selection available to the machine learning methods. To this end, we first forecast using the structural method. Then, the forecasted estimates are fed to the machine learning models as an additional feature variable. We show the performance of this combined technique in Figure 8 alongside the structural forecast and random forest methods alone.

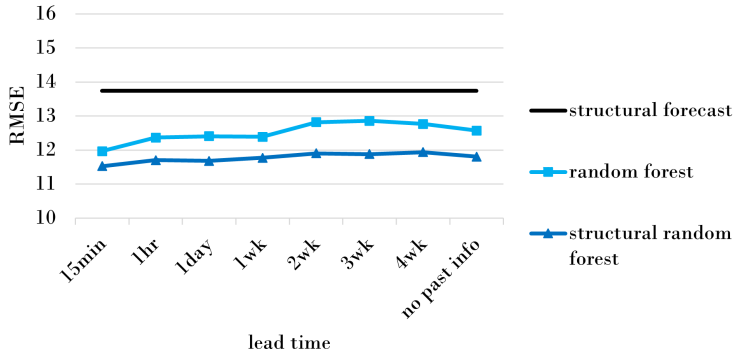

Figure 8 Performance of (i) basic structural forecasting, (ii) random forest, and (iii) a combined method.

By combining the structural and machine learning techniques, we allow the structural estimate to adapt to lagged arrival information and incorporate relevant temporal variables. Unlike the traditional machine learning techniques, however, incorporating the structural estimates relieves the machine learning method from discovering the relationship between flight patterns and arrivals; instead, this relationship is effectively summarized by a single informative covariate. As seen in Figure 8, this results in improved performance over each method used in isolation.

REMARK 1. Numerical analysis suggests that the parameter estimates of the show-up profile quickly converge. Even after only using one day of training data (in this example, April $12, 13,$ and 14 of 2022), we have the estimates of $60.0, 69.7, 65.7$ and $42.4, 34.3, 32.4$ minutes, respectively, for the mean and standard deviation compared to 66.0 and 36.6 minutes when using four and a half months of training data (January 1, 2022 to April 14, 2022).

5. Analysis of Factors Impacting Show-Up Profiles

In the preceding sections, we utilize our capability to estimate show-up profiles exclusively to generate predictive models. Nevertheless, the capability to estimate show-up profiles that vary with flight characteristics such as time of day and destination (as demonstrated in Section 4.4) opens the possibility of using our approach to perform structural empirical analysis on the factors impacting passenger's arrival times before their flights. These insights, developed specifically for an airport, can help managers with activities such as shaping passenger behavior (e.g., through targeted public service announcements encouraging passengers to arrive early) and scheduling flights, where understanding how adjacent flights might impact crowding at common service resources could be valuable. Our goal in this section is not to provide a comprehensive empirical analysis of show-up behaviors, but rather to demonstrate our method's usefulness for such an analysis. In particular, we examine how flight characteristics—departure time, destination distance, and destination purpose (leisure, business, or mixed)—affect the average earliness of passengers.

As in Section 4, we assume that the show-up profile of passengers follows a normal distribution truncated between 0 and 5 hours. We also assume that all passengers share a show-up profile with a common standard deviation σ and a mean that can depend on characteristics of their respective flight. As before, we estimate the show-up profile using the maximum likelihood estimate defined in equation (2) using the same training set as in Section 4. We begin with an analysis of the effect of early morning departure times in Section 5.1 and continue with an extended analysis in Section 5.2 that considers multiple destination features.

5.1. The Effect of Early Morning Flights

Consider the effect that an early morning departure (departures at or before $9{:}00$ a.m.) can have on the show-up profiles of passengers. We hypothesize that early morning departures result in passengers arriving significantly later, closer to their departure time. To motivate this hypothesis, we may consider that the passenger deciding when to arrive at the airport balances the cost of arriving early (which implies less time engaging in other activities) with the cost of missing their flight. For an early morning flight, the cost of arriving early may be inflated because it requires waking up early and navigating morning traffic. In fact, differences in arrival behavior for early morning flights are considered in the literature; see, for example, Chun and Mak (1999) and Postorino et al. (2019), which train show-up profiles based on observations of passenger earliness. Both papers show differences in the distribution based on time of flight departure. For example, Postorino et al. (2019) find differences in show-up profiles of passengers, in particular between evening and morning flights. They find that those flying in the morning arrive closer to their departure time than those flying in the evening. As indicated in Figure 6, early morning and later show-up profiles appear to differ in our data.

To test this hypothesis, we assume that for each passenger arriving for a respective flight with departure time t, the mean of their show up profile is equal to

$$\mu = \mu_{\text{pre9am}} + \alpha_{\text{post9am}} \cdot \mathbb{I}\{t > 9{:}00 \text{ a.m.}\},$$

where $\mathbb{I}\{s\}$ equals 1 if statement s is true and 0 otherwise. Assuming that passenger arrivals are independent and identically distributed, we estimate μ_{pre9am} and α_{post9am} along with the standard deviation σ using the maximum likelihood estimate in equation (2). Table 4 gives the estimate for each parameter along with the standard error, estimated by using the Hessian of the likelihood function. We also report the t-statistic that corresponds to testing if α_{post9am} is different from zero, i.e., the mean of the show-up profile for early morning flights differs from that for non-early morning flights.

Show-up Profiles Split by Departure Time

	estimate	standard error	t-statistic
pre-9am mean, μ_{pre9am}	53.72^{***}	0.24	223.83
post-9am effect, α_{post9am}	21.50^{***}	0.31	69.12
standard deviation, σ	36.93^{***}	0.18	205.17

Table 4 Parameter estimates and standard errors for show-up profiles differing by early morning departures. Significance levels of 0.10, 0.05, and 0.01 are indicated by $*$, $**$, and $***$, respectively.

The analysis in Table 4 is equivalent to the training of the show-up profiles in Section 4.4. The post-9am effect, α_{post9am}, is significantly greater than zero. Compared to pre-9am passengers, post-9am passengers are estimated to arrive on average 21.5 minutes earlier for their flights.

5.2. The Effect of Destination Characteristics

Building on the improved predictive performance observed in Section 4.4 by allowing the show-up profile mean to vary by destination, we explore factors associated with destination that may affect average earliness.

We may hypothesize that leisure travelers, who tend to be less comfortable with navigating the airport, arrive earlier than passengers traveling for business. This is supported by conversations with our airport contacts and appears in the literature as well (Chun and Mak 1999, Ashford et al. 2011, Al-Sultan 2018). Ashford et al. (2011) recommends that airport planning personnel consider important factors such as flight purpose, whether travel is for business or leisure, when developing a show-up profile. Chun and Mak (1999) and Al-Sultan (2018) consider show-up profiles that differ based on the known class (first class or economy) of each passenger. Chun and Mak (1999), in particular, uses earliness observations to show that first class passengers are expected to arrive closer to their flight times on average than those in other classes. Although we cannot determine the characteristics of passengers, we classify destinations as either *business*, *leisure*, or *mixed*. For this analysis, we use the following classifications for the 14 different destinations seen in the training or test set: *business* for DAL, MDW, and STL; *leisure* for FLL, LAS, MCO, MIA, and TPA; and *mixed* for ATL, AUS, BNA, BWI, DEN, and PHX.

A second hypothesis is that the destination distance affects the show-up profile of passengers. On one hand, passengers who are traveling a long distance may incur greater costs of missing their flights. In fact, Ashford et al. (2011) claim that "generally the longer the distance traveled, the greater the time allowed by passengers prior to time of scheduled departure." On the other hand, passengers who are traveling a long distance are more likely to check bags, which may delay the time for them to reach the TSA checkpoint

area. In the literature, distance is commonly accounted for by simply determining different show-up profiles for domestic versus international travel (see, for example, Ashford et al. 1976, Park and Ahn 2003, Bruno et al. 2019). We specifically focus on a terminal that only handles domestic travel to see if the distance to destination affects the show-up profile.

Third, we may consider whether a passenger's flight is to their final destination or whether they will be connecting to another flight. We may hypothesize that a passenger with a connecting flight may incur a higher cost of missing their initial flight. Again, this is something we cannot see in the data at the individual passenger level. However, we can capture this effect by categorizing destinations into those that commonly have connections and those that do not. During the dates of our training data, only one airline, Southwest Airlines, departed from the terminal studied. Southwest Airlines does not have official hubs. However, according to Russell (2020) and Ahlegren and Hardiman (2022), the following airports are commonly used for connections: ATL, BNA, BWI, DEN, MDW, PHX, and STL. We label these airports as *connecting* destinations. In the test data, there are also a few flights on Spirit Airlines to the destinations MCO and MIA. We continue to consider these destinations as non-connecting destinations for both airlines when evaluating test performance.

Taking into account all of the following variables along with the effect of early-morning departures, we model the average earliness as

$$\begin{aligned}\mu = \mu_0 &+ \alpha_{\text{post9am}} \cdot \mathbb{I}\{t > 9{:}00 \text{ a.m.}\} + \alpha_{\text{mixed}} \cdot \mathbb{I}\{p = \text{mixed}\} + \alpha_{\text{business}} \cdot \mathbb{I}\{p = \text{business}\} \\ &+ \alpha_{\text{connecting}} \cdot \mathbb{I}\{c = \text{connecting}\} + \beta \cdot d/100\end{aligned}$$

for a passenger on a flight with departure time t and destination with purpose p, connecting status c, and distance d. Table 5 displays the estimate for each parameter and the corresponding standard error, as well as t-statistics as in Table 4.

	estimate	standard error	t-statistic
mean, μ_0	63.91^{***}	0.54	118.35
post-9am effect, α_{post9am}	22.86^{***}	0.35	64.84
mixed destination effect, α_{mixed}	-6.47^{***}	0.72	-8.99
business destination effect, α_{business}	-9.01^{***}	0.74	-12.15
connecting effect, $\alpha_{\text{connecting}}$	-5.66^{***}	0.57	-9.88
distance effect, β ($\times 100$ miles)	-0.035	0.0391	-0.898
standard deviation, σ	38.73^{***}	0.19	203.84

Table 5 Parameter estimates and standard errors for show-up profiles varying based on flight characteristics. Significance levels of 0.10, 0.05, and 0.01 are indicated by $*$, $**$, and $***$, respectively.

As in Section 5.1, the early-morning effect is significant. In fact, the effect is estimated to be similar to that reported in Section 5.1. Relative to post-9am passengers, pre-9am passengers are estimated to arrive on average 23 minutes later for their flights.

We also now consider the effect of destination by looking at (1) whether this destination is classified as leisure, business, or mixed; (2) whether this destination serves as a hub for Southwest Airlines; and (3) the distance to the destination. We find that passengers traveling to leisure destinations arrive significantly sooner (i.e, further in advance of their flight) on average than those traveling to mixed or business destinations. This agrees with our hypothesis. On average, we see that the leisure destination passengers arrive about 6.5 minutes earlier than the mixed destination passengers and 9 minutes earlier than the business destination passengers.

We also find that those traveling to destinations that work as hubs for Southwest Airlines arrive later on average (specifically, an average of 5.66 minutes later). This is in contrast to what we expected. One possible explanation for this observation is that passengers who are more likely to connect are also more likely to check their bags. Because we measure the time until reaching the TSA checkpoint area, checking a bag delays a passenger's arrival to our sensors.

Finally, we find that destination distance, when looking at domestic travel only, does not appear to have a significant effect when controlling for other factors.

6. Conclusion and Future Research Directions

This paper demonstrates the estimation of show-up profiles from airport data that does not match arriving passengers with departing flights. Such data can be collected using sensors that are relatively easy to install and maintain. We present an estimation approach based on maximum likelihood that is consistent and efficient. Such an approach allows managers to estimate show-up profiles that are customized to particular settings, that can be updated frequently, and that account for relevant covariates. The resulting show-up profiles can be used to generate lead-time-independent structural forecasts whose quality rivals that of machine learning approaches based on a large set of time-series covariates in our airport setting. Furthermore, we show that by incorporating the structural forecasts into machine learning approaches, we obtain adaptive forecasts that significantly improve upon both approaches. It is clear that the structured information provided by the show-up profile model adds value even to sophisticated forecasting approaches. Show-up profiles bring the advantages that they are easy for airport managers to understand and to manipulate when conducting what-if analyses and when forecasting in new environments. We demonstrate that our estimation approach can also be used to test hypotheses on how passenger behavior depends on factors of interest.

As mentioned previously, while we tested our approaches in an airport setting where the notion of show-up profiles is known to be useful, we believe that show-up profiles and our methods have a broader set of potential applications in a more general set of *scheduled service* applications in which arrivals are driven by scheduled events. These include applications in other transportation settings, entertainment, and healthcare. A related problem is where customer arrivals are easily observed or predicted and customers remain in an environment according to a *linger profile*; that is, a distribution of times in system. Variations of our methods

may be used to estimate linger profiles and to forecast the population of customers in such a system over time.

Section 5 of our paper tests some basic insights into what factors correlate with passenger earliness. We believe our methods can be applied to perform more detailed empirical research on passenger earliness, for example to better understand heterogeneity among passengers and the mechanisms driving their arrival choices.

Acknowledgments

The authors are grateful for the support provided by the Raleigh-Durham Airport Authority and would like to thank in particular Bill Sandifer, Executive Vice-President and Chief Development Officer, and Delia Chi, Vice President of Planning and Sustainability, for their continued help throughout this research project. The authors also acknowledge helpful feedback from various seminar and conference participants.

References

Ahlegren, L. and Hardiman, J. (2022). Which airport does each major US carrier call home? *Simple Flying*. https://simpleflying.com/major-us-carrier-hubs/ (Accessed on September 26, 2024).

Ahyudanari, E. and Vandebona, U. (2005). Simplified model for estimation of airport check-in facilities. *Journal of the Eastern Asia Society for Transportation Studies*, 6:724–735.

Al-Sultan, A. T. (2018). Simulation and optimization for modeling the passengers check-in system at airport terminal. *Review of Integrative Business and Economics Research*, 7(1):44.

Alodhaibi, S., Burdett, R. L., and Yarlagadda, P. K. (2019). Impact of passenger-arrival patterns in outbound processes of airports. *Procedia Manufacturing*, 30:323–330.

Ashford, N., Hawkins, N., O'Leary, M., Bennetts, D., and McGinity, P. (1976). Passenger behavior and design of airport terminals. *Transportation Research Record*, 588(3):18–26.

Ashford, N. J., Mumayiz, S., and Wright, P. H. (2011). *Airport Engineering: Planning, Design, and Development of 21st Century Airports*. John Wiley & Sons.

Bergin, M. (2024). Raleigh-Durham International Airport reports record passenger traffic in January 2024. *WRAL*. https://www.wral.com/story/raleigh-durham-international-airport-reports-record-passenger-traffic-in-january-2024/21285965 (Accessed on August 14, 2024).

Bernstein, F., Keskin, N. B., Mersereau, A., Wood, M., and Ziya, S. (2024). Data-driven population tracking in large service systems. *Working Paper*.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3.1. *Statistics Department, University of California, Berkeley*, 1(58):3–42.

Bruno, G., Diglio, A., Genovese, A., and Piccolo, C. (2019). A decision support system to improve performances of airport check-in services. *Soft Computing*, 23:2877–2886.

Caro, F. and Sadr, R. (2019). The internet of things (iot) in retail: Bridging supply and demand. *Business Horizons*, 62(1):47–54.

Cheng, L. (2014). *Modelling Airport Passenger Group Dynamics Using an Agent-based Method*. PhD thesis, Queensland University of Technology.

Chun, H. W. and Mak, R. W. T. (1999). Intelligent resource simulation for an airport check-in counter allocation system. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 29(3):325–335.

Dantas, L. F., Fleck, J. L., Oliveira, F. L. C., and Hamacher, S. (2018). No-shows in appointment scheduling—A systematic literature review. *Health Policy*, 122(4):412–421.

Gao, Y. (2021). ACRP problem statement 454: Describing, modelling and predicting passengers' arrival patterns. Technical report, Airport Cooperative Research Program.

Guo, X., Grushka-Cockayne, Y., and De Reyck, B. (2020). London Heathrow airport uses real-time analytics for improving operations. *INFORMS Journal on Applied Analytics*, 50(5):325–339.

Guo, X., Grushka-Cockayne, Y., and De Reyck, B. (2022). Forecasting airport transfer passenger flow using real-time data and machine learning. *Manufacturing & Service Operations Management*, 24(6):3193–3214.

IATA (1995). *Airport Development Reference Manual*. International Air Transport Association.

Ibrahim, R. and L'Ecuyer, P. (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. *Manufacturing & Service Operations Management*, 15(1):72–85.

Isarsoft (2024). What is a passenger show up profile? https://www.isarsoft.com/knowledge-hub/show-up-profile (Accessed on October 15, 2024).

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.

Josephs, L. (2024). Air travel demand is breaking records. Airline profits aren't. *CNBC*. https://www.nbcnews.com/business/business-news/air-travel-demand-breaking-records-airline-profits-are-not-rcna160663 (Accessed on August 14, 2024).

Jouini, O., Benjaafar, S., Lu, B., Li, S., and Legros, B. (2022). Appointment-driven queueing systems with non-punctual customers. *Queueing Systems*, 101(1):1–56.

Kennon, P., Hazel, R., Ford, E., and Hargrove, B. (2013). ACRP report 82: Preparing peak period and operational profiles—Guidebook. Technical report, Transportation Research Board of the National Academies, Washington, D.C.

Kim, S.-H., Vel, P., Whitt, W., and Cha, W. C. (2015). Poisson and non-Poisson properties in appointment-generated arrival processes: The case of an endocrinology clinic. *Operations Research Letters*, 43(3):247–253.

Kim, S.-H., Whitt, W., and Cha, W. C. (2018). A data-driven model of an appointment-generated arrival process at an outpatient clinic. *INFORMS Journal on Computing*, 30(1):181–199.

Koeleman, P. and Koole, G. (2012). Appointment scheduling using optimisation via simulation. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–5. IEEE.

Lalita, T. and Murthy, G. (2022). The airport check-in counter allocation problem: A survey. Working paper.

Laudanski, K., Moon, K., Singh, A., Chen, Y., and Restrepo, M. (2021). The characterization of the toll of caring for coronavirus disease 2019 on icu nursing staff. *Critical Care Explorations*, 3(4):e0380.

Lee, H. S., Kesavan, S., and Deshpande, V. (2021). Managing the impact of fitting room traffic on retail sales: Using labor to reduce phantom stockouts. *Manufacturing & Service Operations Management*, 23(6):1333–1682.

Liu, N., Ziya, S., and Kulkarni, V. G. (2010). Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management*, 12(2):347–364.

Lu, Y., Musalem, A., Olivares, M., and Schilkrut, A. (2013). Measuring the effect of queues on customer purchases. *Management Science*, 59(8):1743–173.

Martínez-de Albéniz, V. and Belkaid, A. (2021). Here comes the sun: Fashion goods retailing under weather fluctuations. *European Journal of Operational Research*, 294(3):820–830.

Martínez-de Albéniz, V. and Valdivia, A. (2019). Measuring and exploiting the impact of exhibition scheduling on museum attendance. *Manufacturing & Service Operations Management*, 21(4):761–779.

Musalem, A., Olivares, M., and Schilkrut, A. (2021). Retail in high definition: Monitoring customer assistance through video analytics. *Manufacturing & Service Operations Management*, 23(5):1025–1042.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Park, Y. and Ahn, S. B. (2003). Optimal assignment for check-in counters based on passenger arrival behaviour at an airport. *Transportation Planning and Technology*, 26(5):397–416.

Parlar, M. and Sharafali, M. (2008). Dynamic allocation of airline check-in counters: a queueing optimization approach. *Management Science*, 54(8):1410–1424.

Perdikaki, O., Kesavan, S., and Swaminathan, J. M. (2012). Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, 14(1):145–162.

Postorino, M. N., Mantecchini, L., Malandri, C., and Paganelli, F. (2019). Airport passenger arrival process: Estimation of earliness arrival functions. *Transportation Research Procedia*, 37:338–345.

Rauch, R. and Kljajić, M. (2006). Discrete event passenger flow simulation model for an airport terminal capacity analysis. *Organizacija*.

Russell, E. (2020). Does Southwest Airlines have hubs? Yes, but don't call them that. *The Points Guy*. https://thepointsguy.com/news/does-southwest-airlines-have-hubs/ (Accessed on September 26, 2024).

Sayın, M. G., Aktaş, D. Y., Bolat, M., Çelenli, M. K., Dursun, B., Koç, G., and Üçkardeş, K. S. (2023). A study of predicting arrival patterns of airport passengers to the counters on the basis of international terminal. *Avrupa Bilim ve Teknoloji Dergisi*, (51):63–74.

Shi, X. (2022). Lecture 2: Some useful asymptotic theory. https://users.ssc.wisc.edu/~xshi/econ715/Lecture_2_some_asymptotic_theorems.pdf (Accessed on October 15, 2024).

TransSolutions (2011). ACRP report 55: Passenger level of service and spatial planning for airport terminals. Technical report, Transportation Research Board of the National Academies, Washington, D.C.

U.S. Department of Homeland Security (2007). Privacy impact assessment for the boarding pass scanning system. https://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_tsa_bpss.pdf (Accessed on September 26, 2024).

van der Vaart, A. and Wellner, J. A. (2023). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer.

Electronic Companion for "Show-Up Profiles for Scheduled Services: Estimation and Applications"

Appendix EC.1: Extension of the Maximum Likelihood Estimator

This appendix extends the setting in Section 3 to scenarios in which the flight times do not coincide with interval boundaries and in which flight schedules on different days are not necessarily identical. To this end, we now suppose that there are K^m scheduled flights on day m, where flight $k \in \{1, \ldots, K^m\}$ is scheduled to depart at time $\tau_k^m \in [a, T]$. As before, N_k^m, the number of passengers scheduled for flight k on day m, is known. The total number of passengers to arrive on each day m is then equal to $\sum_{k=1}^{K^m} N_k^m$. We assume that the first flight of each day occurs at time a or later, so that all arrivals are observed. Also as before, each day is divided into I non-overlapping sensor intervals each of length $r > 0$, and sensor interval $i \in \{1, \ldots, I\}$ is defined as $(t_{i-1}, t_i]$, where $t_0 = 0$, $t_{i+1} = t_i + r$ for some $r > 0$, and $t_I = T \geq \max_m\{T_{K^m}\}$. We let X_i^m denote the number of arrivals observed during the ith interval $(t_{i-1}, t_i]$ on day m.

Without loss of generality, consider a random passenger who has a seat on flight $k \in \{1, \ldots, K^m\}$, let $s_k = \inf_{i \in \{1,\ldots,I\}}\{t_i - \tau_k : t_i - \tau_k \geq 0\}$ denote the time between the scheduled departure time of flight k and the end of the time interval this time falls into. For example, for a flight k that departs at time 10:35 a.m. in interval (10:25, 10:40], we have $s_k = 5$. For any $i \in \{1, \ldots, I\}$ and $s \leq r$, let

$$B_\theta(s, i) = D_\theta((i-1)r - s) - D_\theta(ir - s)$$

be the probability that a passenger arrives in the ith sensor interval before their flight. For any $s \in [0, r)$, $\{B_\theta(s, i),\ i = 1, \ldots, I\}$ is a probability mass function, which we call the binned distribution. Figure EC.1 presents a modification of the binned distribution example in Figure 4 to show a flight that departs at the non-interval boundary time of 7:20. In the example in Figure EC.1, the remaining time until the end of the interval is $s_1 = 10$ minutes.

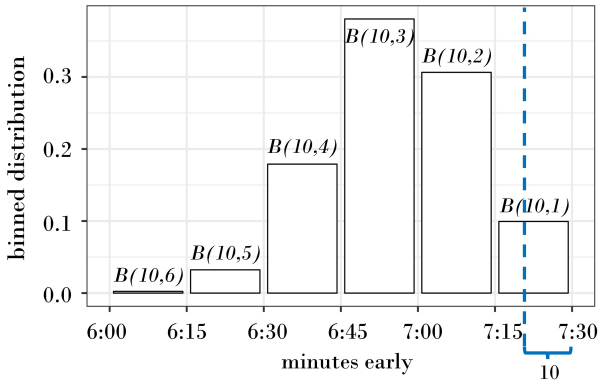

Figure EC.1 Example of a binned distribution with interval size $r = 15$ minutes, flight departure time τ_1 of 7:20, and remaining time $s_1 = 10$ minutes, where the x-axis represents the time of day. The binned distribution's dependence on θ has been suppressed for clarity of presentation.

The probability that a passenger from day m arrives in sensor interval i is

$$P_\theta^m(i) = \frac{\sum_{k=1}^{K^m} B_\theta(\text{int}^m(k) - i + 1) N_k^m}{\sum_{k=1}^{K^m} N_k^m},$$

where $\text{int}^m(k)$ denotes the sensor interval in which flight k on day m departs. As in Section 3.2, we consider a surrogate problem where each passenger is assumed to arrive independently. Then, in the surrogate problem, the counts of arrivals across all intervals on day m, $\tilde{X}^m = \{X_i^m,\, i = 1, \ldots, I\}$, follow a multinomial distribution with parameters $\sum_{k=1}^{K} N_k^m$, I, and $\{P_\theta(i),\, i = 1, \ldots, I\}$, corresponding to the number of trials, number of mutually exclusive events, and event probabilities, respectively. This results in the following likelihood function of parameter $\vartheta \in \Theta$:

$$L\left(\vartheta;\ \{\tilde{x}^m\}_{m=1}^M\right) = \prod_{m=1}^{M} \left(\frac{\left(\sum_{k=1}^{K} N_k^m\right)!}{\prod_{i=1}^{I} \tilde{x}_i^m!} \prod_{i=1}^{I} P_\vartheta^m(i)^{\tilde{x}_i^m} \right).$$

The maximum likelihood estimate of the surrogate problem is then

$$\bar{\theta}_M = \arg\max_{\vartheta \in \Theta} \left\{ \sum_{m=1}^{M} \sum_{i=1}^{I} \tilde{x}_i^m \log P_\vartheta^m(i) \right\},$$

which results in the following estimate:

$$\hat{\theta}_M = \arg\max_{\vartheta \in \Theta} \left\{ \sum_{m=1}^{M} \sum_{i=1}^{I} x_i^m \log P_\vartheta^m(i) \right\}.$$

Appendix EC.2: Consistency Results

As noted in Section 3.3, not every show-up profile is identifiable from its corresponding binned distribution. This implies that there could be scenarios in which the estimate in (1) is not consistent. Example EC.1 illustrates one such setting. The example relies on the interval sizes r being large relative to the support of the show-up profile.

EXAMPLE EC.1. Suppose that there is only one flight, departing at time 2, and intervals are of size 1. Consider a show-up profile that has a truncated normal distribution with mean earliness 1 that has known bounds 0 and 2. Two examples of resulting arrival distributions are depicted in Figure EC.2, where the standard deviations of the underlying show-up profiles differ. However, under any standard deviation value, we have that $P_\theta(1) = P_\theta(2) = 0.5 = B(1) = B(2)$. Therefore, in this example, we cannot recover the exact distribution of the earliness distribution from the available observations.

In fact, the distribution of what one would observe, i.e., the distribution of the arrivals (X_1, X_2), would be the same regardless of the standard deviation. If we aim to learn the mean and standard deviation of the show-up profile, even though we can determine that the mean must be one, the standard deviation cannot be identified. □

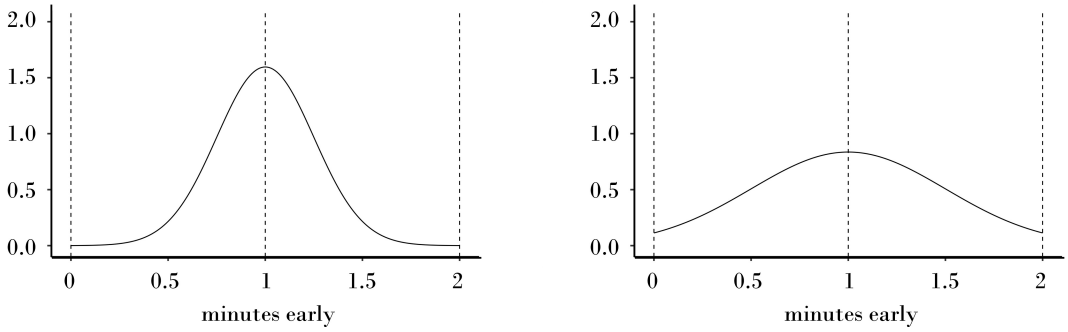

Figure EC.2 Example of two possible arrival distributions that only differ by the standard deviation of the underlying show-up profile. The show-up profile on the left panel has a standard deviation of 0.25, whereas the one on the right has a standard deviation of 0.5. The cutoffs for the intervals, $t_0 = 0$, $t_1 = 1$, and $t_2 = 2$, are displayed as dotted lines.

However, there are many other scenarios where the binned distribution does uniquely determine the parameter θ. While Example EC.1 suggests that two intervals of data may not be enough to uniquely determine the parameter θ, we conjecture that any distribution is identifiable given a sufficiently large number of intervals. Example EC.2 below motivates such an idea. In fact, relaxing the truncated normal distribution to a normal distribution and increasing the number of intervals from two to three guarantees identifiability.

EXAMPLE EC.2. Suppose that the underlying show-up profile has a normal distribution with mean μ and standard deviation θ. (While show-up profiles in practice have truncated distributions, a normal distribution is a good approximation as long as the truncation limits are sufficiently large and there is an arbitrarily small number of arrivals to occur after the flight departure time.) Rather than having two intervals as in Example EC.1, now consider three intervals: $(0, t_1]$, $(t_1, t_2]$, and $(t_2, a]$. Witnessing three intervals guarantees the consistency of the estimators for the binned distribution $\{B_\theta(1), B_\theta(2), B_\theta(3)\}$. This is equivalent to knowing two c.d.f. values, $F_\theta(t_1) = B_\theta(3)$ and $F_\theta(t_2) = B_\theta(2) + B_\theta(3)$. Hence, $t_1 = F_\theta^{-1}(B_\theta(3))$ and $t_2 = F_\theta^{-1}(B_\theta(2) + B_\theta(3))$. Normalizing, we have the two following equations:

$$\frac{t_1 - \mu}{\sigma} = F_Z^{-1}(B_\theta(3)) \text{ and } \frac{t_2 - \mu}{\sigma} = F_Z^{-1}(B_\theta(2) + B_\theta(3)),$$

where F_Z^{-1} is the inverse c.d.f. for a standard normal distribution. These can be solved resulting in parameter values

$$\mu = \frac{t_1 F_Z^{-1}(B_\theta(2) + B_\theta(3)) - t_2 F_Z^{-1}(B_\theta(3))}{F_Z^{-1}(B_\theta(2) + B_\theta(3)) - F_Z^{-1}(B_\theta(3))} \text{ and } \sigma = \frac{t_2 - t_1}{F_Z^{-1}(B_\theta(2) + B_\theta(3)) - F_Z^{-1}(B_\theta(3))}. \quad \square$$

Appendix EC.3: Robustness

In this appendix, we assess the robustness of the results from Section 4 by exploring different training and testing set splits. Specifically, we analyze the impact of shorter training periods and variations in the transition date between training and testing observations.

In Section 4, we evaluate the predictive performance of structural forecasting using arrival observations from January 1, 2022, to May 30, 2022. Observations before April 15 are used to train the show-up profile,

while the performance of structural forecasting is assessed on observations after that date. This setup results in 16 weeks of training data (January 1, 2022, to April 14, 2022) and approximately 7 weeks of testing data (April 15, 2022, to May 31, 2022). Each day is divided into ninety-six 15-minute periods, providing arrival counts for each interval.

To evaluate the robustness of our numerical results from Section 4, we apply the same forecasting method described in that section and present updated versions of Figures 5, 7, and 8.

EC.3.1 Shorter Training Period

We first examine the robustness of our results to shorter training periods. While Section 4 considers 16 weeks of training data, we now analyze performance using only 1 or 2 weeks of training data. Figures EC.3-EC.5 illustrate this comparison: in each of these figures, the left panel shows the RMSE based on 1 week of training data (April 8-14, 2022), and the right panel shows the RMSE based on 2 weeks of training data (April 1-14, 2022).

As shown in Figures EC.3–EC.5, the main findings from Section 4 remain consistent with shorter training periods. Specifically, the performance of the structural method is comparable to that of the displayed machine learning methods (Figure EC.3), structural forecasting exhibits improved performance when multiple show-up profiles are trained (Figure EC.4), and combining machine learning methods—random forest in this case—with the structural method yields further performance gains (Figure EC.5).

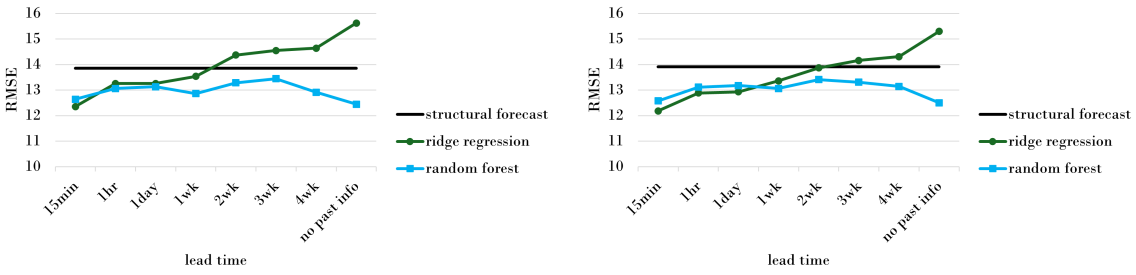

Figure EC.3 Performance of (i) structural forecasting with a truncated normal show-up profile, (ii) ridge regression, and (iii) random forest. All RMSE are similar to one another with random forest performing the best. The left panel displays the results using one week of training data, while the right panel uses two weeks of training data.

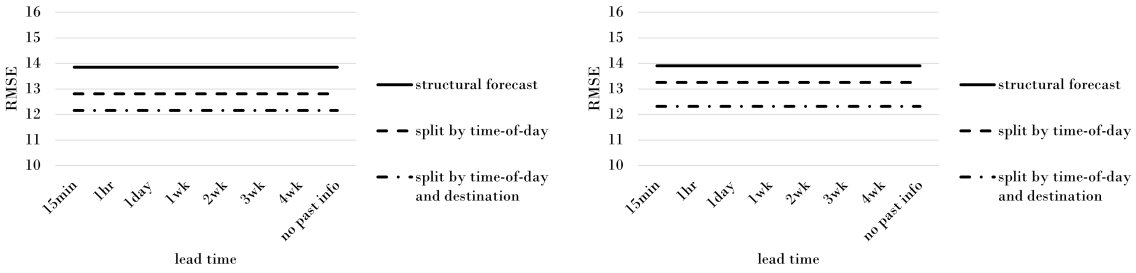

Figure EC.4 Performance of (i) basic structural forecasting, (ii) structural forecasting with pre-9am and post-9am profiles, and (iii) structural forecasting with profiles varying by time of day and destination. The left panel displays the results using one week of training data, while the right panel uses two weeks of training data.

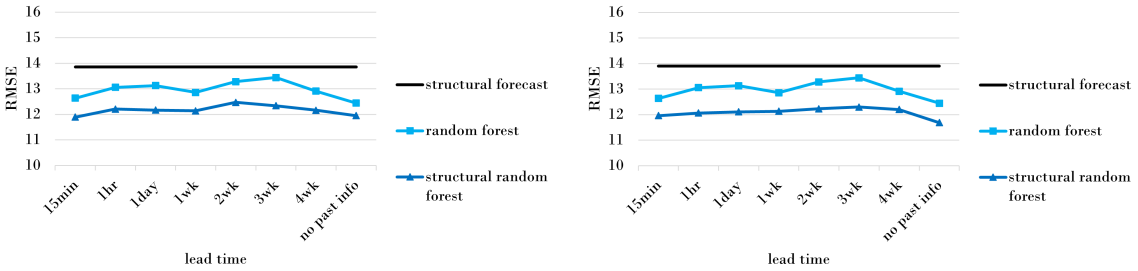

Figure EC.5 Performance of (i) basic structural forecasting, (ii) random forest, and (iii) a combined method. The left panel displays the results using one week of training data, while the right panel uses two weeks of training data.

EC.3.2 Shifted Training and Testing Sets

Section 4 uses April 15, 2022, as the first date of the test set, with all available information prior to this date (January 1, 2022, to April 14, 2022) used as training data. Here, we analyze the effect of shifting this transition date. Specifically, we evaluate performance when the first day of the test set is shifted to February 15, 2022, or March 15, 2022.

As shown in Figures EC.6–EC.8, the main findings from Section 4 remain consistent even with these changes in the transition date between training and testing data. In each figure, the left panel shows the RMSE based on training data from January 1, 2022, to February 14, 2022, while the right panel shows the RMSE based on training data from January 1, 2022, to March 14, 2022. The key insights from Section 4 hold under these conditions as well.

A few additional insights can been gained from these results. First, in Figure EC.6, we note that ridge regression outperforms random forest. This further validates the idea that these performances can all be seen as being similar. Next, in left panel of Figure EC.7, splitting the show-up profile further by destination results in a decrease in test performance in comparison to splitting only by time. This highlights the possibility of overfitting if too many show-up profiles are trained. Thus, we recommend using a validation set when splitting into many show-up profiles or using splits that are justified from empirical analysis (as discussed in Section 5).

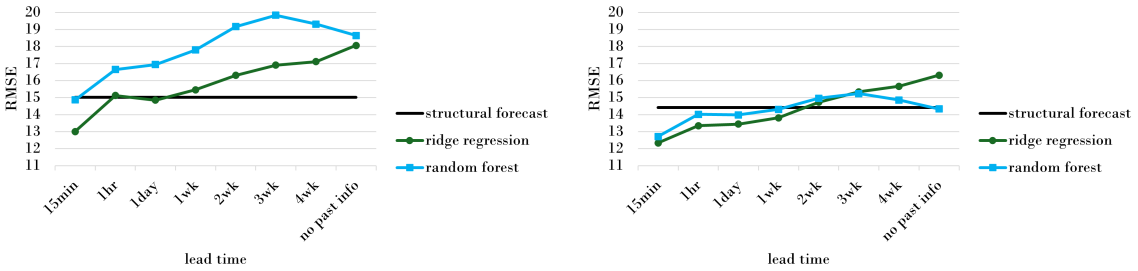

Figure EC.6 Performance of (i) structural forecasting with a truncated normal show-up profile, (ii) ridge regression, and (iii) random forest. The transition date between training and testing observations is February 15, 2022 for the left panel, whereas for the right panel, the transition date is March 15, 2022.

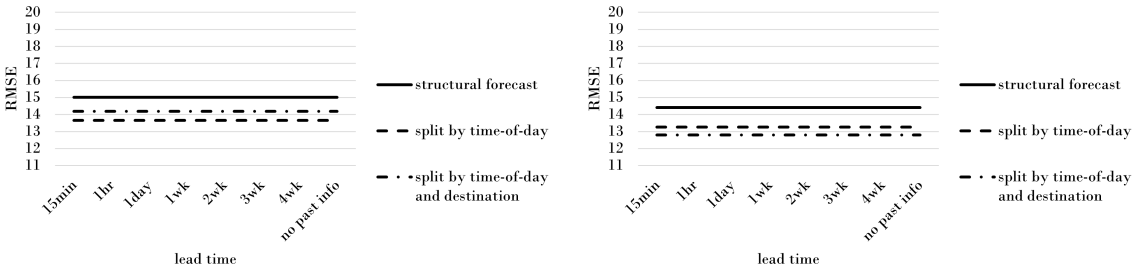

Figure EC.7 Performance of (i) basic structural forecasting, (ii) structural forecasting with pre-9am and post-9am profiles, and (iii) structural forecasting with profiles varying by time of day and destination. The transition date between training and testing observations is February 15, 2022 for the left panel, whereas for the right panel, the transition date is March 15, 2022.

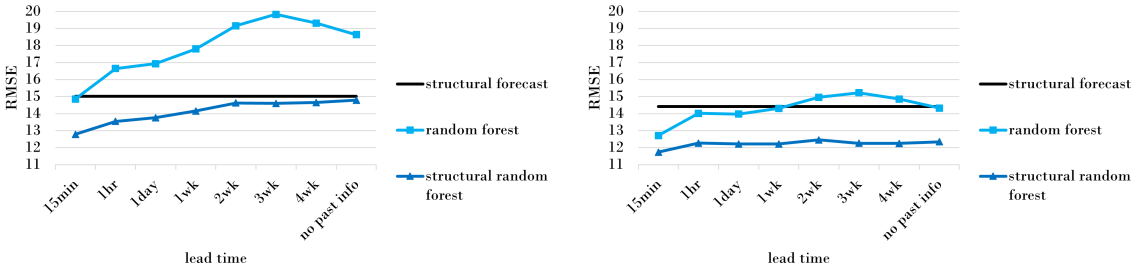

Figure EC.8 Performance of (i) basic structural forecasting, (ii) random forest, and (iii) a combined method. The transition date between training and testing observations is February 15, 2022 for the left panel, whereas for the right panel, the transition date is March 15, 2022.