ABSTRACT

Title of Document: DETERMINING THE NUMBER OF SLOTS TO SUBMIT TO A MARKET MECHANISM AT A SINGLE AIRPORT

Andrew Michael Churchill, M.S., 2007

Directed By: Professor David J. Lovell
Department of Civil and Environmental Engineering

In this thesis, several stochastic optimization models are presented to determine the number of airport arrival slots that should be made available for distribution via a market mechanism. Considerable attention is paid to the structure and mathematical properties of each of these models, with regards to obtaining integer-valued solutions. Calibration of the various parameters is undertaken using historical data. In addition, an analysis of the average pecuniary valuations assigned to each slot is presented, as this is an essential input to these models. Several methods are suggested by which each of these values can be estimated. The models are intended to be taken in a general context, but extensive computational examples making use of data for LaGuardia Airport are provided as a case study in the application of the various techniques presented herein.

DETERMINING THE NUMBER OF SLOTS TO SUBMIT TO A MARKET MECHANISM AT A SINGLE AIRPORT

By

Andrew Michael Churchill

Thesis 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
Master of Science
2007

Advisory Committee:
Professor David J. Lovell, Chair
Professor Michael O. Ball
Professor Paul M. Schonfeld

Acknowledgements

Of course, I would like to thank Prof. David Lovell for introducing me to world of aviation operations research, for advising me in my studies and research, and for providing advice on so many other topics.

Thanks also go to Prof. Michael Ball for being my temporary and unofficial advisor, and my permanent and official source of aviation knowledge. I am indebted to Dr. Robert Hoffman and Dr. Avijit Mukherjee for providing me with help on so many topics. Thanks of course must go to the other NEXTOR students for always providing advice, support, and snacks.

Thanks to Profs. Lovell, Ball, and Schonfeld for serving on my thesis committee, and for reading through my initial thesis drafts.

Thanks also to my friends for asking what happened when their flight was delayed, or how to find the cheapest ticket prices. While you all may not know what I really do, asking me these questions tells me that you consider me, if just a bit, knowledgeable about the world of aviation.

I would like to thank my family for supporting me in so many ways as my seemingly endless years of schooling have continued. The countless things that you (and the dogs) have done to help me cannot be repaid. Thanks also to my grandmothers, for always asking what it is that I work on.

Finally, I have to thank Alison Smith. Between cooking dinners, requiring mandatory work-times, and forcing me to go on crazy German-style marches, you have been wonderful. Thank you so much, my love!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

At most airports in the United States, air carriers are free to schedule as many flights as they would like into and out of those airports with very few restrictions. Some airports have late-night curfews, and this is particularly likely for airports with flight paths that come very close to residential areas. A few airports, however, are so congested, that allowing this kind of scheduling freedom would lead to massive congestion. In many cases, this congestion alone has not served as a deterrent to over-scheduling, probably because of the "tragedy of the commons" phenomenon: there are multiple competing airlines, and no single one of them has anything to gain by unilaterally reducing their schedule, since any benefits of doing so would help all affected carriers, but the acting carrier would yield market share. It airports such as this, control over the numbers of arrivals and departures has been effected by means of slot controls. The "slots" are landing rights at the airport, and a fixed number of them are allocated. They are distributed amongst airlines in some manner. Each corresponds to the right to schedule a flight into that airport. Carriers are not allowed to schedule flights for which there are not corresponding slots owned.

This chapter introduces the subject discussed at length in this thesis, posed here as a question: *Given various economic concerns, and conditions specific to the airport under consideration, how many arrival slots should be made available to the market in a given time period*?

Each of the issues brought forth in this question will be addressed in detail in this thesis, by means of theoretical analysis, and computational experiments. The problem will be posed and discussed in a variety of fashions, but the underlying goal remains the same. Likewise, the economic challenges necessarily involved in this analysis will not be addressed by a single omniscient solution. A variety of techniques and approaches will be explored, and several will be applied.

The majority of the analysis in this thesis lends itself to a case study of New York's LaGuardia Airport (LGA). This does not mean, however, that the techniques and models proposed in this thesis apply only to this facility. This is not an engineering study of a technique to be applied only at LGA. Rather, it is a collection of models and techniques that ought to be considered, and adapted, for use at a variety of airports experiencing extreme congestion.

The remainder of this chapter is devoted to providing background material on the issue of slot-controlled airports in the United States, motivation for the application of the techniques espoused in this thesis, and an introduction to this document and its organizational structure.

1.1 Slot Control Background

The control of arrival opportunities by means of slots ("slot control"), or limitations on the number of arrival permitted at an airport, began in the United States in 1969 when the High Density Rule (HDR) was enacted. The rule was initially designed to regulate five airports – Chicago O'Hare International Airport, Washington National Airport (now Ronald Reagan Washington National Airport), Newark International Airport (now Newark Liberty International Airport), John F. Kennedy International Airport, and LaGuardia Airport. The HDR was originally intended as a temporary solution to congestion problems, but after several short extensions it was extended indefinitely.

Prior to the deregulation of the airline industry in 1978, the Civil Aeronautics Board (CAB) exercised significant control over which carriers operated at a given airport, the number of flights that they operated, and the fares they were permitted to charge. At the HDR airports, scheduling committees consisting of airline and CAB representatives met regularly to divide the available slots. This process was only possible because of the high degree of regulation of the industry under the CAB, and the anti-trust exemptions provided thereby.

After airline deregulation, however, carriers were no longer bound by most of the restrictions placed upon them by the CAB. They were able to fly into the airports of their choosing when they preferred to do so. The HDR, however, remained in effect, but the scheduling committees could no longer be effective. In addition, a number of new carriers were created, many of which wanted to fly into the HDR airports. As a result, competitive pressures forced the carriers holding slots to guard them zealously, thereby preventing the new carriers from gaining access to these highly desirable airports.

Obviously, these developments brought about by deregulation created pressure on the Department of Transportation, the FAA, and Congress to remove or restructure the HDR to allow new carriers to gain access to the high density airports. A succession of policies and laws attempted to address this issue. These included rules to permit a secondary market for slots, to discourage carriers from hoarding slots, and to provide exemptions for several categories, including new entrant carriers, new international flights, and flights serving small communities.

An extreme rule change which was tested to alleviate the problems with the HDR was enacted by Congress in 2000. One of the provisions of the "Aviation Investment and Reform Act for the 21st Century" (AIR 21) removed slot controls and allowed the FAA to grant slot exemptions. The impact of this change was particularly troublesome at LGA. Delays and cancellations increased dramatically, and customers were extremely dissatisfied. Because of the extremely tight connectivity in many carriers' networks, the severe delays experienced at LGA were propagated through the NAS. The Port Authority of New York and New Jersey was eventually forced to impose a moratorium on new flights, effectively rescinding some of the slot exemptions offered under previous policies. The existing exemptions were re-distributed under a lottery. Although several other forms of exemptions have been granted since, the airport continues to operate in much the same fashion as it did once the slot controls were re-exerted.

At present, the HDR is discussed primarily in the context of LaGuardia Airport (LGA), as the rest of the airports included in this group have been able to deal with congestion by other means. Newark International Airport was eventually exempted from the rule. Ronald Reagan Washington National Airport (DCA) and John F. Kennedy International Airport (JFK) do not experience congestion at the same levels as LGA, and thus, do not cause the HDR to come into play. Chicago O'Hare International Airport (ORD) makes uses of the HDR, but the regulating authorities exercise control in a slightly different fashion. Because ORD is dominated by two hub carriers (United Airlines and American Airlines), the regulators are able to more easily demand reductions in flight operations. The threat is evident that if these two carriers do not curtail their operations to a level acceptable to the authorities, that administrative action will be taken to force reductions. In this case, if the two carriers cooperate, they can exact significant change on the landscape of the airport. At airports served by a more diverse group of carriers, this effect is not evident. Any reduction by a carrier at one of these more diversified airports (LGA, for example) damages them significantly, while benefiting the other carriers. No single carrier has any strong motivation to reduce flights, as their marginal contribution would be insignificant to the whole. This is an example of the "tragedy of the commons" parable frequently cited to explain seemingly irrational self-destructive behavior in various problems of unregulated resource usage, such as human population growth, environmental exploitation, etc. (see Hardin, 1968).

At LaGuardia, interim policies and laws have been applied to try to react to the sequence of new carriers wanting access, old carriers going out of business, and to the myriad changes in the aviation business landscape. At all high density airports, however, and at those expected to become so in the near future, what this sequence of events has illustrated is the need for a long-term slot allocation policy that is dynamic and robust. The players will evolve over time, and carriers will come and go. These airports are far too important to be used less than efficiently; thus, policies that allow inefficient and anti-competitive uses for slots in lieu of providing access to competing carriers are unacceptable. Ball et al. (2006) and Gleimer (1996) each provide a more detailed explanation of the evolution of slot allocation regimes since deregulation. An other reference covering the time period 1968–2000 (i.e., until the slot lottery) is contained in a background memorandum prepared for the 2000 hearing of the House

Subcommittee on Aviation on the slot lottery at LaGuardia (see U.S. House of Representatives, 2000).

On a related note, airports in much of Europe are slot-controlled in similar manners. While the exact methods by which these allocations are done seem to be proprietary, it seems that many airports do, in fact, overschedule during the most desirable times of the day, while providing delay recovery periods during the less desirable times. In this thesis, examples will be presented for LGA, but it certainly seems reasonable that this methodology could be applied to these European airports to compare the results of these models with the practices currently in place.

1.2 Motivation for Determining the Number of Slots to Offer

Allocating airport arrival slots has long been a topic of interest. Primarily, this work has focused on how to distribute the slots, given that the number of slots to distribute is known. Much of the work has focused on deriving various market mechanisms to distribute these slots, beginning with the proposal by Rassenti et al. (1982) for a sealed-bid combinatorial auction. The work by Grether et al. (1989) on auction airport arrival slot was initially commissioned as a consulting study, and actually predated Rassenti's, but was not published until much later. The primary focus on this work was to prove that several alternatives existed to manage the distribution of slots, as practically any method would have been more efficient than the scheduling committees used previously.

Analysis in this area has expanded significantly in recent years. Advances in computing have made the combinatorial versions of the auctions proposed far more tractable. Ball et al. (2006) provide a discussion of the developments in this area, and slot auctions in general. In addition, case studies have been put forth regarding individual airports, including Le et al. (2004) and Day (2004). For obvious reasons, this is not a field of study limited to the United States. DotEcon (2001) provides detailed coverage of these issues in a European context.

This analysis has not solely been limited to proposing auctions as a means to regulate airport slots. Congestion pricing has appeared as a solution in the economics literature, including in Daniel (1995), Brueckner (2002), and Mayer and Sinai (2003). Carlin and Park (1970) proposed another method for regulating runway congestion, far earlier than most of the other work in this field.

There is extensive literature relating to controlling airport arrival resource allocation. Most of these studies, however, have focused on specific allocation schemes for these slots. In this thesis, an argument is made for controlling the precise number of these resources that are made available to whatever market mechanism is used.

There are myriad reasons why the number of slots at congested airports should be regulated. These arguments cold be supported by empirical evidence from scheduling, considering the interests of travelers, and reviewing the relevant literature in queuing theory.

First, while rough estimates of arrival capacity under both good and bad conditions are known for a given airport, it is not sensible to fully schedule the airport using either value. Using the upper bounds at all times will result in horrendous delays in even mildly bad weather. Using the lower bounds will leave the airport highly underutilized. Some value in between these two bounds could be arbitrarily chosen, but it seems more reasonable to use some scientific techniques and numerical analysis to determine this exactly.

Second, it is well-known that certain times of day provide more valuable opportunities in which to operate a flight. Simply examining the schedule of any busy airport should provide evidence of this conclusion: there are more flights scheduled at certain times, and there must be some reason for this. Carrier network effects certainly play a role in this, but nonetheless, certain times of the day have a larger number of flights to more desirable destinations. Additionally, specific times of day are more desirable to the most valuable airline customer: the business traveler. For many business travelers, it is desirable to attend meetings on the same day as travel takes place, or even to make a single-day trip. These customers are highly schedule-sensitive, and because they are valuable, the airlines try to accommodate this desire.

Third, any regulations of this type must necessarily be made by the regulating authorities. The carriers operating these flights, in most cases, have little motivation to voluntarily reduce their number of operations, as this would benefit their competitors at their expense, and likely not contribute markedly to overall system performance. Additionally, unallocated resources at a congested airport would not be permitted by the rest of the market to remain unused, even if such an effort could benefit performance. These unallocated slots would be quickly taken and used. As a result, the burden of making these determinations falls directly on the regulating authorities.

An additional argument in favor of scientifically determining the number of slots to offer is provided by stochastic queueing theory. It is well-known that delays early in a time period (e.g. a day) have a far greater effect on overall system performance than those taking place later in that same time period. This argument contributes to the motivation that the number of slots must be carefully controlled, particularly at the beginning of the operating day. If, however, slot valuations dictate that a large number of slots be offered very early in the day, potentially leading to heavy delays, then it could be highly beneficial to provide some type of nominally under-utilized recovery period immediately afterwards, during which these delays can be mitigated.

Significant literature exists analyzing the eventual allocation of airport arrival resources by market mechanisms. In this section, a compelling case has been made that the quantity of these resources to be distributed (i.e. slots) should be carefully, and judiciously, regulated.

1.3 Organization

This document is organized into six related chapters. The first introduced the issue of slot-controlled airports, and provided the motivation needed for undertaking this study. The second chapter will describe a variety of discrete optimization models that can be applied, with the appropriate input data, to determine the number of slots to make available. The mathematical properties of these optimization models will be addressed at length. The third chapter will discuss the process of calibrating the input parameters used in the various optimization models already described. This is complex procedure that makes use of extensive historical data, in an effort to provide as strong a fit as is possible. The fourth chapter will address the issue of the value of an arrival slot in a given time period. It is reasonable to assume that the value of these slots is not constant across a single day, and as such, great attention must be paid to the economic analysis necessary for estimating these values. The fifth chapter will describe various techniques by which the relationship between the value and the quantity offered of an item could be incorporated into the models described thus far. Application of these methods could help the solution techniques described previously to reach a more stable equilibrium between quantity and value. The final chapter will present further results from the computational experiments conducted as part of this analysis, including discussions of the size of the various models, and their sensitivity to variations in various parameters.

Chapter 2: Model Structure

In this chapter, the problem of determining the proper number of slots to offer in each time period is posed as a series of stochastic optimization models. While each has the same general goal, they are distinguished by some important assumptions. There is a trade-off present in each model between realism and tractability, and the three models presented here represent three distinct points on this spectrum

First, the various input data required for the models will be presented. Then, three different possible formulations of the model will be discussed. The first is the Base model, which is so named because it is the simplest formulation, and its constraint structure is encapsulated in each of the two other models. Considerable attention is paid to this model, as it serves as the jumping-off point from which discussion of the other models can be handled most efficiently. The second model is the Consolidated formulation, which has a slightly modified structure, relative to the Base Model, in that certain aspects of the network structure are collapsed and consolidated in the interest of tractability. The third model, which will be used for the analysis in the sections following these discussions, is the Parametric model. This model does not share the computational ease of the previous two models, but instead is much more economically realistic and useful. Its name is derived from the fact that it incorporates important economic factors that can change across airports and carriers, and might be subject to some discretion at the hands of decision-makers.

The discussion of each of the models in this section follows the same pattern: model structure and constraints are described, then polyhedral and integer solution results are proven.

2.1 Input Data Common to All Models

A significant amount and variety of data are required to solve each of the models presented in this chapter. As will be obvious as this document progresses, much of the data is common to each and every formulation of this optimization problem. In this section, the necessary assumptions and background information for each type of data will be presented. In later chapters, some of these parameters will be examined in greater detail, but they are introduced here to facilitate the discussion of the various optimization models.

2.1.1 Assumptions

Several assumptions are common to all three forms of the model. In particular, some assumptions about the model structure and the decision variable bounds are common to every model.

The average day that the model considers is broken into discrete time slices. No specific duration for these time slices is forced – the number of time slices necessary to represent an entire day is a parameter of the model. It can be assumed, therefore, that any time resolution chosen for the model is both appropriate for the solution, and is a resolution at which the necessary input data are known. It is not necessary, for the situations under consideration, to model the entire 24-hour day, because no airport is busy for this period of time. In fact, carriers do not normally schedule arrivals into late hours, so these time periods can be used for congestion recovery on days with considerable irregular operations. The number of time slices to be considered on a day is denoted by T, and the index $t \in \{1, \ldots, T\}$ is used to represent a single time slice when subscripted to other variables.

Given this discretization of time, this problem will now be posed as a network optimization problem, in which the objective is to maximize the total value of all objects passing through the network. The objects on the network are the flights themselves. The entries into the network represent the act of scheduling flights, while the exits represent acceptance (landing) or rejection (cancellation) of that particular flight, with the recognition that a flight might not necessarily be accommodated in the same time slice that it initially desired. That is, flight delays are permitted. There are arcs in the network that represent the "movement" of delayed flights from one time period to the next.

It is assumed that there are minimum and maximum numbers of slots to be offered in each time period, denoted D_{min} and D_{max}, respectively. The maximum number available might be related to the arrival capacity of the airport, although it is important to consider that it is not necessarily efficient to cap slot offerings at the airport's VMC capacity value. Due to the stochastic nature of aircraft arrivals, some over-scheduling (compared to the average capacity) could be acceptable, although it might need to be coupled with a recovery period that is forcibly under-scheduled. Such a strategy may lead to better airport utilization during the most highly valued time periods. Of course, the implicit assumption in this analysis is that there is enough latent demand to make use of all slots offered in each time period. Because the context of this study is high density airports, this assumption is reasonable.

The minimum number of slots to be offered could be zero, or any other number less than D_{max}. It is likely that there would be some necessary political or economic reasons for requiring a higher number. Finally, while the results presented here use only a single value for each of these parameters across the day, these parameters could be time-dependent as well. This would not affect the mathematical structure of the problem.

The problem, then, is to determine what number of slots between these two values should be offered. These decision variables of primary interest are denoted $\{Z_t\}$, and require the obvious constraint shown in (2.1):

$$D_{\min} \le Z_t \le D_{\max} \quad \forall t \in \{1,\ldots,T\} \tag{2.1}$$

2.1.2 Capacity Scenarios

The primary constraint that prevents the maximum number of flights from being able to access the airport is the arrival capacity of the airport. This is not a single number, however, or even a single set of numbers over the course of a random day, given the stochastic nature of the factors, primarily weather, which influence airport capacity. However, any given airport will usually experience a finite number of capacity "scenarios" over the course of a year.

It is possible to analyze historical data about a given airport and use clustering methods to determine a limited number of capacity scenarios and their relative frequencies. Such a study was conducted by Liu et al. (2005). This analysis made use of the K-means clustering technique to determine these "average days." Results shown in Figure 1 are for New York's LaGuardia Airport. The operations at LaGuardia were clustered into six different capacity scenarios, each with an associated frequency of occurrence. Four scenarios had some appreciable number of matching days assigned to them while two strange, but unique, days never repeated. In Figure 1, only the 4 clusters with multiple days assigned to each are shown.

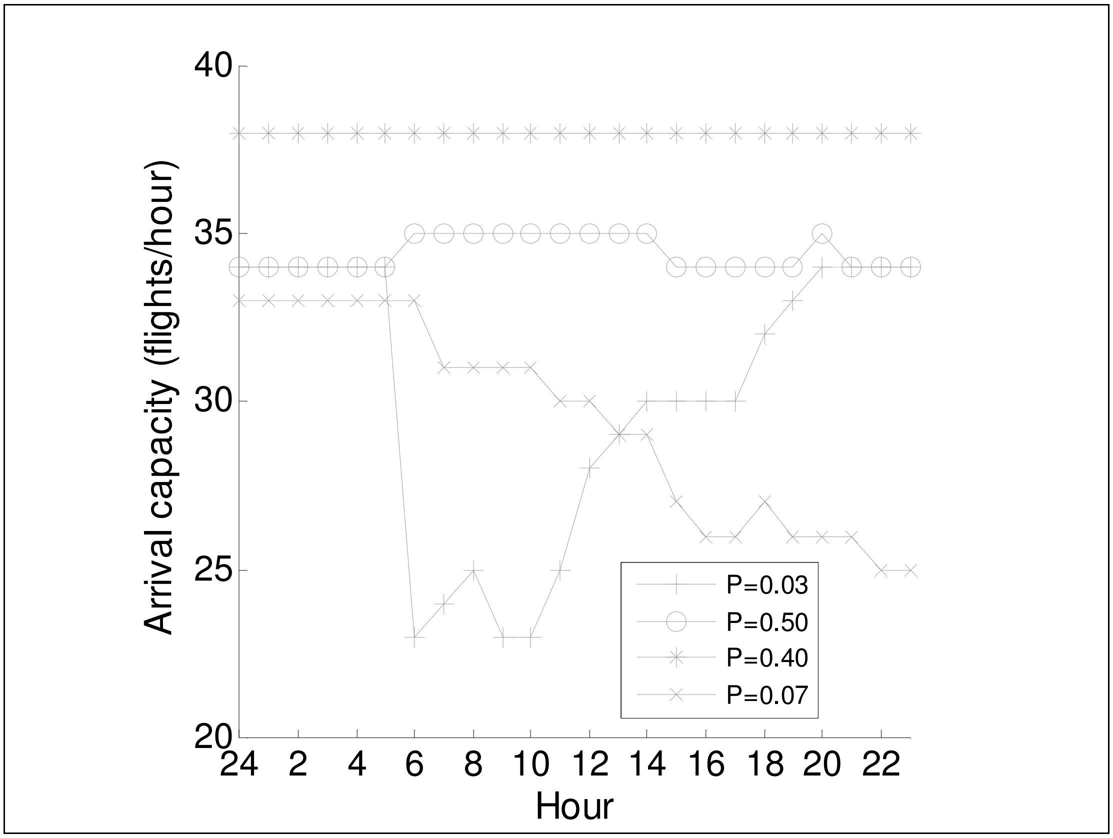

Figure 1 – LGA capacity scenarios

Figure 2 shows the cluster data for Hartsfield-Jackson Atlanta International Airport (ATL). Figure 3 shows similar data for Chicago O'Hare International Airport (ORD). In each of these analyses, there were five discrete scenarios, each with an appreciable number of days assigned to it. One issue of note regarding the data for ATL and ORD is the appearance of some periodic phenomenon in several of the scenarios. The origin of this phenomenon is the data source, the Federal Aviation Administration's Aviation System Performance Metrics (ASPM) database, used in the clustering process. In general, the data in the ASPM system is of high quality. The source of this problem, however, is that the AAR is declared only once for each hour, while ASPM reports quarter-hourly. These are determined by a scheme that divides the declared hourly rate by four, but rounds in such a way so as to guarantee integer solutions that sum to the declared value. As a result, the quarter-hour declared values occasionally exhibit some periodic behavior. Because the period of this cycle is one hour, which is equal to the lower bound of the maximum permitted delay length used in this analysis, this should not significantly affect the results presented.

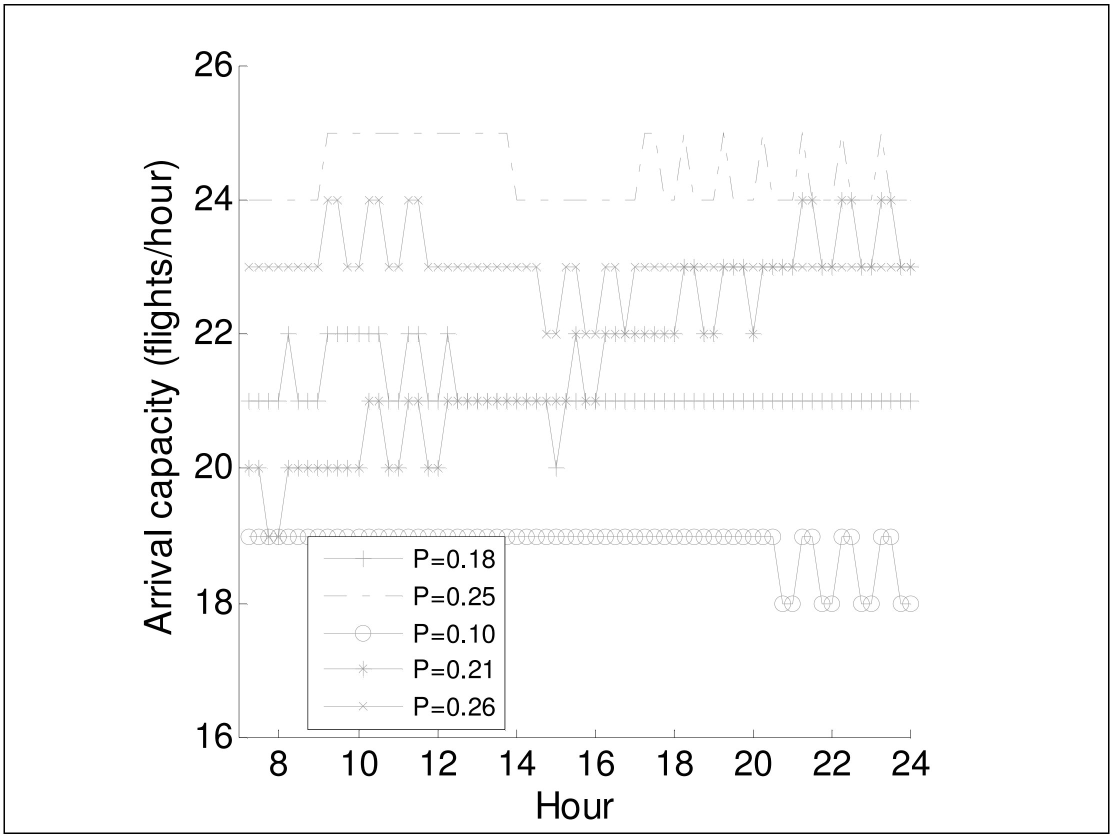

Figure 2 – ATL capacity scenarios

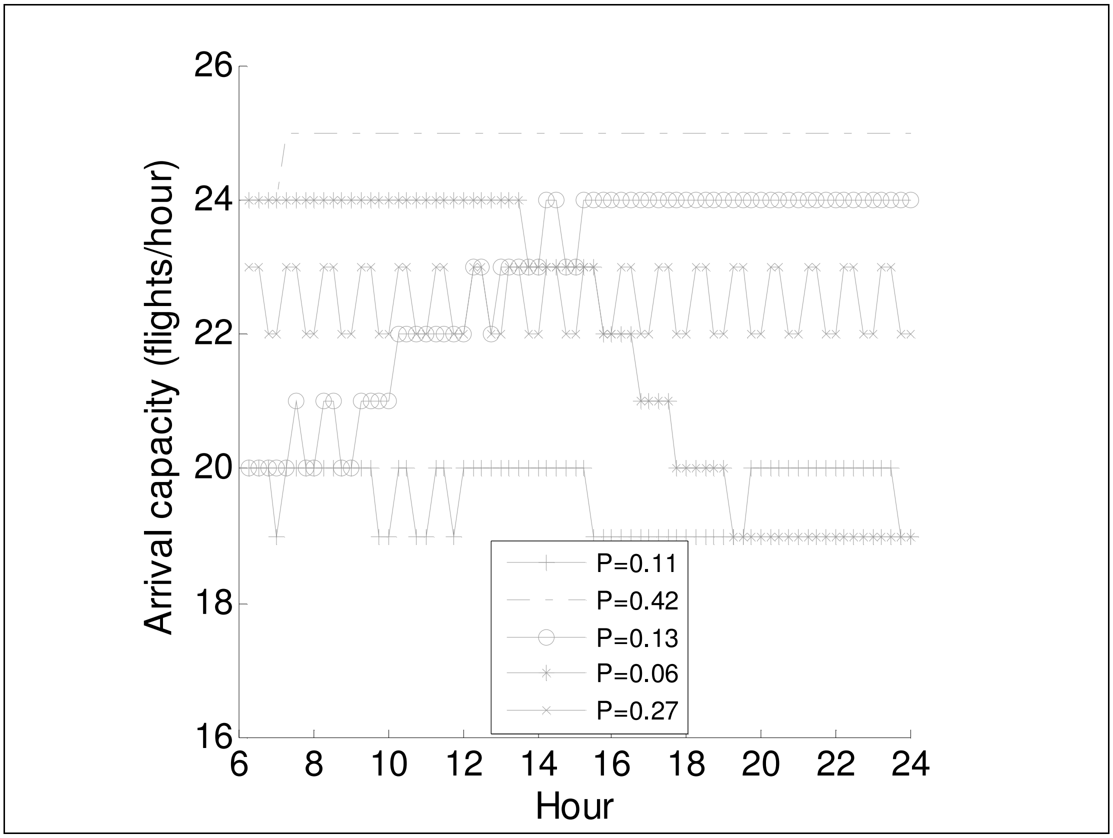

Figure 3 – ORD capacity scenarios

Assuming that such historical patterns of capacity will continue into the future, each capacity scenario is applied to the problem on a conditional basis, with the ultimate levels of delay and cancellation being the expected values over the full range of possibilities. The weights used for each scenario are the associated likelihood (i.e. historical frequency). Thus, the index q used in the remainder of this thesis represents an entry into the set of capacity scenarios that have been defined for a given airport, with range $q \in \{1, \ldots, Q\}$.

2.1.3 Delay Costs

For the models presented herein, no specific cost for flight delays is assumed. Rather, all other values in the problem are normalized into units of delay. That is, the cancellation costs $\{\lambda_i\}$ do not represent monetary costs, but rather the number of time slices of delay at which a carrier would be indifferent between accepting the delay and canceling the flight. This structure is essential to the formulation of the model. This is obviously a major assumption, and to make use of it, it must be acceptable to believe that such trade-offs can be determined explicitly and without regard to individual carrier preferences. Because the motivation of this study is to help create policy for an entire airport, which should serve many carriers, such an assumption will suffice.

In the Base and Consolidated models (but not the Parametric), the slots values are also expressed in terms of the equivalent number of units of delay. This assumption makes these two models very difficult to implement, because it requires that the slot values be determined solely using information regarding flight delays, or that delay be assigned a monetary value, by which the flight delays can be normalized. In either case, further significant assumptions would be required, rendering the model less useful, from a policy-making perspective. That the Parametric model does not require such assumptions is, perhaps, the strongest argument for using it.

2.1.4 Cancellation Costs

The cancellation of a flight is an extreme and infrequent, but sometimes necessary, measure. Modeling this decision is a difficult task, largely because the process by which each carrier determines if and when to cancel flights is a closely guarded secret. Data or even anecdotes about important causal mechanisms are not generally available, and carriers' business strategies might differ from each other significantly. In the present context, however, and at any level where multiple airlines and their respective operating strategies are in play, it is reasonable to assume that neither the internal data nor the operational strategies of the carriers themselves will be publicly known, and, as such, none of the carrier-centric predictive cancellation models developed to date are functional in the framework of this model.

Therefore, the modeling of cancellations is approached from a macroscopic point of view, under the belief that, in the aggregate, airlines use voluntary cancellations primarily to hedge against excessive delays.

As explained previously, the cancellation costs in all three formulations are expressed in units of delay. The process by which this trade-off is estimated will be further explained in Chapter 3. It is reasonable to suppose that the marginal cost of canceling an additional flight increases as the number of flights already canceled increases. For this reason, cancellation costs are expressed at several levels, based on the number of flights which have already been canceled. The index i stratifies the cancellation costs along the various cancellation arcs – the notation λ_i is used to represent the cancellation cost on arc i. Additionally, each cancellation arc has a capacity P_i, which essentially limits the number of cancellations that can be accommodated at that cost. The combination of the costs $\{\lambda_i\}$ and capacities $\{P_i\}$ allow for the specification of a marginal cancellation cost curve that is a step function – presumably monotonically non-decreasing. The total number of cancellation arcs is denoted by *N*. The flow variable $\theta_{t,q}^{i}$ represents the number of flights that are cancelled during time slice *t* and suffer cancellation cost level λ_i, under capacity scenario *q*. The cancellation arc capacities are enforced as shown in (2.2).

$$\theta_{t,q}{}^{i} \le P_i \quad \forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\} \tag{2.2}$$

Figure 4 shows a non-specific example of a piecewise-linear total cost function that would have such a marginal cost function as its derivative. This function can be specified with whatever resolution is required, provided that sufficient data are available to calibrate the parameters of such a function. Of course, increasing the number of cancellation arcs increases the size of the problem, which in turn affects its solution time, so the best choice of *N* would be the smallest that allows the minimum desired resolution in the cancellation cost function.

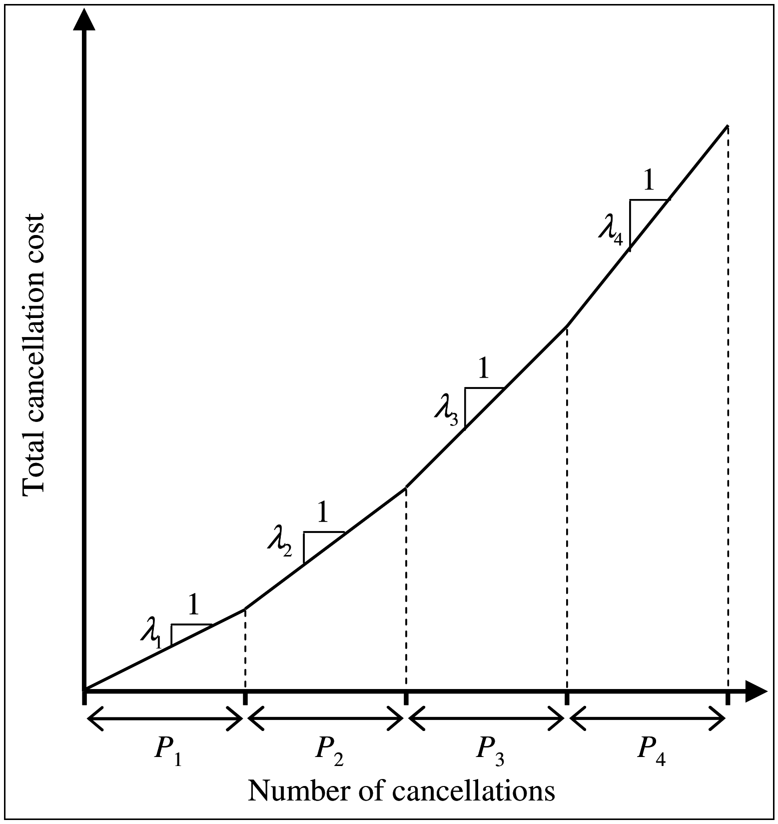

Figure 4 – Convex piecewise-linear cancellation cost curve

2.2 Base Model

The model presented in this section determines the number of slots to be allocated in each time period by optimizing a function that represents the difference between the value of the slots offered and the costs incurred from the resultant levels of delay and cancellation. In some sense, this can be thought of as the net benefit to all the airlines operating at the airport. The levels of delay and cancellation occurring under this slot offering are not explicitly regulated: the model will drive them to some "system-optimal" values, based in large part on the relative costs of all elements involved.

2.2.1 Structure

Figure 5 shows a network representation of the Base Model. The input nodes are white, and are located at the top of the diagram. They have a maximum permissible input of D_{max} for each of the time slices $t \in \{1,\ldots,T\}$. The flows that are permitted from the input nodes to the next stage in the network are the slots offered in each time slice, represented by the decision variables $\{Z_t\}$.

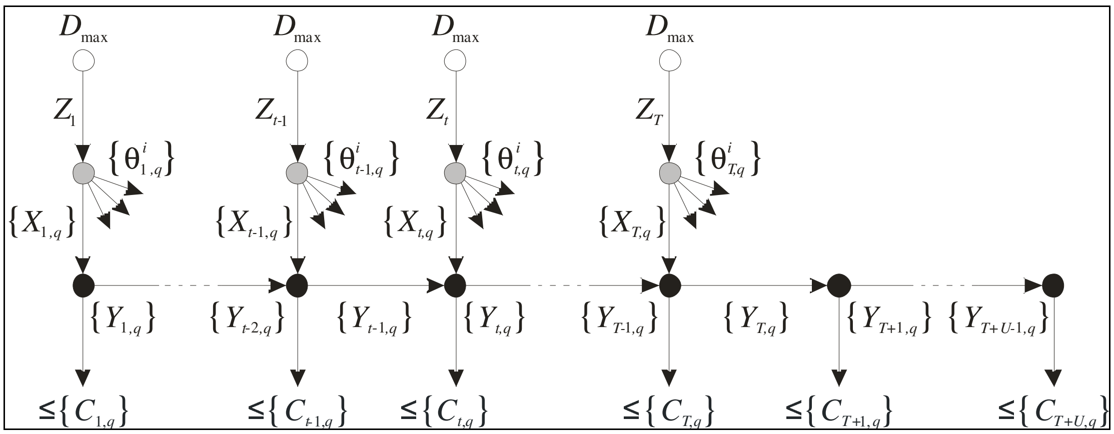

Figure 5 – Base model flow diagram

Given that a slot has been offered (i.e., the slot will be utilized), a flight can have one of three possible dispositions: it can be cancelled, it can be delayed but ultimately admitted to land at the airport, or it can be admitted without delay. The gray nodes at the second horizontal level of the network represent the carrier's decision of whether to cancel a flight or not. The diagram becomes slightly more complicated at this point, as it is now stratified in two additional dimensions, represented by the indices *i* and *q* representing the capacity scenarios and cancellation cost levels described previously. All variables that are grouped together in strata that are not explicitly shown in the figure are indicated in braces.

The number of flights that are not canceled in time slice t under capacity scenario q is $X_{t,q}$. Because each set of capacity constraints will produce a different optimal flow across the network, all flow variables exiting the gray nodes are stratified by capacity scenario. Recall that the flow variable $\theta^i_{t,q}$ represents the number of flights that are cancelled during time slice t and suffer cancellation cost level i, under capacity scenario q. Flow must be conserved at the gray nodes, as shown in (2.3):

$$Z_t - X_{t,q} - \sum_i \theta_{t,q}{}^i = 0 \quad \forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\} \tag{2.3}$$

Flights that are not cancelled proceed through the network from the gray nodes to the black nodes, where the decision to be made is whether they are allowed to land immediately, or with delay. If they are delayed, they are transferred to later time slices by moving in the right-hand direction in Figure 5. The number of flights delayed from time slice t to time slice t+1 under capacity scenario q is denoted $Y_{t,q}$. Notice that flights can be delayed into time slices later than the latest scheduled demand. In fact, the maximum number of time slices that a flight can be delayed is a parameter denoted U. Thus, an additional quantity, U, of black delay nodes are required after time slice T to allow for flights that may be delayed past the scheduled portion of the day. If an airport with a curfew were being considered, the number of these nodes after the end of the scheduled day could be reduced to match the time available for the curfew.

Flow must be conserved at each of the black nodes. The exact structure of the constraint governing each node depends on the time period in which the node lies. These constraints are shown in (2.4)-(2.6):

$$X_{1,q} - Y_{1,q} \le C_{1,q} \quad \forall q \in \{1,\ldots,Q\} \tag{2.4}$$

$$X_{t,q} + Y_{t-1,q} - Y_{t,q} \le C_{t,q} \quad \forall t \in \{2,\ldots,T\},\ q \in \{1,\ldots,Q\} \tag{2.5}$$

$$Y_{t-1,q} - Y_{t,q} \le C_{t,q} \quad \forall t \in \{T+1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\} \tag{2.6}$$

Once it is time for them to land, non-canceled flights travel on the arcs going downward from the black nodes. These represent landings during the given time period t, obviously subject to the airport capacity constraints discussed previously. Thus, the constant $C_{t,q}$ represents the airport capacity during time slice t, under capacity scenario q. Capacity constraints must be observed on these arcs extending down from the black nodes, as shown in (2.7):

$$Y_{T+U-1,q} \le C_{T+U,q} \quad \forall q \in \{1,\ldots,Q\} \tag{2.7}$$

There is a necessary constraint on the maximum number of delay arcs that a given flight can traverse, which is given by the constant U described above. Because this is a network flow model, the entities are not labeled; i.e., there is no ability to track a specific flight as it travels through the model. However, this constraint can still be enforced in the aggregate by specifying capacities for the delay arcs. Given a solution that satisfies these constraints, a mapping of individual flights that would violate the constraint at the individual level could likely be constructed; however, with the same aggregate solution, flight labels could be swapped and that solution transformed into one where the constraints were satisfied even at the individual flight level, with no change in optimality conditions. In order to restrict the maximum delay of any flight to U units, the capacities of the horizontal arcs – denoted W_t – are given by the sum of airport arrival capacities during the subsequent U time periods, except at the end of the day, as shown in (2.8):

$$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_{i,q} \quad \forall t \in \{1,\ldots,T+U-1\},\, q \in \{1,\ldots,Q\} \tag{2.8}$$

The delay arc capacities are then enforced in a very straightforward manner, as shown in (2.9):

$$Y_{t,q} \le W_{t,q} \quad \forall t \in \{1,\ldots,T+U-1\},\, q \in \{1,\ldots,Q\} \tag{2.9}$$

The objective of the program is to maximize the net benefit provided by offering these slots. This net benefit is the difference between the value of the slots offered and the penalties imposed for those delays and cancellations. The costs of cancelling a flight, in units of time periods of delay, have already been described as $\{\lambda_i\}$. The value of a slot offered in time slice t is denoted as V_t. Again, this must be expressed in units consistent with the other costs used in this formulation of the model: periods of delay. Obviously, this requires a value assessment on the part of the entity involved in setting the policy for slot availability. Some values of $\{\lambda_i\}$ are estimated in Chapter 3, and some possible techniques to obtain values of $\{V_t\}$ are presented in Chapter 4.

The final, and obvious, constraint is that the decision and flow variables must be non-negative and integer-valued, as shown in (2.10).

$$X_{t,q}, Y_{t,q}, Z_t, \theta_{t,q}{}^{i} \in \mathbb{N}^{+} \quad \forall t \in \{1,\ldots,T\},\, q \in \{1,\ldots,Q\},\, i \in \{1,\ldots,N\} \tag{2.10}$$

Each set of variables stratified by the index q corresponds to a specific capacity scenario, as described previously. The probability (or frequency) expected for scenario q is given by p_q. To find the expected delay and cancellation costs, a weighted average (i.e., expectation) of the costs according to the weights $\{p_q\}$ is used. Under this construction, the objective function for the current problem becomes (2.11). The constraints for this problem are repeated in Table 1 for clarity.

$$\max\left\{\sum_t V_t Z_t - \sum_q p_q\left[\sum_t\left(Y_{t,q} + \sum_i \lambda_i \theta_{t,q}{}^i\right)\right]\right\} \qquad (2.11)$$

Subject to

$$\theta_{t,q}{}^i \le P_i \qquad \forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\} \qquad (2.2)$$

$$Z_t - X_{t,q} - \sum\nolimits_i \theta_{t,q}{}^i = 0 \qquad \forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\} \qquad (2.3)$$

$$X_{1,q} - Y_{1,q} \le C_{1,q} \qquad \forall q \in \{1,\ldots,Q\} \qquad (2.4)$$

$$X_{t,q} + Y_{t-1,q} - Y_{t,q} \le C_{t,q} \qquad \forall t \in \{2,\ldots,T\},\ q \in \{1,\ldots,Q\} \qquad (2.5)$$

$$Y_{t-1,q} - Y_{t,q} \le C_{t,q} \qquad \forall t \in \{T+1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\} \qquad (2.6)$$

$$Y_{T+U-1,q} \le C_{T+U,q} \qquad \forall q \in \{1,\ldots,Q\} \qquad (2.7)$$

$$Y_{t,q} \le W_{t,q} \qquad \forall t \in \{1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\} \qquad (2.9)$$

$$D_{\min} \le Z_t \le D_{\max} \qquad \forall t \in \{1,\ldots,T\} \qquad (2.1)$$

$$X_{t,q}, Y_{t,q}, Z_t, \theta_{t,q}{}^i \in \mathbb{N}^+ \qquad \forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\} \qquad (2.10)$$

Where

$$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_{i,q} \qquad \forall t \in \{1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\} \qquad (2.8)$$

Table 1 – Base model formulation

2.2.2 Mathematical Properties

Structurally, the model given by equations (2.1) - (2.10) is an integer linear program (IP). For two very simple versions of the problem, the constraint matrix is totally unimodular (TU). This is a useful property, as it ensures that a linear programming (LP) relaxation of the problem, if solved using a vertex method like the

Simplex algorithm, will produce optimal integer solutions without the need for IP-specific solution techniques such as branch-and-bound. To prove that the versions of the problem in question have TU constraint matrices, a well-known result presented in, among other places, Nemhauser and Wolsey (1988), Section III.1 is used:

Lemma 1: *An integer matrix is totally unimodular if and only if for every possible subset of its rows, a partition of those rows can be found, for which in each column, the column sums on opposite sides of the partition differ by no more than 1.*

Corollary 1: *A (-1,0,1) matrix is TU if every column contains at most two non-zero entries, and if the full matrix under consideration can be partitioned in such a way that in every column whose two non-zero entries are the same, they must sit on opposite sides of the partition, while if they are very different (i.e., a 1 and a* -1*), they must sit on the same side of the partition.*

Notice that in the case of the corollary, only the full matrix needs to be checked – if this condition is true for the full matrix, then any arbitrary subset of rows can be arrived at by deleting rows from the partition that worked for the full matrix, but since each column contains at most two non-zero entries, deleting rows cannot produce a violation of the column sum condition. Since total unimodularity is invariant under matrix transposition, this entire discussion is also true when all instances of the words "row" and "column" are interchanged.

Proposition 1. *Any version of the base problem with Q = 2 has a constraint matrix that is totally unimodular.*

***Proof*:** Suppose the constraints for a problem with $Q = 2$ are expressed in the canonical form $A\mathbf{x} \le \mathbf{b}$, where A is the constraint matrix. First, we pare down the matrix A by removing rows and columns whose consideration is unnecessary. The validity of these steps is well-known and is stated mathematically in places like Nemhauser and Wolsey (1988); the reasoning is stated here verbally because it assists with problem understanding. The equality constraints (2.3) are converted to a matched set of inequality constraints, one set representing an upper bound and the other a lower bound. The rows in A for the upper bound will be the negative of those for the lower bound, and only one set of these rows need be considered, for the following reason. Suppose r is a row representing a lower bound for constraint (2.3), and r' is its matching upper bound. Any submatrix involving only row r (not r') will have a determinant whose value is the negative of what would have been obtained if r were replaced with r'. Any submatrix involving both rows will be singular, because two of its rows are obviously linearly dependent. Therefore, for the purposes of evaluating whether a matrix is TU or not, equality constraints can be imagined to be a single set of inequality constraints instead.

Next, it is well known that any row or column containing only a single entry of 1 or -1 need not be considered, because any submatrix involving that row or column, when evaluated using the Laplace expansion, will involve only a single (unsigned) cofactor which is the determinant of a smaller matrix that needs to be considered anyway. Thus, the row or column in question is redundant. Using these last two results, constraints (2.1), (2.2), and (2.9) can be safely ignored, and constraints (2.3) can be treated as inequalities. Furthermore, all columns representing variables $\left\{\theta_{t,q}^{i}\right\}$ can be removed, because once the constraints (2.2) are ignored, each variable in $\left\{\theta_{t,q}^{i}\right\}$ appears in exactly one constraint. When working through the time slices backwards, constraints (2.6) and (2.7) can be removed, as can all columns corresponding to $\left\{Y_{t,q}\right\}$ variables for $T \leq t \leq T+U-1$. This last activity must be done step-wise, since at first only the last row from (2.7) has one non-zero entry, but removal of that row yields a column with only one non-zero entry, whose removal leaves the next-to-last entry of (2.7) with only one non-zero entry, and so on, until all of the specified rows and columns have been deleted. This step-wise deletion process must stop at constraints (2.5), as each of those rows contains three non-zero entries.

For the case Q = 2, the matrix that remains will consist entirely of columns containing exactly two non-zero elements. For example, each of the columns for the variables $\left\{Z_{t}\right\}$ will contain 1's in the rows for constraints (2.3) for the same value of t, one corresponding to q = 1 and the other to q = 2. Additionally, each of the columns for the remaining variables corresponds to a flow in the network that has to exit one node and enter a subsequent node; hence the column will have a single 1 and a single -1 only. If the matrix is partitioned according to q, then each column for a variable in $\left\{Z_{t}\right\}$ will have a 1 on each side of the partition. Because the flows are separable across capacity scenarios, the 1 and -1 in every other column belong to rows associated with the same value of q. This partition clearly satisfies the column sum condition mentioned above. ■

Proposition 2**:** *Any version of the base problem with* T *= 2 has a constraint matrix that is TU.*

Proof**:** The same paring scheme and partition argument as above is used, except that the base partition is formed differently. The constraint matrix is partitioned by columns instead of rows in this case, and segregated according to the variable t instead of q. It is easy to verify that each row will satisfy the row sum condition across this partition. Furthermore, an arbitrary set of columns can be deleted on either side of the partition without violating this condition. ■

Computationally, these results are useful because in these cases, optimal solutions can be found on an integer lattice using linear programming (LP) relaxation with a vertex algorithm. For problems that need to be executed many times (e.g., in real time, in a simulation setting, or in an equilibrium-seeking iterative loop), computational efficiency is very important. This concern is not an issue if the problem needs only be solved once, and not in real time.

Unfortunately, it is not likely that the capacity spectrum at real airports could be modeled adequately with only two scenarios, nor is it likely that only two time slices would suffice to represent the temporal dynamics at an airport. For any more realistic problems, the next result shows that the constraint matrix is *not* TU, which does not necessarily prohibit integral optimal solutions coming from the LP relaxation, but likely makes the proof of such results more complicated.

***Proposition 3*:** Any version of the base problem with $Q \geq 3$ and $T \geq 3$ has a constraint matrix that is not TU.

***Proof*:** For the smallest problem that meets this description, i.e., with T = 3, Q = 3, N = 1, and U = 1, a 13×13 submatrix can be found whose determinant is -2, as shown in Table 2.

Constraint	Label	Z_1	Z_2	Z_3	$X_{1,2}$	$X_{1,3}$	$X_{2,1}$	$X_{2,2}$	$X_{3,1}$	$X_{3,3}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{2,1}$	$Y_{2,3}$
(2.3)	A	1			-1									
(2.3)	B	1				-1								
(2.3)	C		1				-1							
(2.3)	D		1					-1						
(2.3)	E			1					-1					
(2.3)	F			1						-1				
(2.4)	G				1						-1			
(2.4)	H					1						-1		
(2.5)	I						1						-1	
(2.5)	J							1			1			
(2.5)	K											1		-1
(2.5)	L								1				1	
(2.5)	M									1				1

Table 2 – Smallest non-unimodular submatrix

Any problem larger than this, in conjunction with any combination of the dimensions *T, Q*, *N*, and *U*, will contain this matrix as a submatrix. Therefore, its constraint matrix cannot be TU. ■

Because each row of the matrix in Table 2 has exactly two non-zero entries, one way to analyze this matrix is to form a network connectivity graph, whose nodes represent problem variables, and for which arcs exist between two nodes any time the two appear together in a constraint row. For this particular non-TU submatrix, the fact that its determinant is -2 is related to the fact that in the graph so described, the arc cycle shown in Figure 6 is present.

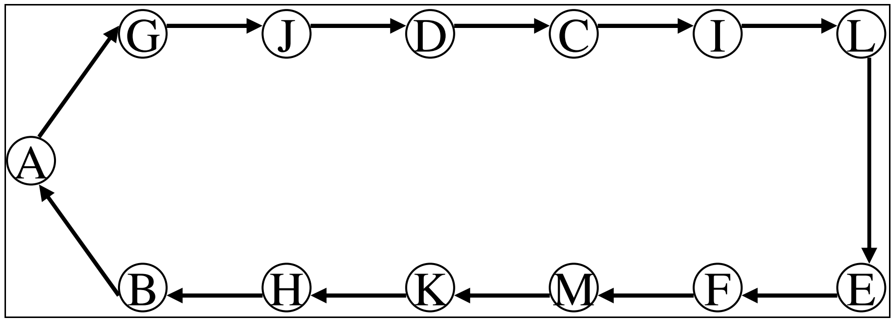

Figure 6 – Arc cycle present in constraint matrix

The fact that for these realistically sized problems the constraint matrix is not TU does not necessarily mean that LP relaxation will not produce integral optimal solutions. In fact, for certain types of capacity values, in fact, it can be guaranteed to do so, if the constraint matrix conforms to a condition that is a generalization of the notion of TU. To prove this, some very recent results on a natural extension of the ideas of total unimodularity are required, as reported by Appa and Kotynek (2004). Because they are new and perhaps unfamiliar to some readers, they are repeated here. The presentation of Appa and Kotnyek (2004) has been modified to reflect the notation and vocabulary of this thesis.

Definition 1**:** *A rational matrix is called* k-*regular if for each of its non-singular square submatrices R, kR^{-1} is integral, where $k \geq 2$ is an integer*.

Lemma 2**:** *If for each non-singular square submatrix R of a matrix A,* $\det(R) \in \{\pm 1, \pm k\}$*, then A is k-regular.*

Lemma 3: *The matrix A is* k-*regular if and only if the polyhedron* $P(A;k\mathbf{b}) = \{\mathbf{x} \mid \mathbf{x} \geq \mathbf{0}; A\mathbf{x} \leq k\mathbf{b}\}$ *is integral for each integral vector* ***b***.

The important thing to recognize from Lemma 3 is that if A is a k-regular constraint matrix and $\mathbf{b}$ is a right hand side vector consisting of integers that are multiples of the integer k, then the feasible region $P(A;k\mathbf{b})$ has extreme points that can only be integer-valued vectors. Thus, the linear programming relaxation will produce an optimal point that is integral. To complete the proof, therefore, it must be shown that any matrix A from the base problem meets the conditions of Lemma 2 for $k = 2$, establishing that the matrices in question are 2-regular. As of this the completion of this thesis, this fact has not been determined with certainty; however we include a sketch proof here and highlight the missing step which, if supplied, would complete the proof. Importantly, having tested a very large number of constraint matrices, we have not found one that did not conform with 2-regularity, so our suspicion is strong that the proof can ultimately be successful. If it is, then as long as even-valued right hand side vectors are applied, the feasible polyhedron will be integral. What this implies, practically, is that in order to retain the guarantee of integrality provided by this theorem, the capacities in each scenarios must be limited to even values, since, other than the number zero (which is even of course), these are the parameters that appear in the right-hand side of the pared-down set of constraints on which the question rests.

Begin by considering a matrix A from the base problem that has been pared down following the steps in Proposition 1. While not stated explicitly here, Appa and Kotnyek (2004)0 also shows that these exact paring steps do not change the k-regularity of a matrix. Next we show that the matrix A satisfies the requirements for Lemma 2 for $k = 2$.

Lemma 4**:** *Any non-singular square submatrix R of A has* $\det(R) \in \{\pm 1, \pm 2\}$.

Proof**:** This proof is not complete, but we include it anyway under the assumption that it can be made complete. This missing logical connection will be highlighted where it appears in the proof. We begin by showing that any non-singular 2×2 submatrix of A has determinant ± 1. In this problem, no pair of variables ever appears simultaneously in two distinct constraints; hence every 2×2 submatrix contains at least one zero. Evaluating the determinant of any 2×2 submatrix by Laplace expansion along a row containing one element in $\{-1, 0, 1\}$ and one zero must yield a determinant of either zero or ± 1.

Next, we show that any non-singular 3×3 submatrix has determinant in $\{\pm 1, \pm 2\}$. Any square submatrix of *A* has at least one row or column with at most two non-zero entries for the following reasons: a) the only columns with more than two non-zero entries are those associated with $\{Z_t\}$, but for any of those that are present, the row from constraint (2.3) for ordered pair (*t*,*q*) must also be present for some value of *q*, and that row contains at most two non-zero entries; b) similarly, the only rows with more than two non-zero entries are those associated with constraints (2.5) for $2 \leq t < T$. Each of these, however, has some non-empty subset of the following: a 1 in $X_{t,q}$, a 1 in $Y_{t-1,q}$, and a -1 in $Y_{t,q}$. The columns associated with these variables all have at most two non-zero entries. Thus, the determinant of any non-singular 3×3 matrix can be evaluated by Laplace expansion along some row or column with at most two non-zero entries, each of whose minors is either -1, 0, or 1, so the determinant must be in $\{\pm 1, \pm 2\}$.

We would now like to complete the proof of Lemma 4 by induction on the size of the square submatrix. Having shown that non-singular 3×3 matrices have determinants in $\{\pm1,\pm2\}$, the induction hypothesis is that non-singular $n\times n$ matrices have determinants in $\{\pm1,\pm2\}$. The remaining step would be to show that this implies that the same is true for matrices of size $(n+1)\times(n+1)$. Suppose *R* is an arbitrary $(n+1)\times(n+1)$ submatrix of *A*. Again, exploit the fact that *R* must have at least one row or column with *at most* two non-zero entries. Furthermore, the induction hypothesis requires that if $|\det(R)|>2$ is at all possible, then all rows and columns of *R* must have *at least* two non-zero entries.

Any non-singular submatrix *R* with $|\det(R)|>2$ must have at least one column corresponding to a variable in $\{X_{t,q}\}$. The reasoning for this is as follows. If a column from $\{Z_t\}$ is included, then it has non-zero entries in rows corresponding to constraints (2.3). In order for these rows to have at least two non-zero entries, then corresponding columns from $\{X_{t,q}\}$ variables must also be present in the matrix. Similarly, if a variable from $\{Y_{t,q}\}$ is included, it must be matched with another column in order to present rows with at least two non-zero entries. The other column might come from $\{X_{t,q}\}$ or it might come from a different variable in $\{Y_{t,q}\}$, from a different time slice. The process then has to be repeated for that new $\{Y_{t,q}\}$ variable, so again its matching column can be from $\{X_{t,q}\}$ or from a different variable $\{Y_{t,q}\}$, moving in the same direction time-wise as the previous step. Importantly, however, this sequence of necessary $\{Y_{t,q}\}$ variables cannot proceed forever, because ultimately this process with end with a $\{Y_{t,q}\}$ variable either from a constraint of type (2.4), if moving backwards in time, or of type (2.5) with $t = T - 1$, if moving forwards in time. In either case, the only way to provide the second non-zero entry in the row in question is to include a column from $\{X_{t,q}\}$.

In the pared matrix we are considering, columns from $\{X_{t,q}\}$ have exactly two 1's as entries, one from a constraint of type (2.3) and one from either (2.4) or (2.5). In any event, the determinant of this matrix can be found by Laplace expansion along a column of $\{X_{t,q}\}$ with exactly two 1's. The rows of the matrix can be interchanged until the rows with those two 1's are adjacent, and this affects the sign but not the magnitude of the determinant.

In order for the proof to succeed, what needs to be demonstrated at this point is that the two (unsigned) cofactors involved in that Laplace expansion are of equal sign. Then, since the rows are adjacent, the determinant of the submatrix in question is their difference. By the induction hypothesis, the unsigned cofactors must lie in $\{0,1,2\}$, the difference between any pair of these numbers is at most 2 in absolute value. At this point, the proof would be complete. The missing part is showing that the unsigned cofactors have the same sign. The $n \times n$ matrices used to determine these cofactors are identical except for one row, and this is probably critical to the proof. In general, non-singular matrices that differ in only one row can have determinants of opposite sign, so what is sought is a condition imposed by the structure of this problem that prevents this for the matrices in question.

Proposition 4: *Any version of the base problem even-valued capacities will solve optimally to integer values using LP relaxation.*

Proof. If Lemma 4 turns out to be provable, then *A* can only have submatrix determinants in $\{0,\pm1,\pm2\}$. Thus, by Lemma 2, *A* is 2-regular, and by Lemma 3, the proof is complete. ■

2.3 Consolidated Model

It is possible to alter the structure of the Base model in such a way that the constraint matrix can be guaranteed to be TU. This requires consolidation of the cancellation and delay decisions to the same nodes, and so the model is called the Consolidated model. This formulation of the model maximizes the same objective as the Base model. The cost for this gain in computational tractability is that this model requires economic assumptions that may be strict and unrealistic. At the very least, however, the structure of the Consolidated model provides an important context for discussing the third formulation.

2.3.1 Structure

It is possible to pose the slot determination problem in a slightly different way that can be shown to be TU, even for large numbers of time slices and capacity scenarios. The trick is to change the order in which decisions are made. In the Base formulation, the cancellation decision is made first, and then only those flights not cancelled are considered for delay. In this formulation, the cancellation and delay decisions are consolidated in the same (gray) nodes. See Figure 7 for a network diagram that illustrates this problem. The white nodes remain the source nodes for flights, and all variables are defined as before.

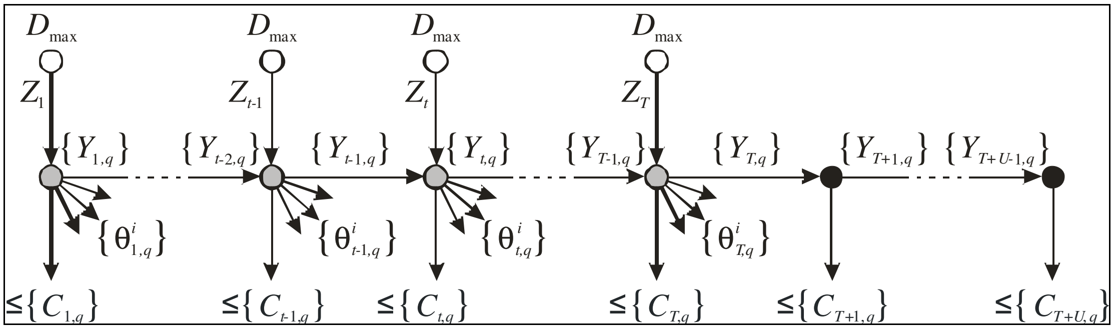

Figure 7 – Consolidated model flow diagram

In this problem, the Base formulation constraints (2.1), (2.2), and (2.6)-(2.9), including the pre-processing step (2.8), and the objective function (2.11) remain the same. The flow conservation constraints must be replaced, with the airport capacity incorporated explicitly, as shown in (2.12) and (2.13).

$$Z_1 - Y_{1,q} - \sum_i \theta_{1,q}{}^i \leq \mathrm{C}_{1,q} \quad \forall q \in \{1,\ldots,Q\} \tag{2.12}$$

$$Z_t + Y_{t-1,q} - Y_{t,q} - \sum_i \theta_{t,q}{}^i \leq \mathrm{C}_{t,q} \quad \forall t \in \{2,\ldots,T\},\ q \in \{1,\ldots,Q\} \tag{2.13}$$

As the $X_{t,q}$ variables are removed, the integrality and non-negativity constraints now become (2.14).

$$Y_{t,q}, Z_t, \theta_{t,q}{}^i \in \mathbb{N}^+ \quad \forall t \in \{1,\ldots,T\}, q \in \{1,\ldots,Q\}, i \in \{1,\ldots,N\} \tag{2.14}$$

The objective function remains the same as in the Base model, as shown in (2.15). The entire Consolidated formulation is repeated in Table 3.

$$\max\left\{\sum_t V_t Z_t - \sum_q p_q\left[\sum_t\left(Y_{t,q} + \sum_i \lambda_i \theta_{t,q}{}^i\right)\right]\right\} \quad (2.15)$$

Subject to:

$\theta_{t,q}{}^i \le P_i$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\}, i \in \{1,\ldots,N\}$	(2.2)
$Z_1 - Y_{1,q} - \sum_i \theta_{1,q}{}^i \le \mathrm{C}_{1,q}$	$\forall q \in \{1,\ldots,Q\}$	(2.12)
$Z_t + Y_{t-1,q} - Y_{t,q} - \sum_i \theta_{t,q}{}^i \le \mathrm{C}_{t,q}$	$\forall t \in \{2,\ldots,T\},\ q \in \{1,\ldots,Q\}$	(2.13)
$Y_{t-1,q} - Y_{t,q} \le C_{t,q}$	$\forall t \in \{T+1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.6)
$Y_{T+U-1,q} \le C_{T+U,q}$	$\forall q \in \{1,\ldots,Q\}$	(2.7)
$Y_{t,q} \le W_{t,q}$	$\forall t \in \{1,\ldots,T+U-1\}, q \in \{1,\ldots,Q\}$	(2.9)
$D_{\min} \le Z_t \le D_{\max}$	$\forall t \in \{1,\ldots,T\}$	(2.1)
$Y_{t,q}, Z_t, \theta_{t,q}{}^i \in \mathbb{N}^+$	$\forall t \in \{1,\ldots,T\}, q \in \{1,\ldots,Q\}, i \in \{1,\ldots,N\}$	(2.14)

Where

$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_{i,q}$	$\forall t \in \{1,\ldots,T+U-1\}, q \in \{1,\ldots,Q\}$	(2.8)

Table 3 – Consolidated model formulation

2.3.2 Mathematical Properties

The advantage of the consolidated formulation is that the constraint matrix is always totally unimodular, so an integer optimal solution is guaranteed by the LP relaxation with a vertex method. The proof of this claim is presented here.

Proposition 5: *Any version of the consolidated problem has a constraint matrix A that is totally unimodular.*

***Proof*:** Begin by using the same paring steps as in Proposition 1. This process will yield a matrix containing only rows for constraints (2.12) and (2.13). Each of these rows corresponds to a unique ordered pair (t,q). Define $R_{t,q}$ as the row of the pared constraint matrix A associated with ordered pair (t,q). Row $R_{t,q}$ contains a 1 in the column for variable Z_t and a -1 in the column for variable $Y_{t,q}$. Additionally, for $t > 1$, row $R_{t,q}$ also contains a 1 in the column for variable $Y_{t-1,q}$.

Now it will be shown that the constraint matrix is TU by a row partitioning argument. Imagine any arbitrary subset R^* of the set of rows $\{R_{t,q}\}$. Form the members of R^* into open chains according to the recipe shown in Table 4.

1. Set $t = 1$.
2. Find the lowest value of q for which $R_{t,q} \in R^*$ and $R_{t,q}$ has not yet been assigned to a chain, if any exist. If none exist, go to step 4. Otherwise, set $s = t$ and continue to step 3.
3. Notice that if it is also true that $R_{s+1,q} \in R^*$, then rows $R_{s,q}$ and $R_{s+1,q}$ must occupy the same side of any valid row partition of R^*, because the column associated with variable $Y_{s,q}$ contains exactly one 1 and one -1. Thus if $R_{s+1,q} \in R^*$, connect row $R_{s+1,q}$ onto the end of the chain containing row $R_{s,q}$, set $s = s + 1$, and go back to step 3. If $R_{s+1,q} \notin R^*$, then the chain containing $R_{s,q}$ is terminated at this point. Go back to step 2.
4. Set $t = t + 1$. If $t < T$, go to step 2; otherwise, go to step 5.
5. Set all unassigned rows to singleton chains. These will only correspond to rows where $t = T$.

Table 4 – Procedure for partitioning R^* into open chains

Upon completion of this procedure, every row $R_{t,q}$ will be assigned to exactly one chain (possibly a singleton). Every chain represents a set of rows that can be moved across any row partition of R^* without affecting the column sum condition in any column for $\{Y_{t,q}\}$, since all possible conflicts in those columns have been consolidated and rectified in the construction of the chain.

Now, form a new matrix R' that has a row for each chain, and a column for each value of t. Each row (chain) represents only a single value of q. In each row of R', put 1's in those cells corresponding to values of (t,q) that are represented in the chain, and zeros everywhere else. Because links in the chain can only be formed for consecutive values of s, the matrix R' has 1's in consecutive columns only; thus, it is the transpose of an interval matrix. Furthermore, a row partition of R' has column sums that correspond exactly to the column sums for the variables $\{Z_t\}$ in R^*. Because R' is the transpose of an interval matrix, it is totally unimodular. Therefore, it has a row partition that satisfies the column sum requirement to be TU. Two rows of R^* will be considered equivalent if they belong to chains on the same side of this partition of R'. This equivalence relation forms a row partition of R^* that satisfies the column sum condition for all columns. Thus, A is TU. ■

Now that the Base and Consolidated model formulations have been introduced and their mathematical properties analyzed, attention is turned to a slightly modified version of these models. It will concentrate not on finding a system optimal level of delay and cancellation, but rather, will determine the number of slots to make available given these levels of system delay and cancellation as parameters. Hence, it is called the Parametric model. Much of the analysis in the remainder of this thesis will focus on the application of this model, for reasons to be explained.

2.4 Parametric Model

The network underlying the Parametric formulation of this problem is structurally identical to the Base model. The added complication in this model is a pair of side constraints that cap the levels of delay and cancellation, as expected over all capacity scenarios. That is, the solutions presented have certain maximum delay and cancellation levels as parameters.

Unlike the Base and Consolidated models, which present their solutions at system-optimal levels of delay and cancellation, the Parametric model determines the number of slots to offer at any given combination of cancellation and delay parameters. It is important to keep in mind that there are a number of stakeholders in this decision, and they will likely not agree on what "system-optimal" means anyway, so departing from this presumption and providing more user controls might be better anyway. Thus, with the Parametric model, the decision about what levels of delay and cancellation are acceptable is not left to the model. Rather, these decisions are left open to debate by policymakers.

In this section, a fairly brief discussion of the model structure will be provided, along with an analysis of the mathematical properties of the Parametric model. Although this section may be brief, it is important to remember that much of the work has been presented in discussing the first two formulations of the model: the Parametric model merely builds on the work already developed.

2.4.1 Structure

The structure of the underlying network in this formulation is identical to the Base formulation. It is repeated in Figure 8 for clarity.

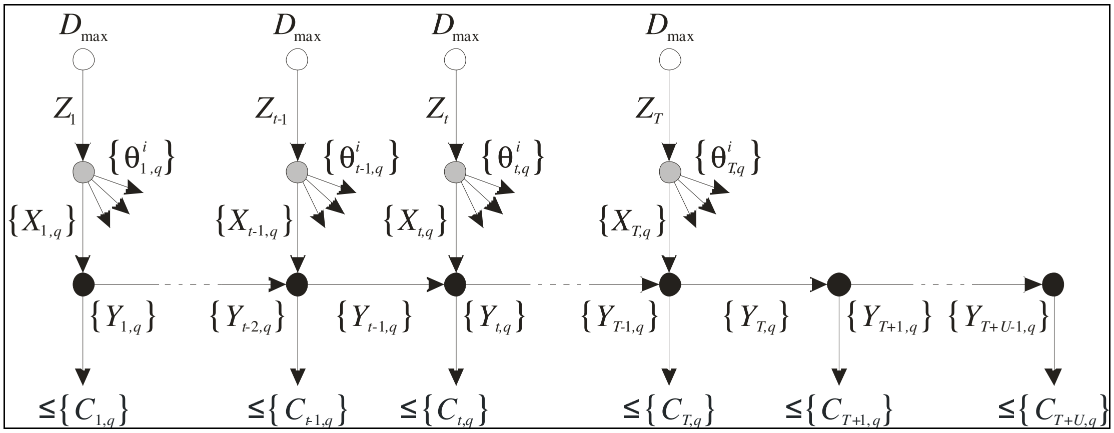

Figure 8 – Parametric model flow diagram

The white nodes are again the source of flights in the network. At the gray nodes, the cancellation decision is made, using the piecewise linear cost function. At the black nodes, the decision between landing immediately and experiencing a delay is made. The $\{Z_t\}$ decision variables represent the number of slots offered in each time period *t*. The flow variables are again $\{\theta^i_{t,q}\}$, $\{X_{t,q}\}$, and $\{Y_{t,q}\}$, representing cancellations, uncanceled flights, and delayed flights, respectively.

The unique features of the Parametric formulation, however, are the side constraints that regulate the levels of delay and cancellation. These create conditional, or parametric, solutions. The expected delay per flight over all capacity scenarios, γ, is calculated as shown in (2.16).

$$\gamma = \sum_q p_q \left(\frac{\sum_t Y_{t,q}}{\sum_t Z_t} \right) \tag{2.16}$$

Because, however, integer solutions are required for this problem, inserting (2.16) into our formulation would likely cause problems with feasibility, depending on the specific value of γ selected. As a result, (2.16) is converted to a linear inequality, and is rearranged as shown in (2.17).

$$\sum_q p_q \sum_t Y_{t,q} - \gamma \sum_t Z_t \le 0 \tag{2.17}$$

In a parallel fashion, the constraint limiting the level of cancellations permitted can be developed. The metric used is the expectation over all scenarios of the percentage of flights cancelled per day, referred to here as ρ. This constraint is shown in (2.18).

$$\sum_q p_q \sum_i \sum_t \theta_{t,q}^i - \rho \sum_t Z_t \le 0 \tag{2.18}$$

As a result of these two constraints, the objective function is simplified by removing the penalties associated with delays and cancellations in units equivalent to slot values. Thus, the goal is to maximize the total economic value of the slots offered. The new objective function becomes simply (2.19). This removes the obvious equivocation in pecuniary terms between slot values and delays and cancellations.

$$\max \left\{ \sum_t V_t Z_t \right\} \tag{2.19}$$

Obviously the constraints shown in (2.17) and (2.18) are inequalities, and as such, are not guaranteed to be binding. However, it should be obvious that allowing some small and finite marginal amount of delay and cancellation would permit an extra slot to be offered. Thus, by using this objective function, the resultant solution should yield delay and cancellation levels as close to those specified as permitted by the integrality constraints.

It is important to note that the previously discussed cancellation cost vector $\left\{\lambda_q^i\right\}$ appear neither in the constraints (2.17) and (2.18), nor in any other portion of the Parametric formulation. Initially, this may seem to be a seriously failing in the model, as it would make the model ambivalent between cancellation and delay. That is, the decision to cancel or to delay one period would be equally costly. While it is true that this assumption does result in each of these decisions being equally costly, the total deleterious effect of delays and cancellations is not of concern in this formulation. Rather, because this is a maximization problem, the model will tend to use as many of the cancellation and delay arcs as are permitted by the side constraints, in order to make more slots available and increase the value of the objective function. Obviously this implies that careful consideration must be paid to the number (N) and capacity (P_i) of delay arcs, the maximum delay length (U), and the maximum number of slots that can be made available in each period (D_{max}).

To complete this mathematical program, constraints (2.1) – (2.10) are retained from the base problem. The entire formulation for the Parametric model is repeated in Table 5 for clarity.

$\max\left\{\sum_t V_t Z_t\right\}$		(2.19)
Subject to		
$\theta_{t,q}{}^{i} \le P_i$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\}$	(2.2)
$Z_t - X_{t,q} - \sum_i \theta_{t,q}{}^{i} = 0$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\}$	(2.3)
$X_{1,q} - Y_{1,q} \le C_{1,q}$	$\forall q \in \{1,\ldots,Q\}$	(2.4)
$X_{t,q} + Y_{t-1,q} - Y_{t,q} \le C_{t,q}$	$\forall t \in \{2,\ldots,T\},\ q \in \{1,\ldots,Q\}$	(2.5)
$Y_{t-1,q} - Y_{t,q} \le C_{t,q}$	$\forall t \in \{T+1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.6)
$Y_{T+U-1,q} \le C_{T+U,q}$	$\forall q \in \{1,\ldots,Q\}$	(2.7)
$Y_{t,q} \le W_{t,q}$	$\forall t \in \{1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.9)
$D_{\min} \le Z_t \le D_{\max}$	$\forall t \in \{1,\ldots,T\}$	(2.1)
$X_{t,q}, Y_{t,q}, Z_t, \theta_{t,q}{}^{i} \in \mathbb{N}^{+}$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\}$	(2.10)
$\sum_q p_q \sum_t Y_{t,q} - \gamma \sum_t Z_t \le 0$		(2.17)
$\sum_q p_q \sum_i \sum_t \theta^{i}_{t,q} - \rho \sum_t Z_t \le 0$		(2.18)
Where		
$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_{i,q}$	$\forall t \in \{1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.8)

Table 5 – Parametric model formulation

2.4.2 Mathematical Properties

With any realistic set of capacity scenarios, (2.17) and (2.18) introduce two row to the constraint matrix *A* that renders it not totally unimodular. Because there is more than one scenario, the probabilities $\{p_q\}$ are themselves not integral, nor are they, in all likelihood, identical. Thus, the delay and cancellation level constraints cannot be scaled to produce coefficients exclusively in the set {-1,0,1}, so there will be submatrix determinants that are also not confined to this set of values.

The matrix is also not *k*-regular in any sense, and thus the results presented for the Base model cannot apply to this model. As a result, the feasible region for this problem is not guaranteed to be bounded by an integral polyhedron, and it is likely the problem will not solve to integral optimality using its linear programming relaxation. Computational results presented later will confirm this conclusion that conventional integer programming techniques such as branch-and-bound are necessary. As with previous problems, provided the size of the problem is manageable and that real-time solutions are not required, this should not be an onerous requirement. Problems of realistic size can be solved in a matter of minutes.

2.5 Formulation Discussions

Three alternative formulations of the problem of determining the number of slots to make available for a market mechanism have been presented. In this section, they will be compared, based on their relative strengths and weaknesses in the areas of prime concern: solution time and economic assumptions required. Various potential modifications to each will be addressed.

2.5.1 Model Solution Times

Having quick solution times may or may not be very important, depending on the further context in which this model is used. If it is to be used once, and with a fixed set of slot values, then solution time is not important, given the problem size. The model would solve in a matter of minutes. If, however, this were to be implemented in a simulation environment, or in an iterative model to account for the price-quantity dynamic of the number of slots offered, then solution time becomes considerably more important.

In terms of solution times, the competition in this category is quite simple. The Consolidated model has a constraint matrix which is totally unimodular for each and every set of input values. The Base model has a TU constraint matrix for limited sets of right hand side values. The Parametric model does not have a TU constraint matrix, or any other properties that assure integrality under any useful conditions, and thus tends to require traditional integer programming techniques to solve to optimality.

That is not to say, however, that a formulation that does not solve to an integer optimal solution using the linear programming relaxation will not solve fairly quickly, as these models do not constitute huge integer programs. In fact, under reasonable assumptions, these problems will have somewhere between several hundred and several thousand variables and constraints. Problems of this size can be solved in reasonable time. Issues relating to problem size and computation time will be further explored in Chapter 5.

2.5.2 Structural Assumptions

All three models suffer from many of the same structural assumptions. Primarily, these relate to the method by which cancellation decisions are made. These assumptions are reasonable, for modeling, given the tremendous complexity of the cancellation decision. In addition, there is a strong, but necessary, assumption in all three models that each and every flight has the same maximum amount of delay it can suffer that does not vary between carriers, or time of day, or by load factor, etc.

The Consolidated model, however, has an addition complication. Because the order of decision-making within the model has been altered, it would now be possible for this model to produce a solution that requires that a given flight be delayed for some number of time slices and subsequently cancelled. One could imagine that this occurs in practice, but in this case the model is not "aware" of the coincidence between the delay and the cancellation. Thus, the proper amount of "hesitance" to do this is not being observed by the model. Of course, the model only works with flows and not individual flights, so such a result is not obvious. But, it is possible that a solution could be produced where the only possible mappings of flows to flights included such aberrations. This could only be strictly prevented by requiring that the marginal cancellation cost be less than the unit delay cost. Obviously this assumption is not realistic, and would lead to excessive numbers of cancelled flights.

2.5.3 Economic Assumptions

The implicit and explicit economic assumptions in these models present several challenges to determining a solution to the problem being studied. Primarily, these challenges relate to the nature of the assumptions about the various costs and values used in the model. Coping with these challenges is certainly not an insurmountable problem, but is one that requires economic expertise and access to data.

The primary economic challenge presented in the Base and Consolidated models is the requirement that all costs be expressed in equivalent units. The modeling assumption was made that a delay lasting one time period t was the unit cost. Cancellation costs and slot values are then be expressed relative to this unit cost. Obviously these values need not be estimated in terms of delays, but rather could be determined in monetary units and normalized. The challenge present in this assumption does not lie in estimating each of these quantities in monetary units, as they each obviously have some monetary value. Rather, the challenge lies in extracting that monetary value in compatible units. Given certain (likely proprietary) data, the values of each of these quantities could certainly be estimated in terms useful to the Base and Consolidated models, at least from the perspective of a single carrier. The advantage to using these models, because all quantities are specified in compatible units, is that the system-optimal levels of delay and cancellation are revealed.

The Parametric model, which is a slightly modified version of the Base model, addresses this challenge by separating the problem of estimating the costs and values in the problem. Because delay and cancellation costs are specified in units independent of the slot values, considerable freedom is gained in estimating each of these relationships, and less proprietary data are required.

2.5.4 Objective Function Assumptions

There are two different objective functions discussed for the three formulations presented in this chapter. The objective function for the Base and Consolidated models is the same, and includes pecuniary equivocations between slot values, delays, and cancelled flights, while that for the Parametric does not.

In the objective function utilized for the Base and Consolidated models, the quantity being maximized is the difference between the total value of the slots offered, and the resultant levels of delay and cancellation. Any choice of cost translation parameters amounts to a value judgement about how society (dis)benefits, and in what relative measures, from slot opportunities, delays, and cancelled flights. Thus, while it is colloquial to call such an objective function a “sytem-opimal” pursuit of net benefit to society, it is important to remember that there would not be unanimity of opinion on the coefficients used to perform this equivocation.

In the Parametric objective function, the total value of all slots being offered is maximized. In this case, the quantity being maximized is the total benefit to the airport operating authority, and presumably to the carriers as a whole (but probably not to any individual carrier). The operating authority is receiving the maximum possible value of its goods, subject to the delay and cancellation cap constraints. In truth, airports are more likely to discuss measures of performance such as passenger enplanements, and load factors are not included in this model. At congested airports, however, it might be safe to assume that there is enough latent demand to fill the flights, and that even if this were not the case, that part of the value for each slot is derived from the passenger demand that would seek to use that slot; therefore an argument can be made that the explicit goal of the airport operating authority is encapsulated implicitly in the slot value functions.

2.5.5 Incorporating Other End-of-day Effects

Several considerations could be made about operations at the airport regarding what takes place as operations cease, relating to the amount of time the airport can stay open after the scheduled operations end, and the actual airport capacity at this time.

Some airport have a strict curfew after the end of the scheduled operations, either taking effect immediately afterwards, or with some short lag. In either of these cases, it would be possible to modify the model formulations to incorporate this effect by changing the assumptions about the maximum delay length parameter U. Because this parameter is assumed constant across the entire day, the delay portion of the model structure is extended past the end of the slot offerings (see Figure 5 for an example). At the end of the day, this maximum delay parameter could be reduced from its nominal value. If it were reduced to zero, then the extra delay arcs past the scheduled day shown in the flow diagrams would be eliminated. Alternatively, at an airport that allowed some flexibility in breaking this end-of-day condition, this parameter could take on a value greater than 1, but less than the nominal maximum delay length value.

In either case, this consideration would require modifying the flow structure of the model to remove some periods past the end of the day. In addition, the quantity W_t, which defines the capacity of the delay arcs would be changed as shown in (2.20), where U^* is defined as the number of time periods beyond the end of the scheduled day during which the airport is still open for arrivals. Potentially this value is zero.

$$W_{t,q} = \sum_{i=t+1}^{\min\{t+U, T+U^*\}} C_{i,q} \tag{2.20}$$

The other consideration regarding the end-of-day conditions is related to the arrival capacity of the airport in the case in which it is open past the end of the scheduled day. In the examples presented in this thesis, the capacity values used for these hours have been those values from the corresponding hours from the capacity scenario analysis. However, it may be more reasonable to consider that the airport, at these late hours, is not able to operate at the full declared capacity (AAR values). For reasons relating more to operations, such as availability of gates, ground crew, and airport staffing, the capacity of the airport to accept these late arrivals may be significantly less than the values specified. As a result, in some scenarios at some airports, it may be reasonable to heuristically define lower capacity values for these unregulated hours.

2.5.6 Conclusions

There are many instances where the computational costs associated with these models are not paramount. In these cases, the economic and structural concerns may be more important drivers of the choice of model formulation. As mentioned, these concerns are primarily derived from the availability of data to calibrate various parameters in compatible units. Because of the data available in conducting this study, the Parametric model will be used for much of the remainder of the analysis undertaken. This will produce solutions at various levels of delay and cancellation that will be specified a priori, and will not reveal the system optimal number of slots and delay and cancellation levels. Analysis using the other two models would produce interesting but likely not significantly different results, as compared to the Parametric model.

The calibration procedure presented next in Chapter 3 is applicable to all three formulations of this problem, and does not rely on any of the aforementioned formulations in the results presented. A slightly modified version of the Base model will be applied for calibration. The Base and Consolidated models make use of all the parameters estimated. The Parametric model uses the estimates of the structural parameters, but does not require the cancellation cost vector, for reasons described previously. The results presented in Chapters 4 and 5 will rely on the use of the

Parametric model. Consider that portion of the analysis a case study in applying the Parametric model to a specific airport, as sufficient data are not presently available to apply the other two models under consideration.

Chapter 3: Model Calibration

In this section, the procedures are described by which most of the parameters used in the various models described in Chapter 2 are calibrated. First, the modeling framework used for calibration with historical data will be described. Then, the assumptions about the time resolution of the modeling presented in this thesis will be discussed in further detail. Next, the parameters relating to flight delay (maximum permissible length of delay U) and cancellation (number of discrete cancellation arcs N, capacity of each cancellation arc P_i, cost of each cancellation arc λ_i) will be described, with particular attention paid to the assumptions and procedure used for the cancellation parameters. The discussion of the slot value parameters will be reserved for discussion in the following chapter.

3.1 Parameter Calibration Model

In this section, the process by which the model parameters are calibrated is discussed. The model described is structurally similar to those in Chapter 2, but it does not serve the same purposes. That is, it cannot be used to estimate the number of slots that should be made available to a market mechanism. The presentation of the Calibration model in this section is structured in the same fashion as the main models (Base, Consolidated, and Parametric) in this paper: discussions of the structure of the model, and of the mathematical properties relating to finding integer solutions.

3.1.1 Structure

This calibration process is cast as a minimum cost network flow model, where the objects flowing through the network are the flights themselves, as before. The objective of this model is to minimize the cost of the delays and cancellations experienced, presuming that, all other things being equal, carriers would prefer to suffer as little as possible from these deleterious effects. The network structure of the model is nearly identical to the Base model, as shown in Figure 9, absent the stratification due to capacity scenarios.

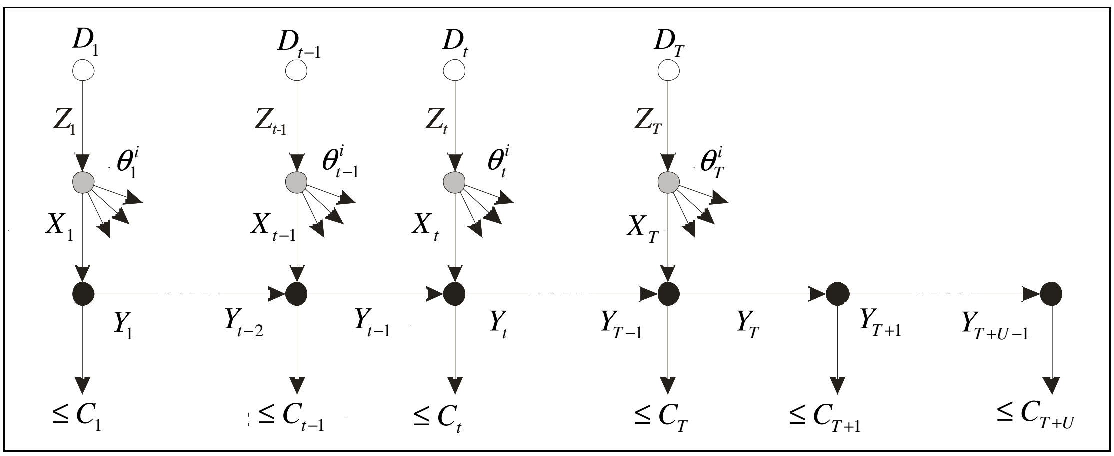

Figure 9 – Parameter calibration model flow diagram

The demand source nodes reflect the true number of scheduled flights on the day being considered, and the capacity sink nodes are the declared capacity of the airport. In other words, the demands and capacities on any given historical day are known, so the goal of the model is not to predict, but rather to help choose a combination of parameter values that make the cancellation and delay results from the model match most closely with historical results.

As before, the demand is shown at the top-most level of the diagram at the white nodes. The amount is denoted as D_t, but is not a bound as was previously used.

Rather, the entire supply must flow through the network and exit either as a cancellation, or as a landing. Thus, in this model, the variable Z_t is retained in the network flow diagram shown in Figure 9, but is not needed in the formulation. The notation is retained to demonstrate the similarity in structure between this calibration model and the Base model.

At the gray nodes, various flights are cancelled. As before, cancellation costs are stratified, to model the increasing marginal cost when canceling many flights. These various costs are denoted $\{\lambda_i\}$, with N the number of cancellation arcs in each time period, and $i \in \{1,\ldots,N\}$ again referring to the cancellation arc being used. Each cancellation arc i is subject to a capacity P_i. This leads to the constraint shown in (3.1).

$$\theta_t^i \le P_i \quad \forall t \in \{1,\ldots,T\},\ i \in \{1,\ldots,N\} \tag{3.1}$$

The flights that are not canceled and thus permitted to land at some time are again denoted as X_t. To maintain flow conservation at the gray nodes, the constraint shown in (3.2) is needed.

$$D_t - X_t - \sum_i \theta_t^i = 0 \quad \forall t \in \{1,\ldots,T\} \tag{3.2}$$

At the black node, the flights X_t may either land immediately, or may be delayed, and permitted to land at some later time. The flights that are delayed to be considered in the next time period are denoted Y_t. Obviously the number of flights permitted to land in time period t must be less than, or equal to, C_t. These observations lead to the constraints shown in (3.3)-(3.5). Each differs slightly, as before, because of the boundary conditions. Recall that flights can be delayed at most U units of time past the time at which demand ends.

$$X_1 - Y_1 \le C_1 \tag{3.3}$$
$$X_t + Y_{t-1} - Y_t \le C_t \quad \forall t \in \{2,\ldots,T\} \tag{3.4}$$
$$Y_{t-1} - Y_t \le C_t \quad \forall t \in \{T+1,\ldots,T+U-1\} \tag{3.5}$$

Because no flights can be delayed more than U time periods, and the demand ends at time period T, all flights permitted to land must do so by the very end of the day, as shown in (3.6).

$$Y_{T+U-1} \le C_{T+U} \tag{3.6}$$

The delay arc capacities are calculated as previously, with the removal of the q subscripts, as shown in (3.7). They are then enforced in a very straightforward manner, as shown in (3.8).

$$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_i \quad \forall t \in \{1,\ldots,T+U-1\} \tag{3.7}$$
$$Y_t \le W_t \quad \forall t \in \{1,\ldots,T+U-1\} \tag{3.8}$$

Because of the network-flow formulation of the model, integer-valued solutions are guaranteed by using linear programming, thus while non-negativity should be stated explicitly, integrality need not be, yielding (3.9).

$$X_t, Y_t, \theta_t^i \ge 0 \quad \forall t \in \{1,\ldots,T\}, i \in \{1,\ldots,N\} \tag{3.9}$$

A proof that the constraint matrix is totally unimodular, and hence that LP relaxation produces integer optimal solutions, is offered in Section 3.1.2 of this thesis.

As mentioned above, the objective of this problem is to minimize the total cost experienced by all flights operating at the airport, as shown in (3.10). The constraints for this model are repeated in Table 6 for clarity.

$\min\left\{\sum_{t=1}^{T+U-1} Y_t + \sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i\theta_t^i\right\}$		(3.10)
Subject to:		
$D_t - X_t - \sum_i \theta_t^i = 0$	$\forall t \in \{1,\ldots,T\}$	(3.2)
$X_1 - Y_1 \le C_1$		(3.3)
$X_t + Y_{t-1} - Y_t \le C_t$	$\forall t \in \{2,\ldots,T\}$	(3.4)
$Y_{t-1} - Y_t \le C_t$	$\forall t \in \{T+1,\ldots,T+U-1\}$	(3.5)
$Y_{T+U-1} \le C_{T+U}$		(3.6)
$Y_t \le W_t$	$\forall t \in \{1,\ldots,T+U-1\}$	(3.8)
$\theta_t^i \le P_i$	$\forall t \in \{1,\ldots,T\},\ i \in \{1,\ldots,N\}$	(3.1)
$X_t, Y_t, \theta_t^i \ge 0$	$\forall t \in \{1,\ldots,T\}, i \in \{1,\ldots,N\}$	(3.9)
Where		
$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_i$	$\forall t \in \{1,\ldots,T+U-1\}$	(3.7)

Table 6 – Calibration model formulation

3.1.2 Mathematical Properties

Given the calibration procedure being undertaken, it is highly desirable that this model be solvable rapidly. One way to guarantee reasonable solution times is to prove that the linear programming relaxation will yield integer solutions by demonstrating that the constraint matrix defining the problem in totally unimodular.

A special case of the row partition argument that was used previously in this thesis to demonstrate that a matrix is TU can be invoked when a column has at most two non-zero elements. In previous formulations, the presence of the $\{Z_t\}$ columns prevented this method of proof. For the calibration model, however, we can exploit it as follows. Per Proposition 3.2 in Nemhauser and Wolsey (1988), one possible set of sufficient conditions for a matrix A to be totally unimodular is shown in Table 7

1. Each element a_{ij} is in the set {-1,0,1}
2. Each column contains, at most, two nonzero elements.
3. There exists some partition of the rows such that the column sums (only of those columns with two nonzero elements) of each partition are equal.

Table 7 – Conditions for total unimodularity

The first step in evaluating the total unimodularity is to convert the equality constraints in this formulation into inequalities. As described in the discussion of the mathematical properties of the Consolidated model formulation, treating equalities as inequalities has no effect on the TU properties of the constraint matrix. Then, examining the formulation of the Calibration model presented in Table 6, it is obvious that the first condition is met. The second and third conditions are demonstrated to be met simultaneously. First, remove from consideration columns containing Z_t and θ_t^i, as they flow to or from source or sink nodes, and thus appear in only one row. The remainder of the variables must appear in exactly two rows, to satisfy mass balance considerations. Because of mass balances, these two coefficients must always be a pair {1,-1}, representing the flow out of one node and the flow into another node. An obvious partition of the rows would be to include all rows in one set, and none in the other. Thus, by the previous statement, each column in the non-empty partition must sum to 0, while those in the empty partition necessarily sum to 0. ■

3.2 Time Resolution

While any arbitrary division of time is structurally permissible in the Calibration model, very few divisions are truly feasible, due to several practical concerns.

First, it is necessary that each time period be of equal length. First, this makes the problem logically tractable. Second, the values of $\{\lambda_i\}$ are relative to a single-period delay cost. As such, if the time periods were of different length, the costs for using the delay arcs would have to be scaled to be relative to the length of the time period under consideration. This would add an unnecessary complication. The third concern is the nature of the historical data available for calibrating the parameters. Although individual flight records could be grouped into any arbitrary division of time, more accessible sources, including the FAA's Aviation System Performance Metrics (ASPM) data collection, have data available in both hourly and quarter-hourly bins.

It is also important to consider the structure of the Chapter 2 models in choosing the time resolution used in the Calibration model. The requirement relating the time scales of the two models is that the parameters estimated in the calibration procedure be scalable to the time resolution used in the Parametric model. As such, it seems reasonable to use the briefest time periods possible in the Parametric model, in order to provide the greatest flexibility in choosing the time scale of the Parametric model.

Given these requirements, and the available sources of data, a reasonable time scale to use in calibrating the model is quarter-hours, and this is the approach taken in this thesis. This choice will permit the time scale of the Parametric model used in this thesis to be any multiple of a quarter-hour. If hourly time periods are chosen for the analysis, then the maximum delay length parameter U set in the calibration procedure must be some multiple of four quarter-hours, as flights can only be delayed an integer number of time periods.

3.3 Calibration Procedure

Because the structure of this model is explicitly dependent on the choice of several of the parameters being calibrated, it does not lend itself to a traditional optimization routine to select the optimal parameter set. It would be best if U, N, P, and λ could all be incorporated as decision variables into some structured mathematical program, but this is just not feasible because several of these parameters partially define the structure of the model. In particular, U determines the number of time slices at the end of the day, and N determines the number of cancellation arcs to include, as illustrated in Figure 9.

As a result, the calibration procedure undertaken is more heuristic in nature. The basic premise of the process is to run the model using a test data set and a large number of possible parameter combinations, and then to select the best combination based on several metrics. By combining manual, but careful, evaluation of incremental results and bounds on the parameter space, this heuristic is able to produce a strong fit between the observed and predicted data.

3.3.1 Parameter Bounding

The first step of this heuristic search is to identify appropriate bounds for the parameters. The first constraint to consider is that three of the four parameters being calibrated are necessarily restricted to integer values. Because the maximum delay length U and number of cancellation arcs N are structural parameters, they must be integer valued. In addition, the capacity P_i of each cancellation arc $i \in \{1, \ldots, N\}$ should logically be integer valued, as an integer number of units are flowing through the network. The cancellation arc costs $\{\lambda_i\}$, however, need not necessarily be integer valued. In this analysis, they will be restricted to conform to some pre-specified lattice of points $\{\zeta_i\}$, as shown in (3.11). The parameter α will be calibrated in this analysis, subject to some assumption about the structure of the vector $\{\zeta_i\}$.

$$\{\lambda_i\} = \alpha\{\zeta_i\} \tag{3.11}$$

Considering the data being used in the calibration procedure, the bounds on the variables set forth in Table 8 were established. Obviously these bounds are fairly arbitrary, but a common sense evaluation of the underlying data justifies them.

Parameter	Lower Bound	Upper Bound	Additional Constraints
U	4	20	Multiple of 4
N	1	5	
P	2	6	
α	1	10	

Table 8 – Parameter bounds

It is likely that some combinations of the parameter values in the ranges above will not produce feasible linear programs. This will likely arise from scenarios which combine low numbers of cancellations arcs N and cancellation arc capacities P_i. In the worst (most limited) capacity scenarios, there will not be sufficient capacity at the demand nodes (arrivals and cancellations) to accommodate all of the supply. As there are no other release points from the network, the problem will have no feasible solutions.

There are obviously an infinite number of combinations of values to uniquely define the vector $\{\zeta_i\}$. As mentioned earlier, the premise of the various cancellation arc costs was that as the number of cancellations increases, the marginal cost of cancellation should also increase. The vector $\{\zeta_i\}$ should thus be defined to satisfy the conditions shown in (3.12) (initial condition), (3.13) (increasing sequence), and (3.14) (increasing marginal cost).

$$\zeta_1 = 0 \tag{3.12}$$

$$\zeta_i < \zeta_{i+1} \quad \forall i \geq 1 \tag{3.13}$$

$$(\zeta_i - \zeta_{i-1}) < (\zeta_{i+1} - \zeta_i) \quad \forall i \geq 2 \tag{3.14}$$

Given knowledge of the problem structure and the assumptions presented thus far, a form can be assumed for $\{\zeta_i\}$, as defined in (3.15). The difference between each successive pair of numbers is one unit greater than the difference between the previous pair. This sequence of integers is called the central polygon numbers, or the "lazy caterer's sequence." Obviously, in any given test run, only the first N elements of this sequence will be used.

$$\zeta = \{1, 2, 4, 7, 11, \ldots\} \tag{3.15}$$

3.3.2 Calibration Metrics

For a given combination of parameters values, the calibration model is run for each day in the calibration data set. The most interesting outputs from the calibration model are the vectors describing the number of cancellations taking place at each cancellation cost level $\{\theta_t^i\}$. Obviously, the real data to which the model is being calibrated do not specify cancellations at the artificial cost levels output by the model, so for each time period t, the cancellations on each arc must be summed to produce a vector $\{\psi_t\}$ in which each element describes the total number of cancellations in that time period, as shown in (3.16).

$$\psi_t = \sum_{i=1}^{N} \theta_t^i \tag{3.16}$$

Now, the real cancellation data Q_t can be compared with the predicted cancellation data ψ_t. Several metrics provide interesting information about the relationship between these two quantities. These metrics will be grouped into two categories: trend and profile. The trend metrics are of prime interest in this analysis, but the profile metrics could be highly applicable in other situations.

The trend metrics explain the strength of the relationship between the daily cancellation levels observed and predicted. Recall that the results discussed thus far reflect some division of each day. That is, an additional index could be added to Q_t and ψ_t indicating the day drawn from the calibration data set. For the trend metrics, the observed cancellation percentage is averaged across each day, as is the model output ψ_t. These operations produce two new vectors describing the average observed and predicted cancellation levels for each day in the calibration data set, as in (3.17) and (3.18).

$$Q^d = \sum_{t=1}^{T} Q_t^d \tag{3.17}$$

$$\psi^d = \sum_{t=1}^{T} \psi_t^d \tag{3.18}$$

The trend metrics are then derived from the relationship between the vectors $\{Q^d\}$ and $\{\psi^d\}$. The relationship is examined using ordinary least squares (OLS) regression with the observed daily cancellation percentage Q^d as the dependent variable, and predicted daily cancellation percentage ψ^d as the independent variable. This relationship is represented in (3.19), with ε^d representing the disturbance term in this regression equation.

$$Q^d = \gamma_0 + \gamma_1 \psi^d + \varepsilon^d \tag{3.19}$$

Considerable information may be gleaned from the regression model estimated in (3.19). First, the resultant coefficient of determination R^2 explains the amount of variance in the observed data that is being explained by the predicted data. Having a higher R^2 value with one set of parameters indicates that the Calibration model is better able to explain the variance in the observed cancellation data than it can with another set of parameters. In some sense, this is a measure of how well the model is predicting cancellations.

In addition, the coefficients γ_0 and γ_1 in (3.19) can be useful in determining the quality of a given set of calibration parameters. The intercept value γ_0 indicates what percentage of cancellations are taking place for reasons other than the congestion explicitly accounted for in the Calibration model. Empirical examination of historical cancellation data at any airport will reveal this effect, as carriers cancel flights for myriad other reasons, even on good weather days, including system connectivity, crew availability, weather elsewhere, and maintenance issues. This effect is further discussed in Ball et al. (2006) and Mukherjee et al. (2007).

The coefficient γ_1 represents the slope of the line describing the relationship between the predicted and observed cancellation data. For a proper set of parameters, this value should approach 1.0, indicating that input data that yield a prediction of a single additional cancellation will have a corresponding historical data point in which the number of cancellations increased by one. The relationship between the two quantities should not only be linear, but with a unit slope.

Another trend metric that will be examined for the daily-level predicted cancellations is the percentage of days in the calibration data set on which the model predicts no cancellations. Examination of the output of the outputs produced by the calibration procedure suggests that this may be an issue that would otherwise not be apparent from the other trend metrics, as such a scenario could yield a regression model with reasonable coefficients and an acceptable correlation. Examination of the scatter plot would clearly suggest that such a parameter set produces unacceptable results.

The other category of metrics mentioned previously was the profile metrics. These metrics are not used in the analysis described herein, but are discussed for completeness. A more extensive examination of the applications of such metrics to aviation data, particularly delays and cancellations, can be found in Ball et al. (2006).

This class of metrics for calibrating a cancellation prediction model was developed to examine the temporal distribution of cancellations over the course of a day. The question being examined is: how well does the time profile of the predicted cancellations throughout the day match that of the observed cancellations? Obviously this is a difficult question to answer for a single day, given the large number of other potentially influential factors. However, if data are averaged across a sufficiently large period of time, the profile of observed and predicted cancellations for the "average" day should match sufficiently well.

Once these data were aggregated, a simple numerical metric was used to quantify the similarity between the observed and predicted cancellation profiles in each period and, more important, to estimate the single vertical offset (or translation) that provides the best superimposition of the two profiles for each period. "Best" in this case is defined as the superimposition that produces the least sum of squared differences between the vertices (i.e., the hourly values of average delay) of these two piecewise-linear profiles. The associated offset can also be thought of as a rough indicator of the average amount of background (non local congestion based) cancellations taking place for the various causes mentioned earlier.

3.3.3 Calibration Data

Data for this calibration procedure were drawn from the Analysis section of FAA's ASPM system. Quarter-hourly observations for various airports were used. The data for the demand in the Calibration model were taken from the "Scheduled

Arrivals" (`SCHARR`) field, denoted here as D_t. These data are taken directly from the airport schedules published in the Official Airline Guide (OAG). The capacity data were taken from the "Airport Arrival Rate" (`AAR`) field, denoted here as C_t. These data are declared by the airport and reported in, among other places, the ASPM system. The historical cancellation data were taken from the field "Cancelled Arrivals" (`CANCARR`), denoted here as Q_t. As mentioned elsewhere, these data could have been derived independently, but the simplicity and accessibility of the ASPM system make such an analysis unnecessary.

3.4 Calibration Results

The primary driver of this calibration procedure is matching the predicted cancellations with the observed cancellations. As described, metrics derived from this relationship are those that determine which set of parameters is best. As such, the delay parameter being calibrated (maximum permissible delay length) is estimated as a part of calibrating the models for cancellation performance.

To demonstrate the calibration process and its utility for varied airports, data for several airports will be examined. The informal hypothesis being examined by using data for several airports is whether the calibration procedure will yield results for each airport consistent with outside knowledge of the operations and procedures in effect at that airport. The airports that will be examined are New York's LaGuardia Airport (LGA), Chicago O'Hare International Airport (ORD), and Hartsfield-Jackson Atlanta International Airport (ATL). The datasets used in this test are summarized in Table 9. The missing observations occur almost exclusively during the overnight hours between 0200-0400 local time, during which traffic levels are negligible. Thus, these missing data are not of concern to this calibration effort.

Airport	Year	Number of Days	Number of Observations	Number of Missing Observations
ORD	2005	365	34634	406
ATL	2005	365	33985	1055
LGA	2005	365	28157	6883

Table 9 – Datasets used in calibration procedure

3.4.1 Results for ORD

The ASPM data for Chicago O'Hare International Airport in 2005 were tested using the calibration process previously described. The five best parameter sets, as determined by highest R^2 value are presented in Table 10, while the five best, as measured by nearness of regression slope γ_1 value to 1.0 are shown in Table 11.

Parameters			Metrics			
U	λ	P_i	R^2	**Percent predicted=0**	**Slope** (γ_1)	**Intercept** (γ_0)
20	4	3	0.821	85.7%	2.281	0.017
20	4	4	0.810	85.7%	1.899	0.018
20	6	3	0.808	88.7%	2.326	0.018
20	[4 8]	2	0.806	85.7%	1.926	0.018
16	4	5	0.799	85.7%	1.677	0.018

Table 10 – Best (by R^2 value) potential parameter sets for ORD

Parameters			Metrics			
U	λ	P_i	R^2	**Percent predicted=0**	**Slope** (γ_1)	**Intercept** (γ_0)
4	[4 8]	6	0.774	85.7%	1.320	0.019
4	[4 8 16]	4	0.774	85.7%	1.356	0.019
4	[4 8 16]	5	0.774	85.7%	1.337	0.019
4	[4 8 16]	6	0.774	85.7%	1.322	0.019
4	[4 8 16 28]	3	0.774	85.7%	1.396	0.019

Table 11 – Best (by slope value) potential parameter sets for ORD

These results seem fairly unsuitable, based upon the percentage of days that were predicted to have no cancellations. Unfortunately, this problem seems to be evident, to a greater or lesser degree, in all of the scenarios tested for this airport. Fortunately, for the analysis presented in the remainder of this thesis using the Parametric model, some of these values (λ, in particular) are not needed. The remaining values can be estimated by other heuristic approaches. In a study in which the Base or Consolidated model were to be applied, great care would obviously have to be taken to ensure better results for this analysis. This analysis is, however, not without merit. In Mukherjee et al. (2007) a very similar process was used to calibrate these models with better results.

A scatter plot of the observed and predicted daily cancellation percentages for the scenario $U = 4$, $P_i = 6$, $\lambda = [4\ 8]$ is shown in Figure 10. The regression line for this data is also shown. As can be seen, there is a strong correlation between the observed and predicted data. Unfortunately, much of the data lies on the y-axis. In addition, the y-intercept is present at a level consistent with the background cancellation rate.

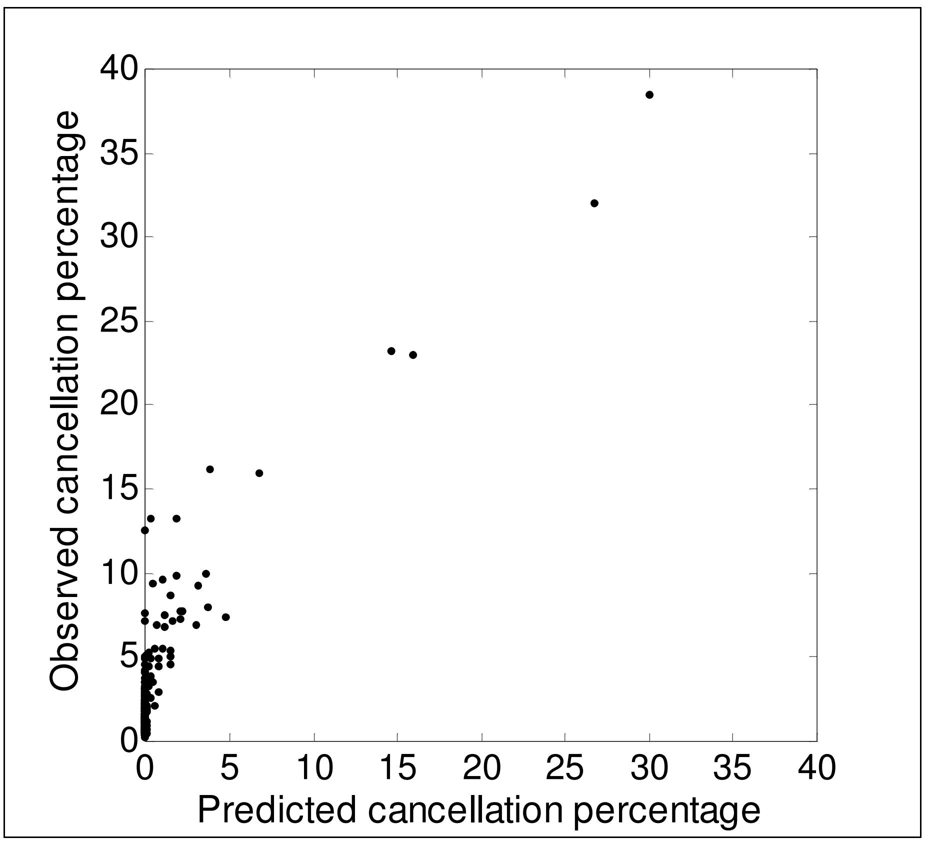

Figure 10 – Sample scatter plot for ORD

3.4.2 Results for ATL

The ASPM data for Hartsfield-Jackson Atlanta International Airport in 2005 was tested using the calibration process previously described. The five best parameter sets, as determined by highest R^2 value are presented in Table 12, while the five best, as measured by nearness of regression slope γ_1 value to 1.0 are shown in Table 13. The results are again fairly unacceptable.

Parameters			Metrics			
U	λ	P_i	R^2	**Percent predicted=0**	**Slope** (γ_1)	**Intercept** (γ_0)
4	4	6	0.408	46.4%	2.581	0.017
4	[4 8]	6	0.408	46.4%	2.581	0.017
4	[4 8 16]	6	0.408	46.4%	2.581	0.017
4	[4 8 16 28]	6	0.408	46.4%	2.581	0.017
4	[4 8 16 28 44]	6	0.408	46.4%	2.581	0.017

Table 12 – Best (by R^2 value) potential parameter sets for ATL

Parameters			Metrics			
U	λ	P_i	R^2	**Percent predicted=0**	**Slope** (γ_1)	**Intercept** (γ_0)
4	4	3	0.388	46.4%	2.6791	0.017
4	4	4	0.390	46.4%	2.5833	0.017
4	4	5	0.399	46.4%	2.5689	0.017
4	4	6	0.408	46.4%	2.5809	0.017
4	[4 8]	2	0.375	46.4%	2.6493	0.018

Table 13 – Best (by slope value) potential parameter sets for ATL

3.4.3 Results for LGA

The Calibration model was test for the ranges of parameters described in Table 8, and the cancellation cost structure vector in (3.15). Unfortunately, initial results suggested that the model was unsuitable for this task, as it could not produce fits with an R^2 value greater than 0.1. Clearly, a stronger correlation is necessary, and has been achieved for other airports, as described in Mukherjee et al. (2007).

There were three possibilities for the source of this problem. First, it could have come from a poorly chosen parameter set. This effect can be discounted, as tests exceeding even the previously specified parameter ranges suggested that the model was overpredicting. Second, the demand being considered could be greater than the true demand. This can also be discounted, as the demand being considered is simply the schedule, as recorded in the Official Airline Guide. Errors in this data seem highly improbable. The final source of error could be in the capacity values being used. As mentioned previously and in Churchill et al. (2006), the capacity values at LGA may not be reflective of the capacity of the airport as it used. That is, the number of actual operations conducted frequently, and for extended periods of time, exceeded the declared Airport Acceptance Rate.

To account for this capacity information problem, a simple capacity-inference heuristic was developed, using data available in the ASPM system. The extra data used was the field reporting "Arrivals for Metric Computation" (`METRICARR`), denoted here as M_t. This is the actual count of arrivals that took place during the time period under consideration. Thus, if an airport is truly highly congested and operating at capacity, these data could be used to infer the arrival capacity of the airport. The heuristic described in Table 14 was developed to infer airport capacity.

1. Set $t = 1$.
2. If $M_t > C_t$, then set $C_t^{\bullet} = M_t$. Else, set $C_t^{\bullet} = C_t$. Set $t = t + 1$.
3. If $t < T$, go to step 2. Otherwise, go to step 4.
4. Set $C = C^{\bullet}$

Table 14 – Procedure for heuristically correcting declared capacities

The model was tested for every combination of parameters as bounded in Table 8 using capacity values corrected by the heuristic in Table 14. These results are summarized in Table 15 and Table 16. The first of these shows the five best combinations, as determined by the highest R^2 values. Similarly, the second shows the five best combinations, as determined by the nearness of the slope value to 1.0.

Parameters			Metrics			
U	λ	P_i	R^2	Percent predicted=0	Slope (γ_1)	Intercept (γ_0)
12	1	2	0.661	38.7%	2.500	0.015
16	1	2	0.661	38.7%	2.500	0.015
20	1	2	0.661	38.7%	2.500	0.015
8	1	3	0.647	38.7%	2.166	0.017
12	1	3	0.647	38.7%	2.166	0.017

Table 15 – Best (by R^2 value) potential parameter sets for LGA

Parameters			Metrics			
U	λ	P_i	R^2	Percent predicted=0	Slope (γ_1)	Intercept (γ_0)
4	1	4	0.634	38.7%	2.029	0.017
4	1	5	0.635	38.7%	2.016	0.017
4	1	6	0.634	38.7%	2.011	0.017
4	[1 2]	2	0.622	38.7%	2.061	0.018
4	[1 2]	3	0.631	38.7%	2.035	0.017

Table 16 – Best (by slope value) potential parameter sets for LGA

The results are again disappointing, for reasons unknown. In the remainder of this thesis, whenever a value is needed for a numerical example, it will be selected heuristically, based on judgment and values commonly accepted in the literature regarding the tradeoffs between delay and cancellation.

Now that the method by which model parameters are to be estimated has been discussed, the attention of this thesis will be turned towards the slot valuations. This is a difficult and multi-faceted issue with which to cope. The next chapter will concentrate and listing and discussing various methods by which the slot valuations might be estimated. An extensive example will be given using the technique deemed most practical.

Chapter 4: Slot Valuation

This section discusses several methods by which the value of an airport arrival slot may be inferred or estimated. These slot valuations are critical to the modeling effort undertaken in this thesis. Unfortunately, they can also be very challenging to estimate. Given the framework used in this thesis, these slot valuations must represent average values across all carriers seeking to utilize the airport, recognizing that this results in a loss of detail when compared to individual carrier values.

The methods suggested here do not comprise the totality of the various methods by which slot values could be estimated. Rather, they are suggestions for several feasible methods. In any application of the models described in this thesis, an extensive economic analysis would have to be undertaken, using all available data specific to the airport under consideration, to ascertain these, or similar, values. This analysis would comprise the major portion of the work at any airport at which these models were implemented.

The first part of the presentation describes several necessary assumptions about the slot values. Then, several possible sources of slot values will be described in varying detail, including obtaining information from carriers, inferring information from ticket price data, and making use of the results of the initial rounds of a slot auction. The third method will be described in greater detail, as it is deemed most practical to implement in an auction setting. Computational experiments were conducted using the final technique, and will be discussed as an illustrative example of this method.

4.1 Necessary Assumptions

In this section, several claims are made regarding assumptions that are either strictly necessary, or at the least, very helpful, regarding the slot values used in these models.

The first of these assumptions is that the value of slots is time-variant. That is, slots during different time periods of the day are more valuable than those in other time periods. Without such an assumption, the framework presented in this thesis would likely be inappropriate, and ought to be replaced by an analysis making extensive use of queueing theory. Obviously then, this is a reasonable assumption for the models presented herein. This assumption can be confirmed empirically by myriad different methods. First, simply examining the published schedules of carriers at any moderately busy airport should suggest that certain times of day have more flights scheduled. Conversations with any scheduled carrier would also confirm this assumption. In addition, it is useful to consider when many travelers generally need or want to travel. There are periods during the day during which it is more beneficial to travel, in order to best make use of time. No assumptions are made about the nature of this time-variance, other than that it exists, and has a measurable effect on the results of these models.

Another assumption is that each carrier has similar preferences for slots in a given time period. Carrier slot valuation may differ for many reasons, including operating aircraft of different sizes, and for different purposes (e.g. shuttle service vs. leisure-oriented service). These differing motivations could clearly yield different value judgments. The models presented in this thesis assume a single value for a slot in any given time slice. This can be interpreted to mean that carriers have the same value functions, or, preferably, that enough slots are available and enough carriers to spread them amongst in each time slice that these numbers could represent average valuations across carriers without concern for differences between carriers. This assumption is reasonable because it is common for market mechanisms such as this to limit the percentage of the total resource owned by any individual entity (and in this case, in any individual time slice) to some maximum number, thereby assuring that monopolistic forces cannot be invoked. It is assumed, therefore, that the slot values described in this section attempt to capture the mean values across all carriers, and that the values of each carrier do not differ significantly from their averages. It is important to recall that the quantity of interest in the Parametric model is not the absolute slot value for each time period, but rather the relative slot values across all periods. These assumptions thus seem more reasonable when applied to the Parametric model.

In addition, each slot in a given hour is assumed to be identical in terms of its utility for operating a flight. No explicit considerations are made regarding aircraft size, and the resultant separation and runway occupancy time requirements. In the analysis presented in this thesis, it is assumed that such assumptions have already been incorporated into the capacity values specified for each time period. This assumption is aided by the fact that the capacity values commonly used for such analysis are the Airport Arrival Rates, as specified by the FAA. These values are not specified conditionally on the fleet mix being used at the airport under consideration, but rather include these characteristics implicitly.

4.2 Revelation of Proprietary Information

Those entities owning airport arrival slots (primarily airlines) necessarily have some idea of the value of these resources. This information is useful to them in utilizing their resources, setting fares, and participating in the secondary slot market. Airlines are understandably quite controlling of proprietary financial information in this competitive industry. However, if someone were able to extract some of this information from one or more slot owners, then these values could be used to drive the models determining how many slots to make available. For the Base and Consolidated models, these values would necessarily have to be monetized in comparable units. This would obviously present significant challenges, depending on the number of entities willing to share their valuations. For the Parametric model, only the relative values of these quantities would be of interest.

The situation that is potentially most valuable is one in which a single entity owns many slots across the entire day. In this case, concerns about monetizing the values are negated. In addition, if the Parametric formulation of the model is being applied, then the absolute values of these slots is not of paramount import. Rather, the relative value of these slots is the quantity of interest. This case seems most likely to see any modicum of success in garnering data from the airlines. The ideal case would be to find a carrier that purchased a large number of slots at some time in the past, and to convince them to provide the modeler with just the relative values of these slots. The carrier might be less reluctant to part with such data, as it would not precisely reveal such delicate financial information.

In addition, if the airline were sufficiently well motivated, it seems likely that they would be able to provide information about the tradeoff between the number of slots and the value of offering a flight in a given time period. Again, this is likely information that they would have some grasp on, as it likely helps to drive pricing decisions.

4.3 Inference from Ticket Prices

Another method by which slot values could be inferred is to make use of information about actual itineraries purchased, or available for purchase, by revenue passengers. One source of such historical data is the US Department of Transportation's Airline Origin and Destination Survey (DB1a). This database provides a 10% random sample of itineraries flown in the US. While this database provides excellent information about itineraries flown, including the market and the total paid, it does not provide any flight-specific information that could aid in determining what time of day the flight operated (i.e. what slot was used). Fortunately, similar datasets, such as the Market Information Data Transfer (MIDT) database, are available for purchase that provide information about routes flown, time of day, and fare information for a sample of passengers. Unfortunately, such data are prohibitively expensive for the purposes of writing a thesis, but might be exploited in a real application of these methods.

If one possessed the resources necessary to purchase such information, several methods could be employed that would enrich the slot valuation analysis. Of particular utility would be the scenario in which the airport under consideration is involved in a shuttle service with another airport. If that were the case, one could analyze the ticket data available for itineraries flown on this shuttle route. Given the time of day of a flight (i.e. slot(s) used), and information about the fares paid to fly during that time, trends could be identified that suggested certain times of day as being able to produce more revenue, i.e. having a higher value to the airline. Such an analysis implicitly assumes that the revenue derived from a flight is directly proportional to the value of the slot used to earn said revenue. Myriad issues could confound such an analysis, such as data quality, data sampling, how far before travel a ticket was purchased, yield management on the part of the carrier, and random variations in demand for particular flights. Despite these potential pitfalls, it seems reasonable that such an analysis could provide some useful information about the time-variance of the slot valuations.

Another piece of information that one could attempt to infer from such ticket price sample data is the nature of the relationship between the quantity of slots offered and their values. Such information could be inferred by examining the same market (city-pair) operated by a given airline on several days on which the number of flights operated in the market differed. Given the assumptions that the flights were operated at similar cost to the airline, and that the revenue derived from a flight is directly proportional to the value of the slot used to earn said revenue, one could estimate a relationship between the number of slots used, and their values. Obviously it would be necessary to normalize this data using the time-variance information estimated previously. This analysis could be repeated for several markets to estimate an average parameter for the airport under consideration representing this tradeoff. Given this parameter and the value of each slot estimated previously, a decreasing function could be used to represent this price-quantity relationship. The application of such a function (in particular, a negative exponential function) will be discussed in Section 5.1.2.

An alternative to acquiring the expensive data discussed in this section would be to use data mining techniques to extract similar data from fares currently offered for sale for future flights. The upside of such data is that they are readily available at no cost over the Internet. The obvious downsides, however, are that it is confounded by so many different factors that the validity of any results obtained with it would automatically be suspect.

4.4 Price Discovery Mechanism

The final method addressed in this thesis by which slot valuations could be inferred is the one that holds the greatest promise for practical application, given the current state of the aviation industry. The premise of this technique is to take advantage of information revealed as a result of a multi-year auction of a portion of the slots at a given airport. Utilizing this information should provide the most reliable estimate of the slot values of the techniques described, as they come directly from the price-discovery results in the auction. Given this information, slots can be identified for removal over a period of years, to ameliorate the impact of these changes.

First, the background information that could permit and justify this methodology will be discussed. Then, the specific procedures will be described, making a case study of the potential redistribution of slots at LaGuardia Airport. This procedure is in no way specific to LGA, but makes use of it merely as a pedagogical tool.

4.4.1 Background and Justification

As background, the FAA has undertaken efforts to move toward allowing market forces to regulate the ownership of airport slots, particularly at New York's LaGuardia Airport. According to the Federal Register (2006), the first step in this process is to assign expiration dates to the arrival slots between 2010 and 2019. A similar number of slots, approximately 10% of the total, will expire in each of those years, and will then be redistributed, potentially via a market mechanism (e.g. auction). It is planned that the 10% of slots that expire each year will be spread across the course of an entire day, rather than clustered in one particular time period. Once these slots are redistributed, they will have a lifetime of ten years. After this initial allocation period (2010-2019), 10% of slots will continue to expire and be redistributed every ten years.

For the numerical example presented in this section, it was assumed that at LGA there were currently 40 slots in circulation for each hour between 0600 and 2359. Thus, there are a total of 720 arrival slots for a given day. Each of these slots must be assigned in one of the ten years of this initial allocation procedure. Two objectives were to be met in allocating these slots to years: an approximately equal number of slots are auctioned each year for 10 years, and in each year, the slots that are auctioned are distributed approximately uniformly across time periods. Fortunately, the division is simple in this case: dividing these 720 slots by the ten years and 18 hours of the process yields an average number of slots to be distributed per year and hour of 4. Thus, the initial plan for distribution of slots is four per hour per year.

4.4.2 Proposed Procedure

The essential idea of this technique is to make use of the prices paid during the market mechanism distribution. From this point on, it will be assumed that the market mechanism utilized is some type of auction. Under this scheme, the value for each slot will necessarily be discovered as the price paid to purchase the ten-year lease on it. These values will then be used to determine the number of slots that should be made available in each time period. This process is continued year after year, until all slots have either been purchased, or removed. The process by which these values are used is envisioned as outlined in Figure 11.

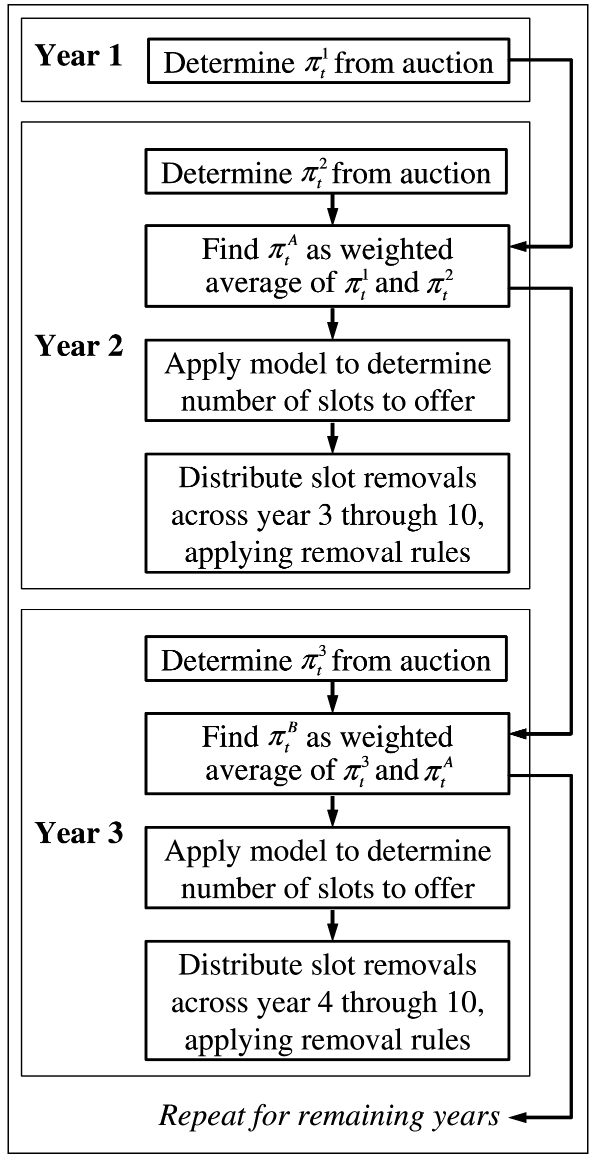

Figure 11 – Price discovery flow chart

The most noticeable feature of the flow chart in Figure 11 is that the process begins simply by observing the prices paid for slots in the first year, denoted here as π_t^1. The current running average of prices paid is denoted as π_t^A. The superscript will increment for each year.

No action is taken in this first year based upon these prices for several reasons. First and foremost, the process will be new to involved participants and as such, prices paid may not truly reflect a well-functioning market. Additional time will be offered to allow the market to stabilize. In addition, it is not strictly necessary at this point in the process to remove slots from circulation. Although slots should be removed as expeditiously as possible to mitigate severe delays, if it is determined that too many are currently available, it is not yet strictly necessary. At this point in the process 90% of the slots are still candidates for removal. This lessens the pressure for immediate, and potentially hasty, action. As time progresses, it becomes more difficult to remove a slot, given the fewer available for removal.

At this point, plots of the style of Figure 12 will be introduced. These will be used throughout the remainder of the discussion of the application of this technique at LGA. The plot on the left shows the average prices paid for slots in this time period, and the overall average price including what was paid in the current time period. The right plot shows the number of slots auctioned in the current and past years, as a portion of the total number of slots. The slot valuation data is necessarily fictitious, as the distribution of slots and the resultant auction have not yet occurred in reality. In the right plot, the gray bars represent the initial allocation of slots offered in year 1. None of the bars are reduced in overall height, as no model has yet been applied to determine the reduced number of slots.

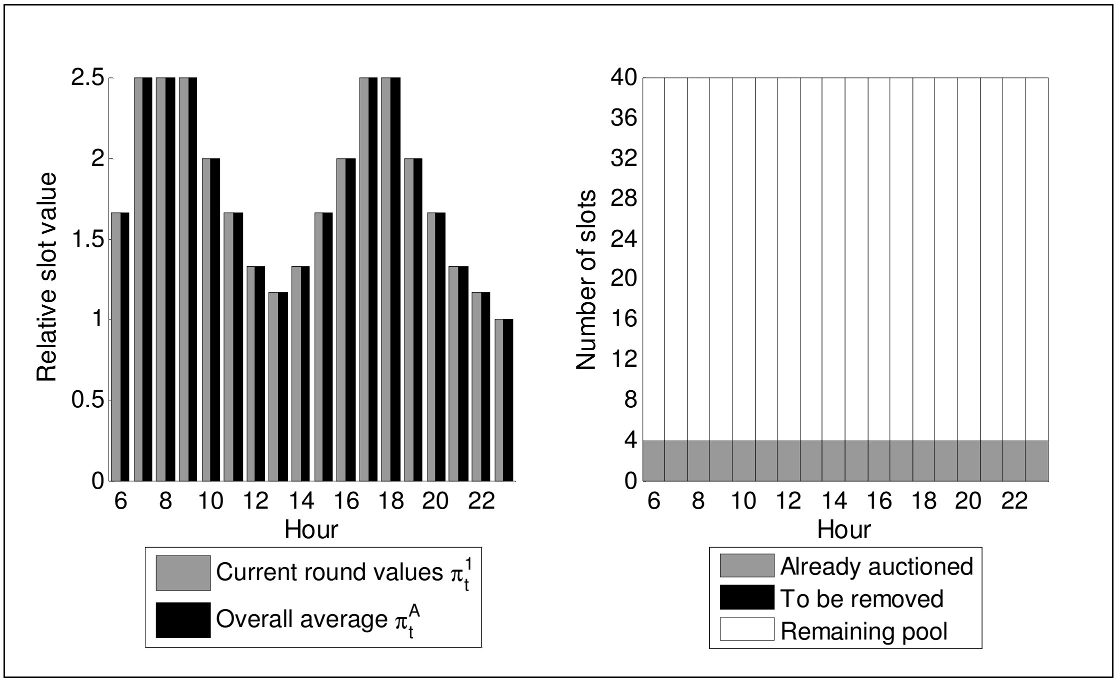

Figure 12 – Year 1 prices and allocation

In later figures, the plot shown on the right will reflect the actual number of slots to be made available, as determined by using the Parametric model, with parameters as specified in Table 17. These represent reasonable values in the spectrum of irregular operations at LaGuardia Airport.

Parameter Name	Symbol	Value
Delay level	γ	15 minutes per flight
Cancellation level	ρ	3.0%
Day length	T	18 hours
Number of capacity scenarios	Q	4 scenarios
Capacity scenario probabilities	p_q	[0.03 0.49 0.40 0.07]
Capacity scenario capacities	$C_{t,q}$	*(see Chapter 2)*
Cancellation costs	λ	*(not needed)*
Maximum delay length	U	3 hours
Number of cancellation arcs	N	3 arcs
Cancellation arc capacity	P_i	4 flights/arc
Upper bound on slot offering	D_{max}	40 slots/hour
Lower bound on slot offering	D_{min}	30 slots/hour
Slot valuation	V_t	*(selected arbitrarily)*

Table 17 – Parameter values

The process continues in the second year with the second 10% of arrival slots being auctioned. With two years of experience in the process, it is now more reasonable to assign some significance to the prices paid for each slot. Thus, after the second year is complete, a new price vector π_t^B is calculated as a weighted average of π_t^1 and π_t^2. The weights assigned to each of these price vectors are the number of slots auctioned during each year. During the course of the second year, this new average valuation function is applied to one of the models discussed in Chapter 2. The resulting output is a vector dictating the total number of slots to be made available in each time period of the day, likely suggesting that some hours have fewer slots than they presently do. Figure 13 shows the resulting prices and required slot removals after the second year of the auctions.

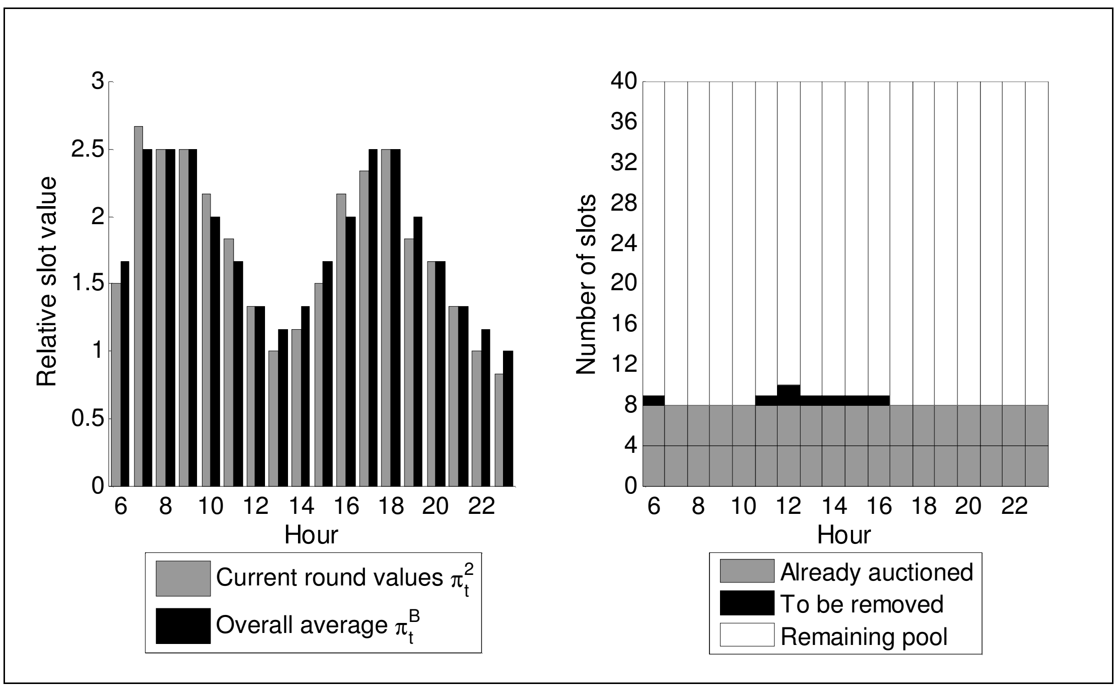

Figure 13 – Year 2 prices and allocation

The black bars shown on the right side of Figure 13 represent the slots marked for removal after the second year auctions. In order to facilitate an orderly transition to the new number of slots, several rules must be applied to distribute these deletions. The motivation for this was presented earlier: after the initial allocation, each of these slots will be leased on a 10 year basis, resulting in approximately 10% of the slot leases expiring each year.

Each year of the initial allocation, beginning with year 2, the model will be solved using the averaged price vector. The number of slots to be offered in each time period will likely be different, at least for the first several years, than what is currently forecast to be offered over the life of the initial allocation. As a result, some slots must be deleted. These deletions must be spread out to limit them to one, or if absolutely necessary two, per hour per year in order to help approximately maintain the 10% per year balance. Limiting removals to one (or two) per hour per year helps to retain some stability and predictability in the marketplace for the carriers involved.

In the third year, the auctions again take place. This is, however, the first year in which some hours may have had slots deleted from the initial allocation of slots to be offered. As can be seen in Figure 14, this was indeed the case in the LGA example. In the 0600, 1100, and 1200-1600 hours, the slots marked in black in Figure 13 were removed from the total. This can be seen as the shortened vertical height of the bars in those columns.

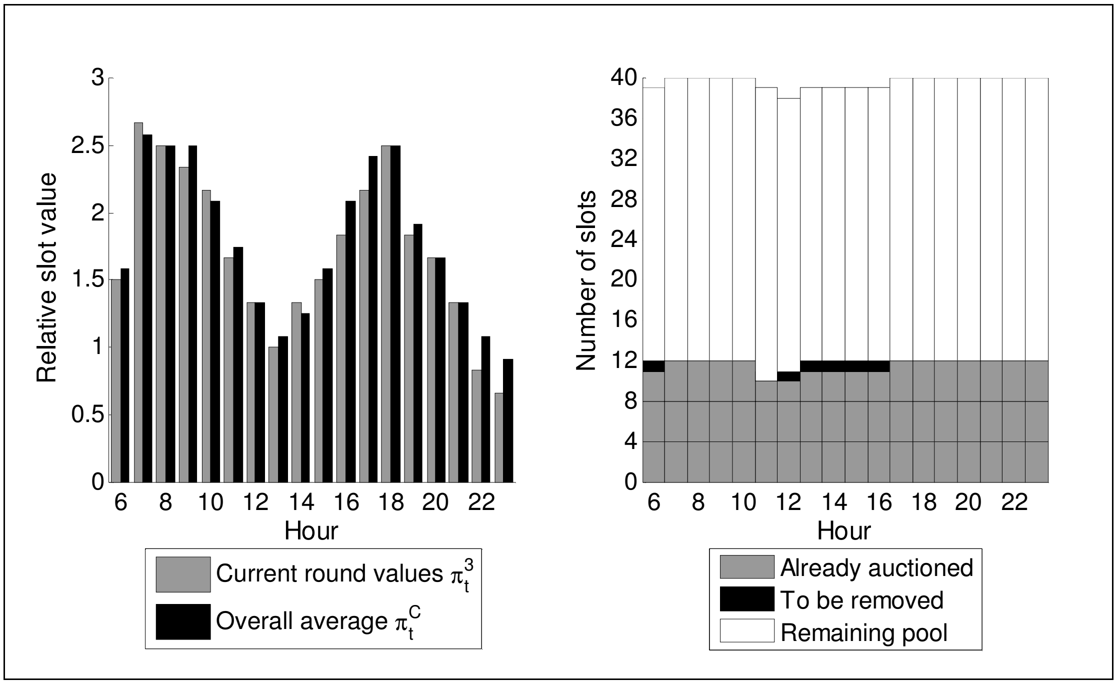

Figure 14 – Year 3 prices and allocation

This process described will continue for each of the remaining years in the initial ten year round of auctions. The results for the remaining years are shown in Figure 15 through Figure 21. The number of slots to be offered in each hour may change slightly as the time progresses, resulting in corrections and slots being returned to the pool. This is a natural result of the volatility in the prices paid to acquire the slots. If it is determined that the model has been overly cautious and removed too many slots from the offerings, some may be added back to the pool of available slots. As shown in Figure 15, an extra slot was added back to the 1100 hour that had been previously removed. As time goes on, it would be reasonable to predict that the prices stabilize, so this was assumed for the case study, implying that neither removals nor additions were required between years 9 and 10.

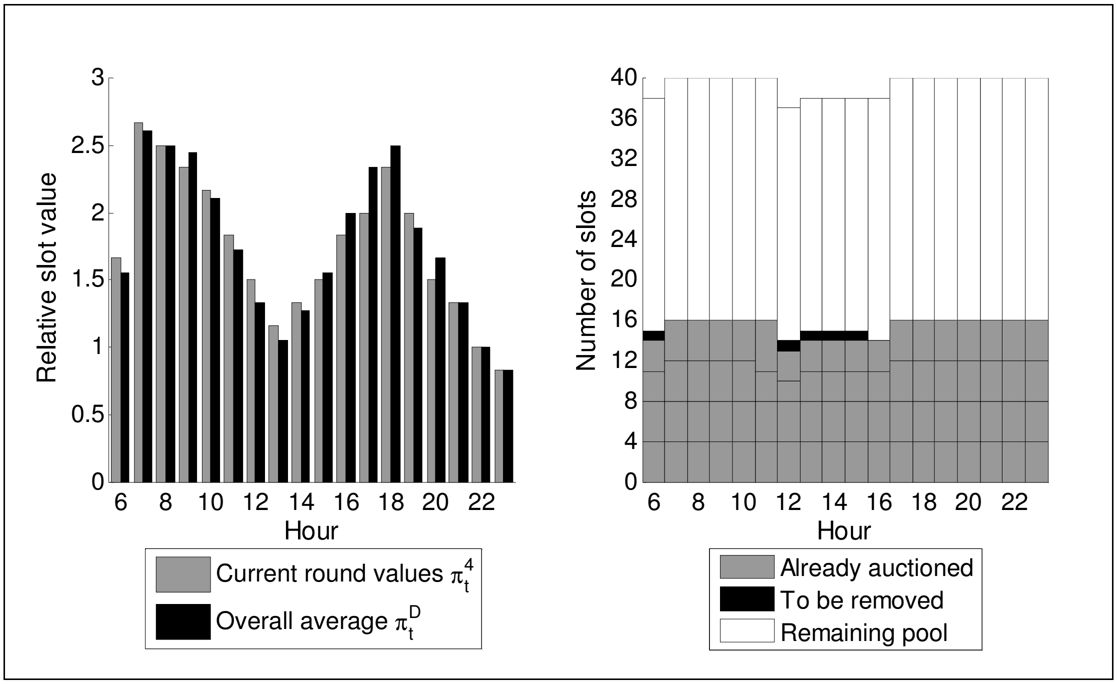

Figure 15 – Year 4 prices and allocation

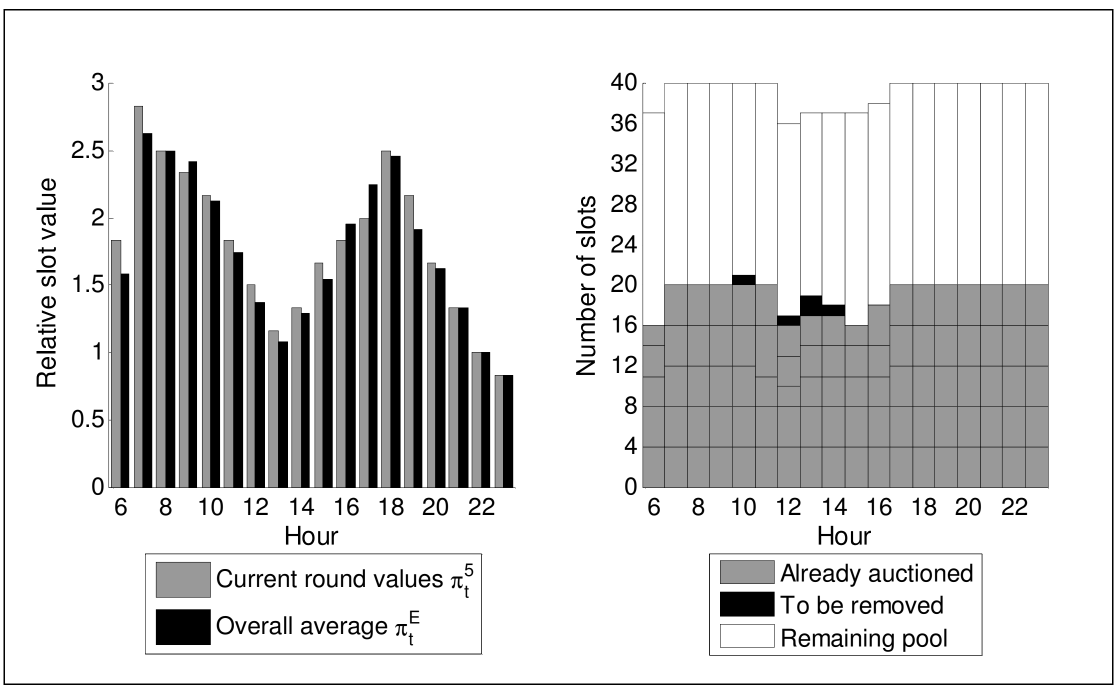

Figure 16 – Year 5 prices and allocation

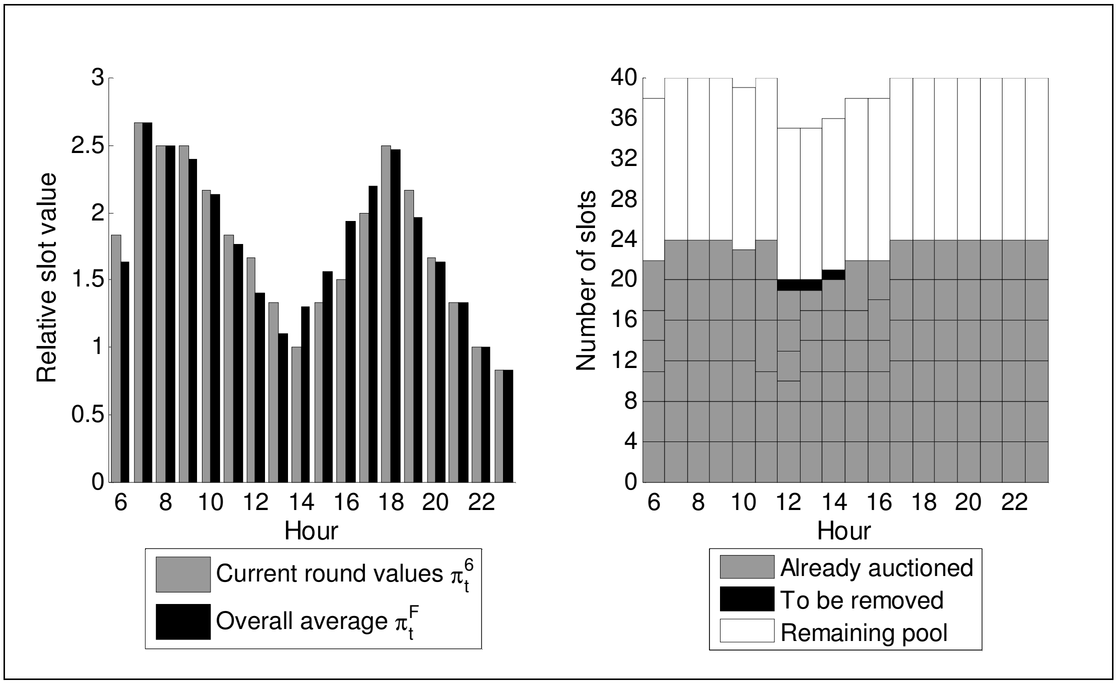

Figure 17 – Year 6 prices and allocation

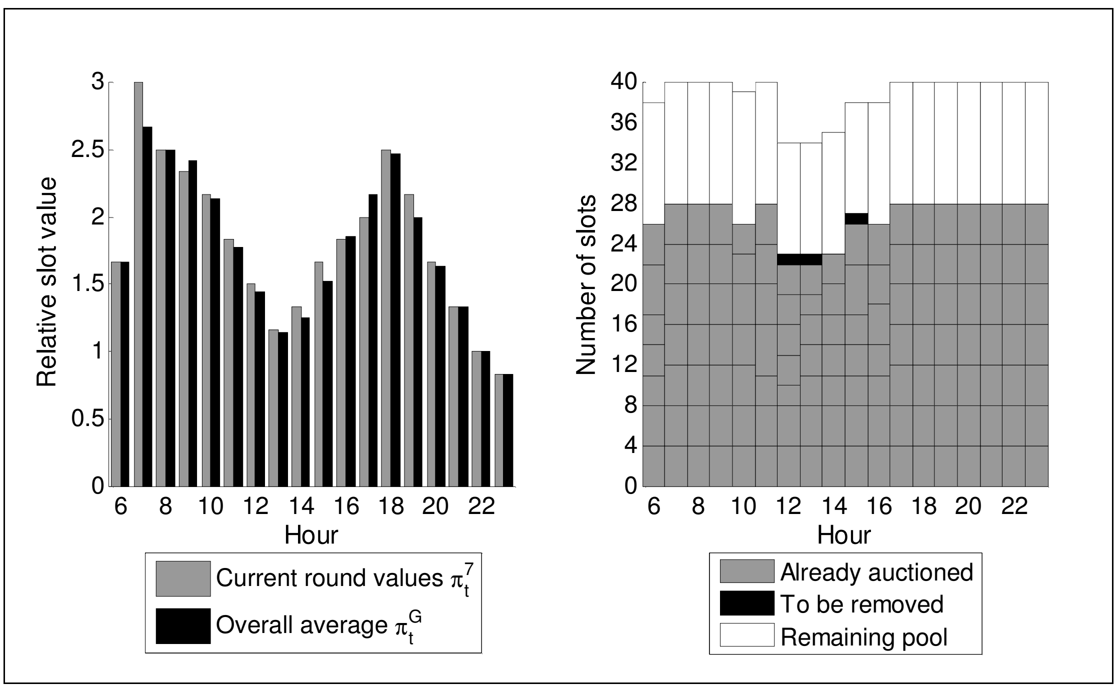

Figure 18 – Year 7 prices and allocation

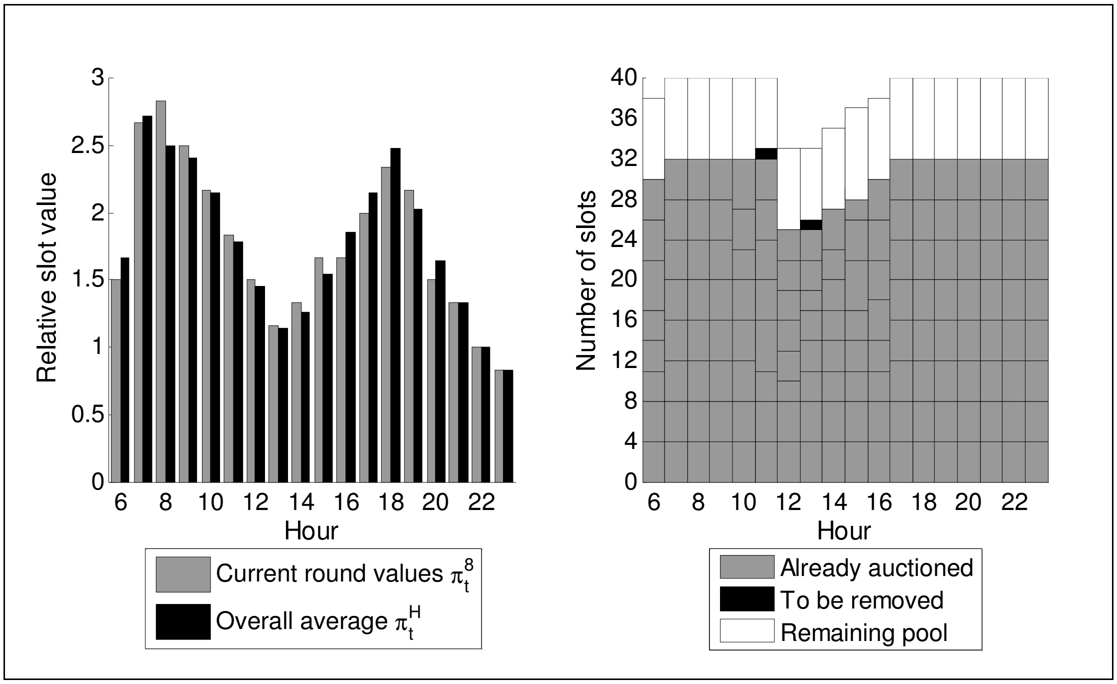

Figure 19 – Year 8 prices and allocation

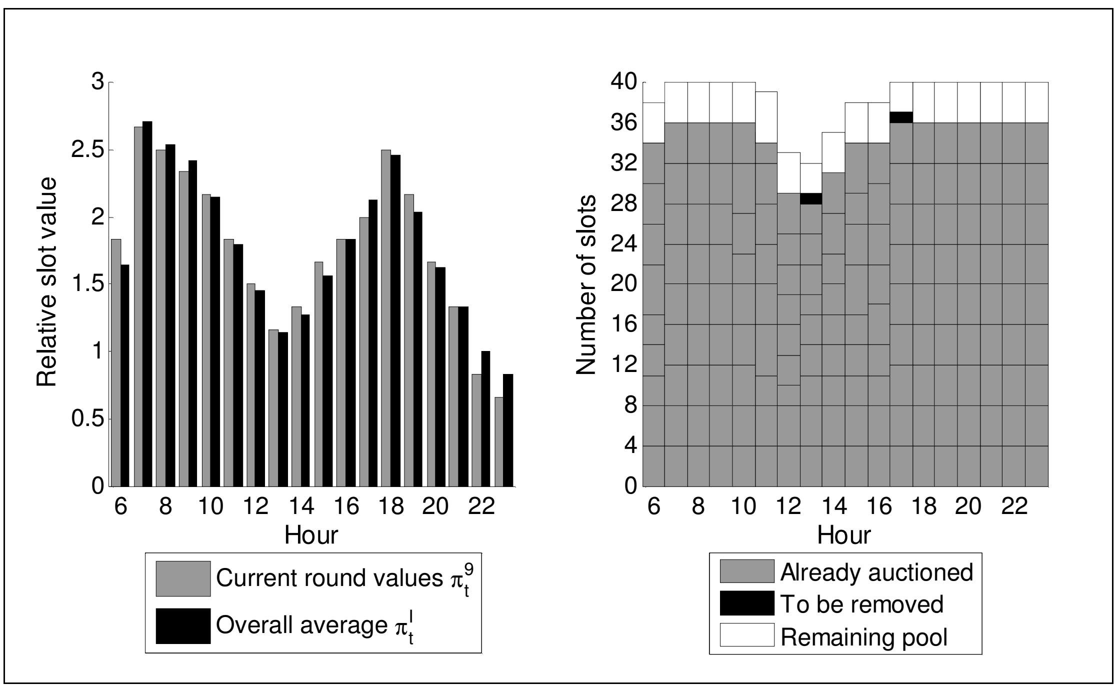

Figure 20 – Year 9 prices and allocation

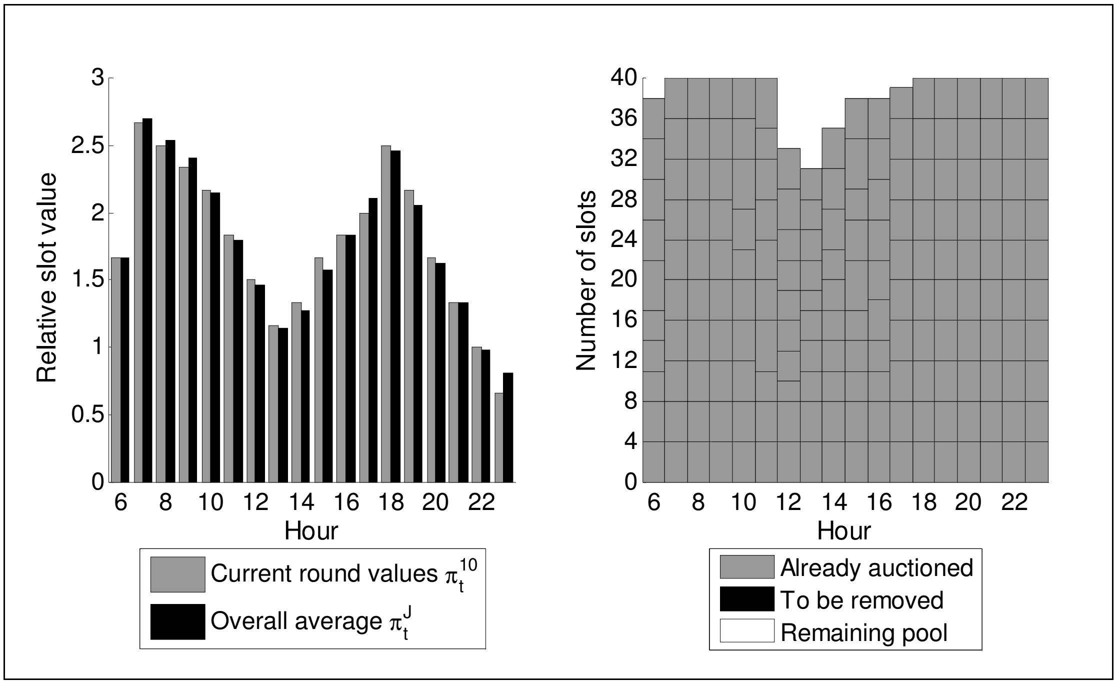

Figure 21 – Year 10 prices and final allocation

At the end of this proposed initial allocation process, all slots have been auctioned off, except those few that it was deemed expedient to remove from circulation. The process performs as expected, as illustrated in the example, and hold promise for being useful in future slot allocation and distribution problems.

The model could continue to be used incrementally after the initial 10 year period ended. The model could be particularly useful to respond to drastic changes in slot valuations. For example, if slots at 0900 became tremendously more valuable (higher prices paid at auction) at some point in the future, this information could be input to the model, and the number of slots offered in each period subsequently adjusted to reflect this change in valuation.

Chapter 5: Incorporation of the Price-Quantity Dynamic

Given information about slot values gleaned from one the sources previously discussed, it is possible to utilize various modeling techniques to examine the relationship between the number of slots offered and their respective values. While this analysis is not strictly necessary for producing a feasible solution, it should provide interesting insight into the economic implications of the efforts undertaken.

Two categories of techniques will be discussed. The first class of techniques proposed is those that permit the model to be formulated as described, but embedded into an iterative loop which converges when the number of slots and their values are consistent with the valuation function assumed. The second category of techniques includes those which add to or alter the formulation of the underlying network structure of the models used. Each technique has its strengths and weaknesses, and could be incorporated into the larger framework of the models presented thus far. Both will be discussed in detail and compared, with computation results presented for the structural model modification.

5.1 Modeling Assumptions

Several considerations must be made regarding the incorporation of the price-quantity relationship into the models presented in this paper. First, the nature of the average values used in this analysis will be described. Then, the functional relationship and a procedure by which it could be estimated will be shown. Clearly, the functional form presented in this document is not the only one that could be used in applying these models, but was chosen for its robustness and ease of calibration.

5.1.1 Nature of Average Values

An important observation regarding the price-quantity relationship discussed in this modeling effort must be addressed first. Each additional slot in these models is not valued at an amount lower than the penultimate slot added, i.e. this is not an issue of decreasing marginal value. Rather, each additional slot decreases the value of each and every other slot offered in that time period. As an extreme example, at a congested airport, if fifty slots are offered in a given time period, each of those is clearly less valuable than if ten slots are offered in that time period. Additionally, if one slot is added to that allocation of fifty, this 51^{st} slot is not less valuable than the other fifty: its value is the same reduced value as the remaining fifty.

5.1.2 Functional Form of Price-Quantity Relationship

Regardless of which technique is employed in modeling this effect, it is necessary to make some assumptions about the functional form of the relationship between slot quantities and valuation. The primary requirement is that slot values be a strictly non-increasing function of the number of slots offered. This is necessary, as the assumption is that additional slots do not enrich the total value of slots offered in one time period.

A huge number of potential functional forms could be used for this analysis. While a form with many parameters could be used to incorporate the nuances of the individual airport being studied, it is important here that the parameters be calibrated with the information available. In another application of these techniques, the resources would be available to define this function with greater detail.

The most natural functional form to use here is the negative exponential. This will be strictly decreasing, and will have few parameters to calibrate. Ideally, this would require calibration of two parameters, one for the "shape" and one for the magnitude. The data for this process would come from one of the analysis techniques discussed in Chapter 2. This is the form that will be used in this thesis. The exponent used will be -0.03. This value was selected heuristically, and should be selected with care after a careful economic study for another application of this model. A plot of this function is shown in Figure 22 for the range under consideration.

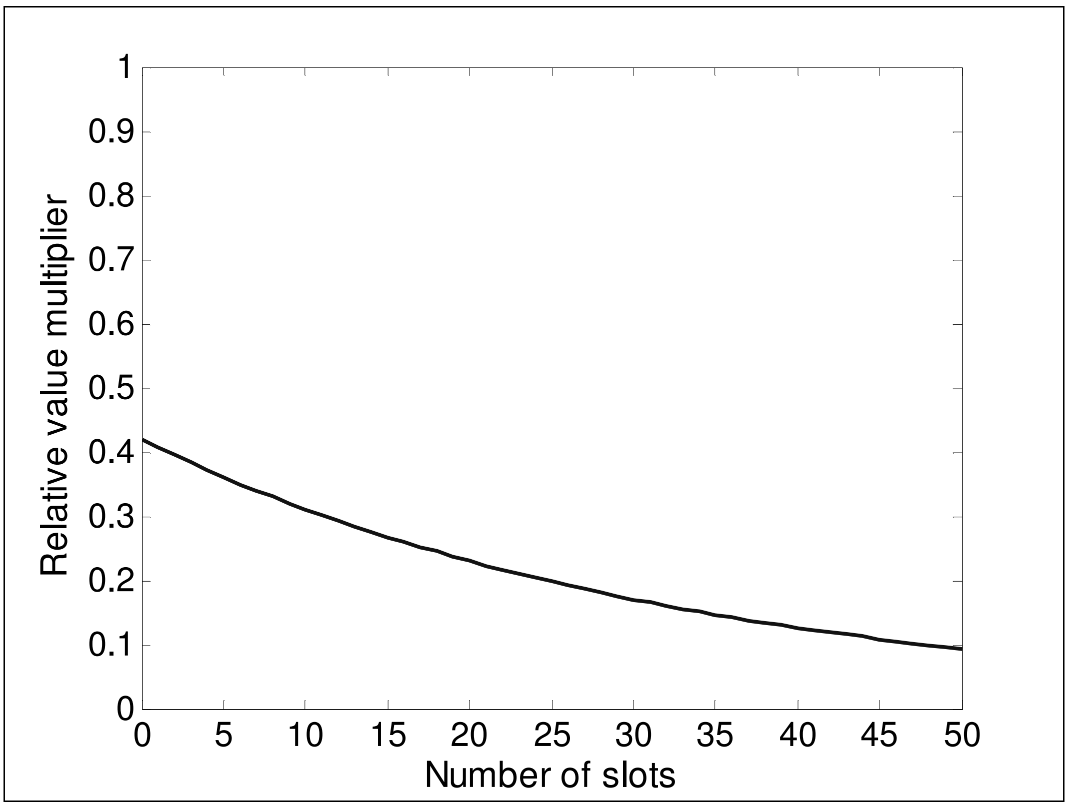

Figure 22 – Proposed price-quantity tradeoff multiplier

The function shown in Figure 22 is used as a coefficient for the slot values previously determined for each hour. Thus, the relationship shown in (5.1) is derived.

$$V_t^* = V_t e^{-0.03 D_t} \tag{5.1}$$

When this relationship is applied to each hour, sets of curves such as the ones shown in Figure 23 are generated. The example shown here makes use of the final averaged values after the tenth year of the auction mechanism proposed in Chapter 4. An important trend to note in this figure is that, because of the nature of the function used, the higher value curves tend to decrease more quickly as the number of slots increase. This relationship seems intuitive to represent reality.

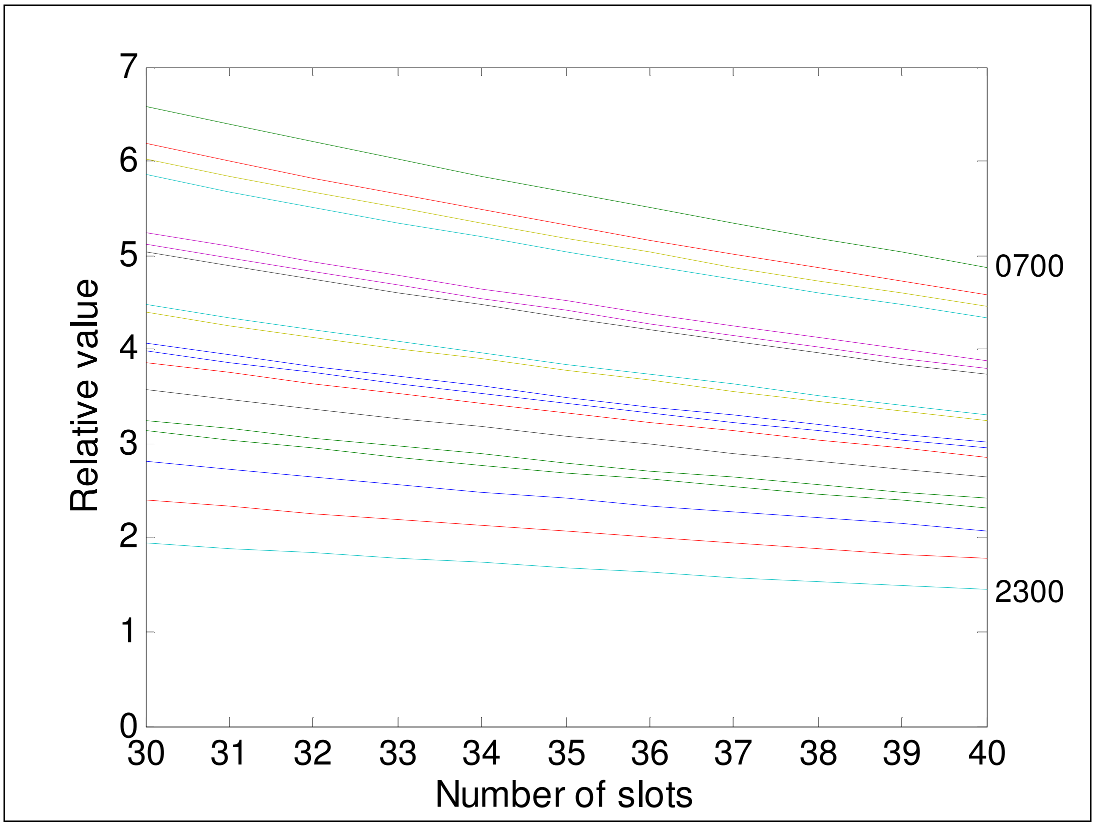

Figure 23 – Sample price-quantity curves for many hours

The curves shown in Figure 23 are applied in the two approaches described in this chapter by examining the relative values at integer-valued number of slots. Each of these curves is discretized to generate a sequence of points consisting of a relative slot value and a quantity of slots. The minimum and maximum value curves are labeled. The other hours obviously lie between these extremes.

5.2 Iterative Equilibrium Model

The first method described to examine the slot value-quantity relationship is discussed in this section. The essence of this technique is that one of the previously discussed models (Base, Consolidated, or Parametric) is embedded in a loop, which is then iterated. The slot valuation is updated based on the number of slots to be made available, as dictated by the chosen model. This process is outlined in Figure 24. The new superscripts on the variables represent the iteration number in the process.

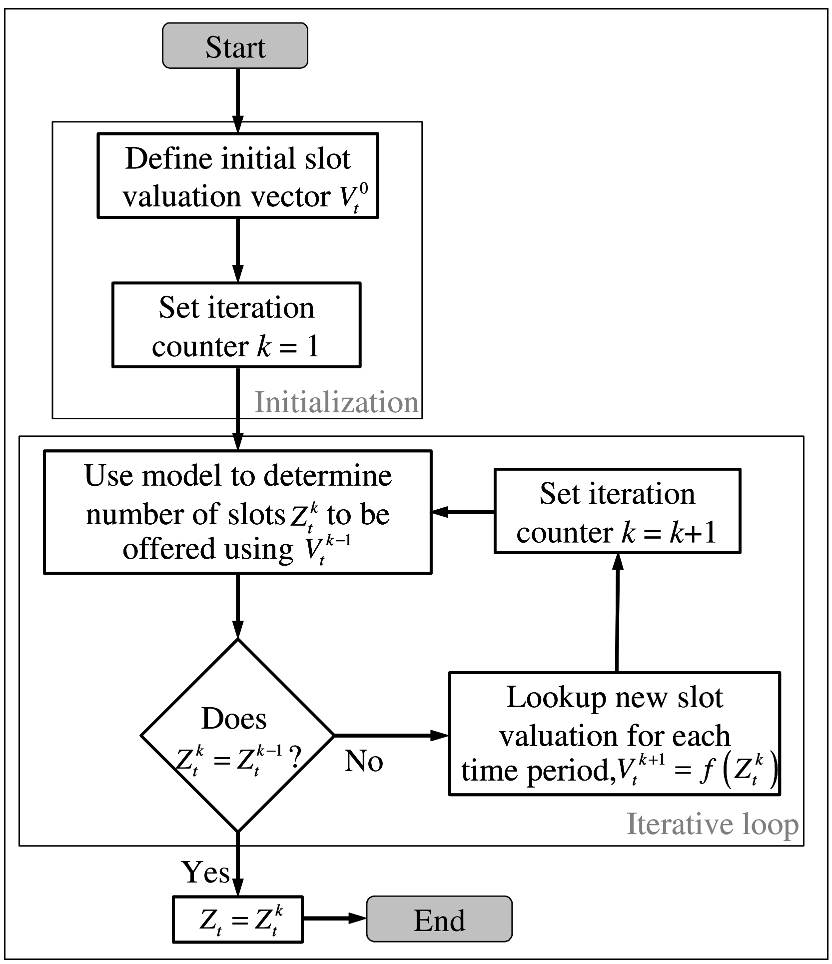

Figure 24 – Iterative model flow chart

As shown in Figure 24, the first step in this process is to define some initial slot valuation vector V_t^0. The most reasonable source of this starting point is the original source of the slot valuation information, whichever was used, as discussed in the previous section.

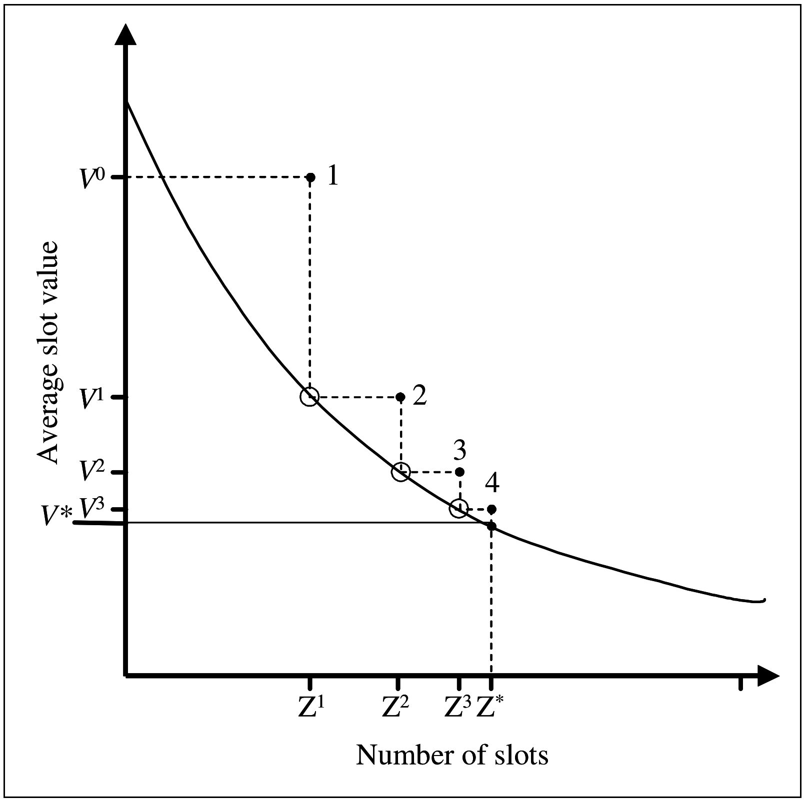

Figure 25 – Iterative model sample results

Using the initial valuation vector, the chosen model is solved to optimality. This yields a vector defining the number of slots to make available in each time period, Z_t^1. The ordered pair $\left(Z_t^1, V_t^0\right)$ very likely does satisfy the function representing the quantity-value tradeoff previously defined for all *t*. It is marked at location 1 in Figure 25. Assuming that one or more of these ordered pairs do not lie on, or sufficiently near to, the curves previously defined, new valuations are found as shown in general form (5.2). The function $f(*)$ refers to the quantity-value function, while the superscript *k* refers to the iteration number.

$$V_t^{k+1} = f\left(Z_t^k\right) \tag{5.2}$$

This process of computing the number of slots to be offered, then computing the associated slot valuation to generate the ordered pair $\left(Z_t^{k+1}, V_t^k\right)$ continues until convergence. In the example shown in Figure 25, the process finds the correct value V_t^1 for Z_t^1, then computes the new number of slots Z_t^2 and the new valuations V_t^2. This process continues until iteration five, at which time it cannot improve further, and the ordered pair $\left(Z_t^5, V_t^4\right)$ generated lies on the curve which properly defines this relationship.

In general, the process will converge when the number of slots to be offered, as computed by the model, is equal for two iterations. At this point, the process is unable to move to the next integer slot value. The value and quantity for that time period have reached equilibrium. Obviously, this process described in this example must take place for each time period simultaneously. As a result of the interactions present between time periods in the model, it may take a large number of iterations for the model to converge completely, with fewer and fewer time periods changing as the number of iterations increases.

This model illustrates the confluence of numerical techniques and optimization. Obviously the results generated for a single airport are not extensible to any other airport – when this procedure is applied, it is decidedly an engineering solution. The alternative method for incorporating the value-quantity dynamic, presented next, is perhaps a more general method by which this phenomenon could be examined.

5.3 Structural Modifications

It is possible that this value-quantity relationship could be incorporated directly into the structure of the model. The most natural way to accomplish this would be to replace the arcs connecting the source node to the cancellation decisions on which Z_t flows with several arcs $j \in \{1, \ldots, J\}$, each of varying value. A flow diagram illustrating this technique is shown in Figure 26.

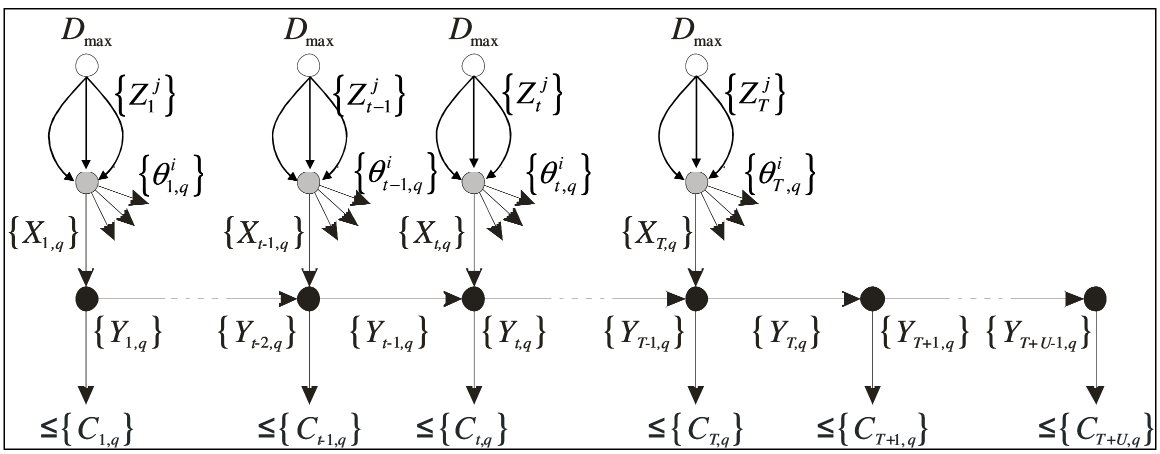

Figure 26 – Multiple slot value arc model flow diagram

5.3.1 Model Structure

While it would be tempting to assume that each of these many arcs has a very low capacity and a value marginally less than the previous, this assumption would be incorrect in the framework being used. The total number of slots to be offered in a time period t would be calculated as shown in (5.3). This scenario would provide control only over the value of a marginal slot, but not over the mean value of each slot in a time period. As explained, a marginal slot does have a value lower than the others in its time period; rather, its addition lowers the value of itself and all others.

$$Z_t = \sum_{j=1}^{J} Z_t^j \tag{5.3}$$

The correct way to incorporate the price-quantity dynamic into the structure of the model could make use of the same flow diagram shown in Figure 26. Each of the arcs *j* on which Z_t^j flowed would have capacity equal to D_{max}^j. The supply function D_{max}^j would define the same for all *t*, but would be an increasing function with *j*. That is, each slot valuation arc would have higher capacity than the previous one.

However, a side constraint would be needed to permit only one of each of these arcs to be used in each time period. This could be accomplished using binary variables $\{S_t^j\}$ that would equal 1 if slot value arc *j* were used in time period *t*, otherwise 0. These binary variables are constrained as shown in (5.4).

$$S_t^j \in \{0,1\} \quad \forall t \in \{1,\ldots,T\}, j \in \{1,\ldots,J\} \tag{5.4}$$

Only a single $\{S_t^j\}$ may be 1 for each t, as constrained by (5.5). The objective function will pressure the model to make as many of these one as possible.

$$\sum_{j=1}^{J} S_t^j = 1 \quad \forall t \in \{1,\ldots,T\} \tag{5.5}$$

Assuming that the capacity of each of the new parallel arcs is incremented by 1 unit, each occurrence of Z_t can be replaced by (5.6), where D_{max}^j represents the number of flights using arc *j*.

$$Z_t = \sum_{j=1}^{J} S_t^j D_{max}^j \tag{5.6}$$

. It is important to consider in what fashion each of the new slot valuation arc capacities D^j_{max} and values $\{V_t^{\,j}\}$ are be defined. As discussed, some functional form must be defined to model the quantity-value dynamic. From this function, unique ordered pairs (D,V) could be extracted that satisfied this function. Each of these pairs will be used to set the value and capacity for an arc in each time period. Necessarily, the exact parameters of the function defining these ordered pairs may be different in each time period, to account for differences in average slot value.

Thus, the constraints on the program are pruned to reflect this change. The previously defined bounds on Z_t are removed. Anywhere else that Z_t appears, it must be replaced by (5.6). This includes the flow balance equations for the gray nodes shown in Figure 26. The new constraint is shown in (5.7).

$$\sum_{j=1}^{J} S_t^{\,j} D^j_{max} - X_{t,q} - \sum_i \theta_{t,q}^{\;i} = 0 \tag{5.7}$$

Additionally, for the Parametric model, the side constraints limiting the delay and cancellation levels must be augmented to the new expression for Z_t shown in (5.6). These updates are shown in (5.8) and (5.9).

$$\sum_q p_q \sum_t Y_{t,q} - \gamma \sum_t \sum_{j=1}^{J} S_t^{\,j} D^j_{max} \le 0 \tag{5.8}$$

$$\sum_q p_q \sum_i \sum_t \theta^i_{t,q} - \rho \sum_t \sum_{j=1}^{J} S_t^{\,j} D^j_{max} \le 0 \tag{5.9}$$

While the objective functions in the Base and Consolidated models are different from in the Parametric model, the term $\sum_t V_t Z_t$ is common in each of them, and must be modified to accommodate this change. The new term is shown in (5.10).

$$\sum_t \sum_{j=1}^{J} V_t^{\,j} D_{max}^{j} S_t^{\,j} \tag{5.10}$$

The Parametric formulation with these modifications is shown in Table 18.

$\max\left\{\sum_t \sum_{j=1}^{J} V_t^{\,j} D_{max}^{j} S_t^{\,j}\right\}$		(5.10)
Subject to		
$\theta_{t,q}^{\;\;i} \le P_i$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\}$	(2.2)
$\sum_{j=1}^{J} S_t^{\,j} D_{max}^{j} - X_{t,q} - \sum_i \theta_{t,q}^{\;\;i} = 0$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\}$	(5.7)
$X_{1,q} - Y_{1,q} \le C_{1,q}$	$\forall q \in \{1,\ldots,Q\}$	(2.4)
$X_{t,q} + Y_{t-1,q} - Y_{t,q} \le C_{t,q}$	$\forall t \in \{2,\ldots,T\},\ q \in \{1,\ldots,Q\}$	(2.5)
$Y_{t-1,q} - Y_{t,q} \le C_{t,q}$	$\forall t \in \{T+1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.6)
$Y_{T+U-1,q} \le C_{T+U,q}$	$\forall q \in \{1,\ldots,Q\}$	(2.7)
$Y_{t,q} \le W_{t,q}$	$\forall t \in \{1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.9)
$\sum_{j=1}^{J} S_t^{\,j} = 1$	$\forall t \in \{1,\ldots,T\}$	(5.5)
$X_{t,q}, Y_{t,q}, \theta_{t,q}^{\;\;i} \in \mathbb{N}^{+}$	$\forall t \in \{1,\ldots,T\},\ q \in \{1,\ldots,Q\},\ i \in \{1,\ldots,N\}$	(2.10)
$S_t^{\,j} \in \{0,1\}$	$\forall t \in \{1,\ldots,T\},\ j \in \{1,\ldots,J\}$	(5.4)
$\sum_q p_q \sum_t Y_{t,q} - \gamma \sum_t \sum_{j=1}^{J} S_t^{\,j} D_{max}^{j} \le 0$		(5.8)
$\sum_q p_q \sum_i \sum_t \theta_{t,q}^{i} - \rho \sum_t \sum_{j=1}^{J} S_t^{\,j} D_{max}^{j} \le 0$		(5.9)
Where		
$W_{t,q} = \sum_{i=t+1}^{\min\{t+U,T+U\}} C_{i,q}$	$\forall t \in \{1,\ldots,T+U-1\},\ q \in \{1,\ldots,Q\}$	(2.8)

Table 18 – Parametric model formulation with structural modifications

5.3.2 Mathematical Properties

The new upper bound on the $Z_t^{\,j}$ term will introduce coefficients not in the set {-1,0,1} into the constraint matrix. Using this fact, and the results presented for the Base model's mathematical properties in Chapter 2, the constraint matrix for this problem will not be totally unimodular. Thus, by the results presented for the Consolidated and Parametric model, their constraint matrices will also not be TU. As a result, integer-valued optimal solutions are not guaranteed by the linear programming relaxation.

Because of the framework in which this model has been particular version of the model has been posed, obtaining integer-valued solutions using the linear programming relaxation is not of paramount concern. Because a model that makes use of the structural modifications described in this section is not intended to be run in an equilibrium loop, requiring extra solution time is not of concern. This issue is particularly nullified by the reasonably small problem size.

5.4 Comparison of Value-Quantity Modeling Techniques

Each of the two techniques described to incorporate the price-quantity dynamic has promise for producing useful and interesting results. Either is certainly feasible to implement, given that sufficient analysis has been undertaken regarding estimating the price-quantity function common to both approaches.

In terms of computational effort required for a solution, it seems likely that the structural model should be more effective. It requires the solution of an integer program using but a few additional variables onto a program that already solves extremely rapidly, for reasons discussed previously. The iterative method may take a large, and potentially infinite, number of iterations before it converges to a solution.

If the iterative model is unable to converge to a solution in which all variables obtain integer values, then it would be necessary to define less stringent convergence criteria. This adds an additional complication to this model.

However, this very iterative nature may make this model more attractive than the structural modifications. In the structural model, the slots values are estimated using the tradeoffs in one fell swoop – during the solution of the integer program. In the iterative model, they are permitted to move more, as the process iterates.

Regardless of which method is applied, interesting results should be found to address the concern that the price-quantity tradeoff issue be included in such an analysis as is conducted in this thesis. In the following section, the structural model will be tested, for the very reason first espoused – it is simpler to implement.

5.5 Computational Results

The structural modifications to the Parametric model described in Section 5.3 were implemented and tested in a computational experiment. The parameters used in this analysis are shown in Table 19.

Parameter Name	Symbol	Value
Delay level	γ	15 minutes per flight
Cancellation level	ρ	3.0%
Day length	T	18 hours
Number of capacity scenarios	Q	4 scenarios
Capacity scenario probabilities	p_q	[0.03 0.49 0.40 0.07]
Capacity scenario capacities	$C_{t,q}$	*(see Chapter 2)*
Cancellation costs	λ	*(not needed)*
Maximum delay length	U	3 hours
Number of cancellation arcs	N	3 arcs
Cancellation arc capacity	P_i	4 flights/arc
Upper bound on slot offering	D_{max}	40 slots/hour
Lower bound on slot offering	D_{min}	30 slots/hour
Slot valuation	V_t	

Table 19 – Parameters used in structural modification experiment

The model solved to optimality rapidly. The resultant number of slots to offer in each hour is shown in Figure 27. They indicate that the optimal number of slots to offer in each time period, when balancing the value-quantity tradeoff is nearly constant. While these results stand in contrast to the results obtained without considering this assumption, they indicate the differing results that may be obtained by considering different objectives in solving this problem. It seems that trying to find the optimal balance between price and quantity dampens the effect previously observed wherein the model results mimic the slot valuations.

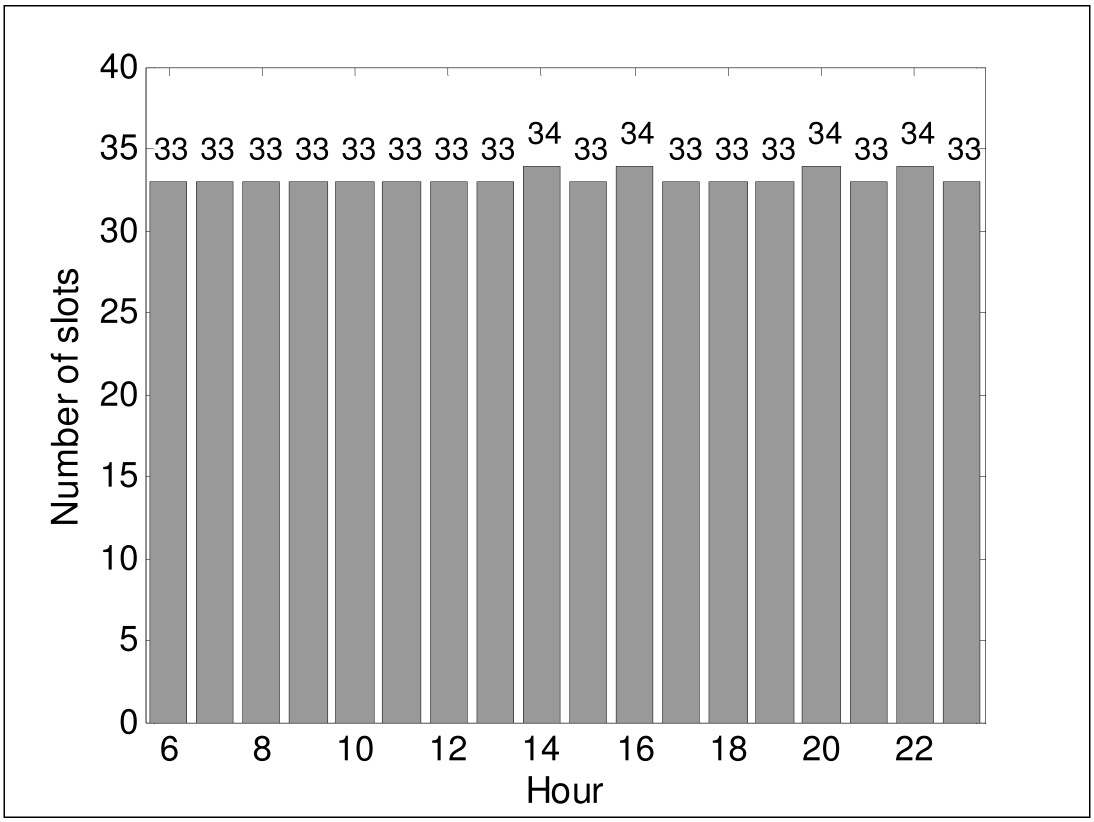

Figure 27 – Structural model results

Chapter 6: Number of Slots Computational Results

The aim of this section is to provide a summary of the results produced in this study. Obviously a large number of possible combinations of parameters are possible, but those presented here are included to show: a) sensitivity of the model to various parameter changes, and b) a comparison to the current numbers of slots offered in each time period. Discussions about problem size and solution time will also be included.

6.1 Problem Size

A reasonable practical problem might use either hourly or quarter-hourly time slices, from 0600 to 2359 local time. This would imply that $T = 18$ or $T = 72$, respectively. The number of distinct capacity scenarios at any airport that correspond to recognizable and repeatable runway and weather combinations should be on the order of $Q = 4$. This is the number used in this study based on the work on Liu et al. (2005) but it is reasonable that at other airports this value could vary. One might imagine a cancellation cost function with several steps, as shown in Figure 4. For this brief analysis, $N = 3$ cancellation arcs will be used. Finally, a reasonable policy decision would be to assume for the purposes of this model that flights will be cancelled rather than being delayed more than 3 hours. This would imply that $U = 3$, or $U = 12$, depending on whether hourly or quarter-hourly data was being used.

Table 20 and Table 21 illustrate the "size" of the problem in general, and for these illustrative numbers in particular. Because the number presented is the count of inequality constraints, the equalities in (2.3) must be double-counted. Also, note that

(2.8) is really a pre-processing step related to problem constants, and does not represent an independent constraint.

Variable	**Number of variables**		
	General	**Hourly, *T* = 18, *Q* = 4, *N* = 3, *U* = 3**	**Quarter-hourly, *T* = 72, *Q* = 4, *N* = 3, *U* = 12**
$\{X_{t,q}\}$	TQ	72	288
$\{Y_{t,q}\}$	$(T + U - 1)Q$	80	296
$\{Z_t\}$	T	18	72
$\{\theta^i_{t,q}\}$	TQN	216	864
Total	$TQN + 2TQ + UQ + T$	386	1520

Table 20 – Base problem size, number of variables

Constraint	**Number of constraints**		
	General	**Hourly, *T* = 18, *Q* = 4, *N* = 3, *U* = 3**	**Quarter-hourly, *T* = 72, *Q* = 4, *N* = 3, *U* = 12**
(2.1)	$2T$	36	144
(2.2)	TQN	216	864
(2.3)	$2TQ$	144	576
(2.4)	Q	4	4
(2.5)	$(T - 1)Q$	68	284
(2.6)	$(U - 1)Q$	8	8
(2.7)	Q	4	4
(2.9)	$(T - \mathrm{U} - 1)Q$	56	272
Total	$2T + TQN + 4TQ + 2UQ - Q$	536	2156

Table 21 – Base problem size, number of constraints

Results such as these could be derived for the consolidated model, but they would be of considerably less interest, given that the Consolidated model solves to optimality with integer solutions using the linear programming relaxation.

The results for the Parametric model are not presented separately, but can be derived easily from the Base model results. Recall that the differences between the two formulations are the objective function (not relevant in this discussion) and the addition of two constraints in the Parametric model. Thus, the number of variables present in the Parametric model is equal to the Base model, while the number of constraints in the Parametric model is two greater. That is not to say that the Parametric model necessarily requires a similar amount of time to solve to optimality: recall that the Base model constraint matrix can often be shown to be totally unimodular, or 2-regular, whereas the Parametric model constraint matrix cannot be.

The structural modifications proposed in Section 5.3 would also change the problem size. All of the T Z_t variables would be removed, but TQ binary S_t^j variables would be added. The number of constraints would decrease overall, as the T upper and T lower bounds on the Z_t variables would be removed. However, T constraints would be added to limit one S_t^j to be one in each time period. The results for these changes, relative to the Parametric model, are summarized in Table 22 and Table 23.

Model Formulation	Number of Variables		
	General	**Hourly, $T = 18, Q = 4, N = 3, U = 3$**	**Quarter-hourly, $T = 72, Q = 4, N = 3, U = 12$**
Base	$TQN + 2TQ + UQ + T$	386	1520
Parametric	$TQN + 2TQ + UQ + T$	386	1520
Structurally-modified Parametric	$TQN + 3TQ + UQ$	440	1736

Table 22 – Summary of problem sizes, number of variables

Model Formulation	Number of Constraints		
	General	**Hourly, $T = 18, Q = 4, N = 3, U = 3$**	**Quarter-hourly, $T = 72, Q = 4, N = 3, U = 12$**
Base	$2T + TQN + 4TQ + 2UQ - Q$	536	2156
Parametric	$2T + TQN + 4TQ + 2UQ - Q + 2$	538	2158
Structurally-modified Parametric	$T + TQN + 4TQ + 2UQ - Q + 2$	520	2086

Table 23 – Summary of problem sizes, number of constraints

As explained in Chapter 2, the Parametric formulation does not guarantee integer-valued optimal solutions using the linear programming relaxation. As a result, the branch-and-bound algorithm was implemented by the software package. Exact statistics are not shown for every case tested in this thesis, but solution times on a modern desktop computer were on the order of tenths of a second. This model solves very fast, and would likely be well suited to be used in the iterative equilibrium model suggested in Chapter 5.

6.2 Parameter Sensitivity Analysis

It is important to understand the sensitivity of this class of models to the various input parameters required. In this section, an informal analysis will be undertaken to evaluate the performance of the model under various conditions. The various delay and cancellation parameters will be varied, and the resulting model outputs examined. Data for LaGuardia airport were used for this example, as before. This was only chosen to be representative of any congested airport.

In this section, only the Parametric formulation of the model was tested. The primary reason for doing so was to separate any effects that specifying the slot values and various costs relative to one another may have on the model performance. That is, the Parametric model creates a disconnect between the various values in the model, so that the variations brought on by each of these parameter classes can be more carefully examined. The parameters used in the tests in this section are shown in Table 25. The slot values used in the test are the final ones derived after the end of the proposed ten-year auction period discussed in Chapter 4.

Parameter Name	Symbol	Value
Delay level	γ	15 minutes per flight
Cancellation level	ρ	3.0%
Day length	T	18 hours
Number of capacity scenarios	Q	4 scenarios
Capacity scenario probabilities	p_q	[0.03 0.49 0.40 0.07]
Capacity scenario capacities	$C_{t,q}$	*(see Chapter 2)*
Maximum delay length	U	*(varies)*
Number of cancellation arcs	N	3 arcs
Cancellation arc capacity	P_i	3 flights/arc
Upper bound on slot offering	D_{max}	40 slots/hour
Lower bound on slot offering	D_{min}	30 slots/hour
Slot valuation	V_t	*(see Figure 21)*

Table 24 – Parameters used in slot valuation analysis

6.2.1 Delay Parameters

The only parameter relating explicitly to delay in these models is the maximum permitted delay length. In this analysis, this parameter U was varied between 1 and 5 hours to examine the effects that it will have on model performance. Table 25 shows several summary statistics about each of the cases tested in this analysis. The results in this table seem reasonable, and suggest that the model is performing properly- namely, it is using the entire allowance for delay and cancellation that it is permitted. Additionally, the number of slots offered under each scenario decreases with the maximum delay length permitted, as was expected.

Scenario	Total number of slots to offer each day	Expected average delay length (minutes/flight)	Expected number of daily cancellations	Expected cancellation percentage
$U = 1$	671	15.00	20.07	2.991%
$U = 2$	686	14.99	20.57	2.999%
$U = 3$	688	14.99	20.64	3.000%
$U = 4$	688	14.99	20.64	3.000%
$U = 5$	688	15.00	20.64	3.000%

Table 25 – Delay parameter sensitivity analysis statistics for LGA

Figure 28 shows the number of slots to make available in each of these cases, while Figure 29 and Figure 30 show the expected number of delayed flights and cancellations, respectively, in each of these cases. Obviously the delay and cancellation numbers should be integer-valued, but their expectations across capacity scenarios may not be.

The results in Figure 28 showing the number of slots to make available in each time period seems reasonable. The primary difference to note is that the most restricted model, U = 1, allows more flights during the middle of the day, but is forced to curtail operations at the end of the day because of its inability to delay flights much beyond that period. The other models are able to take greater advantage of this end-of-day effect.

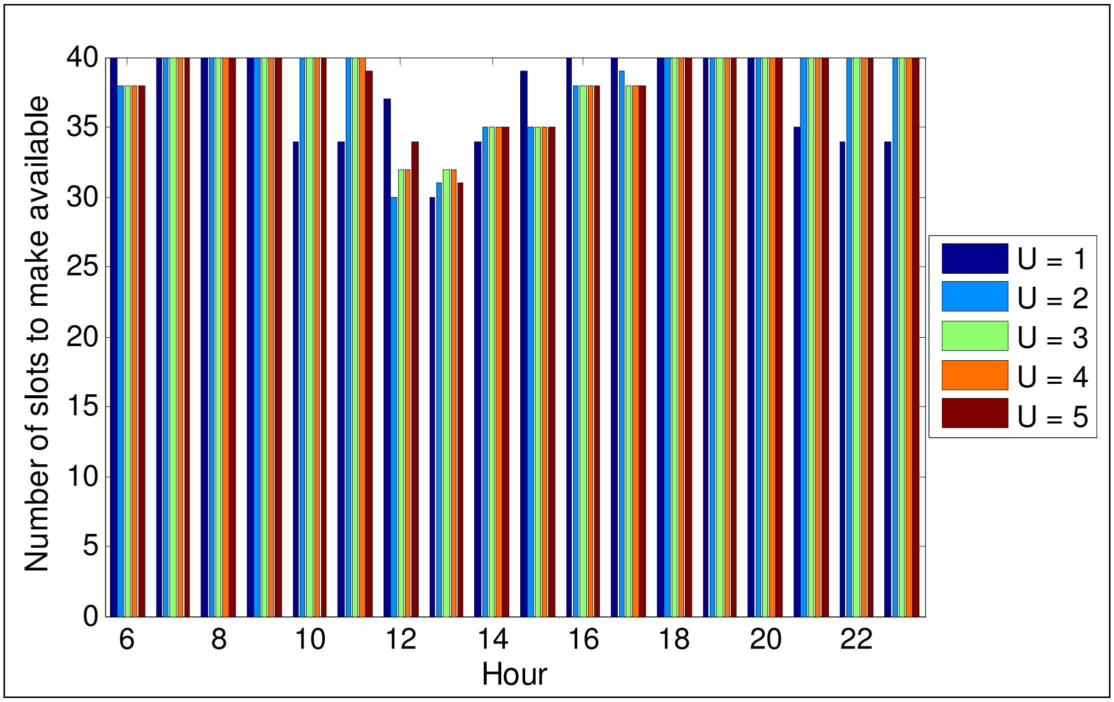

Figure 28 – Number of slots to make available for various *U* values at LGA

The results in Figure 29 also seem reasonably intuitive. The primary difference to note again involves the $U = 1$ case. It is more apt to take advantage of the middle of the day to conduct operations where other models are more apt to use this as a recovery period. This trend is evidenced by the higher number of delayed flights during the middle of the day by the $U = 1$ model. It should be noted that that delays do build near the very end of the day under most scenarios, but these delays are likely not very long, as evidenced by the small number of flights delayed beyond the end of the operating day (i.e. past hour 24). It is impossible to track the individual delay length of a given flight, however, given the network assumptions used.

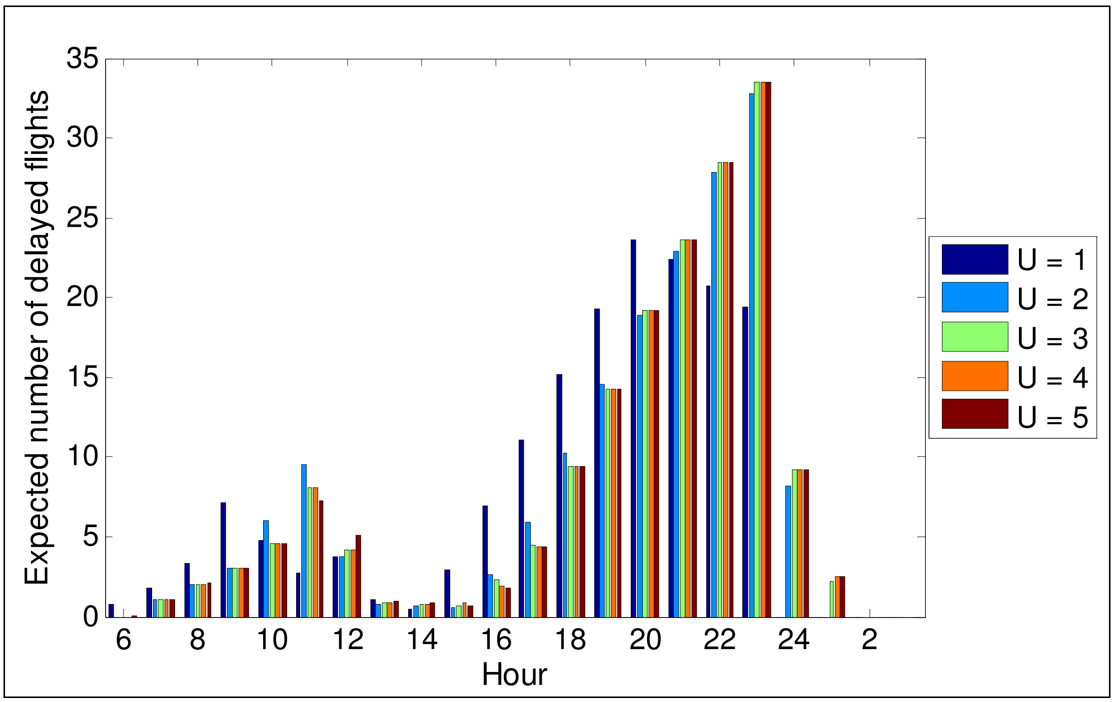

Figure 29 – Expected number of delayed flights for various U values at LGA

As in the previous figures, the results shown in Figure 30 are as expected. The primary feature of note is that the models tend to cancel more flights at the beginning of the day. This trend is also observable in the historical data for operations at LGA. Additionally, it is, in some sense, easier to cancel a flight during those hours simply because there are more flights operating than in other hours. Another trend to note is that as the day nears its end, only the very test runs in which a low value of the maximum delay length is used are forced to use cancellations to control the model performance. In these cases, the models with a higher maximum delay length are better able to take advantage of this delay.

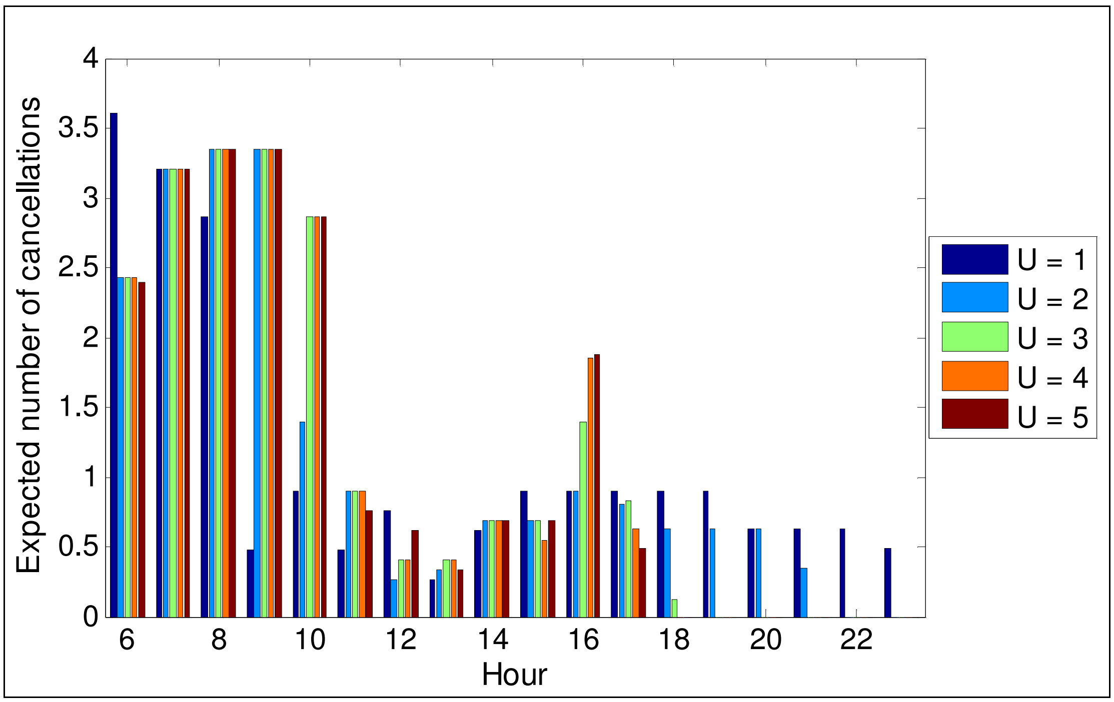

Figure 30 – Expected number of cancellations for various *U* values at LGA

The trivial case of $U = 0$ was not examined in this analysis, but the results of such a test are trivial: the model would function as if each time period were independent. The highest value time periods would use the maximum number of slots and cancellations permitted under the Parametric conditions specified. The remainder of the time periods would only provide as many slots as permitted by the landing capacity in that time period alone. None of the delay allocation would be utilized by the model, as those arcs representing delay in the network flow diagrams would essentially not exist.

6.2.2 Cancellation Parameters

A sensitivity analysis concerning the cancellation parameters used in these models only makes sense under certain conditions. The first among these concerns is that the analysis be conducted using the Base or Consolidated formulations, but not the Parametric formulation. The reasoning for this condition is that the actual cancellation level permitted is specified explicitly in the Parametric model. In the Base and Consolidated formulations, the model finds the system optimal level, which is necessarily influenced by the parameters chosen. This all assumes that sufficient capacity is provided to the Parametric model to allow it to make full use of the cancellation level allowed it. Because the example analysis conducted in this paper was focused on applying the Parametric model, and did not involve properly specifying the parameters in a means appropriate for applying the other models, results for a cancellation parameter sensitivity analysis will not be presented.

6.3 Slot Valuation Sensitivity Analysis

Several different sets of fictitious slot valuation vectors were tested to evaluate the sensitivity of the model. Each of these scenarios was chosen for specific reasons that will be explained. The various scenarios are listed in Table 26. The Parametric model was again used exclusively, in order to isolate the effects of the slot value variation. The magnitude of these values is not significant when used in the Parametric formulation- only the relative values are needed.

	Scenarios									
Hour	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
6:00	1	1	1	1	1	1	18	600	500	600
7:00	1	1	2	2	1	2	17	800	800	900
8:00	1	1	1	3	2	3	16	800	800	900
9:00	1	1	2	2	2	4	15	800	800	900
10:00	1	1	1	1	1	5	14	600	500	600
11:00	1	1	2	2	1	6	13	600	500	600
12:00	1	1	1	3	2	7	12	800	800	900
13:00	1	1	2	2	2	8	11	800	800	900
14:00	1	1	1	1	1	9	10	600	500	600
15:00	1	1	2	2	1	10	9	600	500	600
16:00	1	1	1	3	2	11	8	800	800	900
17:00	1	1	2	2	2	12	7	800	800	900
18:00	1	1	1	1	1	13	6	800	800	900
19:00	1	1	2	2	1	14	5	800	800	900
20:00	1	1	1	3	2	15	4	600	500	600
21:00	1	1	2	2	2	16	3	600	500	600
22:00	1	1	1	1	1	17	2	275	275	275
23:00	1	1	2	2	1	18	1	275	275	275

Table 26 – Slot valuation scenarios tested

The parameters used in testing the model are shown in Table 27. While the cancellation cost vector λ is of a non-standard form, relative to the others used in this thesis, this is not a concern. As was discussed earlier, the Parametric formulation of this model is not concerned with the total, or even relative cost, of decisions, as it is simply trying to maximize the total value of slots offered by using all the allowable resources.

Parameter Name	Symbol	Value
Delay level	γ	4-20 minutes per flight
Cancellation level	ρ	1.0% - 5.0%
Day length	T	18 hours
Number of capacity scenarios	Q	4 scenarios
Capacity scenario probabilities	p_q	[0.03 0.49 0.40 0.07]
Capacity scenario capacities	$C_{t,q}$	*(see Chapter 2)*
Maximum delay length	U	3 hours
Number of cancellation arcs	N	3 arcs
Cancellation arc capacity	P_i	3 flights/arc
Upper bound on slot offering	D_{max}	40 slots/hour
Lower bound on slot offering	D_{min}	20 slots/hour
Slot valuation	V_t	*(see Table 26)*

Table 27 – Parameters used in slot valuation analysis

As a summary, the results in this section suggest that the model performs as it is expected to with respect to various slot valuations. In general, the profile of slots to be offered follows the profile of slot valuations used. The exception to this trend takes place at the end of the day, at which time the model is able to permit more flights to be operated, with the implicit assumption that some are likely to be delayed beyond the end of the scheduled day.

6.3.1 Scenarios A and B: Uniform Slot Values

The first two scenarios (A and B) are presented together. The slot valuation vector used for each was uniform for all hours. Upon first examination, some of the results shown in Figure 31 and Figure 32 seem reasonable, while others may seem counterintuitive.

At this point, plots of the style of Figure 31 will be introduced. The subplots in these figures show the number of slots to make available, and the relative slot values, as a function of the time of day for various combinations of delay and cancellation levels. The various rows in these plots correspond to delay levels varying from 4 to 20 minutes per flight, while the columns correspond to cancellation percentages varying from 1% to 5%. The black lines are the number of slots, while the dotted gray lines are the normalized relative slot values.

First, the general trend in these plots is that the number of slots to offer increases with time. Given that slot values are uniform, this strategy is reasonable, and should best control delays, i.e. prevent delays from propagating.

However, upon further examination, these two plots do exhibit a large amount of spurious behavior. The reason that two different plots are shown for the same slot valuation vector is that multiple optimal solutions exist for this case, each with the spurious disturbances in different places. These different plots were obtained by tweaking the branching behavior of the branch-and-bound algorithm used to solve the integer program.

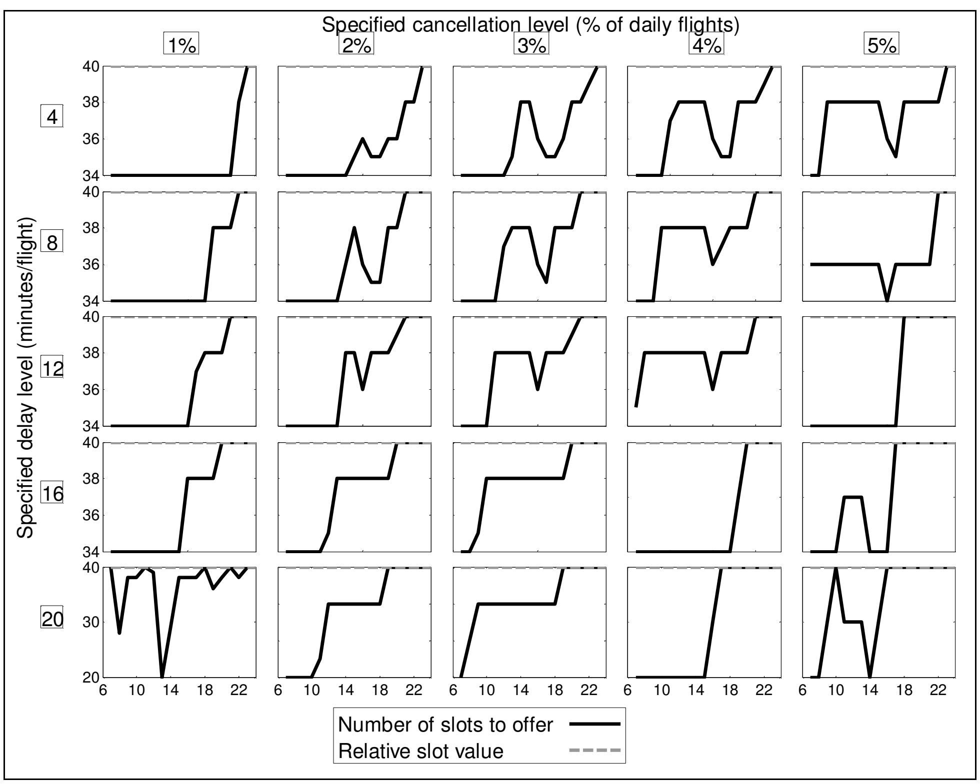

Figure 31 – Scenario A: Uniform slot values

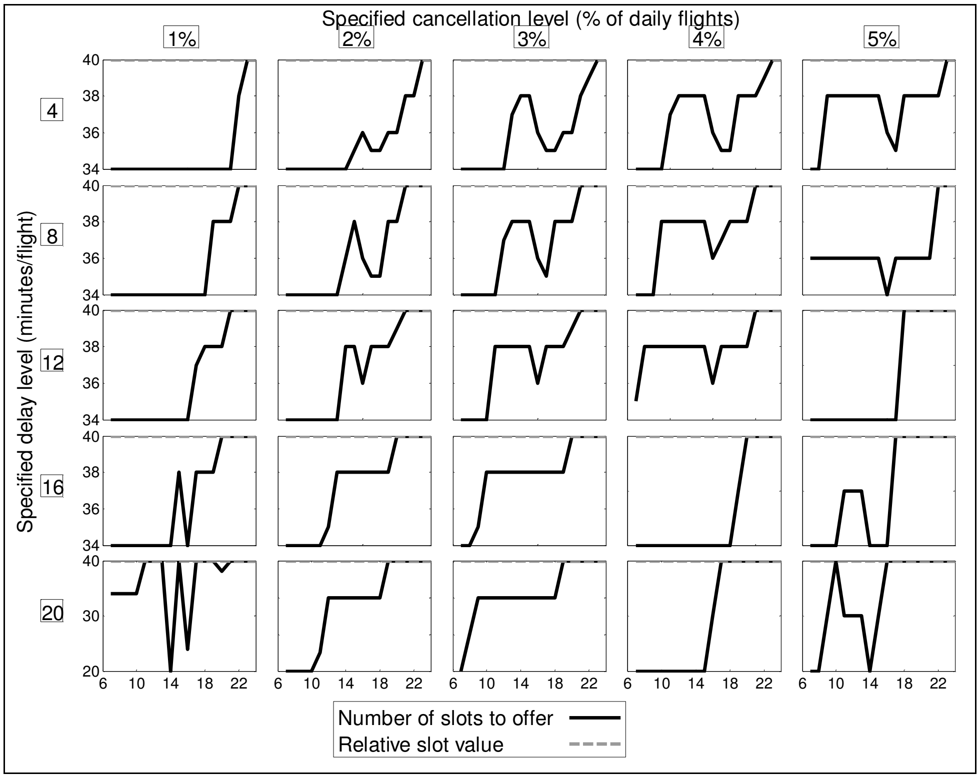

Figure 32 – Scenario B: Uniform slot values

6.3.2 Scenarios C, D, and E: Periodic Slot Values

Various sets of period slot values were tested to demonstrate the model's performance. The results of these tests are shown in Figure 33, Figure 34, and Figure 35. As was postulated previously, this model is highly sensitive to variations in slot valuation. The number of slots to be made available generally follows the period trends in slot values that were tested. The notable feature in these results, however, is that as the day progresses, the number of slots to be offered generally does not dip as far as it would earlier in the day for the same dip in slot valuation. This effect occurs because the model is able to take advantage of the three hour period at the end of the day to which it can delay flights that cannot be accommodated by the end of the scheduled day.

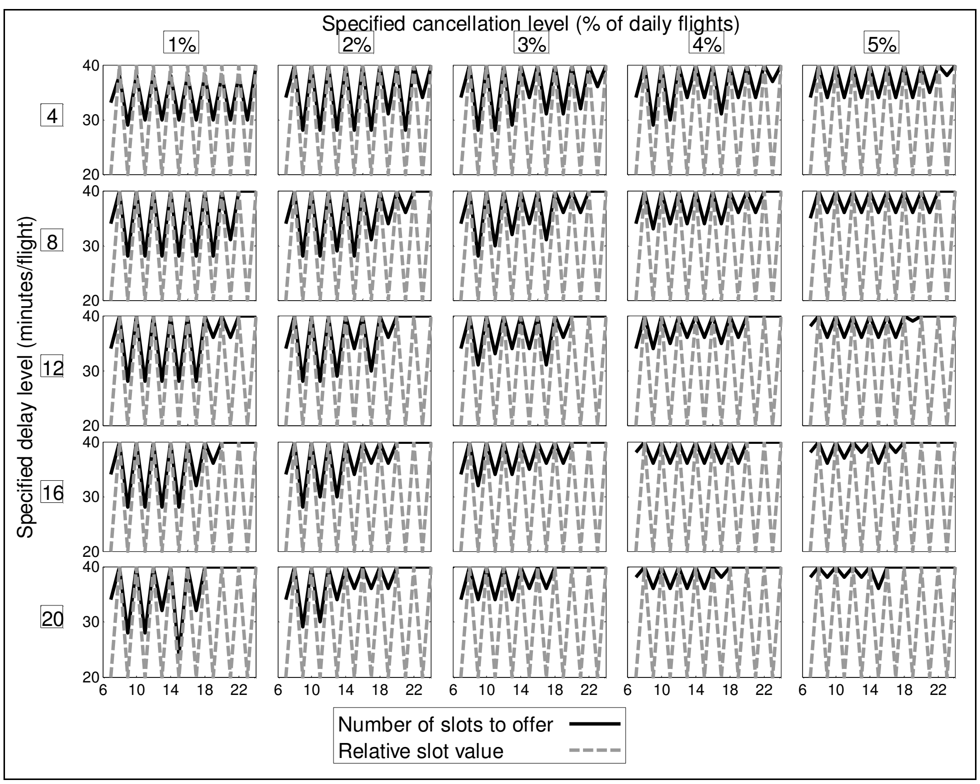

Figure 33 – Scenario C: Short period slot values

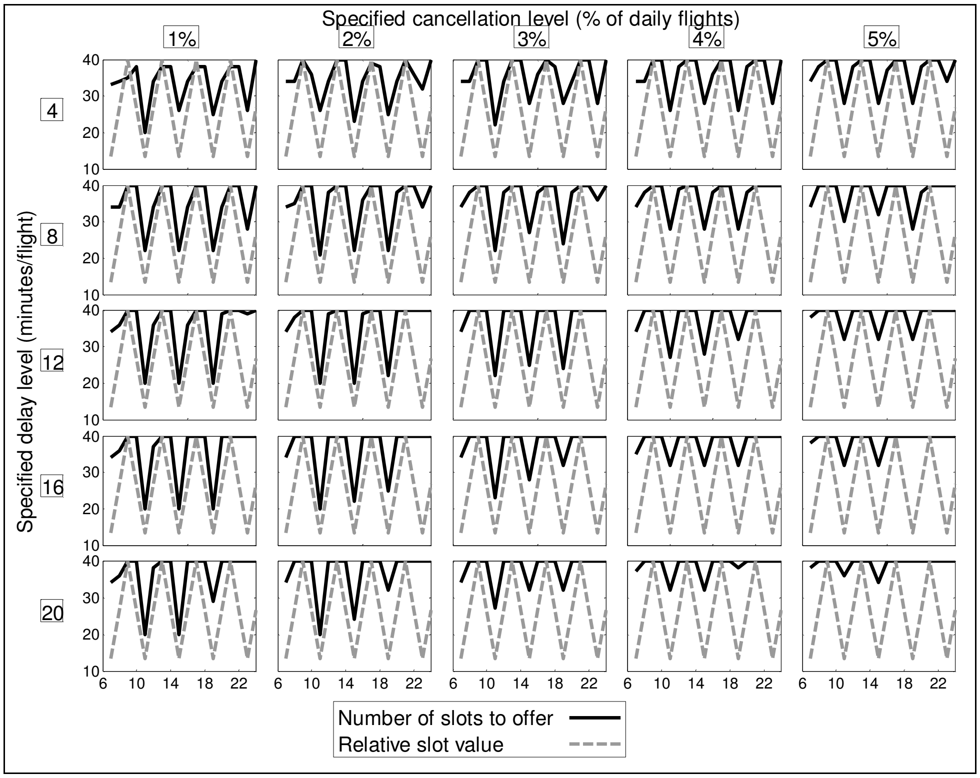

Figure 34 – Scenario D: Long period slot values

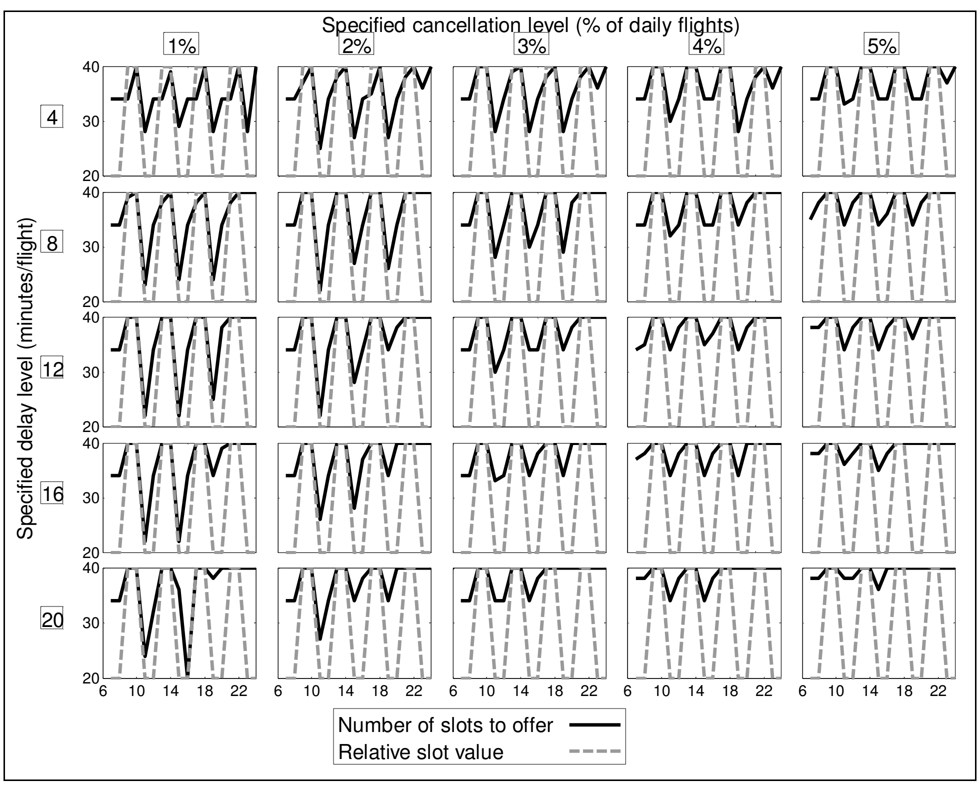

Figure 35 – Scenario E: High-low slot values

6.3.3 Scenarios F and G: Strictly Increasing/Decreasing Values

The cases of strictly increasing and decreasing slot values are shown in Figure 36 and Figure 37. As expected, the curve showing the number of slots to be offered generally increases as the time of day (i.e. slot value) increases. These results are consistent with previous expectations, but are not overly strong, because near the end of the day, the model would tend to offer more slots anyways, as it can take advantage of the unscheduled portion of the day. Spurious results are again observed in the spikes, particularly in Figure 37, likely indicating more alternate optimal solutions.

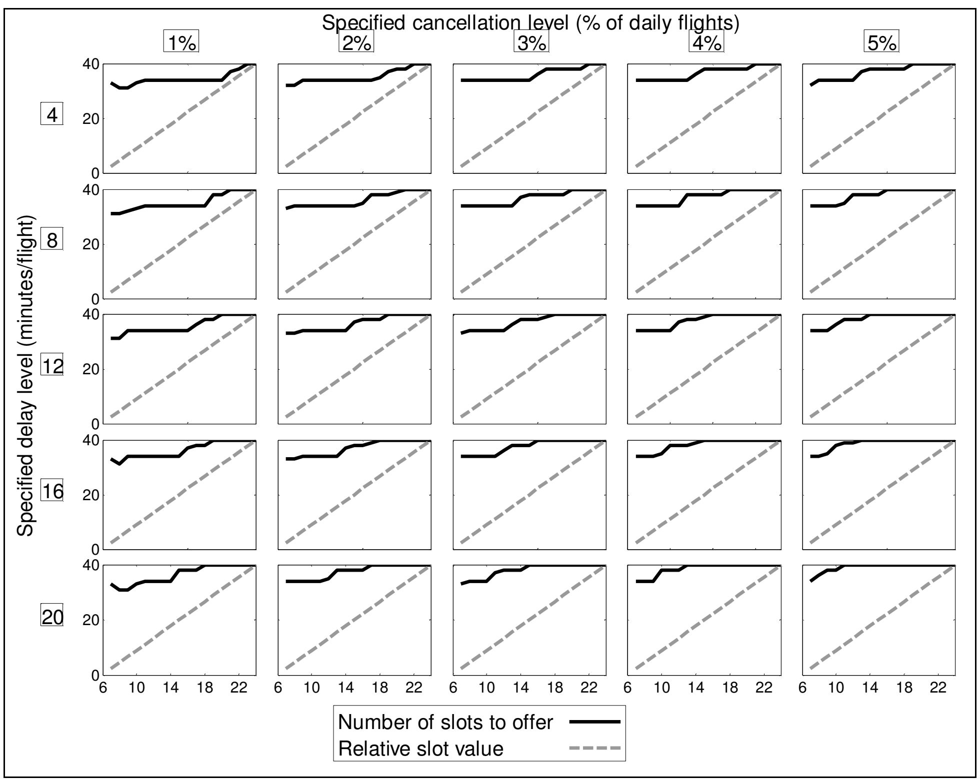

Figure 36 – Scenario F: Strictly increasing slot values

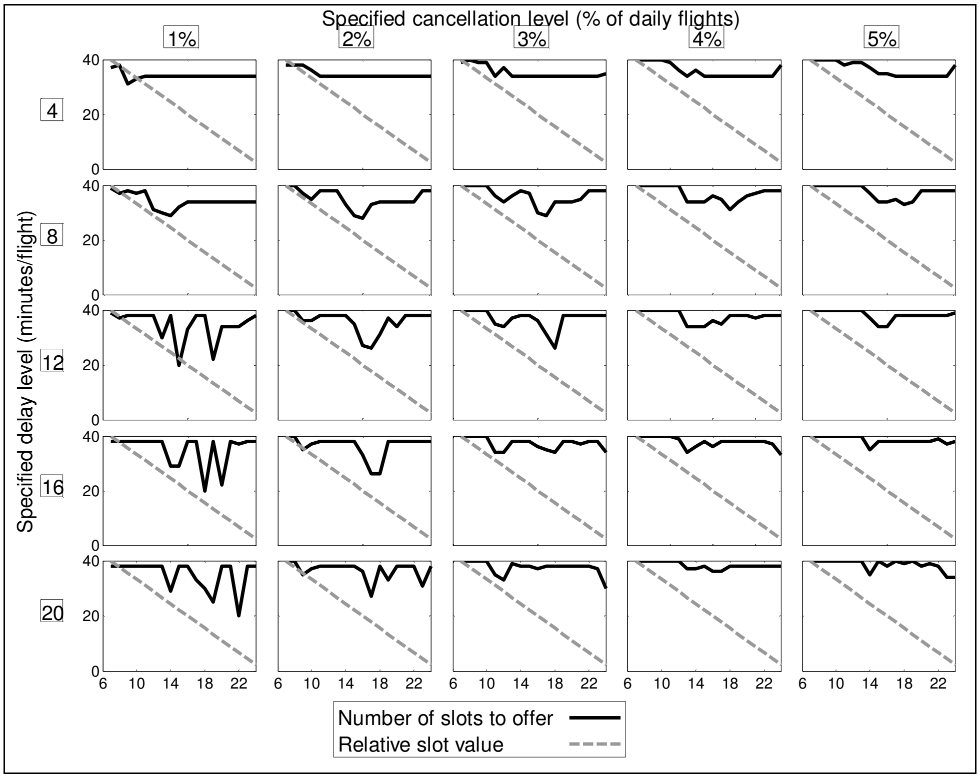

Figure 37 – Scenario G: Strictly decreasing slot values

6.3.4 Scenarios H, I, and J: Realistic Values and Perturbations

The final three slot valuation scenarios tested were chosen to examine the performance of the model in regards to perturbations to a more reasonable and realistic profile of slot valuation.

Figure 38 shows the results for the base case. In general, the profile of the number of slots to be offered follows the profile of the slot valuations. Again, the model takes advantage of the unscheduled period to allow more flights to arrive at the very end of the day.

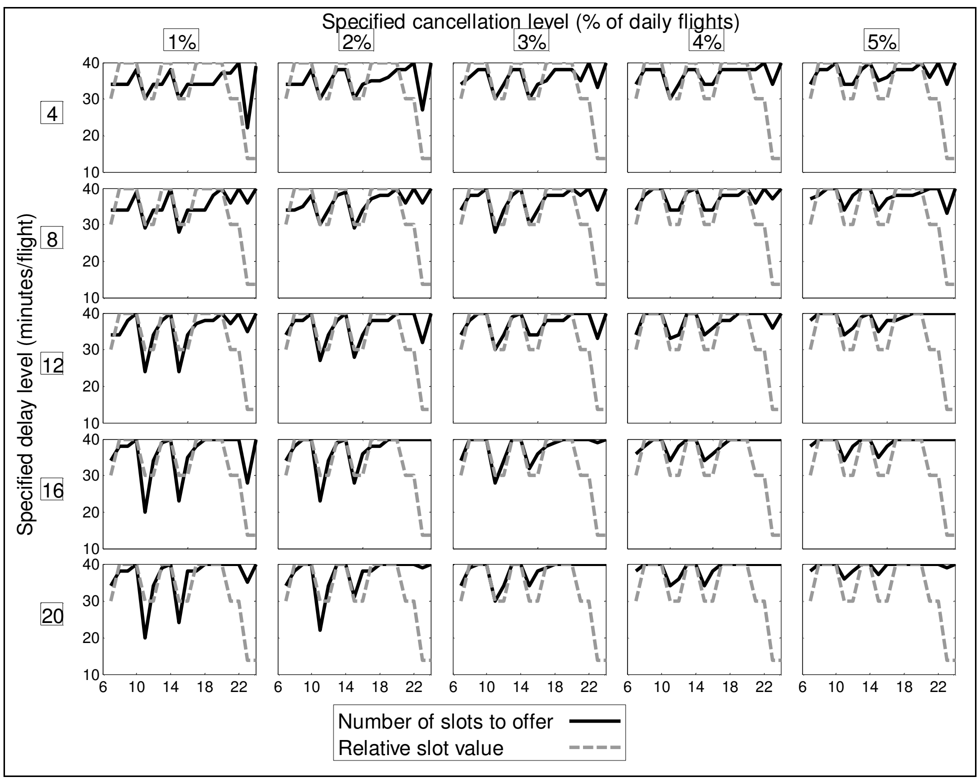

Figure 38 – Scenario H: Realistic base slot values

The case tested in Figure 39 is that in which the lower slot values were decreased by a small amount. The model responded as was expected, by lowering the number of slots to make available by a corresponding proportional amount.

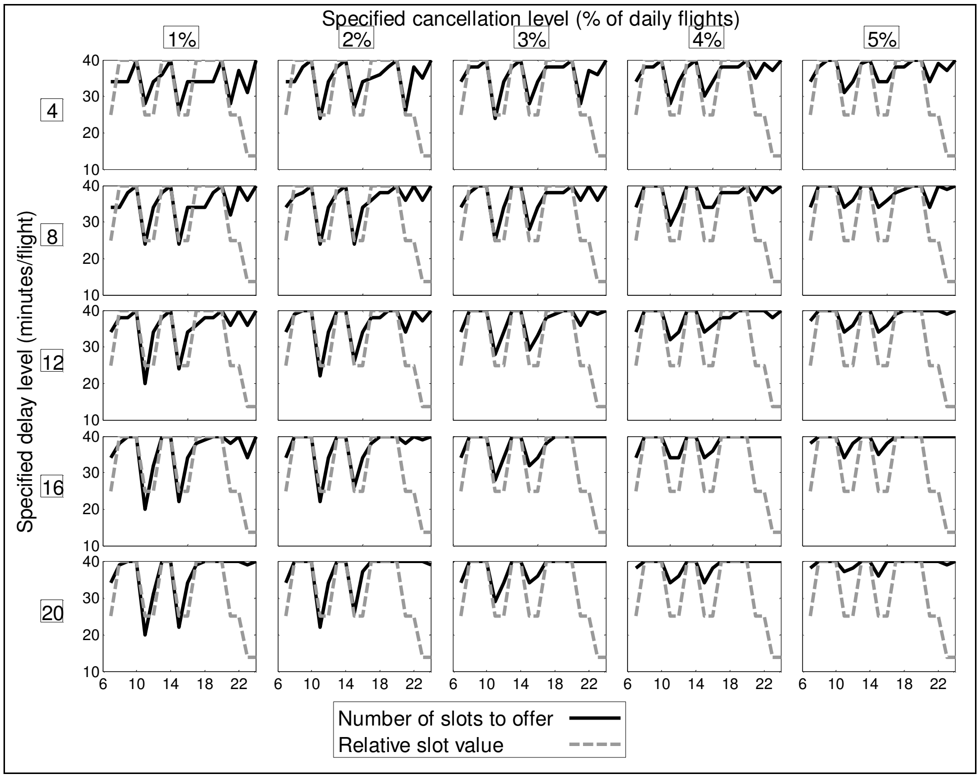

Figure 39 – Scenario I: Realistic lower bound variation slot values

The case tested in Figure 40 is that in which the highest slot values were increased by a small amount. The net effect of this change was similar to that shown in the previous case: the difference between the high and low values increased. Again, the profile of the slots to be offered follows fairly closely with the profile of the slot values.

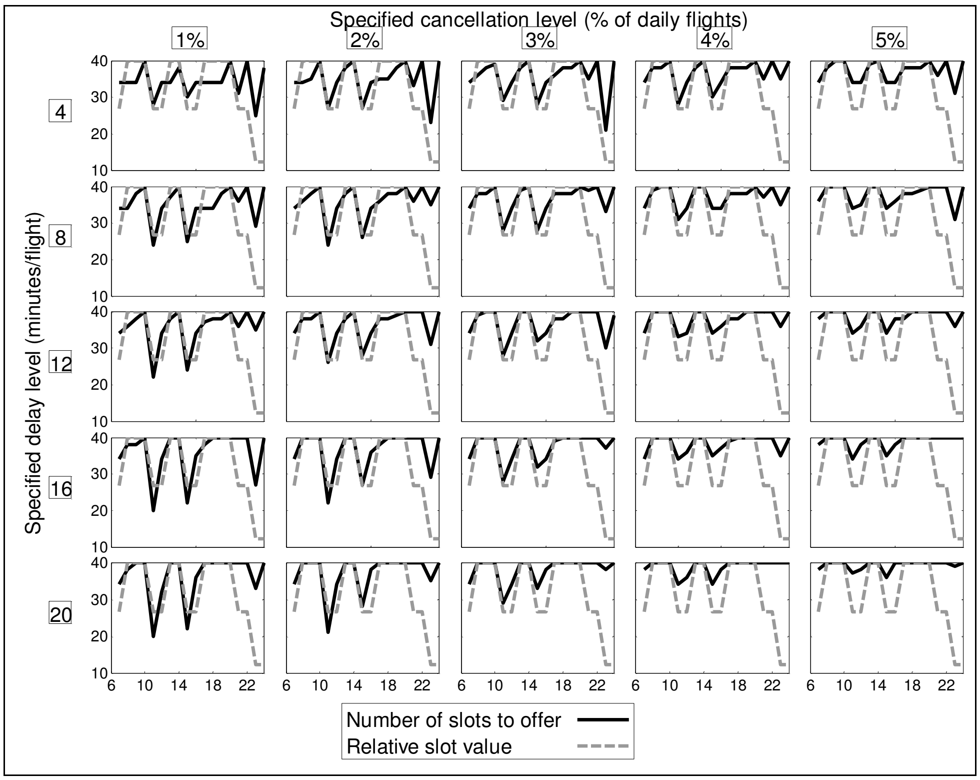

Figure 40 – Scenario J: Realistic upper bound variation slot values

6.4 Comparison of Results to Current Number of Slots

A baseline schedule, as published in the OAG, is shown in Figure 41. The reference period used is October 1, 2006 through October 6, 2006. This is the same period used in the analysis conducted by the FAA and its partners to derive the proposed rule discussed in the Federal Register (2006). The numbers shown above each bar represent the maximum number of flights scheduled in a given hour across all days in this week. In general, this maximum is the same for all weekdays in this period.

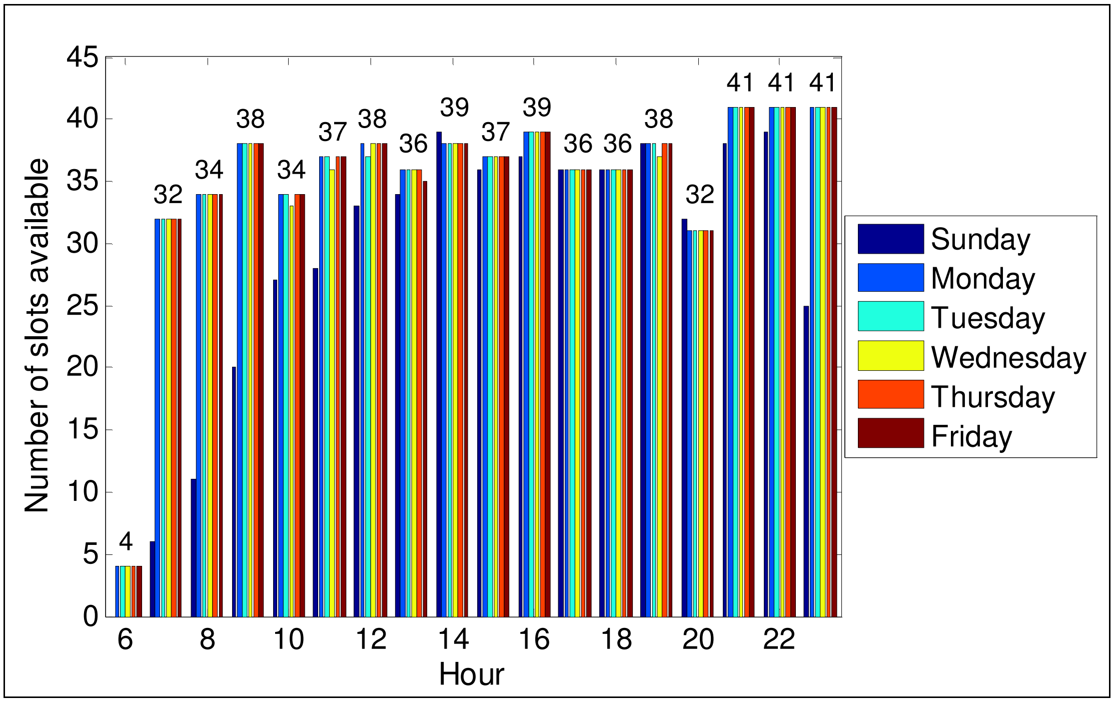

Figure 41 – Baseline LGA Schedule

Several trends are evident in this plot. The first is that the number of flights currently scheduled to arrive in the 0600 hour is much smaller than that suggested by the model. The basic reason for this is that few airports are located near enough to LGA to allow arrivals at such an early hour. This is a trend that ought to be incorporated into future analysis using this model. Another evident trend is that the number of slots at the very end of the day is fairly large, again for the reason that flights can be delayed and make use of the unscheduled overnight period of the airport for arrivals.

The best estimate for the number of slots to offer in each period- the division after the tenth year of the auction scheme proposed in Chapter 4- is shown in Figure 42. Comparing each of these two figures suggest several conclusions- each of which was postulated at the beginning of this thesis. First, some recovery period needs to be provided to help mitigate delay propagation that is not present at the current time. The model suggests this, but the current schedule does not feature it. In addition, these results suggest that, in those hours of the day that are obviously more valuable, more slots should be offered. This conclusion is best illustrated by comparing the results for the 0800, 0900, and 1000 hours in each figure. These are obviously, and necessarily, valuable hours for travelers, and as such, should be valuable for carriers. The number of slots currently offered do not permit more flights to be offered in these time periods, as do the results from the Parametric model.

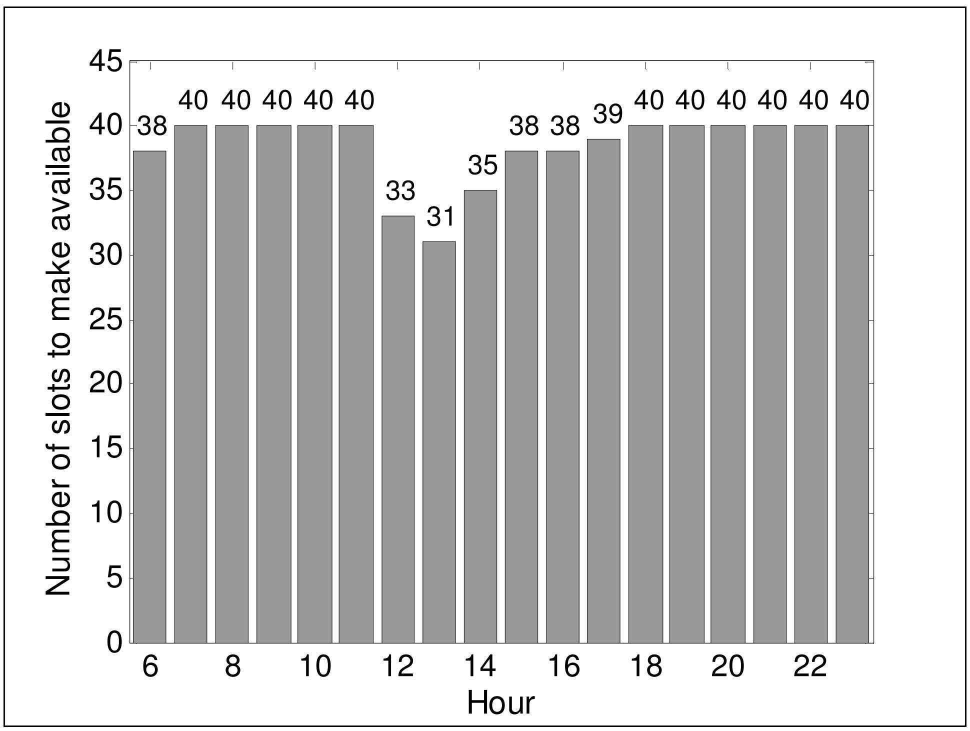

Figure 42 – Best estimate for number of slots to make available

6.5 Conclusions

In this thesis, several methods were suggested by which the number of airport arrivals slots to make available for distribution could be determined. In addition to the base methods suggested to solve this problem, several enhancements to this procedure were suggested, including methods by which the true value of airport arrival slots could be revealed, and methods by which the relationship between the number of slots to offer and their value could be estimated. The techniques suggested were illustrated using New York's highly congested LaGuardia Airport as an example.

The most important element of this thesis is not the LaGuardia Airport case study used illustratively throughout, but rather, is the methodology suggested for realistic applications. Because the problem of congested airports is one that is not likely to soon disappear, the ideas and techniques suggested in this thesis should warrant consideration at any airport which is so congested as to require formalized slot controls.

List of Abbreviations

AAR	Airport Arrival Rate
FAA	Federal Aviation Administration
HDR	High Density Rule
IP	Integer program
LP	Linear program
NAS	National Airspace System
TU	Totally unimodular
ATL	Hartsfield-Jackson Atlanta International Airport
LGA	LaGuardia Airport
ORD	Chicago O'Hare International Airport

References

Appa, G. and Kotnyek, B. (2004). Rational and integer k-regular matrices. *Discrete Mathematics*, 275, pp. 1-15.

Ball, M.O., Donohue, G.L., and Hoffman K. (2006). Auctions for the safe, efficient and equitable allocation of airspace system resources. In Cramton, P., Shoham, Y. and Steinberg, R., eds., Combinatorial Auctions, MIT Press, Cambridge, Chapter 22, pp. 507-538.

Brueckner, J.K. (2002). Airport congestion when carriers have market power. *American Economic Review*, 92, pp. 1357-75.

Carlin, A. and Park R.E. (1970). Marginal cost pricing of airport runway capacity. *American Economic Review*, 60, pp. 310-318.

Churchill, A.M., Ball, M.O., Lovell, D.J., and Mukherjee, A. (2006). Modeling Arrival Queueing, INFORMS Annual Meeting 2006, Pittsburgh, PA, USA.

Daniel, J.I. (1995). Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues. *Econometrica* 63(2), pp. 327-370.

Day, R. W. (2004). Expressing preferences with price-vector agents in combinatorial auctions. Ph.D. Dissertation, University of Maryland, College Park, MD.

DotEcon Ltd. (2001). Auctioning airport slots: A report for HM Treasury and the Department of the Environment, Transport and the Regions.

Federal Register (2006). Congestion management rule for LaGuardia Airport; Proposed Rule, 71 Federal Register 167 (August 29, 2006), pp. 51360-51380.

Gleimer, E.M. (1996) Slot regulation at high density airports - How did we get here and where are we going? *Journal of Air Law and Commerce*, 61(4), pp. 877-931.

Grether, D.M., Isaac, R.M., and Plott, C.R. (1989). The Allocation Of Scarce Resources: Experimental Economics And The Problem Of Allocating Airport Slots. Underground Classics in Economics. Westview Press.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162 (3859), pp. 1243-1248

Le, L., Donohue, G., Chen, C. (2004). Auction-based slot allocation for traffic demand management at Hartsfield Atlanta International Airport: A case study. *Transportation Research Record 1888*, pp. 50-58.

Liu, P.B., Hansen, M., and Mukherjee, A. (2005). Scenario-based air traffic flow management: Developing and using capacity scenario trees. *Transportation Research Record* 1951, pp. 113-121.

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

Mukherjee, A., Lovell, D.J., Ball, M.O., Churchill, A.M., Odoni, A.R. (2007). Estimating long-term average flight delays and cancellation rates for an airport. NEXTOR Working Paper, University of Maryland, College Park, MD.

Mukherjee, A., Lovell, D.J., Ball, M.O., Odoni, A.R., and Zerbib, G. (2005). Modeling delays and cancellation probabilities to support strategic simulations, Proceedings of 6th USA/Europe Air Traffic Management R&D Seminar, Baltimore, MD, USA.

Nemhauser, G. and Wolsey, L. (1988). Integer and Combinatorial Optimization, Wiley-Interscience, New York.

Rassenti, S., Smith, V.L., and Bulfin R.L. (1982). A combinatorial auction mechanism for airport time slot allocation. *Bell Journal of Economics*, 12(2), pp. 402-417.

U.S. House of Representatives (2000). Hearing on the slot lottery at LaGuardia Airport, December 5, 2000. Available online at http://www.house.gov/transportation/aviation/hearing/12-05-00/12-05-00memo.html.

Wolsey, L. (1998). Integer Programming, Wiley-Interscience, New York.