

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Quantifying the Impact of Flight Predictability on Strategic and Operational Airline Decisions

Permalink

<https://escholarship.org/uc/item/9jv2d432>

Author

Hao, Lu

Publication Date

2015

Peer reviewed|Thesis/dissertation

**Quantifying the Impact of Flight Predictability on Strategic
and Operational Airline Decisions**

by

Lu Hao

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark Hansen, Chair

Professor Joan Walker

Professor Rhonda Righter

Spring 2015

Abstract

Quantifying the Impact of Flight Predictability on Strategic and Operational Airline Decisions

by

Lu Hao

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Mark Hansen, Chair

In this thesis, we examine how the predictability of travel time affects both the transportation service providers' strategic and operational decisions, in the context of air transportation. Towards this end, we make three main contributions. The first is the development of accurately measuring predictability of travel time in air transportation to best model airline decision behavior. The measure is sensitive to the different nature that's driving the decision. The second is an empirical investigation of the relationship between the best-measured travel time predictability and the transportation service providers' strategic and operational decisions to gain insights into the significance of the impact of predictability. The third contribution is proposing an algorithm to improve predictability in order to save cost in the strategic decision process through re-sequencing the departure queue at the airport.

We consider the strategic decision as the setting of the scheduled travel time for each trip that typically happened six months before the travel date. On the operational side, we investigate into the decision of the amount of fuel loaded to each flight in the daily operation. We assume that the decisions are based on the predictability of historical travel time performance. When quantifying predictability, it is important to realize that the service providers have different priority of considerations when making the strategic (scheduling) and the operational (fuel loading) decisions. Therefore, we apply different metrics for predictability to modeling the different decision behaviors and prove that the best-fitting measure of predictability is not uniform across different type of decisions. Regarding the strategic decision making, the profit-driven nature of the service provider encourages discounting the effects of extremely long historical travel times. Therefore, segmenting the historical travel time distribution is crucial in our effort of measuring predictability. On the other hand, when making day-to-day operational decisions, specifically fuel loading decisions in this study, the safety-driven nature of the service provider prevails over others and it pays more attention to extreme events. Therefore a metric capturing the tail effects such as the variance and standard deviation is a more appropriate measure of predictability in this context.

In modeling the relationship between predictability and scheduled travel time setting, we seek both analytical insights and empirical evidences. Firstly this relationship is studied with empirical data and multiple regression models. We develop the "percentile model" where the distribution of the historical travel time for an air trip is depicted by the difference between every 10th percentiles. We find that gate delay plays a minor role in setting scheduled travel time and that scheduled travel times have decreasing sensitivity

to historical travel times toward the right tail of the distribution. To specifically link schedule setting with the trip's on-time performance, a scheduled travel time adjustment model is further developed. Poor on-time performance leads to increased scheduled travel time in the next planning period. With the behavior model results showing that both the median travel time and the "inner right tail" of the distribution affect schedule setting, an impact study is conducted to validate these impacts with evidence in the historical data. This impact from behavioral modeling is validated with real data in year 2006-2008 and 2009-2011, and their corresponding scheduled travel times in the later period. Furthermore, by studying the travel performance difference based on different changes in scheduled travel time, we conclude that ignoring the impact on schedule changes when considering potential benefits of improved travel time distribution could lead to inaccurate results.

We complement the strategic behavioral modeling findings with proposing a practical algorithm that optimizes the sequence of departure queue on the airport to improve travel time predictability. The end objective is to reduce scheduled travel time through improved predictability and thus save cost for travel service providers. We present algorithms to sequence departures on a daily basis. For the objective function, scheduled travel time is viewed as a cost for airlines to be minimized. For each flight, the assigned slot generates a new travel time and this time contributes proportionally to the future scheduled travel time, as revealed in estimating the "percentile model". Assuring that the on-time performance is not greatly sacrificed is also important. Therefore the objective function also includes delaying the flight's arrival performance as part of the "cost of assignment". In this way, we develop a multi-objective algorithm to sequence departure flights to improve predictability, reduce airline scheduled travel time, and increase on-time performance.

To investigate the relationship between predictability and fuel loading decisions, we develop a set of multiple regression models considering clusters of standard deviation of the estimates. The unpredictability under performance may cause decision makers (airline dispatchers) to load more fuel onto aircraft, and thus causing extra fuel consumed to carry this excessive fuel. We acquired a large and recent dataset with flight-level fuel loading and consumption information from a major US airline. With this data, firstly the relationship between the amount of loaded fuel and travel time predictability performance is estimated using statistical model. Predictability is measured with metrics such as standard deviation of travel time so that the tail effect of the distribution is properly captured. We find that one minute of standard deviation in airborne time within a month for the same OD pair and shift of day would lead to 0.95 minute increase in loaded contingency fuel and 1.85 minute loaded contingency and alternate fuel. Then, the impact of predictability on loaded fuel is translated into fuel consumption and ultimately, fuel cost for US domestic operations. If there is no unpredictability in the aviation system, the reduction in the loaded fuel would be 6.4 and 12.5 minute per trip, respectively. This ultimately translates into a cost to US domestic air carriers on the order of \$88 – \$345 million per year.

to
my family

Contents

List of Tables	iv
List of Figures	v
Acknowledgements	vi
1. Introduction	1
1.1 Problem Statement	1
1.2 Recent Trend of Flight Predictability	3
1.2.1 Definition of Flight Predictability	4
1.2.2 Trends in Flight Predictability	4
1.2.3 Trends of Predictability Performance	6
2. Strategic Decision Modeling	9
2.1 Literature Review	10
2.1.1 Travel Time Reliability in Ground Transportation	10
2.1.2 Airline Scheduling Strategies	11
2.1.3 Industry Practice	13
2.2 Mean-variance Model	14
2.2.1 Model Specification	15
2.2.2 Estimation Results	16
2.3 Percentile Model	17
2.3.1 Data and Modeling	18
2.3.2 Estimation Results	21
2.4 SBT Adjustment Model	25
2.4.1 Model Specification	26
2.4.2 Estimation Results	27
2.5 Impact Analysis	29
2.5.1 Methodology	29
2.5.2 Results	31
2.6 Conclusion	34
3. Improving Predictability through Airport Departure Queue Sequencing	37

3.1	Departure Queue Management.....	37
3.2	Departure Queue Optimization Model.....	40
3.2.1	Problem Formulation.....	40
3.2.2	Cost of Assignment	42
3.2.3	Benefit Mechanism of the Cost of Assignment.....	44
3.2.4	Optimization Result and Impact on SBT.....	45
3.3	The Effect of On-time Performance.....	51
3.3.1	Updated Objective Function.....	51
3.3.2	Sensitivity Analysis	54
3.4	Conclusion.....	58
4.	Operational Decision Modeling.....	60
4.1	Literature Review	61
4.2	Airline Fuel Loading Practice	63
4.3	Modeling Fuel Loading Behavior	65
4.3.1	Data Collection	65
4.3.2	Estimation Methodology	66
4.3.3	Estimation Results	69
4.4	Cost-to-Carry Analysis.....	72
4.5	Conclusion.....	75
5.	Conclusion.....	77
5.1	Contributions.....	77
5.2	Future Work	79
6.	Reference	82

List of Tables

Table 2.1 Analogy of the Travel Time Reliability Concept between Ground Transportation and Air Transportation	10
Table 2.2 Estimation Results for the Mean-variance Model	16
Table 2.3 Estimation Results for the Percentile Model	23
Table 2.4 Estimation Results for the SBT Adjustment Model	28
Table 2.5 Summary of Nine Scenarios and Their Counts and Frequency in the Dataset	30
Table 2.6 Changes in SBT and Various Performance Metrics for Different Scenarios, 2008 vs. 2011	33
Table 3.1 Partial Estimation Results from the Percentile Model in Section 2.3.....	40
Table 3.2 Percentile Values and Parameters for the Hypothetical Flight	44
Table 3.3 Optimization Results of the Departure Queue Re-sequencing Algorithm	46
Table 3.4 Optimization Results of the Departure Queue Re-sequencing: Basic and Updated Objective Functions.....	52
Table 3.5 Results of the Sensitivity Analysis Changing Relative Weights in the Updated Objective Function: JFK and SFO.....	55
Table 4.1 Estimation Results for Fuel Uplift Models	70
Table 4.2 Annual Cost to Carry Additional Contingency and Alternate Fuel from for Unpredictability in Terms of Fuel lbs, Fuel cost (*: in millions), and CO2.	74

List of Figures

Figure 1.1 Sources of EFT Variance, Aggregated over 2004 to 2010.....	5
Figure 1.2 Trends of EFT, FT, Departure Delay Variances, 2004 to 2010.	6
Figure 1.3 Trends of FT, Taxi-out Time, Airborne Time and Taxi-in Time Variance, 2004 to 2010	7
Figure 2.1 Scheduled Block Time (SBT) in the Context of Flight Time Decomposition. (Deshpande, V., Arikan, M., 2012.).....	9
Figure 2.2 Empirical CDF of Actual Block Time for Representative Flights under the Nine Scenarios	32
Figure 3.1 Taxi-out Time before and after Re-sequencing: JFK, 01/02/2011.....	47
Figure 3.2 Departure Queue Diagram before and after the Optimization for JFK, DFW and SFO, 6am-9am on 01/02/2011.	49
Figure 3.3 Taxi-out Time before and after Re-sequencing: SFO, 01/02/2011.	50
Figure 3.4 Taxi-out Time before and after Re-sequencing, with the Original and Updated Optimization: JFK, 01/02/2011.	53
Figure 3.5 Departure Queue Diagram for the Original and Updated Optimization for JFK, 2pm-4pm on 01/02/2011.....	56
Figure 3.6 Sensitivity Analysis Frontier: JFK and SFO	57

Acknowledgements

I would never have been able to complete my doctoral work at the University of California, Berkeley without the guidance, support and help I gained throughout my graduate study. I am grateful to those who dedicated their time and expertise for me along the journey, some of whom are acknowledged here.

My dissertation research was adeptly advised by Professor Mark Hansen, to whom I would like to express my deepest appreciation and gratitude. Lao Tzu once said, “To lead people, walk behind them”. During my time at Berkeley, Mark has guided my growth both professionally and personally by perceiving my struggles and obstacles from behind me, and always being supportive. Whenever I was frustrated with my research, he is the mentor who helped direct me onto the path to solutions with inspiring insights. Moreover, he is also the friend who understood my frustration and respected my space and my decisions. The many times walking out of Mark’s office feeling lightened and hopeful again after gaining research insights as well as advises for life were among the best moments of my graduate study.

I also would like to thank my other committee members, Professor Joan Walker and Professor Rhonda Righter for guiding my dissertation research. I am also grateful for the guidance and education I received from all Berkeley faculty members, whether from a course, a class project, an exam committee, or just casual conversations.

I have received considerable help along the way from my colleagues in the Berkeley transportation community and the research group NEXTOR. I derived great benefit from discussions with colleagues including David Lovell, Mike Ball, Yu Zhang, Amy Kim, Yoonjin Yoon, Bo Zou, Gurkaran Buxi, Yi Liu, Mogeng Yin and Lei Kang. Specifically, Megan Ryerson sets a role model for me as a researcher, an educator and a woman in engineering. Her insights contributed tremendously to this research, and her sympathetic advices for life helped me through struggles for the past several years. Michael Seelhorst was always generously sharing his time, expertise and SAS codes with me. When working with him, I am always inspired by his diligence and patience to achieve higher goal; when at leisure, his master-level piano playing has brought me so much joy. As a fellow Tsinghua undergraduate, Xiaofei Hu has always been the first person I turned to for sharing my happiness and sorrow (as well as countless firsthand gossips). Her calmness and humor have prominently helped me keep myself together through all the ups and downs these years.

I am also grateful for a group of close colleagues who generously exposed me to the aviation community and helped me ground my research in practicality. Frank Ketcham always has sharp operational insights about the results I derived from data and modeling. I would specifically like to thank Murray Auger from Delta Air Lines, and Monica Alcabin from the Boeing Company for providing me precious industry experiences. The VIP tour on the ground of the Boeing factory and the opportunity to fly an aircraft in the Delta pilot simulator are the perks of my graduate study that were out of my wildest imagination before I embarked on the journey.

During my PhD study, I am also thankful for the financial supports of the Federal Aviation Administration, NASA Ames, Delta Air Lines and the SafeTREC Research Center.

My parents have been extremely supportive of my decisions. I'm deeply grateful to them for never pressuring me to take any particular path and providing me both financial and emotional supports. Especially, I feel lucky to have my father, a mechanical engineer and a super hero in my heart, who fostered my love of engineering from an early age. I dedicate my dissertation to my parents for their unconditional support and love that enabled me to reach goals unachievable without them. Lastly, I would like to thank my best friend, Wenlong Zhang, who is the best thing happened to me at Berkeley. It was once said that being deeply loved by someone gives us strength, while loving someone gives us courage. Thank you Wenlong for all the strength and courage you have made possible, on which I leaned through every challenge I faced these years.

1. Introduction

1.1 Problem Statement

The idea of predictability—also referred to as reliability or (inversely) as variability—is not a new idea in the field of ground transportation and there is extensive research in that domain on predictability concepts, measurement, and valuation. In that literature, (un)reliability mainly refers to the unpredictable variations in travel time and is thus directly related to uncertainty of travel time (Carrion and Levinson, 2012). Operationally, reliability or predictability is inversely related to dispersion of travel times between individual OD pairs or on specific routes, metrics for which include variance, standard deviation, mean absolute deviation, and inter-quartile range, to name a few. A rapidly growing body of literature addresses measurement and valuation of travel time variability, and the goal of enhancing reliability seems to be an increasing priority for policy makers (Börjesson, M., Eliasson, J., Franklin, J.P., 2012). It is also now standard that transit operators regularly publish statistics on reliability (Börjesson, M., Eliasson, J., Franklin, J.P., 2012). In the realm of commercial air transportation, the percentage of flights arriving within 15 minutes of their scheduled times is tracked by DOT and widely published on online flight booking sites. Despite the resemblance between air transportation and other scheduled transportation modes, predictability is still a relatively new concept in the realm of air transportation. The Federal Aviation Administration (FAA), like most air navigation service providers, continuously seeks to better understand and address customer requirements, and improve the quality of service provided. Metrics for quality of the service has long been centered on delay. Thus reducing delay has been the major service quality objective. Recently, however, the concept of predictability has received more attention in service quality assessments. Therefore, understanding how to best measure predictability and how to assess the potential benefit of enhanced predictability makes a substantial contribution to the research society and will be the major focus of this body of research.

The majority of the literature on predictability in transportation assesses predictability by measuring variability in the “travel time,” which could be a road trip travel time, gate-to-gate time of a given flight, or taxi-out time of an aircraft on the airfield. There is a variety of variability measurements: difference between actual trip time and scheduled trip time (Kho et al., 2005), standard deviation of travel time distribution (Bates et al., 2001; Lomax et al., 2003; Ettema and Timmermans, 2006; Riikka and Paavilainen, 2010), standard deviation over the mean travel time (Taylor 1982; Lomax et al., 2003), difference between travel time percentiles (Bolczak et al., 1997; Ettema and Timmermans, 2006; Gulding et al., 2009) and the difference in expected and actual travel delays (Cohen and Southworth, 1999; Liu and Hansen, 2014). None of these studies explicitly consider the temporal aspect of predictability. In contrast, Ball et al. (2000) find that error in predicting flight departure time decreases as the departure of a flight approaches. They proposed a metric termed integrated predictive error that takes this effect into account.

Other studies are not concerned with predictability per se, but rather focus on methods for predicting travel time on the basis of the information available prior to the

commencement of the travel. These studies focus on road networks and are motivated by the increasing use of routing and navigation decision support tools (Borokhov et al., 2011). Linear regression, based on a combination of the current information—system variables—and historical travel time information, has served as one of the main methodologies (Kwon et al., 2000; Zhang and Rice, 2003; Rice and Zwet, 2004). Considering the importance of prediction timeliness in the application, the algorithms are usually designed to be simple, fast and scalable (Rice and Zwet, 2004). In these studies, the travel time predictions are intended to guide travel decisions but are not linked to performance measurement. In the broader literature on systems, a concept closely related to predictability is entropy. Entropy has been used to measure unpredictability of a set of possible events since its introduction into information theory by Shannon (1948). Studies have been carried out to validate the application of entropy analysis in stochastic processes (Cover and Thomas, 1991; Ciuperca and Girardin, 2005; Jacquet et al., 2008).

In the realm of air transportation, block time is analogous to travel time in ground transportation. Block time is the interval that commences when an aircraft moves under its own power for the purpose of flight and ends when the aircraft comes to rest after landing. Block times for specific flights—e.g. United 364 from San Francisco to Washington Dulles—vary from day to day. Therefore, metrics that capture the variation of flight block time over time would be an appropriate approach to measure predictability in the aviation system. It is also worth our attention that in commercial air transportation, there are various aspects where decisions made by service providers can be influenced by block time (un)predictability. The different processes of making these decisions might be driven by contradicting considerations underlying the service providers' overall objective of maximizing profitability. For example, on one hand the service providers (airlines) might want to operate at minimal cost, on the other hand they would be willing to spend more than the minimal cost to assure the quality of the service provided. These different considerations often co-exist in the decision making process but the trade-offs between them may vary for different kinds of decisions. Therefore attentions should be paid when quantifying predictability to see its impact on these different kinds of decisions in commercial air transportation.

There are two types of decisions to be investigated in this research: the strategic¹ and the operational airline decisions. We consider the strategic decisions as the setting of the scheduled travel time for each trip (the scheduled block time) that typically happened six months before the travel date. Choosing the scheduled block time is similar to travelers' choice of departure time when they have a preferred arrival time. Various researchers in ground transportation have shown that travel time reliability is a significant factor that affects traveler's departure time decision. Therefore, it is natural to assume an analogous relationship between scheduled block time and block time reliability. There are, however, few studies of how scheduled block time is decided and how the concept of predictability (reliability) is incorporated into this decision.

¹ "Strategic" decisions, as used here, refer to planning decisions made several months prior to the day of operation, as opposed to decisions made on the day of operation.

Scheduled block time is an important airline cost driver. Again compared to the ground transportation studies where the travel time reliability is found to have a strong effect on departure time scheduling, block time reliability is expected to be a significant factor in deciding scheduled block time. If the relationship can be understood, there may be opportunities for the FAA and other air navigation service providers to allocate resources to make block time scheduling more efficient, through improved predictability. For example, a clearer understanding of the link between block time variability (predictability) and scheduled block time might lead to the development of innovative air traffic management practices that will help improve predictability and thus allow shorter scheduled block times, while also furthering the FAA's goal of improving predictability.

On the operational side, we investigate the decision of the amount of fuel to be loaded to each flight in the daily operation. Unlike ground transportation, fuel loading contributes to the overall fuel consumption of a trip and thus to the airline cost. Flight dispatchers load contingency fuel in order to ensure that a flight can complete its mission without using any of its 45 minute reserve. The amount of contingency fuel loaded depends upon the dispatcher's perception of the risk of unexpected delays or reroutes, a part of which can be captured by flight time (un)predictability. Thus, the amount of fuel loaded may depend upon flight time variability. Extra boarded fuel results in additional weight of the aircraft and thus additional fuel burn cost.

This body of research aims to contribute to the understanding of how predictability affects airline decision making in various contexts and how these relationships can be used to save cost for air transportation service providers. In particular, we investigate two mechanisms through which improving predictability may reduce airline costs. The first is the scheduled block time effect. We conjecture that certain changes in the distribution of realized block times may reduce scheduled block times, and that the scheduled block time change may, in turn, reduce the airline scheduling cost. The next two chapters examine this, first by considering, in Chapter 2, the relationship between scheduled block time and flight time predictability and then, in Chapter 3, the air traffic management algorithms to adjust flight time predictability in order to save scheduled block time. The second mechanism is related to fuel loading. Here, we hypothesize that the distribution of realized block times affects the amount of contingency fuel airlines load onto flights, which in turn affects aircraft weight and fuel burn. This is the subject of Chapter 4. Chapter 5 offers conclusions and future research directions.

1.2 Recent Trend of Flight Predictability

Before investigating into the relationship between airline decisions and predictability, in this section a general trend of flight predictability for the past few years is presented. This helps better understand the historical trend of the metric of interest, as well as its relationship with other factors affecting aviation system performance, such as flight delay, flight on-time performance and on a larger scale, the overall economy.

1.2.1 Definition of Flight Predictability

We employ the Bureau of Transportation Statistics (BTS) Airline Service Quality Performance System (ASQP) database to characterize airline schedule and operations. This database contains detailed performance information for individual flights by major US air carriers between points within the United States. These flight records are aggregated to define flight predictability. The aggregation of flights is by specific airlines, flight numbers, origins, and destinations: e.g. AA 112 from ORD-LGA. The time unit for aggregation is quarter.

For each quarter, we assume that there is a uniform scheduled block time (denoted as SBT, hereinafter) for each individual flight, which is the elapsed time between the scheduled departure and the scheduled arrival. In the actual dataset, the condition where SBT is not uniform is rare. The median value of SBT in the quarter is used for such situations. The actual flight block time (denoted as FT) for each flight is the time from actual departure to actual arrival. FT has three components, taxi-out, airborne, and taxi-in. Another important component of the flight phase is the departure delay of a flight. Departure delay (or gate delay) refers to the elapsed time from scheduled departure time to actual departure time, the time that the aircraft leaves the gate. In some occasions a flight leaves the gate earlier than its scheduled departure time, and the departure delay for such flight is thus negative. To include the departure delay into total flight time, we also define effective flight time (EFT). EFT is defined as the time duration from scheduled departure time to actual arrival time. EFT can be decomposed into departure delay, taxi-out time, airborne time and taxi-in time.

In this section, flight predictability is defined as the variability of flight time for an individual flight, defined as above in terms of airline, OD, and flight number. The quarterly variance of EFT, FT, and the four components for each individual flight are calculated and used to measure flight predictability. For each quarter, the average variance across all the flights flown in that quarter is computed to represent the aggregated predictability level of that quarter. To conduct a thorough examination of the trends of flight time reliability over recent years, we collected the data from 2004 to 2010, quarterly. We consider flights flown on weekdays only, and, to guarantee robustness of our dataset, only the flights flown at least 50 times on weekdays in a quarter are considered. There are in total 28 quarters that we studied and roughly 10,000 qualifying individual flights that flew 50 or more times on weekdays in each quarter.

1.2.2 Trends in Flight Predictability

Variability in effective flight time (EFT), flight time (FT), and their components are all indicators for predictability. In this section, we use variance to measure variability. Let T_{ift} be the time for component $i \in \{departure\ delay, taxi\ out, airborne, taxi\ in\}$ of flight $f \in F$ on day $t \in T$ of a given quarter. The variance for component i for the quarter is defined as

$$VAR(T_{if}) = \frac{1}{|T|-1} \sum_t (T_{ift} - \bar{T}_{if})^2 \quad (1.1)$$

, where $\overline{T_{if}}$ is the average value for T_{if} over the $|T|$ days in the quarter. Since the effective flight time is the sum of the four components, we also have:

$$VAR(EFT_f) = \sum_i VAR(T_{if}) + \frac{1}{2} \sum_i \sum_{j \neq i} COV(T_{if}, T_{jf}) \quad (1.2)$$

, where the second term includes the covariances between the various flight components.

To summarize variability for a quarter, we take the average of the results for the individual flights. Thus, for example, $VAR(EFT) = \frac{1}{|F|} \sum_f VAR(EFT_f)$. Figure 1.1 shows the results after further averaging these values across all the 28 quarters in our data set. From Figure 1.1, we see that the average variance of effective flight time is slightly over 1000 min^2 , which corresponds to a standard deviation of just over 30 min. The variance of departure delay is the largest source of EFT variance and accounts for over 80% of the total EFT variance. The second largest source is taxi-out time variance, accounting for about 8%. Airborne time variance accounts for 5.5% while the variance of taxi-in time and covariances are very minor contributors to EFT variability. The dominance of departure delay as a contributor reflects the effects of delay propagation, airline internal factors, and air traffic management. Propagated delay from an upstream flight will often generate departure delay on a downstream leg with the same aircraft (this is called late aircraft delay in the departure delay). Delays can also propagate if flights are held for late connecting passengers or crews. Airline internal delays, including boarding delays and mechanical problems, also primarily manifested in departure delays. Finally, traffic management initiatives (TMI's), such as ground delay programs and airspace flow programs, are designed to shift delays from the air to the ground, and much of the delay resulting from TMI's takes the form of departure delay.

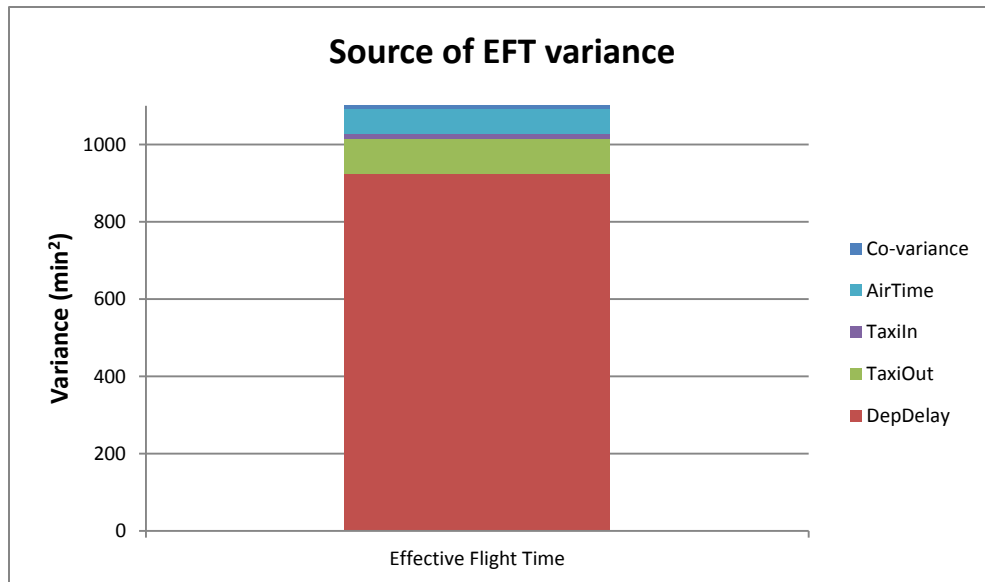


Figure 1.1 Sources of EFT Variance, Aggregated over 2004 to 2010

Taxi-out time is also a significant contributor to flight time variability. Much of the variability in taxi-out time is caused by departure queues, which can be affected by many unpredictable factors, such as traffic levels and weather conditions. Also, part of the departure delay effect is sometimes shifted from the gate to taxi-out time, because delayed flights may push back from the gate so that it can be reported that they departed “on-time.” Airborne time variability, the third most important source of EFT variability, is mainly the result of en route weather and wind conditions. On the other hand, taxi-in time is quite stable, and covariances between different EFT components are negligible sources of variation in effective flight time.

1.2.3 Trends of Predictability Performance

From the above section we gain a general sense of the magnitude and composition of EFT variability. We now focus on trends in variability. Figure 1.2 shows the trends of the quarterly variances of EFT, FT and departure delay from year 2004 to 2010. Firstly, we can see that the trends of EFT variance and departure delay variance are highly consistent. Just as Figure 1.1 shows that departure delay is the major source of EFT variability, Figure 1.2 reveals that it is the dominant driver of changes in variability over the years. The variance of FT is much smaller and seems to remain flat over the study period. EFT and departure delay variance increased starting in 2004 and peaked in the year 2007. The peak extended to 2008 and then experienced a sharp decrease. This trend is consistent with the air traffic volume across the years: 2007 was the busiest year in history and the subsequent economy recession caused a decrease in air transportation demand and flight traffic. It thus appears that flight predictability is highly correlated with air traffic volume and other performance metrics, such as delay. This is expected because unreliability tends to increase when the traffic is at higher level.

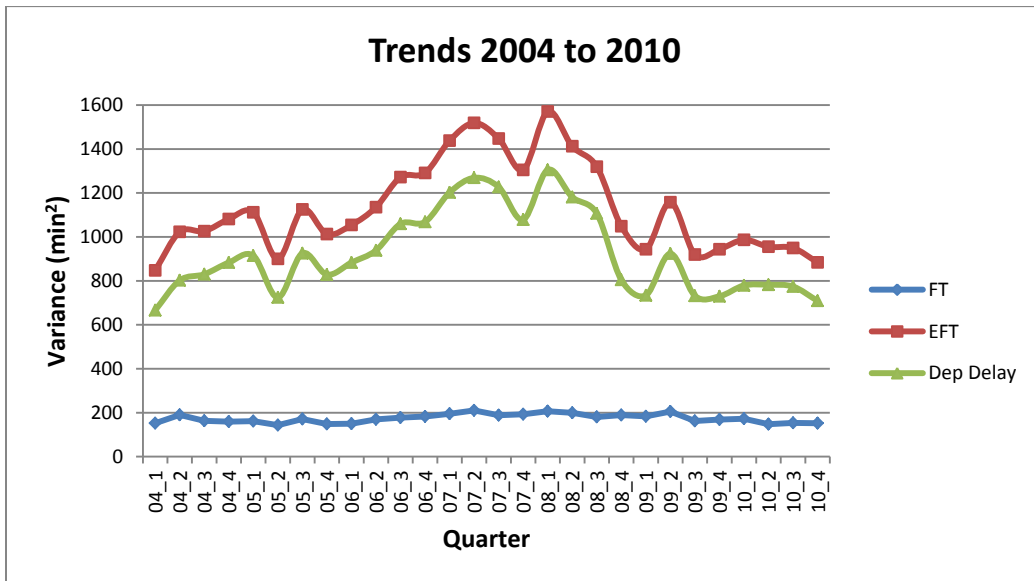


Figure 1.2 Trends of EFT, FT, Departure Delay Variances, 2004 to 2010.

Figure 1.3 provides a further decomposition of the variance. On a smaller scale, we more clearly observe the trends of FT variance and other components. The trend of FT variance is similar to that of EFT and departure delay in Figure 1.2, where the peak comes around 2007 and decreases occurred after 2008. Taxi-out time variance has a consistent trend with the FT, suggesting that taxi-out time is a major source of FT reliability and drives the trend to a large extent. On the other hand, the trend of taxi-in time remains flat over the years, validating our conclusion that taxi-in time is the most stable component of the four. The variance of airborne time exhibits a different pattern, changing periodically with a cycle of a year, peaking in the first and fourth quarters and taking the lowest value in the third quarter. Airborne time predictability is largely unaffected by the surges and declines in traffic that influence EFT and FT. The periodic pattern might be attributed to the jet stream phenomena. Jet streams are fast flowing, narrow air currents found in the atmospheres. They are caused by a combination of the earth's rotation and atmospheric heating. Within North America, the time needed to fly east across the continent can be decreased by about 30 minutes if an airplane can fly with the jet stream, or increased by more than that amount if it must fly west against it. The effect of jet stream on aircrafts is larger in cold air, so the disruption of airborne time appears to be higher in winter. Seasonal effects can also be observed in the trend of FT and taxi-out variability, although with a pattern of fluctuation that is nearly opposite, with variability higher in the summer. This is probably because summer is the busiest season leading to more crowded ground situation leading to more crowded ground situation. Also convective weather (weather caused by the current in the atmosphere induced by heated air, mainly thunderstorms) in the summer could generate severe ground delays, resulting in less reliable ground time.

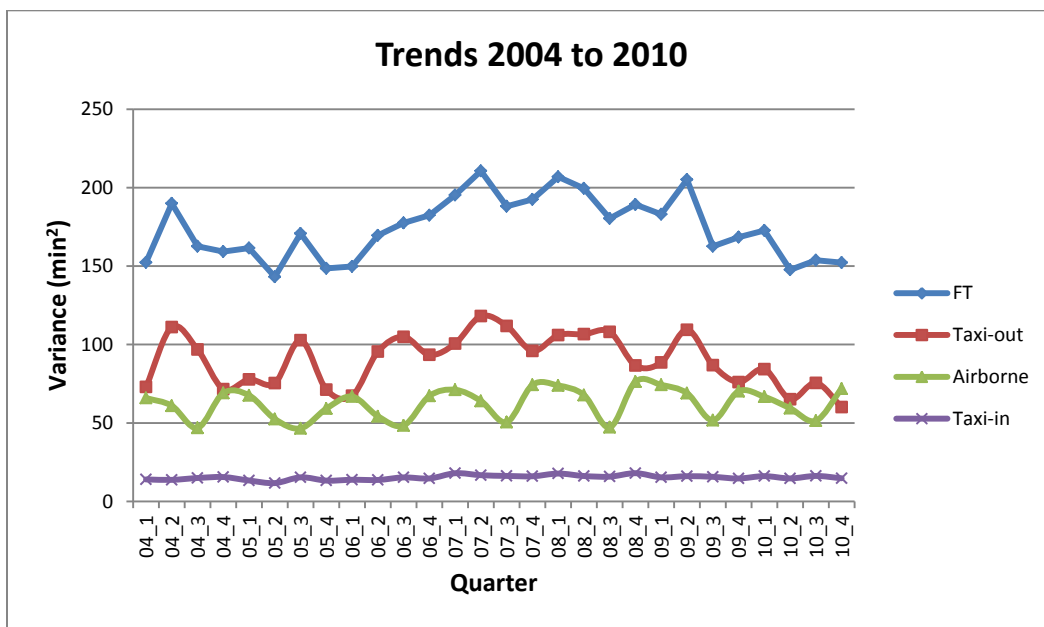


Figure 1.3 Trends of FT, Taxi-out Time, Airborne Time and Taxi-in Time Variance, 2004 to 2010

In summary, the concept of predictability in air transportation is measured in terms of the the variability of effective flight time and its various components in this section. Throughout this body of research, we will further investigate how such metrics capturing the variability of flight time affect different airline decisions. In this section, we also analyzed trends in predictability performance over the past few years by applying these metrics to a large set of US domestic flights. It is found that predictability performance has similar trends with traffic volume and flight delay. The time-based metric is further decomposed to see proportion of different contributions. We found that departure delay (gate delay) is the component that contributes the most to the variability of total effective flight time and it largely drives EFT performance.

2. Strategic Decision Modeling

In this chapter, the relationship between flight predictability and airline strategic decisions, i.e., the SBT decision is modeled using various modeling techniques. Figure 2.1 illustrates SBT in the context of flight time decomposition. SBT is the time duration between the scheduled (computer reservation system, or CRS) departure and scheduled arrival time. The actual block time (FT) is the time between actual departure and arrival time and varies from day to day for the same flight. The block time can be further decomposed into taxi-out, airborne and taxi-in time. The time between scheduled and actual departure time is defined as departure delay, or gate delay.

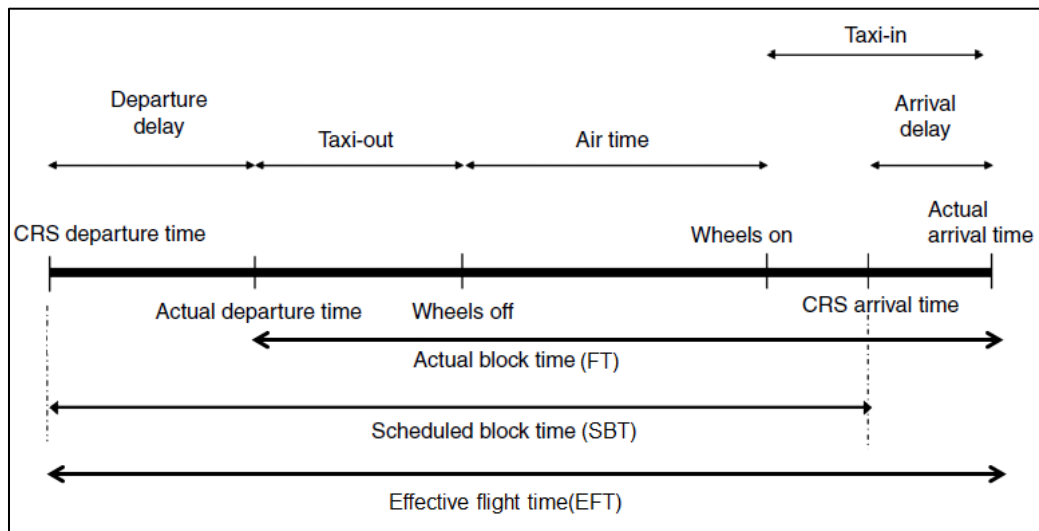


Figure 2.1 Scheduled Block Time (SBT) in the Context of Flight Time Decomposition. (Deshpande, V., Arikan, M., 2012.)

Airlines face a difficult set of trade-offs in setting SBT. They must balance their cost saving motive against their desire for good schedule adherence. The choice of SBT is somewhat similar to travelers' choice of departure time when they have a preferred arrival time, except the decision must be made much further in advance. While traveler choice is made based on considerations including time saving and the disutility of a late arrival, SBT setting is mainly driven by market forces and the profit motive. Various researches in ground transportation have shown that travel time variability significantly affects travelers' departure time decisions. Therefore, despite the difference in motivation, it is natural to assume an analogous relationship between SBT and block time variability in air transportation.

2.1 Literature Review

2.1.1 Travel Time Reliability in Ground Transportation

The research into surface travel time reliability has followed many branches. As mentioned above, setting SBTs is roughly analogous to urban travelers choosing departure time. When travellers have a preferred arrival time, such as work start time during the morning commute, they choose departure time based on prior knowledge of travel time on the pre-selected route. The time duration between the selected departure time and preferred arrival time is then fixed and serves as an implicit “scheduled” travel time. The actual travel time may deviate from this “scheduled” time from day to day due to traffic conditions, reflected by the variation in actual arrival time. This analogy is further developed in Table 2.1.

Table 2.1 Analogy of the Travel Time Reliability Concept between Ground Transportation and Air Transportation

Concept	Ground transportation	Air transportation
Decision	Departure time	Block time
Scheduled travel time	Preferred arrival time – Selected departure time	Scheduled block time
Actual travel time	Actual arrival time – Selected departure time	Actual block time
Prior knowledge	Historical travel times	Historical block times
Cost of earliness/excessive scheduled block times	Lost utility from reduced time at origin	Excess labor expense, reduced aircraft utilization
Costs of lateness/insufficient block times	Late penalty, work constraints	Degraded on-time performance, traveler inconvenience, delay propagation

In ground transportation, traveler costs due to early or late arrivals are assumed to influence departure time decisions. A traveler’s preferred arrival time (PAT) serves as a reference point that determines whether an arrival at a particular time is early or late. The classic example is the morning commute, in which the work start determines the PAT. Gaver is one of the earliest advocates for this approach. In (Gaver, 1968) a framework for

explaining variability in trip-scheduling decisions is introduced given a delay distribution and the costs of arriving early or late. Vickrey (1969) considered the tradeoff faced by travelers between queue delay and schedule delay of late or early arrivals. (Knight, 1974) and (Pells, 1981) posited the existence of a “safety margin”. Knight (1974) proposed the hypothesis that departure time for morning commute trips is chosen when the marginal utility of time spent at home is equal to the expected marginal utilities of arriving early to work and arriving late to work. Pells (1981) further added the need to minimize the frequency of late arrivals as an influential factor. Another important contribution is Small (1982). His theoretical model addresses departure time choice, taking into account workplace considerations (such as a policy linking departure times and working hours with merits or penalties to the wage rate) in the traveler utility function. The influence of workplace constraints on value of time is then assessed. The model proposed in Small (1982) is typically estimated using a discrete choice model based on the above utility function. To expand modeling traveler’s choices to include uncertainty, Noland and Small (1995) expressed this uncertainty in the form of a random variable with given probability density function.

$$E(U(t_d, T_r)) = \int_0^{\infty} U(t_d, T_r) f(T_r) dT_r = \gamma_1 E(T) + \gamma_2 E(SDE) + \gamma_3 E(SDL) + \gamma_4 P_L \quad (2.1)$$

In Equation 2.1, the utility function is determined by the choice of departure time (t_d) as well as travel time (T_r). Reliability is reflected by setting travel time as a random variable rather than a given value. The elements of the above equation are the scheduling cost for early (SDE) and late (SDL) arrivals. P_L is the probability of late arrival, which also reflects travel time unreliability because this probability is affected by the dispersion of travel time. The source of the travel time dispersion (or variability) is assumed to be non-recurrent congestion. Moreover, the dispersion may increase the tendency of early arrivals, thus high earliness costs can be incurred. This indicates that variability and scheduling costs are related. Noland and Small (1995) found that the uncertainty in travel time affects both the departure time choice and the expected costs. As uncertainty increases, travellers shift their departure time earlier to avoid late arrivals, corresponding to an increase in the analogous “Scheduled block time”. More recent work (Fosgerau and Karlstrom, 2010) proved mathematically the statement in Bates et al. (2001), that the term $\gamma_2 E(SDE) + \gamma_3 E(SDL)$ approximates the impact of standard deviation in the utility function.

2.1.2 Airline Scheduling Strategies

One attempt to predict SBT using historical data is by Coy (2006). A two-stage statistical model of airlines’ SBT is applied in the paper. Realized block time is found to be an effective predictor of SBT, having a parameter very close to 1. In addition, arrival times, airport utilizations, and poor weather conditions are found to be significant predictors of block time (Coy, 2006). The variability (inversely reliability) of block time, however, is not directly considered in this study. In fact, while the idea of making flight schedules more “robust”—immune to the disruptive impact of delays—has emerged in the past decade and applied in a wide range of scheduling decisions, there is little literature that takes predictability (reliability) into account in the analysis of SBTs. Airline schedule

development has always been one of the most challenging planning activities for airlines. A critical component of the schedule activity is the choice of SBTs, which depend on several factors. According to some airline schedulers, many airlines decide SBTs based on fixed percentiles of actual block time distributions built from historical data (Sohoni et al., 2011). Sohoni et al. argue, however, such techniques have not resulted in good on-time performance (OTP) of the schedule during operations. According to the U.S. Department of Transportation, a flight is on time if it arrives at its destination gate less than 15 minutes after its scheduled arrival time. The OTP is computed based on SBTs and airlines perceive their OTP as an important operational measure of their schedule reliability (Sohoni et al., 2011). However airlines face a key trade-off between increasing flight SBTs to improve schedule reliability and additional planned costs, so they often fail to adequately adjust block times and typically do not incorporate uncertainty in their planned schedule. Sohoni et al. define two service-level metrics for an airline schedule to incorporate reliability and develop a stochastic integer programming formulation to adjust existing schedule by changing departure time to maximize expected profit, while ensuring the two service levels.

Chiraphadhanakul and Barnhart (2013) focused on schedule slack, defined as the additional time allocated beyond the expected time required for each aircraft connection, passenger connection, or flight leg. Considering the complexity of robust scheduling, they studied how to more effectively utilize the existing slacks rather than simply having more slacks to achieve a more robust schedule. Slacks can absorb delay to keep the system more reliable, however at a very high cost per minute. They developed the concept of effective slack (the total aircraft/passenger slack after accounting for the historical arrival delay) with a certain upper bound, as an optimization objective.

Mayer and Sinai (2003) explored the factors affecting SBT using data on nearly 67 million flights between 1988 and 2000. They found that average SBT is almost exactly equal to the median block time excluding departure delay, notwithstanding that flights on average leave about 10 minutes after scheduled departure time. The (un)predictability of block time is in some way captured in their SBT model where certain percentile of the historical block time distribution other than merely the median (50th percentile) is incorporated as explanatory variable. However, their SBT model is behaviorally unrealistic since it relates the SBT decision in a given month, which is set several months beforehand, to realized flight times in the same month. In addition to an unrealistic temporal structure, the model is incorrect in assuming that SBTs are based on operating results for a single month, and in aggregating taxi-out and flight time components, which, as discussed below, airlines consider separately. Therefore, the need to understand the relationship between flight predictability, captured in historical block time performance, and airlines' SBT setting behavior with a deeper and more practical approach is still unsatisfied.

The impact of airline schedules is profound for both airlines and the FAA. The direct impact of SBT on flight delays would naturally influence airline on-time performance. Deshpande and Arikan (2012) analyzed empirical flight data published by the Bureau of Transportation Statistics to estimate the scheduled on-time arrival probability of each commercial domestic flight. To calculate scheduled on-time arrival probability, they used

both the DOT's 15-minute on-time metric and also defined their own on-time metric without the 15 minutes buffer. They claimed that definition for on-time performance is crucial and questioned the DOT's 15-minute on-time metric. The structural estimation approach from econometrics is then used to impute the ratio of leftover (overage) cost to shortage (underage) cost for each flight. Their results regarding the distribution of the cost ratio show that airlines systematically "underemphasize" flight delays, i.e., the implied flight delay costs are less than the implied costs of early arrivals for a large fraction of flights. This is interesting and different from what we have learned from ground transportation, while consistent with Sohoni et al. (2011), where they find from their conversations with airline planners that airlines tend to have a shorter SBT to save cost, willing to incur more delay and thus less on-time reliability.

Other research has considered the impact of SBT on airlines' cost. In Zou and Hansen (2012), the conventional delay characterization is extended into two distinct sets: delay-buffer and time based. These operational performance measures are then incorporated into the airline cost models, using an aggregate, statistical cost estimation approach. In the delay-buffer model, schedule buffer is defined to be the difference between the scheduled flight time and the 5th, 10th, and 20th percentiles of all observed flight time. Results from estimating a variety of delay-buffer models reveal that both delay and schedule buffer are important cost drivers. The coefficients suggest 0.6% increase in variable cost would occur if there is a 1-min increase in average delay against schedule or a similar increase in schedule buffer. Since a large portion of SBT is schedule buffer, the ability to reduce SBTs (without increasing delay against schedule) could thus lead to significant cost savings.

2.1.3 Industry Practice

Airline schedule planning is typically conducted in four sequential (and sometimes iterative) stages, namely schedule generation, fleet assignment, aircraft routing and crew pairing/rostering. The major focus in this chapter is the schedule generation process. This stage of work involves demand modeling, market forecasting and initial schedule establishment. Airlines often start route planning well ahead of operations to set up a preliminary timetable to generate maximum profits with limited resources such as aircraft fleets, capital investments and human resources (Wu, 2010). The schedule construction phase begins with the route system. Strategic development focuses on future schedules which may range from a few months to ten years depending on the air carriers' policies, while tactical strategies focus on short-term changes to the schedule and routes, sometimes on a daily basis (Bazargan, 2010).

We interviewed the block time setting group from a major US airline in order to gain a more detailed understanding of the SBT-setting process. The group confirmed that historical performance data is the primary basis for setting SBT. Schedulers at the airline categorize the data by quarter, origin-destination pair, departure time-of-day window, and aircraft type. The time window is based on the frequency of flights and is normally 15-30 minutes. After the historical data is grouped, the primary basis for choosing SBT for a flight is the Block Time Reliability (BTR). For commercial flights, the percentile at which the SBT lies among actual block times is reported as BTR. In other words, the BTR for a

certain flight group is the percentage of realized flights whose block time is shorter than or equal to its SBT. BTR is different from the DOT-reported on-time performance, which compares scheduled and actual arrival time, and is thus affected by gate delay at the origin airport. Also, the DOT metric, unlike BTR, counts a flight as “on-time” if it is less than 15 minutes late. DOT on-time performance is not specifically considered by the block time group, but is considered an important objective by a different part of the organization, the flight network group. The network group works with the SBT provided by the block time group and gives feedback for SBT adjustment if they feel on-time performance will be unsatisfactory. There are intensive discussions between the two groups and the adjustment is basically reflected in the choice of the target BTR.

The target BTR for this airline is normally in the 65%-75% range. Adjustments are made according to airport and flight characteristics as well as feedback from other internal groups. In the case of airlines operating hub-and-spoke networks, schedulers may set a lower BTR for flights into hub airports because these airports have high gate utilization, which makes early arrivals highly disruptive. The airline we interviewed sets a BTR as low as 65% for its major hub airport, in order to reduce early arrivals. Regarding the flight-specific characteristics, for long-haul flights, whose block time distributions tend to be more dispersed, the BTR for setting SBT is in general lower, in order to reduce average earliness. A frequent request from the network planning group is for the block time group to a lower SBT, both to be more competitive with other airlines and so that there can be longer scheduled turn times. Lastly, it is worth noting that when this airline sets SBT, it gives little consideration to gate delay, even though it affects on-time performance. This may be because historical gate delay is not considered predictive of future gate delay.

Besides the extensive interview with one specific airline, colleagues interviewed other airlines and an airline consultant regarding SBT setting behavior. The findings confirm that most carriers employ a similar target BTR-based SBT setting process. The range of percentiles is generally in 65%-75%. There are other variations regarding how much historical data to use, how to break the year into different season units to set SBT, and what factors are considered in deciding the specific target BTR for a given flight. After SBT is set, all the carriers interviewed monitor the on-time performance of flights and adjust SBTs when that performance is too low.

2.2 Mean-variance Model

As a first-step effort to capture predictability in historical block time and its relationship to SBT setting, a traditional mean-variance model is applied to capture both the centrality and dispersion of actual block time in this section. Variance or standard deviation of historical block time is used to define and measure flight (un)predictability. The underlying relationship between SBT and flight predictability (reliability) is empirically investigated using a statistical approach based on past operating experience of individual flights.

2.2.1 Model Specification

The mean-variance model intends to include the mean and variance (or standard deviation) together as factors that affect the setting of SBT. This approach is originated from the risk-return models in finance for decision making. In the context of ground transportation, the key idea is that not only travel time is a source of impact on travellers' choice, but also is travel time variability (or unreliability). In the literature, mean-variance is the usual name for the approach in transportation literature, despite the fact that various measures for travel time centrality and dispersion are actually used besides mean and variance, such as standard deviation.

Adopting the form of mean-variance model into the realm of block time scheduling in air transportation, we assume that both the centrality and dispersion of EFT have influence on SBT. To examine the impact of EFT deeper, in the model EFT is decomposed into its four components: departure delay, taxi-out time, airborne time and taxi-in time. Although in section 1.2 the flight predictability is measured by variance, to achieve a consistent unit in this model the standard deviation is used rather than the variance. The pattern of standard deviation of each component should be similar to the trends in section 1.2. Similarly, T^{ay}_{ift} is the time for component $i \in \{departure\ delay, taxi\ out, airborne, taxi\ in\}$ of flight $f \in F$ on day $t \in T$ of quarter $a \in \{1, 2, 3, 4\}$ in year $y \in Y$. The standard deviation for component i for quarter a in year y is defined as

$$sd(T^{ay}_{if}) = \sqrt{\frac{1}{|T^{ay}|} \sum_t (T^{ay}_{ift} - \overline{T^{ay}_{if}})^2} \quad (2.2)$$

Where $\overline{T^{ay}_{if}}$ is the average value for T^{ay}_{ift} over the $|T^{ay}|$ days in quarter a of year y . In the mean-variance model, both the average value and the standard deviation of each of the four components serve as explanatory variables for SBT setting. The formulation of the model is:

$$SBT^{a,y+1}_f = \sum_i \alpha_i \times \overline{T^{ay}_{if}} + \sum_i \beta_i \times sd(T^{ay}_{if}) + \sum_{a=2}^4 \gamma_a \times Q^y_a + const \quad (2.3)$$

Where $SBT^{a,y+1}_f$ is the scheduled block time of flight $f \in F$ in quarter a of the year $y+1$. Intercept is kept in the model. Also, to capture the seasonal variation, dummy variables Q^y_a are included besides the intercept and set to 1 if the variable is observed in quarter $a \in \{2, 3, 4\}$ in year $y \in Y$, and 0 otherwise.

In the formulation, we assume that schedulers set SBT for a flight with the knowledge of actual flight information of the same quarter in the previous year. This setting implies that schedulers focus on flight experience during the same season for which they are scheduling. In this section, the year 2009 and 2010 are chosen to be studied (i.e., $y=2009$). Thus SBT in 2010 is modeled with the actual flight data from the same quarter in 2009.

2.2.2 Estimation Results

Estimation results are shown in Table 2.2. From the R^2 results, the mean-variance model explains almost 100% of the variation in SBT. Consistent with our expectation, the average of each component of effective flight time has a positive influence on the SBT of the same quarter in the following year. However, the values of coefficients vary a lot. The coefficients imply that an additional 1 minute of average taxi-out time and airborne time will result in 1.05 minute and 1.03 minute in next year's SBT, respectively. 1 minute of actual taxi-in time will lead to a smaller value of 0.79 minute in SBT. The most interesting fact is that our results suggest that the influence of average actual departure delay is vanishingly small—only 0.01 minute per minute of average departure delay—as well as insignificant. On first glimpse, this result is quite surprising because departure delay is a major source of flight time variation in effective flight time. It appears that airlines ignore historical departure delay experience in setting SBTs. One reason may be that a large component of departure delay is late aircraft delay, and, since this depends on aircraft rotations that vary from quarter to quarter, the results of a previous quarter are assumed not to be predictive. This has been suggested in Deshpande and Arikani (2012). The fact that average taxi-out time, which is affected by some of the same factors as departure delay but not by delay propagation, has a much larger impact on SBT, provides some additional support for this theory.

Table 2.2 Estimation Results for the Mean-variance Model

Variable	Description	Estimate	SE	p-Value
Intercept		0.7280	0.206	0.0004
$\overline{D^{ay}_f}$	Mean departure delay	0.0119	0.008	0.1358
$\overline{TO^{ay}_f}$	Mean taxi-out time	1.0458	0.010	<0.0001
$\overline{A^{ay}_f}$	Mean airborne time	1.0314	0.001	<0.0001
$\overline{TI^{ay}_f}$	Mean taxi-in time	0.7921	0.024	<0.0001
$sd(D^{ay}_f)$	Standard deviation of departure delay	0.0313	0.006	<0.0001
$sd(TO^{ay}_f)$	Standard deviation of taxi-out time	-0.1599	0.012	<0.0001
$sd(A^{ay}_f)$	Standard deviation of airborne time	-0.2051	0.019	<0.0001
$sd(TI^{ay}_f)$	Standard deviation of taxi-in time	0.1209	0.031	0.0001

Q_2^y	Dummy variable for Quarter 2	-0.1986	0.111	0.0070
Q_3^y	Dummy variable for Quarter 3	-0.6625	0.112	<0.0001
Q_4^y	Dummy variable for Quarter 4	-0.0423	0.110	0.7004
<i>R</i> -square		0.9962		

Estimation results for standard deviations of the components are contrary to expectation. Capturing the unreliability of actual flight time, the standard deviations are expected to have a positive impact on SBT. However, it is found that the standard deviations of taxi-out time and airborne time, which are the major sources of block time unreliability, both have negative coefficients, suggesting that an increase in unreliability would reduce SBT, all else equal. On the other hand, the standard deviations of departure delay and taxi-in time have a positive coefficient, implying that 1 minute increase in the standard deviation will increase the SBT by 0.03 and 0.12 minute, respectively, for departure delay and taxi-in time. It is worth noting that mean departure delay and taxi-in time have a coefficient that is less than one, while mean taxi-out time and airborne time coefficients that are larger than 1. It appears that schedulers create buffers by adding fixed percentages to mean values for the major flight time components rather than directly taking into account standard deviations.

From these results, we can conclude that while SBT is highly influenced by historical average flight times, when these historical averages are pulled up as a result of high dispersion, the effect of dispersion is discounted. Put another way, given two flights with the same historical average flight time, but one with a greater flight time dispersion than the other, the flight with more dispersion will have a lower SBT, because the “far right tail” flights that are bringing up both the mean and the dispersion and not considered. Similar results are found in Mayer and Sinai (2003) regarding the impact of standard deviation of historical block time on SBT. Therefore, the distribution of block time must be characterized in a more detailed way than simple second-moment metrics if its impact on SBT setting is to be understood. To examine more closely how the “right tail” affects SBT setting, we developed the percentile model in the next section.

2.3 Percentile Model

As a measure of travel time variability in ground transportation, most studies have used either the standard deviation or the average delay relative to scheduled arrival time (Börjesson et al., 2012). However, both Mayer and Sinai (2003) and our mean-variance model in section 2.2 find that the standard deviation of actual block time reduces SBT. This is probably because airlines disregard extremely long actual flight times when setting SBTs to maintain competitiveness and efficiency. For the major airline we interviewed, the rule for SBT setting seems to be a specific BTR (block time reliability) target, i.e. a

certain percentile of the historical block time distribution. Thus, we developed a model with the percentile statistics of the actual flight time. The huge amount of historical data in the field of air transportation allows us to employ this approach to empirically investigate SBT setting behavior.

2.3.1 Data and Modeling

In this section, the relationship between block time distribution and SBT setting is modeled empirically, using multiple regression in order to understand the relationship between SBT and past operational experience. The variables capture the difference in percentile of historical block time; therefore the model is called the percentile model. The percentile model is a generalization of the BTR target model, and assumes that, because of the adjustments to the BTR that airlines make based on on-time performance, competition, and other factors, the SBT is influenced by more than a single percentile of the historical block time distribution. For the same reason, other variables than the historical block time distribution that might also affect the SBT decisions are also included in the model. It should also be noted that, in contrast to Deshpande and Arikan (2012), these models do not attempt to rationalize airline behavior by reference to a cost function for lateness and earliness. The structural estimation approach in Deshpande and Arikan (2012) links block time decisions to different factors that affect the relative cost of earliness and lateness. Our aim is to develop a model that comports more directly with stated airline practice which is expressed in terms of BTR rather than cost function minimization. Such a model can be used to predict how SBT will respond to changes in the distribution of realized block times, especially over the short term where model coefficients can be expected to be fairly stable.

The data on which the percentile model is estimated are collected from three sources: the Airline On-time Performance dataset, the air carrier statistics data from US DOT T-100 Domestic segment with U.S. carrier, Form 41 database, and the aircraft type information from a combined dataset including B43, OAG and FAA registry aircraft. We referred to three sources of aircraft type information to guarantee a master dataset that could cover most tail number (an identification number painted on an aircraft, frequently on the tail, that represents an aircraft registration number) information in the BTS dataset. The details about matching these multiple datasets will be discussed shortly. The first two datasets are both acquired from the Bureau of Transportation Statistics (BTS). We employ the Bureau of Transportation Statistics (BTS) Airline On-time Performance data to characterize airline schedule and operations. This database contains detailed performance information for individual flights by major US air carriers between points within the United States. These flight records are aggregated to capture the distribution of historical flight time. The aggregation of flights is by specific airline, origin-destination pair, 30-minute departure time window, and aircraft type. For instance, ATL BOS 20 B757 DL means the group of flights from ATL to BOS at departure time window between 10:00 to 10:30 am flying Boeing 757 by Delta Air Lines. While more cumbersome than simple flight number tracking as used in Mayer and Sinai (2003), this method allows is not dependent on airlines maintaining the same flight numbering from one time period to the next. The time unit for the aggregation is quarter. Each flight group is referred to as an individual flight in this research.

For each quarter, we assume that there is a single SBT for each individual flight, which is the elapsed time between the scheduled departure and the scheduled arrival time. However, since there are occasional variations within the quarter, the median value of SBT in the quarter is used as the dependent variable. The distribution of actual flight time is captured by calculating differences in percentiles of the historical flight time data. To separate the effect of different flight phases, we distinguish taxi-out time and the non-taxi-out time, which includes airborne and taxi-in time (hereafter this is called NTO time, for brevity). Also, because gate delay is expected to have a different effect than flight time, we include the difference in percentiles for gate delay separately. This differentiation by flight phase contrasts with Deshpande and Arikan (2012), which subtracts the delay caused by late arriving aircraft but otherwise combines the times in all three phases. The variables capturing the distribution of gate delay could include the effects from late arriving aircraft, as well as other factors. For individual flight f in day t , the 50th to 100th percentiles of the different components of block time and gate delay are calculated. The 50th percentile or median taxi-out time (NTO time, gate delay), denoted as $TO_{0.5}^{f,q,y}$ ($nonTO_{0.5}^{f,q,y}$, $dep_{0.5}^{f,q,y}$) of individual flight $f \in F$ in quarter q of year y are all included in the model. The variability of flight time is further captured by the differences between every 10th percentiles from 50th to 100th. For example, $dTO_{56}^{f,q,y} = TO_{0.6}^{f,q,y} - TO_{0.5}^{f,q,y}$ is the difference between the 50th and 60th percentile of taxi-out time for flight f . This approach depicts the distribution of components of flight time information in a manner that can represent the industry practice of BTR (block time reliability)-based block time setting found in our interview. The different segments of percentiles capture how SBT is influenced by successively rarer but higher realized flight time values, reflecting the reliability of historical flight time. While flight time percentiles are also considered in Mayer and Sinai (2003), there the actual percentiles rather than the differences between percentiles are used. This makes it difficult to interpret results in terms of how much emphasis carriers place on different regions of the flight time distribution.

Competition with other airlines flying the same market may motivate a shorter SBT so that the airline appears to offer faster service, or a longer SBT so that it appears more reliable. Therefore, we include variables that depict the OD pair competitiveness in the model. To capture competition for the OD pair, the Herfindahl index (also known as Herfindahl–Hirschman Index, or HHI) is applied. It is an economic concept widely applied in areas such as competition law, technology management (Liston-Hayes and Pilkington, 2004), and so on. It measures concentration in a market and is defined as the sum of the squares of the market shares of the 50 largest firms (or summed over all the firms if there are fewer than 50) within the industry, where the market shares are expressed as fractions. Increases in the HHI generally indicate a decrease in competition and an increase of market power. For the purpose of our analysis, the market share of a carrier in an OD pair can be expressed as the portion of number of seats provided in the total number of seats serving this market. For market od , the HHI can be calculated as:

$$HHI_{od} = \sum_{i=1}^N \left(\frac{s_{odi}}{s_{od}} \right)^2 \quad (2.4)$$

where S_{odi} is the number of seats provided by carrier i flying this OD pair, S_{od} is the total number of seats provided in this OD pair, and N is the number of carriers in this OD pair. Thus, in a market with two carriers that each provides 50 percent of seats, HHI equals $0.5^2 + 0.5^2 = 1/2$. A smaller HHI indicates a more competitive route. The US DOT T-100 database provides number of seats for domestic OD pairs and carriers to calculate the HHI.

Moreover, based on our interview with the industry, airlines give special consideration to their hubs, where they have the majority of the gates and the traffic. Most airlines adopt some variation of a hub-and-spoke system. Major carriers operate up to five hubs; while smaller ones typically have one hub located at the center of the region they serve (Bazargan, 2010). The airline we interviewed claims to set shorter SBTs for flights into their major hub airports to avoid early arrivals that can be highly disruptive to their operations. Therefore, in the percentile model we include dummy variables HUB_O and HUB_D that are airline and airport specific. They indicate whether the origin/destination airport is a major operation hub for the specific carrier, for each individual flight.

Lastly, air fare and load factor information of the individual flight might also affect the SBT decision. Airline schedulers might assign longer SBT for more expensive flights to ensure more reliable service, or shorter SBT to appear more attractive to high value passengers. Some industry observers also suggest that airlines set higher SBTs for flights with higher load factors. Thus load factor and fare variables are included in the percentile model as well. The information is merged into the major dataset by OD pair, carrier and quarter. The variables are denoted as LF_{od} and F_{od} .

In the formulation, we assume that schedulers set SBT for a flight with the knowledge of actual flight information and the HHI competition index of the same quarter in the previous two consecutive years. This implies that schedulers focus on flight experience during the same season for which they are scheduling. In this section, the years 2009, 2010 and 2011 are chosen to be studied, with the SBT in 2011 modeled based on the actual flight data from the same quarter in 2009 and 2010. The actual flight information in the two years are aggregated together to calculate the percentiles. The resulting model, with $y=2011$, and $h(y)$ indicating the two years prior to y , in this case years 2009 and 2010, is:

$$\begin{aligned}
SBT^{q,y}_f &= \alpha_1 \times TO_{0.5}^{f,q,h(y)} + \alpha_2 \times nonTO_{0.5}^{f,q,h(y)} + \alpha_3 \times dep_{0.5}^{f,q,h(y)} + \sum_{i=1}^5 \beta_i \times dTO_{i+4,i+5}^{f,q,h(y)} \\
&+ \sum_{i=1}^5 \gamma_i \times dnonTO_{i+4,i+5}^{f,q,y,h(y)} + \sum_{i=1}^5 \lambda_i \times ddep_{i+4,i+5}^{f,q,h(y)} + \gamma_1 \times HUB_O + \gamma_2 \times HUB_D + \chi \times HHI^{h(y)}_{od} \\
&+ \tau \times LF^{h(y)}_{od} + \omega \times F^{h(y)}_{od} + \pi \times Dist + \sum_{i=1}^3 \mu_i \times Quarter_i + const
\end{aligned} \tag{2.5}$$

The flight data are filtered to exclude Saturday flights, which is also based on the airline interview from which we learned that Saturday flights are treated separately. The BTS dataset contains 12,900,424 flights for year 2009 and 2010. Excluding Saturday flights

reduces it to 11,322,930 observations. Tail number (representing an aircraft registration number) is used to match the aircraft type information into the main database from the three databases mentioned above, starting with B43 dataset and using the other two datasets mentioned above if there was no match. For flights in the BTS dataset with missing tail numbers, the aircraft type information of all the flights with the same OD pair and same airline in the same month is used to fill the missing information. For cases where there were multiple aircraft types in the group, the aircraft type for the majority of flights is used to fill in the missing observations. With this dataset, the aggregation by OD pair, airline, time window, aircraft type, and quarter is conducted, generating 457,496 individual flights. The HHI variable is merged into the dataset identified by OD pair and quarter. The number of observation after this merging is 148,869. To assure robustness in the data, we only include the flights that are frequently flown in a quarter in the two years. Therefore, to be included in the data set, an individual flight must have been flown at least 50 times in a given quarter over 2009 and 2010, not including Saturdays. After this filter is applied, the estimation data set consists of 42,625 observations, each corresponding to an individual flight with a given departure time window, flying a given aircraft type, operated by a given airline, between a given origin and destination. Lastly, the load factor and fare information are merged into the dataset identified by OD pair, carrier and quarter information. This merging reduces the number of observations to 34,809 individual flights, which is the final dataset on which our analysis is performed.

2.3.2 Estimation Results

The estimation results on the whole dataset are shown in Table 2.3. The R^2 explains almost 100% of the variation in SBT. Distance is positively related to SBT, suggesting that there is more unpredictability in longer flights that is not reflected in the historical block time distribution and SBTs are set to be longer—at a rate of 4 min per 1000 miles—to take this into consideration. We can see that the coefficients for the gate delay distribution are all quite small and some are not significant. SBT increases 0.2 minutes for every 1 minute increase in median gate delay, but, surprisingly, decreases as the difference between the median and the 60th percentile increases. These results confirm that historical gate delay is not a strong consideration in setting SBT, but also suggest that it does have some effect.

The coefficient on median NTO time is 0.986, which is close to 1, indicating that this is a major determinant of SBT. The $d_{i,i+1}^{f,q,y}$ variables are intended to capture the variability of non-taxi-out flight time over the right tail of the distribution where it exceeds the median value. The 1-minute increase in the interval between 50th and 60th percentile generates a 0.48 minute increase in SBT. The coefficient decreases to 0.22 minutes for the interval between 70th and 80th percentile and to essentially 0 (-0.0089 minutes) for the far right end tail of the distribution. These results show that SBT is strongly affected by the left tail of the NTO flight time distribution (as reflected by the median), while the “inner right tail” has a moderate effect, whereas the effect of the outer right tail above the 70th percentile has a rather small effect. This is somewhat consistent with the airline practice described in section 2.1.3, in so far as airlines claim to choose SBT for a BTR target of around 70%. Thus, it is expected that more weight is put on the inner right tail (below 70th percentile) and the far right tail (above 70th percentile) is down-weighted. There are, however,

significant differences between these results and a “pure” BTR target model, as will be discussed below.

The pattern is similar for the taxi-out component of flight time, but the coefficients are somewhat smaller. For example, the median taxi-out time has a coefficient 0.709 (as compared to 0.986 for the NTO time). This is probably because, as indicated in our interview, airlines give more consideration to terminal and gate assignment changes and less to historical data in predicting taxi out times. The right tail of the distribution also has a rapidly decreasing pattern, from 0.5 for $d56$ to 0.0007 for $d90$. For the gate delay variables, 1-minute increase in median gate delay increases 0.24 minute in SBT. This impact is relatively small compared to the flight time variables. The right tail of the distribution has even smaller impacts that are generally insignificant.

The HHI variable has a negative coefficient. Higher HHI indicates lower competitiveness for the OD pair. Thus, a negative coefficient means that if the OD market is highly competitive, airlines will increase SBT. This suggests that competition drives airlines toward improving on-time performance instead of publishing a shorter SBT. Regarding the effect of airline hub airports, the dummy variable for origin hub airport is not quite significant, while the effect of destination hub airport is marginally significant, with a coefficient of -0.633. Schedulers thus tend to set a shorter SBT for flights from and into the airline’s hubs, more for destination. This is consistent with the airlines interviewees’ statement that they set shorter SBT for their hub airports in an effort to avoid early arrivals. However, the magnitude of this adjustment is quite small—about 0.5 minute.

The load factor is not a significant factor for setting SBT. The fare for the flight with the same OD pair and carrier in a quarter will affect SBT setting. Schedulers set a longer SBT for more expensive flights. A one dollar increase in the fare for the individual flight leads to 0.0046 minute increase in SBT, which is a quite small impact.

The percentile model represents airlines’ composite SBT-setting behavior, in a manner that explicitly shows the weight they place on different regions of the historical distribution of realized block times. Far less weight is put on the extreme right tail of the distribution compared to the median, or even the inner right tail. In essence, these results reveal that airlines plan for the “normal” scenario, not the few “extreme” historical cases with exceedingly long realized block times, lying around the far right tail of the distribution. Therefore, our results reveal the airlines’ emphasis for efficiency in their schedule while tolerating that a certain portion of flights will have longer block times than scheduled and will be delayed. To further interpret the results of the percentile model, two hypothetical models for the SBT setting process are shown in the last two columns in Table 2.3 to compare with our estimation results.

The first hypothetical model (termed HM1) assumes that the SBT is solely determined by the average historical block time. In a CDF plot, the area above the plot corresponds to the mean value of the variable. Now consider a model where the mean value of realized flight time solely determines SBT. In this hypothetical model the coefficient of mean flight time would be 1. Using the CDF plot, we can translate the mean flight time into an expression based on percentile differences. If we divide the plot into 50th, 60th... 100th percentiles and

assume the plot is piecewise linear between percentiles, then the mean value can be expressed as the sum of the areas above the CDF plot between each percentile line. For example, the area between 0 and 50th percentile value corresponds to the contribution to the mean of the median flight time value, and can be calculated using the percentile value as the area of a trapezoid. This can be repeated for each 10th percentile interval of the tail above the 50th percentile of the distribution. The specification for hypothetical model 1 thus becomes:

$$SBT = 0.75 \times Q_{0.5} + 0.45 \times d56 + 0.35 \times d67 + 0.25 \times d78 + 0.15 \times d89 + 0.05 \times d90 \quad (2.6)$$

Table 2.3 Estimation Results for the Percentile Model

	Percentile Model		HM 1	HM 2
<i>Variable</i>	<i>Estimate</i>	<i>p-Value</i>	<i>Coefficient</i>	<i>Coefficient</i>
Intercept	6.961	<.0001	-	-
<i>dist_{od}</i>	0.002	<.0001	-	-
<i>nonTO_{0.5}^{f,q,y}</i>	0.986	<.0001	0.75	1
<i>dnonTO₅₆^{f,q,y}</i>	0.481	<.0001	0.45	1
<i>dnonTO₆₇^{f,q,y}</i>	0.428	<.0001	0.3	1
<i>dnonTO₇₈^{f,q,y}</i>	0.221	<.0001	0.25	0
<i>dnonTO₈₉^{f,q,y}</i>	0.069	<.0001	0.15	0
<i>dnonTO₉₀^{f,q,y}</i>	-0.0089	<.0001	0.05	0
<i>TO_{0.5}^{f,q,y}</i>	0.709	<.0001	0.75	1
<i>dTO₅₆^{f,q,y}</i>	0.550	<.0001	0.45	1
<i>dTO₆₇^{f,q,y}</i>	0.432	<.0001	0.3	1
<i>dTO₇₈^{f,q,y}</i>	0.211	<.0001	0.25	0
<i>dTO₈₉^{f,q,y}</i>	0.077	<.0001	0.15	0
<i>dTO₉₀^{f,q,y}</i>	0.00066	0.5710	0.05	0

$dep_{0.5}^{f,q,y}$	0.235	<.0001	-	-
$ddep_{56}^{f,q,y}$	-0.102	<.0001	-	-
$ddep_{67}^{f,q,y}$	0.017	0.1018	-	-
$ddep_{78}^{f,q,y}$	-0.0094	0.1146	-	-
$ddep_{89}^{f,q,y}$	0.0019	0.4371	-	-
$ddep_{90}^{f,q,y}$	-0.00195	<.0001	-	-
$Quarter_1^y$	-0.904	<.0001	-	-
$Quarter_2^y$	-1.524	<.0001	-	-
$Quarter_3^y$	-0.591	<.0001	-	-
HHI_{od}	-0.633	<.0001	-	-
HUB_O	-0.373	<.0001	-	-
HUB_D	-0.484	<.0001	-	-
LF_{od}	-0.281	0.4103	-	-
F_{od}	0.0046	<.0001	-	-
R-square	0.9964			

Hypothetical model 2 (HM2) is a pure version of the airlines' BTR-based behavior. It assumes that SBT is equal to a certain percentile of the historical block time; here we assume the 70th percentile. Then the parameters of the median and the difference between 50th and 60th, 60th and 70th percentiles would be 1, since the sum of these variables is exactly the 70th percentile value, and the coefficients for the differences above 70th percentile would be 0, indicating that the airline doesn't consider the far right tail. The equation of HM2 is thus:

$$SBT = 1 \times Q_{0.5} + 1 \times d56 + 1 \times d67 + 0 \times d78 + 0 \times d89 + 0 \times d90 \quad (2.7)$$

Table 2.3 compares the results between the percentile model and the hypothetical models. HM1 only considers the mean value of flight time. In the estimated percentile model, the coefficient for the median NTO flight time ($nonTO_{0.5}$) is larger in the percentile model. The coefficients for the differences from the 50th to 100th percentile decrease at a faster

rate in the estimated model than they do for HM1. This clearly shows that, compared to HM1, SBT schedulers place more weight on the left side of the flight time distribution while down-weighting the far right tail, particularly above the 70th percentile. For the taxi-out component, the median value is closer to HM1. However, every interval above the 50th percentile has a smaller coefficient than under HM1. These findings are broadly consistent with previous literature where the implied flight delay costs are less than the implied costs of early arrivals for a large fraction of flights (Deshpande and Arikan, 2012). Put another way, airlines tend to be “optimistic” when they choose the SBT, tolerating longer delays in order to realize the advantages of shorter SBTs. However, our results go further in showing that airlines specifically discount delays associated with the roughly 20% of flight realizations with the longest durations, while paying some attention to the inner right tail for the block time distribution.

HM2 assumes SBT is solely based on the 70th percentile of actual block time and thus ignores flight times beyond these values. In the estimated percentile model, the coefficients for the median values are close to 1, as in this hypothetical model. In contrast to that model however, the inner right tail parameters are less than 1 and outer right tail parameters are greater than 0. Thus, compared to HM2, the estimated percentile model shifts weight from the inner right to the outer right tail. One interpretation of this is that the regression model, when estimated for a large diverse set of flights, captures a composite of different BTR standards. Thus, for the NTO component, 97% of flights have a standard at or above 50%, 44% have a standard at or above 60%, and so forth. However, it is also possible that the different regions of the block time distribution are indirectly taken into account through the various adjustments airlines make to the nominal BTR standard. This seems the more likely explanation for the small but significant influence of the far right tail, since we have heard no reports of airlines setting the BTR threshold at 80%, 90%, or 100%.

Returning to the comparison with the morning commute, we observe from these results that airlines are more willing to be late than most workers. While most workers would not want to be late 20% of the time, airlines pay little attention to block times over the 80th percentile. When doing this they accept that 20% of the flights would have block times longer than the schedule they set. In exchange for this, they reduce earliness and avoid the high costs of setting longer block times.

2.4 SBT Adjustment Model

The percentile model in the above section reveals that different parts of the historical block time distribution have different impacts on SBT setting. However, as discussed in section 2.1.3, SBT is sometimes adjusted based on other considerations. One common situation is that if a lower on-time performance is observed for a certain flight, there will be incentives for the schedulers to adjust SBT to improve on-time performance. The percentile model is designed to reflect the SBT setting in a static setting, thus the adjustment in response to on-time performance and various other factors is difficult to

capture. In this section, we model SBT adjustment to depict the relationship between SBT setting and various schedule adherence metrics, including on-time performance.

2.4.1 Model Specification

The SBT adjustment model analyzes changes in SBT for individual flights in a certain quarter between two consecutive years. We assume that the schedulers adjust SBT according to historical flight performance data. In other words, based on past performance, schedulers either increase or decrease the SBT, or leave it unchanged from the same quarter of the previous year. For two consecutive years y and $y+1$, the difference in SBT in quarter q for each flight $f \in F$ is denoted as: $\Delta SBT_f^q = SBT_f^{q,y+1} - SBT_f^{q,y}$ and is used as the dependent variable.

In our SBT adjustment model, the schedulers are assumed to base their judgment for a certain flight on the past on-time performance and the distribution of positive and negative deviation realized block times from the SBT. We consider two on-time performance metrics: A0 and A14. A14 is based on the DOT definition: a flight is on time if it arrives at its destination gate less than 15 minutes after its scheduled arrival time. Let $o_{f,i}^{q,y} = 1$ if the i^{th} realization of the flight $f \in F$ that flew a total of N times in quarter q of year y is on time, and 0 otherwise. Then the explanatory variable depicting the

quarterly on-time performance is: $A14_f^{q,y} = \frac{\sum_i o_{f,i}^{q,y}}{N^{q,y}}$. The A0 on-time performance is stricter than A14, counting a flight as on time only if it arrives no later than its scheduled arrival time. The calculation for A0 is similar to that of A14, and is denoted as $A0_f^{q,y}$. A0 and A14 are included in two separate models because we found that they are too strongly correlated to allow satisfactory results from including them in the same model.

In addition to on-time performance, positive and negative deviations between realized and scheduled block times may also lead to SBT adjustments. Thus, terms that characterized these deviations are included as explanatory variables. The negative deviation (ND) for the i^{th} realization of the flight $f \in F$ that flew in quarter q of year y is $\max(SBT_f^{q,y} - ABT_{f,i}^{q,y}, 0)$ where $ABT_{f,i}^{q,y}$ indicates the actual block time of this single realized flight. Similarly, the positive deviation (PD) is $\max(ABT_{f,i}^{q,y} - SBT_f^{q,y}, 0)$ for the i^{th} realization of flight f . Note that these deviations only reflect the block time, not gate delay. The latter is, however, reflected in the on-time performance variables A0 and A14.

A percentile-based approach similar to that used in the percentile model is applied to create the deviation variables used in this model. As discussed in section 2.3, the inner and outer right tails of historical distribution exert different impacts on SBT setting. Therefore, we calculated the median and two parts of the right tail distribution of the positive and negative deviations. The median ND (PD) for individual flight $f \in F$ is denoted as $ND_{0.5}^{q,y}(PD_{0.5}^{q,y})$ over the N realized flights in the quarter. The inner right tail is the difference between the 75th percentile and the median value. For negative deviation,

$ND_{inner}^{q,y} = ND_{0.75}^{q,y} - ND_{0.5}^{q,y}$, and similarly for positive deviation. The outer right tail is calculated as the difference between the 100th percentile and the 75th percentile of the distribution. For example, $ND_{outer}^{q,y} = ND_{1}^{q,y} - ND_{0.75}^{q,y}$ is the outer tail for negative deviation. These variables are calculated for each individual flight over a quarter and included as explanatory variables.

In addition, certain airport and OD pair characteristics might also affect the adjustment to SBT. Therefore, the HHI, HUB, load factor and fare variables defined in the percentile model are also matched and included in this SBT adjustment model. We estimated two separate models, each including one of the on-time variables, A0 or A14, but otherwise identical.

$$\begin{aligned}
\Delta SBT_f^q &= \alpha_1 \times A14^{q,y}_f + \beta_1 \times ND_{0.5}^{q,y}_f + \beta_2 \times ND_{inner}^{q,y}_f + \beta_3 \times ND_{outer}^{q,y}_f \\
&+ \gamma_1 \times PD_{0.5}^{q,y}_f + \gamma_2 \times PD_{inner}^{q,y}_f + \gamma_3 \times PD_{outer}^{q,y}_f + \sum_{i=1}^3 \mu_i \times Quarter_i \\
&+ \lambda_1 \times HUB_O + \lambda_2 \times HUB_D + \chi \times HHI_{od}^y + \tau \times LF^{h(y)}_{od} + \omega \times F^{h(y)}_{od} + \pi \times Dist + const
\end{aligned} \tag{2.8}$$

$$\begin{aligned}
\Delta SBT_f^q &= \alpha_1 \times A0^{q,y}_f + \beta_1 \times ND_{0.5}^{q,y}_f + \beta_2 \times ND_{inner}^{q,y}_f + \beta_3 \times ND_{outer}^{q,y}_f \\
&+ \gamma_1 \times PD_{0.5}^{q,y}_f + \gamma_2 \times PD_{inner}^{q,y}_f + \gamma_3 \times PD_{outer}^{q,y}_f + \sum_{i=1}^3 \mu_i \times Quarter_i \\
&+ \lambda_1 \times HUB_O + \lambda_2 \times HUB_D + \chi \times HHI_{od}^y + \tau \times LF^{h(y)}_{od} + \omega \times F^{h(y)}_{od} + \pi \times Dist + const
\end{aligned} \tag{2.9}$$

In the estimation of the SBT adjustment model, the flights are grouped in the same way as in section 2.3 for the percentile model to calculate the respective variables. Years 2009, 2010 and 2011 are again chosen to be studied. The difference between the SBT in 2010 and 2011 is modeled with the actual flight performance data from the same quarter in years 2009 and 2010, combined.

2.4.2 Estimation Results

Estimation results for the A0 and A14 SBT adjustment models are shown in Table 2.4. The estimation results from the two models are quite similar. The R²'s indicate that both models explain about 18% of the variation in the adjustment of SBT from year 2010 to year 2011. This is well below the R² for the percentile model presented earlier. While the root mean square error for both the SBT percentile and SBT adjustment models is about 4.7 minutes, the overall variance in SBT is much greater than that in SBT adjustment, leading to a larger R². From the results, we can see that the intercept is around 6 to 7 minutes. The coefficients for the on-time performance variables are -3.93 and -3.67 for A14 and A0, respectively, showing, as expected, that downward adjustments in SBT are associated with higher on-time performance. All else equal, a flight with perfect on-time performance, as measured by A0, would have an SBT adjustment 3.7 minutes lower than a flight that is always late. For A14, which is a more lenient on-time performance measure, this difference is a slightly larger 3.9 minutes.

Turning to the deviation variables, we find that the coefficients on negative deviation (ND) have stronger impacts than the coefficients on the corresponding positive deviation (PD) coefficients. For example, in the A0 model the coefficient on median ND is -0.318 and compared to a coefficient on the median PD of 0.278. Thus, as suggested in the earlier results, airlines react to SBT's in excess of realized times (negative deviations) more strongly than realized times in excess of SBTs (positive deviations). Additionally, we find that while the coefficients decrease as we move along the right tails of both the ND and PD distributions, they do so much more rapidly in the case of PD. Put another way, experience with unusually early flights is more likely to shift SBTs down than experience with unusually late flights to shift SBTs up.

Table 2.4 Estimation Results for the SBT Adjustment Model

A14 Model			A0 Model		
<i>Variable</i>	<i>Estimate</i>	<i>p-Value</i>	<i>Variable</i>	<i>Estimate</i>	<i>p-Value</i>
<i>Intercept</i>	7.015	<.0001	<i>Intercept</i>	6.064	<.0001
$A14^{q,y}_f$	-3.932	<.0001	$A0^{q,y}_f$	-3.672	<.0001
$ND_{0.5}^{q,y}_f$	-0.352	<.0001	$ND_{0.5}^{q,y}_f$	-0.318	<.0001
$ND_{inner}^{q,y}_f$	-0.136	<.0001	$ND_{inner}^{q,y}_f$	-0.132	<.0001
$ND_{outer}^{q,y}_f$	-0.020	0.0020	$ND_{outer}^{q,y}_f$	-0.021	0.0011
$PD_{0.5}^{q,y}_f$	0.304	<.0001	$PD_{0.5}^{q,y}_f$	0.278	<.0001
$PD_{inner}^{q,y}_f$	0.031	0.0229	$PD_{inner}^{q,y}_f$	0.0295	0.0297
$PD_{outer}^{q,y}_f$	-0.0024	0.0346	$PD_{outer}^{q,y}_f$	-0.0014	0.2051
$Quarter_1^y$	-1.984	<.0001	$Quarter_1^y$	-1.961	<.0001
$Quarter_2^y$	-2.608	<.0001	$Quarter_2^y$	-2.567	<.0001
$Quarter_3^y$	-1.461	<.0001	$Quarter_3^y$	-1.414	<.0001
HHI_{od}	1.081	<.0001	HHI_{od}	0.996	<.0001
HUB_O	-0.661	<.0001	HUB_O	-0.700	<.0001

<i>HUB_D</i>	-0.391	<.0001	<i>HUB_D</i>	-0.375	<.0001
<i>dist_{od}</i>	0.00047	<.0001	<i>dist_{od}</i>	0.00044	<.0001
<i>LF_{od}</i>	-1.323	0.0003	<i>LF_{od}</i>	-1.533	<.0001
<i>F_{od}</i>	0.0027	0.0001	<i>F_{od}</i>	0.0027	0.0001
<i>R-square</i>	0.1795		<i>R-square</i>	0.1813	

The intercept, combined with quarterly dummy variables, suggest a tendency to increase SBTs between 2010 and 2011. If negative and positive deviations were both zero, and on-time performance were 1, then the change in SBT under the A0 model would range 2.4 minutes for the fourth quarter to 0.4 minutes for the 2nd quarter. One possible explanation is that fuel prices were considerably higher in 2011, encouraging airlines to operate at a lower cost index (the ratio of time cost to fuel cost used by flight management systems to set economical speeds).

2.5 Impact Analysis

The percentile model in section 2.3 shows the impact of the distributions of historical block time on SBT setting. Different segments of the distribution have varying impacts on the SBT, with left and inner right tails of the distribution the most influential. In the real system block time distributions are constantly changing, and SBTs updated in response to the changing distributions, as well as other factors. It is of interest to observe these changes over a period of time, and in particular to observe the contributions of the changing distributions and SBT adjustments to changes in schedule adherence metrics. In this section, we perform such an analysis. Its aim is not to estimate a model, but simply to document how observed changes in SBT and the distribution of realized block time work together to change schedule adherence.

2.5.1 Methodology

We observe changes in actual block time distributions and SBTs for individual flights (as defined above) between two time epochs. For this comparison, year 2007 is of particular interest because it is an extremely busy year with a large amount of delay. It is a reasonable speculation that the highly dispersed block time, observed in section 1.2.3, will lead to significantly different behavior in SBT setting compared to years in which the system was less congested. It is also suspected that various flight performance metrics will also exhibit different patterns. Since our percentile model requires two years of data for setting SBT, we include year 2006 and 2007 as one group in the impact study. The data from year 2009 and 2010 are used as the other group for comparison.

For the two-year data set, the flights are aggregated in the same manner as before: by OD pair, departure time window, aircraft type, carrier and quarter; each distinct combination of these attributes is treated as an individual flight. Merging the two groups of two-year data together, there are 8,353 observations in total with at least 50 realizations in each period. From section 2.3, we learnt that the inner right tail of the block time distribution has the most impact on SBT setting, in addition to the median actual block time. Therefore, in a process similar to that described in section 2.4, we calculated the inner right tail of the historical block time distribution as the difference between the median and the 75th percentile as well as the median value. The average median block time decreased 0.66 minutes between 2006-2007 and 2009-2010; the standard deviation of this change was 4.92 minutes. For the inner right tail, the average change is -0.21 minutes, with a standard deviation of 2.84 minutes. The before and after time periods used for the impact study are close together, suggesting that there was little change in how block times are set between the two periods. They are also close to the time period used for estimating the percentile model, so that the behaviors shown in that model, such as the focus on the inner right tail of the flight time distribution, should also be in effect.

We want to compare the changes in block time distribution to changes in SBT. Therefore, the dataset is divided into nine separate “scenarios” where the median block time and the inner right tail could increase, decrease, or remain the same across these two time periods. If the change in median or inner right tail is less than one standard deviation from the mean change, then the flight is assigned to the “Average” group for that variable. If there is a change greater than one standard deviation above the mean change, then the observation is categorized as an “Increase”, and conversely for “Decrease”. (We use the terms “Increase” and “Decrease” somewhat loosely, since they are defined in terms of deviation from the average change; however, since the average change is close to zero these terms are appropriate.) Table 2.5 below lists the counts and frequency (in the brackets) for each of the nine scenarios in the 8,353 observations. Around 5,000—61%—of the observations fall into the “Average” category for both median and the inner right tail of the block time distribution. However, there are still reasonably large counts for each scenario. Notably, this is the case even for scenarios involving an increase in one metric and a decrease in the other. The margins of Table 2.5 show the marginal counts and frequencies.

Table 2.5 Summary of Nine Scenarios and Their Counts and Frequency in the Dataset

		Median BT			
		<i>Increase</i>	<i>Average</i>	<i>Decrease</i>	<i>Total</i>
Inner Right Tail	<i>Increase</i>	226 (0.027)	598 (0.072)	142 (0.017)	966 (0.116)

of BT	<i>Average</i>	657 (0.079)	5125 (0.614)	733 (0.088)	6515 (0.781)
	<i>Decrease</i>	88 (0.011)	521 (0.062)	263 (0.031)	872 (0.104)
	<i>Total</i>	971 (0.117)	6244 (0.748)	1138 (0.136)	8353 (1.00)

For each scenario, we are interested in the changes in SBT and various schedule adherence metrics in the years immediately after 2006-07 and 2009-10: i.e. the years 2008 and 2011. Firstly we study the changes in SBT. For each scenario in both years, the median SBT for each individual flight is calculated, as defined in section 2.3 and the average SBT for each scenario is recorded and compared for the two years. Moreover, the on-time performance metrics—A0 and A14—and average ND and PD of the flights for each scenario are also calculated and compared. Lastly, we investigate how the change in SBT affects these metrics, by calculating their values under the counter-factual scenario in which the SBT in 2011 is the same as that in 2008.

2.5.2 Results

Table 2.6 shows the impact analysis results. The nine scenarios are numbered as 1 to 9 from the upper left cell horizontally to the bottom right cell in Table 2.5. The second major column in Table 2.6 presents the results from comparing SBTs in year 2008 and year 2011—the years immediately following the two-year periods considered in Table 2.5. The upper half shows the average SBT values, where the bottom part shows the average change in SBT and its standard deviation. The largest increase (decrease) of SBT happens where both median and inner right tail of previous years’ actual block time increased (decreased). Median block time change is clearly the major determinant of whether SBT increases or decreases. However, the effect of the inner right tail is also significant. Comparing scenario 1 and 7, where median BT increased and the inner right tail increased and decreased respectively, the change in SBT has a 3.5 minutes greater in the former case. These differences are 3.3 and 3.2 minutes when we make similar comparisons for the scenarios with the same median BT (scenarios 2 and 8) and a reduced median BT (scenarios 3 and 9). Differences among the various median scenarios for a given inner right tail scenario are also fairly consistent—around 9 minutes. These results further validate the finding that—contrary to previous studies such as Mayer and Sinai (2003)—the inner right tail of the distribution matters in SBT setting. Finally, in scenario 5, where both the median and the inner right tail are in the “average” category, average SBT decreases 0.35 minutes. This is in line with the small reductions in the average median and average inner right tail between the two periods.

To further illustrate the changes in actual block time distribution and the change in SBT, one representative individual flight is picked from each scenario and the empirical CDF of historical block time in year 2006-2007 and 2009-2010 is plotted in Figure 2.2. The vertical lines in each graph denote the SBT for the earlier year (2008—dashed line), and the newer year (2011—solid line). The graph shows patterns similar to those found Table 2.6. The middle column is especially interesting as it illustrates scenarios where median BT does not change while the inner right tail varies. In the top graph where inner right tail increased, the SBT also increased, and conversely for the graph in the middle bottom.

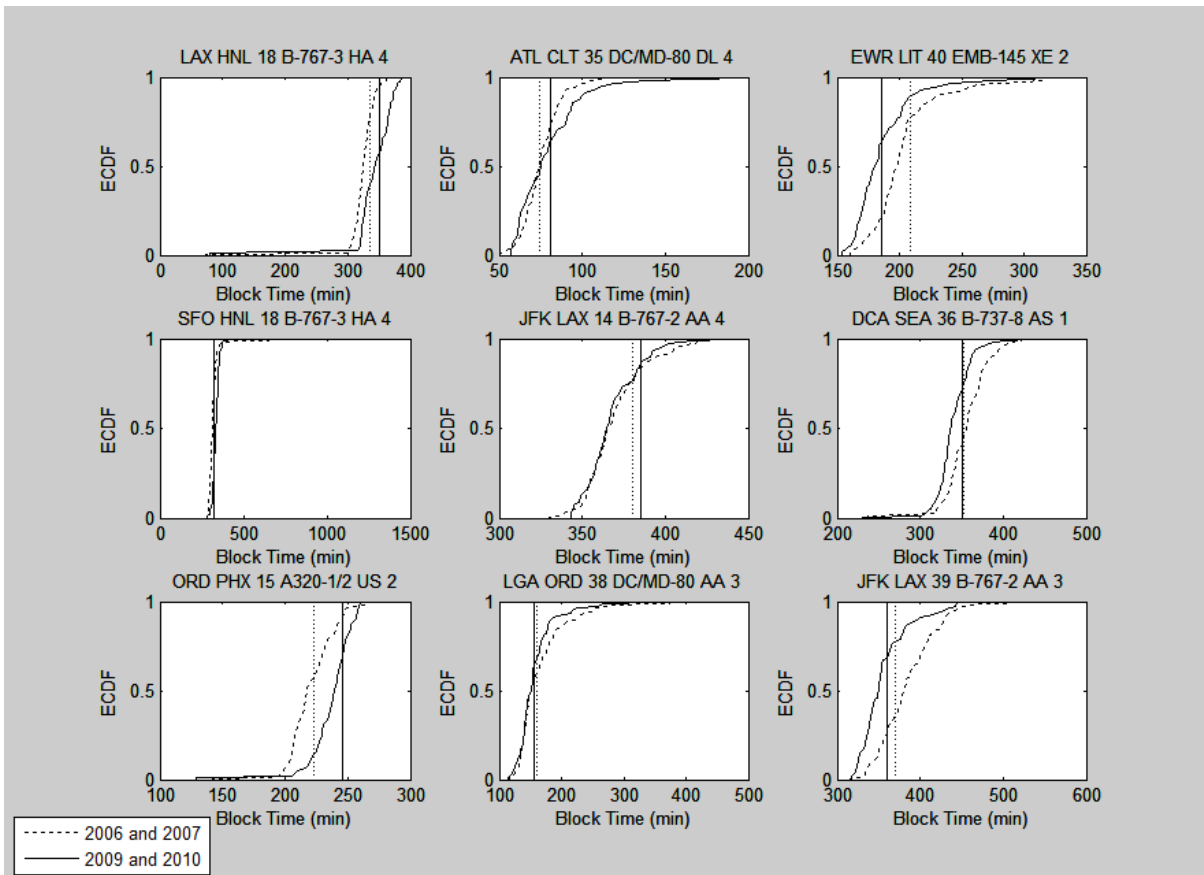


Figure 2.2 Empirical CDF of Actual Block Time for Representative Flights under the Nine Scenarios

Regarding the performance metric results in Table 2.6 below, we note first that there is overall improvement in on-time performance and reduction in average PD between 2008 and 2011. Average ND also increased for most scenarios. These overall results derive from changes in block time distributions (and in the case of the on-time metrics, gate delay) combined with changes in SBT. The impact of the latter is isolated by comparing the 2011 results with the 2011' results, which show what the 2011 performance would have been if SBTs had not changed from their 2008 values. We see that changes in SBT have sizable impacts. For example in Scenario 1—median increases and right tail increases—a large increase in on-time performance between 2008 and 2011, as well as a large reduction in

average PD (and increase in average ND) are mainly due to a 6 minute increase in SBT. On the other hand, in Scenario 9—median and right tail decrease—reductions in SBT virtually eliminate what would otherwise be substantial increases in on-time performance and decreases in average PD. More generally, the magnitudes of differences in on-time performance between 2011 and 2011', which reflect only the impact of SBT change, are comparable to the magnitudes of the overall differences 2008 and 2011. In the case of average PD, changes resulting from SBT adjustment are of somewhat larger magnitude than the overall changes observed. In sum, changes in SBT are as large or larger a driver of change in schedule adherence between 2008 and 2011 as changes in the underlying operational performance.

Table 2.6 Changes in SBT and Various Performance Metrics for Different Scenarios, 2008 vs. 2011

Scenario	Scenario Description	SBT (min)		A0			A14		
		2008	2011	2008	2011	2011'	2008	2011	2011'
1	Med +, IRTail +	150.591	156.425	0.526	0.679	0.557	0.762	0.842	0.803
2	Med avg., IRTail +	153.834	155.107	0.546	0.631	0.598	0.776	0.822	0.820
3	Med -, IRTail +	218.986	214.877	0.517	0.599	0.675	0.716	0.794	0.828
4	Med +, IRTail avg.	150.112	153.750	0.550	0.669	0.571	0.797	0.857	0.832
5	Med avg., IRTail avg.	118.388	118.040	0.593	0.632	0.638	0.818	0.845	0.853
6	Med -, IRTail same	181.443	176.533	0.554	0.540	0.647	0.757	0.777	0.828
7	Med +, IRTail -	199.307	201.574	0.509	0.657	0.584	0.757	0.851	0.817
8	Med avg., IRTail -	158.923	156.873	0.539	0.581	0.621	0.742	0.788	0.810
9	Med -, IRTail -	184.304	176.956	0.491	0.511	0.620	0.668	0.705	0.763
		SBT (min)	Change	ND (min)			PD (min)		

Scenario	Scenario Description	<i>mean</i>	<i>s.t.d</i>	2008	2011	2011'	2008	2011	2011'
1	Med +, IRTail +	5.834	6.323	4.939	8.966	5.414	6.397	3.579	5.861
2	Med avg., IRTail +	1.273	6.429	5.456	7.322	6.624	5.069	3.642	4.218
3	Med -, IRTail +	-4.109	7.968	8.313	8.871	11.833	6.139	4.524	3.377
4	Med +, IRTail avg.	3.638	6.513	4.654	7.186	5.148	5.336	3.219	4.818
5	Med avg., IRTail avg.	-0.348	5.113	4.722	5.293	5.595	3.554	2.870	2.825
6	Med -, IRTail avg.	-4.909	6.637	7.596	6.352	9.879	4.343	4.312	2.930
7	Med +, IRTail -	2.267	7.995	6.817	9.132	7.933	6.250	3.744	4.812
8	Med avg., IRTail -	-2.050	6.733	6.739	7.196	8.704	5.078	4.050	3.508
9	Med -, IRTail -	-7.348	8.024	9.396	7.629	13.219	5.827	5.030	3.272

2.6 Conclusion

In this chapter, the impact of predictability on airlines' strategic decision—scheduled block times setting—is studied by means of econometric models that relate scheduled block times to previous operational experience, as well as through validation with the real-world historical data. We developed three types of multiple regression models to incorporate the effect of predictability, each with different definitions and metrics for predictability. The first type is the mean-variance model and our results indicate that departure delay has little influence on scheduled block-time, although it is a major contributor to effective flight time in real operations. Schedulers do not consider historical departure delay information, perhaps because they believe that historical departure delay information does not reliably predict the future. Also, in the mean-variance model, the effect of flight time variability turns out to be negative. This is contrary to expectation based on the urban trip scheduling literature. It appears that block time schedulers account for variability by adding a certain percentage to the mean block time rather than considering historical block time variance/standard deviation (one metric for flight predictability) directly.

The second model is called the percentile model and is intended to capture flight predictability in a different approach. According to one airline, SBT is set using a BTR (block time reliability, the percentile at which the SBT lies among the actual block times) target. We developed the percentile model in order to capture airlines' BTR-based practice. The variability in block time is captured by increments between every 10th percentile above the median (50th percentile). This enables us to observe how different regions of the historical block time distribution are considered in SBT setting. The different components of the block time, i.e., the taxi-out and non-taxi-out phase of the flight are also separately treated in the model. Other variables, such as gate delay, distance, airport hub status, and market concentration are also included in the model. The estimation results of the percentile model suggest that the entire right tail of the block time distribution is considered when setting SBT, but that the inner right tail up to 70th percentile receives by far the most consideration. In general, airlines are willing to experience occasional severe delays in exchange for a shorter SBT. Other notable results include that again historical gate delay is virtually ignored, that historical taxi-out time is given somewhat less weight in the SBT setting compared to the non-taxi-out component of flight, that airlines with hubs tend to set shorter SBTs for their hub-bound flights, and that competition encourages longer SBTs. A comparison between the percentile model and an assumed model that only considered the mean flight time value suggests that airlines tend to put more weight on earliness than lateness as they decide their scheduled block time. Thus it suggests that airlines are willing to experience delays to realize the efficiency benefits of scheduling shorter SBTs.

The third econometric model developed in this chapter is the SBT adjustment model, which captures the impact of on-time performance that is not reflected in the percentile model. We also analyze cross-year adjustments in SBTs based on the distributions of earliness and lateness relative to schedule, in addition to on-time performance. This model models the adjustment of SBT's across year and reveals that airlines adjust their SBTs for individual flights in response to their historical on-time performance. It also reveals that SBTs are adjusted to mitigate persistent earliness and lateness, but that the earliness (negative deviation) effect is stronger.

Finally, to investigate changes in block time distributions and associated scheduled block times that actually result from a change in operating conditions in the NAS, we compare the high traffic period around 2007 with the period of curtailed traffic around 2010 to analyze the impact of the historical distributions of actual block times on SBT. Real data are used to demonstrate that when changes in block time distributions—in particular the median and inner right tail—occur, significant adjustments in SBTs often result, and that the impacts of these adjustments on schedule adherence is great or greater than the changes in the underlying operational performance. With the efforts to fully understand the complete cycle of SBT setting through estimating the three behavioral models, this piece of analysis explicitly links the relationship between the changes in actual block time distribution and SBT. This fills a missing piece of the NEXTGEN benefit analysis by relating changes in flight time distributions with changes in SBT.

This body of research is critical to assessing the consequences of improvements to the National Airspace System (NAS), such as those contemplated under NEXTGEN. Such

improvements will affect the distributions of realized block times for individual flights. This change in distribution may affect both the SBT and deviations of actual block times from the SBT. It is important to consider both effects, since they have different economic implications. Changes in SBT influence a host of related costs including crew time and aircraft ownership, as well as the earliness and lateness of flights relative to the schedule. In current practice, however, the impact of a NAS improvement on SBT is not explicitly considered. In essence, it is assumed that any reduction in realized block time has the same economic value regardless of its impact on SBT. In the future, NEXTGEN is expected to be a major source of change in operational performance, and therefore SBTs, and finally of deviations between scheduled and realized times. It is clear from our results that knowledge of the change in average block times is not sufficient to understand these impacts, since a given change in the average can arise from many different changes in the distribution. This suggests that business cases for NAS improvements should pay more attention to impacts on the distribution of block times, instead of the average. Broadly speaking, improvements that push in the far right tail of the distribution will affect delays and on-time performance but not the SBT, while improvements that shift the inner right tail will effect scheduled block times but have limited impact on on-time performance. There is benefit from either change, but the nature of the benefit is fundamentally different, and it is important that NEXTGEN business cases recognize this.

Beyond this specific focus, the study in this chapter provides a perspective on how the phenomenon of transport system reliability is manifested in the specific mode of scheduled air transport. As suggested above, setting SBT is somewhat analogous to scheduling the morning commute. However, there are important differences because the SBT must be set well in advance, and also in the perceived penalties of earliness and lateness. As we shall see, these differences cause airlines to focus on a particular part of the block time distribution when setting SBT. The innovative methodology required to reveal this behavior is a further contribution of our work in this chapter.

3. Improving Predictability through Airport Departure Queue Sequencing

In section 2.3, through estimating the percentile model we have found a relationship between SBT and the distribution of realized times, by phase of flight. In this chapter, we use those estimation results and seek a practical approach to achieve the benefit of predictability, in the form of reduced SBT. More specifically, we consider how departure queue sequencing at the airport surface can be used to change the block time (especially taxi-out time) distribution for flights. Through this day-of-operations practice, the benefit of the adjustment would be reflected on savings in future SBT. In addition, we also investigate the trade-off between block time (taxi-out time) adjustment and flight on-time performance in the optimization process.

3.1 Departure Queue Management

The need for a more automated system to better manage and improve the operation of aircraft departure queues at the airport has been recognized for a long time. All major airports have a Ramp Control or Ground Control procedure for the management of separation of all surface movement on airport taxiways, inactive runways, holding areas, transitional aprons, and intersections (ADFMS report). However, the procedures to establish a departure queue are usually on a “first-come, first-serve” basis, whose major goal is to ensure separation and safety, but not to improve efficiency (ADFMS report). One main cause of inefficiency in this domain is that, under high traffic conditions, multiple aircraft pushback at around the same time and contest for the runway. This leads to many aircraft taxiing to the runway simultaneously and long runway queues as well as congestion effects on taxiways. (Liu et al., 2014)

One methodology to address this problem is departure metering (Malik et al., 2010; Brinton et al., 2011; Nakahara et al., 2011; Simaiakis et al., 2011; Simaiakis, 2013). One approach to achieve departure metering is N-control, which controls the queue length by metering pushbacks from the gate, so as to maintain but not exceed levels of airfield occupancy that allow for efficient runway utilization (Simaiakis et al., 2011; Simaiakis, 2013). The suggested pushback rates generated by the N-control algorithm are provided to ground controllers, adding to their current responsibility of maintaining separation and a smooth flow of aircraft on taxiways. A second approach is called Collaborative Departure Queue Management (CDMQ). This approach assigns flight operators taxiway entry slots according to ration-by-schedule principle (Brinton et al., 2011), to manage the length of the runway queue. A method similar to CDQM was developed and implemented at JFK airport in 2010. More recently, the method called Airport Departure flow Management System (ADFMS) has also been developed and proposed to be implemented at PHL airport. ADFMS achieves departure queue sequencing through implementing two important functions: Departure Slot Scheduling and Departure Queue Management. The former meters demand to match departure capacity and enables airlines to trade departure

slots among themselves. The latter divides the departure queue into physical and virtual components, assigning expected pushback times that allow aircraft to remain at their gates as long as possible while still meeting their assigned departure slot times.

All the methods mentioned above manage runway queue length by controlling flights in the ramp area only. Another surface traffic management system, known as the Spot And Runway Departure Advisor (SARDA), extends automated decision support to other areas including taxiways, queue areas, and the runway (Liu et al., 2014). Specifically, SARDA advises on actual pushback time, sequence and timing for spot release, sequence for take-offs, and sequence for active runway crossings (Jung et al., 2010; Malik et al., 2010; Gupta et al., 2012; Hoang et al., 2011). Unlike the tools mentioned above, the SARDA advisories are provided to both ground and local controllers in the control tower. In such cases, the local controller is responsible to keep the runway operations, including take-off, landing and runway crossings, both safe and efficient by providing separations between aircrafts and safely sequencing departures and arrivals. The ground controllers are responsible for the ground movement of aircraft taxiing or vehicles operating on taxiways or inactive runways and preventing runway incursion is their primary responsibility.

In the existing literature on performance evaluation of these automation technologies, attention has been focused on throughput increases, delay reductions, and fuel savings (Simaiakis and Balakrishnan, 2009; Simaiakis et al., 2011; Nakahara and Reynolds, 2012; Gupta et al., 2013; Hao et al., 2015). In addition to efficiency concerns, emission reduction is another major objective because aircraft taxiing on the surface contribute significantly to the fuel burn and emissions at airports (Simaiakis and Balakrishnan, 2009).

More recent work has proposed predictability metrics to evaluate runway operations (Liu et al., 2014). Unpredictability is measured as the integration over time t of standard deviation of remaining taxi-out time at time point t over time. Results from a human-in-the-loop simulation show that SARDA is able to reduce unpredictability of taxi-out time through improving surface operations, by 46% and 39% under high and medium traffic levels, respectively (Liu et al., 2014). Another study used the term “Flexibility” as a key performance metric, which essentially reflects the predictability of the system (Wojcik et al., 2013). Over a long period of time, predictability can be measured with long-term statistics of how well the originally intended schedules of flights were met in actual operations (Wojcik et al., 2013). Flexibility is provided to operators in their decision making process to mitigate disruptions of the schedule. Thus, increased flexibility may be used to limit the unpredictability, especially under unfavorable operating conditions. Through the use of virtual queuing (VQ) in departure operations, operators are provided with additional flexibility in prioritizing flights for departure and allowing flights with more expensive delays to skip ahead in the departure queue. Flexibility metrics derived from delay recovered with VQ relative to physical queuing (PQ) are compared under a variety of operational scenarios. The benefit of the predictability, i.e., the additional flexibility, is the reduction of operating cost because the cost of departure queuing delay can vary widely among flights. Fast-time simulations assuming a variety of operator cost functions and optimization objectives are developed and the results indicate that inter-operator exchanges to reduce the size of small physical queues could substantially

improve operator flexibility performance, measured in the average and the standard deviation of positive delay per flight (Wojcik et. al, 2013).

In this research, the percentile model proposed in section 2.3 captures the impact of flight predictability on SBT setting better than the traditional mean-variance model. The distribution of the historical block time— specifically the taxi-out time component—for a flight is depicted by the difference between every 10th percentiles. It is found that SBTs have decreasing sensitivity to historical taxi-out times toward the right tail of the distribution. In other words, airlines put different weights on different segments of the historical taxi-out time distribution when deciding SBTs for the next planning period, with more weight on the median and inner right tail of the distribution, and downplayed weight on the outer right tail of the distribution (beyond 80th percentile). Their “optimistic” scheduling behavior is rooted in the profit-driven nature of the airline operation.

With the existing linkage between surface operations and predictability improvements, and between increased predictability and benefit in saving SBT, it is a reasonable expectation that the surface operation management could, if so desired, be used to improve strategic predictability of taxi-times in a manner that could ultimately allow reductions in SBTs. Airlines consider SBT to be very expensive. Thus, saving in SBT through improved surface operations can be an additional substantial benefit of the departure queue management. When air traffic controllers manage aircraft in the departure queue, they seldom consider the impact of their decisions on flight predictability reflected through the actual taxi-out time of the flight. This is because improving predictability, i.e., the distribution of taxi-out time, does not carry any immediate operational benefit. Rather, the benefit would be realized in the long run when historical block time distributions affect the choice of SBT.

This chapter aims at filling the missing link between surface operations and strategic planning, through the improvement of flight predictability. It investigates the impact of flight predictability on SBT and the potential of improvement for SBT that can be attained through improved flight predictability. In this chapter, we propose to build on the estimation results of the percentile model that reveals the relationship between SBT and flight predictability, especially the distribution of historical taxi-out time. We propose to use departure queue sequencing to achieve adjustments in realized taxi-out times and, over time, in the distribution of these times. The adjustments can then be translated into changes in SBT using the relationship revealed by the percentile model. Through this analysis, we are able to determine the potential benefits of operational improvements at a strategic planning level. This adds to the set of mechanisms through which improved surface traffic management can generate value. Therefore, this chapter applies the results from the percentile model by directly taking the improvement on SBT into consideration when analyzing the impact of a NAS operational improvement. In this way, we are able to explore a heretofore unrecognized mechanism for realizing benefit from surface operation improvements.

3.2 Departure Queue Optimization Model

There are two things about the percentile model in section 2.3 that specifically encourage us to use it to help construct the optimization model in this chapter. Firstly, the percentile model aggregates flights in a way that best replicates the industry SBT setting practice (by origin-destination pair, aircraft type, 30-minute departure time window, and airline), among other studies. Therefore, we have more confidence in developing our optimization with the purpose of helping airlines reduce SBT. Secondly, the percentile model includes the distribution of the taxi-out time and the non-taxi-out time components of block time separately. In this way, the impact of taxi-out time on airline’s future SBT setting can be more directly predicted. Given that this study focuses on airport surface operation that adjusts only the taxi-out time of the flights, this characteristic of the percentile model is ideal.

Section 2.3 gives a complete set of coefficients from their percentile model describing the impact of the block time distribution, separately for each component of block time, on airline SBT setting in Table 2.3. For the interest of the study in this chapter, the optimization only adjusts the taxi-out time for each flight. Therefore, we are only interested in the taxi-out time portion of the percentile model. Table 3.1 lists the estimation results from percentile model for taxi-out time, which is a partial representation of Table 2.3.

Table 3.1 Partial Estimation Results from the Percentile Model in Section 2.3

<i>Variable</i>	Percentile Model	
	<i>Estimate</i>	<i>p-Value</i>
$TO_{0.5}^{f,q,y}$	0.709	<.0001
$dTO_{36}^{f,q,y}$	0.550	<.0001
$dTO_{67}^{f,q,y}$	0.432	<.0001
$dTO_{78}^{f,q,y}$	0.211	<.0001
$dTO_{89}^{f,q,y}$	0.077	<.0001
$dTO_{90}^{f,q,y}$	0.00066	0.5710

3.2.1 Problem Formulation

In this section, we develop a set of methodologies for the daily airport surface operation to optimize the block time distribution for each flight, over a certain time period, with a final objective of reducing the SBT.

As mentioned before, SBT is the time duration between the scheduled departure and scheduled arrival time. The actual block time is the time between actual departure and arrival time and varies from day to day for the same flight. The block time can be further decomposed into taxi-out, airborne and taxi-in time. For this study, the changes made to

surface operation are aimed at adjusting the taxi-out time for each flight. The taxi-out time for a flight is the time duration between the flight's actual departure time, and the wheels-off time (the time the aircraft takes off and leaves the runway). In the percentile model, the block time is decomposed into the three components and the distribution of each component is separately included in the model. Therefore, it is logical to adjust the taxi-out component through optimizing surface operations, and analyze the impact on SBT using the estimation results from the percentile model.

The optimization of surface operations is realized through re-sequencing the departure queue at the airport. After flights pushback from the gate, they enter the ramp area, where aircrafts form a queue for departure at the runway threshold. Air traffic controllers receive the departure requests, and establish the sequence of departing aircraft by requiring them to adjust ground operations, if necessary, to achieve proper spacing. (ORDER JO 7110.65V, 2014). Other information, such as departure delay, runway configuration and coordination with arrival aircrafts, are also considered in the process. However, the major objective, aside from ensuring safety, is to handle requests to enter the taxi-way system in roughly the order they are received. Once established within a queue on a taxiway, the queue cannot be reordered if the width and configuration of the airport aprons, taxiways, and runways cannot support the simultaneous movement of aircraft within what is normally restricted space (ADFMS report). This study develops an algorithm to re-sequence the departure queue, with an objective of improving flight predictability. The algorithm is applied to real airline performance data from the Bureau of Transportation Statistics (BTS) Airline On-time Performance database, which contains detailed performance information for individual flights by major US air carriers between points within the United States. The re-sequencing process can be achieved in practice by the "virtual queue" concept. The optimization generates the departure queue sequence, and flights are granted the information regarding the time to leave the gate. In this way they can leave the gate and enter the departure queue in the desired sequence.

The optimization is performed on a daily basis and is formulated as an assignment problem. All the flights departing in a day form a flight pool, I . For each flight $i \in I$, there is a corresponding wheels-off time slot j , which is the time when the aircraft takes off. All the slots comprise a slot pool, J , which has the same size as I . In this study, we assume that the wheels-off slots are known in advance, based on historical flight performance information. On a given day, the flights are re-assigned to the slots under our algorithm to achieve better predictability, resulting in a re-sequenced departure queue. The assignment results are presented by matrix X , where $X_{ij}=1$ if flight i is assigned to wheels-off slot j , and is zero otherwise. The optimization can be expressed as:

$$\text{Min } \sum_{i,j} X_{ij} \times C_{ij}$$

$$\text{s.t. } \sum_i X_{ij} = 1, \forall j$$

$$\sum_j X_{ij} = 1, \forall i$$

$$\sum_i X_{ij}(\text{dep}_i + \text{minTO}_i) \leq \sum_j X_{ij} \text{Woff}_j, \forall i, j$$

The objective of the optimization is to minimize the total “cost of assignment”, summed over all the assignments, i.e., the C_{ij} s, throughout the day. The specification for the “cost of assignment” will be explained in greater detail shortly. To guarantee the feasibility of the re-sequencing, there are several constraints. The first two constraints are to assure that all the flights in the day are assigned to one and only one slot, and that all the slots are assigned to one and only one flight. The last constraint is included to guarantee the operational feasibility of the assignment. It is not possible for a flight to take off at a wheels-off time slot earlier than the time it leaves the gate. There is also a minimum taxi-out time required for the flight, between its departure time and wheels-off time. In this study, the minimum taxi-out is calculated as the minimum of the 20th percentile of the taxi-out time distribution of the same flight i over a month, and the actual taxi-out time for the flight at the day of operation. The third constraint says that flight i can only be assigned to a wheels-off slot that is after flight i 's departure time by at least the minimum taxi-out time for this flight.

The definition for an individual flight i is more complicated than merely categorizing by flight number. The aggregation of flights is by specific airline, origin-destination pair, 30-minute departure time window, and aircraft type. This is the same aggregation used to calculate the block time distribution variables in the percentile model (see Section 2.3.1). This method is identified by airline block time setting groups, and they choose the SBT for each of the individual flights as a group. Therefore, it is a reasonable and consistent aggregation method for defining predictability if we want to build on the percentile model and study the impact of the departure sequence optimization on SBT. The time unit for the aggregation in this chapter is month.

3.2.2 Cost of Assignment

All the flights departing a certain airport on a single day can potentially be re-sequenced under certain feasibility constraints. For each flight, its actual block time starts as it leaves the gate and enters the departure queue. Then it proceeds through the taxi-out, airborne and taxi-in phase, the time durations of which are the three components of the block time. The percentile model gives a complete set of coefficients describing the impact of the block time distribution, separately for each component of block time, on airline SBT setting. The coefficients for the different segments of the taxi-out time distribution are listed in Table 3.1. Besides the interpretation of the results described in section 2.3.2, another way to interpret these results is on a flight level. For example, for a given flight i on a specific day d , its actual taxi-out time TO_i^d makes a certain contribution to the SBT setting in the next planning period when schedulers are using this period of historical performance as reference. The SBT set for this flight is SBT_i and the contribution is $\Delta SBT_{i,d}$. The contribution depends on which segment of the distribution this actual taxi-out time data point falls in and can be quantified by the estimation results presented in Table 3.1. The contribution is the difference between the actual taxi-out time and its nearest lower-bound percentile in the historical distribution, plus all the 10th percentile segments below the actual taxi-out time. These segments are weighted by the coefficients given in Table 3.1, denoted as β_j 's. The mathematical formulation is:

$$\Delta SBT_{i,d} = \alpha_1 \times TO_{i,50} + \sum_{j=1}^5 \beta_j \times \min(\max(TO_i^d - TO_{i,j+4}, 0), TO_{i,j+5} - TO_{i,j+4}) \quad (3.1)$$

where $TO_{i,j}$ is the value of the j^{th} percentile of the historical taxi-out distribution for flight i , where $j=5, \dots, 10$. The percentiles are calculated using two years of historical taxi-out times, consistent with the way the data is processed in section 2.3 for the percentile model. The aggregation method is also the same as in section 2.3, and the time unit is by quarter. For this interpretation, the underlying assumption is that the historical distribution for calculating $TO_{i,j}$ is stable through time and is going to repeat for the existing time period when TO_i^d happens, providing the percentile values as parameters in the objective function. Even though the re-sequencing process changes the actual taxi-out time on a daily basis, and thus changes the distribution, in this study we still use the original distribution to extract the parameters in the optimization and keep them the same for the case study conducted in this chapter. The parameters are only indicating percentile values we used in the objective function; and they do not reflect the true on-going taxi-out times during the optimization period. This is reasonable because in this research, we only conducted the optimization daily for a month. If the time horizon of the optimization is to be expanded, the parameters (percentile values) extracted from historical distribution will also need to be updated to consider the newly adjusted taxi-out times. Even when just considering one month, in the real-world setting, this scenario of stable monthly block time distribution is still hard to achieve. We make this assumption more reasonable by selecting multiple years of historical data when calculating the different percentile values. The detailed method for this calculation will be discussed in section 3.2.3 shortly.

For example, suppose flight i has a distribution of historical taxi-out time characterized with the 50th through the 100th percentile. If on a specific day d , the flight experiences a taxi-out time TO_i^d that falls between the 70th and 80th percentile of the flight's historical taxi-out time distribution, denoted as $TO_{i,70}$ and $TO_{i,80}$. Then, the contribution of this flight's taxi-out time to the SBT being set during the next planning period would be: the difference between TO_i^d and $TO_{i,70}$, plus the difference between the 50th and 60th, and 60th and 70th percentile of this flight's historical taxi-out time distribution, and plus the median value of the distribution. Again all these components should be weighted by the coefficients in Table 3.1 as:

$$\Delta SBT_{i,d} = \alpha_1 \times TO_{i,50} + \beta_1 \times (TO_{i,60} - TO_{i,50}) + \beta_2 \times (TO_{i,70} - TO_{i,60}) + \beta_3 \times (TO_i^d - TO_{i,70}) \quad (3.2)$$

In the context of this study, we are interested in adjusting the distribution of actual taxi-out time to reduce SBT. Therefore, the cost of the assignment C_{ij} that we are trying to minimize in the optimization should be the contribution to its future SBT, for the flight i assigned to a wheels-off time slot j . On a certain day, flight i is assigned to slot j through the optimization, generating a new taxi-out time which is the new wheels-off time minus the original departure time of the flight, $TO_{ij} = \text{WheelsOff}_i - \text{Dep}_i$. The ‘‘cost of assignment’’ associated with this assignment is thus measured by the contribution to SBT from the new taxi-out time, TO_{ij} , denoted as:

$$C_{ij} = \sum_{j=1}^5 \beta_j \times \min(\max(TO_{ij} - TO_{i,j+4}, 0), TO_{i,j+5} - TO_{i,j+4}) \quad (3.3)$$

3.2.3 Benefit Mechanism of the Cost of Assignment

Through the optimization process, the total daily “cost of assignment”, i.e., the total contribution to future SBT summed over all flights in the day, is minimized. In fact, by re-assigning flights based on such an objective function daily for a period of time, we are actually able to change the historical distribution of taxi-out time for each flight. In this way, the parameters (percentile values) extracted from the updated actual taxi-out times are therefore adjusted in the favored direction, i.e., towards reducing the future SBT. In this section, a simplified example is provided to illustrate the mechanism of SBT reduction through the optimization with the “cost of assignment” specified in section 3.2.2.

Table 3.2 Percentile Values and Parameters for the Hypothetical Flight

Taxi-out Time	Hypothetical Value (min)	Variable in Percentile Model	Estimate
TO_{50}	15	$TO_{0.5}^{f,q,y}$	0.709
TO_{60}	17	$dTO_{56}^{f,q,y}$	0.550
TO_{70}	18	$dTO_{67}^{f,q,y}$	0.432
TO_{80}	20	$dTO_{78}^{f,q,y}$	0.211
TO_{90}	22.5	$dTO_{89}^{f,q,y}$	0.077
TO_{100}	26	$dTO_{90}^{f,q,y}$	0.00066

Table 3.2 above shows the values for different percentiles of the taxi-out time for a hypothetical flight. For purposes of this example, assume that the taxi out times can take one of these six values listed—50% of the time the value is 15, 10% of the time it is 17, 10% of the time it is 18, and so on. Now suppose that, through re-sequencing, the 22.5 minute taxi-out times can all be changed to 20 minute taxi-out times. This brings the 90th percentile taxi out time down from 22.5 to 20. The resulting change in scheduled block time would be $\beta_4 \times dTO_{89} = 0.077 \times (TO_{90} - TO_{80}) = 0.077 \times 2.5 = 0.19 \text{ min}$. This is exactly the difference in the assignment cost in Equation 3.3 between an assignment to a 22.5 min taxi-out time and the assignment to a 20 min taxi-out time. By similar logic, it can be shown that if the 22.5 taxi-out times all became 18 minutes, the block time reduction would be:

$$\begin{aligned} \beta_4 \times dTO_{89} + \beta_3 \times dTO_{78} &= 0.077 \times (TO_{90} - TO_{80}) + 0.211 \times (TO_{80} - TO_{70}) \\ &= 0.077 \times 2.5 + 0.211 \times 2 = 0.61 \text{ min} \end{aligned}$$

This is again the difference in the cost of assignment using Equation 3.3. Although in actuality taxi-out times vary continuously and it is not possible to make the same substitutions from day to day, the same reasoning applies to the “one-off” substitutions that are made using our assignment model. In summary, through the simplified example in this section, we have how the optimization we designed captures the impact of re-sequencing the taxi-out time distribution, and consequently on future SBT.

3.2.4 Optimization Result and Impact on SBT

The cost of assignment defined in section 3.2.2 directly represents the contribution of the taxi-out time to SBT, in the unit of minute. Therefore, the optimization minimizes the total contribution from all the flights’ taxi-out times to their corresponding SBTs in the future, on a daily basis. The daily optimization is conducted for 26 days in January 2011, at John F. Kennedy International Airport (JFK), excluding Saturday operations, which is consistent with the data processing method used in section 2.3.1. By comparing the difference between the original value of the objective function and the optimized value, we would acquire the direct benefit of improved predictability, realized through the optimization. For each day, a significant reduction in the objective value is achieved. Dividing this reduction by the number of flights in the day, the per flight saving for SBT can then be calculated. Averaging over the 26 days of our optimization, the optimization is able to reduce the contribution of taxi-out time to future SBT by 0.61 minutes per flight.

JFK is notorious for its busy air traffic, especially the busy surface operations which often result in excessive taxi-out times. Given the fact that JFK is among the ten airports that have the longest taxi-out times (BTS, 2007), the benefit from departure queue re-sequencing is expected to be significant for JFK. To compare the effect of the proposed algorithm, we apply the optimization to other two less congested airports: the Dallas/Fort Worth International Airport (DFW), and the San Francisco International Airport (SFO), during the same 26 days in January 2011. Respectively, the re-sequencing process described above reduces the contribution of taxi-out time to future SBT by 0.27 minutes per flight for DFW, and 0.23 minutes per flight for SFO.

Table 3.3 presents the result of the optimization for the three airports in greater detail. Column 2 and 3 show that DFW airport has the largest number of daily operations, as well as the largest portion of flights that are re-sequenced through the process. Among the three airports, the proportion of flights changed in the daily optimization is perfectly correlated with the number of daily operations at the airport. JFK is the airport with the most congested surface operation among others, and the optimization produces the greatest improvement in the future SBT (the largest reduction). For the other two airports, the reduction is much smaller, about 1/3 to 1/2 of the reduction in JFK.

The last two columns show another important performance metric—the on-time performance—of the three airports, before and after the optimization. Although on-time performance is not directly considered in the objective function of the optimization, changing the wheels-off time of the flight will consequently change the arrival time of the flight, thus possibly affecting the on-time performance. The objective function aims at pulling the inner right tail of the distribution closer to the center, and therefore we would

expect the optimization might also improve the consequent on-time performance of the flights. In our study, the on-time performance is calculated on a per airport, daily basis. For flight i departing airport j in day d , $OT_i=1$ if the flight arrives less than 15 minutes after its scheduled arrival time at the destination airport (excluding 15 minutes). Then for day d and airport j , the on-time performance is $\frac{\sum_{i=1}^{N_{j,d}} OT_i}{N_{j,d}}$. $N_{j,d}$ is the total number of

flights in day d at airport j . The new actual arrival time of the flight after it is re-sequenced in the queue depends solely on the newly assigned wheels-off time. In this study we consider the airborne time and taxi-in time of the flight remain unchanged.

Table 3.3 Optimization Results of the Departure Queue Re-sequencing Algorithm

Airport	Average No. Flights per day	Proportion of Flights Changed	SBT Reduction (min)	Original On-time Performance	On-time Performance Improvement
DFW	441	60.31%	0.273	0.609	0.0016
JFK	236	52.84%	0.611	0.434	0.0096
SFO	129	46.84%	0.240	0.782	-0.0059

In Table 3.3, the average original on-time performance across the 26 days and the average improvement on the on-time performance through the optimization is shown. Based on the original data, the on-time performance is the lowest at JFK, which is not surprising given its surface congestion. Meanwhile, JFK is also the airport that sees the greatest improvement after the re-sequencing: the on-time performance is increased by 0.0096, or 1.7%. The on-time performance is also slightly increased for DFW; however it is decreased for SFO, which is also the airport with the smallest reduction of SBT through the process. Overall, Table 3.3 shows that re-sequencing can reduce scheduled block time, not hugely but tangibly, and that flights from airports with greater level of surface congestion benefit the most from the re-sequencing proposed in this study.

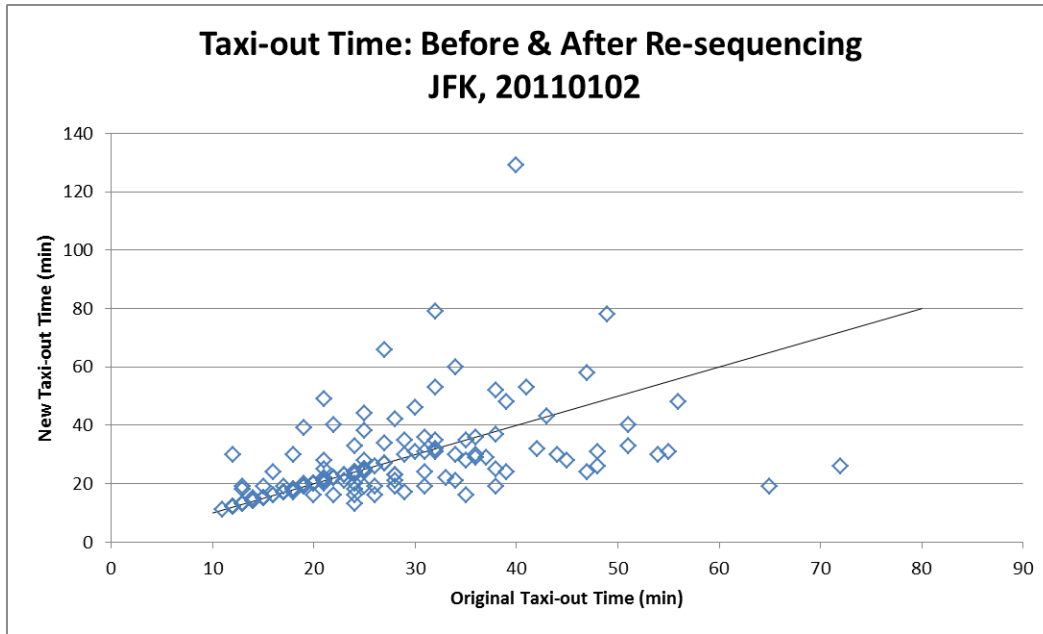
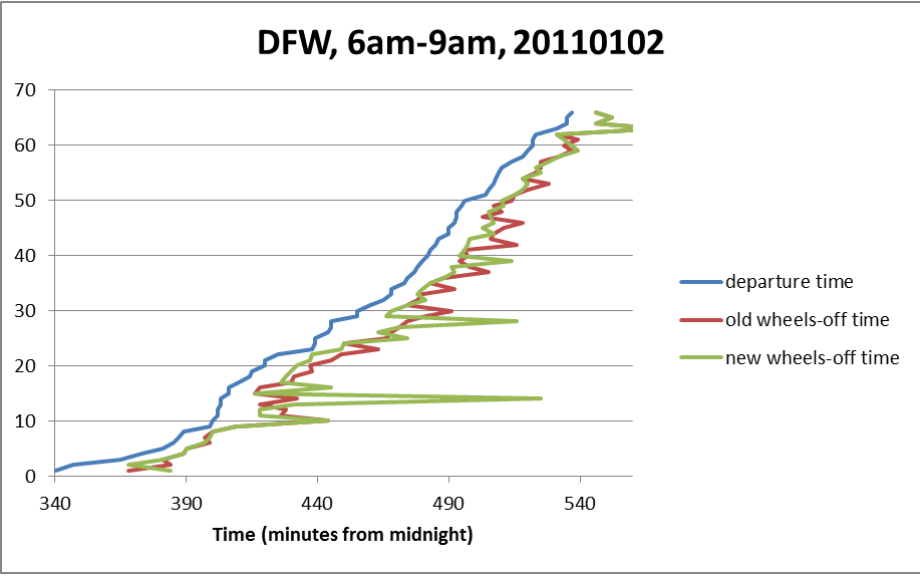
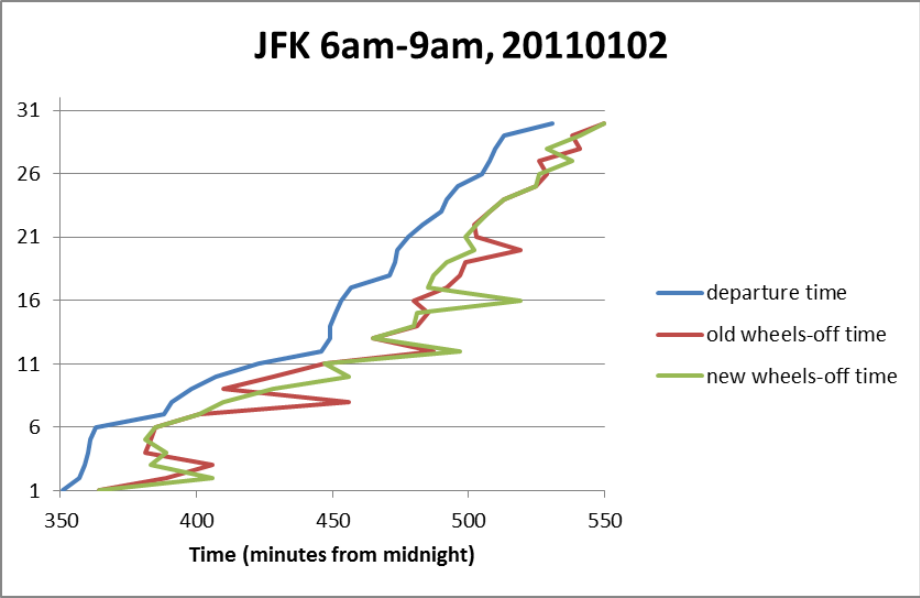


Figure 3.1 Taxi-out Time before and after Re-sequencing: JFK, 01/02/2011.

Figure 3.1 provides a closer look at how taxi-out times for individual flights are changed by the re-sequencing process, at JFK on a specific day, 01/02/2011. For each flight on this date, X axis represents its original taxi-out time, and Y axis shows its new taxi-out time after the re-sequencing. The straight line is a 45 degree line and the observations falling on this line have unchanged taxi-out times, i.e., these flights are not re-sequenced through the process. Of the 152 flights on 01/02/2015 at JFK, 71 flights fall on this line (46.7%). Observations falling above the straight line in Figure 3.1 are flights with a longer taxi-out time after the re-sequencing; and 35 flights (23%) fall into this category. The average increase of taxi-out time per flight is 11.3 minutes. 46 observations (30.3%) are below the straight line, meaning that they have a shorter new taxi-out time, and the average reduction is 8.8 minutes. The daily optimization process switches the sequence in the queue for over half of the flights; and the flights ending up with longer taxi-out times are slightly more than those with shorter taxi-out times after the re-sequencing. The magnitude for the increase is also higher on average than the average decrease in taxi-out time. From Figure 3.1 this difference in the magnitude of the change can also be observed as flights are scattered further above than below the straight line.



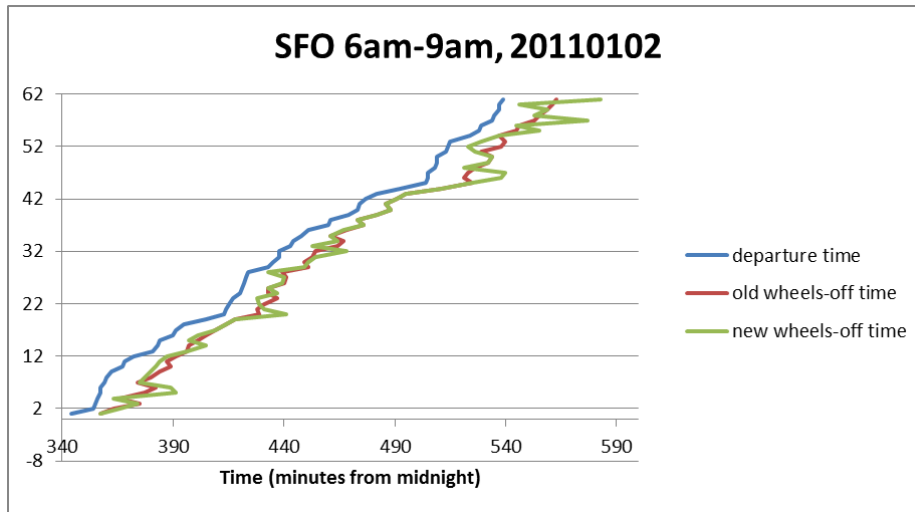


Figure 3.2 Departure Queue Diagram before and after the Optimization for JFK, DFW and SFO, 6am-9am on 01/02/2011.

Figure 3.2 provides the results from the departure queue sequencing, during certain time duration in a particular day, for the three airports. The actual departure times of the flights included fall between 6am and 9am. The blue line reflects the gate departure time of each flight, which is not changed in the optimization process. The red and green line represent the wheels-off time for each flight, before and after the re-sequencing respectively. For all the three airports, it can be observed that a significant amount of flights during the three hours are switched through the optimization. Since the optimization is re-assigning the flights to existing wheels-off slots, we can see a balanced outcome where certain flights are switched to earlier wheels-off times, while other flights are pushed back in the departure queue. JFK is the airport with the fewest number of operations (30 departing flights). Different patterns of how the optimization changes the queue sequence are observed. For example, for the 6th flight to the 11th flight in the queue, originally the earlier flight is assigned to a later wheels-off time slot, which is contrary to the first-in-first-served practice. The re-sequencing yields a queue that's more consistent with first-in-first-out pattern, where the flights leaving the gate earlier gain an earlier wheels-off time in the departure queue. On the contrary, at a later time for the 16th to 19th flight in the queue, originally the flights are sequenced in the same order as how they leave the gate. The re-sequencing process mixed up the order and the new queue is no longer in a first-in-first-out fashion.

Comparing these two segments, the earlier one has a short average taxi-out time for the flights, where the departure time is close to the wheels-off time; while in contrast the later segment has a much longer average taxi-out time for the flights. This indicates that the optimization process may work differently for time periods with busy surface operations and less busy ones. For peak hours with busy operations and longer taxi-out times, the flights' taxi-out times are more likely to fall on the right tail of their distribution (beyond the median value), and the optimization process tends to interrupt the first-in-first-out

pattern and switch flights around in the queue to achieve the best adjustment to the actual taxi-out times, and thus to the future SBT. On the other hand, during less busy hours, the taxi-out times of the flights are shorter and the re-sequencing process tends to smooth out the variation in the queue and make it more in line with a first-in-first-served pattern.

For DFW, we can also observe the effect of the optimization that pulls the queue towards both the left and right side at different positions. We can also observe the two patterns in the JFK case where the queue is adjusted differently, but more often we see cases where flights are adjusted to be more consistent with the first-in-first-served pattern (for example, the 31th to 38th, and the 42th to 48th flight in the queue). In general, these situations again happen where the average taxi-out times are shorter; and the optimization results in a pattern that is contrary to the first-in-first-out pattern when taxi-out time are greater (for example, the 10th to 20th flights in the queue). There are also a few more extreme events where the flight is significantly delayed and thus experiences a very long taxi-out time after the optimization. Compared to JFK and DFW, fewer flights are re-sequenced drastically in the departure queue at SFO. Figure 3.3 below provides a similar illustration of the original and new taxi-out times for the flights on 01/02/2011 at SFO. Among the 231 flights on that day, 104 remained the same taxi-out times. Again over half of the flights are re-sequenced in the departure queue through the process. 52 flights (22.5%) ended up with longer taxi-out times, and the average increase is 7.71 minutes for these flights. The remaining 75 flights (32.5%) are re-sequenced to have shorter taxi-out times, and the average reduction is 5.35 minutes. Similar as the same results from JFK, the magnitude of the increase is higher. However, the scale of the change is much smaller compared to JFK, for both positive and negative changes.

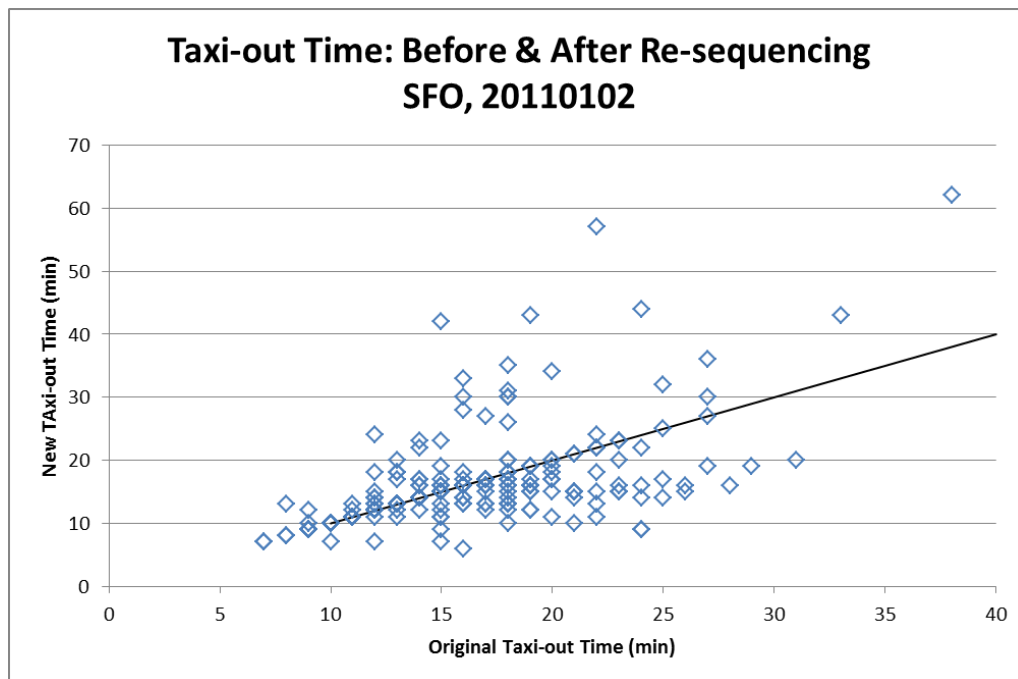


Figure 3.3 Taxi-out Time before and after Re-sequencing: SFO, 01/02/2011.

3.3 The Effect of On-time Performance

The formulation of the “cost of assignment” identified in section 3.2 optimizes the contribution of actual taxi-out time to future SBT. However, the sole focus on reducing future SBT makes it less realistic, since there are other objectives the airlines consider in the re-sequencing process. For example, as briefly mentioned in section 3.2.3, the process would definitely affect the on-time performance, which is an important performance metric in airline operations. Since the percentile block-time model gives little weight to the far right tail of the distribution, it is possible that the optimization will deliberately assign extremely long taxi-out times to certain flights where the contribution to future SBT is negligible to ensure other flights gain more benefits (SBT reduction) from reduction at the “inner right tail” of the distribution. In these cases, the on-time performance (fraction of flights that arrive less than 15 minutes behind schedule) of the delayed flights may be compromised. Additionally on-time performance is affected by gate delay, which plays a very small role in scheduled block time setting (see Chapter 2). Therefore, in this section, the trade-off between impact on SBT and the on-time performance of the flight is taken into consideration when designing the “cost of assignment”.

3.3.1 Updated Objective Function

In this section, a second piece in the objective function will be added to account for on-time arrival performance as part of the “cost of assignment” as well. In this way, we develop a multi-objective algorithm to sequence departure flights to improve predictability, reduce airline scheduled block time, while maintaining on-time performance. On a certain day, flight i is assigned to slot j through the optimization, generating a new taxi-out time which is the new wheels-off time minus the original departure time of the flight, $TO_{ij} = WheelsOff_j - Dep_i$. In addition to the contribution to future SBT of this new taxi-out time, as described in section 3.2, we also calculated the new on-time performance of the flight with the updated taxi-out time. As mentioned in section 3.3, we assume that the airborne time Air_i and taxi-in time $TaxiIn_i$ for flight i remain the same after being re-sequenced. The new actual arrival time of the flight i assigned to slot j is therefore $Arr_{ij} = WheelsOff_j + Air_i + TaxiIn_i$. If the flight arrives less than 15 minutes after its scheduled arrival time at the destination airport, then the on-time performance of this flight, $OT_{ij}=1$. Otherwise, $OT_{ij}=0$. In our study, we minimize the objective function, which should exert cost to airline operations. Therefore it is the dummy variable reflecting the flight not being able to arrive on time that should be included in the “cost of assignment”. We then develop the variable $Delay_{ij}=1-OT_{ij}$ that reflects whether the flight is delayed or not and incorporate that into our objective function to be minimized in the optimization. The updated “cost of assignment” is now:

$$C_{ij} = \sum_{j=1}^5 \beta_j \times \min(\max(TO_{ij} - TO_{i,j+4}, 0), TO_{i,j+5} - TO_{i,j+4}) + Delay_{ij} \quad (3.4)$$

The new “cost of assignment” has two components that airlines try to minimize in their operations: the contribution to future SBT, and the on-time performance of the flight. As in section 3.2, the daily optimization is conducted for 26 days in January 2011, at JFK, DFW and SFO airports (excluding Saturday operations as in estimating the percentile model). Since the new objective function contains two parts with different units and magnitudes, we cannot simply compare the original value of the objective function and the optimized value as in section 3.2. Instead, we calculate the change in the contribution to SBT in minutes, and the change in daily average on-time performance separately. Table 3.4 presents the results from the optimization, both with the original objective function in section 3.2, and the updated one incorporating on-time performance proposed in this section. Firstly, the proportion of re-sequenced flights in the queue remains the same as before. Nearly or over half of the flights are re-sequenced in the new optimization model as well, and this proportion is positively correlated with the number of daily operations.

Table 3.4 Optimization Results of the Departure Queue Re-sequencing: Basic and Updated Objective Functions

Airport	Average No. Flights per day	Proportion of Flights Changed	SBT Reduction (min)	Original On-time Performance	On-time Performance Improvement
	Basic objective function				
DFW	441	60.31%	0.273	0.609	0.0016
JFK	236	52.84%	0.611	0.434	0.0096
SFO	129	46.84%	0.240	0.782	-0.0059
	Incorporating on-time performance				
DFW	441	60.42%	0.270	0.609	0.0142
JFK	236	53.04%	0.608	0.434	0.0232
SFO	129	46.87%	0.235	0.782	0.0080

A significant per flight reduction in the contribution to future SBT, as well as a significant increase in daily on-time performance (except SFO) is achieved. Compared to the basic objective function, the reduction in SBT is slightly reduced. The difference in SBT reduction between the basic and the updated objective function is the largest for JFK (around 3.6% reduction), and smaller for DFW (1% reduction) and SFO (2% reduction).

In exchange of the less reduction in future SBT, the daily on-time performance of the daily flights is improved, since the current objective function now directly incorporates the number of delayed flights. The greatest absolute improvement in on-time performance compared to the prior optimization happens at JFK, where the on-time performance improvement is increased by 0.0136, while the improvement is 0.0126 for DFW. For SFO, originally the optimization would decrease the on-time performance of the airport compared to the original sequencing. The new optimization considers late arrivals as a cost in the objective function being minimized and yields positive improvement from the original on-time performance by 0.008, which is an increase of 0.0139 from the original version of optimization. Compared to JFK and DFW, SFO has less issue on the surface; it has other issues, such as foggy weather limiting runway operations that are more crucial in impacting its flight performance. Therefore, the on-time performance is less sensitive to departure re-sequencing, and the proposed optimization is less effective in improving the on-time performance at the airport compared to others; however it is still effective in increasing the daily on-time performance in the updated optimization version.

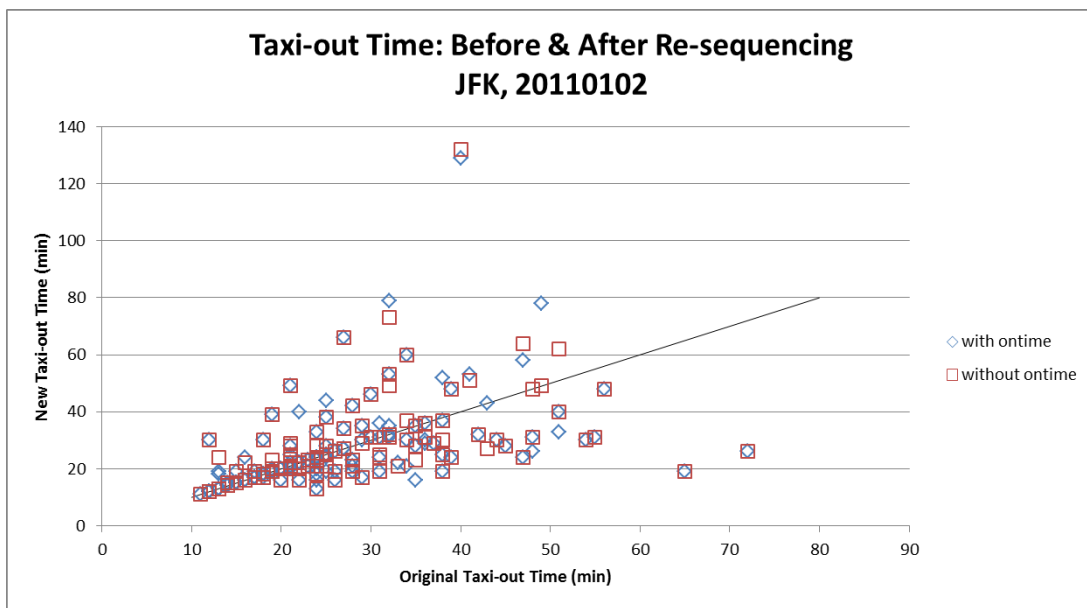


Figure 3.4 Taxi-out Time before and after Re-sequencing, with the Original and Updated Optimization: JFK, 01/02/2011.

Figure 3.4 provides the comparison of the changes in taxi-out times for both the original and updated version of the re-sequencing. Similarly as before, the straight line in the graph indicates unchanged taxi-out times through the re-sequencing. Observations above the line have longer new taxi-out times and those below the line have shorter new taxi-out times. The two versions of the optimization have mostly similar outcome and many observations overlap in the graph. For the original optimization version, for those flights with a shorter taxi-out time, the average reduction in taxi-out time is 8.8 minutes, while it is 11.2 minutes for the updated optimization version with on-time performance. For the flights with a

higher taxi-out time after the re-sequencing, the average increase for the original optimization is 11.3 minutes, and 14.3 in the updated version. The updated version of optimization increases both the positive and negative change in taxi-out times. It more aggressively adjusts the taxi-out times for the flights, reducing taxi-out times for a better on-time performance, while maintaining the increase in taxi-out times at acceptable level and not further interrupting on-time performance.

3.3.2 Sensitivity Analysis

The optimization conducted in section 3.3.1 considers both the reduction in future SBT and the improvement in daily on-time performance as benefit from the departure queue re-sequencing. However, on a per flight level, the algorithm weighs the reduction in SBT and the on-time performance equally. In this section, we propose to adjust the relative weight of the two components in the objective function to investigate the sensitivity of the re-sequencing process. The new formulation of the “cost of assignment” is now:

$$C_{ij} = \sum_{j=1}^5 \beta_j \times \min(\max(TO_{ij} - TO_{i,j+4}, 0), TO_{i,j+5} - TO_{i,j+4}) + \alpha \times Delay_{ij} \quad (3.5)$$

By changing the value of α , we will be able to adjust the cost of delay relative to the cost of contribution to future SBT in the objective function and see how the optimization result varies. In this section, we only consider the two airports with the most and least effectiveness from the optimization in section 3.2 and 3.3: JFK and SFO. Table 3.5 shows the result of the sensitivity analysis with different values of α . For both airports, the trends for SBT reduction and on-time performance improvement as α increases are similar to each other. As α gets larger, the optimization focuses more on reducing delayed flights, and thus is less inclined to strategically delay certain flights to the further right tail of the taxi-out time distribution. Therefore, the reduction of the contribution to future SBT decreases as α gets larger. In exchange, the improvement in on-time performance grows significantly larger as more weight is allocated to this factor. For SFO, originally the optimization decreases the on-time performance. Adding the on-time performance piece into the objective function turns the trend and now the on-time performance is increased after the re-sequencing, and the increase gets larger as more weight is allocated to the on-time performance factor. Using the updated objective function, from Table 3.5, we observe the trade-off between maintaining on-time performance and adjusting the distribution to reduce future SBT. The adjustment is restrained more (especially the cases where taxi-out time is strategically prolonged) as on-time performance plays a more important role in the objective function.

Table 3.5 Results of the Sensitivity Analysis Changing Relative Weights in the Updated Objective Function: JFK and SFO

α	JFK			SFO		
	Proportion of Flights Changed	SBT Reduction (min)	On-time Performance Improvement	Proportion of Flights Changed	SBT Reduction (min)	On-time Performance Improvement
0	52.84%	0.611	0.0096	46.84%	0.240	-0.0059
1	53.04%	0.608	0.0232	46.87%	0.235	0.0080
2	52.50%	0.601	0.0276	46.87%	0.232	0.0108
5	52.77%	0.581	0.0336	46.47%	0.231	0.0112
10	52.83%	0.576	0.0345	46.68%	0.231	0.0112
50	52.47%	0.576	0.0345	46.64%	0.231	0.0112

Another interesting finding from Table 3.5 is that while the improvement of the on-time performance through the optimization is significant, it is still limited. Beyond certain value of α (50 for JFK and 10 for SFO), the changes remain the same and more weight for on-time performance no longer change the result of optimization. These limits indicate the maximum room for improving on-time performance through re-sequencing. In other words, using all the available resources to re-sequence the departure queue, these limits are the best we can improve upon the on-time performance, which is about 8% increase for JFK, and only 1.5% for SFO. One possible explanation about this limited impact of re-sequencing is that flights are delayed in batches. If one flight is about to be delayed (arriving more than 14 minutes after its scheduled arrival time), it is highly likely that the other flights departing at similar times are also delayed. It is hard for this flight to get out of the “delayed pool” of flights, without resulting in another flights being delayed. Therefore, the proportion of the delayed flights for the day has limited room to be decreased. Another reason is that taxi-out is less dominant a component determining whether the flight is delayed. When leaving the gate, the departure delay is already incurred by the flight and the re-sequencing only adjusts the taxi-out time of the flight after it leaves the gate. Gate delay is a more dominant component and changing the taxi-out time doesn’t have an overwhelming power to change the expected delay situation of the flight.

Figure 3.5 illustrates the results of the optimization, with the original and updated objective function, for JFK, during a 2-hour period in one specific day. Comparing the

two graphs, one significant change of the sequence happens around the 75th flight, where the updated optimization with on-time performance incorporated into the objective function has effectively prevented delaying this flight in the queue, most likely in an effort to maintain the on-time performance of this flight. This example helps explain the difference of the outcome lying between the two optimization processes.

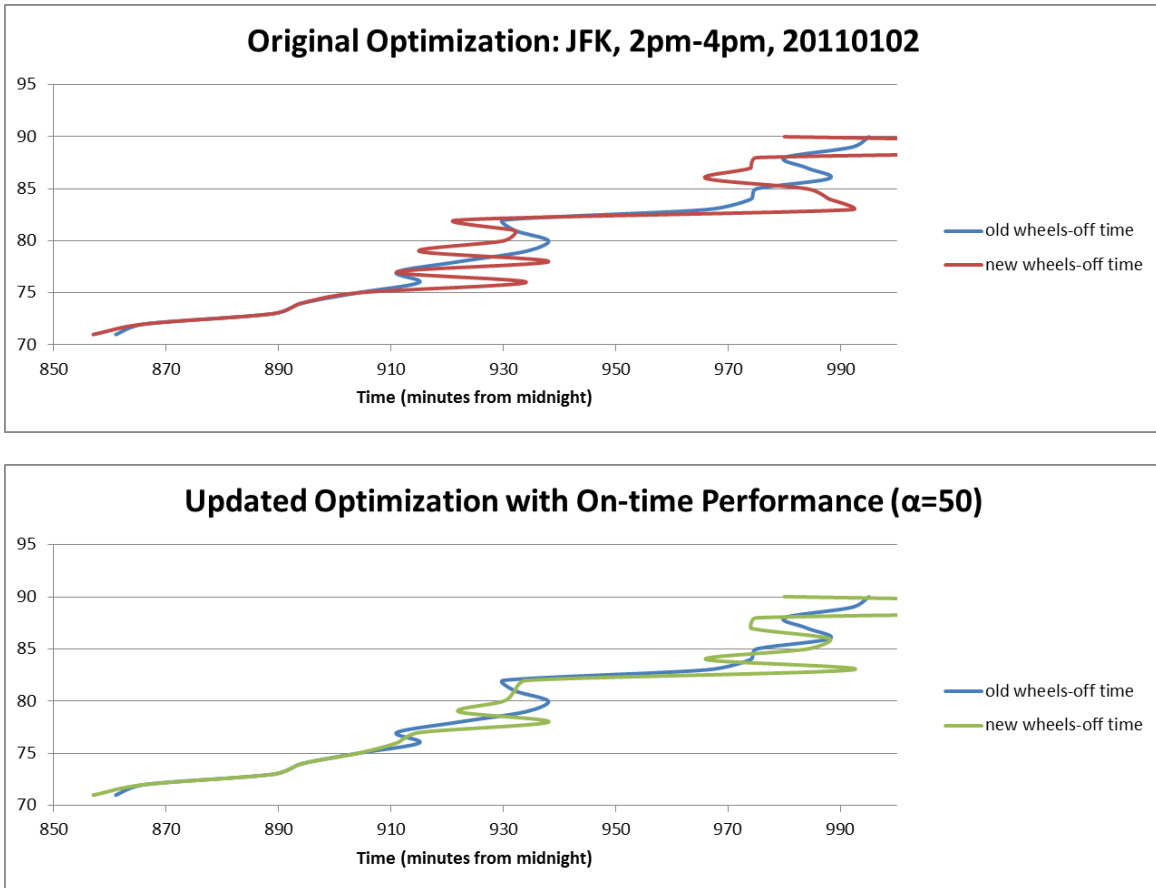


Figure 3.5 Departure Queue Diagram for the Original and Updated Optimization for JFK, 2pm-4pm on 01/02/2011.

From Table 3.5 we learned that the trade-off between SBT reduction and on-time performance improvement varies as weight given to on-time performance is adjusted. However, the scales of the changes in the two components in the objective function are quite different. Figure 3.6 plots the trade-off between the two dimensions and provides a more direct illustration of the scale of changes, for both JFK and SFO. The values of α increase as the plot goes from left to right. From the figure, we can clearly see that the reduction in SBT decreases while the improvement of on-time performance increases as more weight is assigned to the cost of on-time performance. For both JFK and SFO, the plot displays an “elbow-shape” frontier. At first great on-time improvement is achieved

with slight compromise of the SBT reduction for incremental increase in α ; after a certain point, we need to sacrifice greatly in reducing SBT to trade for a marginal increase in on-time performance. For both airports, the elbow is when α is in the range between 1 and 2 (derived from combining the results in Table 3.5 and Figure 3.6). One last thing to notice is that the points on the plots overlap at the left end. This is consistent with the observation from Table 3.5 that beyond certain value of α , no improvement in on-time performance can be obtained.

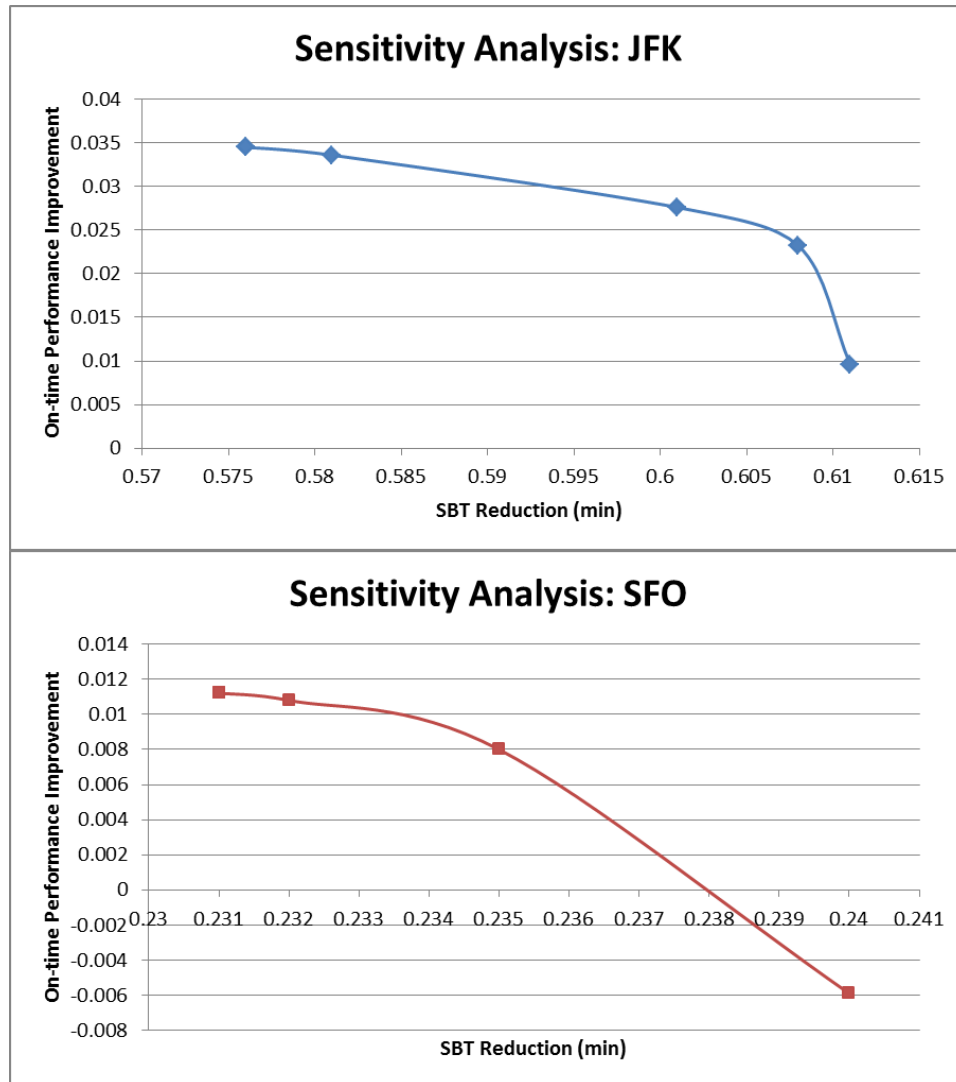


Figure 3.6 Sensitivity Analysis Frontier: JFK and SFO

These results show that while incorporating the on-time performance compromises the efficiency in reducing SBT to some extent, this can lead to a small but measurable gain in on-time performance. Overall, the sensitivity analysis shows that the proposed optimization is still able to reduce future SBT significantly, while also maintaining prior levels of daily on-time performance for flights departing the airport.

3.4 Conclusion

In this chapter, the benefit of improved predictability reflected on reduced SBT is quantified. Predictability is improved through sequencing the departure queue at the airport, with the objective of adjusting the distribution of taxi-out time in a way that can allow airlines to set shorter SBTs in the future. We build on the results of the percentile model in section 2.3 that reveals the relationship between SBT setting and historical block time (more specifically, taxi-out time) and define objective functions that capture the contribution of the daily taxi-out time of each flight to its future SBT. The sequencing process re-assigns the flights in a day to the existing take-off slots at the airport and minimizes the contribution to future SBT, with constraints that guarantee the feasibility and the one-to-one matching of the assignment.

The first form of the objective function we applied strictly replicates the different contributions to SBT from each segment of the taxi-out time distribution estimated from the percentile model. In an effort to also take on-time performance into consideration, an alternative form of the objective function is also developed. A case study is conducted with three airports: JFK, DFW and SFO in January 2011. The daily optimization is conducted for the 26 non-Saturday days in the month; and the average saving in future SBT per flight is calculated. On average, through the optimization airlines can save 0.61 minutes of future SBT per flight in JFK, 0.27 minutes in DFW, and 0.23 minutes in SFO. This shows that the benefit of reduced SBT of the proposed re-sequencing varies across airports and is more valuable at airports with more congested surface operations.

The proposed objective function focuses solely on reducing future SBT, which is not the only objective of concern to airlines. As a result, it is possible that some flights' taxi-out times are strategically prolonged because the contribution to additional SBT is negligible on the far right tail of the taxi-out time distribution. This may compromise the on-time performance of the flight, which is also an important metric for the airlines. Therefore, we also develop a multi-objective optimization that aims at both reducing future SBT and improving airport daily on-time performance. A case study is conducted for JFK and SFO, as well as a sensitivity analysis. A trade-off between the two components in the objective function is observed. While the SBT reduction is slightly compromised with on-time performance added (around 0.6%), the increase in the improvement of on-time performance is huge (around 141.8%). As the weight given to on-time performance grows larger in the sensitivity analysis, the SBT reduction continues to get smaller, in exchange of an increase in the on-time performance; and the difference between the percentage changes of the two components become less drastic. Beyond certain weight given to on-time performance, the optimization becomes stable and is no longer sensitive to this weight.

These results show that if reducing SBT is the only goal of the airport surface management, then the departure queue sequencing process should further delay the "hopeless" flights whose expected taxi-out time is predicted to be much longer than historical performance. When adding on-time performance into the objective function,

the reduction in SBT is compromised to some extent, but with greater gain in improved on-time performance. However, airport departure queue management involves multiple objectives including delay, fuel burn, etc. This study contributes to the existing literature by pointing out another potential factor to consider, which is predictability, and demonstrated the benefit in the form of reduced future flight SBT. Neglecting this benefit might lead to suboptimal airport surface management decisions and reduced value for airlines and other stakeholders. In future research, incorporating other objectives in addition to the single predictability objective function considered in this chapter would be a promising research direction.

As a final remark, the benefit quantified in this chapter, especially in the form of reduced SBT, is contingent to the length of application of this re-sequencing process. In this chapter, re-sequencing is conducted consecutively over the period of a month. The reduction of future SBT on a per flight level specified by the optimization process is under the assumption that the daily optimization is performed for a month and only the taxi-out times during this month is considered as historical performance for future SBT. In reality, multiple years of historical block time (taxi-out time) information is used aggregately by the airlines to determine their future SBTs. Therefore, in order to truly achieve the benefit, the re-sequencing process should be repeated for a period much longer than only one month as in this study. Extending the optimization period will also undermine the assumption of the historical block time distribution being stable and thus using the same percentile values for the extended optimization would no longer be valid. The optimization process changes the actual taxi-out times and the parameters should be updated for the later periods because the distribution has been changed and is no longer stable throughout time. Since in the percentile model, flights are aggregated by quarter, it is reasonable to update the taxi-out time distribution and the parameters for the optimization on a quarterly basis. Despite the fact that we only conducted the optimization for a limited period of time in this study, the findings in this chapter still demonstrate that there is substantial benefit in airlines' long-term strategic planning from departure re-sequencing. The study in this chapter contributes to the aviation community by proving that among all the benefit objectives of the surface management tools in the pipeline, there is the additional possibility of using these tools to allow airlines to set shorter SBTs. How to adjust the one-month optimization and make it more practical for implementation is a promising direction for future research.

4. Operational Decision Modeling

In chapter 2 and 3, we focus mainly on the relationship between flight time predictability and the scheduled block time setting decision, which is major airline cost driver. In this chapter, another aspect of the airline decision, i.e. the operational decision that happens on a daily basis for the airlines is considered. In the scope of this study, we investigate an additional benefit mechanism from improved predictability—fuel savings, which is another major cost driver for airlines.

On first consideration, it may seem implausible that increasing predictability will affect fuel consumption. An automobile commute, after all, burns the same amount of gas in following a certain route with a certain speed profile, whether that speed profile is highly variable or very consistent from day to day. The difference between an automobile commute and a flight by an aircraft is the role of the loaded fuel. For both modes, it is not uncommon to arrive at the destination with extra loaded fuel in the tank as a form of contingency against risk. However, while for automobiles the weight of the tank is nearly negligible compared to the vehicle, fuel loaded to an aircraft accounts for a large portion of the total aircraft weight. Moreover, loading this fuel is not without penalty. Again unlike the automobile commute where the vehicle's fuel burn rate is quite stable, aircraft fuel burn is highly sensitive to aircraft weight. In summary, the fact that extra loaded fuel causes significant additional fuel burn differentiates a flight by aircraft from an automobile commute. Considering this, in aviation the decision of the amount of fuel loaded to the aircraft for an flight, termed fuel uplift, is a much more delicate economic tradeoff than automobile trips, where this is a simple matter of occasionally filling up the tank. On one hand, airlines must load a sufficient quantity of fuel for flights to avoid any risk of fuel exhaustion and to reduce the likelihood of a fuel-related diversion. In practice, this means that considerably more fuel is loaded than is likely to be burned, and thus that most flights land with considerable fuel in their tanks.

Even the most stingy and courageous flier would be willing to pay for some extra fuel to ensure that a flight has enough fuel on-board to complete its mission. Indeed, toward this end federal regulations stipulate minimum fuel reserves that must be boarded to each flight, and in some conditions also require sufficient fuel to fly to an alternate airport. In addition to reserve and alternate fuel, contingency fuel is also boarded. The amount of contingency fuel loaded is discretionary, and reflects the airline dispatcher's assessment of the "downside" risks that may lead to additional fuel burn beyond what is projected by the flight plan. Contingency fuel, together with the decision of carrying extra fuel to fly to an alternate airport when it is not required by federal regulation, thus represents the dispatcher's hedge against unpredictability. This, in combination with the effect of fuel uplift on fuel burn, suggests a connection between unpredictability and fuel consumption.

Recognizing the link between fuel consumption and predictability allows for both the monetization of predictability and the identification of new strategies for reducing aviation fuel consumption. There is intense focus on reducing fuel consumption from all stakeholders both to preserve the financial health of the airline industry and minimize environmental impact. Airlines are moving aggressively to reduce fuel consumption

because of rising fuel costs, which have gone from \$21 billion in 2009 to \$31 billion in 2012, and now account for 27% of airline operating costs, based on Bureau of Transportation Statistics data (BTS, 2009; 2012). Higher fuel costs force airlines to increase their ticket prices, which in turn suppresses demand. While fuel prices decreased in 2014, the price fluctuation only showcases the instability of the international fuel market. Consider that in 2008, jet fuel prices reached levels more than three times those of 2004, followed by a sharp decrease in 2009. Many interpret the current 2015 period as a “bust” and argue that the long term trend is still toward higher prices (The State of the Global Markets Report, 2015). In addition to issues with supply and prices, in the future, climate change policies and environmental attitudes of potential air travelers may further increase or destabilize effective fuel prices (Ryerson et al., 2013). Thus, by economizing on fuel airlines reduce their exposure to an economic “wild card.”

In this chapter, we investigate the relationship between flight time predictability and fuel consumption for a major US carrier. As discussed before, there is a direct physical relationship between fuel uplift and fuel consumption. Moreover, the fuel uplift determined by dispatchers may be affected by the flight time unpredictability. This might be the underlying mechanism to explain the results found by Ryerson et al. (2014) that when airlines add additional “buffer” time to flight schedules, fuel consumption increases. In this study, we exploit a large and recent flight-level dataset provided by a major US airline and merge this dataset with other publicly available datasets that incorporate NAS operating characteristics. The data included for each flight are the amount of loaded fuel, fuel burn and its rate, scheduled, planned, and actual flight times, and delays. We measure unpredictability of a flight by the standard deviation of airborne time among all the flights between a specific OD pair, departure time bank and month. This dataset enables us to estimate the relationship between unpredictability and fuel uplift, while controlling for other relevant factors such as terminal weather and traffic. Then by exploiting established relationships between fuel uplift and fuel consumption, we are able to evaluate the value of predictability in terms of cost savings from less consumed fuel.

4.1 Literature Review

To investigate the impacts of flight predictability on fuel loading or any other aspect of airline behavior, the first step is to define and measure flight predictability. The idea of reliability or (inversely) variability as an equivalent to predictability is not new in the field of ground transportation, where (un)reliability mainly refers to the unpredictable variations in travel time and is thus directly related to uncertainty of travel time (Carrion and Levinson, 2012). As a measure of travel time variability in ground transportation, most studies have used either the standard deviation or the average delay relative to scheduled arrival time (Börjesson et al., 2012), although some studies include both, while other studies use percentiles of the travel time distribution (Brownstone and Small, 2005). The most common approach for non-scheduled services with relatively high travel time variability, such as car trips and urban high-frequency transit trips, seems to be the so-called “mean-variance” approach, where the formulation (with a linear-additive form) contains only the mean and standard deviation of the travel time (Noland and Polak, 2002;

Hollander, 2006). Hollander (2006) further argues that “mean-variance” formulation may in practice not be able to capture the full disutility of travel time variability.

For low-frequency scheduled services with relatively low travel time variability, such as long-distance train or air trips, using the “average delay” as the variability measure seems to be the most common approach (Börjesson and Eliasson, 2011). By assuming that all travelers’ preferred arrival times are equal to the scheduled arrival time, the form expressing variability with “average delay” can be derived. Wardman (2001), followed by Abrantes and Wardman (2011) include a meta-analysis of British valuations of average delay. However, Börjesson and Eliasson (2011) argue that the value of “average delay” is not proportional to delay risk, suggesting average delay underestimates the value of unreliability, using survey results conducted on long-distance train passengers.

The analogy between travel time unreliability in ground and air transportation leads to the use of similar standard deviation metrics in air transportation, as employed in this chapter. In the realm of air transportation, the block time, commonly referred to as flight time, for a flight is the analogue for travel time in ground transportation. It is defined as the time between when an aircraft pushes back from the gate for the purpose of flight and when the aircraft comes to rest at the gate after landing. The block time can be further decomposed into taxi-out, airborne and taxi-in time. Among the three components, the airborne time is generally the largest component. In section 2.3 we use the different percentiles of block time as metrics for block time predictability. Block times percentiles are found to better predict scheduled block time than other metrics, including standard deviation.

Previous research on predictability valuation in an aviation context is fairly limited. Hansen et al. (2001) use factor analysis to devise a predictability metric whose cost impact is estimated by including the metric as an argument in an airline cost function. More recent studies, including the study in chapter 2 in this body of research, focus on the relationship between flight time dispersion and SBT, which is a major cost driver for airlines. Other mechanisms through which predictability generates value for airlines are more difficult to document and quantify. One approach is to identify airline planning and operating behaviors that both influence system cost and are influenced by predictability. In an effort to investigate how different components of delay impact fuel consumption, Ryerson et al. (2014) use simulated and actual airline fuel consumption data and find that fuel attributed to planned delays accounts for about 20% of the fuel that can be attributed to unplanned delays. Ryerson et al. (2014) conjecture that additional fuel loaded in the planning phase adds weight to the aircraft, thus increasing fuel consumption. The present study focuses on this effect.

Despite the important role fuel loading plays in fuel consumption, there is little literature specifically on fuel loading practice as a driver of airline fuel efficiency. Many other avenues for reducing fuel burn in the aviation sector have received far more attention. These include substituting all connecting flights in the US with non-stop flights (Jamin et al., 2004), substituting narrow body jets with turboprops (Ryerson and Hansen, 2010), and implementing Continuous Descent Approaches (CDA) and Airspace Flow Programs (AFP) aiming at coordinating ground and air operations (Clarke et al., 2004). There is

significant work investigating ground-based fuel savings measures such as single-engine taxi (Khadilkar et al., 2012) and delayed pushback procedures (Simaiakis et al., 2012).

4.2 Airline Fuel Loading Practice

The determination of fuel uplift for a specific flight is an important and safety-critical aspect of airline flight planning. During the flight planning phase, flight plans are created by dispatchers at the airline operations control center. There is a small body of literature regarding airline dispatch and flight planning; see Karisch et al. (2012) for a comprehensive look at the topic. In the following section, we provide a short background on flight planning and fuel loading.

As flight crew members on the ground, flight dispatchers perform a number of duties to ensure the safe operation of a flight from its origin to destination. Dispatchers prepare a flight plan for each flight, about two hours prior to departure. Each dispatcher typically works a 9 or 10 hour shift during which, for domestic dispatchers, about 40 flights will be planned. The flights for a particular dispatcher's desk are typically organized by geographic region. Dispatchers typically work the same desk, and thus the same set of flights, from day to day. In addition to flight planning, dispatchers perform other duties to ensure the safe operation of a flight from origin to destination. These include providing pilots with real-time updates, coordinating between various parties to resolve maintenance issues, and continuously monitoring the flight from takeoff to landing.

The information typically considered by dispatchers when planning flights – and loading the fuel – includes current and forecasted weather conditions at the origin, destination, and en route, restrictions or notifications from air traffic control, and specific flight routings. Each flight plan identifies characteristics of the flight, such as the trajectory from origin to destination, and critically for our study, the quantity of fuel to be loaded. The fuel uplifted is in quantities classified as mission fuel, reserve fuel, and discretionary fuel (sometimes termed contingency fuel by U.S. carriers); while in practice the fuel is uplifted as a single quantity, the classification of fuel into these three categories allows us to investigate any additional fuel uplift that could be reduced for fuel savings. Flight planning and dispatching is greatly different between flights within and outside the continental US (CONUS); the following discussion, and forthcoming analysis, will focus flights within the CONUS.

Mission and Reserve Fuel: U.S. Federal Aviation Regulations (14 C.F.R. § 91, E-CFR 2014) (FARs) require a domestic commercial flight to uplift enough fuel to complete the flight to the intended destination airport, miss the landing approach at the intended destination airport, fly from the destination airport to the alternate airport (if required by weather conditions), and hold in the air for 45 minutes at normal cruising speed (Federal Aviation Administration, 2008). This mandated fuel quantity is broken into two categories: mission fuel and reserve fuel. The mission fuel is calculated by the flight planning system (FPS) upon selection of a route by the dispatcher. The reserve fuel is the quantity of fuel an aircraft needs to fly for 45 minutes at normal cruising speed,

presumably to enter a holding pattern above either the destination airport or an alternate airport or to enter a holding pattern en route in the case of reduced airport or airspace capacity. The reserve fuel is not input by the dispatcher but rather calculated by FPS.

Alternate Fuel: The hold fuel, or alternate fuel, is the quantity of fuel that would be needed to fly from the destination airport to the alternate airport. The designation of an alternate airport, however, requires input from the dispatcher. An alternate airport is an airport in the general vicinity of the destination airport that will serve as the designated destination in the event of some flight disruption at the original destination, such as adverse weather, congestion, airport closures, etc. When an alternate airport is listed on a flight release, the FPS calculates the additional fuel needed to miss a landing approach at the original destination and then fly to the alternate airport.

If a dispatcher adds an alternate airport to a flight release, it is for one of two reasons. The first is that the designation of the alternate airport (and the fuel loading that it requires) is required by the FARs because of weather conditions. The FARs require a flight to carry enough fuel to travel to an alternate airport if the weather conditions are such that visibility is less than 3 miles and the ceiling at the destination airport (defined as the distance above the earth's surface of the lowest layer of clouds (e-CFR)) is less than 2000 feet at the flight's Estimated Time of Arrival (ETA) \pm 1 hour. The second reason a dispatcher might add an alternate to the flight release is that the dispatcher wants to provide extra buffer in the case where capacity is unexpectedly reduced and the flight needs additional fuel to complete its mission (or divert to the alternate). The dispatcher might have access to a different (and possibly internal) weather forecast or they might perceive there will be high levels of congestion when the flight in question enters the destination terminal airspace. In this case, the act of adding an alternate airport to a flight release is similar to adding contingency fuel, except the dispatcher chooses an alternate airport instead of a number of contingency fuel minutes.

Sometimes a second alternate airport is added to the flight plan, although this is never required by FARs. A dispatcher may add a second alternate if the first alternate airport is predicted to have marginal weather and the dispatcher does not feel fully secure by only having one alternate.

Contingency Fuel: Discretionary fuel, often termed contingency fuel in the U.S. (not to be confused with the portion of European required reserve fuel termed “contingency fuel”), may be uplifted onto a flight. Contingency fuel is, in essence, a reflection of expected operational degradation, or unpredictability. Contingency fuel is for use in the case of unexpected conditions during the flight, such as additional time on the ground or in the air from unexpected delays, re-routing, flight level changes or airborne holding. It is typically measured in minutes based on the fuel burn rate of the aircraft in normal cruise conditions. This allows dispatchers to think in terms of flight time uncertainty rather than fuel burn uncertainty. Airline policy may dictate a minimum amount of contingency fuel for a domestic flight (for example, 10-15 minutes) regardless of flight conditions, but dispatchers usually add more. As explained by Karisch et al. (2012), dispatchers may be presented with guidance regarding the historical distribution of actual fuel burn relative to planned fuel burn for similar flights to help them determine

contingency fuel . In practice, however, dispatchers uplift more contingency fuel than even the far right tail of this distribution would suggest, in order to provide extra protection against unforeseen circumstances. Airline management at our study airline, and anecdotally at many others, encourages dispatchers to reduce contingency fuel, since it is extremely rare for a flight to divert because of low fuel.

4.3 Modeling Fuel Loading Behavior

4.3.1 Data Collection

To estimate the impact of unpredictability on fuel uplift, data are collected from three sources: fuel and flight statistics from a major United States-based air carrier, weather information from the National Oceanic and Atmospheric Administration (NOAA), and airport data from the FAA Aviation system Performance Metrics (ASPM) database.

The US carrier used in our analysis operates an extensive domestic and international network serving all continents except Antarctica. The dataset provided from the airline includes all domestic flights between April 2012 and May 2013, inclusive. There are altogether 810,227 flights during the 14 months for which data is collected. The airline dataset contains flight-by-flight data on planned and actual fuel consumption, fuel uplift in all categories (taxi, contingency, and alternate) in units of minutes, as well as flight information such as equipment, origin and destination, planned and actual flight times (the so called OOOI times: Out from the gate, Off the origin runway, On the destination runway, and In at the destination gate), and delay information. It also provides actual fuel burn data from gate to gate.

The weather data collected from the NOAA database include both the actual weather and weather forecast (TAFs) information for major US airports. The actual and forecast weather information contains ceiling, visibility as well as indicators of the presence of thunderstorms, snow, and visibility conditions by hour, date, and airport. Visibility condition is expressed in terms of whether visual flight rules (VFR) or instrument flight rules (IFR) are in effect; these respectively indicate overall favorable or unfavorable terminal weather conditions. The weather data was matched with the flight-level airline dataset to recreate the conditions seen by dispatchers during the time of flight planning. Dispatchers typically make fuel loading decisions around the time the flight plan is created. Small deviations in this time do occur from flight to flight, but typically these decisions are made 2 hours prior to the flight's scheduled departure time. We will refer to this time as the dispatch time. We merge the weather data with the flight-level data to recreate the real-time and forecast weather that was available at the dispatch time for each flight. For the real-time weather, we find the actual weather at the origin and destination at the dispatch time. For the forecast weather, first we have to find the most recent forecast that was issued prior to the dispatch time and refer to the forecast conditions for the origin at the planned departure time and for the destination at the planned arrival time. This allows us to identify the following weather conditions for each flight: 1. the actual weather at the origin and destination airports two hours prior to the scheduled flight

departure time (as the actual weather at time of flight planning could influence dispatcher fueling decisions) and 2. the most recent forecasted weather (as forecasted two hours prior to the flight departure) for the origin and destination airports at the scheduled times of arrival and departure. These time frames for defining weather variables are consistent with the flight planning process revealed by our on-site observations and discussions with flight dispatchers.

The FAA ASPM database includes quarter hourly data for the 77 large airports in US on arrival traffic conditions. It contains average arrival delay for each quarter hour of day for each airport to depict level of congestion at the airport.

The three datasets are merged in a manner that will be described below. After merging and some filtering to keep the dataset robust that will be explained shortly, there are 448,660 flights in the dataset during the 13-month time period.

4.3.2 Estimation Methodology

We seek to statistically estimate the contribution of flight unpredictability to contingency fuel uplift. As noted above, in addition to loading contingency fuel per se, dispatchers sometimes add fuel by adding alternates. We therefore estimate two separate models: one with contingency fuel uplift as the dependent variable, and the other with contingency plus alternate fuel uplift as the dependent variable. As explained above, in both cases we express this fuel quantity in minutes, which is a common practice in fuel loading. We denote $CF(orig, dest, m, s, d, t_{orig}, t_{dest}, q, dis)$ as the contingency fuel uplifted on the individual flight from airport $orig$ to airport $dest$ on date d , in month m , with planned departure and arrival times t_{orig} and t_{dest} , planned by dispatcher dis working in shift s . The times t_{orig} and t_{dest} specify the hour and quarter-hour of planned departure and arrival. The empirical definition of predictability (discussed below) will also require including the month m and shift s when the flight departs as arguments. We divide the departure time of the flight into three shifts in a day: 5am to 3pm as the morning shift, 3pm to 10pm as the afternoon shift, and 10pm to 5am as the midnight shift, based on the actual shifts used for dispatchers' work schedules. To capture variation across dispatchers' fuel uplift practices we also index the identity of the dispatcher who planned this specific flight, denoted as dis . Similarly, we define variable $TOT(\cdot)$ with the same arguments as $CF(\cdot)$, which is the sum of contingency and alternate fuel, reflecting the total amount of fuel uplift for contingencies/unplanned events.

The variables $CF(\cdot)$ and $TOT(\cdot)$ are the dependent variables in our models. The independent variables will capture flight predictability and other variables that affect contingency and alternate fuel uplift, such as weather and traffic demand. As a metric for predictability, we use the standard deviation of actual airborne time for flights serving the same $orig-dest$ pair with a departure time in the same shift s taking place in the same month m . We segment flights in this manner based on interviews with the airline dispatchers who indicated that fuel loading judgment is most greatly impacted by month, shift, and $orig-dest$ pair. In contrast, the aircraft type and specific flight number are not specifically considered because most dispatchers do not pay attention to these details in fuel loading decisions. The calculation of the standard deviation of actual airborne time

takes place on these aggregated sets of flights segmented by m , s , $orig$, and $dest$. This is denoted as $stdair(orig, dest, m, s)$. To ensure robustness, only combinations of these arguments with more than 25 flights are kept in our dataset. (After applying this filter, our data set contains 7589 unique combinations of m , s , $orig$, and $dest$.) By examining the effect of dispersion in historical flight performance, we seek to capture the impact of the unpredictability in flight time on dispatcher uplift of contingency fuel and alternate fuel. We also include the associated mean value $avgdair(orig, dest, m, s)$ to capture uncertainty that arises from increasing average flight duration, which may result in less reliable forecasts of conditions en route and at the destination airport. Note that only airborne time is considered, because it is the largest component of the total flight time and an even more dominant source of fuel burn.

In addition to the dispersion of airborne time, the difference between actual airborne time and planned airborne is also a reflection of unpredictability. In everyday operation, the actual airborne time is sometimes different from the planned airborne time due to unforeseen en route conditions. Thus in this model, we include the mean and standard deviation of the difference in airborne time from flight plan. The variable $dif = actual\ airborne\ time - planned\ airborne\ time$ is calculated for each flight, and used to calculate the mean and standard deviation $avgdif(orig, dest, m, s)$ and $stdif(orig, dest, m, s)$.

We assume that dispatchers consider experience in the recent past when making fuel uplift decisions. Thus we specify our models so that decisions for flights taking place in month m are based on the above flight performance metrics for the previous month $m-1$.

As noted in section 4.2, the addition of an alternate on a flight plan, and the related fuel consumption, may be mandated by federal regulations due to certain weather conditions. In an effort to separate the impact of predictability and the presence of weather, we define variables to capture the weather conditions that trigger the mandated addition of an alternate. The weather information includes two parts: the actual and forecasted information for both the origin and destination airport. Even though dispatchers are only supposed to base their fuel loading decisions on the forecasted weather at the departure and arrival times of the flight, we discovered, from our observation and interview with dispatchers in the airline's control center, that they also consider current weather conditions at the time of their flight planning. Thus current weather information at the time the flight was planned is also included in the model. Moreover, the weather conditions at both the destination airport (denoted as $dest$) and the origin airport (denoted as $orig$) are included, since both may affect flight time and fuel burn.

For the forecasted weather at destination airport $dest$ (or origin airport $orig$), the hourly NOAA data is merged to the realized flight by the time t_a in which the flight is planned to arrive (depart) in the flight plan. Variable $lowc_F(dest, d, t_{dest})$ is 1 if the forecast ceiling at airport $dest$ in day d and time t_{dest} , is lower than 2000 feet and 0 otherwise. Similarly, $lowv_F(dest, d, t_{dest})$ is 1 if the forecasted visibility at airport $dest$ in day d and time t_{dest} is lower than 3 nautical miles and 0 otherwise. These criteria for low ceiling and low visibility are the thresholds specified in the Federal Aviation Regulations (FAR) for alternate requirements—FAR sections 121.619 and 135.223. To further

capture overall conditions, the variable $IFR(dest, d, t_{dest})$ is 1 if the airport $dest$ has IFR conditions at time t_{dest} of day d , and 0 otherwise. IFR stands for instrumental flight rules, which is an overall indicator of poor visibility for airport operations. It is also acquired from the NOAA weather dataset. The NOAA weather also provides dummy variables indicating snow and thunderstorms at the airport. Variable $snow_F(dest, d, t_{dest})$ ($Tstorm_F(dest, d, t_{dest})$) is 1 if snow (thunderstorms) is forecasted at airport $dest$ in day d and time t_{dest} , and 0 otherwise. Variables describing forecast weather at the origin airports are defined in a similar way, with origin airports denoted as $orig$, and based on the time t_{orig} . In addition to the forecasted weather condition, the actual weather conditions at dispatch time are also included. The variables are similar to the forecasted ones, but without the suffix $_F$. The actual weather variables are merged to the flights based on the assumed dispatch time—two hours prior to the planned departure time of the flight, i.e. $t_d - 2$.

Lastly, to capture the effect of congestion at the destination airport on contingency and alternate fuel uplift, we include a variable depicting arrival delay at the destination airport. Variable $Arr(dest, d, t_{dest})$ is the average arrival delay per flight in minutes at destination airport $dest$ in day d and time t_{dest} . The quarter hourly average arrival delay information is obtained from the ASPM dataset.

As the airline dataset includes flight-specific information regarding the dispatcher identity and there exists substantial variation in fuel loading behavior across dispatchers, fixed effects for individual dispatchers are included in the model. Each flight has a specific dispatcher dis , with the fixed effect captured in variable ρ_{dis} . We additionally define a fixed effect for each month. The purpose of this variable is to capture seasonality effects not captured explicitly by our terminal congestion and weather variables. For example, en route weather conditions are highly subject to seasonality and greatly influence fuel loading decision. However, the en route weather conditions are not captured in the terminal weather variables that we include in the model. The monthly fixed effects are denoted as γ_m .

Using the variables defined above, we assume a linear specification for the CF and TOT models and estimate it using OLS. The choice of a linear specification is mainly for convenience. Also, the linear model can be viewed as a first order approximation of a more general model around the mean values in the data. Other model specifications may be explored in future research.

As both models have the same specification, we present them using a single equation whose dependent variable, Y , may be either CF or TOT . The model formulations are as follows.

$$\begin{aligned}
Y(\text{orig}, \text{dest}, m, s, d, t_{\text{orig}}, t_{\text{dest}}, \text{dis}) = & \alpha_0 + \alpha_1 \times \text{avgair}(\text{orig}, \text{dest}, m-1, s) + \alpha_2 \times \text{stdair}(\text{orig}, \text{dest}, m-1, s) + \\
& \alpha_3 \times \text{avgdif}(\text{orig}, \text{dest}, m-1, s) + \alpha_4 \times \text{stdif}(\text{orig}, \text{dest}, m-1, s) + \\
& \sum_{i \in \{\text{dest}, \text{orig}\}} (\beta_{1i} \times \text{lowc}(i, d, t_i) + \beta_{2i} \times \text{lowv}(i, d, t_i) + \beta_{3i} \times \text{IFR}(i, d, t_i) + \beta_{4i} \times \text{snow}(i, d, t_i) + \beta_{5i} \times \text{tstorm}(i, d, t_i)) + \\
& \sum_{i \in \{\text{dest}, \text{orig}\}} (\lambda_{1i} \times \text{lowc_F}(i, d, t_i) + \lambda_{2i} \times \text{lowv_F}(i, d, t_i) + \lambda_{3i} \times \text{IFR_F}(i, d, t_i) + \lambda_{4i} \times \text{snow_F}(i, d, t_i) + \lambda_{5i} \times \text{tstorm_F}(i, d, t_i)) + \\
& \pi \times \text{Arr}(\text{dest}, d, q) + \rho_{\text{dis}} + \gamma_m + \varepsilon
\end{aligned} \tag{4.1}$$

4.3.3 Estimation Results

The estimation results are shown in Table 4.1. Note that because our assumption is that the fuel uplift is impacted by the flight performance data of the previous month, the model is estimated on 13 out of the 14 months covered by our data set. The left portion of Table 4.1 shows the estimation results for Equation 4.1 where $Y=CF$, meaning that the dependent variable is contingency fuel in minutes. The top part of the table shows the intercept and the predictability metrics. Average airborne time over the past month has a small positive coefficient, indicating that longer flights are loaded with slightly more contingency fuel than shorter flights—about 1.5 minutes more for a 100 minute increase. The coefficient on the standard deviation of airborne time, *stdair*, is 0.88, indicating that one minute of variation in the airborne time in previous month will lead to the uplift of an average of almost one minute of additional contingency fuel. This indicates that an increase in the unpredictability of the daily flight operation, based on our metric, will lead to an increase in contingency fuel uplift at an almost one- to-one ratio. The deviations captured in variables *avgdif* and *stdif* both have significant coefficients as well. Variable *avgdif* has a small negative coefficient, indicating that an increase in the average disparity between actual and planned flight time slightly decreases fuel uplift. A possible explanation of this seemingly counter-intuitive result is that when there is en route convective weather, planned routings are very conservative, often resulting in time-saving reroutes during the flight. Thus *avgdif* may be capturing the degree to which a flight is subject to en route convective weather, with a lower value indicating that a flight is more subjective. The positive coefficient on *stdif* could also reflect this, as well as other sources of uncertainty that make some flights more difficult to be planned accurately than others. In any case, the most important message from these results is that dispatchers fueling decisions are much more strongly influenced by the day-to-day variability in flight time, as measured by *stdair*, than by patterns of discrepancy between planned and actual times.

The middle part of Table 4.1 shows the estimates for the weather variables, separated by destination and origin airport. Most weather dummy variables have a positive coefficient, indicating that adverse weather at the airports will increase contingency fuel uplift. Thunderstorms have the largest impact and low ceiling the smallest. Overall, forecasted weather has a larger impact on the uplift of contingency fuel than the actual weather, indicating that dispatchers' fueling decisions are more affected by the weather forecast than the actual weather at time of flight planning. However the actual weather conditions still have significant effects, validating our observation on dispatcher behavior. Moreover, the weather conditions at the destination airport, both actual and forecasted, are greater

contributors to contingency fuel than the weather conditions at the origin airport. As the arrival is the more unpredictable phase of the flight and could induce a large amount of fuel consumption from holding or rerouting due to adverse weather, these results are consistent with our expectations. Lastly, regarding the traffic conditions at the destination airport, a 1-minute increase in the average arrival delay at the destination will lead to a 0.14 minute increase in the contingency fuel loaded. Arrival delay can derive from arrival queuing at the destination airport, which may lead to airborne holding, but it may also derive from gate delay at the origin, which has little impact on fuel burn. This probably explains why the arrival delay coefficient is considerably less than 1.

Table 4.1 Estimation Results for Fuel Uplift Models

<i>Variable Name</i>	<i>Variable Abbreviation</i>	Contingency min		Total min	
		<i>Estimate</i>	<i>P-value</i>	<i>Estimate</i>	<i>P-value</i>
Intercept		20.601	<.0001	20.883	<.0001
Average Airborne Time	avgair	0.0153	<.0001	0.0103	<.0001
Standard Deviation of Airborne Time	stdair	0.883	<.0001	1.657	<.0001
Average Difference in Airborne Time from Flight Plan	avgdif	-0.0279	<.0001	0.0938	<.0001
Standard Deviation of the Difference in Airborne Time from Flight Plan	stdif	0.0096	<.0001	0.0361	<.0001
Low Ceiling Indicator at Destination Airport	lowc_dest	-0.0377	0.7191	10.399	<.0001
Forecasted Low Ceiling Indicator at Destination Airport	lowc_dest_forecast	1.686	<.0001	14.057	<.0001
Low Visibility Indicator at Destination Airport	lowv_dest	1.858	<.0001	16.041	<.0001
Forecasted Low Visibility Indicator at Destination Airport	lowv_dest_forecast	3.483	<.0001	24.774	<.0001
Thunderstorm Indicator at Destination Airport	tstorm_dest	3.267	<.0001	22.457	<.0001
Forecasted Thunderstorm Indicator at Destination Airport	tstorm_dest_forecast	8.875	<.0001	51.952	<.0001
Snow Indicator at Destination Airport	snow_dest	1.0046	<.0001	11.896	<.0001
Forecasted Snow Indicator at Destination Airport	snow_dest_forecast	3.322	<.0001	26.517	<.0001
IFR Condition Indicator at Destination Airport	IFR_dest	0.658	<.0001	16.886	<.0001
Average Quarter Hourly Arrival Delay at Destination Airport	Arrival delay_dest	0.136	<.0001	0.514	<.0001

Low Ceiling Indicator at Origin Airport	lowc_ori	-0.145	0.1785	-0.0621	0.8434
Forecasted Low Ceiling Indicator at Origin Airport	lowc_ori_forecast	0.0527	0.6710	-0.672	0.0635
Low Ceiling Indicator at Origin Airport	lowv_ori	0.470	0.0002	0.175	0.6297
Forecasted Low Ceiling Indicator at Origin Airport	lowv_ori_forecast	0.793	<.0001	2.023	<.0001
Low Visibility Indicator at Origin Airport	tstorm_ori	1.625	<.0001	1.824	0.0188
Forecasted Low Visibility Indicator at Origin Airport	tstorm_ori_forecast	5.627	<.0001	8.381	<.0001
Snow Indicator at Origin Airport	snow_ori	0.734	0.0002	-1.135	0.0497
Forecasted Snow Indicator at Origin Airport	snow_ori_forecast	1.263	<.0001	0.0933	0.8536
IFR Condition Indicator at Origin Airport	IFR_ori	0.643	<.0001	1.997	<.0001
January Dummy	month 1	-1.279	<.0001	-4.679	<.0001
February Dummy	month 2	-0.631	<.0001	-3.080	<.0001
March Dummy	month 3	-1.219	<.0001	-8.636	<.0001
April Dummy	month 4	1.018	<.0001	-1.784	<.0001
May Dummy	month 5	1.068	<.0001	1.633	<.0001
June Dummy	month 6	2.037	<.0001	2.811	<.0001
July Dummy	month 7	5.079	<.0001	14.467	<.0001
August Dummy	month 8	3.559	<.0001	9.974	<.0001
September Dummy	month 9	1.9999	<.0001	3.799	<.0001
October Dummy	month 10	0.912	<.0001	0.131	0.6888
November Dummy	month 11	-1.785	<.0001	-8.920	<.0001
December Dummy	month 12	0.000	-	0.000	
R-square		0.2623		0.4163	

The right-hand columns of Table 4.1 include the estimation results using the sum of contingency minutes and the alternate fuel in minutes as the dependent variable. The estimation results are generally similar to the contingency fuel model on the left-hand side. The intercept of 20 minutes is similar to the intercept from the contingency fuel model. Since most adverse weather conditions are considered, the intercept depicts a typical amount of fuel uplift for a good weather day when alternates are not required. Therefore, it is reasonable to see a similar intercept for the two models. Most other coefficients are larger than the *CF* model. The coefficient estimate for standard deviation of airborne time is twice as large as the coefficient from the contingency-only model, implying that dispatchers uplift approximately 1.66 minutes more fuel for a 1-minute increase in the standard deviation of airborne time. This suggests that alternate fuel is a major component of dispatchers' hedge against uncertainty. Again the impact of airborne time unpredictability is much larger than the impact of flight plan unreliability, the mean and standard deviation of which now both have positive, albeit very small, impacts.

Since the dependent variable now includes alternate fuel, the weather conditions at the destination airport have a more dominant impact. The coefficients for all the destination airport weather variables are significant, for both actual and forecasted weather, and are around 9 times larger in magnitude than the corresponding coefficients from the contingency-only model. For origin airports, the coefficients are similar to those in the contingency-only model, and some variables are not statistically significant, such as real-time low visibility and low ceiling.

The bottom part of Table 4.1 shows the monthly fixed effects. For ease of presentation the estimates of the dispatcher fixed effects are excluded. December is chosen as the baseline month to which all other months are compared. We see that summer months, especially July and August, have larger coefficients than other months. These two months are during thunderstorm season, which greatly impacts contingency and alternate fuel uplift. Although the adverse weather variables, in particular the thunderstorm variables, are included to account for the impact of convective weather, there are still effects from en route thunderstorms and other weather conditions that vary seasonally which cannot be captured by the airport weather variables. The absolute difference in contingency fuel due to monthly fixed effects is almost 7 minutes, with the largest fuel load in July and the smallest in November. If we consider alternate fuel as well, the scale is much larger, ranging from -9 to 15 minutes for November and July, respectively, but the trend is much the same.

4.4 Cost-to-Carry Analysis

The previous section established that unpredictability in airborne time will lead to an increase in fuel uplift. The significance of this additional uplift is both financial and environmental. The rate of fuel consumption increases with weight; said another way, you spend fuel to carry fuel. There is a measurable cost to carry, or the pounds of fuel consumed per pound of fuel carried per mile. This rate varies across aircraft types and flight lengths with a general rule of thumb being that it costs about one-quarter to one half-pound of fuel to carry a pound of fuel (Leigh, 1995). There is therefore an additional amount of fuel consumed that can be attributed to the additional contingency and alternate fuel uplifted as a result of unpredictability. In this section, we quantify this added fuel and then translate it into fuel consumption, and then into costs in terms of purchase expense and emissions of carbon dioxide (CO₂), the most abundant of the Greenhouse Gas (GHG) contributing to climate change.

In an aviation system with no operational unpredictability, the standard deviation in airborne time would be zero. It follows that the coefficient α_2 on the variable $\text{stdair}(\text{od}, \text{m}, \text{s})$ represents the fuel penalty of unpredictability. For the 448,660 realized flights during the 13 months for which data is collected, there are 7,589 groups of od_month_shift . For each flight observation, we calculate $\alpha_2 * \text{stdair}(\text{od}, \text{m} - 1, \text{s})$ as the contribution of variation in airborne time to the loaded fuel of this flight. The average contribution of the total 448,660 flights is then calculated. In a perfect scenario where no variation in airborne time exists, the loaded contingency fuel would be reduced by 6.12

minutes per flight. If we consider the sum of contingency fuel and alternate fuel, the reduction would be 11.28 minutes per flight. Again the loaded fuel is in the unit of minutes, which is a common practice in flight planning. While a perfect predictability scenario might include other differences, such as the elimination of differences between planned and actual flight times, we ignore those here because their impacts are small compared to that of standard deviation.

To estimate the savings in cost to carry from perfect predictability, we translate the excess minutes of contingency and alternate fuel of each individual flight into pounds of fuel using the fuel consumption per minute rates provided by the airline. These rates are specific to a particular flight, based on information in the flight plan such as equipment type and weather conditions. We next translate this into a quantity of fuel burned due to the loading of additional fuel using the airline cost to carry rates in units of lb per lb per mile. The results are presented in Table 4.2 in four categories. The first is the average fuel consumed (in lbs) due to additional uplift per operation. The second is the total amount of additional fuel consumed (in lbs) over the entire set of flights for which data are available. Due to reporting difficulties and the manual method some aircraft require for fuel reporting, the master airline dataset covers about 80% of the total operations. As such, the third category is the total amount of additional fuel consumed, in lbs, across the airline. We collected monthly domestic flight counts from the BTS T-100 Segment Database and extrapolated our results to these monthly counts. The fourth category is the total amount of additional fuel consumed (in lbs) across all airlines for all domestic operations. We collected domestic operational counts for all US carriers (those with \$20M or more in revenue per year) from the BTS T-100 Segment Database and extrapolated our results to this operational count again on a monthly basis.

The fuel quantified in the first two rows of Table 4.2 can be translated into airline monetary cost and environmental externalities of fuel consumption. Reducing fuel consumption is a major initiative of the aviation industry as a whole. It is a way to reduce costs, and environmental impacts particularly related to climate change, manage the risk related to fuel price fluctuations and uncertainty surrounding a future environmental policy, and improve consumer perceptions of “greenness”. As such there are many initiatives being considered in the form of policies, operational changes, and technology deployments. These ranges come from airline driven changes such as emphasizing single engine taxi procedures (Simaiakis and Balakrishnan, 2010; Nikoleris et al., 2011) and the federally-driven Next Generation Air Transportation System (NextGen) which promises significant fuel consumption reduction. We can translate the fuel savings from increased predictability into costs in terms of fuel prices and CO₂ emissions. As fuel prices fluctuate throughout the year and airlines have their own fuel contracts that may change the fuel cost they see, we estimate the cost to carry for three fuel prices: the average for the study year, about \$3.00/gallon; and plus/minus \$1.00/gallon (Airlines for America, 2013). To convert excess fuel into lbs of CO₂, we utilize the U.S. Environmental Protection Agency conversion factor for Jet Fuel (EPA, 2013a). The benefit results are presented in Table 4.2 below.

Table 4.2 Annual Cost to Carry Additional Contingency and Alternate Fuel from for Unpredictability in Terms of Fuel lbs, Fuel cost (*: in millions), and CO2.

Metric		Mean per Operation	Sum over Operations in the Dataset	Sum Extrapolated Over the Airline	Extrapolated over all Domestic Operations
Fuel (lbs)	Contingency	48.35	1.07×10^7	3.56×10^7	4.04×10^8
	Contingency and Alternate	90.73	2.01×10^7	6.69×10^7	7.58×10^8
CO2 (lbs)	Contingency	155.11	3.44×10^7	1.14×10^8	1.30×10^9
	Contingency and Alternate	291.07	6.46×10^8	2.14×10^8	2.43×10^9
Fuel cost at \$2/gallon	Contingency	\$14.43	\$3.20*	\$10.64*	\$120.55*
	Contingency and Alternate	\$27.08	\$14.41*	\$19.96*	\$226.22*
Fuel cost at \$3/gallon	Contingency	\$21.65	\$11.07*	\$15.95*	\$180.82*
	Contingency and Alternate	\$40.62	\$21.62*	\$29.94*	\$339.33*
Fuel cost at \$4/gallon	Contingency	\$28.86	\$14.76*	\$21.27*	\$241.10*
	Contingency and Alternate	\$54.17	\$28.82*	\$39.91*	\$452.43*

The results in Table 4.2 provide us with the value, in terms of monetary costs and environmental externalities, of predictability. On a per flight basis, this value is \$14.43 - \$54.17 depending on fuel prices and whether alternate fuel is considered. Across all domestic flights, this value ranges from \$120.55 – \$452.43 million per year. To put these results in perspective, we first consider that in 2011 (a close proxy for our time frame) the total amount of jet fuel consumed was 12.1 billion gallons (EPA, 2013b). Therefore, the total amount of fuel consumed due to the lack of predictability in the system is about 1%. One percent may seem like a small number, however, it is in line with current initiatives branded as fuel saving “green” initiatives. Consider that the 1% translates to about 50-100 lbs per flight burned due to excess uplift (Table 4.2). These values are also comparable to the savings estimated from use a continuous descent approach as compared to a conventional step-down approach (Cao, et al., 2011).

There are numerous efforts and research taking place to reduce fuel consumption on the ground and during descent and these involve investment, institutional change management, and the cooperation of federal, state, local, and private stakeholders as well as the traveling public (Cao, et al., 2011). Our results show that efforts to increase predictability should be included under the same umbrella. Additionally, reductions in fuel uplift and associated cost to carry, both monetary and environmental, should be included in the benefit assessments of programs such as NEXTGEN that are likely to improve predictability by making the NAS less vulnerable to adverse conditions.

4.5 Conclusion

In this chapter, the relationship between flight predictability and contingency fuel loading, which is a major airline operational decision, is studied using detailed empirical data. Flight predictability is mainly depicted by the variation of airborne time because that most strongly affects dispatchers' decision on contingency fuel loading. Flights are grouped by OD pair, departure time of day and month based on our observation of dispatchers' fuel loading behavior. We found 1 minute of standard deviation of airborne time would lead to an additional 0.88 minutes of contingency fuel loaded to each flight. If we also consider alternate fuel, there would be a 1.66 minutes increase in the sum of contingency and alternate fuel for a 1 minute increase in standard deviation of airborne time. Our other findings show that the deviation from planned flight time has a much smaller but significant impact on loaded fuel than overall airborne time variability, indicating dispatchers are more influenced by overall variability rather than flight plan accuracy; the forecasted weather at the destination airport is the most influential weather factor for fuel loading, among which thunderstorm is the largest contributor, and low ceiling is the smallest. Also, there are significant seasonal fixed effects, with dispatchers loading fuel more fuel in the summer (July and August), probably to account for en-route thunderstorms and other weather factors not included in our model.

To further quantify the impact of flight predictability on fuel loading, we calculate the value of flight predictability assuming a hypothetical scenario where there is no variability of airborne time for flights in a given OD pair, shift, and month. If the standard deviation of airborne time for all the flights are zero, on average there is a reduction of 6.12 minute per flight of contingency fuel, and 11.28 minute per flight of the sum of contingency and alternate fuel. This extra boarded fuel requires additional fuel to carry. Based on our calculation, for an average flight 48.35 lbs of fuel is consumed to carry the extra contingency fuel and 90.73 lbs to carry the extra contingency and alternate fuel that results from flight time variability. This translates into a cost to US domestic airlines on the order of \$120.55 – \$452.43 million per year. Social costs from additional emissions of GHG, not explicitly estimated here, add to this total value. Of course, it is not realistic to assume that all variability in flight times can be eliminated. The figures presented should be viewed as a heretofore unrecognized potential benefit from increasing predictability. Individual projects large or small that improve predictability can tap into this potential benefit, and in some cases this may tip the balance for the project business case. For purposes of comparison, a recent study estimates that flight

delay, which likewise can never be eliminated but is the major motivation for NEXTGEN, costs the US airline industry around \$10 billion per year (Zou and Hansen, 2012).

Our analysis establishes a behavioral link between flight time variability and fuel loading. Given that link, we have estimated the potential cost and emission savings from reducing variability. With our analysis, the FAA's proposed study on flight predictability is provided with a more complete benefit motivation. The improvement in flight operational predictability will benefit not only the operational performance, but also the airlines' long-term fuel cost, which is rather the strategic planning aspect. Neglecting the less obvious benefit manifested in this study would lead to underestimating benefit in improving predictability. Lastly, another way to attain these savings is to change dispatcher behavior, since it is not obvious that there is a sound operational reason to load more contingency fuel because flight time varies, since much of this variation is captured in the flight plan.

5. Conclusion

5.1 Contributions

This body of research presents a methodology to quantify the impact of flight predictability on airlines' strategic and operational decisions. It also proposes a methodology to improve predictability through managing the aircraft queue at the departure airport, the benefit of which is reflected mainly on the saving in the airlines' strategic decisions. The strategic decision that we consider is the scheduled block time, which is typically decided six months before the travel date. On the operational side, we investigate the decision of the amount of fuel loaded to each flight in the daily operation. Our concept of predictability is based on the variability of flight times and the different components of flight time.

In this research we develop two sets of empirical models for the two types of airline decisions. The dependent variables are the different decisions made by the airlines: the scheduled length of the flight (scheduled block time, SBT) and the amount of fuel to be loaded. To explore the impact of flight predictability on these decisions, the performance of historical flight times is measured and included as explanatory variables in the two sets of models. Given the different natures of the two types of decisions, the quantification of the historical flight time performance is also different for the two models.

For the SBT setting model, we develop three empirical models to capture this behavior. The most accurate model among the three is the "percentile model" that segments the distribution of historical flight travel time (block time) in great detail. Flight predictability is captured by increments between every 10th percentile above the 50th of historical flight time, by different flight phase. For the two components of the block time, airlines consistently have a skewed focus on the left and inner right tail of the distribution and almost neglect the far right tail (beyond the 80th percentile). In general, airlines are willing to experience occasional severe delays in exchange for a shorter SBT. This represents their "optimistic" behavior in schedule design which is largely driven by the profit-seeking nature of the airline business. Notable results from the three empirical models include that historical gate delay is virtually ignored, that airlines with hubs tend to set shorter SBTs for their hub-bound flights, that competition encourages longer SBTs, and that airlines adjust their SBTs in response to the flight's historical on-time performance, as well as the persistent earliness and lateness, with earliness having a stronger effect. Also, the impact analysis with the airline performance data demonstrates that changes in block time distributions—in particular the median and inner right tail—often cause significant adjustments in SBTs, and that the impacts of these adjustments on schedule adherence is greater than the changes in the underlying operational performance. This leads to the next piece of this research, where an algorithm is proposed to improve predictability with reducing SBT as one of the objectives in a practical setting.

In this thesis, we present algorithms and practical rules to sequence departures at the airport to adjust historical flight time distribution and improve flight predictability on a daily basis. The mechanism is that the existing wheels-off times for the flights in a certain

day on a certain airport form a slot pool. In the optimization, we reassign these flights into the wheels-off slots, based on a certain objective and constraints that assure the feasibility of the assignment. The objective function is expressed in the form of the “cost of the assignment”. SBT is viewed as a cost for the airlines to be minimized. For each flight, the assigned slot generates a new taxi-out time and this time contributes proportionally to the future SBT, based on results from the percentile model. In addition to reducing SBT, assuring that the on-time performance for the flights is not greatly sacrificed is another important aspect of the optimization. Therefore a second piece in the objective function is added to include on-time arrival performance as part of the “cost of assignment” as well. In this way, we develop a multi-objective algorithm to sequence departure flights to improve predictability, reduce airline scheduled block time, while maintaining on-time performance. Through this real-time sequencing decision, we observe that future SBT can be reduced about half a minute, on a per flight basis. If the on-time performance consideration is added to the objective function, the reduction in scheduled block time is compromised to some extent, but in return the on-time performance of the flights can also be improved. Overall, our results for this part of the research show that there are opportunities to reduce scheduled block times and increase on-time arrival performance through re-sequencing airport departure queue, but the magnitude of the potential gains is modest. Further research is required to compare the value of the gains with the costs of implementing a re-sequencing process through which they could be realized.

Finally, the relationship between flight predictability and another aspect of the airline decision making—fuel loading on the day of operation—is investigated using empirical modeling and the benefit of predictability is quantified through constructing hypothetical scenarios. The dependent variable being modeled is the amount of contingency and alternate airport fuel loaded on the flight. We depict flight predictability by the variation of airborne time because that most strongly affects dispatchers’ decision on contingency fuel loading. We found one minute of standard deviation of airborne time would lead to about 0.88-1.66 minutes of additional contingency and alternate fuel loaded by airline dispatchers. We also find that the deviation from planned flight time has a much smaller but significant impact on uplifted fuel than overall airborne time variability, indicating the influence of (un)predictability on dispatchers is higher than that of flight plan accuracy. We calculate the value of flight predictability assuming a hypothetical scenario where there is no variability of airborne time for all flights. Under this scenario, on average there is a reduction of 6-11 minutes per flight of loaded fuel. This extra uplifted fuel requires additional fuel to carry and it translates into a cost to US domestic airlines on the order of \$121-\$452 million per year. Social costs from additional emissions of GHG, not explicitly estimated here, add to this total cost. This research provides one of the first analyses of the link between flight predictability and fuel consumption, and opens a new research frontier on the fuel loading and fuel management process.

In this research, we present two different mechanisms of how flight predictability is quantified and how it affects airline decision. Historical flight time performance data are utilized for the quantification of flight predictability. Viewing the two sets of models focusing on different aspects of the airline operation from a higher level, we gain a more systematic insight about how to best quantify predictability to measure its impact on

airline decisions. In chapter 2, the standard deviation of historical block time is shown not to be a good explanatory variable to explain the airline SBT setting behavior. The percentile based segmentation of the distribution is a better measure of predictability and captures the stronger effect of the “inner right tail” than the far right tail. Rooted in this “optimistic” behavior is the airlines’ profit-driven motivation, from which they tolerate delay for more efficient SBT. On the other hand, in chapter 4, the standard deviation of historical block time is actually an accurate measure to depict the airline fuel loading behavior. This means that the airlines’ fuel loading behavior considers the extreme cases that happened in the past and it shows the airlines’ safety concerns, which is another important aspect of this business. Viewing these two types of decisions and their relationships with predictability, we should be aware that when facing the complicated nature of airline operation with conflicting drivers, there is no single predictability metric that is appropriate for all situations. Different metrics are required for different contexts within the whole value chain of airline business.

For the two different mechanisms, this research also takes a system-level view to quantify the potential benefits of predictability. With our analysis, the decision in the aviation community to target flight predictability as a performance goal is provided with a more complete benefit motivation. For strategic decisions, we sought to understand the connection between SBT and the historical flight time distribution. With the behavioral link, this study contributes to the existing literature by demonstrating the benefit of predictability in the form of reduced future flight SBT through airport surface queue management, without much loss in other performance metrics. On the operational decision side, our analysis establishes a behavioral link between flight time variability and fuel loading. Given that link, we have estimated the potential pool of fuel cost and emission savings from improved predictability. Our models will help decision makers understand the impact of predictability in multiple areas such as airline scheduling, fuel cost and environmental externalities. This research has identified such benefit mechanisms and quantified their potentials. From this research, we conclude that the increased emphasis on predictability as a dimension of aviation system performance is well founded, and that improvements in predictability, like reductions in delay, can be monetized.

5.2 Future Work

The research presented in this thesis can be extended in many directions. For the SBT setting models, the role of gate delay in the decision process is worth further investigation. For the airport surface management algorithm, the scale of the optimization can be extended to longer time periods, as well as more airports nationwide. Moreover, alternative forms of the “cost of assignment” can be utilized to improve the practicability of the optimization. For the airline fuel loading study, more advanced cost-to-carry analysis than simply applying airline conversion rate (such as flight fuel burn simulation) can be applied. It is also worth further investigation regarding how to effectively change dispatcher behavior to reduce excess fuel loading.

In the behavior models in section 2 we found that airlines virtually ignore historical gate delay in setting SBTs. On the other hand, gate delay is the dominant source of variation in effective flight time (observed in section 1.2). In future research it may be desirable to decompose gate delay by cause—later arriving aircraft, ground delay programs, etc.—and analyze the effects separately. Also, it may be possible to improve on-time performance significantly by adjusting SBT to reflect expected gate delay, again possibly disaggregated by cause, to the extent it can be predicted from past experience.

The optimization in chapter 3 is conducted only over a period of one month and only on three airports. A natural extension of this research is to apply this algorithm to re-sequence departure queues over a longer time period, and on more airports. As mentioned in section 3.4, the benefit in reduced SBT revealed in chapter 3 can only be achieved if the optimization is conducted consistently for years, generating changed distribution of historical block time. Thus extending the study temporally will relax the one month constraint and help observe the benefit more realistically. Moreover, understanding and comparing the benefit of re-sequencing on multiple airports in US will provide a holistic view and help prioritize airports for improving predictability. Also, it would be interesting to understand the relationship between magnitude of benefit and other airport characteristics to build a model that better predicts benefits of predictability on the airport level in the future.

The formulation of the “cost of assignment” identified in chapter 3 does a good job in accurately optimize the contribution of actual taxi-out time to future SBT. However, the calculation of the objective function depends too much on the estimation results from the SBT percentile model. The percentile model uses only three years of historical performance data; therefore using the exact coefficients of the percentile model might not be generalized enough for the daily optimization, especially for future time periods. Therefore, another future research topic is developing an objective function that takes advantage of the trends along the historical taxi-out time distribution reflected by the percentile model (e.g. inner right tail matters, far right tail has little impact on future SBT setting), without relying too much on the specific estimation results. Such research would help the optimization to rely less on the estimation results of the percentile model, and therefore to be easier to implement in the airport daily operation and less subject to estimation inaccuracies.

The cost-to-carry analysis in chapter 4 directly applies the airline’s fuel burn rate when translating additional loaded fuel due to lack of predictability to fuel consumption. This is an accurate methodology for processing the airline specific dataset. However, different airlines might have different aircraft configuration and thus different fuel burn rate, which makes the analysis in this study less convincing when extrapolated to all domestic operations. One potential improvement is to use simulation models to directly estimate cost-to-carry for different aircraft types. Another potential research path is to focus on practical ways to achieve these savings. One way is to change dispatcher behavior. While it is sensible to increase fuel uplift in response to flight time variability, many believe that overall dispatchers load too much contingency fuel and sometimes include alternates that are not called for. This may include fuel related or unrelated to unpredictability, measured as flight time variability in this study. Whether, and how, it may be possible to

persuade dispatchers to attenuate the influence of flight time variability in their fuel loading decisions, or reduce fuel loads motivated by other factors, is another subject of future research.

6. Reference

- Abrantes, P., Wardman, M., 2011. Meta-analysis of UK Values of Travel Time: an Update. *Transportation Research Part A: Policy and Practice*, 45 (1), 1–17.
- ADFMS report, Airport Departure Flow Management System Final Report, 2010.
- Airlines for America, Annual Crude Oil and Jet Fuel Prices. <http://www.airlines.org/Pages/Annual-Crude-Oil-and-Jet-Fuel-Prices.aspx>. Last accessed, July 2013.
- Ball, M., R.L. Hoffman, D. Knorr, J. Wetherly and M. Wambsganss, 2000. Assessing the Benefits of Collaborative Decision Making in Air Traffic Management. *3rd USA/Europe Air Traffic Management Research and Development Seminar*, Napoli, Italy.
- Bates, J., J. Polak, P. Jones, and A. J. Cook, 2001. The Valuation of Reliability for Personal Travel. *Transportation Research Part E: Logistics and Transportation Review*, 37(2), 191-229.
- Bazargan, M., 2010. Airline Operations and Scheduling.
- Bolczak, C. N., J. H. Hoffman, A. J. Jensen, and W. W. Trigeiro, 1997. National Airspace System Performance Measurement: Overview. MITRE Technical Report W0000035, The MITRE Corporation, McLean, VA, USA.
- Börjesson, M., Eliasson, J., 2011. On the Use of “Average Delay” as a Measure of Train Reliability. *Transportation Research Part A: Policy and Practice*, 45 (3), 171-184.
- Börjesson, M., Eliasson, J., Franklin, J.P., 2012. Valuations of Travel Time Variability in Scheduling versus Mean-variance Models. *Transportation Research Part B: Methodological*, 46 (7), 855-873.
- Borokhov, P., S. Blandin, S. Samaranayake, O. Goldschmidt, and A. Bayen, 2011. An Adaptive Routing System for Location-aware Mobile Devices on the Road Network. *14th IEEE Conference on Intelligent Transportation Systems*, Washington, D.C, USA.
- Brinton, C., Provan, C., Lent, S., Prevost, T., Passmore, S., 2011. Collaborative Departure Queue Management. *9th USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.
- Brownstone, D., Small, K., 2005. Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. *Transportation Research Part A: Policy and Practice*, 39 (4), 279-293.

Bureau of Transportation Statistics (BTS). <http://www.transtats.bts.gov/>

BTS 2007, Table 2: Ten Longest Average Taxi-Out Times by Airport in 2007. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/special_reports_and_issue_briefs/special_report/2008_008/html/table_02.html. Last accessed, March 2015.

Cao Y., T. Kotegawa, J. Post, 2011. Evaluation of Continuous Descent Approach as a Standard Terminal Airspace Operation. *9th USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.

Carrion, C., Levinson, D., 2012. Value of Travel Time Reliability: A Review of Current Evidence. *Transportation Research Part A: Policy and Practice*, 46(4), 720-741.

Chiraphadhanakul, V. and C. Barnhart, 2011. Robust Flight Schedules through Slack Re-Allocation. Paper submitted to Anna Valicek Medal 2011. www.agifors.org/award/submissions2011/VirotChiraphadhanakul.pdf. Last accessed, July 2012.

Clarke J. P., N. T. Ho, R. Liling, J. A. Brown, K. R. Elmer, K. Tong, et al., 2004. Continuous Descent Approach: Design and Flight Test for Louisville International Airport. *Journal of Aircraft*, 41 (5), 1054-1066.

Ciuperca, G. and V. Girardin, 2005. On the Estimation of the Entropy Rate of Finite Markov Chains. *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis*, Brest, France.

Cohen, H. and F. Southworth, 1999. On the Measurement and Valuation of Travel Time Variability due to Incidents on Freeways. *Journal of Transportation Statistics*, 2(2), 123-131.

Cover, T., and J. Thomas, 1991. *Elements of Information Theory*. John Wiley and Sons, Inc., New York, NY, USA.

Coy, S., 2006. A Global Model for Estimating the Block Time of Commercial Passenger Aircraft. *Journal of Air Transport Management*, 12(6), 300-305.

Deshpande, V. and M. Arikan, 2012. The Impact of Airline Flight Schedules on Flight Delays. *Manufacturing and Service Operations Management*, 14(3), 423-440.

Electronic Code of Federal Regulations, Title 14: Aeronautics and Space, section 1.1. <http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=91ac073c1ab333fde1e47f9d87fa244f&rgn=div8&view=text&node=14:1.0.1.1.0.1.1&idno=14>. Last accessed, July 2012.

- EPA, 2013a. Emission Factors for Greenhouse Gas Inventories, US Environmental Protection Agency. <http://www.epa.gov/climateleadership/documents/emission-factors.pdf>. Last accessed, July 2013.
- EPA, 2013b. U.S. Transportation Sector Greenhouse Gas Emissions. EPA, September, 2013 <http://www.epa.gov/otaq/climate/documents/420f13033a.pdf> Last accessed, July 2013.
- Federal Aviation Regulations, section 91, section 121.619, section 135.223. <http://www.risingup.com/fars/info/part121-619-FAR.shtml>. Last accessed, December 2014.
- Ettema, D. and H. Timmermans, 2006. Costs of Travel Time Uncertainty and Benefits of Travel Time Information: Conceptual Model and Numerical Examples. *Transportation Research Part C: Emerging Technologies*, 14, 335-350.
- Fosgerau, M. and A. Karlstrom, 2010. The Value of Reliability. *Transportation Research Part B: Methodological*, 44(1), 38-49.
- Gaver, D, 1968. Headstart Strategies for Combating Congestion. *Transportation Science*, 2, 172-181.
- Gulding, J., D. Knorr, M. Rose, J. Bonn, P. Enaud, and H. Hegendoerfer, 2009. US/Europe Comparison of ATM-related Operational Performance. *8th USA/Europe Air Traffic Management Research and Development Seminar*, Napa, California.
- Gupta, G., Malik, W., Jung, Y., 2012. An Integrated Collaborative Decision Making and Tactical Advisories Concept for Airport Surface Operations Management. *12th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference*, Indianapolis, Indiana, USA.
- Gupta, G., Malik, W., Tobias, L., Jung, Y.C., Hoang, T., Hayashi, M., 2013. Performance Evaluation of Individual Aircraft based Advisory Concept for Surface Management. *10th USA/Europe Air Traffic Management Research and Development Seminar*, Chicago, Illinois, USA.
- Hansen M., D. Gillen, R. Djafarian-Tehrani, 2001. Aviation Infrastructure Performance and Airline Cost: a Statistical Cost Estimation Approach. *Transportation Research Part E: Logistics and Transportation Review*, 37(1), 1-23.
- Hao L., M. Ryerson, L. Kang, M. Hansen, 2015. Hold Me If You Can: Estimating the Fuel Burn Impacts of Taxi-out Delay and the Implications for Gate-hold Benefits. Under review, *Transportation Research Part C: Emerging Technologies*.

- Hoang, T., Jung, Y.C., Holbrook, J.B., Malik, W., 2011. Tower Controllers' Assessment of the Spot and Runway Departure Advisor (SARDA) Concept. *9th USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.
- Hollander, Y., 2006. Direct versus Indirect Models for the Effects of Unreliability. *Transportation Research Part A: Policy and Practice*, 40 (9), 699-711.
- International Civil Aviation Organization (ICAO), 2005. Report: Global Air Traffic Management Operational Concept, 1st Edition.
- Jacquet, P., G. Seroussi, and W. Szqankowski, 2008. On the Entropy of a Hidden Markov Process. *Theoretical Computer Science*, 395, 203-219.
- Jamin S., A. Schäfer, M. E. Ben-Akiva, I. A. Waitz, 2004. Aviation Emissions and Abatement Policies in the United States: A City-pair Analysis. *Transportation Research Part D: Transport and Environment*, 9(4), 295-317.
- Jung, Y.C., Hoang, T., Montoya, J., Gupta, G., Malik, W., Tobias, L., 2010. A Concept and Implementation of Optimized Operations of Airport Surface Traffic. *10th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference*, Fort Worth, Texas, USA.
- Karisch, S. E., Altus, S. S., Stojković, G., Stojković, M, 2012. Chapter Operations, in Quantitative Problem Solving Methods in the Airline Industry (Vol. 169), 283-383. Springer US.
- Khadilkar, H., H. Balakrishnan, 2012. Estimation of Aircraft Taxi Fuel Burn using Flight Data Recorder Archives. *Transportation Research Part D: Transport and Environment*, 17(7), 532-537.
- Kho, S. Y., J. S. Park, Y. H. Kim, and E. H. Kim, 2005. A Development of Punctuality Index for Bus Operation. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 492-504.
- Knight, T, 1974. An Approach to the Evaluation of Changes in Travel Unreliability: A Safety Margin Hypothesis. *Transportation*, 3, 393-408.
- Kwon, J., B. Coifman, and P. Bickel, 2000. Day-to-day Travel-time Trends and Travel-time Prediction from Loop-detector Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1717, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 120-129.

- Leigh R.J.,1995. Economic Benefits of Terminal Aerodrome Forecasts (TAFs) for Sydney Airport, Australia. *Meteorological Applications*, 2, 239-247.
- Liston-Hayes, C., Pilkington A., 2004. Inventive Concentration: An Analysis of Fuel Cell Patents. *Science and Public Policy*, 31(1), 15-25.
- Liu Y., M. Hansen, G. Gupta, W. Malik, Y. Jung, 2014. Predictability Impacts of Airport Surface Automation. *Transportation Research Part C: Emerging Technologies*, 44(2014), 128-145.
- Liu, Y. and M. Hansen, 2014. Ground Delay Program Performance Evaluation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2400, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 54-64.
- Lomax, T., D. Schrank, S. Turner, and R. Margiotta, 2003. Selecting Travel Reliability Measures. Texas Transportation Institute, College Station, TX, USA.
- Malik, W., Gupta, G., Jung, Y.C., 2010. Managing Departure Aircraft Release for Efficiency Airport Surface Operations. *American Institute of Aeronautics and Astronautics Guidance, Navigation, and Control Conference*, Toronto, Canada.
- Mayer, C., Sinai, T., 2003. Why Do Airlines Systematically Schedule Their Flights to Arrive Late?" Working paper, The Wharton School, University of Pennsylvania
- Nakahara, A., Reynolds, T.G., White, T., Maccarone, C., Dunsky, R., 2011. Analysis of a Surface Congestion Management Technique at New York JFK airport. *11th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference*, Virginia Beach, VA, USA.
- Nakahara, A., Reynolds, T.G., 2012. An Approach for Estimating Current and Future Benefits of Airport Surface Congestion Management Techniques. *12th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference*, Indianapolis, Indiana, USA.
- Nikoleris, T., G. Gupta, M. Kistler, 2011. Detailed Estimation of Fuel Consumption and Emissions During Aircraft Taxi Operations at Dallas/Fort Worth International Airport. *Transportation Research Part D: Transport and Environment*, 16 (4), 302-308.
- Noland, R. and K. Small, 1995. Travel-time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1493, Transportation Research Board of the National Academies, Washington, D.C., 1995, pp.150-158.

- Noland, R., Polak, J., 2002. Travel Time Variability: a Review of Theoretical and Empirical Issues. *Transport Reviews* 22 (1), 39-54.
- ORDER JO 7110.65V, 2014. Air Traffic Organization Policy (SUBJ: Air Traffic Control). Federal Aviation Administration, U.S. Department of Transportation.
- Pells, S, 1987. The Evaluation of Reductions in Travel Time Variability. PhD thesis, University of Leeds (UK), 1987.
- Ryerson, M.S., Churchill, A., 2013. Aircraft Rerouting due to Abrupt Facility Outages: A Case Study of the 2011 Great Tōhoku Earthquake. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2336, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp.27-35.
- Ryerson M. S., Hansen, M., 2010. The Potential of Turboprops for Reducing Aviation Fuel Consumption. *Transportation Research Part D: Transport and Environment*, 15(6), 305-314.
- Ryerson, M.S., Hansen, M., 2013. Capturing the Impact of Fuel Price on Jet Aircraft Operating Costs with Engineering and Econometric Models. *Transportation Research Part C: Emerging Technologies*, 33, 282-296.
- Ryerson M.S., Hansen M., Bonn J., 2014. Fuel Consumption and Operational Performance. *Transportation Research Part A: Policy and Practice*, 69, 286-298.
- Rice, J. and E. V. Zwet, 2004. A Simple and Effective Method for Predicting Travel Times on Freeways. *IEEE Transactions on Intelligent Transportation Systems*, 5(3), 200-207.
- Riikka, S. and J. Paavilainen, 2010. Measuring Railway Traffic Punctuality from the Passenger's Perspective. *12th World Conference in Transportation Research*, Lisbon, Portugal.
- Shannon, C. E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.
- Simaiakis, I., H., Balakrishnan, 2009. Queuing Models of Airport Departure Processes for Emissions Reduction. *AIAA Guidance, Navigation, and Control Conference*, Chicago, Illinois, USA.
- Simaiakis I., H. Balakrishnan, 2010. Impact of Congestion on Taxi Times, Fuel Burn, and Emissions at Major Airports. *Transportation Research Record: Journal of the*

Transportation Research Board, Transportation Research Board of the National Academies, Vol. 2184, 22-30.

Simaiakis, I., Sandberg, M., Balakrishnan, H., and Hansman, R. J. 2012. Design, Testing and Evaluation of a Pushback Rate Control Strategy. *Proceedings of the International Conference on Research in Air Transportation*.

Simaiakis, I., Khadilkar, H., Balakrishnan, H., Reynolds, T.G., Hansman, R.J., Reilly, B., Urlass, S., 2011. Demonstration of Reduced Airport Congestion through Pushback Rate Control. *9th USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany.

Simaiakis, I., 2013. Analysis, Modeling and Control of the Airport Departure Process. Doctoral Dissertation. Massachusetts Institute of Technology.

Small, K., 1982. The Scheduling of Consumer Activities: Work Trips. *American economic Review*, 72, 467-479.

Sohoni, M., Y. Lee and D. Klabjan, 2011. Robust Airline Scheduling Under Block-Time Uncertainty. *Transportation Science*, 45, 451-464.

Taylor, M.A.P., 1982. Travel Time Variability – the Case of Two Public Modes. *Transportation Science*, 16(4), 507-521.

The State of the Global Markets Report, 2015. Crude Oil Price Ten Year Forecast to 2025: Importers Set to Receive a \$600 Billion Refund. <http://www.marketoracle.co.uk/Article48969.html>. Last accessed, March, 2015.

Vickrey, W., 1969. Congestion Theory and Transportation Investment. *The American Economic Review*, Vol. 59, 1969, pp. 251-260.

Wardman, M., 2001. A Review of British Evidence On Time and Service Quality Valuations. *Transportation Research Part E: Logistics and Transportation Review*, 37, 107-128.

Wojcik L.A., S. L. Mondoloni, S. J. Agbolosu-Amison, P. Wang, 2013. Flexibility Metrics and Their Application to Departure Queue Management. *10th USA/Europe Air Traffic Management Research and Development Seminar*, Chicago, Illinois, USA.

Wu, C., 2010. Airline Operations and Delay Management : Insights from Airline Economics Networks and Schedule Planning .

Zhang, X. and J. A. Rice, 2003. Short-term Travel Time Prediction. *Transportation Research Part C: Emerging Technologies*, 11, 187-210.

Zou, B., Hansen M., 2012. Impact of Operational Performance on Air Carrier Cost Structure: Evidence from US Airlines. *Transportation Research Part E: Logistics and Transportation Review*, 48(5), 1032-1048.