ABSTRACT

Title of dissertation: PROBLEMS AND MODELS IN STRATEGIC AIR TRAFFIC FLOW MANAGEMENT

Prem Swaroop, Doctor of Philosophy, 2013

Dissertation directed by: Professor Michael O. Ball
Robert H. Smith School of Business
and Institute for Systems Research

The thesis comprises of three essays.

The first essay is titled "Do more US airports need slot controls? A welfare based approach to determine slot levels." It analyzes the welfare effects of slot controls on major US airports. We consider the fundamental trade-off between benefits from queuing delay reduction and costs due to simultaneous schedule delay increase to passengers while imposing slot limits at airports. A set of quantitative models and simulation procedures are developed to explore the possible airline scheduling responses through reallocating and trimming flights. We find that, of the 35 major US airports, a more widespread use of slot controls would improve travelers' welfare. The results from our analyses suggest that slot caps at the four airports that currently have slot controls (Washington Reagan, Newark, New York LaGuardia, New York John F. Kennedy) are set too high. Further slot reduction by removing some of the flights at these airports could generate additional benefits to passengers. Slot controls can potentially reduce two thirds of the total system delays caused by

congestion. A number of implementation and design issues related to the use of slot controls are also discussed in the paper.

The second essay is titled "Designing the Noah's Ark: A Multi-objective Multi-stakeholder Consensus Building Method." A significant challenge of effective air traffic flow management (ATFM) is to allow for various competing airlines to collaborate with an air navigation service provider (ANSP) in determining flow management initiatives. This challenge has led over the past 15 years to the development of a broad approach to ATFM known as collaborative decision making (CDM). A set of CDM principles has evolved to guide the development of specific tools that support ATFM resource allocation. However, these principles have not been extended to cover the problem of providing strategic advice to an ANSP in the initial planning stages of traffic management initiatives. In the second essay, we describe a mechanism whereby competing airlines provide "consensus" advice to an ANSP using a voting mechanism. It is based on the recently developed Majority Judgment voting procedure. The result of the procedure is a consensus real-valued vector that must satisfy a set of constraints imposed by the weather and traffic conditions of the day in question. While we developed and modeled this problem based on specific ATFM features, it appears to be highly generic and amenable to a much broader set of applications. Our analysis of this problem involves several interesting sub-problems, including a type of column generation process that creates candidate vectors for input to the voting process.

The third essay is titled "Strategic Opportunity Analysis in *COuNSEL* – A Consensus-Building Mechanism for Setting Service Level Expectations." The consensus-building mechanism described in the second essay has been accepted as a technically viable solution for the stated problem – although many practical challenges still remain before it may be deployed in operations. A key issue worthy of further investigation is its strong strategy-resistance – as claimed by the authors of Majority Judgment, the voting procedure embedded in *COuNSEL*. Using the broad ideas of Nash Equilibria, we characterize the necessary and sufficient conditions for an airline to benefit from unilaterally deviating from truthfully grading one or more candidates. The framework provides the airline with all the other airlines' grades on a set of candidates, and allows it an opportunity to present new grades. The analysis is repeated over multiple instances, and likelihood of beneficial strategic opportunity is presented.

Problems and Models in Strategic Air Traffic Flow Management

by

Prem Swaroop

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Prof Michael O Ball, Chair/Advisor
Prof Martin Dresner
Prof Subramanian Raghavan
Prof Yi Xu
Prof David Lovell, Dean's Representative

In memory of

Ma and Papa

Acknowledgments

I owe my deepest gratitude to my adviser, Prof Michael O. Ball. His unique insights, thorough guidance, invaluable advice, and most of all, patience have been instrumental in shaping this dissertation. Despite his shouldering increasingly larger administrative roles over the last five years, he has always been available for reviews and meetings. The flexibility and support that he extended to me are sincerely appreciated.

I wish to thank my committee members, Prof Martin Dresner, Prof Wedad Elmaghraby, Prof David Lovell, Prof S. Raghavan, and Prof Yi Xu, for their time and guidance on my dissertation.

I would also like to thank the sponsors at the FAA, in particular Rich Jehlen and his staff, for the opportunity to work and collaborate on an interesting and challenging problem.

I am grateful to my co-authors on the first essay Prof Mark Hansen and Prof Bo Zou, for sharing their research and contributions with me. Thanks also for hosting us for the wonderful stay at Berkeley.

This dissertation has benefited from numerous internal presentations that I had the fortune of delivering to fellow students, followed by thoughtful critiques and discussions. I am thankful to all my colleagues, faculty and staff at the business school as well as the NEXTOR group at ISR.

Special thanks are also due to Dr Robert Hoffman, Prof David Lovell, Prof Louiqa Raschid, and Prof Thomas Vossen for providing insightful critiques, encouragement, and helping gain broader audiences.

My ambitious dream of pursuing doctoral studies could materialize thanks in most part to my wife, Prashanthi Krishna. Words fall short in acknowledging her love and support that made it all possible! Sincere thanks to her mother, Hemalatha, for being such a bastion of strength and service. I would also like to thank the entire extended family, friends and neighbors for extending their wishes and support; thanks a ton to Sudha and Anshu; Suchi and Vanu; Samir, Shilpa, Shrenik, Aunty and Uncle!

And finally, many thanks to the little bundles of joy, Sahaj and Shreya, for coming into our lives and giving us a new meaning.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Air transportation is among the most complex man-made systems. It touches millions of lives every day, with over two million daily passenger enplanements in the US alone (BTS 2013). The overall economic activity generated by civil aviation supported over 10 million jobs, and accounted for 5.2% of total US GDP in 2009 with \$1.3 trillion in total output (FAA 2011). Of this, airline and airport operations contributed to over 2.5 million jobs, and \$375 billion of output (1.4% of GDP).

The Federal Aviation Administration (FAA) is the Air Navigation Service Provider (ANSP) in the US. Its Air Traffic Organization (ATO) is primarily tasked with safely and efficiently coordinating air traffic over the National Airspace System (NAS) (FAA 2013). The scale of the operations in this Air Traffic Flow Management (ATFM) is enormous: over 7000 airplane operations (takeoff and landing) per hour, at about 800 airports through the country (29 of these classified as "major"), 50000 flights every day, operated by over 50 passenger airlines (15 with over \$20 million annual operating revenue) and 25 cargo airlines. Employing 35000 air traffic controllers and other support personnel, the ATO operations are executed through 22 Air Route Traffic Control Centers, 27 Terminal Radar Approach Control Facilities (TRACON), and 133 Airport Traffic Control Towers.

The role of strategic planning in ATFM cannot be over-emphasized. The sheer nature of the ATFM operations involves a large network of personnel and equipment, subject to uncertain weather events as well as market forces. This dissertation focuses on some strategic aspects of ATFM in the US.

1.1 Do More U.S. Airports Need Slot Controls?: A Welfare-Based Approach to Determine Slot Levels

The year 2007 was marked with an unprecedented demand for air travel in the US. It was also the year when the on-time performance for US airlines reached historic lows (Figure 1.1). The issue caught national attention, with the Joint Economic Committee of the US Congress issuing a report titled "Your flight has been delayed again: Flight delays cost passenger, airlines, and the US economy billions." The report put its economic estimate of the delays on the overall economy at $41 billion (JEC 2008). The FAA instituted a more scientific and comprehensive study with the NEXTOR (roughly, "National Center of Excellence for Aviation Operations Research") consortium of (then-) five universities. The NEXTOR report estimated the impact of delays at $31.2 billion (Ball et al. 2010).

This essay deals with a specific congestion management approach, namely slot controls. Widely prevalent in Europe, slot controls have historically been used much more sparingly in the US.

Generally speaking, as the number of scheduled operations from an airport increases, passengers would get more service options. This would result in a decrease

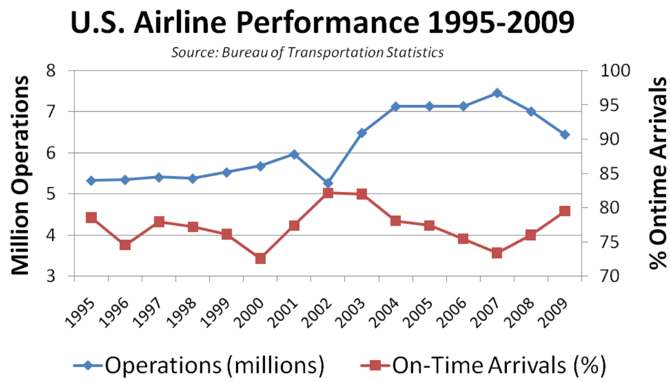


Figure 1.1: Unprecedented delays and demand for US air travel in 2007.

in average schedule delay – which is the difference between a passenger's ideal departure (or arrival) time and the nearest scheduled time for departure (or arrival) by any airline. Increase in scheduled operations would also increase the capacity utilization at the airport. High capacity utilization at busy airports would lead to an increase in the realized delays, as queuing theory would predict.

Slot controls, in effect, reduce the peak capacity utilization at an airport. By implementing slot controls at a busy airport, the increase in schedule delay cost can be traded-off against the decrease in the queuing delay cost. Of course, slot controls at a particular level can be recommended only if the net benefit is positive. This key insight is operationalized in the essay.

Ours is the first prescriptive work that justifies a more widespread use of slot controls in the US through a welfare-based analysis. It not only identifies which airports in the US should be instituted with slot controls, but also specifies the slot levels that should be implemented at each airport to achieve maximum benefit. Using domain-specific data and novel models, we find that 16 airports would benefit with slot controls even at their current capacity levels, yielding net annual benefit of $237 million. The annual benefit nearly triples to $629 million if our recommended slot levels are implemented at these airports. Furthermore, 12 of the sixteen airports would continue to serve the current demand despite the recommended slot controls. Significantly, the recommendations could potentially eliminate two-thirds of overall congestion-related delays.

The essay was co-authored with researchers at University of California, Berkeley. Specifically, I have no intellectual claim on Section 2.4.2 "Passenger Queuing Delay Cost".

1.2 Designing the Noah's Ark: A Multi-objective Multi-stakeholder Consensus Building Method

A significant challenge of effective air traffic flow management (ATFM) is to allow for various competing airlines to collaborate with an air navigation service provider (ANSP) in determining traffic flow management initiatives (TMIs). In the US, this challenge has led over the past 15 years to the development of a broad approach to ATFM known as collaborative decision making (CDM). A set of CDM principles has evolved to guide the development of specific tools that support ATFM resource allocation. However, these principles have not been extended to cover the problem of providing strategic advice to the ANSP in the initial planning stages of traffic management initiatives.

In this research, we seek to develop a framework that addresses the strategic

Figure 1.2: NextGen Service Process at Aggregate NAS level

level planning in advance of the design of a TMI. Specifically, we propose a mechanism whereby the airlines provide "consensus" advice to an ANSP using a voting mechanism. To more precisely state its role in the ATFM decision-making processes, let us contrast the current and the proposed state of collaboration among the stakeholders in the U.S. Currently, the airlines influence ATFM decision-making of the Federal Aviation Administration (FAA) – the ANSP in the U.S. – in several ways. To obtain strategic planning input, the FAA Air Traffic Control System Command Center (Command Center) holds one or more daily *Strategic Planning Teleconns*. In many cases, these are augmented by various ad-hoc calls between the Command Center or regional FAA facilities and various airlines. While collecting the airlines' views is certainly desirable, the current system allows for ad-hoc and inequitable representation – even in the structured teleconns. Moreover, the focus of the discussions is on the TMIs, and not necessarily on the service performance expectations desired by the airlines. In contrast, the architecture of the Next Generation Air Transportation System (NextGen) envisages a service process that focuses entirely on the service expectations (see Figure 1.2, JPDO (2007)). The collaboratively agreed service expectations are to be taken as input to the latter service processes of devising an operational plan, and executing and evaluating the same.

We set out the following as a list of desirable outcomes for our proposed mechanism:

(i) *consensus-building.* The winning vector should have maximum acceptability among the airlines.

(ii) *single winner determination.* The mechanism should result in a single winning vector.

(iii) *practical.* The procedure should be easy to administer, and not involve time-consuming information gathering and / or processing steps by the airlines as well as the ANSP.

(iv) *equitable.* The mechanism should be perceived to be fair to all parties involved from the outset.

(v) *confidential.* The private information requirements from the airlines should be minimal.

(vi) *strategy-resistant.* As far as possible, the mechanism should discourage gaming, and encourage truth-telling behavior.

These are consistent with the principles of mechanism design, and also take into account some specific needs of our application environment.

Our research team, in collaboration with the Federal Aviation Administration (FAA), considered several mechanisms to address the basic requirements listed above. An initial proposal viewed the process as one of allocation of capital among "investment" alternatives, wherein the airlines may purchase the service expectation metrics. A related mechanism is that of Combinatorial Auction, wherein bundles of the metrics may be offered for bidding. However, this paradigm suffers from various problems. Firstly, it requires creation of an artificial "currency," that would be used by the airlines for the investments. Secondly and more fundamentally, the service expectation metrics are not really goods being split up, as is the assumption in combinatorial auction. Rather, each airline's value (performance) is derived from a single, mutually agreed upon vector. Strategic behavior from the airlines becomes unavoidable. Specifically, this mechanism is especially prone to the well-known free-rider problem.

To precisely model strategic behavior among the airlines, the framework was modeled as a Multi-player Non-cooperative Game by other members of the research team (see Ball et al. (2011)). This approach successfully modeled the airline strategic behavior well; the existence of unique Nash equilibrium under certain conditions was also established along with a computational method. However, the ultimate solutions were not viewed as practical in that only extreme solutions were generated (with a clear winner and loser). Solutions where the various stakeholders (airlines) compromise – viewed as highly desirable by the research group – were not generated using this approach.

We then turned to voting procedures, which would seem to constitute a natural way to model the decision making paradigm here. However, challenges – and opportunities – exist in modeling the framework as a voting mechanism. We considered two alternatives: a variant of the Instant Runoff Voting, and Majority Judgment. Majority Judgment is a recently proposed procedure (Balinski and Laraki 2011), that "bypasses" Arrow's Impossibility Theorem – a result that rules out existence of any preference ranking aggregation procedure over three or more candidates, that has certain desirable properties. And hence, its authors claim it to be "a better alternative to all other known voting methods, in theory and in practice."

The choice of Majority Judgment helped meet four of the six desired outcomes: consensus-building, single winner determination, confidentiality, and strategy-resistance. In this essay, we extend the basic Majority Judgment procedure in many ways to address the remaining two. To address equity, the airlines are assigned weights in proportion to the likely impact of the weather, respecting the well-accepted notion of proportional representation.

Our most significant contributions have been three-fold that makes the proposed mechanism practical. First, the proposal allows a continuous candidate space that is constrained by the physics of ATFM parameters. Second, the airlines' preferences are modeled using multi-attribute valuation theory, and estimated over multiple rounds. Three, we develop a novel integer program that identifies the consensus winner over the continuous candidate space, given the airlines' true preference functions. Alternately, given their estimated preference functions, it generates new candidates that approximate the true winner over rounds. In our simulation experiments, we found the optimality gap of the procedure to be 0.13% – which makes our proposal a sound recommendation.

The architecture of the proposed mechanism is presented in Figure 1.3. The mechanism will be initiated by the ANSP with a small (possibly empty) consideration set of feasible candidates. All the candidates will be governed by the physical

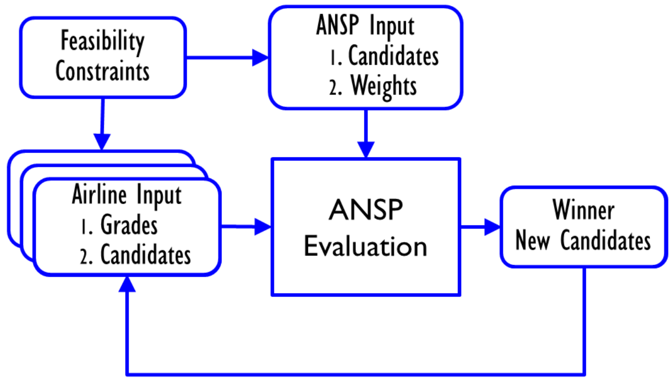


Figure 1.3: Mechanism Architecture

feasibility constraints. The airlines will provide two kinds of inputs: (a) a grade for each candidate in the consideration set, and optionally, (b) one or more feasible candidates. The FAA would use pre-assigned weights for each airline to determine a winner. It would also update its estimate of the airlines' preference functions, and use them to generate new candidates. The consideration set will be updated, and unless some stopping criteria are met, the procedure would repeat.

A shared perception of a common, imminent, unavoidable, impactful threat or opportunity oftentimes leads even fierce competitors to seek consensus solutions. The mythical Noah's Ark is indeed witnessed in the real-world of business, key examples are technology standards bodies like the American National Standards Institute and the Internet Society. We believe our proposal has larger application areas in multi-stakeholder strategic decision making contexts, like capital budgeting, and collaborative forecasting.

1.3 Strategic Opportunity Analysis in *COuNSEL* – A Consensus-Building Mechanism for Setting Service Level Expectations

The consensus-building mechanism described in the second essay has been accepted as a technically viable solution for the stated problem – although many practical challenges still remain before it may be deployed in operations. The underlying models and software tools have been named *COuNSEL*: *CON*sensus *S*ervice *L*evel *E*xpectation Setting. Currently, the research team is actively seeking user feedback from the airlines, and is also preparing software to facilitate Human In The Loop experiments.

A key issue worthy of further investigation is its strong strategy-resistance – as claimed by the authors of Majority Judgment, the voting procedure embedded in *COuNSEL*. In this essay, we seek to verify their claim through simulations.

Following the notions behind Nash Equilibrium, we explore beneficial strategic (that is, untruthful) grading opportunities for each airline *after* they are allowed to see everyone else's grades. This idea has been of prevalence in analyzing mechanism design implementations (Maskin 1999). Of course, such opportunities will not exist in practice, and one may hurt oneself without the exact knowledge of others' grades. Thus, this framework of analysis allows us to characterize the worst case strategy-proneness of the procedure.

Our key contributions are two-fold. One, we characterize the necessary and sufficient conditions under which an airline may benefit from unilateral strategic grading. Using these, we define three measures for strategy-proneness of a given setup. Two, we compute these measures over a variety of simulation experiments, starting from the basic Majority Judgment to more complex procedures that are near our application. Our general finding is that the likelihood for such beneficial strategic grading opportunity for a player via one or more candidates is quite low, in the region of 2% or below. Further, differentially weighting the players does not significantly change the strategy-proneness. Moreover, the reasonable assumption of convex preference structure significantly reduces the strategy-proneness.

1.4 Methodology

A variety of methods from the general areas of Operations Research and statistics were deployed in each of the essays. We provide a summary of the tools and analysis methods in this section.

The foremost tools used in the first essay are economic modeling, linear / integer programming, and simulations. The essay describes two models to predict the likely aggregate response of slot controls at an airport. *FlightTrim* assumes that some service will be dropped from the airport, while *FlightMove* assumes no such drop in service.

FlightMove involves generating new hypothetical schedules that may result after slot controls are implemented at the airports. The airline response to slot controls being not entirely predictable, we treat it as an inherently stochastic process. A large-scale simulation was designed that perturbs the current schedule into one that follows the slot controls, where the small random perturbations are made incrementally over several rounds. A transportation model with a specialized objective function is then used to determine the passenger disutility cost attributable to the new schedule, which yields the marginal schedule delay cost.

Since *FlightTrim* drops the passengers in the new schedule, the marginal cost of schedule delay for the dropped passengers is essentially undefined in the sense described above. Hence, an economic model is developed from the first-principles to estimate the same for this model. A specialized algorithm is designed and implemented to compute the marginal cost.

The second essay employs aspects of voting theory, multi-attribute valuation theory, linear, integer, and non-linear programming, statistical estimation, and simulations. The primary voting method finally used in the essay is Majority Judgment, although it was extended in many directions for the final proposal in *COuNSEL*. Instant runoff voting was another voting method that was also explored in the initial stages.

The airlines' grade functions are motivated from multi-attribute valuation theory. As the grade functions are assumed to be globally concave, non-linear programming tools are used to determining conditions that ensure the same. As a result, statistical estimation of the coefficients of the airlines' grade function involves a constrained linear regression. As no standard libraries are available for such custom-specified regressions, a quadratic program was developed and deployed for the estimation.

A novel integer program is developed to determine the majority judgment winner over a continuous feasible candidate space, given the airlines' true grade functions. In the absence of true grade function coefficients, the same integer program is used for generating new candidates using the estimated coefficients. A similar linear programming formulation achieves the same results, but needs enumeration of majoritarian sets as an input. The same linear program finds the grade-maximizing candidate for a given airline if the argument has only a single airline instead of a majoritarian set.

Finally, large-scale simulations are used to test the validity of the entire proposal. The simulations bring together all the components of the proposal, including aspects of carefully selecting airlines' grade function coefficients, and different weighting schemes. An acceptance sampling based approach is developed for selecting the airlines' grade function coefficients that follow some intuitive guidelines.

The third essay uses some aspects of mechanism design, voting theory, and simulations. The framework for analysis is inspired from Nash equilibrium concept, much like in implementation theory of mechanism design. A logical analysis based on the framework is applied to Majority Judgment which leads to identification of the conditions for unilateral strategic grading that may benefit a player via one or more candidate.

A design of experiments is laid out that proceeds from the basic Majority Judgment to a scenario that closely resembles the *COuNSEL* proposal. Simulating the range of scenarios helps establish the key conclusions from the essay.

Chapter 2: Do more U.S. airports need slot controls? A welfare based approach to determine slot levels

This paper analyzes the welfare effects of slot controls on major U.S. airports. We consider the fundamental tradeoff between benefits from queuing delay reduction and costs due to simultaneous schedule delay increase to passengers while imposing slot limits at airports. A set of quantitative models and simulation procedures are developed to explore the possible airline scheduling responses through reallocating and trimming flights. We find that, of the 35 major U.S. airports, a more widespread use of slot controls would improve travelers' welfare. The results from our analyses suggest that slot caps at the four airports that currently have slot controls (Washington Reagan, Newark, New York LaGuardia, New York John F. Kennedy) are set too high. Further slot reduction by removing some of the flights at these airports could generate additional values to passengers. Slot controls, if optimally implemented, could yield a net benefit of 0.8 billion dollars for the U.S. air transportation system in 2007, and help reduce two thirds of the total system delays caused directly by congestion. A number of implementation and design issues related to the use of slot controls are also discussed in the paper.

2.1 Introduction

Air transportation delays in the U.S. and around the world represent a well-known burden to society and are the subject matter of both technical and public policy debates. A recent study (Ball et al. 2010) estimated the total economic impact of air transportation delays on the U.S. economy in the year 2007 to be \$31.2 billion. The most obvious and often called-for actions are investments in the expansion of system capacity either in the form of infrastructure, e.g. new runways and airports, or new capacity-enhancing technologies. The other option to curtail delay is through demand management. While investing in capacity can be lumpy, expensive, politically contentious, and sometimes technically challenging, demand management – often realized in the form of either slot control or congestion pricing at the airport level – seems cheaper, more flexible, and effective in the short run. To be sure, the second approach involves altering the behavior of individuals or companies, resulting in social and political hurdles. Appropriate evaluation of the benefits from demand management is, therefore, critical to justify the implementation of airport demand management strategies and to inform the inevitable public policy debates. This paper focuses on slot control, the most widely implemented form of airport demand management, and develops a method to investigate the fundamental trade-offs between costs and benefits from restricting flight schedules. The next section provides a background of recent slot control policy and practice in the U.S., and establishes our basic premise for the research. In Section 2.3 we review the existing research. Section 2.4 presents our models for computing the cost (increase in "passenger schedule delay") and benefit (savings in "delay against schedule") of implementing slot controls at an airport. The application of the models to U.S. airports and results are presented in Section 2.5. Section 2.6 concludes and presents further discussion.

2.2 Background

In the U.S. delays reached a high point in the year 2000 only to recede in the advent of the 9/11 tragedy and related changes to the air transportation system and economy. Subsequent demand growth coincided with the return to levels of delay at and even beyond the 2000 level in the year 2007. Very recently a softening of demand has once again led to a reduction in delays. While delay trends have seen fluctuations, few would argue that this is not a long-term problem in need of government investment and action.

2.2.1 Recent Slot Control Policy and Practice in the U.S.

The history of slot rules in the U.S. dates back to the inception of the High Density Rule (HDR) in 1969 and has had many twists and turns over the past decade (Berardino 2009). The passage of the Wendall H. Ford Aviation Investment and Reform Act (AIR-21) in 2000 called for the elimination of slot controls at New York's John F Kennedy International Airport (JFK) and LaGuardia Airport (LGA) by January 1, 2007 and at Chicago O'Hare Airport (ORD) beginning July 1, 2002. In anticipation of delay increase after the expiration of the HDR, the Federal Aviation

Administration (FAA) has proposed alternatives in an effort to avoid exorbitant delays in a post-HDR era. In a proposed 2006 rulemaking (FAA 2006), the FAA sought to require airlines serving LGA to maintain a certain average gauge (seat capacity); airlines failing to attain the average gauge standard would lose slots for their smaller-gauge flights until the standard was attained. While based on the idea that larger aircraft allow the access of more passengers to the airport, this proposal was strongly opposed by airlines and the Port Authority of New York and New Jersey (PANYNJ), arguing that it was overly disruptive and prescriptive, and did not take into account airport-specific constraints (PANYNJ 2008).

The FAA then proposed a slot allocation policy for LGA, and soon after JFK and Newark Liberty International Airport (EWR), based primarily on grandfather rights, but with auctioning of a limited number of slots (FAA 2008d,b). In its final rule for LGA, each carrier currently holding slots would have lost 15 percent of its slots in excess of 20 (FAA 2008c). These slots would be relinquished over a five-year period, with two thirds of them auctioned and the remaining one third retired, decreasing the hourly cap from 75 to 71. Similar rules, albeit with relinquishment of 10% of slots in excess of 20 and no retirements, were set forth for JFK and EWR (FAA 2008a). These rules were challenged in court by the Air Transport Association and PANYNJ, who argued that FAA lacked legal authority to conduct slot auctions. The DC Court of Appeals issued a stay delaying the plan, and this, in combination with Congressional action, caused FAA to rescind the rule in 2009. However, the FAA did feel compelled to implement simple caps on the number of operations at the three major airports in the New York Region (FAA 2008e,f). These caps remain in effect as of 2011.

Despite the many debates throughout the above slot control rulemaking process, there remain important gaps to fill before making any improved decision making. The benefits and costs from slot controls have not yet been systematically quantified and well understood. Neither the setting of caps nor the allocation of slots at the slot controlled airports was based on rigorous economic analysis. A further issue left unaddressed is whether policy makers in the U.S. should consider more widespread use of slot control, as exists in Europe where it has been implemented at virtually all major airports. Filling these gaps constitutes the major motivation in the present paper.

2.2.2 The Fundamental Tradeoff: Economic Justification for Slot Controls

This paper focuses on the fundamental tradeoff in determining the socially optimal level of operations for an airport. Figure 2.1 illustrates this tradeoff. The x-axis is given in units of the fraction of available capacity at which the operations level is set, e.g. by the hourly slot levels. The curve that decreases from left to right is the ex-ante *schedule delay cost*. Schedule delay is a well-known phenomenon in transportation systems. It measures the degree to which passengers must adjust their planned departure time to accommodate the schedule offered by a transportation service. For example, if a passenger wished to depart at 9 AM but there were only flights offered at 8 AM and 10 AM, then that passenger might choose the 8 AM

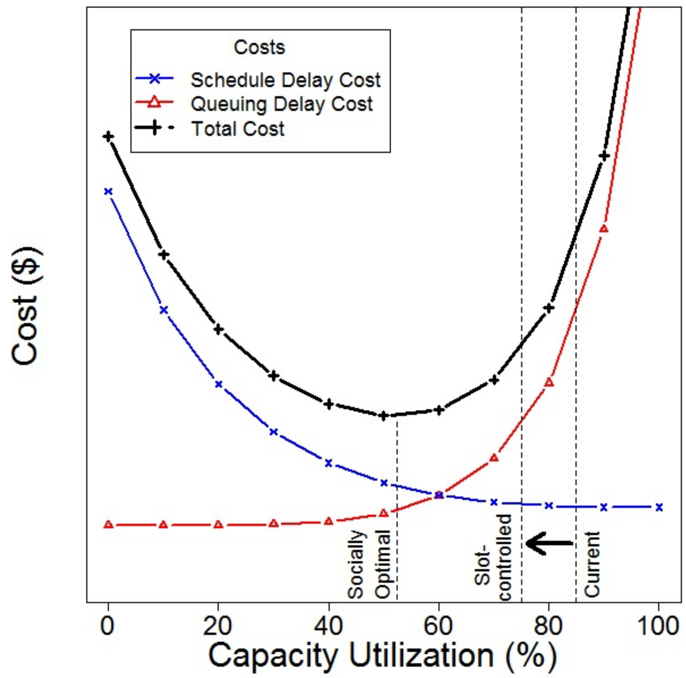


Figure 2.1: Schematic representation of cost curves for queuing delay and schedule delay vs. capacity utilization.

flight and we would say the passenger suffered one hour of schedule delay. The curve that increases rapidly as the airport capacity is approached represents the "classic" delay-against-schedule cost that passengers experience when a flight arrives late, a flight is canceled or a connection is missed. We refer to this delay component as *queuing delay cost* since it results from system congestion and increases at a greater than linear rate as system demand approaches system capacity. As the level of operations at an airport is restricted, airlines will be forced to reduce flight frequency in certain city pairs (in the following we refer to a city pair as a market), increasing passenger schedule delay, but in the meantime resulting in reduction in passenger queuing delays. The optimal level of operations is identified by the lowest point on the total cost curve, as shown in Figure 2.1, which is the sum of costs due to passenger schedule delay and queuing delay. Of course, an extreme case would be an air carrier exiting one market entirely. Consideration of this will be explicitly incorporated into the subsequent analysis. In this paper we argue that many US airports currently operate at a point far to the right of this optimal point. This would seem to give strong support for instituting slot controls at more airports and for setting the existing slot controls at lower levels than they are now set.

To support this argument, we estimate the slopes of the two curves illustrated in Figure 2.1. The relevant models are described in Section 2.4. The models to estimate schedule delay employ a combination of economic modeling, simulation, and optimization. In order to compute the cost of changes in the level of operations these models produce estimates of plausible – albeit not necessarily optimal – changes in flight schedules that would result. The model to estimate the delay against schedule is an econometric model. It directly estimates flight delays, and from this, passenger delays and costs are derived. We consider two possible scheduling responses from airlines: either by moving flights from peak periods to less congested time windows, or eliminating certain flights from time windows where slot control is active.

The fundamental tradeoff may also involve fares and competition effects. Specifically, if operations are restricted in some way, then increased resource scarcity could lead to higher fares. To the extent that such restrictions allow one or more air carriers to increase market power, this could move fares even higher. While these effects are often cited as a major detriment of airport access controls, they are not included in the present study. Among other challenges, the degree to which there is an anti-competitive effect depends very much on how controls are implemented. For example, administrative slot controls that are based primarily on grandfather rights would tend to preserve existing market structure and restrict new entrants from entering the markets served by the airport. Mechanisms that allow for some slot re-allocation, e.g. via auctions, would support a more vibrant competitive environment and lower fares.

The fundamental tradeoff described in this section will determine when access controls *have the potential* to improve welfare. Poorly implemented controls, on the other hand, can negate or greatly reduce the overall benefits by moving farther away from the minimum point on the total cost curve.

2.3 Summary of Literature

Before proceeding to discussing the models that characterize the fundamental tradeoffs and identify the welfare-improving slot control levels, it would be useful to have a review of existing research in related fields. Any airport congestion management scheme, including slot control, essentially deals with flight schedule change and in particular de-peaking of airport traffic. One consequence of flight schedule change is variation of schedule delay perceived by passengers. The flight schedule, the single most important product of an airline, is originally developed in a manner that best accommodates travelers' departure time preferences (Proussaloglou and Koppelman 1999). However, no matter how flights are scheduled, schedule delay-as measured by the difference between a traveler's desired departure time and the nearest flight's scheduled departure time-always exists, and contributes to the passenger generalized cost for a trip. Because each individual's preferred travel time is generally unknown, researchers often resort to statistical, aggregate approach to quantify the relationship between schedule delay and frequency (Douglas and Miller 1974, Abrahams 1983). Provided passenger demand and flight spacing can be reasonably assumed to be uniform, simplified schedule delay functional forms have been derived (e.g. Brueckner and Flores-Fillol (2007)). When airport congestion management schemes are implemented, some flights may be forced to fly at less convenient times and therefore increase the overall passenger schedule delay. Quantification of such schedule delay change, however, has garnered only limited attention (Hansen 2002).

Impacts on schedule delay notwithstanding, airport congestion management schemes are incentivized by the reduction of queuing delay through de-peaking. Queuing delay models abound in literature, falling primarily into three categories: stochastic (e.g. Kivestu (1976)), deterministic (Hansen (2002), Hansen et al. (2009)), and econometric (Hansen and Wei (2006), Morrison and Winston (2008), Hansen et al. (2010)), and have been utilized to compute marginal delay cost for flights. Carlin and Park (1970), Morrison (1983), Hansen (2002), and Ashley and Savage (2010) found these marginal costs are higher than the actual airport charges. Researchers have also looked at different trade-offs between queuing delay and other pertinent elements. Using a greedy algorithm, Hansen (2002) demonstrated eliminating five flights could save 1570 seat-hours of congestion delay while incurring only 51 seat-hours in additional schedule delay. Flores-Fillol (2010) analyzed a simple network model incorporating flight frequency choices and congestion, in which the tradeoff between congestion and schedule delay was explicitly presented in the computation of congestion tolls. Focusing on connecting hub operations, Daniel (1995) examined the tradeoff between delay, additional layover time, and "interchange encroachment" time. Coogan et al. (2010) considered additional trades, including increased travel time from shifting short-haul flight traffic to surface modes, and having private aircraft shift operations away from busy commercial airports.

While our analysis primarily focuses on passenger schedule and queuing delay costs and benefits, slot control also incurs other consequences, including changes in carrier profitability, load factor, air fare, and aircraft size. Research by Vaze and Barnhart (2011) and Le (2006) attempted to understand the impact of slot controls on these variables, focusing on New York LaGuardia (LGA). These studies, at least implicitly, considered the queuing delay – schedule delay tradeoff as well as other impacts and tradeoffs. Both studies attempted to answer the question, "What would happen if more restrictive slot controls were put in place at LGA?" They considered not only changes in service frequency but also impact on airline costs and profits. Their conclusions are consistent with ours, namely that tighter slot controls at LGA would lead to a net benefit to society.

Research also extends to comparing different congestion management schemes. Brueckner (2009b) found that atomistic pricing, which charges each flight its marginal congestion cost even though some of that cost is borne by flights of the same airline, is less efficient than slot controls so long as the number of slots is optimally chosen. He further pointed out congestion pricing, despite its economic justification, might be politically infeasible because small carriers would fiercely oppose a rule that appears to subject them to an unfair burden (Brueckner 2009a). Czerny (2010) argued that demand and congestion cost uncertainty, which may lead to a suboptimal choice for the number of slots or the congestion price, favors congestion pricing, i.e. the pricing errors that result from imperfect information are less harmful than errors in setting the number of slots. Basso and Zhang (2010) considered a model with perfectly elastic air travel demand and found slot auctions will outperform congestion pricing when airport profits matter from a social viewpoint. Ball et al. (2007) reported on gaming simulations of congestion pricing and slot auction policies for LGA. While the simulation indicated that both schemes are feasible, it also showed the challenge in setting congestion prices. Congestion pricing, on the other hand, was seen to bear certain advantages, including increased carrier scheduling flexibility and reduced incentive for airlines to hoard slots. These different views notwithstanding, on balance many observers find slot-based approaches conceptually much more intuitive and easier to manage than congestion pricing (Berardino 2009). As discussed in Section 2.2.1, some of the recent slot control proposals in the U.S. included a provision to auction some slots. Ball et al. (2006) provides a broad framework for airport slot auction. It gives many of the key features that should be included in an auction design and also discusses the relationship between the problem of allocating long-term access rights and performance on a given day-of-operations.

2.4 Models for Estimating Schedule Delay and Queuing Delay

We now present models for computing the incremental costs (i.e. schedule delay costs), and incremental benefits (i.e. savings in queuing delay), of implementing slot controls. Estimation is done at three slot levels, namely the peak airport capacity, and its 90% and 80% levels, measured in the number of arrival operations per quarter hour. Setting slot levels to peak airport capacity may be the most straight-forward recommendation if the benefits justify the costs, as it may be simpler to communicate and gain agreement on than any lower slot level. However, as we hypothesize in section 2.2.2, further benefits may be possible with slot level set below the peak capacity, justifying the need to investigate the costs and benefits at lower levels.

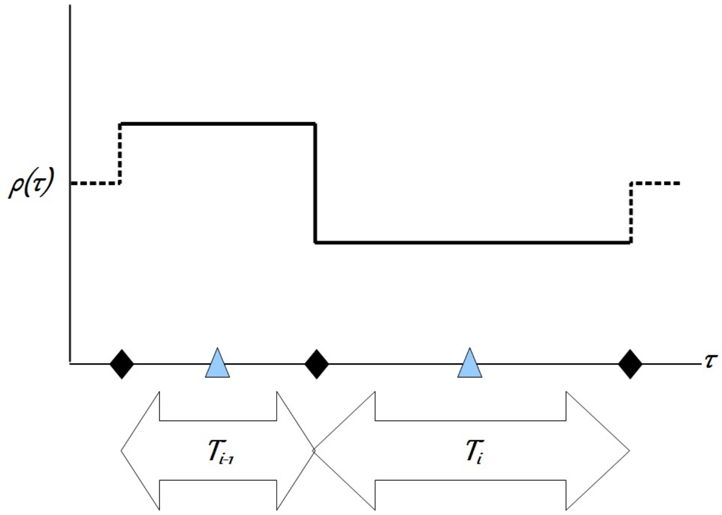
2.4.1 Passenger Schedule Delay Cost

Although the definition of schedule delay is straightforward, evaluating it is challenging. We need to first estimate the "ideal" passenger demand, which will be compared with the original flight schedule to determine passenger schedule delay. To convert schedule delay into cost, we also need an estimate of passenger valuation of time. The estimation of incremental change in passenger schedule delay cost requires a procedure to emulate airlines' responses to the imposition of slot control, and the resulting new flight schedule. In the present study, we propose two schedule delay models, *FlightMove* and *FlightTrim*, to characterize possible airline responses to slot control and the associated change in passenger schedule delay. In the *FlightMove* model, we assume airlines would reschedule, but not drop, existing flights when slot controls are imposed at an airport. The *FlightTrim* model assumes there would be a reduction in the number of scheduled flights when the slot levels are made more stringent. The core for both models is the schedule delay cost function, which we first present below.

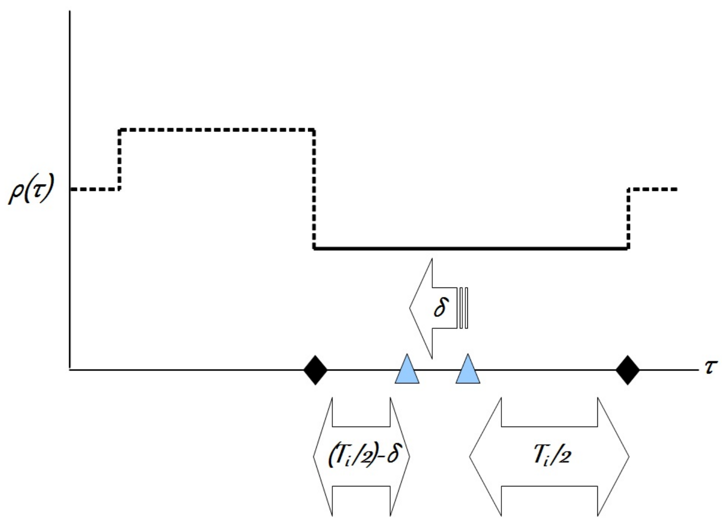
2.4.1.1 Cost Function

As *FlightMove* and *FlightTrim* represent two distinct responses from airlines to slot control, the calculation of their respective schedule delay change would also be different. In the *FlightMove* model, we specify a cost function in terms of flight perturbation from the current schedule. In this case, the model computes the incremental cost directly using this function. In the *FlightTrim* model, we derive schedule delay cost from first principles as a function of the number of flights. Incremental cost is then derived by iteratively deleting an average flight from the total number of flights.

Let $\rho(\tau)$ be the density function of the number of passengers who would ideally like to travel at time τ for a market (refer Figure 2.2a). Integrating $\rho(\tau)$ over a time-interval T gives total demand in the interval. Assuming that airlines place flights so that each captures an equal amount of demand, and that demand is uniform over each interval covered by a flight, $\rho(\tau)$ would have the form illustrated in Figure 2.2a. While it seems reasonable and pragmatic to use the observed flight schedule time to derive passengers' preferred travel time, this may not be accurate as airlines' schedules are subject to many operational constraints, such as those from terminal capacity, coordination of a hub-and-spoke network, and competition. Some passengers' schedule delay would certainly be overestimated whereas others underes-



(a) Ideal departure time function



(b) Perturbation to ideal departure times

Figure 2.2: Schematic for schedule delay cost function.

x-axis represents time, y-axis represents the passenger demand density. Triangles denote the flight time, diamonds demarcate the time intervals over which the flight services its demand.

timated . The extent of this deviation, however, is difficult to gauge. To circumvent these uncertainties, in the following we focus on the average effect of scheduling by assuming the flights are evenly placed, with T being the headway. This assumption leads to the constant demand density function $\rho(\tau) = \rho = K/T$, where K is the average number of passengers per flight, and simplifies the subsequent schedule delay calculation.

Incremental Schedule Delay Cost on Moving a Flight. Consider scheduled flight arrivals for a given market at the airport of our interest. Let us perturb the current flight placement by moving a flight δ time-units earlier within the time-interval T_i (i.e. from the triangle on the right to the left one), as shown in Figure 2.2b. We assume that the perturbed flight continues to serve its demand density. In other words, passengers do not change their flight due to the change in schedule. This would change schedule delay for three sets of passengers differently:

(i) Passengers whose ideal travel time is later than the original flight will see an increase in their schedule delay by δ. The number of such passengers $= (K/T_i)(T_i/2) = K/2$; hence the total increase in schedule delay $= K\delta/2$.

(ii) Passengers whose ideal travel time is between the new and original flight times, as a group, will experience a total schedule delay change of zero. If we split interval δ evenly into two parts, then the increase in schedule delay for passengers on the right part will be offset by equal decrease in schedule delay for passengers on the left part.

(iii) Passengers whose ideal travel time is before the new flight time will see a decrease in schedule delay by δ. The number of such passengers $= (T_i/2 - \delta)(K/T_i)$; hence the total decrease in schedule delay $= K\delta/2 - K\delta^2/T_i$.

Thus, the increase in passenger schedule delay for a flight perturbation equals

$(K/T_i)\delta^2$. If f flights were so perturbed, the total change in schedule delay in passenger-hours would be $(K/T_i)\delta^2 f$. Under the simplification of a single headway T_d for market d, the incremental schedule delay cost from identically moving f flights leftward by δ_d is given by:

$$MSDC_d^{FM}(\rho_d, \delta_d, f_d) = \Lambda(K_d/T_d)\delta_d^2 f_d = \Lambda\rho_d\delta_d^2 f_d.$$

where superscript FM denotes *FlightMove* model; Λ the value of passenger schedule delay, measured in $\$/hr$.

Similar results can be derived by moving flights rightward and beyond the original time-interval.

Incremental Schedule Delay Cost On Deleting a Flight. Under the assumption of constant demand density ρ and flight headway T, the schedule delay for each flight can be calculated as:

$$SD(T) = 2\int_0^{T/2} \tau\rho d\tau = \rho T^2/4.$$

In the case of *FlightTrim*, we assume N_{dt} flights for market d in a four-hour period t. In this four-hour period, demand can be more reasonably assumed to be uniform than in other longer time periods . Then the headway for market d in period t, $T_{dt} = 4/N_{dt}$. Total schedule delay cost across all flights for market d in period t is:

$$SDC_{dt} = \Lambda N_{dt}\rho_{dt}T_{dt}^2/4 = 4\Lambda\rho_{dt}/N_{dt}.$$

Assuming that the removal of one flight in market d in a four-hour time interval t redistributes the passenger demand over the remaining flights flying the same market in the same time-interval, incremental cost of schedule delay caused due to the removal of flight to its passengers is given by:

$$MSDC_{dt}^{FT} = 4\Lambda\rho_{dt}\left(\frac{1}{N_{dt}-1} - \frac{1}{N_{dt}}\right).$$

where the superscript FT denotes *FlightTrim* model.

2.4.1.2 *FlightMove* Model.

The aim of this model is to determine the expected delay cost from modifying a flight schedule so that it conforms to quarter-hourly slot limits. To determine this, we must consider how the schedule is modified to satisfy the limits. Unfortunately, this cannot be known with any certitude. While it is tempting to assume some form of maximizing behavior, this is unlikely at the aggregate level since multiple airlines are involved. Competitive models might also be employed, but there are so many cost and revenue considerations at play that this would be difficult for even a single airport, and prohibitively so for the large set of airports we are considering here.

In light of this, we opt for a different, more agnostic, approach. Essentially, we randomly generate a series of small perturbations to the existing schedule that eventually yields a new schedule that conforms to the slot limits. We then assume that the changes leading from the prior schedule to the new one minimize the aggregate amount of flight schedule change. For example, if a schedule includes one less flight to a given destination between 8 and 8:15am, and one less flight between 5 and 5:15 pm, while the 8:15-8:30am and 4:45-5pm periods have one more flight to the same destination, then we assume the morning and evening flights were each moved by one period, rather than having the morning flight moved to the evening slot and vice versa.

This process is clearly stochastic; we therefore simulate it multiple times and average the results. Also, in some cases there may not be enough slots to accommodate all the flights no matter how they are rescheduled. This necessitates the use of the *FlightTrim* model discussed later.

We begin with an "average" daily schedule aggregated to the origin, time period level for an airport. Suppose now the airport is subjected to slot controls limiting the number of scheduled arrivals in a given quarter-hour period. The *FlightMove* simulation now proceeds as follows. First, it randomly selects a time window t with total flights above the slot level, then determines the number of flights f_{dt} scheduled in this time window to move for each market d, and then assigns each flight a move to either $t-1$, t (i.e. no move), or $t+1$. $t=1$ and $t=64$ are handled specially to ensure moves are not made outside of the time range. This is repeated until the aggregate schedule is within the slot level at each time-point in the range $1,\ldots,64$. Schedule delay costs are computed for the predicted schedule, which completes one simulation run. Further details about the simulation are provided in A.1. This procedure is then repeated multiple times. We report the mean cost over the simulation runs for each airport in the subsequent analysis, and the range of results in terms of z-scores, i.e., number of standard deviations from the mean for each airport in A.2.

Cost Determination Using Transportation Model. The simulation procedure yields a predicted schedule that may have moved the same flight segment multiple times over the iterations. Since the cost function $MSDC_d^{FM}$ is not linear in the length of a move it would not be accurate to simply add the cost of the independent moves, which could involve multiple moves to the same flight. We employ a costing model that minimizes the total cost of all moves. This can be viewed as a lower bound of the total cost, given the new schedule, but also as the most likely way in which the new schedule would be reached.

A linear programming-transportation model determines minimum cost flight moves for each destination, given the original and predicted schedules. Its objective function value gives the total increase in schedule delay upon perturbing the original schedule into predicted schedule, or, the incremental schedule delay cost for the airport.

Define non-negative decision variables:

$f_{dik} =$ number of flights moved from a 15-minute time-interval i to k for destination d,

and parameters:

$\rho_d =$ demand density for destination d,

$\Lambda =$ passengers' valuation of time for schedule delay,

$N_{dj} =$ number of flights originally scheduled during time-interval j for destination d,

$P_{dj} =$ number of flights in the predicted schedule during time-interval j for destination d.

Then the formulation is:

Minimize

$$\text{Incremental schedule delay cost, } MSDC^{FM} = \sum_{dik} \Lambda \rho_d (i-k)^2 f_{dik}$$

subject to:

$$\sum_k f_{djk} \le N_{dj} \qquad \forall \quad d, j \qquad \text{(Supply)}$$

$$\sum_i f_{dij} \ge P_{dj} \qquad \forall \quad d, j \qquad \text{(Demand)}$$

$$f_{dik} \ge 0 \qquad \forall \quad d, i, k \qquad \text{(Non-negativity)}$$

"Supply" constraints ensure that the total flights moved *out* from any time-point across all destinations are below the originally scheduled flights at the time-point. "Demand" constraints are the converse; these ensure that total flights moved *into* any time-point across all destinations meet the predicted scheduled flights at that time-point. The cost minimizing objective function makes sure that these are met at equality, as any excess flights would come at an avoidable positive cost. "Non-negativity" constraints make sure that decision variables, here the flights moved, are zero or positive.

2.4.1.3 *FlightTrim* Model.

We now present an alternate approach to estimating incremental schedule delay cost upon imposition of slot controls. Here we assume that flights will necessarily be trimmed from the current schedule when slot controls are imposed at an airport, and we estimate the schedule delay incurred due to the removal of "average" flights.

This procedure approximates the average outcome that would be obtained from simulating a random sequence of flight removals analogous to the random sequence of flight schedule changes modeled in *FlightMove*.

The model we now describe is applied to each four-hour period t over the course of a day. In our case we employ a sixteen-hour day so we consider $t = 1, \ldots, 4$. We initially compute for each t, the number of flights to be dropped $DF_t = \max\left(0, \sum_d N_{dt} - SL\right)$, that is, the excess if any, of scheduled flights over the slot level in the period. Note that for this model, t denotes a larger time-period than *FlightMove* model.

Algorithm. We adopt an algorithm that successively trims one average flight from the overall schedule in each iteration for the period t.

Initialize. Set iteration counter:

$$i \leftarrow 1.$$

Repeat steps 1–3 while there are flights to trim in iteration i, i.e., total number of flights remaining is larger than the slot level:

$$\sum_d N_{dt}^{(i)} > SL.$$

Step 1. For all destinations that have greater than two flights remaining in iteration i, trim $1/\sum_d N_{dt}^{(i)}$ fraction of flights for destination d at period t and update number of flights remaining as below:

$$N_{dt}^{(i+1)} \leftarrow N_{dt}^{(i)} \cdot \left(1 - \frac{1}{\sum_{d'} N_{d't}^{(i)}}\right).$$

Note that we drop a total of $\sum_d \{N_{dt}/\sum_{d'} N_{d't}^{(i)}\} = 1$ flight over all the destinations in each iteration.

Step 2. Compute incremental schedule delay of dropping the flights in Step 1 as:

$$MSD_{dt}^{(i)} \leftarrow 4\rho_{dt}\left(\frac{1}{N_{dt}^{(i)}} - \frac{1}{N_{dt}^{(i-1)}}\right).$$

Step 3. Update iteration counter:

$$i \leftarrow i + 1.$$

We need $\lfloor DF_t \rfloor + 1$ iterations, where $\lfloor \cdot \rfloor$ is the floor operator, yielding the largest integer smaller than or equal to the operand. The last iteration is to interpolate MSD_{dt} for the fractional part of DF_t.

The incremental passenger schedule delay cost for the airport is then:

$$MSDC^{FT} = \sum_{tid} \Lambda \cdot MSD_{dt}^{(i)},$$

It is implicitly assumed that flight moves within each four-hour period are cost-less, and are not possible beyond the specific period. If the flights within a period are more than the slot level, then those are trimmed instead of being moved out to another four-hour period. Further, the model preserves smaller markets: destinations having less than two flights in each period are not trimmed at all. This can be viewed as a type of policy prescription. However, we should also point out that the model becomes unstable for markets with less than one flight in the average schedule. We note that such markets do exist (they receive some service during a week but not daily service). In fact, eliminating the only flight that serves a markets cannot really be modeled using schedule delay since this amounts to a loss of service.

2.4.1.4 Data.

We use August 2007 as the target month for our analysis of the 35 Operational Evolution Partnership (henceforth OEP35) airports in the U.S. The name and code of the airports included in OEP35 are listed in A.3. To avoid irregularities in schedules over weekends, we use only Tuesday, Wednesday, and Thursday. Daily schedules are computed by averaging over the relevant days. The daily schedule is composed of sixty-four 15-minute periods spanning 6 AM to 10 PM to capture the busy period.

The Aviation System Performance Metrics (ASPM) database, maintained and published by the U.S. Federal Aviation Administration (FAA), was used for computing aggregate schedules. Market-based schedules are computed using Official Airline Guide (OAG) data. We use arrivals data for all computations.

Finally, we use passengers' valuation of time for schedule delay, Λ = \$15.77 per hour. Adler et al. (2005) report fare substitution values for a number of service variables for business and leisure travelers. The mean values for an hour of scheduled arrival time difference are respectively \$30.3 and \$4.8. As their study has 43% business trips and 57% leisure trips, we arrive at average passengers' valuation of time for schedule delay, Λ, as \$15.77 per hour. This is the most recent estimate; another by the classic Proussaloglou and Koppelman (1999) reports the valuations to be \$40 and \$10 per hour for business and leisure travelers, respectively. Hsiao and Hansen (2011) present evidence that the value of schedule delay as trended downward since 2000. As the exact share of business vs. leisure travelers at an airport is in general unknown, in the subsequent analysis we also use one third and twice of the \$15.77 per hour value as passengers' valuation of schedule delay to test the sensitivity of the results. These two values would represent the extreme cases that an airport is used only by either leisure, or business passengers.

2.4.2 Passenger Queuing Delay Cost

Under the assumption that the queuing delay experienced by a passenger is the same as the queuing delay of his/her flight, this sub-section adopts a two-step approach to quantify flight queuing delays. In the first step, we construct a deterministic queuing diagram at each of the US OEP35 airports. The calculated queuing delays and their higher order terms are then included-together with other explanatory variables-in an econometric model which is estimated using data for 2007. This two-step, hybrid approach enhances the model's capability of predicting queuing delays at current levels and producing credible delay results under different airport slot control scenarios.

2.4.2.1 Deterministic Queuing Delay.

We derive the deterministic queuing delay at each airport by constructing a deterministic queuing diagram, which illustrates the operational demand and supply relationship at an airport. The deterministic queuing diagram is based on the time profile of scheduled flight demand and airport capacity over the course of a day, and thus is capable of capturing temporal characteristics of scheduled demand, such as

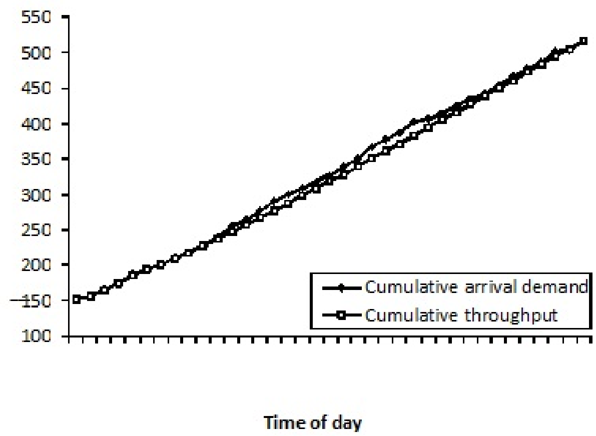


Figure 2.3: Queuing diagram of arrivals (EWR, Jan 2, 2007).

peakedness. Two curves in the queuing diagram are pertinent to the calculation of queuing delays: cumulative scheduled arrivals and cumulative throughput. As an example, Figure 2.3 illustrates the deterministic queuing diagram at Newark Liberty International Airport (EWR) on January 2, 2007.

Specifically, the cumulative scheduled arrival curve is constructed using the ASPM quarter-hour scheduled flight arrival information. Let $D_{it,l}$ denote the cumulative demand at airport i on day t by the lth quarter hour. The cumulative throughput curve can be constructed using $D_{it,l}$'s and the quarter-hour airport acceptance rate $AAR_{it,l}$'s. For time period l on day t, the cumulative throughput at airport i is the minimum of the cumulative scheduled arrival and the sum of cumulative throughput in the preceding period and contemporaneous AAR:

$$C_{it,l} = min(D_{it,l}, C_{it,l-1} + AAR_{it,l}).$$

Employing Little's Law, total deterministic queuing delay ($TDQD$) is calculated as the area between the two curves:

$$TDQD_{it} = 15 \sum_{l} (D_{it,l} - C_{it.l}) \qquad \text{(in minutes).}$$

Daily average queuing delay per flight, Q_{it}, is obtained by dividing the $TDQD_{it}$ by the total scheduled arrivals. This procedure is repeated for each day in 2007 and each of the OEP35 airports. While it is possible for delays on a given day to spillover to the next day and consequently calculate the queuing delay continuously over the entire analysis period (Hansen and Kwan 2010), we find the queuing delay results and the subsequent delay model estimation do not significantly differ. In addition, because in the study only Tuesdays/Wednesdays/Thursdays are concerned, we still stick to computing queuing delays for each airport-day pair. Previous studies have revealed that average deterministic queuing delay is highly correlated with the observed average flight delay (Hansen and Hsiao 2005, Hansen and Kwan 2010).

2.4.2.2 Econometric Model.

In the second step of modeling queuing delay, we propose and estimate the following econometric model:

$$\begin{aligned} D_{it} = \beta_0 &+ \beta_1 Q_{it} + \beta_2 Q_{it}^2 + \beta_3 Q_{it}^3 \\ &+ \beta_4 IFR_{it} + \beta_5 IFR_{it}^2 + \beta_6 Wd_{it} + \beta_7 Temp_{it} \\ &+ \beta_8 AAR_{it} + \beta_9 Connect_{it} \\ &+ \sum_k \omega_k q_k(t) + \sum_j \lambda_j m_j(i) + \epsilon_{it}, \end{aligned}$$

where:

$D_{it} =$ Average positive arrival delay against schedule per flight, in minutes, at airport i during day t;

$Q_{it} =$ Average deterministic arrival queuing delay per flight, in minutes, at airport i during day t;

$Q_{it}^2 =$ The square of average deterministic arrival queuing delay per flight;

$Q_{it}^3 =$ The cube of average deterministic arrival queuing delay per flight;

$IFR_{it} =$ The portion of time during day t in which airport i operated under Instrument Flight Rules (IFR) conditions;

$IFR_{it}^2 =$ The square of the portion of time during day t in which airport i operated under Instrument Flight Rules (IFR) conditions;

$Wd_{it} =$ Average wind speed, in knots, at airport i during day t;

$Temp_{it} =$ The average temperature, in Fahrenheit, at airport i during day t;

$AAR_{it} =$ Airport arrival acceptance rate (number of arrivals per day) at airport i during day t;

$Connect_{it} =$ The number of non-stop flight segments connected to airport i during day t;

$q_k(t) =$ Dummy variable for month q, i.e. $q_k(t) = 1$ if day t belongs to month k and 0 otherwise;

$m_j(i) =$ Dummy variable for airport j, i.e. $m_j(i) = 1$ if $j = i$ and 0 otherwise;

$\beta_0 \ldots \beta_9, \omega_k, \lambda_j =$ dummy coefficients to be estimated;

$\epsilon_{it} =$ stochastic error term.

The dependent variable D_{it}, the average positive arrival delay against schedule per flight, is a standard measure of flight delays available from the FAA's ASPM database, and represents one of the official performance metrics adopted by the U.S. FAA Air Traffic Organization (ATO). This delay metric only reflects positive delays: flights that arrive earlier than schedule are assigned a zero delay value in the calculation. The deterministic queuing delay Q_{it} described above is included in the econometric model as an explanatory variable. We further consider the second and third order terms of Q_{it} as they can help capture other schedule disruptive phenomenon such as flight cancellations in response to exorbitant delays and the effect of delay propagation.

In addition to deterministic queuing delays, adverse weather at an airport will increase flight time by causing air traffic controllers to increase aircraft separation within airspace around the airport. Although in theory the weather effect could be reflected in the deterministic queuing and AAR variables (as discussed below), such variables may not fully capture this effect (Hansen and Hsiao 2005). As a consequence several weather variables are explicitly introduced in the model. The first two variables are the proportion of quarter-hours under Instrument Flight Rules (IFR) conditions in a day and the quadratic term of this proportion. Daily average wind speed is also included because of either the direct effect of wind itself or the associated conditions such as wind shear that may impact airport capacity and not be adequately captured by the recorded AAR. Furthermore, we include temperature as it has proven to be another causal factor to airport delay (Hansen and Wei 2006).

We also hypothesize that flight delay can be affected by the size and network connectivity of an airport, and include AAR and the number of non-stop flight segments connected to the airport under study ($Connect$) as two separate explanatory variables. The AAR variable is included because high AAR's tend to be set more conservatively – that is, at a lower level relative to the absolute maximum throughput – than low AAR's (Neufville et al. 2003).

The $Connect$ variable is calculated based on the number of airports to which the observed airport has commercial non-stop flights on a given day. We expect high connectivity to complicate aircraft turnaround operations and increase the exposure of the airport to delay propagated from other airports, and may therefore make the airport more susceptible to delays. Airline hubs are especially prone to such effect as the integrity of flight schedule is more fragile due to connecting banks.

Finally, to account for monthly and airport-specific effects that are not captured by the above explanatory variables, a set of dummy variables are employed. The model includes 11 monthly dummies, with December used as the baseline month. As an example, February day would have the February dummy set to 1 and all other monthly dummies to 0. Similarly, there are 34 airport dummies with TPA as the baseline airport.

One may argue that airport concentration may exhibit as well some effect on flight delays to the extent that the delay that airlines cause themselves is internalized (Morrison and Winston, 2008). An airport Herfindahl-Hirschman Index

(HHI) variable is often introduced in statistical delay models to serve this purpose. However, the delay impact of HHI results from its impact on the flight schedule, characteristics of which are already accounted for in this model. The debate on whether hub airlines (fully) internalized their delays (Daniel 1995, Brueckner 2002, Mayer and Sinai 2003, Daniel and Harback 2008, Rupp 2009), therefore, does not seem to be relevant, and such a variable is not included in our model.

2.4.2.3 Data, Model Estimation and Results.

The delay model is estimated on a daily dataset of the OEP35 airports covering the year 2007. The variables included in the model are constructed using two data sources: the FAA ASPM database and the US Bureau of Transportation Statistics (BTS) Airline On-time Performance database. The former provides quarter-hour based information about flight schedule, runway capacities, and meteorological conditions at major US airports. All variables except for the *Connect* are obtained from the ASPM database. The BTS Airline On-time Performance database documents individual flight information, such as the scheduled and actual departure and arrival time, and origin and destination airports, for each domestic flight operated by carriers that account for at least one percent of domestic scheduled passenger revenues. Such information is used to construct the *Connect* variable. In the data collection process, we observe a number of airport-days for which some of the required data are missing from ASPM, and are therefore excluded from the dataset. Days in which there was a transition between standard and daylight saving time are also excluded

(March 10-11 and November 3-4 in 2007) considering the frequent reporting errors on such days. In total, the airport-day dataset contains 12,605 usable observations.

In estimating the model several econometric issues need to be considered. First, since airport operations are interdependent in the National Airspace System (NAS), it is important to account for this interdependency in estimating the model. Second, it is likely that individual airports in the panel have features that consistently increase or decrease delay, leading to a need to include airport fixed effects. Third, errors in econometric delay models have been found to be heteroskedastic (Wei and Hansen, 2006). Finally, serial correlation among error terms may persist because, among other reasons, delay at the end of a day could possibly affect the operations of the next day. We therefore perform a Prais-Winsten regression by allowing a first-order autocorrelation between observations for the same airport. Panel corrected standard errors are employed, in which error terms are assumed heteroskedastic and contemporaneously correlated across panels (i.e. errors are correlated across airports at a given point in time). Estimation results are presented in the Table 2.1.

In general, the coefficients have the expected signs as previously discussed. The first-order queuing delay variable has a highly significant coefficient, whose value is very close to one. The coefficients for the quadratic and cubic queuing delay terms are also highly significant, with diminishing magnitude which is natural as the higher-order terms have greater absolute values. Greater prevalence of IFR conditions results in high delay, but the effect is not linear as the second-order IFR term has a negative coefficient. Consistent with previous studies (e.g. Hansen and Hsiao

	Estimate	Std. Err.		Estimate	Std. Err.
$Cons(\beta_0)$	15.7834***	3.0695	$DEN(\lambda_8)$	-5.4005*	2.4175
$Q(\beta_1)$	1.0562***	0.0305	$DFW(\lambda_9)$	-8.7964**	3.3074
$Q^2(\beta_2)$	-0.0073***	0.0004	$DTW(\lambda_{10})$	-15.9171***	2.4457
$Q^3(\beta_3)$	1.53E-05***	1.11E-006	$EWR(\lambda_{11})$	-11.3089***	1.4861
$IFR(\beta_4)$	16.4138***	1.2369	$FLL(\lambda_{12})$	-0.6885	0.7155
$IFR^2(\beta_5)$	-9.5336***	1.3383	$HNL(\lambda_{13})$	0.931	1.4899
$Wd(\beta_6)$	0.1887***	0.0353	$IAD(\lambda_{14})$	0.2958	0.6282
$Temp(\beta_7)$	-0.0476*	0.0192	$IAH(\lambda_{15})$	-9.8916***	2.4193
$AAR(\beta_8)$	-0.0081***	0.0006	$JFK(\lambda_{16})$	-9.3924***	1.1149
$Connect(\beta_9)$	0.2554***	0.0428	$LAS(\lambda_{17})$	-8.2991***	1.1843
$Jan(\omega_1)$	-4.1829**	1.418	$LAX(\lambda_{18})$	-7.0825***	1.1062
$Feb(\omega_2)$	-0.5913	1.4483	$LGA(\lambda_{19})$	-4.9853***	1.0944
$Mar(\omega_3)$	-0.9399	1.4373	$MCO(\lambda_{20})$	-0.7347	0.6904
$Apr(\omega_4)$	-3.6284*	1.4456	$MDW(\lambda_{21})$	-7.3218***	0.9345
$May(\omega_5)$	-3.3868*	1.4678	$MEM(\lambda_{22})$	-3.7089***	0.87
$Jun(\omega_6)$	2.6321	1.5094	$MIA(\lambda_{23})$	10.1735***	0.7077
$Jul(\omega_7)$	1.8042	1.5227	$MSP(\lambda_{24})$	-18.5847***	2.5349
$Aug(\omega_8)$	0.5598	1.529	$ORD(\lambda_{25})$	-16.8544***	3.5873
$Sep(\omega_9)$	-5.2635***	1.4978	$PDX(\lambda_{26})$	-2.8426**	1.0401
$Oct(\omega_{10})$	-4.3928**	1.447	$PHL(\lambda_{27})$	-3.9256***	0.8351
$Nov(\omega_{11})$	-5.5085***	1.4291	$PHX(\lambda_{28})$	-7.8380***	1.3608
$ATL(\lambda_1)$	-20.5981***	4.9125	$PIT(\lambda_{29})$	9.2130***	1.1994
$BOS(\lambda_2)$	-4.9724***	0.7836	$SAN(\lambda_{30})$	-9.5330***	0.9438
$BWI(\lambda_3)$	-7.2301***	0.7373	$SEA(\lambda_{31})$	-8.3987***	0.7468
$CLE(\lambda_4)$	-11.6438***	0.7591	$SFO(\lambda_{32})$	-7.5172***	0.7514
$CLT(\lambda_5)$	-7.9047***	0.5803	$SLC(\lambda_{33})$	-13.2007***	1.911
$CVG(\lambda_6)$	-6.8763***	1.8915	$STL(\lambda_{34})$	-0.6465	0.6538
$DCA(\lambda_7)$	-6.6269***	0.8221			
R^2		0.5488	Autocorr coeff (ρ)		0.3114

Table 2.1: Delay model estimation results.

*** significant at 0.1% level, ** significant at 1% level, * significant at 5% level

2005, Hansen and Kwan 2010), higher delay values are associated with stronger winds and lower average temperature. As expected, larger AAR seems to reduce average delay, whereas greater connectivity contributes to higher delays. Ceteris paribus, the months of February, March, June, July, and August would experience the same level of delays as that in December because of their statistically insignificant dummy coefficients. Interestingly, the bulk of airport dummy coefficients are negative and significant, implying that, all else equal, delays at most airports will be lower than at TPA.

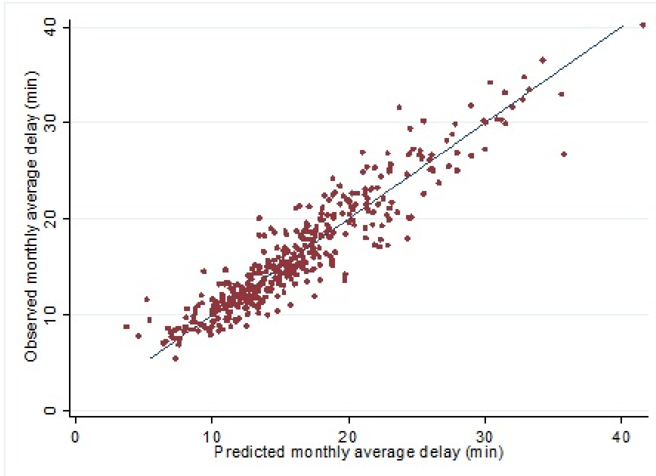


Figure 2.4: Monthly average delays by airport: predicted vs. observed (Mean Absolute Percentage Error, MAPE: 10.75%.

Since monthly average delays are of the major concern in the subsequent analysis, we use the estimated delay model to perform in-sample prediction with monthly averaged data. The predicted airport-month values are plotted against the observed average delays, as shown in Figure 2.4. We observe that most data points are concentrated along the 45-degree line, to some degree validating the prediction power of the model. Mean Absolute Percentage Error (MAPE) of the monthly predictions with respect to the monthly averages observed values is quite small (10.75%). This model will be used in Section 2.5 to predict new monthly average delays brought by changes in the deterministic queuing delay variables given various slot control situations.

2.4.2.4 Computation of Passenger Queuing Delay Cost Savings.

With the deterministic and econometric queuing delay models, quantifying passenger queuing delay cost savings when slot control is introduced involves the following steps. Given an airport and a slot control level, we first generate a new schedule profile given by either the *FlightMove* or the *FlightTrim* model. The new schedule profile enters the deterministic queuing delay model to produce Q_{it}, which then feeds into the econometric model to yield the predicted average queuing delay per flight. This delay value is compared with the predicted average delay in the absence of slot control using the same econometric model. The difference represents the average delay savings per flight at the airport. Following our assumption that the queuing delay experienced by a passenger is the same as the queuing delay of his/her flight, total passenger queuing delay cost savings is the product of the average delay savings per flight, the average number of passengers on a flight, the number of arriving flights at the airport, and the passenger value of travel time. In the case of *FlightMove*, we assume that the number of passengers carried on each flight will be adjusted such that all existing passengers continue to be served after introducing slot control. This assumption is especially pertinent when considering the long run responses of airlines. We use the US Department of Transportation recommended value for the passenger value of travel time, which equals $37.5/hr when inflated to 2007 US dollars (DOT 2003).

2.5 Results of Combined Model: the need for increased slot controls in the US

The schedule delay and queuing delay models developed in Section 2.4 are used to quantify the incremental scheduled delay costs and queuing delay cost savings respectively, when slot control is imposed at an airport. We consider three slot levels for each of the OEP35 airports: the peak airport capacity, and its 90% and 80% levels. The peak airport capacity would generally be the recommendation that most easily would gain community acceptance. However, further benefits (in terms total cost reduction) may be achieved by setting slot level below the peak airport capacity. To this end, we also look at costs and benefits associated with lower slot levels in this analysis.

Table 2.2 provides the results of the *FlightMove* model for all of the OEP35 airports. Fifty predicted schedules generated by using the simulation algorithm in subsection 2.4.1.2 are used for computing queuing delay and schedule delay cost change. The MQDC_80, MQDC_90 and MQDC_100 columns give the mean daily cost savings resulting from a reduction in queuing delay with slot controls imposed at the 80, 90 and 100% of the airport peak capacity respectively. Similarly, the MSDC_80, MSDC_90 and MSDC_100 columns give the mean daily increase in cost associated with schedule delay with slot controls imposed at the 80, 90 and 100% levels respectively. A.2 provides standard deviations and z-scores for the estimates. The small values of standard deviation and z-scores relative to the mean strongly support the robustness of the procedures adopted.

To justify slot control as a realistic option at an airport, the measure must yield savings in queuing delay of significant magnitude and well in excess of the corresponding increase in schedule delay costs. A high margin is required to offset any additional sources of cost that could arise from implementing slot controls. Examples of such cost include basic administrative costs to the government and

	Mean Queuing Delay Cost Savings			Mean Schedule Delay Cost Increase		
Airport	MQDC_80	MQDC_90	MQDC_100	MSDC_80	MSDC_90	MSDC_100
ATL	$220,239	$183,658	$125,873	$347,507	$53,643	$26,317
BOS	$660	$311	$304	$1,779	$301	$102
BWI	$2,505	$2,429	$1,880	$609	$320	$146
CLE	$61,484	$55,995	$45,690	$10,710	$6,935	$4,786
CLT	$120,556	$111,800	$94,128	$21,654	$12,449	$7,845
CVG	$1,663	$1,297	$994	$1,342	$422	$67
DCA	$40,464	$40,399	$32,796	$11,612	$5,449	$3,173
DEN	$2,125	$633	$0	$725	$138	$0
DFW	$0	$0	$0	$0	$0	$0
DTW	$86,706	$71,233	$55,364	$28,239	$16,525	$8,924
EWR	$435,155	$223,109	$68,011	$1,915,486	$306,853	$37,446
FLL	$0	$0	$0	$0	$0	$0
HNL	$0	$0	$0	$0	$0	$0
IAD	$36,785	$28,190	$20,012	$6,720	$4,130	$2,355
IAH	$6,488	$2,814	$0	$1,669	$424	$0
JFK	$217,960	$87,359	$66,631	$449,293	$97,747	$32,114
LAS	$6,694	$4,334	$2,781	$3,444	$800	$181
LAX	$17,328	$12,551	$3,874	$27,119	$3,362	$793
LGA	NA	$131,228	$40,451	NA	$445,154	$8,904
MCO	$0	$0	$0	$0	$0	$0
MDW	$2,666	$1,944	$864	$1,075	$372	$68
MEM	$2,139	$1,609	$353	$376	$162	$29
MIA	$0	$0	$0	$0	$0	$0
MSP	$48,766	$32,074	$16,143	$17,208	$8,331	$4,000
ORD	NA	$78,665	$56,177	NA	$67,904	$8,440
PDX	$0	$0	$0	$0	$0	$0
PHL	$168,166	$124,377	$102,694	$269,887	$47,075	$19,096
PHX	$58,037	$57,575	$49,256	$11,186	$6,332	$4,270
PIT	$0	$0	$0	$0	$0	$0
SAN	$36,661	$36,661	$35,881	$54,931	$7,596	$2,576
SEA	$16,978	$12,540	$8,531	$14,000	$1,836	$717
SFO	$4,134	$2,227	$470	$957	$270	$52
SLC	$5,985	$5,203	$2,996	$3,418	$1,165	$298
STL	$0	$0	$0	$0	$0	$0
TPA	$0	$0	$0	$0	$0	$0

Table 2.2: *FlightMove* model results for OEP35 airports – mean daily values over 50 simulation runs. Sixteen airports (italicized) are selected for further analysis.

costs incurred by the flight operators in planning their response (and subsequently operating in the presence of slot controls). In addition, slot controls could restrict competition, leading to higher air fares. One should therefore set a high benefit threshold for the imposition of slot controls. We chose as a cutoff for consideration a daily queuing delay cost savings of $10,000 at 90% capacity slot level. Based on this criterion, 16 airports realize significant queuing delay savings when slot controls are introduced (highlighted in Table 2.2) and are thus examined further as potential

Airport	MQDC_80	MSDC_80	Drop in Service	MQDC_90	MSDC_90	Drop in Service
EWR	\$473,510	\$121,540	11%	\$432,941	\$43,685	4%
JFK	\$560,851	\$142,433	9%	\$440,361	\$52,422	4%
LGA	\$214,102	\$61,634	15%	\$197,771	\$27,282	7%
ORD	\$165,723	\$101,711	7%	\$141,185	\$18,971	2%

Table 2.3: *FlightTrim* model results for airports expected to drop service on imposing slot controls (daily values)

candidates for this measure.

We assume that as long as *FlightMove* is feasible, airlines would always prefer *FlightMove* to *FlightTrim*, because *FlightMove* preserves baseline demand without requiring changes in fleet mix. *FlightTrim* may be justified under two situations. First, it may be impossible to attain slot limits without reducing flights. We find two cases of this. When slot controls are imposed at 80% AAR, for LGA and ORD it is not possible to service the existing scheduled flights. That is, it is not possible to move the scheduled flights among the various time windows so that all time windows are below the slot limit. The other situation is when there are protracted periods when total flights exceed total slots. Specifically, we find that, when slot controls are imposed at the 90% or 80% level, at least one of the 4-hour time windows at EWR, JFK, LGA, and ORD will encounter insufficient capacity to service the scheduled demand. As a result, *FlightTrim* is applied at 80% and 90% levels at these airports. The other twelve airports do have sufficient capacity even after imposing slot controls at the 80% level; however the schedules will have to be further flattened beyond the 90% level, leading to larger MQDC and MSDC values. Table 2.3 gives the *FlightTrim* results for the four airports in question.

Table 2.4 provides further information on the 16 airports under consideration. It provides net benefits for the three Slot Levels (SLs) as the difference between the cost of imposing the slot control at that level, i.e., MSDC, and the benefits, i.e. MQDC: (i) MQDC_80 – MSDC_80, (ii) MQDC_90 – MSDC_90, and (iii) MQDC_100 – MSDC_100. Next, it gives the incremental benefit of imposing slot controls at a given slot level compared to the next highest slot level, i.e.: (MQDC_80 – MSDC_80) – (MQDC_90 – MSDC_90) in column (iv), and (MQDC_90 – MSDC_90) – (MQDC_100 – MSDC_100) in column (v). Then, it gives a measure of return on implementing slot control as the benefit-cost ratios: (vi) MQDC_80 / MSDC_80, (vii) MQDC_90 / MSDC_90, and (viii) MQDC_100 / MSDC_100. Finally, our recommended slot level is presented in column (ix), the recommended slot level in terms of arrival operations per hour is given in column (x).

For the airports listed, net benefit levels are positive at the 100% and 90% slot levels. This provides support for a more widespread use of slot controls at these airports. When considering whether to restrict slots at the 100%, 90%, or the 80% capacity levels, one should consider the incremental savings achieved by proceeding from the 100% level to the 90% level, and further to the 80% level, i.e. the columns (v) and (iv). We note that (v) is positive for all airport listed except: PHL and SAN. Thus, a direct interpretation of the results suggests that slot control at 100% of the capacity is cost justified at all airports in the Table 2.4 and slot controls at the 90% level are justified at all airports in the table except PHL and SAN. Upon examining column (iv), we note that further benefits are possible by implementing slot controls at the 80% level for six airports: CLE, CLT, DTW, IAD, JFK, and MSP. Of the four airports where the *FlightTrim* model is applicable, only JFK has a positive incremental benefit at 80% level over 90% level, albeit with a higher drop

	Net Benefit(SL) = MQDC_SL − MSDC_SL			Incremental Benefit(SL)		Benefit Ratio(SL) $=\frac{\text{MQDC\_SL}}{\text{MSDC\_SL}}$			Recommended SL	
Slot Level (SL)	80%	90%	100%	80%	90%	80%	90%	100%	%age of AAR	Arrivals/hr
Airport	(i)	(ii)	(iii)	(iv)=(i)-(ii)	(v)=(ii)-(iii)	(vi)	(vii)	(viii)	(ix)	(x)
ATL	-$127,268	$130,015	$99,556	-$257,283	$30,459	0.63	3.42	4.78	90%	103
CLE	$50,774	$49,059	$40,904	$1,715	$8,155	5.74	8.07	9.55	80%	31
CLT	$98,902	$99,351	$86,283	-$449	$13,068	5.57	8.98	12.00	90%	63
DCA	$28,851	$34,950	$29,623	-$6,099	$5,327	3.48	7.41	10.34	90%	32
DTW	$58,467	$54,708	$46,440	$3,758	$8,268	3.07	4.31	6.20	80%	57
EWR*	$351,971	$389,256	$30,565	-$37,285	$358,691	3.90	9.91	1.82	90%	40
IAD	$30,065	$24,060	$17,657	$6,004	$6,403	5.47	6.83	8.50	80%	67
JFK*	$418,418	$387,939	$34,517	$30,478	$353,423	3.94	8.40	2.07	80%	41
LAX	-$9,791	$9,189	$3,081	-$18,980	$6,108	0.64	3.73	4.89	90%	71
LGA*	$152,468	$170,489	$31,547	-$18,021	$138,943	3.47	7.25	4.54	90%	35
MSP	$31,558	$23,743	$12,143	$7,815	$11,600	2.83	3.85	4.04	80%	49
ORD*	$64,012	$122,215	$47,737	-$58,202	$74,477	1.63	7.44	6.66	90%	87
PHL	-$101,722	$77,302	$83,597	-$179,024	-$6,295	0.62	2.64	5.38	100%	50
PHX	$46,851	$51,243	$44,986	-$4,392	$6,257	5.19	9.09	11.54	90%	71
SAN	-$18,270	$29,065	$33,305	-$47,335	-$4,240	0.67	4.83	13.93	100%	24
SEA	$2,977	$10,704	$7,814	-$7,727	$2,890	1.21	6.83	11.90	90%	38

Table 2.4: Summary of Results from Combined Models (daily values).
(*) Using *FlightTrim* model results for MQDC and MSDC for 80% and 90%.

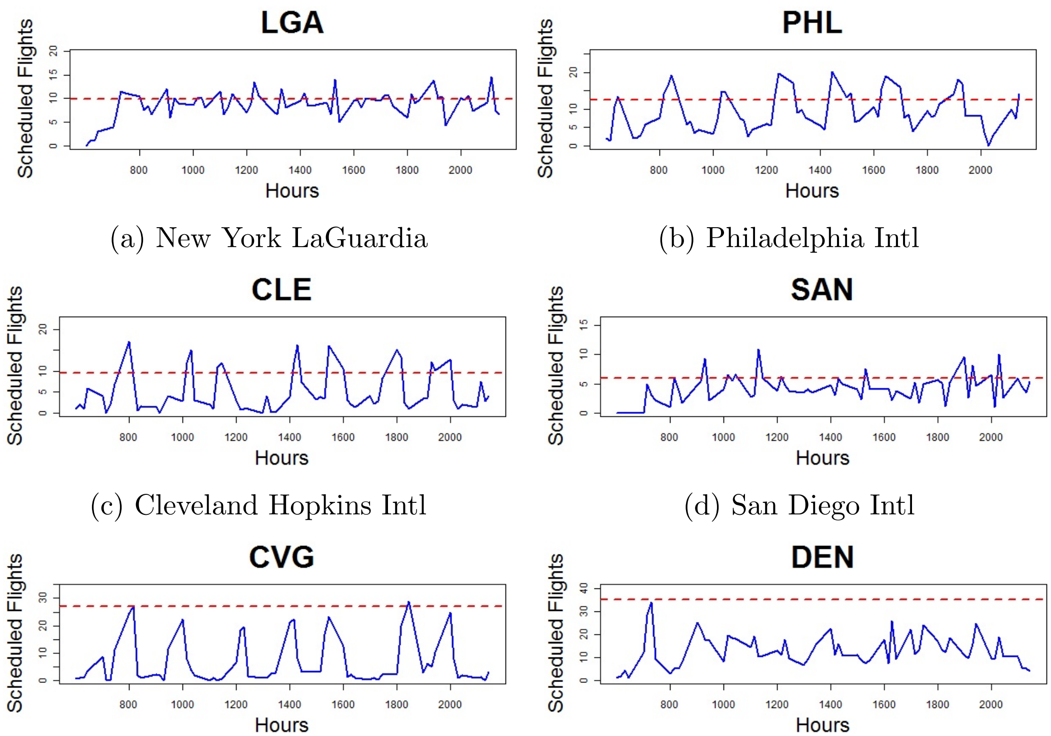
in service.

Furthermore, this list of 16 airports that may benefit by imposing slot controls includes all airports that currently have slot controls: DCA, EWR, JFK and LGA. Our results suggest that the current caps at these airports are set too high: 90% level is most beneficial for DCA, EWR, LGA; and 80% level is most beneficial for JFK. Finally, the list contains airports, such as CLE, MSP, SAN and SEA, which are not normally considered to be highly congested. On the other hand, they do have pronounced peaks; thus the results suggest that imposing controls that would reduce such peaks by spreading flights to less congested periods are worthwhile. This is illustrated in Figure 2.5, which shows schedules for some highly, mildly, and least congested airports.

As an airport can serve primarily leisure or business travelers, the benefit estimates can be different as the two types of passengers involve different values of schedule delay and travel time. To investigate the sensitivity of the above results to different time values, we recalculate the potential benefit gains for two extreme cases. As mentioned in Section 4.1.4, we choose one third and twice the base Λ value as the new value of passenger schedule delay for MSDC calculation, representing the extreme cases of only leisure and only business travelers using an airport. Following DOT (2003),we use values of travel time for leisure and business passengers as 81% and 140% of the average to generate new MQDC estimates.

Overall, the results are rather insensitive to the variations in passenger value of schedule delay and travel time. We observe the same 16 airports that would reap benefits from implementing slot controls. Assuming an airport are used by leisure travelers only, maximum benefits at LGA, CLT, EWR and PHL could be achieved by further reducing the optimal slot level by 10%. For the other 12 airports, the recommended slot levels stay the same. If all air travel is assumed for business purposes, the optimal slot levels at the 16 airports would remain unchanged – except for DTW and JFK, which would have 90% instead of 80% of their respective peak airport capacity as the optimal slot level. Clearly, these small changes, which are associated with very extreme distribution of leisure/business passengers, suggest the robustness of the general conclusions from our analysis to specific passenger composition at the airports.

Using the base case values, we obtain a first-order estimate of annual benefits from queuing delay reduction by summing up the estimates over the 16 airports and multiplying it by 365 (days). When slot controls are implemented *at the current peak capacity* at all the sixteen airports, a net annual benefit of $237 million is indicated by the study. The annual benefits could be significantly increased if our



(a) New York LaGuardia (b) Philadelphia Intl

(c) Cleveland Hopkins Intl (d) San Diego Intl

(e) Cincinnati/Northern Kentucky Intl (f) Denver Intl

Figure 2.5: Aug 2007 Aggregated Arrival Schedules for several airports. Congestion levels decrease from top to bottom. The dotted line shows the peak arrival capacity (AAR).

recommended slot levels are implemented to \$629 million.

Associated with the latter figure is a total queuing delay saving at about \$0.8 billion in 2007. It is important to note that the estimate mainly captures reduction in delay that is within the control of the National Airspace System, or NAS delay. Using individual flight delay records from the BTS On-time Performance Database, we compute the fractions of delay by causes in August 2007 (BTS 2007). Figure 2.6 shows that, NAS delay constitute 24.56% of the total; the other two major causes are air carrier delay and aircraft propagated delay. Air carrier delays can result from mechanical breakdowns and various operational problems not related to congestion, e.g. problems boarding passengers. On the other hand, some air carrier delays are in fact indirectly caused by NAS delays, e.g. crew timeouts. All aircraft propagated delay is the result of an earlier delay that could either be a NAS delay or an air carrier delay. Thus, while congestion *directly* causes 24.56% of recorded delays, it indirectly causes another large portion of the total delay. Ball et al. (2010) find that the total passenger delay cost in the US in 2007 directly caused by flights arriving late equals $4.7 billion. Under the assumption that the distribution of delay causes remains stable throughout the year, an entire elimination of NAS delay would generate a cost saving of .2456 * $4.7 = $1.2 billion (if we further assume that other sources of delays contribute to aircraft propagated delay based on their share of delay minutes, the delay cost saving from slot controls and overall saving will be proportionately larger). Our estimate of $0.8 billion suggests that the proposed slot control policies would help reduce delay cost by 67 % (0.8/1.2). This estimate, however, does not include components such as delay cost reduction due to passenger misconnections or canceled flights which are considered in Ball et al. (2010). As discussed above, it is also the case that air carrier delays and propagated delays would be substantially reduced as a result of reducing NAS delays. Therefore, the $.8 billion figure should be regarded as a lower bound rather than a precise estimate of passenger delay cost reduction. In fact, the discussion above suggests that the proper application of slot control would eliminate 67% of congestion related delays.

While we would recommend a cautious approach to deciding on the exact set of airports where slot controls should be imposed, we feel at an aggregate level strong conclusions could be made. Specifically, these results indicate that slot controls should be imposed more broadly across the U.S. and where they are imposed the

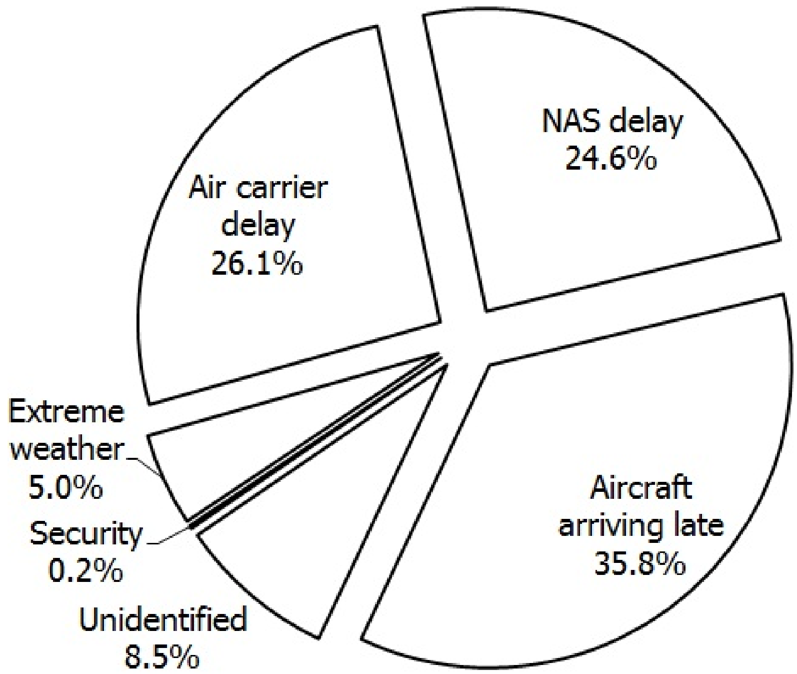


Figure 2.6: Causes of delay published by Bureau of Transportation Statistics for Aug 2007.

slot levels (caps) should be set at lower values than has been the norm.

2.6 Conclusions and Further Discussion

This paper provided a comprehensive investigation of the fundamental trade-offs associated with implementing airport slot controls. The potential cost and benefit to travelers – realized in the form of increased schedule delay and reduced queuing delay – were explicitly examined using a set of quantitative models and simulation procedures. The results from empirically analyzing the OEP35 airports suggest that more widespread use of slot controls in the U.S. would improve traveler welfare. The recommended slot control level varies at different airports depending upon where the maximum net benefits can be achieved. Applying this logic to airports that currently have slot controls, we found the current slot caps at DCA, EWR, JFK, and LGA are set too high. Further slot reduction through deleting some flights that are currently scheduled at these airports could generate additional value to passengers. Robust to demand split between leisure/business passengers at the airports, these findings offer helpful insights and reference for future policy making in airport congestion management.

One needs to, however, bear in mind some uncertainties surrounding these conclusions. First, a typical airport serves origin-destination as well as connecting passengers, for whom penalty from flight schedule perturbation would be realized through the layover time change. Provided that relevant data are available, it would be worthwhile to investigate the sensitivity of our results to this passenger differentiation. Second, in the present study we did not explicitly consider the impact of slot control on airline revenue and cost. Intuitively, decreased queuing delay reduces airline operating cost, allowing airlines to charge lower fare. Empirical evidence implies that such effect is rather minimal (Zou and Hansen 2011). On the other hand, limited slots make air travel more valuable goods, adding potential for price increase, which may partially offset or even reverse the benefits from queuing delay reduction net of schedule delay cost increment for leisure travelers, given their low value of travel time. Business passengers, however, may still be better off from slots because of their much higher valuation of time (such effects has been discussed in road pricing, e.g. Hau (1992). While in a different context, the insights also hold for slots). From the competition perspective, slot control can affect airport concentration by forcing some airlines to cut operations more substantially or even eliminate services in certain markets. As a consequence, carriers with greater market concentration could charge passengers higher fares; whereas the fringe competitors, in order to maintain a foothold at the airport, might be forced to offer lower ticket prices. The overall effect on fare may be neutralized to some extent. Uncertainties over airline cost changes arise primarily from two sources: immediate cost savings due to delay reduction, and potential cost increase associated with schedule adjustment. One possible reason for the latter could be the new aircraft purchase in order to meet the adjusted schedule under slot control. The multiple effects on both revenue and cost sides confound any conclusion about airline profitability. For local airports or pertinent public entities, while slot control generates revenues, additional cost can result from government administration in response to slot controls. The distribution of rents from slot controls among carriers, airports or public authorities would further depend upon the initial allocation procedures. These uncertainties suggest the need to set a relatively high threshold for either the passenger queuing delay savings (as is the case in the paper) or the estimated net passenger benefits in order to make more affirmative recommendations.

Even when slot control is economically justified, several practical issues need to be carefully considered and addressed before any successful implementation of airport slot control schemes. First is the access of small communities. Due to the relative sparse schedule, undifferentiated slot control policies can exert disproportionately adverse impact on small communities. This has been widely acknowledged and used as an argument for opposing slot control (FAA 2006, PANYNJ 2008). Politically acceptable airport demand management schemes have to demonstrate their ability to insure adequate access of these small communities. A second issue concerns who should be responsible for making and implementing demand management policies. While airport authorities might be the natural candidates for this role, several arguments suggest the FAA would be a more appropriate choice. As natural monopolies, airports are highly regulated in the U.S. The revenue-neutral objective combined with arbitrary size of revenue generated by slot control requires that slot control be implemented by an entity other than the airport itself. The FAA has the legal responsibility for the efficient operations of the National Airspace System and therefore, when controlling access at individual airports, can bear a national perspective in mind. An ideal solution would be taking both national and local perspectives into account in devising and instituting slot controls. A third issue is slot allocation, or, more pointedly, the determination of which carriers must eliminate or re-schedule flights. While an auction appears to be a reasonable allocation mechanism, this has proved to be highly problematic in practice. As a fourth issue, the very concept of slot ownership deserves further attention. Through secondary markets, slot owners in the U.S. include non-air-carriers (e.g. banks). Local communities can be part of the slot owners to insure access to the major airport in questions. There could be unintended consequences, however, such as airport opponents purchasing slots in order to retire them. Slot control also needs to be reconciled with international bilateral agreements and should avoid creating substantial inconvenience for international/domestic connections. Finally, slot control, while reducing queuing delay which is a typical signal to indicate the needs for infrastructure investment, should maintain its appropriate signaling mechanism (e.g. high slot prices) so as to not unduly suppress new infrastructure investment.

Even after these challenges are resolved, implementing slot controls at all the airports simultaneously may be a high-risk endeavor. Practical issues like equitable allocation among the airlines of the exact flights to be reduced from the congested time-slots; settlement time for the new schedules to take effect; training of various personnel in the airline industry, airports, and the FAA; adaptation of IT systems and services etc would need to be handled in a careful manner so as to not disrupt the passenger service. Indeed, the entire exercise may seem too daunting to undertake despite the economic benefits, although the same might be said for the alternative of capacity expansion in the case of many airports. Interestingly, the list of airports identified as suitable for slot controls is diverse in many respects: it includes airports spanning the entire geography; of small, medium, and larger capacities; from mildly to highly congested regions; and has recommendations at various slot levels. This could prove useful in de-risking the entire initiative, by phasing the implementation at carefully selected lower-risk pilot airports first. The implementation may be conducted at the pilot airports, and the benefits – as well as challenges – established before taking upon the other airports.

Acknowledgment

Earlier versions of this paper were presented at 2010 and 2011 INFORMS annual meetings in Austin, Texas and Charlotte, North Carolina, and at the INFORMS Transportation Science and Logistics (TSL) Society Workshop in Pacific

Grove, California. The authors would like to thank session participants for their suggestions. We are also grateful to the anonymous reviewers for their very helpful comments.

Chapter A: (Appendix to Chapter 2)

A.1 *FlightMove* Simulation Algorithm

We present a sketch of the simulation procedure here, the cost determination aspect is explained in the following subsection.

Initialize. Set iteration counter:

$$i \leftarrow 1.$$

Repeat steps 1–4 while there are time-points that have flights in excess of the given slot level:

$$\max_t \left(p_t^i\right) > 0,$$

where $\mathbf{p^i}$ is a vector of 64 non-negative real numbers comprising of the excess of flights over the slot level for each time-period:

$$\left\{\mathbf{p^i} \quad | \quad p_t^i \leftarrow \max\left(0, \sum_d N_{dt}^i - SL\right) \quad \forall \quad t \in 1, \ldots, 64\right\}.$$

Step 1. Make a multinomial draw to select time-period to move flights from:

$$\mathbf{s^i} \sim \text{multinomial}\left(64, \mathbf{p^i}\right); \quad \hat{t}^i = \arg\max_t s_t^i.$$

$\mathbf{s^i}$ is a vector of 64 non-negative integers that sum to 64, resulting from a single multinomial draw. $\mathbf{p^i}$ is normalized to sum to 1 before the draws are made. The time-periods where total flights are below the given slot level have 0 probability of being selected for moves. This scheme favors the time-slots with more excess flights to be selected for move. Note that it does so only in probability, as against always selecting the time-period with maximum excess as the target.

Step 2. Make a multinomial draw to compute destination-wise proportions of excess flights to move from the target time-point:

$$\mathbf{m^i} \sim \text{multinomial}\left(|\mathbf{D}|, N^i_{d\hat{t}}\right).$$

$\mathbf{D}$ is the vector of all destinations. This draw generates a vector $\mathbf{m^i}$ of non-negative integers that sum to total number of destinations, using the number of flights to each destination at the selected time-points as probability distribution (after normalizing so as to sum to 1). Next, a logit-link is applied to the draw:

$$\left\{\mathbf{f^i} \quad | \quad f^i_d \leftarrow \frac{exp(m^i_d)}{\sum_d exp(m^i_d)}\right\}.$$

$\mathbf{f^i}$ is a vector with a non-negative real number for each destination, its entries sum to 1. Multiplying this vector with the total excess flights at target time-point results in a vector of flights to move for each destination:

$$\left\{\mathbf{e^i} \quad | \quad e^i_d \leftarrow f^i_d \cdot \left(\sum_d N^i_{d\hat{t}} - SL\right) \quad \forall \quad d\right\}.$$

Step 3. Make a multinomial draw to determine direction of move for each destination:

$$\left\{\mathbf{c^i} \quad | \quad c^i_d \sim \text{multinomial}\left(1, \left\{{}^1/_3, {}^1/_3, {}^1/_3\right\}\right) \quad \forall \quad d\right\}.$$

$\mathbf{c^i}$ is a vector with 3 entries for each destination, with a single entry as 1 and other two entries as 0 – drawn with equal probability for all the three positions. The three positions depict respectively, $\hat{t}-1$ (one time-point earlier than the $\hat{t}$), $\hat{t}$ (implying no move), $\hat{t}+1$ (one time-point later than $\hat{t}$). Finally, the flights are moved in the selected direction for each destination:

$$N_{dj}^{i+1} \leftarrow N_{dj}^{i} + e_d^i \cdot c_d^i \quad \forall \quad d \quad \forall \quad j \in \left\{\hat{t}-1, \hat{t}, \hat{t}+1\right\},$$

followed by move out from target time-point:

$$N_{d\hat{t}}^{i+1} \leftarrow N_{d\hat{t}}^{i} - e_d^i \cdot c_d^i \quad \forall \quad d.$$

Step 4. Update iteration counter:

$$i \leftarrow i+1.$$

There are further technical implementation details to get a quick convergence that we omit here for brevity.

A.2 *FlightMove* Simulation Results

Table A.1 reports the results obtained over 50 simulation runs for *FlightMove* model for each airport in the study. The first set of columns is for incremental queuing delay costs, and the other set is for incremental schedule delay costs. We report here the standard deviations over the runs, followed by z-scores of the minimum and maximum values obtained over the runs (means are reported in Table 2.2). Most of the z-scores are within ∓ 3, indicating robustness over runs.

	Standard Deviations, MQDC			z-Score of (Min, Max), MQDC			Standard Deviations, MSDC			z-Score of (Min, Max), MSDC		
Airport	MQDC_80	MQDC_90	MQDC_100	MQDC_80	MQDC_90	MQDC_100	MSDC_80	MSDC_90	MSDC_100	MSDC_80	MSDC_90	MSDC_100
ATL	371	853	3007	(-2.37,2.01)	(-2.26,2.59)	(-2.26,1.82)	6141	2761	836	(-1.27,2.78)	(-1.49,2.53)	(-1.34,1.86)
BOS	127	68	23	(-2.22,1.69)	(-2,1.74)	(-2.37,2.44)	267	29	11	(-1.42,2.25)	(-1.93,1.61)	(-1.49,1.78)
BWI	0	1	4	–	(-2.03,1.8)	(-1.62,2.02)	31	20	15	(-1.74,1.53)	(-1.72,1.97)	(-1.93,1.59)
CLE	281	399	580	(-2.21,2.91)	(-2.4,2.4)	(-2.22,1.89)	819	447	422	(-1.4,1.93)	(-2.21,1.81)	(-1.33,1.87)
CLT	444	454	1142	(-1.99,1.88)	(-1.81,2.06)	(-1.84,2.85)	569	516	190	(-1.6,2.03)	(-1.8,2.06)	(-1.54,2.31)
CVG	0	1	15	–	(-1.99,2.06)	(-2.15,1.99)	101	49	6	(-1.82,2.21)	(-1.77,2.24)	(-1.73,1.95)
DCA	0	28	384	–	(-2.92,1.52)	(-2.07,2.48)	811	497	340	(-1.81,1.52)	(-1.33,2.41)	(-1.58,1.59)
DEN	5	53	0	(-1.94,2.01)	(-2.15,2.13)	–	89	10	0	(-1.7,1.62)	(-1.26,1.57)	–
DFW	0	0	0	–	–	–	0	0	0	–	–	–
DTW	96	315	375	(-2.29,3.01)	(-2.27,2.76)	(-1.84,3.52)	1353	771	425	(-1.43,1.79)	(-2.09,2.01)	(-1.48,2.36)
EWR	1842	1039	1715	(-1.5,1.53)	(-1.78,2.3)	(-2.59,1.94)	39568	10000	2139	(-2.06,1.6)	(-2.06,1.5)	(-1.37,2.49)
FLL	0	0	0	–	–	–	0	0	0	–	–	–
HNL	0	0	0	–	–	–	0	0	0	–	–	–
IAD	414	437	713	(-1.06,3.39)	(-2.25,2.72)	(-1.76,2.79)	497	377	201	(-1.34,1.9)	(-1.86,2.23)	(-1.45,1.78)
IAH	369	84	0	(-2.06,2.37)	(-1.88,2.59)	–	112	52	0	(-1.74,2.38)	(-1,2.44)	–
JFK	4227	2227	1221	(-2.31,0.93)	(-2.16,2.67)	(-2.49,2.07)	19948	4006	1536	(-1.17,2.18)	(-2.1,1.62)	(-1.5,2.62)
LAS	137	150	127	(-2.28,2.92)	(-2.14,1.29)	(-2.63,2.08)	302	117	15	(-1.33,2.3)	(-1.9,2.62)	(-1.21,2.14)
LAX	0	333	103	–	(-1.75,2.15)	(-1.89,2.74)	1009	178	52	(-1.56,1.6)	(-1.53,1.9)	(-2.17,1.97)
LGA	NA	870	1419	NA	(-1.34,2.45)	(-2.88,1.6)	NA	10863	483	NA	(-1.07,2.25)	(-1.55,1.72)
MCO	0	0	0	–	–	–	0	0	0	–	–	–
MDW	43	47	23	(-2.1,1.64)	(-2.35,2.22)	(-1.98,2.67)	83	22	5	(-2.17,1.88)	(-2.11,1.42)	(-1.42,1.88)
MEM	8	8	5	(-1.32,3.11)	(-1.9,2.32)	(-2.19,1.89)	66	36	7	(-1.77,1.89)	(-1.63,1.22)	(-1.81,1.75)
MIA	0	0	0	–	–	–	0	0	0	–	–	–
MSP	1181	1197	516	(-1.73,1.92)	(-1.84,2.37)	(-2.02,2.25)	750	272	180	(-2.01,1.35)	(-1.77,1.68)	(-1.18,2.82)
ORD	NA	282	1241	NA	(-1.78,2.2)	(-1.99,2.13)	NA	5955	546	NA	(-1,2.43)	(-1.77,1.72)
PDX	0	0	0	–	–	–	0	0	0	–	–	–
PHL	2128	1361	757	(-2.73,2.09)	(-2.94,1.57)	(-2.79,1.66)	11032	3347	485	(-1.97,2.44)	(-1.07,1.74)	(-1.62,1.38)
PHX	43	52	399	(-0.62,3.57)	(-3.71,1.02)	(-3.16,1.12)	836	512	329	(-1.45,1.86)	(-1.44,1.44)	(-1.32,2.02)
PIT	0	0	0	–	–	–	0	0	0	–	–	–
SAN	0	0	216	–	–	(-2.13,1.63)	2719	788	156	(-1.71,1.33)	(-1.67,1.87)	(-1.35,1.94)
SEA	0	298	187	–	(-1.78,1.92)	(-2.51,1.58)	2291	106	26	(-1.64,1.6)	(-2.15,1.67)	(-1.56,1.91)
SFO	349	193	46	(-1.77,2.36)	(-2.5,2.57)	(-2.01,2.25)	99	28	5	(-1.66,1.93)	(-1.63,1.5)	(-1.74,1.82)
SLC	3	36	55	(-2.36,1.8)	(-2.5,1.74)	(-1.51,2.85)	180	108	29	(-1.89,1.92)	(-2.2,1.83)	(-1,1.98)
STL	0	0	0	–	–	–	0	0	0	–	–	–
TPA	0	0	0	–	–	–	0	0	0	–	–	–

Table A.1: *FlightMove* model results for OEP35 airports – spread of results over 50 simulation runs (based on daily values)

A.3 U.S. Operational Evolution Partnership (OEP) 35 Airports

Refer Table A.2.

Airport code	Airport	City	State
ATL	Atlanta Hartsfield Intl	Atlanta	GA
BOS	Boston Logan Intl	Boston	MA
BWI	Baltimore-Washington Intl	Baltimore	MD
CLE	Cleveland Hopkins Intl	Cleveland	OH
CLT	Charlotte Douglas Intl	Charlotte	NC
CVG	Cincinnati-Northern Kentucky Intl	Covington-Cincinnati, OH	KY
DCA	Washington Reagan Natl	Washington	DC
DEN	Denver Intl	Denver	CO
DFW	Dallas-Ft Worth Intl	Dallas-Ft Worth	TX
DTW	Detroit Metropolitan Wayne County	Detroit	MI
EWR	Newark Intl	Newark	NJ
FLL	Ft Lauderdale-Hollywood Intl	Ft Lauderdale	FL
HNL	Honolulu Intl	Honolulu	HI
IAD	Washington Dulles Intl	Washington	DC
IAH	George Bush Intercontinental	Houston	TX
JFK	John F Kennedy Intl	New York	NY
LAS	Las Vegas McCarran Intl	Las Vegas	NV
LAX	Los Angeles Intl	Los Angeles	CA
LGA	La Guardia	New York	NY
MCO	Orlando Intl	Orlando	FL
MDW	Chicago Midway	Chicago	IL
MEM	Memphis Intl	Memphis	TN
MIA	Miami Intl	Miami	FL
MSP	Minneapolis-St Paul Intl	Minneapolis	MN
ORD	Chicago O'Hare Intl	Chicago	IL
PDX	Portland Intl	Portland	OR
PHL	Philadelphia Intl	Philadelphia	PA
PHX	Phoenix Sky Harbor Intl	Phoenix	AZ
PIT	Pittsburgh Intl	Pittsburgh	PA
SAN	San Diego Intl-Lindburgh Field	San Diego	CA
SEA	Seattle-Tacoma Intl	Seattle	WA
SFO	San Francisco Intl	San Francisco	CA
SLC	Salt Lake City Intl	Salt Lake City	UT
STL	Lambert-St Louis Intl	St Louis	MO
TPA	Tampa Intl	Tampa	FL

Table A.2: List of the U.S. Operational Evolution Partnership (OEP) 35 Airports

Chapter 3: Designing the Noah's Ark: A Multi-objective Multi-stakeholder Consensus Building Method

A significant challenge of effective air traffic flow management (ATFM) is to allow for various competing airlines to collaborate with an air navigation service provider (ANSP) in determining flow management initiatives. This challenge has led over the past 15 years to the development of a broad approach to ATFM known as collaborative decision making (CDM). A set of CDM principles has evolved to guide the development of specific tools that support ATFM resource allocation. However, these principles have not been extended to cover the problem of providing strategic advice to an ANSP in the initial planning stages of traffic management initiatives. In the second essay, we describe a mechanism whereby competing airlines provide "consensus" advice to an ANSP using a voting mechanism. It is based on the recently developed Majority Judgment voting procedure. The result of the procedure is a consensus real-valued vector that must satisfy a set of constraints imposed by the weather and traffic conditions of the day in question. While we developed and modeled this problem based on specific ATFM features, it appears to be highly generic and amenable to a much broader set of applications. Our analysis of this problem involves several interesting sub-problems, including a type of column generation process that creates candidate vectors for input to the voting process.

3.1 Introduction

A shared perception of a common, imminent, unavoidable, impactful threat or opportunity oftentimes leads even fierce competitors to seek consensus solutions. The mythical Noah's Ark is indeed witnessed in the real-world of business. For example, technology standards bodies have been the foundation for inter-operability of the products and services offerings of firms competing for the same or similar customers. American National Standards Institute (ANSI), a consortium of industry and researchers, performed this key function during the entire Industrial Age; while the more recent Internet Society serves similarly in the Information Age. In the highly competitive airline industry, we see examples of airline alliances which have helped airlines maximize their offerings and reach through collaborating with other competing airlines. At another level, whenever there is bad weather, the airlines come together with the Air Navigation Service Provider – namely, the Federal Aviation Administration (FAA) in the US – to devise effective means to handle the constrained system resources.

Future visions of Air Traffic Flow Management (ATFM) – both in the U.S. and Europe – support a "performance-based" approach that employs collaboration between the air navigation service providers (ANSPs) and the airlines (ICAO 2005, JPDO 2007, SESAR 2006). A key feature of this outlook is to support the airline operators' business objectives in the ANSP's traffic management initiatives (TMIs), subject only to system-level constraints like safety and security. Our focus is on a performance-based framework that addresses the strategic level planning in advance of the implementation of a TMI. These overarching system performance expectations may then serve as the basis for design and operation of a specific TMI (or a coordinated set of TMIs) that aim to meet the stated expectations.

The framework must (a) be founded upon commonly agreed definitions of service expectations among the several stakeholders, and (b) result in a consensus on the service expectations over independent stakeholders, with possibly conflicting business objectives. We use the Global Air Traffic Management Operational Concept (ICAO 2005) to address the former requirement. Unanimously approved by the U.S. and 187 other States in the eleventh global Air Navigation Conference, it dedicates a section on "expectations of the AT[F]M community." Among 11 performance expectations, three are more specific to the airline operators' business objectives, while the others are more generic to the entire framework – predictability, capacity-utilization, and efficiency. Our focus in this work is on the latter aspect of the framework: given that there is intent to collaborate among the stakeholders, how to design an effective consensus solution that encompasses multiple inter-related objectives.

We postulate six properties as highly desirable for any effective solution for the stated problem: (i) consensus-building, (ii) single solution determination, (iii) practicality, (iv) equitability, (v) confidentiality, and (vi) strategy-proof. These are consistent with the principles of mechanism design (Maskin 2008), and also take into account some specific needs of our application environment.

The first three are desirable for purely pragmatic reasons: it is our stated wish that the method determines an acceptable solution among the multiple stakeholders; the method would be most effective if the method results in an unambiguous solution; and that the method does not take inordinate time and / or effort on the part of the decision makers to yield its solution.

The latter three are higher-level properties. Given that we are dealing with possibly competing stakeholders, we shall like the method to adhere to well-accepted notions of equitability, specifically we shall like the voice of each stakeholder to be fairly represented in the decision making process. Further, as we are likely dealing with independently operating businesses, the method should not require information that may be deemed confidential. Finally, we shall like the method to discourage any strategic behavior among the decision makers.

A recently proposed voting scheme called "Majority Judgment" has many desirable properties (Balinski and Laraki 2011). Of primary interest to us has been its high strategy-resistance – while it does not preclude gaming of the system, the probability of a single player to significantly game the system is severely restricted in this design. We therefore base our proposal on Majority Judgment.

In section 3.2, we describe the problem and present related literature. Section 3.3 focuses on the mathematical models that underpin the proposed mechanism. After reasonably structuring the underlying information, it presents efficient solution methods. Validation is provided in Section 3.4 through simulation experiments on a large dataset motivated from real-life. Section 3.5 concludes.

3.2 General Problem Statement and Related Work

The general context for the problem we address involves a group of n stakeholders, N, who jointly seek to make a decision. It is not necessarily the case that these stakeholders are cooperative or have common goals: in our application, the stakeholders are the flight operators who in fact are competitors. The form of the decision we seek is a numeric vector $\mathbf{m}$ that is subject to a set of feasibility conditions μ so that $\mathbf{m} \in \mu$. The $\mathbf{m}$ we seek should represent a consensus among a majority of the stakeholders. Each stakeholder $i \in N$ has a value or value function $V_i()$ that maps each $\mathbf{m} \in \mu$ to a real number that represents the value of $\mathbf{m}$ to i. The problem we address is to design a mechanism in which a coordinator exchanges information with the stakeholders and produces the desired $\mathbf{m}$. Of course, this is hardly a well-defined problem yet, as in particular, we have not precisely defined a majority consensus $\mathbf{m}$. Nonetheless, this description does allow us to place our problem in a broader context and to discuss the nature of our contributions. In particular, attacking this problem would seem to require key elements from two large bodies of literature: Voting, and Multi-criteria decision making (MCDM).

The case where $\mathbf{m}$ is of dimension 2 and $\mu = \{(1,0),(0,1)\}$ can be viewed as a classic election among two contenders. Each stakeholder would "vote" for either $(1,0)$, expressing a preference for the first candidate or $(0,1)$, expressing a preference for the second candidate. The vector output would indicate the winning candidate. The case of higher dimensional $\mathbf{m}$ with μ consisting of all unit vectors would correspond to an election with several candidates where one must be chosen. The instant runoff voting mechanism and majority judgment represent mechanisms that would produce a single winner candidate/vector.

Voting in particular, and social choice in general, is concerned with aggregating evaluations over a multitude of voters, in ways that the final outcome has appeal to a large section of the decision-makers. Over centuries, investigators devising a fool-proof voting system have been riddled by a result – famously known in social choice theory as Arrow's Impossibility Theorem (Arrow 1951). It states: "when voters have three or more distinct alternatives, no voting system can convert the ranked preferences of individuals into a community-wide (complete and transitive) ranking while also meeting a certain set of criteria, namely: unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives." Majority Judgment is a recently proposed procedure (Balinski and Laraki 2011), that "bypasses" this result. And hence, its authors claim it to be "a better alternative to all other known voting methods, in theory and in practice."

Majority Judgment involves grading – instead of preference rankings – of each candidate, by all voters, in a common language. It is a natural, rich preference elicitation method, already being practiced in spirit in many contests and juries around the world, as well as a few political elections. It has many good properties; among them, high resistance to strategic voting – which makes it appealing for our work. An outline of the procedure with an example follows later.

A key feature of our problem is that μ is very large; in fact, the μ we employ is a polyhedron and so has the structure of the feasible region of a linear programming problem. Our consideration of, and modeling of, this large space of feasible candidate vectors represents the most essential contribution of this paper.

The decision-making framework in the general MCDM involves a decision maker evaluating a set of candidates on the basis of multiple criteria or attributes (Wallenius et al. 2008). A common assumption about the decision maker's or group's actions is consistency with maximization of a utility or value function that depend on the attributes (Raiffa and Keeney 1976). Wallenius et al. (2008) characterize the distinctions between the discrete and the continuous candidate space versions of the MCDM problem. Our work is related to both the versions. Like the continuous version, we iteratively search a continuous candidate space. However, like the discrete version, we do specify a functional form of the decision makers' value functions, and estimate its parameters over several candidates over the iterations.

We generally assume that $V_i()$ is known in some way to each stakeholder i. Thus, we do not devote attention to methods to "discovering" $V_i()$. We note a significant body of research that focuses on this aspect of the problem. In this literature it is generally assumed the stakeholders can provide some preference information, e.g. the ability to choose between pairs of alternatives. The stakeholders are then asked to make various preference decisions to elicit functional forms, e.g. the $V_i()$'s, that allow a decision on a complete option to be made. We note in particular the Analytic Hierarchy Process (AHP), which is a well-regarded tool for multi-criteria decision making (Saaty and Vargas 2012). It relies on pairwise comparisons over a set of alternatives, eliciting preference rankings on several criteria organized in a hierarchy on a nine-point scale. The group version of AHP aggregates the individual scores into group scores using their geometric means – similar to the way it aggregates the scores over the hierarchy of criteria. Any mean is less resistant to extremes (and thus strategic behavior) than median – which is used by Majority Judgment.

Green and Rao (1971) introduced conjoint analysis into marketing literature – which has enjoyed considerable success in marketing applications (Green et al. 2001). A decompositional technique, the method presents respondents with descriptions of alternatives with differing levels on a number of attributes, and records their preference order over the alternatives. For reasons just discussed we do not use these methods to determine utility functions but we do use the functional forms from this literature as part of our estimation process.

As the research progresses, there will be need to approximate the efficient frontier of the feasible candidate space using historical or simulated data on candidate realizations. Hence, on the computational side, research dealing with the problem of approximating the efficient frontier of the continuous candidate space is also relevant to our work, e.g. Data Envelopment Analysis (see Charnes et al. (1978), Cook and Seiford (2009)) and potentially, multi-objective linear programming, (see Ruzika and Wiecek (2005), including methods to approximate the efficient frontier (see Sayın (2000) and Karasakal and Köksalan (2009)).

A multi-criteria decision analysis based approach was adopted in a strategic decision making context by Eurocontrol (Grushka-Cockayne et al. 2008). Similar to our setting, the problem involved the ANSP and the airlines collaboratively arriving at a common decision for selecting operational improvements. Further, the decision was subject to constraints like safety and environmental impact, and was expected to improve on objectives like predictability and efficiency. However, unlike our problem that seeks to evaluate at a day-of-operations level, the Eurocontrol was faced with a one-time strategic decision.

3.2.1 Majority Judgment

Majority Judgment is defined as a social decision function. It involves grading – instead of preference rankings – of each candidate, by all voters, in a common language. It is a natural, rich preference elicitation method, already being practiced in spirit in many contests and juries around the world, as well as a few political elections.

It takes as input the Grades given by the voters, and produces "Majority Grade" of each candidate as an output. These can be used to compute rank-orderings as well (called "Majority Ranking"). Majority Grade of a candidate is the highest grade approved by an absolute majority of the voters. In case of an odd number of voters, it is the median of the grades; if there are even number of voters, then it is the lower middlemost of the grades. Its high resistance to strategic voting primarily results due to this median-seeking property.

Suppose there are six voters, voting on three candidates: C_1, C_2, C_3. They assign one of these five grades to each candidate: Excellent, Very Good, Good, Passable, and Reject. The grades thus obtained by voting are then sorted from

Candidates:	C_1	C_2	C_3	Maj. Gr.
Worst Grade:	Passable	Reject	Reject	MG-5
.	Passable	Passable	Reject	MG-3
.	Good	Passable	Good	MG
.	Very Good	Good	Very Good	MG-2
.	Very Good	Good	Very Good	MG-4
Best Grade:	Excellent	Good	Very Good	MG-6

Table 3.1: Majority Judgment example

worst to best, as given in Table 3.1.

The majority grade for each candidate (marked "MG" in the last column) is the top fourth grade, as majority (four of six) would give at least that grade to the candidate. Row "MG-2" is found similarly after hiding the "MG" row, and so on; these are useful for tie-breaks when ranking candidates. Majority Ranking for the example is: $C_1 \succ C_3 \succ C_2$.

Majority Judgment requires a common language accepted by all the voters for grading the candidates. Grades may be either continuous or discrete (like above). A continuous grading language could be: $\{0, \ldots, 100\}$, where 0 is commonly understood as "unacceptable", and 100 as "most favorable".

The aspect of common language, while being very intuitive and simple to express, is critical to the overall procedure. Any practical implementation has to carefully come up with the common language that is accepted by all the voters. For some applications, the common language is easier to identify as it forms part of the trade, e.g., tea or wine tasting within a company, or assignment grading in classes. In new applications however, specific focus groups with the voters are sometimes conducted to establish the common language. Furthermore, special training and communication procedures are developed to ensure that new entrants to the system are well-conversed with the common language.

3.3 Mechanism Design and Underlying Models

As discussed in the preceding sections, Majority Judgment provides a solution to the challenge we have outlined. However, Majority Judgment cannot be directly applied due to the very large – in fact, infinite – size of the set of "candidates". In this section, we develop an analytical framework and set of models for addressing this issue.

3.3.1 Majority Judgment Winner

Suppose N is the set of n stakeholders – hereafter referred to as players – faced with a potentially infinite set of feasible candidates μ. As discussed in Section 2, each player has a value function V_i that assigns each candidate $\mathbf{m} \in \mu$ a value. In this section we make use of a grade function g_i, which is similarly a function defined on μ. The grade function will be employed by player i to assign a grade to each $\mathbf{m}$ as part of the Majority Judgment process. While g_i is clearly closely related to V_i (and higher V_i values would generally induce higher g_i values), it is possible that a player may consider various strategies for setting g_i based on V_i. However, at this time, we will assume that a simple linear transform is used to produce g_i based on V_i and we will refer to $g_i(\mathbf{m})$ as the value of $\mathbf{m}$ to i. In fact, the only reason that we do not use V_i directly is that we require all grades to use a "common language".

Here this implies $0 \leq g_i(\mathbf{m}) \leq 1$ for all $\mathbf{m}$.

Let b denote a *minimal* majority-forming subset of N, and let β denote the set of all possible minimal majority-forming subsets of N. In a "one-person, one-vote" situation, a majority-forming subset is any set of size $\frac{n}{2}+1$ for n even and $\lceil \frac{n}{2} \rceil$ for n odd. In the weighted case addressed here, each player i is given a weight w_i; the total weight of a minimal majority-forming subset b just exceeds half the total weight of all players:

$$\overline{W} < \sum_{i \in b} w_i; \text{ where } \overline{W} \equiv \frac{\sum_{j \in N} w_j}{2}. \tag{3.1}$$

Requiring the set to be minimal implies that if any element is deleted from b then equation (3.1) would no longer hold. Note that the complement $N - b$ clearly does not form a majority.

The *min grade* for a specific b and candidate $\mathbf{m}$ is $u(\mathbf{m}, b) = \min_{i \in b} g_i(\mathbf{m})$. The Majority Grade $v(\mathbf{m})$ for a candidate $\mathbf{m}$ is the highest grade a majority of players is agreeable to assign it, i.e.

$$v(\mathbf{m}) = \max_{b \in \beta} u(\mathbf{m}, b)$$

A Majority Judgment winner is the candidate $\mathbf{m}^*$ with the highest Majority Grade v^*:

$$v(\mathbf{m}^*) \equiv v^* = \max_{\mathbf{m} \in \mu} v(\mathbf{m}).$$

A Majority Judgment winner $\mathbf{m}^*$ thus guarantees a majority of the players a grade of at least v^*.

Determining a winner over a "small" set of candidates is straight-forward in the presence of a trusted, benign "central planner". The players submit their grades for each candidate to the central planner. The planner then sorts the grades for each candidate, and identifies the median grade for each (lower median in case of even number of players) – this is the *majority grade*. The candidate with the highest majority grade is deemed the winner.

Our challenge is to determine such a winner when the size of μ is very large – perhaps infinite. In fact, the proceeding discussion already implicitly associated a subset of players with the winning candidate. This in turn provides a potential approach to making the candidate search finite in the sense that we could search for the winning minimal majority forming subset rather than the winning candidate. Specifically, if we define for any $b \in \beta$

$$\hat{v}(b) = \max_{\mathbf{m} \in \mu} u(\mathbf{m}, b)$$

then it is easy to see that

$$v^* = \max_{b \in \beta} \hat{v}(b). \tag{3.2}$$

While we have now made a search over a potentially infinite set finite, this reduction depends on the ability to efficiently find $\hat{v}(b)$. The following optimization model can accomplish this:

Subset_Opt(b)

$$\begin{aligned} \hat{v}(b) = \max \quad & z \\ s.t. \quad & z \leq x_i & \forall \quad i \in b \\ & x_i = g_i(\mathbf{m}) & \forall \quad i \in b \\ & \mathbf{m} \in \mu \end{aligned}$$

	C_1	C_2	C_3	C_4
g_1	1.00	.70	.40	.50
g_2	1.00	.90	.70	.85
g_3	.80	1.00	.80	.90
g_4	.60	.75	1.00	.80
g_5	.40	.50	.60	.30

Table 3.2: Sample grade functions for four candidates

We will later show that for applications of interest to us this model can be cast as a linear program.

A special type of minimal majority-forming subset is relevant in Majority Judgment: a *majoritarian set* is a minimal majority forming subset that gives the highest grade to some candidate $\mathbf{m}$. That is,

a $b' \in \beta$ is a *majoritarian set* if there exists an $\mathbf{m} \in \mu$ such that

$$u(\mathbf{m}, b') = \max_{b \in \beta} u(\mathbf{m}, b)$$

To illustrate these concepts consider the example provided in Table 3.2.

Assuming all weights are one, there are $\binom{5}{3} = 10$ minimal majority forming subsets but only two majoritarian sets: $\{1, 2, 3\}$ and $\{2, 3, 4\}$ ($\{2, 3, 4\}$ produces the highest grade for each of candidates C_1, C_2, C_3). Note that player 5 is in no majoritarian set since this player tends to give all candidates a low grade. While the grade functions prevent player 5 from being in any majoritarian set, in the weighted case it is possible that a player could be in no majoritarian set because that player was not in any minimal majority-forming subset. An extreme example could occur if the weight of a single player $\hat{i}$ was greater than $\overline{W}$. In such a case, $\{\hat{i}\}$ would be the only minimal majority forming subset and by necessity the only majoritarian set. All other players could be in no majoritarian set irrespective of how they graded. Of course, we may wish to impose rules or assumptions that prevent some of these extreme cases. For example, we will only consider weighting schemes that do not make a single player a majority and we may also require that each player give at least one candidate a grade of one.

The concept of a majoritarian set can potentially allow us to reduce the search space size since if we define β' to be the set of all majoritarian sets then we can replace Equation (3.2) with:

$$v^* = \max_{b \in \beta'} \hat{v}(b).$$

However, we can in fact reduce the search even more. It should be clear from the preceding discussion that for any $b \in \beta$ there is an $\mathbf{m} \in \mu$ and an $i \in b$ such that $g_i(\mathbf{m}) = u(\mathbf{m}, b) = \hat{v}(b)$, i.e. i is the element of b that assigns $\mathbf{m}$ its minimum grade. In general, a given player i might play such a role for several sets b. We can thus define an optimization problem that determines the highest value of $\hat{v}(b)$ achievable where $i \in b$ and i defines the minimum grade, i.e.

$$\tilde{v}_i = \max\{g_i(\mathbf{m}) : g_i(\mathbf{m}) = u(\mathbf{m}, b), i \in b, \mathbf{m} \in \mu\}$$

We note in general it can be the case that the set optimized over in this expression can be null, in which case $\tilde{v}_i$ is defined to be zero. For example, a player that consistently grades very high could be in many majoritarian sets but might not define the minimum grade for any of them.

We have now developed a new approach to finding v^*, namely:

$$v^* = \max_{i \in N} \tilde{v}_i. \tag{3.3}$$

We now define an optimization model that determines a value closely related to $\tilde{v}_i$ and will allow us to compute v^* using an equation similar to (3.3). This optimization model is defined for any $i' \in N$.

Player_Opt(i')

$$
\begin{aligned}
\tilde{z}_{i'} = \max \quad & x_{i'} & & \\
s.t. \quad & x_{i'} \leq G^{max}(1 - I_i) + x_i & \forall \quad i \in N & \qquad (3.4)\\
& \sum_{i \in N} w_i I_i \geq \overline{W}' & & \qquad (3.5)\\
& I_i \in \mathbb{B} & \forall \quad i \in N & \\
& x_i = g_i(\mathbf{m}) & \forall \quad i \in N & \\
& \mathbf{m} \in \mu & &
\end{aligned}
$$

Here, G^{max} is the maximum grade value and $\overline{W}'$ is the smallest number greater than $\overline{W}$ that can be achieved as the sum of the weights of a subset of players. Note there are two sets of variables. The continuous x_i variables define the grades assigned by each player. The binary I_i variables define the players in the majority forming subset; specifically, $I_i = 1$ implies player i is in the majority forming subset and $I_i = 0$ implies it is not. Constraint (3.4) insures that $x_{i'}$ is the minimum grade in the set. Constraint (3.5) insures that the set has total weight larger than $\overline{W}$.

Proposition 1 *The following hold true:*

1. $v^* = \max_{i \in N} \tilde{z}_i$,

2. *any* i^* *that solves* $\max_{i \in N} \tilde{v}_i$ *also solves* $\max_{i \in N} \tilde{z}_i$,

3. *for any* i^* *that solves* $\max_{i \in N} \tilde{v}_i$*, the corresponding majoritarian set* b *when converted to an* I *vector and the corresponding grade vector when expressed as an* x *vector are an optimal solution to* **Player_Opt**(i^*).

Proof All three results follow from two observations. First, consider any optimal solution to **Player_Opt**(i) for some i and let b^* be the set corresponding to the optimal I vector. Constraint set (3.5) implies that b^* is a majority forming subset. If b^* is not minimal then there is a minimal $b' \subset b^*$ with $\hat{v}(b') \leq \hat{v}(b^*)$. In particular, if $\hat{v}(b') < \hat{v}(b^*)$ then there exists an $i' \in b'$ such that $\tilde{z}_{i'} < \tilde{z}_i$. Second, any majoritarian set b together with a grade minimizing $i \in b$ generates feasible solution to **Player_Opt**(i). $\square$

3.3.2 Structural Assumptions and Efficient Modeling of Feasible Set of Candidates and Grade Functions

We now describe some assumptions regarding the structure of the set of feasible candidates and the grade functions. These are appropriate for our target applications (as well as many others) and also aid in the tractability and modeling of the problem.

Assumption 1 *The feasible candidate space* $\mu \subset \mathbb{R}_+$ *is continuous and has a concave "efficient frontier".*

The concave efficient frontier is a reasonable assumption if: (a) larger values of each individual metric are desirable, and (b) there is a tradeoff required among the metrics – that is, increasing the value of one metric comes at the expense of other(s). The first requirement can be met by suitable transformations if smaller values are more desirable than larger. Tools like Data Envelopment Analysis are dedicated to finding such efficient frontiers among a miscellany of metrics.

Assumption 2 *Each player's value function* $V_i(\mathbf{m})$ *is non-negative, continuous, non-decreasing and concave.*

The non-negativity assumption could be resolved by transformation if the original value function did not have this property. Continuity would seem to be a reasonable assumption (for many applications): very small changes in candidate component values should not induce jumps in value. Non-decreasing is related to the discussion above: higher component values are better. The concavity might perhaps fail in certain settings but in many it could be quite reasonable – expressing a type of diminishing returns property.

Assumption 3 *The common grading language allows for continuous grades in* $\mathbb{G} \equiv [0,1]$*, where a higher grade implies better acceptability by a player.*

This assumption defines a common voting language, which is necessary in Majority Judgment.

Assumption 4 *Each player derives its grade function by a simple linear transformation of its value function. Specifically, define* $V_i^{max} = \max_{\mathbf{m}\in\mu} V_i(\mathbf{m})$*; then* $g_i(\mathbf{m}) = V_i(\mathbf{m})/V_i^{max}$*.*

We also might consider slightly more general transformations. However, in general it is possible (and perhaps profitable) for a player to consider a variety of strategies to set the grade function based on their own value function and also knowledge or assumptions regarding the value functions and/or strategies of the other players. Reducing or eliminating the gain that could be achieved by such "strategic" voting is a very important design consideration. We will address it in future research, currently relying on the strategy-resistance claims of Majority Judgment by its authors.

3.3.2.1 Linear Representation of Feasible Candidate Set and Grade Functions.

The assumptions just described allow us to produce an efficient form of the optimization models previously described. Specifically, Assumption 1 allows us to use a piecewise linear approximation to represent the space of feasible candidates and we can replace $\mathbf{m} \in \mu$ with:

$$\mathbf{c^1} m_1 + \mathbf{c^2} m_2 + \cdots + \mathbf{c^p} m_p \leq \mathbf{c^0} \tag{3.6}$$

where $\mathbf{c}$'s are appropriately defined coefficient vectors.

Assumptions 2 and 3 allow us to use similar piecewise linear approximations in place of the grade functions. We approximate $x_i = g_i(\mathbf{m})$ with

$$\mathbf{d^1}_i m_1 + \mathbf{d^2}_i m_2 + \cdots + \mathbf{d^p}_i m_p + x_i \leq \mathbf{d^0}_i$$

where $\mathbf{d}_i$'s are appropriately defined coefficient vectors. The fact that higher grades are always preferred allows us to replace each equality constraint with a set of inequalities that approximate the grade functions.

3.3.2.2 Grade Function Model.

In the prior Section we showed how to represent the grade functions using linear constraints. However, doing this requires knowledge of the grade functions. In fact, the central planner will only observe the players' voting behavior. Our candidate generation process requires that we approximate player grade function based on these observations. We will do this using statistical models that assume a particular functional form for the grade functions. The functional form we assume is based on well-accepted notions developed by economists and marketing researchers in the fields of choice modeling and multi-attribute valuation (e.g. Meyer and Johnson (1995)).

Each player takes three steps to determine the grade of a given candidate. The first two involve the value function (V_i) and the last converts the value function approximation into the grade function (g_i). First, she determines the value of each individual component of the candidate – holding the other components at constant levels. Second, she integrates the individual valuations of the components into an overall value of the entire candidate. Third, she normalizes the value of the candidate into its grade.

Specific models are now proposed for each step. First, the value of an individual component m_r to i is modeled as a non-decreasing concave function $\nu_{r_i}(m_r)$. The value can be visualized as net profitability gain as the metric value is increased, holding other metrics at constant levels. The concavity assumption models diminishing marginal returns as the metric value increases. Second, the integration step combines the individual value functions as a multiplicative-multilinear function of $\nu_{r_i}(m_r)$'s, modeling complementarities among the valuations over the different component metrics:

$$\tilde{V}_i(\mathbf{m}) = r_{1_i}\nu_{1_i}(m_1) + r_{2_i}\nu_{2_i}(m_2) + r_{12_i}\nu_{1_i}(m_1)\nu_{2_i}(m_2) + \ldots,$$

with non-negative coefficients $r_{1_i}, r_{2_i}, r_{12_i}, \cdots \geq 0$. The non-negativity of the constants implies that higher values are better, and that the individual components are not substitutes to each other. For more than two components, pair-wise interaction terms are added; higher-order interaction terms are ignored. Finally, the normalization step converts the integrated value into a grade, using a simple linear scaling based on the maximum value $\tilde{V}_i^{max}$. The grade function for player i is thus specified as:

$$g_i(\mathbf{m}) = \frac{r_{1_i}\nu_{1_i}(m_1) + r_{2_i}\nu_{2_i}(m_2) + r_{12_i}\nu_{1_i}(m_1)\nu_{2_i}(m_2) + \ldots}{\tilde{V}_i^{max}},$$

Appendix B.1 provides further implementation details.

3.3.3 Iterative Procedure

In practice, the true grade functions $g_i(\mathbf{m})$ will be confidential to the players. We use the functional form just described in a procedure that statistically approximates the grade functions based on each player's observed grades, denoted $\hat{g}_i(\mathbf{m})$. Appendix B.2 provides details on the estimation procedure.

The optimization problems $\mathbf{Subset\_Opt}(b)$ and / or $\mathbf{Player\_Opt}(i)$ can be solved with the estimated grade functions $\hat{g}_i(\mathbf{m})$, for some or all $b \in \beta$ or $i \in N$ respectively. The resultant candidates will be an approximation to those computed

Algorithm 1 Algorithm for Proposed Mechanism

Initialize consideration set of feasible candidates

repeat

Obtain players' grades on the consideration set

Estimate players' grade function coefficients

Generate new feasible candidates and / or ask players for new feasible candidates

Introduce some or all new candidates into consideration set

until stopping criteria met

with the true grade functions. All or a subset of these "generated" candidates are put to vote by the players. This cycle of estimation, new candidate generation, voting is repeated until a stopping criterion is met. Algorithm 1 summarizes the entire mechanism:

3.3.4 Evaluation

The "optimal" candidate $\mathbf{m}^*$ uses the "true" grade functions $g_i(\mathbf{m})$, while the "winning" candidate $\widehat{\mathbf{m}}^*$ emerges after the mechanism run using estimated grade functions $\hat{g}_i(\mathbf{m})$. The two are compared to evaluate accuracy of the procedure.

Deviation between candidates is determined as the Euclidean distance between the two. For p-dimensional candidate space:

$$d_v = \sigma \sqrt{\sum_{s=1}^{p} (\widehat{m}_s^* - m_s^*)^2}.$$

$\sigma = \pm 1$ assigns a sign to differentiate outcomes with negative versus positive deviation.

Recall the majority grade of the optimal candidate is $v(\mathbf{m}^*)$, or v^*. The "true" majority grade of the winning candidate is computed with the true grade functions for the players, and denoted $v(\widehat{\mathbf{m}}^*)$. Deviation in majority grades is determined as:

$$d_g = \left(\frac{v(\widehat{\mathbf{m}}^*)}{v(\mathbf{m}^*)} - 1 \right) \times 100.$$

By definition, d_g cannot exceed 0; however, errors in the piecewise linear approximations of the grade functions may lead to violations.

d_v is an absolute measure, useful in comparing several variants of the mechanism. d_g is relative – akin to "optimality gap", it can be used to assess the overall quality of the mechanism itself.

3.4 Experimental Results

A large simulation experiment was conducted to validate the proposed mechanism using data from real-life operations. The data selection and preparation is explained first. Instead of randomly fixing the "true" grade functions for the different airlines, some judgment was exercised to mimic reality. This intuition was vetted within the research team which has expertise in air traffic flow management. The procedure to draw the coefficients for grade function with quadratic functional form is detailed in the appendix. Determination of each airline's weight is also a practical challenge. Multiple weighting schemes are explained.

3.4.1 Data

October 10, 2007 was selected as the sample date. It was a mid-week (Wednesday), with no exceptional events like holidays or expectations of severe weather. The entire day's scheduled departures were included in the dataset.

In terms of geographical scope, the Chicago area airports – ORD (O'Hare) and MDW (Midway) – were included. Operations of feeder airlines were merged into their main airlines' operations. OAG schedule data was used for calculating the number of flights impacted, the left panel of Table 3.3 sums up results. The setup is representative of real-life: impact of the weather on a part of the National Air Space spanning multiple airports of differing sizes, dominance of a few larger airlines, and a long-tail of smaller airlines.

Heterogeneity in airline operations is evident. The final dataset comprises of 47 airlines, totaling 1603 operations. Six hub-and-spoke airlines make up more than $^3/_4$-th of the operations – 1243 in total. Eight point-to-point airlines make up the next largest group, with 292 operations. 25 international airlines have total 50 operations, three charter airlines have 11 operations, and five cargo airlines have seven operations.

At the airline-level operations, four groups emerge. The first group has three large airlines with large presence: United, American, and Southwest. With over 100 operations each, these make up over 85% of total operations. The second group has five large airlines with small presence: Northwest, Delta, US airways, Continental, and Airtran. With operations between 10 and 100 each, these make up about 8% of total. The third group has between 2 and 9 operations, and comprises of 20 airlines. The fourth group has 19 airlines with a single operation.

3.4.2 Feasible Candidate Space

An adversely impacted day-of-operations will suffer loss in the service performance metrics as compared to a normal day-of-operations. The metrics are interrelated, requiring trade-offs amongst them. For instance, an “aggressive” approach might yield a high capacity-utilization, but at the expense of delaying the time when final decisions on releasing flights are made, thus reducing predictability. On the other hand, a “conservative” strategy may release fewer flights that are closely tracked by the air traffic controllers; thereby yielding a high predictability, but low capacity-utilization. Infinitely many “moderate” strategies can be proposed in the intervening space.

Research conducted by other members of our research team has shown a concave relationship among representative metrics for three performance categories: efficiency, predictability and capacity (Ball et al. 2011). The relationships are developed for a single airport, by varying the time-period during which the airport suffers a reduced capacity due to bad weather. The metrics are normalized to lie between 0 and 1; the infeasible “ideal point” (1,1,1) represents a normal day-of-operations where all the performance metrics are realized at 100% levels. The envelope forms the efficient frontier, while all the interior points serve as feasible region. Two metrics – capacity-utilization and predictability – are used for illustrative purposes here,

Airline	MDW	ORD	Characteristics	Profile	nops	log.2	root.10
United (UA) [a]	10	625	Large hub & spoke, large presence	HL	635 (39.6)	9.31 (10.8)	8.85 (10)
American (AA) [b]		500	Large hub & spoke, large presence	HL	500 (31.2)	8.97 (10.4)	8.16 (9.2)
Southwest (WN) [c]	242		Large point-to-point	LH	242 (15.1)	7.92 (9.2)	6.39 (7.2)
Northwest (NW) [d]	11	23	Large hub & spoke, small presence	SS	34 (2.1)	5.09 (5.9)	3.29 (3.7)
Delta (DL) [e]	6	22	Large hub & spoke, small presence	SS	28 (1.7)	4.81 (5.6)	3.08 (3.5)
US Air (US)		27	Large hub & spoke, small presence	SS	27 (1.7)	4.75 (5.5)	3.04 (3.4)
Continental (CO)	2	17	Large hub & spoke, small presence	SS	19 (1.2)	4.25 (4.9)	2.70 (3.1)
Airtran (FL)	18		Large point-to-point	LH	18 (1.1)	4.17 (4.8)	2.65 (3)
Air Canada		8	International, neighboring regions	LH	8 (0.5)	3 (3.5)	2.02 (2.3)
ExpressJet	3	4	Small point-to-point	LH	7 (0.4)	2.81 (3.3)	1.93 (2.2)
Jetblue		7	Small point-to-point	LH	7 (0.4)	2.81 (3.3)	1.93 (2.2)
Chautauqua		6	Small point-to-point	LH	6 (0.4)	2.58 (3)	1.83 (2.1)
Frontier	6		Small point-to-point	LH	6 (0.4)	2.58 (3)	1.83 (2.1)
Mexicana		6	International, neighboring regions	LH	6 (0.4)	2.58 (3)	1.83 (2.1)
Lufthansa		5	International, business-dominant	HL	5 (0.3)	2.32 (2.7)	1.72 (1.9)
Primaris	5		Small charter	HL	5 (0.3)	2.32 (2.7)	1.72 (1.9)
Alaska		4	Small point-to-point	LH	4 (0.2)	2 (2.3)	1.60 (1.8)
Air Midwest	4		Small charter	HL	4 (0.2)	2 (2.3)	1.60 (1.8)
Aeromexico		3	International, neighboring regions	LH	3 (0.2)	1.58 (1.8)	1.45 (1.6)
British Airways		3	International, business-dominant	HL	3 (0.2)	1.58 (1.8)	1.45 (1.6)
Polar Air Cargo		3	Cargo	SS	3 (0.2)	1.58 (1.8)	1.45 (1.6)
Spirit		2	Small point-to-point	LH	2 (0.1)	1 (1.2)	1.26 (1.4)
Aer Lingus		2	International	SS	2 (0.1)	1 (1.2)	1.26 (1.4)
Air Canada Jazz		2	International	SS	2 (0.1)	1 (1.2)	1.26 (1.4)
Lot - Polish		2	International	SS	2 (0.1)	1 (1.2)	1.26 (1.4)
SAS Scandinavian		2	International	SS	2 (0.1)	1 (1.2)	1.26 (1.4)
Singapore		2	International	SS	2 (0.1)	1 (1.2)	1.26 (1.4)
USA 3000		2	Small charter	HL	2 (0.1)	1 (1.2)	1.26 (1.4)
Air France		1	International	SS	1 (0.1)	–	1 (1.1)
Air India		1	International, economy-dominant	LH	1 (0.1)	–	1 (1.1)
Air Jamaica		1	International	SS	1 (0.1)	–	1 (1.1)
Alitalia		1	International	SS	1 (0.1)	–	1 (1.1)
All Nippon		1	International	SS	1 (0.1)	–	1 (1.1)
British Midland		1	International	SS	1 (0.1)	–	1 (1.1)
Iberia		1	International	SS	1 (0.1)	–	1 (1.1)
Japan International		1	International	SS	1 (0.1)	–	1 (1.1)
KLM-Royal Dutch		1	International	SS	1 (0.1)	–	1 (1.1)
Korean		1	International	SS	1 (0.1)	–	1 (1.1)
Martinair Holland		1	International	SS	1 (0.1)	–	1 (1.1)
Pakistan International		1	International, economy-dominant	LH	1 (0.1)	–	1 (1.1)
Swiss		1	International	SS	1 (0.1)	–	1 (1.1)
Turkish		1	International	SS	1 (0.1)	–	1 (1.1)
Virgin Atlantic		1	International	SS	1 (0.1)	–	1 (1.1)
ABX		1	Cargo	SS	1 (0.1)	–	1 (1.1)
Cargoitalia		1	Cargo	SS	1 (0.1)	–	1 (1.1)
Custom Air		1	Cargo	SS	1 (0.1)	–	1 (1.1)
Kalitta		1	Cargo	SS	1 (0.1)	–	1 (1.1)
TOTAL	307	1296			1603 (100)	86.03 (100)	88.38 (100)

[a]includes several United feeders like Go Jet, YV, Shuttle America, United / Skywest, Trans Air; [b]includes American Eagle; [c]includes ATA; [d]includes Mesaba; [e]includes Skywest, Comair, Atlantic Southeast.

Table 3.3: MDW and ORD airline-wise scheduled departures on 10 Oct, 2007.

the proposed procedures extend to any number of metrics.

3.4.3 "True" Grade Functions

Airlines can be broadly classified along several dimensions. (i) number of operations: large, medium, or small airline. (ii) type of network: hub-and-spoke airline, or point-to-point. (iii) type of operations: cargo or passengers. (iv) customer focus: business-dominant, or economy-dominant, or type-independent. (v) distance of markets: long-haul, or short-haul. (vi) political markets served: domestic, or international.

To make the setup realistic, these differentiating factors should be reflected in the grade function of the airlines. Some judgment was exercised in modeling the airline grading behavior; it was vetted within the extended research team, which has expertise in air traffic flow management.

Between the two metrics, we first assessed how each airline would value the two relatively. The possibilities are: "HL", "LH", "SS", where H indicates High, L Low, and S Same; the letters pertaining to predictability and capacity utilization respectively. It does not matter if absolute levels are either both H or both L, as the normalization process would not differentiate between the two. Airline characterizations and their posited profiles are summarized in the middle panel of Table 3.3.

We posit large hub-and-spoke airlines with a significant presence, United and American in this instance, to have HL profile, as they have a large pool of aircrafts to re-balance the impacted passengers – so long as they know the impact adequately in advance. Hence, they would care a lot more about predictability than capacity utilization. However, this cannot be said of the other large hub-and-spoke airlines with a small presence (Northwest, Delta, US Air, Continental), hence we assign them the neutral SS profile.

The low-cost point-to-point airlines – of any size – are hypothesized to prefer capacity utilization than predictability. Their predominantly economy passengers are likely interested in completing their itinerary, without a significant time-sensitivity. Hence, we assign LH to large point-to-point airlines (Southwest and Airtran), as well as the smaller ones (ExpressJet, Jetblue, Chautauqua, Frontier, Alaska, and Spirit). We posit the opposite should hold for luxury or time-sensitive passenger focused Charter airlines. Primaris, Air Midwest, USA 3000 are, therefore, assigned HL profile.

We treat the international airlines serving the neighboring countries to be similar to the point-to-point operators, and assign Air Canada, Mexicana De Aviacion, and Aeromexico LH profile too. Lufthansa and British Airways are posited to cater to more time-sensitive passengers, hence assigned HL profile, while Air India and Air Pakistan are treated as opposite and therefore assigned LH profile. All the remaining international airlines are assigned the neutral SS profile . Finally, cargo carriers are also posed to value the two metrics similarly – and are assigned SS profile.

Next, we assessed the degree of curvature for the value function of each individual metric. The possibilities are: small curvature (straight-line like) and large curvature (more concave). We posit that the airlines with smaller operations would have a straight-line like curvature, as they would not have as much degree of freedom as the airlines with larger presence. The latter are more likely to observe increasingly diminishing returns, and hence, would have a more concave shape.

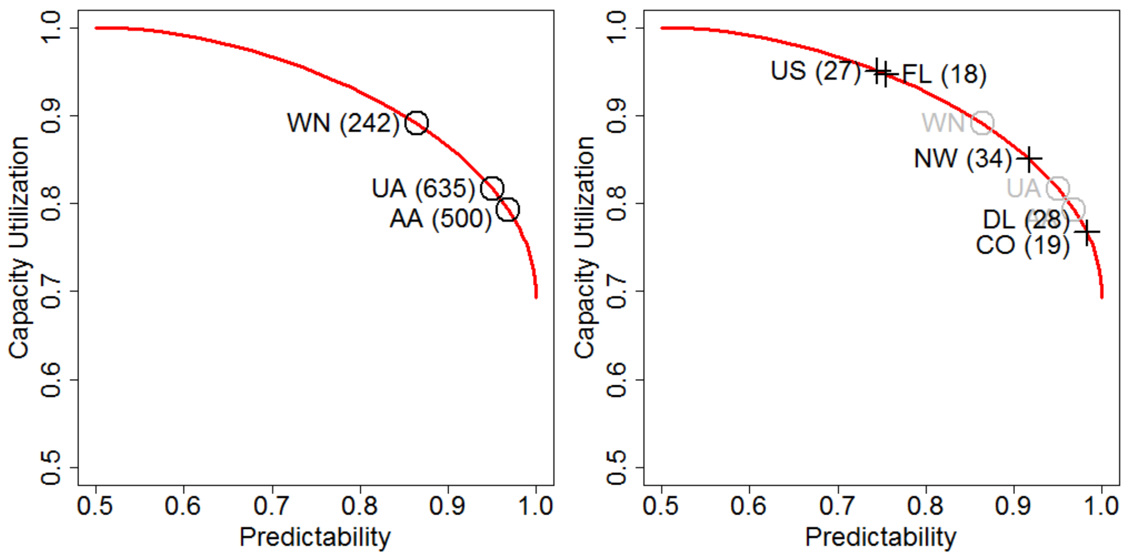
Appendix B.3 explains implementation of this intuition using quadratic functional form for the airlines' grade functions. The grade-maximizing candidates are plotted in Figure 3.1 for the various groups of airlines. The diversity shows the effectiveness of the procedure.

3.4.4 Weights

The democratic "one-person, one-vote" assigns a weight of one to all the airlines ("eqwt"). This may be perceived as inequitable in many practical decision-making contexts though. E.g., in our case, it implies that airlines with a single operation get same representation as those with hundreds of operations. Nonetheless, this is a benchmark for evaluating other weighting schemes.

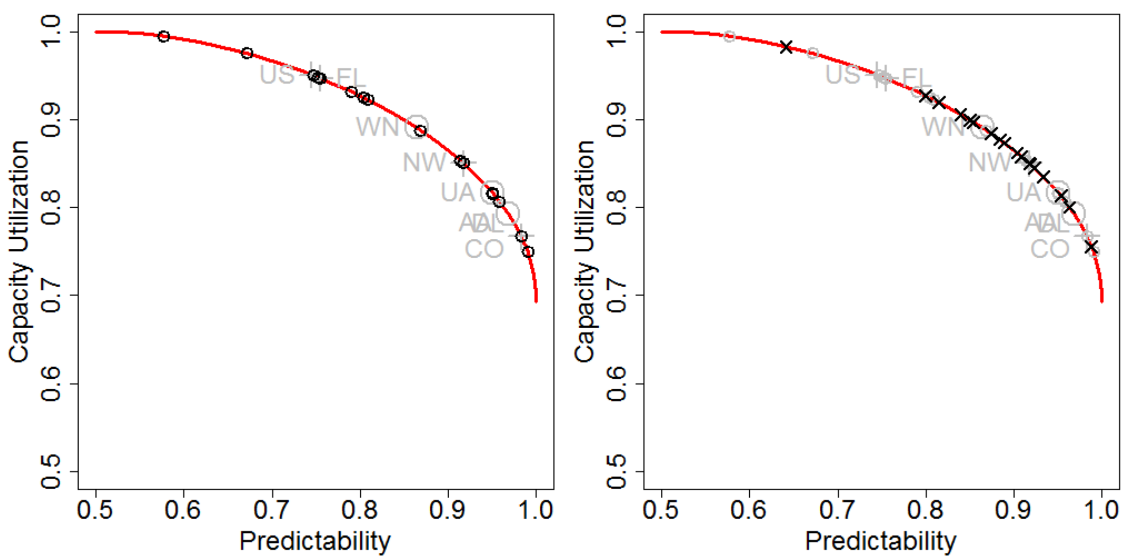
Proportional representation can be achieved by replicating each voter's grade as many times as her weight. A basis is needed for determining the weights. To keep matters simple, practical, and minimal in private information, we use publicly available data on total operations impacted as the basis. It is also a very relevant measure to use in the current context.

The weights are traditionally seen as integers, with the interpretation as given above. In our case though, weights can be fractional. A majoritarian set is formed by a set of voters if the proportion of their combined weight is strictly above 0.5.



(a) Large airlines with large presence

(b) Large airlines with small presence



(c) Small airlines with > 1 operation

(d) Small airlines with one operation

Figure 3.1: Grade-maximizing candidates for different groups of airlines

A simple scheme could use the number of operations as weights ("nops"). However, few voters may get significantly high influence. In this instance, United alone has about 40% operations, the top two airlines make over 70% of total operations. Thus, it may be beneficial to balance the influence of the larger voters.

Logarithmic and power-root transforms on the number of operations would reign in the large positive numbers. However, the choice of base is an open decision. We tried logarithms to three well-known bases: 10, e, and 2, and selected the base 2 for our experiment ("log.2"). The other two bases had lesser differentiation among the airline weights – for the two largest airlines: $\log_{10}(635) = 2.8; \log_{10}(500) = 2.7$ and $\ln(635) = 6.45; \ln(500) = 6.21$. As $\log(1) = 0$, the log transform assigns weight of zero to the airlines with a single operation – which may or may not be desirable. In this instance, the airlines with single operations are mostly international and cargo airlines. If eliminating these is seen as inequitable, a $\log(.) + 1$ would ensure that all airlines have some say in the mechanism.

Alternately, fix the largest airline's proportion of total weight at some desired level, say π_{max}. Power-root transforms can accomplish this. To get π_{max} of 30%, 20%, and 10% ("root.30", "root.20", "root.10") in our example, these are respectively: 1.32585, 1.80390, and 2.96015.

While all of these are valid choices, the exact decision of which one to choose would not be taken at the time of each mechanism run. This decision should be made experimentally, and then left unchanged for a relatively long period of time, until there are reasons to reconsider.

We will evaluate results with four weighting schemes: eqwt, nops, log.2, and root.10. log.2 eliminates the 19 airlines with a single operation. root.10 has similar proportional weight for United as log.2.

3.4.5 Mechanism Design Choices

At this stage, all the inputs for running the procedure are ready. There are a few design choices still to be made though.

3.4.5.1 Initial consideration set.

To initiate the mechanism, the ANSP could provide the airlines a set of candidates. The airlines may heuristically arrive at the grades, through possibly comparing the candidates among themselves.

Alternately, it could communicate the feasible candidate space, and request the airlines to provide their grade-maximizing candidates – to be graded 1. This may be perceived as equitable as the airlines get to submit their most preferred candidates upfront. It also addresses the scaling problem, as the grade of 1 is clearly established for each airline at the outset. However, it does need the airlines to solve a type of profit-maximization problem with feasibility constraints.

Our initial experiments found the former approach converging faster than the latter. Hence, we initialize the consideration set with five or more equally spaced candidates, as there are five coefficients to be estimated for the quadratic value functions.

3.4.5.2 Extent of agreement.

Majority Judgment is a median-seeking procedure. The median has the desirable property that it exactly balances the number of votes that find a candidate's grade too high with those that find its grade too low (Galton 1907). This property will be lost in seeking a non-median based solution, and may encourage strategic behavior.

Having said that, the procedure can be easily extended to allow for any higher (or lower) level of agreement. When seeking a higher (lower) agreement, the Majority Grade of the final candidate could be smaller (higher). Alternate criteria may be explored, for instance, one that seeks a minimum number of airlines to be in the majoritarian set. Any deviations should be subjected to a strategic behavior analysis. In the experiment, the extent of agreement is set at the original, 50% of total weight.

3.4.5.3 Voter input.

At the end of any round, the ANSP may ask for the grade-maximizing candidates from the airlines (if not already done). Alternatively, the ANSP may choose not to ask the airlines for their input. In our experiments, we adopted the latter.

Variants of this alternative may be adopted in practice. For instance, it may be made optional for certain airlines – e.g. those with smaller number of operations, who may possibly not have sufficient infrastructure, and / or stake in the current decision-making context. Furthermore, smaller subsets of airlines may be requested after each round. This would ensure that the consideration set is kept manageable over the rounds. For maintaining equity though, the selection of airlines may be made random, or through a preset procedure.

3.4.5.4 Consideration set update.

At the end of each round, the voter input and the ANSP-generated new candidates are available. A balance has to be made between the size of the consideration set and its quality. Among the new candidates, one could select few candidates with the highest Majority Grades. In our initial experiments, we found that this strategy led to inferior final winners. The inherent error in the estimation of the grade functions is likely the cause.

On the other hand, adding all the new candidates would lead to very large consideration sets. The ANSP may select few diverse candidates among the voter input and new candidates – or it could randomize the selection.

In our experiments, we added all the new candidates generated at the end of each round into the consideration set. This was so we could learn about convergence of the overall procedure with a large input. Results from this experiment would serve to benchmark other strategies in future.

3.4.5.5 Consistency in grading.

We assume the airlines grade every candidate precisely, and report the grades truthfully. In real-life, one or both of the assumptions may not hold, necessitating establishment of consistency rules. In our experiments though, no consistency checks are required. This experiment establishes a benchmark to evaluate the results with different consistency rules.

3.4.5.6 Stopping criterion.

We chose a simple stopping criterion of six rounds for the experiments – in the interest of convergence. More sophisticated stopping criteria should be evaluated against the benchmark established herein.

3.4.6 Mechanism Evaluation

Several runs of the mechanism were conducted with varying parameters. This section reports evaluations in terms of accuracy and technical performance measures.

Figure 3.2 shows the optimal and the winning candidates for different weighting schemes, for one of the runs. In this run, apart from root.10, all the other weighting schemes produced winners very close to the optimal candidates.

As just explained, the initial consideration set is a key parameter of interest. We increased the size of the initial consideration set from 5 through 35, in steps of 10 – the respective runs are called "Init5" through "Init35". Experiments with larger sizes did not yield any significant improvements.

Figure 3.3a plots the percentage deviation in the majority grades of the winning candidate relative to that of the optimal candidate, d_g. The median absolute percentage error is about 0.013%. By complete enumeration of the majority grades

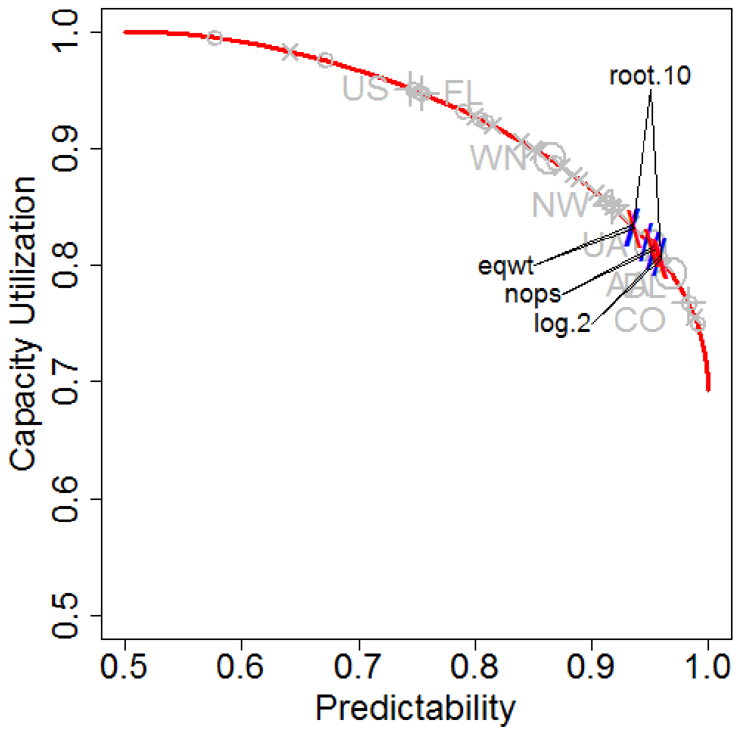


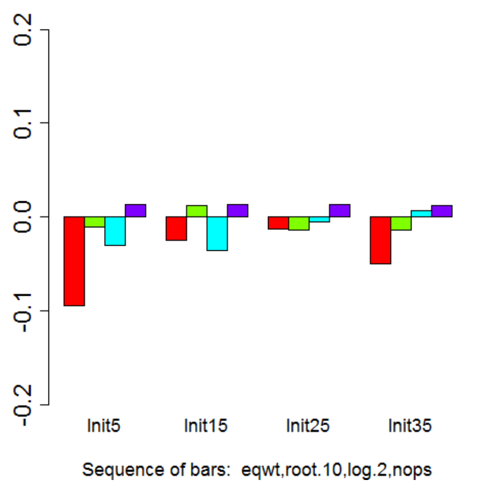
Figure 3.2: Optimal vs. winning candidates for the different weighting schemes

Optimal ("/") and winning ("\") candidates for the different weighting schemes. Initial consideration set size is 15; the winner is declared after six iterations.

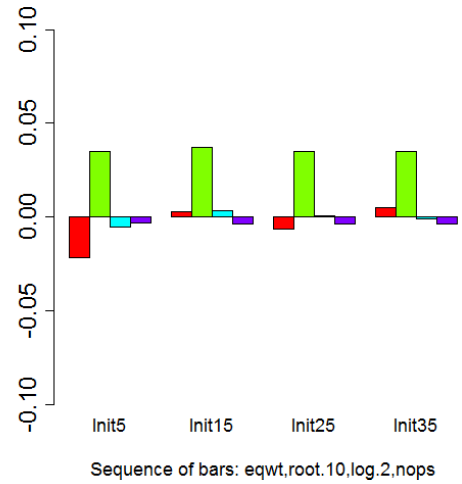
using true grade functions over the entire efficient frontier, we found its range to be (0.88, 0.98) – over all the weighting schemes. Hence, the optimality gap is about $0.013\%/(0.98 - 0.88) = 0.13\%$, which indicates the high quality of the mechanism outcome.

Figure 3.3b plots the signed Euclidean distances between the winning and optimal candidate, d_v. A negative sign was ascribed to d_v if the predictability metric of the winning candidate was less than the optimal candidate's (the winning candidate lay to the "left" of the optimal candidate in Figure 3.2).

We observe that the winning candidates obtained by the mechanism are quite close to the optimal ones. A larger size does not necessarily mean better solutions consistently – only Init5 seems to suffer in overall quality, but the others are quite



(a) Percentage deviation of the majority grade



(b) Signed Euclidean distance

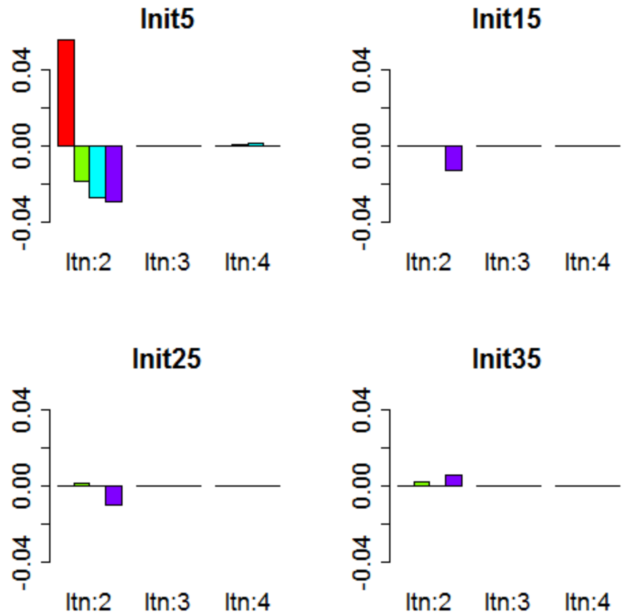
Figure 3.3: Evaluation results over several initial consideration set sizes and weighting schemes

similar. Recall this is after six rounds of grading.

Figure 3.4 reports on convergence over the rounds. It plots the signed distances for winning candidate in each round over the one in the previous round. We note that except for Init5, all the higher initial consideration set sizes practically converge at the end of the first round itself. However, it may still be beneficial to have at least two rounds.

These experiments were conducted on a personal laptop with Intel Celeron Dual-core CPU (1.8 GHz), having 2 GB RAM, running 32-bit Microsoft Windows 7 Home Premium operating system. Computing environment used was R version 13.0, with API Rcplex to interface with the CPLEX 12.0 solver, obtained through IBM Academic Initiative.

Figure 3.5 plots the computational times for running six rounds for the respective weighting scheme-initial consideration set size combination. log.2 scheme



Sequence of bars: eqwt,root.10,log.2,nops

Figure 3.4: Euclidean (signed) distance of winning candidates for each round over the previous round – for several initial consideration set sizes and weighting schemes.

eliminates the airlines with single operation, hence takes the smallest time. nops gives largest weight to the largest airline, hence takes lesser time than the root-transformed schemes. The computational times increase as the largest airline is apportioned smaller weight: root.10 takes longest, followed by root.20, then root.30, which takes about same time as nops. eqwt interestingly does not take the longest, which gives all airlines equal weight. Finally, an interesting observation is that higher initial consideration set sizes take lesser time to compute.

All the computations were run serially. As each airline's process is independent of other's, there is scope for parallelization. In effect, the computational times could be $\frac{1}{47}$-th of those reported. Moreover, for just two rounds, the computation time should further reduce by 67%.

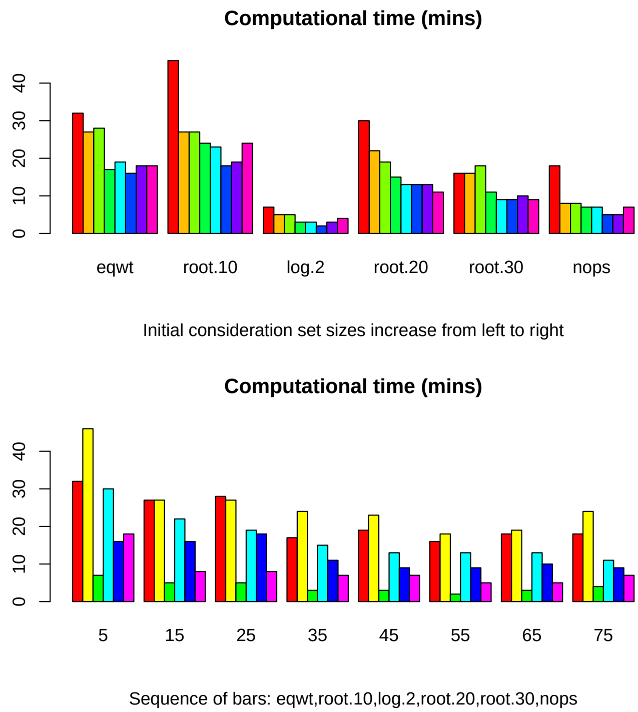


Figure 3.5: Computation time in minutes for several initial consideration set sizes and weighting schemes.

3.5 Conclusions

In this paper, we have described a mechanism for generating a consensus vector for use in strategic planning in air traffic flow management. Our approach is based on Majority Judgment but it employs a novel extension: the ability to handle very large sets of candidates. Our experimental results show the methods developed are very effective and can be efficiently carried out.

Several additional steps are required (and currently being carried out) to achieve practicality in the ATFM context. These include developing intuitive mechanisms for the flight operators to understand the performance vectors and to grade them, development of methods to generate the constraints defining the feasible vector space (μ) based on the current weather conditions and air traffic demand, estimation of benefits and human-in-the-loop experiments.

Of particular importance both to the ATFM application and more general applications is the issue of the potential for strategic grading/voting. Our experiments assumed that flight operators graded in a manner that was consistent with their true value functions. While Majority Judgment is generally (somewhat) immune to gaming, this issue deserves further analysis. For example, it could be the case that certain rules should be put in place to help insure reasonable behavior and outcomes, e.g. rules against collusion seem to be warranted.

Finally we note that we are quite excited about the potential application of this mechanism in other areas. There would seem to be a natural fit for many other application contexts.

Acknowledgment

This work was supported by the Federal Aviation Administration through the NEXTOR-II Consortium.

Chapter B: (Appendix to Chapter 3)

All appendices deal with individual players; subscript i for player is suppressed.

B.1 Grade Function Specification

Without loss of generality, the individual component metrics are normalized to have support in $[0, 1]$. Two components are used for explanation, but the specification easily extends to any number of components.

A quadratic form is specified for the value functions of individual components,

$$\nu_s(m_s) = a_s m_s^2 + b_s m_s,$$

without an intercept. To obtain the desired increasing function over the range of m_s, the values of a_s and b_s need to be constrained such that:

$$-1 \leq a_s < 0; 0 < -2a_s \leq b_s \leq 1 - a_s.$$

This yields:

$$-0.5 \leq \frac{a_s}{b_s} < 0.$$

Substituting ν's into the grade function, normalizing and renaming the coefficients gives:

$$g(\mathbf{m}) = k_1 m_1 + k_2 m_2 + k_3 m_1^2 + k_4 m_2^2 + k_5 m_1 m_2 + k_6 m_1^2 m_2 + k_7 m_1 m_2^2 + k_8 m_1^2 m_2^2,$$

with following constraints for concavity and the integration rule:

$$k_1 \geq 0; k_2 \geq 0; -0.5k_1 \leq k_3 < 0; -0.5k_2 \leq k_4 < 0; k_5 \geq 0.$$

The renaming yields:

$$\frac{k_3}{k_1} = \frac{a_1}{b_1}; \frac{k_4}{k_2} = \frac{a_2}{b_2},$$

and thus:

$$-0.5k_1 \leq k_3 < 0; -0.5k_2 \leq k_4 < 0; -0.5k_5 \leq k_6 < 0; -0.5k_5 \leq k_7 < 0; 0 \leq k_8 \leq 0.25k_5.$$

Note that the normalization involves V_i^{max}, which can be computed using the optimization model provided in **Subset_Opt**(b), as follows. Specify $b = \{i\}$, and replace the constraint $x_i = g_i(\mathbf{m})$ with $x_i = V_i(\mathbf{m})$ – that is, the (un-normalized) value function. This would yield the V_i^{max} for the player i at optimality.

Normalization would only be required if the grade function is specified from the value functions of the individual components of the candidate space. Instead, if the specification with k's is used directly, and the constraints as mentioned above are honored for all k's, then the resulting grade function would automatically have the support in $[0, G^{max}]$. However, it is not possible to recover the original constants a's and b's from the k's. Only global concavity remains to be ensured.

Proposition 2 *Any one of the following constraints is a sufficient condition for global concavity of the grade function as specified above:*

$$k_1 + 3k_3 m_1 \leq 0, k_2 + 3k_4 m_2 \leq 0.$$

Proof. The Hessian matrix of a function being negative definite in a given region is a necessary and sufficient condition for concavity of the function within it. The region of interest here is: $\mathbf{m} \in ((0,0),\ldots,(1,1)]$. Denote the Hessian matrix of the grade function as:

$$\mathbf{H_g} = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22}, \end{bmatrix}$$

where g_{st} is the partial derivative of the $g(\mathbf{m})$ with respect to m_s and m_t:

$$\begin{aligned}
g_{11} &= \frac{\partial}{\partial m_1}\left(\frac{\partial g(m_1,m_2)}{\partial m_1}\right) \\
&= \frac{\partial}{\partial m_1}\left(k_1 + 2k_3m_1 + k_5m_2 + 2k_6m_1m_2 + k_7m_2^2 + 2k_8m_1m_2^2\right) \\
&= 2k_3 + 2k_6m_2 + 2k_8m_2^2 \\
g_{22} &= \frac{\partial}{\partial m_2}\left(\frac{\partial g(m_1,m_2)}{\partial m_2}\right) \\
&= \frac{\partial}{\partial m_2}\left(k_2 + 2k_4m_2 + k_5m_1 + k_6m_1^2 + 2k_7m_1m_2 + 2k_8m_1^2m_2\right) \\
&= 2k_4 + 2k_7m_1 + 2k_8m_1^2 \\
g_{12} &= \frac{\partial}{\partial m_1}\left(\frac{\partial g(m_1,m_2)}{\partial m_2}\right) \\
&= \frac{\partial}{\partial m_1}\left(k_2 + 2k_4m_2 + k_5m_1 + k_6m_1^2 + 2k_7m_1m_2 + 2k_8m_1^2m_2\right) \\
&= k_5 + 2k_6m_1 + 2k_7m_2 + 4k_8m_1m_2
\end{aligned}$$

For $\mathbf{m} \neq \mathbf{0}$, non-negative r's and the above relationships for k's:

$$\begin{aligned}
\mathbf{m}^T \mathbf{H_g} \mathbf{m} &= \begin{bmatrix} m_1 & m_2 \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \\
&= m_1^2 g_{11} + 2m_1 m_2 g_{12} + m_2^2 g_{22} \\
&= 2\left[k_3 m_1^2 + k_6 m_1^2 m_2 + k_8 m_1^2 m_2^2\right] \\
&\quad + 2\left[k_5 m_1 m_2 + 2k_6 m_1^2 m_2 + 2k_7 m_1 m_2^2 + 4k_8 m_1^2 m_2^2\right] \\
&\quad + 2\left[k_4 m_2^2 + k_7 m_1 m_2^2 + k_8 m_1^2 m_2^2\right] \\
&= 2\left[k_3 m_1^2 + k_4 m_2^2\right] + 2\left[k_5 m_1 m_2 + 3k_6 m_1^2 m_2 + 3k_7 m_1 m_2^2 + 6k_8 m_1^2 m_2^2\right] \\
&= 2\left[k_3 m_1^2 + k_4 m_2^2\right] + 2k_5 m_1 m_2 \left[1 + 3\frac{k_3}{k_1} m_1 + 3\frac{k_4}{k_2} m_2 + 6\frac{k_3}{k_1}\frac{k_4}{k_2} m_1 m_2\right] \\
&= \underbrace{2\left[k_3 m_1^2 + k_4 m_2^2\right]}_{<0} + \underbrace{2k_5 m_1 m_2}_{\geq 0} \underbrace{\left[1 + 3\frac{k_3}{k_1} m_1 + 3\frac{k_4}{k_2} m_2 \left(1 + 2\frac{k_3}{k_1} m_1\right)\right]}_{?}
\end{aligned}$$

The first bracketed term is negative as $k_3, k_4 < 0$, and $\mathbf{m} \neq \mathbf{0}$ by hypothesis. Further, as $k_5 \geq 0$, if the final bracketed term is negative, the entire expression $\mathbf{m}^T \mathbf{H_g} \mathbf{m}$ would be negative, and the Hessian would be negative definite. However, it is not guaranteed to be so, as explained below.

Re-express the final bracketed term as: $\left[1 + 3\frac{k_3}{k_1} m_1 + 3\frac{k_4}{k_2} m_2 + 6\frac{k_3}{k_1}\frac{k_4}{k_2} m_1 m_2\right]$

$$= \underbrace{1 + 3\frac{k_3}{k_1} m_1}_{h_1^a} + \underbrace{3\frac{k_4}{k_2} m_2}_{h_2^a} \underbrace{\left(1 + 2\frac{k_3}{k_1} m_1\right)}_{h_3^a} = \underbrace{1 + 3\frac{k_4}{k_2} m_2}_{h_1^b} + \underbrace{3\frac{k_3}{k_1} m_1}_{h_2^b} \underbrace{\left(1 + 2\frac{k_4}{k_2} m_2\right)}_{h_3^b}$$

Recall that $\frac{k_3}{k_1} = \frac{a_1}{b_1}$, hence for $0 < m_1 \leq 1$:

$$-0.5 \leq \frac{k_3}{k_1} < 0 \Rightarrow -1 \leq 2\frac{k_3}{k_1} < 0 \Rightarrow -m_1 \leq 2\frac{k_3}{k_1} m_1` < 0$$

$$\Rightarrow 1 - m_1 \leq h_3^a < 1 \Rightarrow 0 \leq h_3^a < 1.$$

Similarly for $0 < m_2 \leq 1$: $-1.5 \leq h_2^a < 0$. Correspondingly, for $\mathbf{m} \neq \mathbf{0}$: $0 \leq h_3^b < 1; -1.5 \leq h_2^b < 0$. Thus, following hold true for $\mathbf{m} \neq \mathbf{0}$:

$$-1.5 \leq h_2^a h_3^a \leq 0;$$

$$-1.5 \leq h_2^b h_3^b \leq 0.$$

Consider the following two cases for h_1's.

Case 1 $h_1^a \leq 0$ or $h_1^b \leq 0$. This would directly imply that $\mathbf{m}^T \mathbf{H_g} \mathbf{m} \leq 0$, and is thus a sufficient condition for concavity of the grade function.

Case 2 $h_1^a > 0$ and $h_1^b > 0$. It follows then that:

$$1 + 3\frac{k_3}{k_1}m_1 > 0 \Rightarrow 1 > -3\frac{k_3}{k_1}m_1 \Rightarrow -\frac{1}{3} < \frac{k_3}{k_1}m_1 < 0,$$

and, similarly:

$$-\frac{1}{3} < \frac{k_4}{k_2}m_2 < 0.$$

A feasible range exists for h_2h_3's that allows the bracketed term to be positive.

There are other negative terms in the entire expression, which could result in $\mathbf{m}^T \mathbf{H_g} \mathbf{m} > 0$ even in these two Cases. This is why Case 1 conditions are also not necessary; however they do guarantee concavity. Either constraint in the proposition rules out Case 2. $\square$

B.2 Grade Function Estimation Procedure

Note from (B.1) that the grade function $g(\mathbf{m})$ is linear in the parameters k. Further, only five of the eight k's are independent. Treat the observed grade x as the dependent variable, and $m_1, m_2, m_1^2, m_2^2, m_1 m_2$ as five explanatory variables. The observational data over h candidates can be represented as: $X = Mk$, where $X_{(h\times 1)}$ is the vector of observations, $M_{(h\times 5)}$ is the matrix with the five columns computed as above from the graded candidates, and $k_{(5\times 1)}$ is the vector of the coefficients.

The sum of squared errors is: $e(k) = (X - Mk)^T(X - Mk) = X^T X - 2X^T Mk + k^T M^T Mk$. There are additional constraints to be observed on k's, as derived in Appendix B.1. A constrained least-squares procedure is specified as the following quadratic program:

$$\begin{aligned} \min \quad & -X^T Mk + \frac{1}{2} k^T M^T Mk \\ \text{s.t.} \quad & A^T k \geq k_0, \end{aligned}$$

where:

$$A^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -\frac{1}{2} & 0 & 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 1 & 0 \end{bmatrix}$$

and k_0 is vector of seven ϵ's (small positive constant), thus forcing strict inequalities as desired by the constraints.

B.3 Airlines' "True" Coefficients for Quadratic Grade Functions

We fix the coefficients for each airline's value function, following the intuition developed in Section 3.4.3.

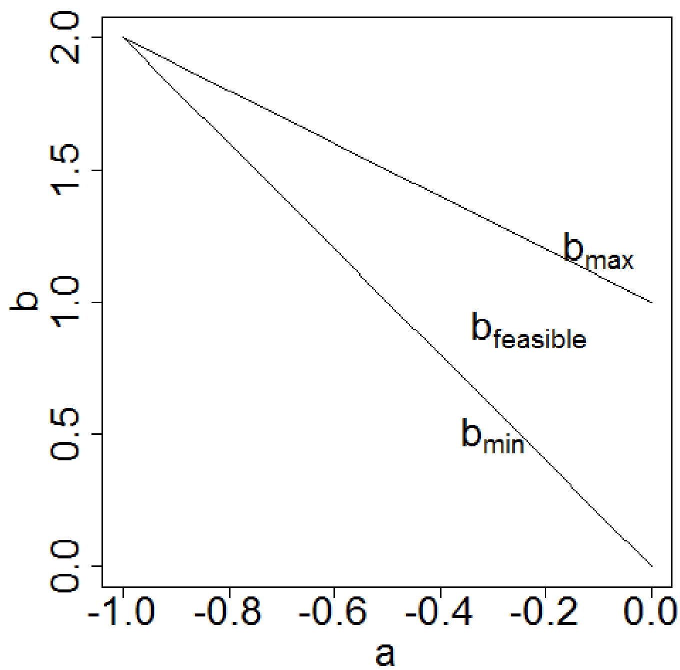


Figure B.1: Feasible values for a and b for value functions for individual metrics

B.3.1 Coefficients for Individual Value Functions

The quadratic value function is: $\nu(m) = am^2 + bm$, where $a \in [-1, 0)$, and $b \in [-2a, 1-a]$ are the coefficients to be fixed. At the higher values of a, that is, near zero, the shape of the value function is similar to a straight line with slope b. On the other hand, the lower end of a's range provides a more concave curvature. Fig B.1 shows the feasible values of b over the range of a. Note that b has a larger feasible range at higher values of a. The lower end of a's range allows a much smaller flexibility in choice of b; indeed, at $a = -1, b = 2$.

The highest possible value obtainable by the airline from a metric (i.e. at $m = 1$), is $a + b$. Hence, an airline with profile "HL" would have $a + b$ of the former metric higher than that of the latter. For an "H"-profile metric, high a would yield a straight-line like value function, while low a would yield a more concave one. On the other hand, since low $a+b$ allows only higher values for a, an "L"-profile metric will be straighter. Once $a+b$ is fixed, only one of the two coefficients has to be chosen, say a.

Three ranges within the support of $a+b$ are defined thus: $\{L : (0,^1/_2], S : [^1/_3,^2/_3], H : (^1/_2, 1]\}$. Following are repeated for each airline and metric. First, $a+b$ is drawn randomly from the designated ranges in accordance with the airline-metric profile. Next, a is drawn according to the relative number of operations of the airline, such that larger operations imply smaller a. We employ an acceptance sampling based approach for achieving this, described below and presented in Algorithm 2. This approach accounts for likely errors in our hypotheses, allowing some airlines to have different preference structures than what we posited. Finally, b is computed, and if not feasible, a is drawn again until a feasible b is found. We summarize this procedure in Algorithm 3.

The acceptance sampling algorithm for drawing values of a takes as input the vector of airline-wise operations A_{orig}, the index i_{orig} of the focal airline whose number of operations are reported as i_{orig}-th entry in A_{orig}, and $num.draws$ for number of draws to return for the focal airline. A_{orig} is sorted, and new position of the i_{orig}-th airline is identified – stored as A and i respectively. If there are multiple airlines with exactly same number of operations, any one of those could be designated as i, as the procedure treats similarly sized airlines in a similar fashion.

A proposal probability distribution from which random variables will be drawn is specified as uniform (0,1), such that each draw has mapping onto the desired coefficient a. In this case, $a = -v$. A proposed draw v for the ith airline will be accepted if it falls within its "valid range". If the ith airline has a unique value for number of operations, then its valid range is the width of the ith interval. If multiple airlines have the same number of operations, then the valid range extends to the width of these contiguous intervals. Thus, the ordering of airlines with same number of operations does not matter – which is desirable, as the sorting order for such airlines would be arbitrary.

We wish to allow some probability of accepting a v that happens to fall outside its valid range. Following scheme is adopted. Another iid random variable r is next drawn. v is accepted if r falls in the valid range. Thus, we accept v if either v or r fall within the valid range. Note that the valid range for r need not be the same as that of v; a different range could be used for fine-tuning the acceptance probabilities.

We show the simulation results for a hypothetical set of airline operations: $A = \{1, 1, 4, 4, 4, 7, 9, 10\}$. The first two airlines should predominantly have higher a, followed by the next three, and so on. The last airline should have predominantly lower values of a. We make 1000 draws and plot the histogram in Fig B.2. The results are clearly as desired.

B.3.2 Coefficients for Integration of Individual Value Functions

The integration rule states: $V = r_1\nu_1(m_1) + r_2\nu_2(m_2) + r_{12}\nu_1(m_1)\nu_2(m_2)$. Recall that r's are all non-negative by assumption. That is, the interaction between the two metrics cannot *decrease* the overall value to an airline. If the value derived

Algorithm 2 Acceptance sampling algorithm for drawing a values

sample.a($A_{orig}, i_{orig}, num.draws$)

$A \leftarrow sort(A_{orig}); i \leftarrow \min\{k|A[k] = A_{orig}[i_{orig}]\}$ {sort and identify new position of i_{orig}}

$n \leftarrow |A|; Acc \leftarrow \{\}$ {initialize}

{compute range for valid draws}

$j \leftarrow \min\{k|A[k] = A[i]\}; t_{min} \leftarrow \frac{j-1}{n}$

$j \leftarrow \max\{k|A[k] = A[i]\}; t_{max} \leftarrow \frac{j}{n}$

{make draws}

for $iter \in \{1, \ldots, num.draws\}$ **do**

  while true do

    $v \sim unif(0, 1)$ {draw (negative) value for a}

    $r \sim unif(0, 1)$ {draw whether to accept v or reject it}

    if $\Big(v \in \{t_{min}, \ldots, t_{max}\}\Big)$**or**$\Big((v \notin \{t_{min}, \ldots, t_{max}\})$**and**$(r \in \{t_{min}, \ldots, t_{max}\})\Big)$ **then**

      $Acc \leftarrow Acc \cup \{-v\}$; break {accept and break out of while loop}

    end if

  end while

end for

return Acc

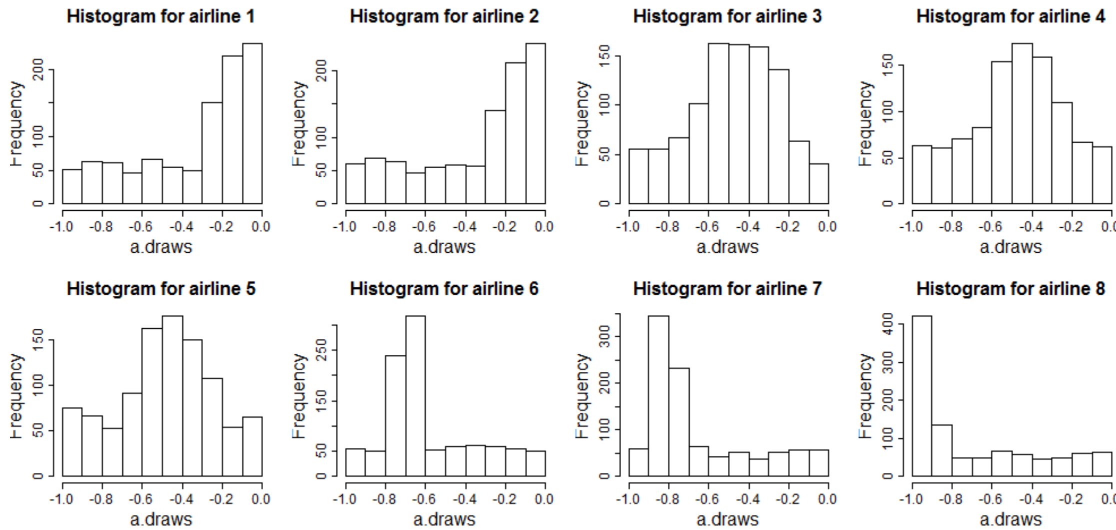


Figure B.2: Acceptance sampling results for hypothetical data of airline operations

from the two metrics are independent of each other to the airline, $r_{12} \rightarrow 0$.

r_1, r_2 have to be fixed with respect to the profile for the metric. As these will finally be normalized by V^{max} for each airline, the same positive range can be used for all the airlines without any loss in generality. The following ranges are used: $\{L : [2,4], S : [2,6], H : [4,6)\}$. The interaction effect is constrained to be smaller than the major effects, hence the range for r_{12} is taken as: $[0,2]$.

To ensure global concavity, the drawn values for a, b, r for each airline have to meet the necessary and sufficient condition over the support of (m_1, m_2), as shown in Appendix B.1:

$$\mathbf{m}^T\mathbf{H_g}\mathbf{m} = 2\left[k_3 m_1^2 + k_4 m_2^2\right] + 2k_5 m_1 m_2 \left[1 + 3\frac{k_3}{k_1}m_1 + 3\frac{k_4}{k_2}m_2 + 6\frac{k_3}{k_1}\frac{k_4}{k_2}m_1 m_2\right] < 0,$$

where,

$$k_1 = r_1 b_1 \frac{1}{V^{max}}; k_2 = r_2 b_2 \frac{1}{V^{max}}; k_3 = r_1 a_1 \frac{1}{V^{max}}; k_4 = r_2 a_2 \frac{1}{V^{max}}; k_5 = \frac{r_{12}}{b_1 b_2}\frac{1}{V^{max}}. \tag{B.1}$$

Since V^{max} is positive by assumption, it has no role in determining the curvature of the grade function. For ensuring concavity, we need to test that the necessary condition below is met at several sample points over the unit square of the metrics:

$$nec(m_1, m_2) =$$
$$\left[r_1 a_1 m_1^2 + r_2 a_2 m_2^2\right] + \frac{r_{12}}{b_1 b_2} m_1 m_2 \left[1 + 3\frac{a_1}{b_1} m_1 + 3\frac{a_2}{b_2} m_2 + 6\frac{a_1}{b_1}\frac{a_2}{b_2} m_1 m_2\right] < 0.$$

Treating $V^{max} = 1$, un-normalized k's are computed using (B.1). The LP corresponding to **Subset_Opt**(b) is solved (with individual airline as input) to determine the airline's grade-maximizing candidate. The associated optimal solution is V^{max} for the airline. This is then used to normalize the k coefficients using (B.1).

B.3.3 Overall Procedure

The overall algorithm for making the draws is now presented in Algorithm 3. The coefficients a, b and r thus drawn are shown in the left panel of Table B.1 – a_1, b_1 are the a, b coefficients for m_1, while a_2, b_2 are for m_2. The grade maximizing candidate and V^{max} for each airline are shown in the middle panel of Table B.1. Finally, the normalized k coefficients for each airline are in the right panel. Only $k_1, \cdots, k_5$ are shown, the other three can be directly computed using these five.

Algorithm 3 Algorithm for drawing a and b values

```
gen.true.abr()
for all airlines in A do
  repeat
    for all metrics do
      lookup profile P for the given metric and airline
      repeat
        if P = "H" then
          a.plus.b ~ unif(1/2, 1))
          r ~ unif(4, 6)
        else if P = "L" then
          a.plus.b ~ unif(0, 1/2))
          r ~ unif(2, 4)
        else if P = "S" then
          a.plus.b ~ unif(1/3, 2/3))
          r ~ unif(2, 6)
        end if
        a ←sample.a(A, i, 1)
        b ← a.plus.b − a
      until b ∈ {−2a, . . . , 1 − a}
    end for
    r_12 ~ unif(0, 2)
  until necessary condition for concavity met at each of several sample (m_1, m_2) points
  determine grade-maximizing candidate and associated V^max
  normalize coefficients using equations B.1
end for
```

Airline	a_1	b_1	a_2	b_2	r_1	r_2	r_{12}	m_1^{max}	m_2^{max}	V^{max}	k_1	k_2	k_3	k_4	k_5
United	-0.55	1.20	-0.06	0.44	5.84	2.79	0.02	0.9500	0.8168	4.67	1.50	0.26	-0.69	-0.03	0.00
American	-0.41	1.04	-0.27	0.70	5.78	3.12	0.30	0.9679	0.7930	4.90	1.23	0.44	-0.48	-0.17	0.04
Southwest	-0.11	0.49	-0.82	1.82	2.69	5.06	1.72	0.8639	0.8906	6.35	0.21	1.45	-0.05	-0.66	0.24
Northwest	-0.45	1.03	-0.25	0.66	3.30	3.10	0.46	0.9172	0.8503	3.15	1.08	0.65	-0.47	-0.24	0.10
Delta	0.00	0.36	-0.59	1.21	5.96	3.86	0.46	0.9829	0.7668	4.43	0.48	1.06	-0.01	-0.51	0.04
US Air	-0.13	0.48	-0.15	0.74	3.55	5.95	0.44	0.7451	0.9506	4.46	0.38	0.99	-0.11	-0.20	0.03
Continental	-0.30	0.88	-0.47	0.97	5.78	3.48	0.94	0.9829	0.7668	5.22	0.97	0.65	-0.33	-0.31	0.15
Airtran	-0.24	0.63	-0.28	1.26	3.37	4.02	1.37	0.7547	0.9468	5.39	0.40	0.94	-0.15	-0.21	0.20
Air Canada	-0.06	0.31	-0.82	1.81	2.62	5.02	1.84	0.8094	0.9219	5.79	0.14	1.57	-0.03	-0.71	0.18
ExpressJet	-0.30	0.77	-0.01	0.87	3.15	5.00	0.50	0.6708	0.9752	5.58	0.43	0.78	-0.17	-0.01	0.06
Jetblue	-0.03	0.11	-0.45	1.19	2.63	5.13	0.18	0.5767	0.9946	3.91	0.07	1.56	-0.02	-0.59	0.01
Chautauqua	-0.15	0.64	-0.71	1.71	2.37	5.42	0.70	0.8045	0.9244	6.55	0.23	1.42	-0.05	-0.59	0.12
Frontier	-0.22	0.59	-0.55	1.47	3.74	5.06	1.73	0.7896	0.9316	6.26	0.35	1.19	-0.13	-0.44	0.24
Mexicana	-0.07	0.29	-0.76	1.56	2.06	5.29	0.02	0.7466	0.9500	4.61	0.13	1.79	-0.03	-0.87	0.00
Lufthansa	-0.69	1.57	-0.17	0.55	5.36	3.00	0.32	0.9501	0.8167	5.80	1.45	0.29	-0.63	-0.09	0.05
Primaris	-0.67	1.58	-0.26	0.57	4.33	3.83	1.86	0.9580	0.8069	5.52	1.24	0.40	-0.53	-0.18	0.30
Alaska	-0.34	0.75	-0.62	1.55	3.22	5.27	1.00	0.7518	0.9480	6.34	0.38	1.29	-0.17	-0.51	0.18
Air Midwest	-0.30	0.98	-0.42	0.91	5.40	3.90	0.06	0.9829	0.7668	5.42	0.98	0.65	-0.30	-0.30	0.01
Aeromexico	-0.25	0.68	-0.61	1.49	2.11	5.93	1.70	0.7552	0.9466	6.44	0.22	1.37	-0.08	-0.56	0.27
British Airways	-0.56	1.19	-0.17	0.43	5.32	3.01	1.72	0.9501	0.8167	4.32	1.47	0.30	-0.69	-0.12	0.21
Polar Air Cargo	-0.58	1.20	-0.29	0.70	2.27	2.84	0.27	0.8688	0.8873	2.56	1.06	0.78	-0.51	-0.32	0.09
Spirit	-0.44	0.93	-0.49	1.00	3.77	5.42	1.23	0.9174	0.8501	4.79	0.73	1.13	-0.35	-0.56	0.24
Aer Lingus	-0.43	1.03	-0.43	1.06	5.27	2.25	0.57	0.9502	0.8166	4.61	1.18	0.52	-0.50	-0.21	0.14
Air Canada Jazz	-0.26	0.91	-0.28	0.76	2.96	3.33	0.59	0.9508	0.8159	3.50	0.77	0.73	-0.22	-0.26	0.12
Lot - Polish	-0.21	0.84	-0.12	0.54	4.08	4.44	0.22	0.9503	0.8164	4.17	0.82	0.58	-0.20	-0.13	0.02
SAS Scandinavian	-0.37	0.88	-0.33	0.72	2.26	2.75	0.73	0.9172	0.8502	2.28	0.87	0.87	-0.37	-0.40	0.20
Singapore	-0.49	1.06	-0.44	0.92	3.06	5.12	0.32	0.9134	0.8536	4.17	0.78	1.13	-0.36	-0.54	0.07
USA 3000	-0.45	1.35	-0.31	0.75	5.13	2.54	0.97	0.9900	0.7494	5.95	1.17	0.32	-0.39	-0.13	0.17
Air France	-0.21	0.80	-0.37	0.87	3.09	5.45	0.17	0.9184	0.8492	4.31	0.57	1.10	-0.15	-0.47	0.03
Air India	-0.23	0.47	-0.25	0.93	2.27	5.85	0.87	0.6411	0.9828	4.50	0.24	1.20	-0.11	-0.33	0.08
Air Jamaica	-0.35	0.86	-0.29	0.84	2.55	4.45	0.55	0.8391	0.9058	3.65	0.60	1.02	-0.25	-0.36	0.11
Alitalia	-0.37	0.78	-0.27	0.72	5.42	5.58	0.92	0.8500	0.8993	4.71	0.89	0.85	-0.42	-0.32	0.11
All Nippon	-0.01	0.57	-0.37	0.96	2.01	3.04	0.16	0.9332	0.8351	2.74	0.42	1.06	-0.01	-0.41	0.03
British Midland	-0.43	0.93	-0.01	0.62	5.77	4.93	0.76	0.7995	0.9268	5.71	0.94	0.53	-0.43	-0.01	0.08
Iberia	-0.31	0.93	-0.40	0.83	3.60	4.68	0.98	0.9530	0.8131	4.36	0.77	0.89	-0.26	-0.43	0.17
Japan Int'l	-0.35	0.98	-0.10	0.65	5.36	2.49	0.07	0.9530	0.8131	4.44	1.18	0.36	-0.42	-0.06	0.01
KLM-Royal Dutch	-0.30	0.91	-0.36	0.75	3.81	2.36	1.24	0.9879	0.7555	3.43	1.01	0.52	-0.33	-0.25	0.25
Korean	-0.49	1.12	-0.18	0.59	4.89	5.01	0.54	0.9035	0.8619	4.99	1.09	0.59	-0.48	-0.18	0.07
Martinair Holland	-0.22	0.82	-0.07	0.50	2.15	3.38	0.04	0.8837	0.8770	2.50	0.71	0.68	-0.19	-0.10	0.01
Pakistan Int'l	-0.42	0.88	-0.52	1.20	2.04	4.26	1.56	0.8144	0.9194	4.14	0.43	1.23	-0.21	-0.54	0.40
Swiss	-0.26	0.74	-0.12	0.54	5.61	4.80	0.58	0.9233	0.8447	4.44	0.93	0.58	-0.32	-0.14	0.05
Turkish	-0.29	0.92	-0.13	0.76	4.46	5.31	0.30	0.8540	0.8969	5.72	0.72	0.70	-0.23	-0.12	0.04
Virgin Atlantic	-0.21	0.80	-0.17	0.62	2.85	5.53	0.21	0.8738	0.8840	3.87	0.59	0.89	-0.16	-0.24	0.03
ABX	-0.37	0.80	-0.39	0.88	4.59	3.98	0.46	0.9084	0.8578	3.92	0.94	0.90	-0.43	-0.39	0.08
Cargoitalia	-0.16	0.65	-0.25	0.83	4.38	4.69	0.06	0.9174	0.8501	4.54	0.63	0.86	-0.15	-0.26	0.01
Custom Air	-0.12	0.67	-0.33	0.67	2.13	4.50	0.23	0.9630	0.8002	2.64	0.54	1.15	-0.10	-0.57	0.04
Kalitta	-0.37	0.98	-0.31	0.87	4.06	5.63	1.75	0.8886	0.8734	5.85	0.68	0.84	-0.26	-0.30	0.26

Table B.1: Table with the draws of a, b, r, grade-maximizing candidate and its value, and the normalized k coefficients

Chapter 4: Strategic Grading Opportunity in *COuNSEL* – A Consensus-Building Mechanism for Setting Service Level Expectations

The consensus-building mechanism described in the second essay has been accepted as a technically viable solution for the stated problem – although many practical challenges still remain before it may be deployed in operations. A key issue worthy of further investigation is its strong strategy-resistance – as claimed by the authors of Majority Judgment, the voting procedure embedded in *COuNSEL*. Using the broad ideas of Nash Equilibria, we characterize the necessary and sufficient conditions for an airline to benefit from unilaterally deviating from truthfully grading one or more candidates. The framework provides the airline with all the other airlines' grades on a set of candidates, and allows it an opportunity to present new grades. The analysis is repeated over multiple instances, and likelihood of beneficial strategic opportunity is presented.

4.1 Introduction

COuNSEL is a multi-objective multi-stakeholder consensus-building mechanism that has several desirable properties. It is based on Majority Judgment voting procedure, in which players provide a grade for each candidate in the consideration set, in a common language. The procedure uses the input of grades to compute a Majority Grade for each candidate; the candidate with the highest Majority Grade is deemed winner. Majority Judgment is described by its authors as being highly strategy-resistant (Balinski and Laraki 2011). We wish to verify this claim using simulations.

Our framework is as follows. Assume each player is provided an opportunity to unilaterally change her grade *after* observing everyone else's grades for a given consideration set of candidates. In practice, such opportunity would not exist – and the likelihood of hurting oneself would deter the players from strategic grading. Thus, this analysis provides the worst-case strategy proneness of the procedure.

The core idea behind this framework for analysis is similar to Nash Equilibria. It has origins in mechanism design, particularly in implementation theory (Maskin 1999). Gibbard and Satterthwaite's impossibility theorem established that true incentive-compatibility is not attainable if there is no restriction on the players' preference structure, unless a player is dictatorial. This realization led to investigation of weaker notions of strategy-proofness. Many solution concepts have been studied, e.g., Bayesian and sub-game perfect equilibria; however Nash equilibrium and Pareto optimality have been of particular interest. Such mechanisms are termed

Nash implementable. Maskin identified two properties that the social choice rule underlying a mechanism with three or more players must possess in order to be Nash implementable: monotonicity and no veto power. These results were tightened later, and extended to two players (Moore and Repullo 1990) – with potential applications in contracting theory, which invariably deals with two-party settings.

The Nash equilibrium solution concept assumes complete information and allows unrestricted domain of preferences – albeit observing convexity, continuity, and monotonicity. Maskin (1985) provides justifications for using such a complete information solution concept for an inherently incomplete information process like these social choice rule mechanisms. First, by definition, Nash equilibrium is a fixed-point among players' strategy spaces. Thus, it represents a stationary point in a process whereby players (with incomplete information) iteratively adjust their preference elicitations, until no unilateral deviation from true preferences benefits any player. Second, Nash equilibrium is a fitting solution concept in cases where the planner has incomplete information (or may not even exist), but the players are well-informed about each others preferences, such as in committee decisions.

Given that complete strategy-proofness is ruled out in any mechanism, it is of interest to quantify the extent of manipulability. This is particularly important in our case, as Majority Judgment is not a traditional voting procedure, and is therefore not as well-studied. Moreover, we intend to use weights for the players, and not the traditional "one person-one vote" setting. Of course, no single player will be given 50% or more of the total weight over all players to disallow dictatorial powers. However, this uneven distribution of decision power is worthy of investigation with regard to strategy-proneness. Finally, while unrestricted domain is of interest in itself, it would be useful to compare against a scenario where the players' preferences are convex.

Untruthful or strategic grading by a player may take several forms. She may increase the grade of one or more candidates, and / or decrease the grade of one or more candidates, possibly leaving grades on some candidates unchanged. Strategic grading is beneficial to a player only if the majority judgment winner is replaced by a candidate that she regards more preferable to it. Indeed, strategic grading can hurt the player if the new winner is less preferred by her than the existing winner. Or, it may not yield any change to the existing winner.

Some consideration sets may inherently be more manipulable than others – depending on the number of players, their grades, and number of candidates. Proportion of manipulable candidates to the total number of candidates is one measure of strategy-proneness. However, that does not imply that each such candidate can be manipulated by all the players. Some players may not have any candidate that they prefer over the current winner – these players will not have an incentive to deviate unilaterally. Among the remaining players, there may be some for whom there are no beneficial opportunities for the candidates that they prefer more than the current winner. These players too would not deviate unilaterally and benefit themselves. The proportion of the players that have any opportunity to benefit from strategic grading is a second measure of strategy-proneness. Another measure of strategy-proneness is the proportion of the total number of such beneficial player-candidate combinations.

Section 4.2 intuitively describes the procedure to identify strategic opportunity within this framework, using an illustrative example. Section 4.3 formalizes the description, and exhaustively identifies the necessary and sufficient conditions for beneficial strategic opportunities for a player. The measures for strategy-proneness, or manipulability, are also formally defined. Results from simulations for six types of scenario configurations are presented in Section 4.4. The first three allow the players unrestricted domain in grading; that is, no preference structure is imposed on the players. The latter three impose a convex grading function for each player. The three scenarios with these two assumptions on preference structures that were simulated are: players have equal weights, 5 players with differential weights, and 25 players with differential weights. The very first scenario, namely players have equal weights, and are allowed unrestricted domain in grading, is the basic Majority Judgment procedure. The last scenario, namely 25 differentiated players with a convex preference structure, is closer to the proposed *COuNSEL* procedure. The progression from the basic Majority Judgment to the last scenario is instructive. Section 4.5 concludes.

4.2 Illustration

Suppose five players (of equal weight) provide grades to three candidates as summarized in Table 4.1a. The grades are unrestricted, that is, no structure is imposed on the preferences. Of course, the grades should be within the allowable range – in this case in $[0 \dots 1]$. The grades are sorted for each candidate, and

Player	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$
1	0.6	0.3	0.2
2	0.1	0.3	**0.5**
3	0.1	0.6	0.6
4	**0.2**	0.7	0.3
5	0.8	**0.4**	0.7

(a) Grades provided by five players to three candidates

	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$
	0.1	0.3	0.2
	0.1	0.3	0.3
M.G.	0.2	0.4	0.5
	0.6	0.6	0.6
	0.8	0.7	0.7

(b) Grades in increasing order for each candidate

Player	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$
1	$[0.1 \ldots 0.2]$	$[0.4 \ldots 0.6]$	$[0.5 \ldots 0.6]$
2	$[0.2 \ldots 0.6]$	$[0.4 \ldots 0.6]$	$[0.3 \ldots 0.6]$
3	$[0.2 \ldots 0.6]$	$[0.3 \ldots 0.4]$	$[0.3 \ldots 0.5]$
4	$[0.1 \ldots 0.6]$	$[0.3 \ldots 0.4]$	$[0.5 \ldots 0.6]$
5	$[0.1 \ldots 0.2]$	$[0.3 \ldots 0.6]$	$[0.3 \ldots 0.5]$

(c) Each players' manipulable range for each candidate.

Table 4.1: Illustrative example.

presented in Table 4.1b. The majority grades are marked as "M.G.". The candidate $\mathbf{m}_3$ is the winner in this example.

We highlight several observations relevant to unilateral grading decisions. First, not all players have an incentive to deviate, as the consideration set does not have better candidate for them. In the example, players 2 and 3 are such players.

Second, to influence the majority grade of any candidate, a player has to grade *towards* its majority grade. In other words, if her grade for a particular candidate is higher (lower) than the current majority grade, then her new grade for it must be smaller (greater) than her current grade to have any chance to change the majority grade. This also implies that if her grade is higher (lower) than the current majority grade, then she can only decrease (increase) the new majority grade. If her grade is same as the majority grade for the candidate, then she can influence it upwards or downwards. Player 1 in this example clearly does not like the current winner, and would rather prefer $\mathbf{m}_1$ as the winner. However, her decreasing the grade on $\mathbf{m}_3$ will not change its majority grade – nor would increasing her grade on $\mathbf{m}_1$. The only way for her to change the new majority grade for $\mathbf{m}_1$ is to decrease her new grade on it, resulting in a lower majority grade; the opposite holds for $\mathbf{m}_3$.

The third observation relates to the extent of strategic grading opportunity available for a given candidate. A player can unilaterally influence the majority grade of a candidate within a specific range determined by the ordering of the grades provided by all the players. If player 1's new grade for $\mathbf{m}_3$ is below the current majority grade of 0.5, the majority grade remains at 0.5. Any grade between 0.5 and 0.6 would become the new majority grade, but any higher than 0.6 would not increase it beyond 0.6. Thus, player 1's "manipulable" range for $\mathbf{m}_3$ is [0.5,0.6]. Similarly for $\mathbf{m}_1$, a new grade by player 1 above the current majority grade of 0.2 will not have any impact. Any grade between 0.1 and 0.2 would become the new majority grade, any lower than 0.1 would keep it 0.1. Player 1's manipulable range for $\mathbf{m}_1$ is [0.1,0.2].

Clearly, player 1 has no opportunity to make her most preferred candidate $\mathbf{m}_1$ as the winner in this example. The fourth observation is regarding comparative grading over multiple candidates. Following the last two observations for $\mathbf{m}_2$, player 1 can only increase its majority grade, and that increase is bounded between 0.4 and 0.6. The range of grades between 0.5 and 0.6 overlaps with that of her manipulable range of $\mathbf{m}_3$, the current winner. Thus, player 1 can provide new grades for the two vectors $\mathbf{m}_2$ and $\mathbf{m}_3$ within [0.5,0.6] such that the grade for the former is less than that of the latter. This would make $\mathbf{m}_2$ the new winner, which she prefers over the current winner $\mathbf{m}_3$. The manipulable ranges for each candidate for the players who have an opportunity to benefit from strategic grading are reported in Table 4.1c.

Building on the previous observation, the fifth observation characterizes strategy-proneness of a given candidate for a player. A candidate is prone to (beneficial) strategy only if its manipulable range has an overlap with that of the current winner for any player. $\mathbf{m}_1$'s manipulable ranges for players 1 and 5 have no such overlap, similarly $\mathbf{m}_2$'s manipulable range for player 4 has no such overlap with those of the winner.

The sixth observation is about the relative position of a player's grade for a candidate with respect to its majority grade – in relation to those of the winner. When the player's grade is not same as majority grade for a candidate, its relation to the majority grade should be same as that for the winner. For player 1, the grade (0.2) for the winner $\mathbf{m}_3$ is below the majority grade (0.5). This is also true for $\mathbf{m}_2$: her grade (0.3) is below the majority grade (0.4) – but not for $\mathbf{m}_1$. The former is manipulable, but the latter is therefore not. The converse also holds, though there is no instance in this example. Such an opportunity also exists when a player provides the same grade as the majority grade for a candidate, and grades the winner lower than its majority grade. For example, player 4 grade for $\mathbf{m}_1$ is its majority grade, while she grades lower (0.3) than the majority grade for the winner (0.5). Another case is when a player grades the same as majority grade for the winner, and has a higher grade for a candidate than its majority grade. There is no instance in this example of this happening. These relationships are established formally in a later section.

Seventh, at an overall level, a candidate would not yield any benefit to any player if no player has an overlap of its manipulable range with that of the winner. In this example, all the candidates have an overlap with the winner's. Consider a candidate whose sorted grades are: $\{0.1, 0.15, 0.2, 0.25, 0.8\}$. Its manipulable range for any player has to be within [0.15,0.25], while the winner's has to be within [0.3,0.6]. Indeed, any candidate for which the grade just above the majority grade (the second highest grade in this example) is lower than the grade just below the winner's majority grade will not yield any benefit to any player. Each candidate in the consideration set should be pre-screened using this observation before analyzing at player-level.

Measures for Strategy-Proneness. Let us analyze the example with regard to strategy proneness. As just noted, all of the candidates (100%) in the consideration set are potentially manipulable. However, that does not mean that each player can unilaterally manipulate the grades to benefit.

We already identified that player 1 can benefit by manipulating $\mathbf{m}_2$ and / or $\mathbf{m}_3$. Also, we noted that the players 2 and 3 already have their most-preferred candidate in the current winner $\mathbf{m}_3$ – and hence do not have incentive to manipulate. Player 4 has an overlap between the manipulable ranges for $\mathbf{m}_1$ and $\mathbf{m}_3$ – but its preference for $\mathbf{m}_1$ being lesser, it has no incentive to manipulate these. There is no overlap for its most preferred candidate $\mathbf{m}_2$ with $\mathbf{m}_3$. Thus, player 4 actually has no opportunity to strategically grade that might benefit her. Similarly, player 5 has only an opportunity with $\mathbf{m}_2$, but since it prefers it less than the $\mathbf{m}_3$, it cannot benefit by manipulating her grades.

Thus, of the five players, only one – 20% – has a beneficial strategic opportunity. Among the 15 player-candidate opportunities, only two – about 13% – are beneficial to any player.

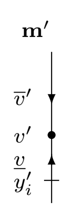
4.3 Conditions for Beneficial Strategic Grading

We formalize the observations regarding beneficial strategic grading opportunities for a player i with respect to a candidate $\mathbf{m}'$, whose majority grade is v'. For ease of exposition, the analysis and development begins with the equally weighted players case, that is, where all the players have the same weight. We relax this restriction later in the section, and explain the approach for the more general case of differentially weighted players.

4.3.1 Equally Weighted Players

Sorted in increasing order, the grade just before the majority grade is denoted $\underline{v}'$, and the grade just after the majority grade as $\overline{v}'$. Player i's grade for $\mathbf{m}'$ is denoted y_i'. Denote the winning candidate as $\mathbf{m}^*$, and the notation regarding it replaces the prime ($'$) with asterisk (*) in above.

A simple line diagram is used extensively in this section, it is explained below.



A candidate $\mathbf{m}'$ is depicted with a vertical bar, which represents the allowable grading range as per the common grading language. The majority grade v' is marked with a circle, and the two neighboring grades $\underline{v}'$ and $\overline{v}'$ are marked with upwards and downwards pointing arrowheads. Player i's grade for the candidate is marked with a horizontal tick marks.

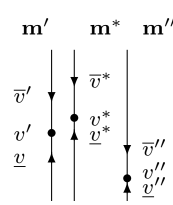
For any strategic grading by i for $\mathbf{m}'$ that changes its majority grade, the manipulable ranges are defined as below.

$$\delta_i' = \begin{cases} [v', \overline{v}'] & \text{if } y_i' < v' \\ [\underline{v}', v'] & \text{if } y_i' > v' \\ [\underline{v}', \overline{v}'] & \text{if } y_i' = v' \end{cases}$$

Looking at each candidate against the winner, the overlap of manipulable ranges between $\mathbf{m}'$ and $\mathbf{m}^*$ is a necessary condition:

$$\overline{v}' > \underline{v}^* \tag{4.1}$$

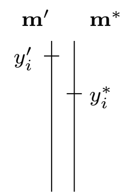
For instance, in the following, $\mathbf{m}'$ is potentially manipulable, but $\mathbf{m}''$ cannot be beneficially manipulated by any player.



The proportion of the candidates in the consideration set that meet the conditions of (4.1) gives an idea of strategy proneness of the setting at an overall level. A strategy-proof consideration set would have no candidate with such an overlap –

although it would be quite unlikely in practice. Indeed, as the seventh observation in Section 4.2 implied, this would be an overly strong measure, and an investigation of player-wise opportunities is required for a better and tighter quantification of strategy-proneness.

At a player level, a necessary condition for player i to strategically grade $\mathbf{m}'$ is that she grades it higher than she does the winner: $y'_i > y^*_i$.



This is not sufficient, as noted in the observations. Specific relationships among her grades for $\mathbf{m}'$ and $\mathbf{m}^*$ are required. We examine all possible relationships in Table 4.2, and summarize the necessary and sufficient conditions.

Combining cases 1 and 9, we see that among candidates that have: $(y'_i \leq v')$ & $(y^*_i < v^*)$, if there exists a candidate with $\overline{v}' \geq v^*$, then player i could increase its grade to anywhere in $(v^*, \overline{v}']$ without changing grades of the rest of the candidates. This is also a sufficient condition for a beneficial strategic grading opportunity for i, as she can only manipulate her grade for a single candidate and benefit herself. Of course, if multiple candidates meet the conditions, then she could only manipulate the candidate that she grades highest amongst these. Hence, one sufficient condition is:

$$(y'_i > y^*_i) \;\&\; (y^*_i < v^*) \;\&\; (y'_i \leq v') \;\&\; (\overline{v}' \geq v^*)$$

Cases 2 and 8 can be combined as: $(y'_i > v')$ & $(y^*_i \geq v^*)$. A candidate could

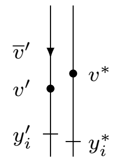
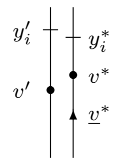
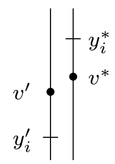
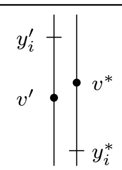
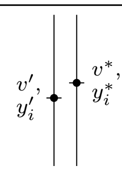
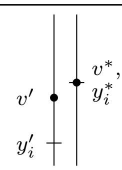
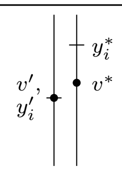
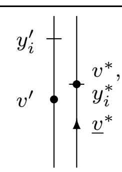
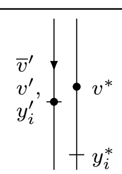
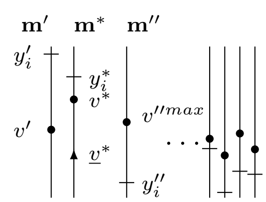
Case	Relative Positions	$\mathbf{m}'$ $\mathbf{m}^*$	Required Condition	Manipulable Range
1.	$(y_i' < v')$ & $(y_i^* < v^*)$.		$\overline{v}' \geq v^*$	$[v^*, \overline{v}']$
2.	$(y_i' > v')$ & $(y_i^* > v^*)$.		$v' \geq \underline{v}^*$	$[\underline{v}^*, v']$
3.	$(y_i' < v')$ & $(y_i^* > v^*)$.		NA	
4.	$(y_i' > v')$ & $(y_i^* < v^*)$.		NA	
5.	$(y_i' = v')$ & $(y_i^* = v^*)$.		NA	
6.	$(y_i' < v')$ & $(y_i^* = v^*)$.		NA	
7.	$(y_i' = v')$ & $(y_i^* > v^*)$.		NA	
8.	$(y_i' > v')$ & $(y_i^* = v^*)$.		$v' \geq \underline{v}^*$	$[\underline{v}^*, v']$
9.	$(y_i' = v')$ & $(y_i^* < v^*)$.		$\overline{v}' \geq v^*$	$[v^*, \overline{v}']$

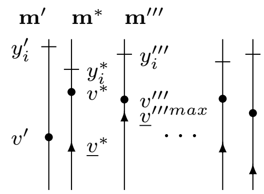
Table 4.2: Examination of relative positions between the majority grade and a player's grade for a non-winner candidate $\mathbf{m}'$ and the winner $\mathbf{m}^*$.

be potentially graded strategically to benefit i if $v' \geq \underline{v}^*$ is also met. However, this is not a sufficient condition. For, the required strategy is to down-grade the winner *as well as* any other candidates whose majority grade lies between v' and v^*, so that their majority grade becomes lower than v'. Such candidates may not be manipulable by the player i. Some more screening conditions need to be added in this case.

First, recall that any candidate with $y_i'' < v''$ cannot be manipulated by i so as to reduce its majority grade. Thus, the highest majority grade among such candidates, say v''^{max} forms a bound below which i cannot reduce the majority grade of the other candidates. For example, examine the following consideration set. Player i prefers $\mathbf{m}'$ the most. Candidate $\mathbf{m}^*$ is currently winning. Now, i can reduce its majority grade down to $\underline{v}^*$, but this will make $\mathbf{m}''$ as the new winner, not $\mathbf{m}'$. While $\mathbf{m}''$ is much to her dislike, she cannot influence its majority grade downwards.



Secondly, for a candidate with $y_i''' \geq v'''$, she could reduce its majority grade to $\underline{v}'''$. If she were to down-grade all of these candidates, the highest majority grade among these, say $\underline{v}'''^{max}$ would form a similar bound as above. Pictorially, examine the following consideration set. Player i likes the $\mathbf{m}'$ over the current winner $\mathbf{m}^*$. She could reduce the majority grade of the winner to lower than v', but she would also need to reduce the majority grade of $\mathbf{m}'''$ and other such candidates to make $\mathbf{m}'$ as the new winner. However, the lowest majority grade she can get for all such candidates is $\underline{v}'''^{max}$.



Finally, the two conditions are combined as follows. To decide whether $\mathbf{m}$ can be made the new winner by i, all the remaining candidates are evaluated. Depending on the relative position of her grade with respect to its majority grade, the candidate is marked as one of $\mathbf{m}''$ or $\mathbf{m}'''$. The bounds v''^{max} and $\underline{v}'''^{max}$ are determined, and the higher of these two is taken as $\underline{w}_i'$:

$$\underline{w}_i' = \max(v''^{max}, \underline{v}'''^{max}).$$

If $v' > \underline{w}_i$, then i can make $\mathbf{m}'$ as the winner, otherwise not. This will form the other sufficient condition for beneficial strategic grading:

$$(y_i' > y_i^*) \ \& \ (y_i^* \geq v^*) \ \& \ (y_i' > v') \ \& \ (v' > \underline{w}_i')$$

Putting it all together, the following is the necessary and sufficient condition that allows beneficial strategic grading opportunity to a player via a non-winner candidate $\mathbf{m}'$:

$$(y_i' > y_i^*) \ \& \left\{ \Big((y_i^* < v^*) \ \& \ (y_i' \leq v') \ \& \ (\overline{v}' \geq v^*) \Big) \Big| \Big((y_i^* \geq v^*) \ \& \ (y_i' > v') \ \& \ (v' > \underline{w}_i') \Big) \right\} \tag{4.2}$$

The first term in (4.2) states simply that the player has to prefer an alternate candidate over the winner. The first two terms of the two groups of conditions within the bracket state the relative positioning of the player's grade with respect to the majority grade for the alternate candidate and the winner. The two groups are mutually exclusive. Note that the beneficial opportunities are only likely if the player's grade is on the *same* side of the majority grade for *both* the candidates. Depending on which side of the majority grade the player's grades fall, specific conditions are required to be met for her to benefit – as stated in the final condition in the two groups of conditions.

In terms of exact strategies, if the player's grades are below the majority grades for both the candidate and the winner, she could simply raise her grade on the alternate candidate all the way to the maximum possible grade, G^{max} (though the majority grade of the candidate would remain at $\overline{v}'$ by her doing so), while keeping the grade on the winner at the same level. Of course, this is the simplest strategy for her; one can imagine several other strategies that would result beneficially to her. For instance, she could increase the grade of the winner too – while ensuring that her grade on the winner is smaller than the grade on the alternate candidate. Or, she could increase the grade of the alternate candidate barely above the winner's majority grade, or may be at any other level above it.

On the other hand, if her grades are both higher than the majority grades, her simplest strategy would be to keep the grade on the alternate candidate at the same level, and give all the other candidates the lowest possible grade. Unlike the previous case, it is necessary that she down-grades the other candidates as well –

as just down-grading the winner does not guarantee that the alternate candidate becomes the winner.

Measures for Strategy-Proneness. Suppose the consideration set comprises of M candidates, and there are N players. We formally define the three measures of interest.

1. Likelihood of manipulability of a candidate, φ^C: the number of non-winner candidates that meet condition (4.1) $\div$ $(M-1)$.
2. Likelihood of manipulability by a player, φ^P: the number of players for whom any candidate meets condition (4.2) $\div$ N.
3. Likelihood of manipulability of the consideration set, φ^S: the number of player-candidate pairs that meet condition (4.2) $\div$ $(M \times N)$.

4.3.2 Differentially-Weighted Players

In *COuNSEL*, the airlines are assigned different weights which are a function of the impact they suffer from the weather. The equally weighted case explained thus far needs four types of modifications to account for the players' weights.

First, the definitions of the Majority Grade v', and its neighbors $\underline{v}'$ and $\overline{v}'$ are modified. Instead of a simple median, a weighted median is sought for identifying v'.

Table 4.3a provides an example with six players, whose grades for a candidate and their weights are listed. The players are then sorted in the increasing order of their grades, as shown in Table 4.3b. In this ordered list, the cumulative weights are computed for each player. π is the proportion of each individual player's weight to the total weight (20 in this example). Π is the cumulative proportional weight in the increasing order of grades. The player whose cumulative weight meets or exceeds half the total weight (20/2=10 in this example) provides the majority grade v' – player B in this example. Coincidentally, if the players had equal weight, the majority grade would have been the same – but this need not be the case, as we shall see shortly. The grades just below and above v' are respectively marked $\underline{v}'$ and $\overline{v}'$, as earlier. The majoritarian set in this example is formed by players B, C, D, and F.

Recall that no player is assigned a weight that is larger than half the total weight, to avoid giving it dictatorial powers. This implies that when the players are ordered in increasing order of their grades for a given candidate $\mathbf{m}'$, the weighted majority grade v' is always flanked by at least one grade on either side. That is, with three or more players, $\underline{v}'$ and $\overline{v}'$ are always defined in the differentially weighted case – just like the equally weighted case.

Aside from this modification, the rest of the procedure for determining the winning candidate over a consideration set remains the same. That is, the weighted majority grade is computed for each candidate in the consideration set, and the candidate with the largest majority grade is declared the winner.

The second modification has to do with manipulability of a candidate $\mathbf{m}'$ by a player i with proportional weight π_i, whose grade for $\mathbf{m}'$ is y_i'. Like in the equally weighted case, to influence the majority grade of $\mathbf{m}'$, i has to provide a new grade *towards* v. However, the equally weighted case ensured that each player could

Players	Grades	Weights
A	0.15	6
B	0.24	5
C	0.96	4
D	0.33	3
E	0.18	1
F	0.63	1

(a) Example grades

Players	Ordered Grades	Weights	Cumulative Weights	π	Π	
A	0.15	6	6	0.30	0.30	
E	0.18	1	7	0.05	0.35	$\underline{v}'$
B	0.24	5	12	0.25	0.60	v'
D	0.33	3	15	0.15	0.75	$\overline{v}'$
F	0.63	1	16	0.05	0.80	
C	0.96	4	20	0.20	1.00	

(b) Players ordered by grades

Table 4.3: Weighted Majority Grade example

Ordered Grades	Players	Weights	Cumulative Weights	π	Π	
A	0.15	6	6	0.30	0.30	
E	0.18	1	7	0.05	0.35	
F	0.20	1	8	0.05	0.40	
B	0.24	5	13	0.25	0.65	v'
D	0.33	3	16	0.15	0.80	
C	0.96	4	20	0.20	1.00	

(a) Player F has provided a reduced grade

Ordered Grades	Players	Weights	Cumulative Weights	π	Π	
A	0.15	6	6	0.30	0.30	
E	0.18	1	7	0.05	0.35	
C	0.20	4	11	0.20	0.55	v'
B	0.24	5	16	0.25	0.80	
D	0.33	3	19	0.15	0.95	
F	0.63	1	20	0.05	1.00	

(b) Player C has provided a reduced grade

Table 4.4: Manipulation in Differentially-Weighted Case: Downwards Revision

influence v – by grading in this fashion, i could move from majoritarian set to the non-majoritarian set and vice-versa. This was possible due to the fact that in the equally weighted case, the majoritarian set is a minimal majority-forming set: if any player moved out, it no more forms the majority. The converse held true for the non-majoritarian set: if any player moved in, it would now have formed a majority.

As weights are "lumpy", this no more holds true for the differentially weighted case. For instance, consider player F in Table 4.3b. She is currently in the majoritarian set for the given candidate. Hence, she has to provide a grade below the majority grade of 0.24 to influence it downwards – she cannot increase it by increasing her grade, and any grade above 0.24 also would not change anything. Suppose she provides 0.20, Table 4.4a is the amended table. Note that the majority grade remains at 0.24, as player F's weight is insufficient to move the new cumulative proportional weight Π to 0.5 or above.

To formalize this observation, denote the cumulative proportional weight of the player that provided the majority grade v' for the candidate $\mathbf{m}'$ as Π'. For the player that provided $\underline{v}'$, it is denoted as $\underline{\Pi}'$; and for the player that provided $\overline{v}'$, it is denoted as $\overline{\Pi}'$. In Table 4.3b, $\underline{\Pi}' = 0.35$, $\Pi' = 0.60$, and $\overline{\Pi}' = 0.75$.

So, for a player i whose grade $y_i' > v'$, the only way to influence the majority grade would now be qualified by the additional condition that $\underline{\Pi}' + \pi_i \geq 0.5$. Player B in Table 4.3b could only get 0.35+0.05=0.40, which being less than 0.5, was not sufficient, as seen in the amended Table 4.4a. Player C, on the other hand, could manipulate its majority grade: 0.35+0.20=0.55 clearly crossed 0.5, as seen in Table 4.4b.

Ordered Grades	Players	Weights	Cumulative Weights	π	Π	
A	0.15	6	6	0.30	0.30	
B	0.24	5	11	0.25	0.55	v'
E	0.30	1	12	0.05	0.60	
D	0.33	3	15	0.15	0.75	
F	0.63	1	16	0.05	0.80	
C	0.96	4	20	0.20	1.00	

(a) Player E has provided an increased grade

Ordered Grades	Players	Weights	Cumulative Weights	π	Π	
E	0.18	1	1	0.05	0.05	
B	0.24	5	6	0.25	0.30	
A	0.30	6	12	0.30	0.60	v'
D	0.33	3	15	0.15	0.75	
F	0.63	1	16	0.05	0.80	
C	0.96	4	20	0.20	1.00	

(b) Player A has provided an increased grade

Table 4.5: Manipulation in Differentially-Weighted Case: Upwards Revision

Conversely, a player with $y_i' < v'$ can influence the majority grade upwards only if $\Pi' - \pi_i < 0.5$. Table 4.5a shows player E could not influence, as 0.60-0.05=0.55 exceeded 0.5; while player A could do so, because 0.60-0.30=0.30 was below 0.5.

Finally, for a player with $y_i' = v'$, manipulability is possible in either direction, so long as $\underline{v}' < v' < \overline{v}'$. Indeed, even if strict inequality does not hold, manipulation by such a player is possible due to differential weights – as we shall see next.

The third modification has to with the manipulable ranges. With differential weights, it is possible that a player can manipulate the majority grade beyond $\underline{v}'$ and $\overline{v}'$. For instance, see Table 4.6. In Table 4.6a, player C reduced her grade further, below that of E – who in the original Table 4.3b had provided $\underline{v}'$. This caused the new majority grade to become lower than the original $\underline{v}'$. Conversely,

Ordered Grades	Players	Weights	Cumulative Weights	π	Π	
A	0.15	6	6	0.30	0.30	
C	0.16	4	10	0.20	0.50	v'
E	0.18	1	11	0.05	0.55	
B	0.24	5	16	0.25	0.80	
D	0.33	3	19	0.15	0.95	
F	0.63	1	20	0.05	1.00	

(a) Player C has provided a further reduced grade

Ordered Grades	Players	Weights	Cumulative Weights	π	Π	
E	0.18	1	1	0.05	0.05	
B	0.24	5	6	0.25	0.30	
D	0.33	3	9	0.15	0.45	
A	0.60	6	15	0.30	0.75	v'
F	0.63	1	16	0.05	0.80	
C	0.96	4	20	0.20	1.00	

(b) Player A has provided a further increased grade

Table 4.6: Manipulation in Differentially-Weighted Case: Larger Revisions

player A in Table 4.6b effectively changed the majority grade above the original $\overline{v}'$. The manipulable ranges are thus not constrained to be within $\underline{v}'$ and $\overline{v}'$.

For player i with $y_i \geq v'$, the lower bound for the manipulable range is given by the grade of the player j with the smallest Π_j, where $\Pi_j + \pi_i \geq 0.5$. Denote this grade as $\underline{u}'_i$ – note that it depends on the particular player i under consideration. Conversely, for player i with $y_i \leq v'$, the upper bound for the manipulable range is given by the grade of the player j with the smallest Π_j, where $\Pi_j - \pi_i \geq 0.5$. Denote this grade as $\overline{u}'_i$.

For any strategic grading by i for $\mathbf{m}'$ that changes its majority grade, the manipulable ranges are defined as below.

$$\delta'_i = \begin{cases} [v', \overline{u}'_i] & \text{if } y'_i < v' \\ [\underline{u}'_i, v'] & \text{if } y'_i > v' \\ [\underline{u}'_i, \overline{u}'_i] & \text{if } y'_i = v' \end{cases}$$

The fourth and final modification updates the necessary and sufficient conditions for manipulability over multiple candidates in the candidate set. The core necessary condition that $y'_i > y^*_i$ remains – i must grade the alternate candidate higher than she grades the winning candidate. The observations made in the first two columns of Table 4.2 continue to hold – the only ways to benefit from strategic grading via a candidate $\mathbf{m}'$ require the player's grades for both the winner and $\mathbf{m}'$ to be on the same side of their respective majority grades. Specifically:

a. $(y^*_i < v^*)$ & $(y'_i \leq v')$, or

b. $(y^*_i \geq v^*)$ & $(y'_i > v')$.

However, the latter columns need an update, as described above.

Case (a) requires a simpler manipulation – grade for only $\mathbf{m}'$ needs to be increased to make it a winner. The sufficient condition in this case is that there is room for benefit:

$$(y'_i > y^*_i) \,\&\, (y^*_i < v^*) \,\&\, (y'_i \leq v') \,\&\, (\overline{u}' \geq v^*).$$

Case (b) requires a complex manipulation – grades for multiple candidates need to be decreased, to make them all losers against $\mathbf{m}'$. Using a similar approach as developed in the equally-weighted case, the rest of the consideration set is split into two categories: (i) with $y''_i < v''$ and (ii) $y'''_i \geq v'''$. Among category (i) candidates, the highest majority grade v''^{max} is the lower bound below which majority grade cannot be decreased by i. This is same as the equally-weighted case. Among category (ii) candidates, there is a modification: the highest $\underline{u}'''^{max}$ forms the lower bound. Thus, $\underline{\underline{w}}'_i$ needs to be updated as:

$$\underline{\underline{w}}'_i = \max(v''^{max}, \underline{u}'''^{max}).$$

Putting it all together, for the differentially-weighted case, the following is the necessary and sufficient condition that allows beneficial strategic grading opportunity to a player via a non-winner candidate $\mathbf{m}'$:

$$(y'_i > y^*_i) \; \& \; \left\{ \Big((y^*_i < v^*) \; \& \; (y'_i \leq v') \; \& \; (\overline{u}' \geq v^*) \Big) \; \Big| \; \Big((y^*_i \geq v^*) \; \& \; (y'_i > v') \; \& \; (v' > \underline{\underline{w}}'_i) \Big) \right\}. \tag{4.3}$$

Measures for Strategy-Proneness. Suppose the consideration set comprises of M candidates, and there are N players. We formally define the three measures of interest.

1. Likelihood of manipulability of a candidate, φ^C: the number of non-winner candidates that meet condition (4.1) $\div$ $(M - 1)$.

2. Likelihood of manipulability by a player, φ^P: the number of players for whom any candidate meets condition (4.3) $\div$ N.

3. Likelihood of manipulability of the consideration set, φ^S: the number of player-candidate pairs that meet condition (4.3) $\div$ $(M \times N)$.

Relative weights of the players	Preference structure: P1: None ("unrestricted domain")	Preference structure: P2: Convex function
R1: Equal weights ("unweighted")	P1R1	P2R1
R2: Differential weights (N=5)	P1R2	P2R2
R3: Differential weights (N=25)	P1R3	P2R3

Table 4.7: Design of experiments for investigation of strategy resistance

φ^C uses the same condition as the equally-weighted case, as it is at the overall consideration set level. φ^P and φ^S are now updated with the modified condition derived in this section.

4.4 Simulation Results

To get a sense of strategy resistance of the procedure, we conducted a number of simulations systematically varying some key parameters. The design of experiments is summarized in Table 4.7.

The intent behind this design has been to contrast the proposed *COuNSEL* procedure with several other plausible implementations. At the simplest extreme, P1R1 is the basic Majority Judgment, as laid out by its authors. At the other extreme lies P2R3, which is closest to the real-life scenarios that *COuNSEL* may be deployed for. The progression in the two directions from P1R1 to P2R3 is instructive. R2 and R3 address the proportional representation aspect of *COuNSEL*, which is a key design element that adds equitability. R2 is a very small setup, and might represent the initial deployment phase of *COuNSEL*, in which fewer airlines may participate. R3 is a more likely setup reflecting the later phases of deployment. P2, on the other hand, addresses the key assumption in structuring of the grade functions. An unrestricted domain would easily lead to inconsistent grading over rounds, which is highly undesirable.

With this broad overview, the specific details for each scenario are now explained. Consideration set sizes of 5, 10, or 15 candidates are simulated in all the scenarios. The players' grades for the consideration set are generated randomly within the grading range of $\{0 \ldots 1\}$ for the unrestricted domain scenarios (P1). An increasing quadratic function of a randomly generated number that restricts the function maxima to be within the grading range is used for convex preference scenarios (P2).

In the equal weight scenarios (R1), number of players is one of 5, 15, 25, 35, and 45. Five different weighting schemes are simulated in the differential weight scenarios. In the differential weight (N=5) scenarios (R2), the number of players is fixed at 5; while the differential weight (N=25) scenarios (R3) have 25 players.

Table 4.8 summarizes the weighting schemes for the R2 scenarios. The first scheme gives all players equal weight for comparison. The proportion of largest weight to the total weight is 0.20 in this case. The Herfindahl-Hirschman Index, or HHI, is reported as a measure of the "market concentration". HHI is computed as sum of the square of the market shares of all players, where market share of a player is the proportion of her weight to the total weight. From scheme 1 through 5, the HHI keeps increasing, as two players (namely A and B) are given progressively higher weights. Player A has the largest weight; its proportion to total weight never crosses 50%, as that would provide it dictatorial power.

Table 4.9 summarizes the weighting schemes for the R3 scenarios. These have

Weighting Scheme	Player Weights A	B	C	D	E	Total Weight	Largest to Total	HHI
1	1	1	1	1	1	5	0.20	0.20
2	2	1	1	1	1	6	0.33	0.22
3	2	2	1	1	1	7	0.29	0.23
4	3	2	1	1	1	8	0.38	0.25
5	3	3	1	1	1	9	0.33	0.26

Table 4.8: Weights for the different weighting schemes for R2 scenarios

Weighting Scheme	Player Weights A	B	C	D	E	F	G	H	I	J	K	L	M	...	Y	Total Weight	Largest to Total	HHI
1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	1	25	0.04	0.040
2	2	2	2	2	2	2	2	2	2	2	2	2	1	...	1	37	0.05	0.045
3	4	4	4	4	2	2	2	2	2	2	2	2	1	...	1	45	0.09	0.054
4	8	8	4	4	2	2	2	2	2	2	2	2	1	...	1	53	0.15	0.073
5	16	8	4	4	2	2	2	2	2	2	2	2	1	...	1	61	0.26	0.107

Table 4.9: Weights for the different weighting schemes for R3 scenarios

25 players each, marked A through Y. The first scheme gives all players equal weight for comparison. The thirteen players marked M through Y have the same weight of 1 for all the schemes. Weights for the initial twelve players are systematically varied, so that the HHI increases as we go down the list. Player A has the largest weight; its proportion to total weight is kept below 50%.

A hundred simulations runs were conducted for each of the four scenarios. The averages of the three metrics for strategy proneness are reported.

4.4.1 P1R1: Unrestricted domain, Equal weights

In this scenario, the numbers of candidates (M) and players (N) are systematically varied; each player having the same weight as others. This is very similar

Number of candidates, M	Number of players, N				
	5	15	25	35	45
5	69.25	43.50	32.00	25.75	22.00
10	69.44	30.44	25.56	16.89	14.67
15	64.21	28.07	22.29	22.29	13.57

(a) P1R1: Mean φ^C (%)

Number of candidates, M	Number of players, N				
	5	15	25	35	45
5	9.60	8.73	6.68	4.43	4.49
10	15.20	10.07	7.52	6.00	6.47
15	19.40	9.67	8.12	7.74	5.87

(b) P1R1: Mean φ^P (%)

Number of candidates, M	Number of players, N				
	5	15	25	35	45
5	2.40	2.01	1.48	1.02	1.00
10	2.46	1.25	0.90	0.65	0.76
15	2.16	0.81	0.66	0.62	0.46

(c) P1R1: Mean φ^S (%)

Table 4.10: Strategy-proneness Measures for P1R1

to the basic Majority Judgment procedure described by its authors. It thus forms a benchmark to which results from the other scenarios are compared.

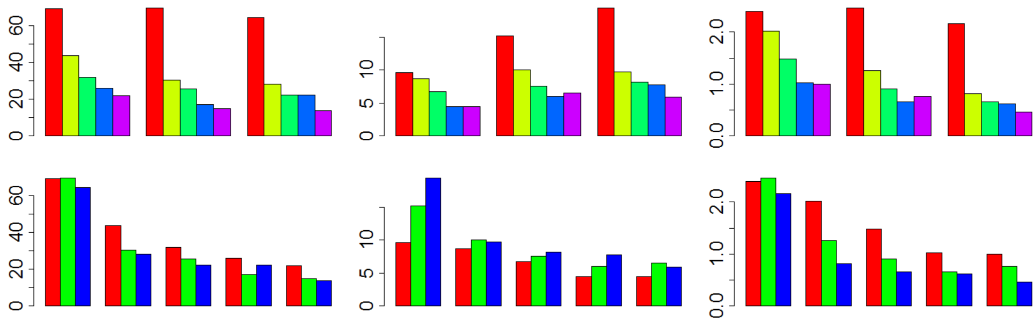
At a broad level, many candidates appear to be manipulable overall, as Table 4.10a reports. However, when it comes to individual players, a much smaller proportion of players actually have beneficial opportunities, as reported in Table 4.10b. More specifically, as each player evaluates each candidate, the likelihoods are even smaller, as reported in Table 4.10c. This is as expected.

Figure 4.1 pictorially depicts the information in the tables. The top row has groups of bars for consideration set sizes, M of 5, 10, 15 respectively. Within each group, the individual bars represent the measures for the different number of players, N of 5,15,25,35,45. The bottom row presents the same information, but groups them in the other way: the broad groups are for various N's, and the individual bars within each group have measures for different M. The patterns are clearly noticeable with this layout.

For a fixed size of consideration set, all the three measures decline as more players are included – as the top row depicts. This effect tapers off as the number of candidates increase. This makes intuitive sense – as more grades are provided by the increased number of players for each candidate, the gap between the $\underline{v}'$ and $\overline{v}'$ narrows. This gap is a decisive factor in manipulability of a candidate.

For a fixed number of players, φ^C and φ^S decrease as the consideration set size increases, while the opposite holds true for φ^P. This trend also tapers off with larger consideration set sizes. The decrease in the φ^C and φ^S has the same intuitive explanation as above. More candidates being available increases chance of a player to look forward to strategic grading – thus, φ^P increases with consideration set size. This has design implications – if *COuNSEL* has to be initiated in a region that has smaller number of airlines, then it should force them to grade more candidates in order to minimize strategic grading opportunities.

Equal weight to each airline is clearly ruled out in the implementation of *COuNSEL*. For, it would imply that airlines with large impact due to the weather would have the same voice in the decision-making as other airlines with perhaps a single impacted flight. However, this forms a benchmark for our investigation, as



(a) Mean φ^C (b) Mean φ^P (c) Mean φ^S

Figure 4.1: Strategy-proneness Measures for P1R1

Measures for proneness to beneficial strategic opportunity in P1R1 scenario. Within each of the three groups of bars in the top row, the consideration set size, M is fixed to one of 5, 10, 15. Across the five groups of bars in the bottom row, the number of candidates, N is fixed to be one of 5,15,2,53,45. Players have equal weight. Players' grades are unrestricted within the allowable range.

the proposed procedure should retain Majority Judgment's key desirable property of strategy resistance.

4.4.2 P1R2: Unrestricted domain, Differential weights with 5 players

In this scenario, the number of players is fixed at $N = 5$, and the consideration set size is one of $M = 5, 10, 15$. Players may have different weights; the five weighting schemes reported in Table 4.8 are simulated. This represents a likely scenario in the initial pilot phase of *COuNSEL*, in which a small number of airlines may be involved. A key difference from *COuNSEL* is the unrestricted domain, as *COuNSEL* assumes a structured grade function for each airline.

Tables and figures similar to those in the equal weights scenario are reported for the three measures. Very broadly, the strategy-proneness measures are not significantly different with differential weighting schemes as compared to the equal weighted scheme. A mild systematic pattern is evident from the top row of Figure

	Weighting Scheme				
Number of candidates, M	1	2	3	4	5
5	69.25	69.00	69.50	60.25	73.25
10	69.44	63.11	64.44	58.44	63.67
15	64.21	59.07	65.21	58.93	58.14

(a) P1R2: Mean φ^C (%)

	Weighting Scheme				
Number of candidates, M	1	2	3	4	5
5	9.60	7.80	14.40	7.20	11.80
10	15.20	13.40	13.40	11.60	17.20
15	19.40	16.40	18.80	15.20	16.40

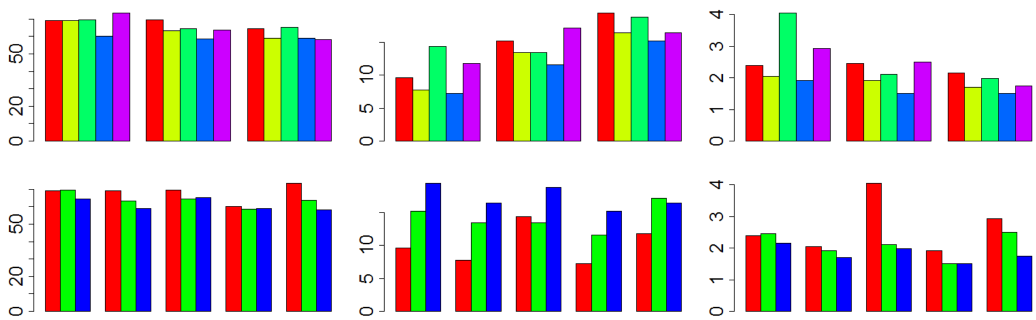
(b) P1R2: Mean φ^P (%)

	Weighting Scheme				
Number of candidates, M	1	2	3	4	5
5	2.40	2.04	4.04	1.92	2.92
10	2.46	1.92	2.10	1.50	2.50
15	2.16	1.69	1.97	1.51	1.73

(c) P1R2: Mean φ^S (%)

Table 4.11: Strategy-proneness Measures for P1R2

4.2. For a fixed size of consideration set, the weighting scheme appears to have smallest strategy-proneness; this scheme has the largest proportional weight to the player A. No clear patterns are visible with respect to HHI. The bottom row continues the pattern with the equal weights scenario. Given a weighting scheme, φ^C and φ^S decrease as the consideration set size increases, while φ^P increases. The intuition behind this remains the same.



(a) Mean φ^C (b) Mean φ^P (c) Mean φ^S

Figure 4.2: Strategy-proneness Measures for P1R2

Measures for proneness to beneficial strategic opportunity in P1R2 scenario. Number of candidates N is fixed at 5. The players have same weight in the first weighting scheme, and different weights in the others. Within each of the three groups of bars in the top row, the consideration set size, M is fixed to one of 5, 10, 15. Across the five groups of bars in the bottom row, the weighting scheme is varied so that HHI increases from left to right. Players' grades are unrestricted within the allowable range.

4.4.3 P1R3: Unrestricted domain, Differential weights with 25 players

In this scenario, the number of players is fixed at $N = 25$, and the consideration set size is one of $M = 5, 10, 15$. Players may have different weights; the five weighting schemes reported in Table 4.9 are simulated. This represents a later deployment phase of *COuNSEL*, whereby several airlines are involved in the decision-making process. Again, the unrestricted domain of preferences is a key difference from *COuNSEL*.

Tables and figures are reported for the three measures. A systematic pattern is evident from the top row of Figure 4.3: for a fixed consideration set size, increasing HHI (which also increases the proportional weight of the largest player in this scenario) tends to reduce strategy-proneness. The bottom row continues the pattern with the equal weights scenario. Given a weighting scheme, φ^C and φ^S decrease as

	Weighting Scheme				
Number of candidates, M	1	2	3	4	5
5	32.00	29.25	25.00	21.75	18.50
10	25.56	22.11	21.44	15.11	9.56
15	22.29	20.43	18.79	13.71	10.71

(a) P1R3: Mean φ^C (%)

	Weighting Scheme				
Number of candidates, M	1	2	3	4	5
5	6.68	5.12	4.56	4.76	6.20
10	7.52	7.04	10.04	7.48	7.32
15	8.12	9.16	10.36	9.32	11.04

(b) P1R3: Mean φ^P (%)

	Weighting Scheme				
Number of candidates, M	1	2	3	4	5
5	1.48	1.11	0.95	0.99	1.32
10	0.90	0.85	1.16	0.80	0.78
15	0.66	0.71	0.89	0.75	0.88

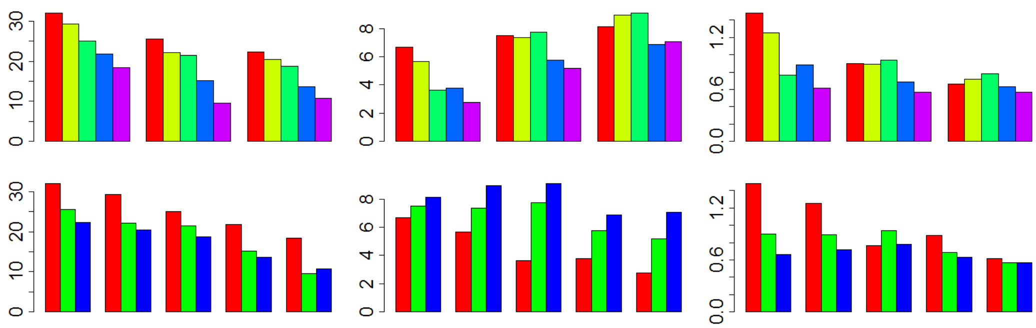
(c) P1R3: Mean φ^S (%)

Table 4.12: Strategy-proneness Measures for P1R3

the consideration set size increases, while φ^P increases. The intuition behind this remains the same.

4.4.4 P2R1: Convex preference structure, Equal weights

In this scenario, the numbers of candidates (M) and players (N) are systematically varied; each player having the same weight as others. The key difference from P1R1 is that the players have a convex grading function of a special type. The mechanics of drawing such convex grades are summarized in Appendix C.1.



(a) Mean φ^C (b) Mean φ^P (c) Mean φ^S

Figure 4.3: Strategy-proneness Measures for P1R3

Measures for proneness to beneficial strategic opportunity in P1R3 scenario. Number of candidates N is fixed at 25. The players have same weight in the first weighting scheme, and different weights in the others. Within each of the three groups of bars in the top row, the consideration set size, M is fixed to one of 5, 10, 15. Across the five groups of bars in the bottom row, the weighting scheme is varied so that HHI increases from left to right. Players' grades are unrestricted within the allowable range.

Compared to P1R1 scenario, the strategy-proneness measures are all dramatically lower. The convex structure forces the grades to be more concentrated near the peaks for each player. This potentially reduces the gap between $\underline{v}'$ and $\overline{v}'$ for all the candidates, leading to reduction in strategy proneness.

The general pattern of reductions in all the strategy-proneness measures within a fixed consideration set size continues, as the top row of Figure 4.4 shows. The tapering off effect is also evident in the top row. The bottom row has similar patterns as P1R1 for φ^C and φ^P – the former is more or less similar within each group having the same number of players, while the latter increases within each group. However, the φ^S measure increases as the consideration set size increases, with fixed number of players. Recall φ^P counts a player as potentially manipulative if she has opportunity via even a single candidate, whereas φ^S counts exact player-candidate pairs that are manipulable. Compared to P1R1, this implies that more candidates are manipulable for the players who have an opportunity to manipulate

Number of candidates, M	Number of players, N				
	5	15	25	35	45
5	30.00	16.75	9.25	7.50	7.25
10	30.67	13.22	7.89	6.00	6.11
15	36.43	15.29	10.14	5.36	4.93

(a) P2R1: Mean φ^C (%)

Number of candidates, M	Number of players, N				
	5	15	25	35	45
5	3.00	2.80	1.16	0.97	0.91
10	9.20	4.93	4.12	3.26	3.24
15	14.20	8.60	8.40	4.71	4.42

(b) P2R1: Mean φ^P (%)

Number of candidates, M	Number of players, N				
	5	15	25	35	45
5	0.64	0.65	0.27	0.19	0.18
10	1.04	0.61	0.48	0.38	0.32
15	1.55	0.68	0.68	0.37	0.31

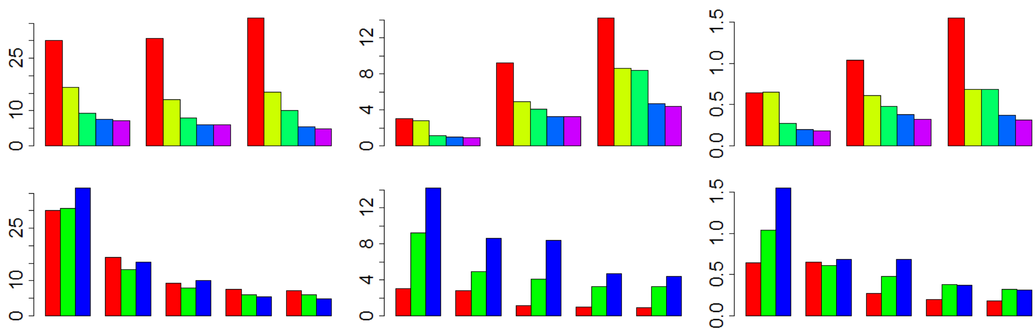
(c) P2R1: Mean φ^S (%)

Table 4.13: Strategy-proneness Measures for P2R1

at all, as the number of candidates increase. Note, however, that the overall levels of φ^P and φ^S are both lower than those in P1R1. All the three measures taper off as number of players increases.

4.4.5 P2R2: Convex preference structure, Differential weights with 5 players

In this scenario, the number of players is fixed at $N = 5$, and the consideration set size is one of $M = 5, 10, 15$. Players may have different weights; the five weighting



(a) Mean φ^C (b) Mean φ^P (c) Mean φ^S

Figure 4.4: Strategy-proneness Measures for P2R1

Measures for proneness to beneficial strategic opportunity in P2R1 scenario. Within each of the three groups of bars in the top row, the consideration set size, M is fixed to one of 5, 10, 15. Across the five groups of bars in the bottom row, the number of candidates, N is fixed to be one of 5,15,2,53,45. Players have equal weight. Players' grades are convex within the allowable range.

schemes reported in Table 4.8 are simulated. The key difference from P1R2 is that the players have a convex grading function of a special type – as specified for the P2R1 scenario.

As with P2R1 versus P1R1, there is a dramatic reduction in the strategy-proneness measures compared to its analogous unrestricted domain, namely P1R2. The main observation continues from P1R2: compared to the equal-weighted scenario with convex grading functions, the measures do not change dramatically due to introduction of weights – especially for φ^C. The patterns for the other two remain similar to those in the equal-weighted scenario as well – as the bottom row of Figure 4.5 shows. The top row shows no systematic patterns are discernible as the HHI changes for fixed consideration set sizes; however, like P1R2, the weighting scheme 4 has the smallest strategy-proneness measures.

Number of candidates, M	Weighting Scheme 1	2	3	4	5
5	30.00	32.25	33.75	26.50	30.50
10	30.67	34.00	33.22	33.89	34.33
15	36.43	33.07	32.21	34.29	33.07

(a) P2R2: Mean φ^C (%)

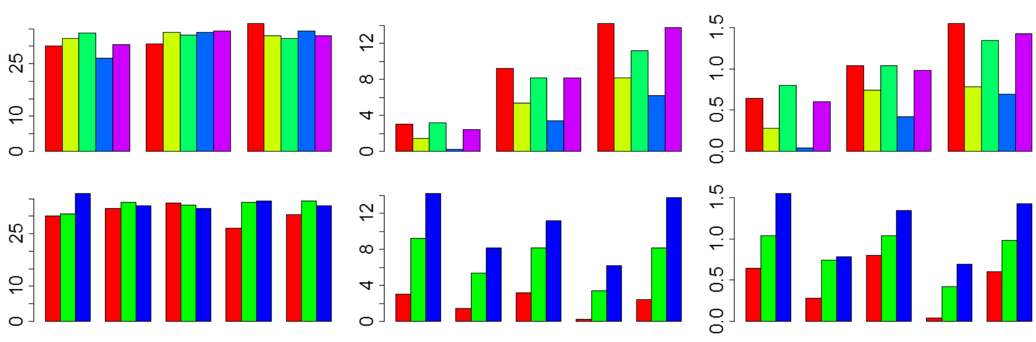
Number of candidates, M	Weighting Scheme 1	2	3	4	5
5	3.00	1.40	3.20	0.20	2.40
10	9.20	5.40	8.20	3.40	8.20
15	14.20	8.20	11.20	6.20	13.80

(b) P2R2: Mean φ^P (%)

Number of candidates, M	Weighting Scheme 1	2	3	4	5
5	0.64	0.28	0.80	0.04	0.60
10	1.04	0.74	1.04	0.42	0.98
15	1.55	0.79	1.35	0.69	1.43

(c) P2R2: Mean φ^S (%)

Table 4.14: Strategy-proneness Measures for P2R2



(a) Mean φ^C (b) Mean φ^P (c) Mean φ^S

Figure 4.5: Strategy-proneness Measures for P2R2

Measures for proneness to beneficial strategic opportunity in P2R2 scenario. Number of candidates N is fixed at 5. The players have same weight in the first weighting scheme, and different weights in the others. Within each of the three groups of bars in the top row, the consideration set size, M is fixed to one of 5, 10, 15. Across the five groups of bars in the bottom row, the weighting scheme is varied so that HHI increases from left to right. Players' grades are convex within the allowable range.

4.4.6 P2R3: Convex preference structure, Differential weights with 25 players

In this scenario, the number of players is fixed at $N = 25$, and the consideration set size is one of $M = 5, 10, 15$. Players may have different weights; the five weighting schemes reported in Table 4.9 are simulated. This scenario closely resembles the likely final deployment phase of *COuNSEL*.

The key difference from P1R3 is that the players have a convex grading function of a special type – as specified for the P2R1 scenario. All the strategy-proneness measures are significantly lower as compared to the unrestricted domain case of P1R3. The top row shows no systematic patterns are discernible as the HHI changes for the smaller consideration set sizes of 5 and 10. However, a decline in the measures is apparent with increase in HHI for consideration set comprising of 15 candidates. The main observation continues from P1R2: compared to the equal-weighted scenario with convex grading functions, the measures do not change dramatically due to introduction of weights – especially for φ^C. Unlike P1R3, where φ^C decreased with increase in HHI, φ^C does not seem to have any pattern. The patterns for the other two remain similar to those in the P2R1 as well as P2R2 – as the bottom row of Figure 4.5 shows. That is, for each weighting scheme, φ^C and φ^S generally increase with increase in consideration set size.

These have implications on the implementation design parameters for the mechanism. As far as possible, consideration set sizes should be kept small, not only for increased cognitive load to the players, but also for strategy-proneness. Ad-

Number of candidates, M	Weighting Scheme 1	2	3	4	5
5	9.25	9.25	8.00	9.25	8.25
10	7.89	8.89	10.89	8.33	7.78
15	10.14	7.93	8.07	8.43	7.43

(a) P2R3: Mean φ^C (%)

Number of candidates, M	Weighting Scheme 1	2	3	4	5
5	1.16	2.00	1.36	0.80	1.56
10	4.12	5.24	4.64	2.48	2.72
15	8.40	7.28	5.44	5.88	4.24

(b) P2R3: Mean φ^P (%)

Number of candidates, M	Weighting Scheme 1	2	3	4	5
5	0.27	0.41	0.27	0.17	0.33
10	0.48	0.59	0.53	0.32	0.31
15	0.68	0.58	0.41	0.48	0.34

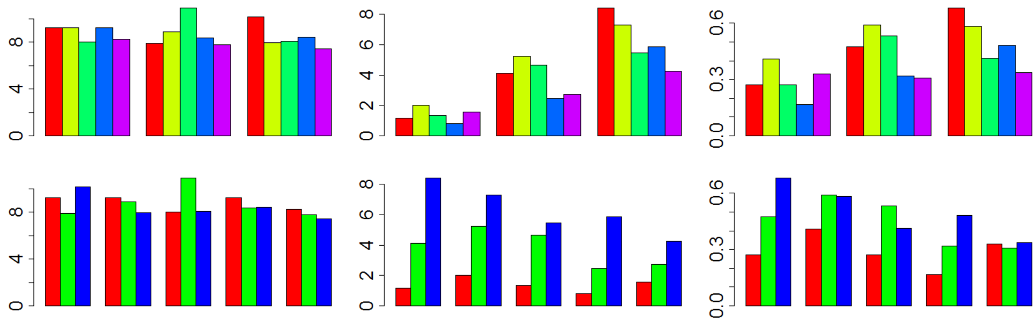
(c) P2R3: Mean φ^S (%)

Table 4.15: Strategy-proneness Measures for P2R3

dition of weights not significantly impacting the strategy-proneness measures is a useful observation in itself. However, these should be investigated for different types of players – as it must be giving larger strategic opportunities to the larger players, while eliminating such opportunities for the smaller players.

4.5 Conclusion

Impossibility results due to Arrow, Gibbard and Satterthwaite, have ruled out existence of strategy-proof mechanisms in which no player has dictatorial pow-



(a) Mean φ^C (b) Mean φ^P (c) Mean φ^S

Figure 4.6: Strategy-proneness Measures for P2R3

Measures for proneness to beneficial strategic opportunity in P2R3 scenario. Number of candidates N is fixed at 25. The players have same weight in the first weighting scheme, and different weights in the others. Within each of the three groups of bars in the top row, the consideration set size, M is fixed to one of 5, 10, 15. Across the five groups of bars in the bottom row, the weighting scheme is varied so that HHI increases from left to right. Players' grades are convex within the allowable range.

ers – especially with unrestricted domain. Majority Judgment is a recent proposal that bypasses this result, and is claimed to be highly strategy resistant by its authors. In this paper, we characterized and quantified the proneness of Majority Judgment-based voting procedure to beneficial strategic opportunities by the players. We employed a framework similar to Nash equilibrium concept, which has been extensively used in mechanism design literature as a solution concept.

Specific to Majority Judgment in general, we developed the necessary and sufficient conditions for a player to benefit by reporting untruthful grades for one or more candidates. The conditions were then used as basis for quantifying three measures of strategy proneness.

Finally, we simulated several scenarios starting from basic Majority Judgment procedure, systematically varying assumptions and key parameters, leading up to scenarios that closely resemble initial and later deployment phases of *COuNSEL*. We found that the most obvious measure for strategy proneness, the one based on proportion of manipulable candidates, is deceptive – it consistently reports very high likelihood of manipulation, typically upwards of 50%. However, the likelihood of an individual player to find a beneficial strategic opportunity drops in the regions of 10% or less. Moreover, as the specific candidates via which the individual players may benefit are also brought into consideration, the likelihood drops to 1-2% levels.

A surprising, though useful, observation has been the rather insignificant impact of attaching weights to the players. Weights are a significant design element in *COuNSEL*, wherein unlike the democratic "one-person one-vote" scenario, it is essential to provide the airlines differential weight in the overall decision-making, for equity reasons.

Another key observation has been the drastic reduction in strategy proneness when the unrestricted domain of grades is replaced with a convex preference structure. Convexity, continuity, and monotonicity have been standard extensions in the literature. These are also reasonable in our application area, whereby players would more likely have a possibly "single-peaked" preference structure over the feasible candidate space.

The results in themselves are quite encouraging. Even with complete knowledge of everyone's grades, and then being provided with an opportunity to benefit oneself, the likelihood of a particular player to find a beneficial opportunity via a candidate is in the region of 2% or below. In real-life, such opportunity would of course not exist. Moreover, untruthful reporting has a good possibility of hurting the player, as it may result with a new winner that is less preferred than the current winner.

These observations are based on simulations with simple preference structure, whereas *COuNSEL* design allows for a more nuanced structure over a multi-dimensional candidate space. Furthermore, it deals with feasibility constraints on the candidate space. Experiments incorporating these details, and with realistic application scenarios should be conducted before finalizing the design parameters of *COuNSEL*.

It should be mentioned here that the simulations assumed that a player had complete knowledge of other players' grades, and then had an opportunity to unilaterally deviate from truthful grading if it led to a more preferable candidate than the current winner. In practice, this will not be the case. There are three implications and possible directions for future research. One pertains to the information dissemination at the end of each round. The FAA could possibly release all the grading information, but that could incentivize airlines to collude among themselves – which would defeat the purpose of the entire mechanism. It could also lead to an information overload. On the other hand, the FAA need not release any information until the final round, but that may call into question the FAA's trust-worthiness. A middle ground that encourages the airlines to productively contribute to the process without divulging unnecessary information needs to be found.

The second implication has to do with the possible strategic uses of the partial knowledge that does get disseminated at the end of each round. As the airlines gain experience, they may be able to anticipate other airlines' behavior probabilistically, and use the information to update their beliefs. Instead of Nash equilibrium, a Bayesian Nash equilibrium may then serve as a more appropriate solution concept.

The modeling details would depend on the type of information released.

Finally, with the probabilistic knowledge of other airlines' grade functions replacing the full knowledge as in this paper, it would be imperative to quantify the expected loss due to strategic grading. We have identified the best case scenarios for an airline to benefit from strategic grading; this investigation would form the worst case for an airline.

Acknowledgment

This work was supported by the Federal Aviation Administration through the NEXTOR-II Consortium.

Chapter C: (Appendix to Chapter 4)

C.1 Convex Preference Structure

The procedure for drawing grades so that they follow a convex structure is detailed in this section.

Suppose the candidates are drawn randomly from a fixed range: $x \sim [0 \ldots 1]$. For a given candidate x, a special quadratic function maps these values into the grade for each player i: $y_i = a_i x^2 + b_i$, where a_i and b_i are player i-specific coefficients. The coefficients for each player are constrained such that: (a) the grade function is convex in the allowable grading range of $[0 \ldots 1]$, (b) the grade function is non-negative in the allowable range, (c) the grade function has its global maxima within the allowable range, and (d) the grade function has its maxima as the largest allowable grade of 1.

(a) and (c) are inter-related for quadratic functions. For it to have a global maximum, following necessary and sufficient conditions must be met (dropping subscript i for ease of notation):

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 0 \Rightarrow 2ax^* + b = 0 \Rightarrow x^* = \frac{-b}{2a},$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} < 0 \Rightarrow 2ax^* < 0 \Rightarrow a < 0.$$

For the maxima to be within the given range as required in (c), we want:

$$0 \leq x^* \leq 1 \Rightarrow 0 \leq \frac{-b}{2a} \leq 1 \Rightarrow 0 \leq b \leq -2a.$$

For the last inequality, recall $-2a < 0$ as required in the previous statement.

Further, recall that the specified function has $y = 0$ at $x = 0$. To satisfy (b), we need to ensure that $y \geq 0$ at the largest allowable value of x – which is 1 in this case. Thus, we get another bounding constraint for b:

$$0 \leq y|_{x=1} \leq 1 \Rightarrow 0 \leq a + b \leq 1 \Rightarrow -a \leq b \leq 1 - a.$$

Putting the two bounding constraints for b, we get:

$$0 \leq -a \leq b \leq -2a \leq 1 - a.$$

The tighter of the bounds require that:

$$-a \leq b \leq -2a.$$

For (d), we evaluate y at the maxima, and set it to the largest allowable grade, that is, 1:

$$y|_{x=x^*} = 1 \Rightarrow \frac{-b}{2a}\left[a(\frac{-b}{2a}) + b\right] = 1 \Rightarrow b = 2\sqrt{-a}.$$

Thus the bounds derived above imply:

$$-a \leq 2\sqrt{-a} \leq -2a \Rightarrow -1 \leq a \leq -4.$$

Some sample grade functions are shown in Figure C.1.

The procedure for generating the coefficients for each player is summarized as follows. For each player i, draw a coefficient: $a_i \sim [-4 \cdots -1]$, and compute $b_i = 2\sqrt{-a}$.

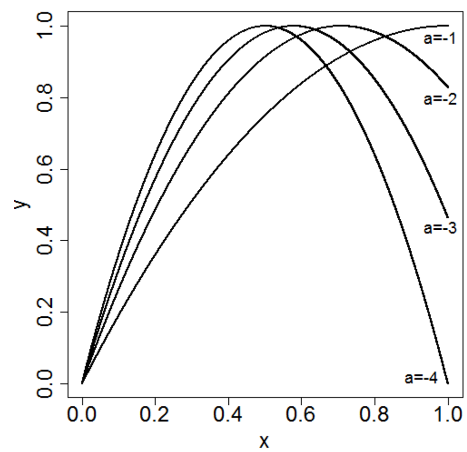


Figure C.1: Sample convex grade functions

Generating the grades for a given consideration set of M candidates is straightforward. Player i's grade for a candidate x is computed as: $y_i = a_i x^2 + b_i x$. For all the M candidates the same coefficients are used for a given player. The procedure is repeated for all N players.

It should be mentioned here that *COuNSEL* employs a more nuanced grading function, and allows for arbitrary number of dimensions for the candidates. It also has the additional complexity of the feasible candidate space. Nonetheless, this simple model allows us to study the strategy proneness with structured preferences, and contrast the results with no structure.

Bibliography

Abrahams, M. 1983. A service quality model of air travel demand: an empirical study. *Transportation Research Part A: General* **17**(5) 385–393.

Adler, T., C. S. Falzarano, G. Spitz. 2005. Modeling service trade-offs in air itinerary choices. *Transportation Research Record: Journal of the Transportation Research Board* **1915** 20–26.

Arrow, K. 1951. *Individual values and social choice*. New York: Wiley.

Ashley, K., I. Savage. 2010. Pricing congestion for arriving flights at Chicago O'Hare Airport. *Journal of Air Transport Management* **16**(1) 36–41.

Balinski, M., R. Laraki. 2011. *Majority Judgment: Measuring, Ranking, and Electing*. The MIT Press.

Ball, M., G. Donohue, K. Hoffman. 2006. Auctions for the safe, efficient, and equitable allocation of airspace system resources. *Combinatorial auctions*. Cramton, P., Y. Shoham and R. Steinberg, eds., MIT Press, 507—538.

Ball, M. O., L. M. Ausubel, F. Berardino, P. Cramton, G. Donohue, M. Hansen, K. Hoffman. 2007. Market-based alternatives for managing congestion at New Yorks LaGuardia Airport. *Proc. of AirNeth Annual Conference*.

Ball, M. O., C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, B. Zou. 2010. Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States. *NEXTOR Report* URL `http://www.nextor.org/pubs/TDI_Report_Final_11_03_10.pdf`.

Ball, M. O., C. Barnhart, A. Evans, M. Hansen, Y. Liu, P. Swaroop, V. Vaze. 2011. Distributed Mechanisms for Determining NAS-Wide Service Level Expectations. *NEXTOR Technical Report, July 2011.* **Year 1 Report**. URL `http://www.nextor.org/rep2011.html`.

Basso, L.J., A. Zhang. 2010. Pricing vs. slot policies when airport profits matter. *Transportation Research Part B: Methodological* **44**(3) 381–391.

Berardino, F. 2009. New US airport slot policy in flux. *Journal of Transport Economics and Policy (JTEP)* **43**(2) 279–290.

Brueckner, J. K. 2002. Airport congestion when carriers have market power. *American Economic Review* **92**(5) 1357–1375.

Brueckner, J. K. 2009a. Airport congestion management: Prices or quantities? *Access* **35**(Fall 2009) 10–15.

Brueckner, J. K. 2009b. Price vs. quantity-based approaches to airport congestion management. *Journal of Public Economics* **93**(5-6) 681–690.

Brueckner, J. K., R. Flores-Fillol. 2007. Airline schedule competition. *Review of Industrial Organization* **30**(3) 161–177.

BTS. 2007. Airline On-Time statistics and delay causes. Accessed Sep 01, 2011 by selecting

Aug 2007 as "Period From" and "Period To". URL http://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp.

BTS. 2013. U.S. Air Carrier Traffic Statistics. *Bureau of Transportation Statistics* URL http://apps.bts.gov/xml/air_traffic/src/index.xml.

Carlin, A., R. E. Park. 1970. Marginal cost pricing of airport runway capacity. *The American Economic Review* **60**(3) 310–319.

Charnes, Abraham, William W Cooper, Edwardo Rhodes. 1978. Measuring the efficiency of decision making units. *European journal of operational research* **2**(6) 429–444.

Coogan, M., M. Hansen, M. Ryerson, L. Kiernan, J. Last, R. Marchi, R. Yatzeck. 2010. Innovative approaches to addressing aviation capacity issues in coastal mega-regions. Tech. rep. URL http://www.trb.org/Publications/Blurbs/163311.aspx.

Cook, Wade D, Larry M Seiford. 2009. Data envelopment analysis (dea)–thirty years on. *European Journal of Operational Research* **192**(1) 1–17.

Czerny, A. I. 2010. Airport congestion management under uncertainty. *Transportation Research Part B: Methodological* **44**(3) 371–380.

Daniel, J. 1995. Peak-Load-Congestion pricing of hub airport operations with endogenous scheduling and Traffic-Flow adjustments at Minneapolis-St. paul airport. *Transportation Research Record: Journal of the Transportation Research Board* **1298**(1) 1–13.

Daniel, J. I., K. T. Harback. 2008. (When) do hub airlines internalize their self-imposed congestion delays? *Journal of Urban Economics* **63**(2) 583–612.

DOT, US. 2003. Revised departmental guidance: Valuation of travel time in economic analysis. *US Department of Transportation, Washington, DC* .

Douglas, G. W., J. C. Miller. 1974. *Economic regulation of domestic air transport: theory and policy*. Brookings Institution.

FAA. 2006. Congestion Management Rule for LaGuardia Airport, Notice of proposed rulemaking, August 29. *Federal Register* **71**(167) 51360–51380.

FAA. 2008a. Congestion Management Rule for John F. Kennedy International Airport and Newark Liberty International Airport, Final Rule, October 10. *Federal Register* **73**(198) 60544–60571.

FAA. 2008b. Congestion Management Rule for John F. Kennedy International Airport and Newark Liberty International Airport, Notice of proposed rulemaking, May 21. *Federal Register* **73**(99) 29626–29651.

FAA. 2008c. Congestion Management Rule for LaGuardia Airport, Final rule, October 10. *Federal Register* **73**(198) 60574–60601.

FAA. 2008d. Congestion Management Rule for LaGuardia Airport, Supplemental notice of proposed rulemaking, April 17. *Federal Register* **73**(75) 20846–20868.

FAA. 2008e. Operating Limitations at John F. Kennedy International Airport; Order Limiting Scheduled Operations at John F. Kennedy International Airport, Notice, January 18. *Federal Register* **73**(13) 3510–3542.

FAA. 2008f. Operating Limitations at Newark Liberty International Airport; Order Limiting Scheduled Operations at Newark Liberty International Airport, Notice, May 21. *Federal Register* **73**(99) 29550–29566.

FAA. 2011. The Economic Impact of Civil Aviation on the U.S. Economy. *Federal Aviation Administration* URL `http://www.faa.gov/air_traffic/publications/media/FAA_Economic_Impact_Rpt_2011.pdf`.

FAA. 2013. Air Traffic Organization. *Federal Aviation Administration* URL `http://www.faa.gov/about/office_org/headquarters_offices/ato/`.

Flores-Fillol, R. 2010. Congested hubs. *Transportation Research Part B: Methodological* **44**(3) 358–370.

Galton, Francis. 1907. One vote, one value. *Nature* **75** 414.

Green, Paul E, Abba M Krieger, Yoram Wind. 2001. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces* **31**(3 supplement) S56–S73.

Green, Paul E, Vithala R Rao. 1971. Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research* 355–363.

Grushka-Cockayne, Yael, Bert De Reyck, Zeger Degraeve. 2008. An Integrated Decision-Making Approach for Improving European Air Traffic Management. *Management Science* **54**(8) 1395 –1409.

Hansen, M. 2002. Micro-level analysis of airport delay externalities using deterministic queuing models: a case study. *Journal of Air Transport Management* **8**(2) 73–87.

Hansen, M., C. Y. Hsiao. 2005. Going south?: Econometric analysis of US airline flight delays from 2000 to 2004. *Transportation Research Record: Journal of the Transportation Research Board* **1915** 85–94.

Hansen, M., I. Kwan. 2010. US flight delay in the 2000's: an econometric analysis. *Paper submitted to the 90th Transportation Research Board Annual Meeting* .

Hansen, M., Y. Liu, I. Kwan, B. Zou. 2010. US/Europe comparison of operational performance: a comparison of US and european airline schedules: size, strength and speed. Tech. rep.

Hansen, M., T. Nikoleris, D. Lovell, K. Vlachou, A. Odoni. 2009. Use of queuing models to estimate delay savings from 4D trajectory precision. *Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM 2009)*.

Hansen, M., W. Wei. 2006. Multivariate analysis of the impacts of NAS investments: A case study of a capacity expansion at Dallas-Fort worth airport. *Journal of Air Transport Management* **12**(5) 227–235.

Hau, T.D. 1992. Economic fundamentals of road pricing. *World Bank Policy Research Working Paper Series WPS* **1070**.

Hsiao, Chieh-Yu, Mark Hansen. 2011. A passenger demand model for air transportation in a hub-and-spoke network. *Transportation Research Part E: Logistics and Transportation Review* **47**(6) 1112–1125.

ICAO. 2005. Global Air Traffic Management Operational Concept (Document 9854). *International Civil Aviation Organization* **First Edition**. URL `http://www.icao.int/Meetings/anconf12/Documents/9854_cons_en[1].pdf`.

JEC. 2008. Your flight has been delayed again: Flight delays cost passenger, airlines, and the US economy billions. *US Congress Joint Economic Committee and others, Washington, DC* URL `http://www.jec.senate.gov/public/?a=Files.Serve&File_id=47e8d8a7-661d-4e6b-ae72-0f1831dd1207`.

JPDO. 2007. Concept of Operations for the Next Generation Air Transportation System. *Joint Planning and Development Office, Washington, DC* **Version 2.0**.

Karasakal, Esra, Murat Köksalan. 2009. Generating a representative subset of the non-dominated frontier in multiple criteria decision making. *Operations research* **57**(1) 187–199.

Kivestu, P. 1976. Alternative methods of investigating the time-dependent M/G/K queue.

SM Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA .

Le, L. 2006. Demand management at congested airports: How far are we from utopia. Ph.D. thesis, George Mason University, Fairfax, VA.

Maskin, Eric. 1985. The theory of implementation in nash equilibrium: A survey. *Social goals and social organization: Essays in memory of Elisha Pazner* 173–204.

Maskin, Eric. 1999. Nash equilibrium and welfare optimality. *The Review of Economic Studies* **66**(1) 23–38.

Maskin, E.S. 2008. Mechanism design: How to implement social goals. *The American Economic Review* **98**(3) 567–576.

Mayer, C., T. Sinai. 2003. Network effects, congestion externalities, and air traffic delays: or why not all delays are evil. *American Economic Review* **93**(4) 1194–1215.

Meyer, R., E.J. Johnson. 1995. Empirical generalizations in the modeling of consumer choice. *Marketing Science* 180–189.

Moore, John, Rafael Repullo. 1990. Nash implementation: a full characterization. *Econometrica: Journal of the Econometric Society* 1083–1099.

Morrison, S. A. 1983. Estimation of long-run prices and investment levels for airport runways. *Research in Transportation Economics* **1** 103–130.

Morrison, S. A, C. Winston. 2008. The effect of FAA expenditures on air travel delays. *Journal of Urban Economics* **63**(2) 669–678.

Neufville, R. De, A. R. Odoni, Knovel (Firm). 2003. *Airport systems: planning design, and management*. McGraw-Hill.

PANYNJ. 2008. Notice of proposed action, August 4. *Port Authority of New York and*

New Jersey .

Proussaloglou, K., F. S. Koppelman. 1999. The choice of air carrier, flight, and fare class. *Journal of Air Transport Management* **5**(4) 193–201.

Raiffa, Howard, R Keeney. 1976. *Decisions with multiple objectives: Preferences and value tradeoffs*. John Wiley and Sons.

Rupp, N. G. 2009. Do carriers internalize congestion costs? empirical evidence on the internalization question. *Journal of Urban Economics* **65**(1) 24–37.

Ruzika, Stefan, Margaret M Wiecek. 2005. Approximation methods in multiobjective programming. *Journal of optimization theory and applications* **126**(3) 473–501.

Saaty, Thomas L, Luis G Vargas. 2012. *Models, methods, concepts & applications of the analytic hierarchy process*, vol. 175. Springer.

Sayın, Serpil. 2000. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming* **87**(3) 543–560.

SESAR. 2006. Air Transport Framework: The Performance Target. *SESAR Consortium* **2**.

Vaze, V., C. Barnhart. 2011. Modeling airline frequency competition for airport congestion mitigation. *Transportation Science* **Under review**.

Wallenius, Jyrki, James S Dyer, Peter C Fishburn, Ralph E Steuer, Stanley Zionts, Kalyanmoy Deb. 2008. Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science* **54**(7) 1336–1349.

Zou, B., M. Hansen. 2011. Assessing benefits from aviation capacity investment: An equilibrium approach. *Presented at the Institute for Operations Research and the Manage-*

ment Sciences (INFORMS) Transportation Science and Logistics Society Workshop. Asilomar, California, Jun. 27, 2011.