On Optimizing Airline Ticket Purchase Timing

WILLIAM GROVES, University of Minnesota
MARIA GINI, University of Minnesota

Airline ticket purchase timing is a strategic problem that requires both historical observations and domain knowledge to solve consistently. Even with some historical information (often a feature of modern travel reservation web sites), it is difficult for consumers to make true cost-minimizing decisions. To address this problem, we introduce an algorithm which is able to optimize purchase timing on behalf of customers. Also it can provide performance estimates of its computed action policy based on past performance. We apply machine learning to recent ticket price quotes from many competing airlines for the target flight route. Our novelty lies in extending this using a systematic feature selection technique incorporating elementary user-provided domain knowledge that greatly enhances the performance of machine learning algorithms. Using this technique, our algorithm achieves much closer to the optimal purchase policy than other proposed decision theoretic approaches for this domain. This approach meets or exceeds the performance of existing feature selection methods from the literature. Applications to other domains for this feature selection process are also discussed.

Categories and Subject Descriptors: I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent Agents*

General Terms: Algorithms, Economics, Experimentation

1. INTRODUCTION

The conventional wisdom of airline ticket purchasing states that it is generally best to buy a ticket as early as possible to avoid the risk that the price may increase. As prices do generally increase dramatically before a flight's departure, it seems generally correct. However, the earliest purchase strategy only occasionally achieves the optimal lowest cost ticket. This paper proposes a model for estimating the optimal policy for future departures. The ultimate application of this model is to autonomously make daily purchase decisions on behalf of airline ticket buyers to lower their costs.

This kind of optimal airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. Prices can vary significantly on a daily basis, and consumers have no information about pricing behaviors of particular routes and airlines. Prices do vary in this domain for a reason for several reasons. Sellers (airlines) make significant long term investments in fixed infrastructure (airports, repair facilities), hardware (planes), and route contracts. The specific details of these long term decisions are intended to roughly match expected demand but often do not match exactly. Dynamic setting of prices is the mechanism that airlines use to synchronize their individual supply and demand in order to attain the greatest revenue.

Author's addresses: W. Groves and M. Gini, Department of Computer Science and Engineering, 200 Union St SE, University of Minnesota, Minneapolis, MN, 55455, USA

DOI:http://dx.doi.org/10.1145/0000000.0000000

The airline ticket domain is characterized by adversarial risk in two contexts: the adversarial relationship between buyers and sellers, and the competitive relationships between the airlines providing an equivalent service. We assume buyers are seeking the lowest price on their travel, while sellers are seeking to keep overall revenue as high as possible to maximize profit. Simultaneously, each seller must consider the price movements of its competitors to ensure that its prices remain competitive to achieve sufficient (but not too high) demand. It is impossible to effectively optimize decision making from the buyer's point of view without also considering both types of adversarial relationships.

A central challenge in airline ticket purchasing is in overcoming the information asymmetry that exists between buyers and sellers. Airlines can mine significant databases of historical sales data to develop models for expected future demand for each flight. Demand for a specific flight is likely to vary over time and will also vary based on the pricing strategy adopted by the airline. For buyers without access to historical price information, it is generally best to buy far in advance of a flight's departure, per the conventional wisdom. However, this is not always best since airlines will adjust prices downward if they want to increase sales.

Given a corpus of historical data and the proposed learning approach, it is possible to compute policies that do much better than the earliest purchase strategy. The success of the proposed method depends on several novel contributions:

(1) We leverage a user-provided hierarchical structure applied to the features in the domain and use automated methods to decide which features to include or prune. This enables us to compute efficiently a more optimal feature set than using existing feature selection methods which use only information from the data set. Feature sets violating basic knowledge of the domain are avoided.
(2) We capture temporal trends in the model by allowing time-delayed observations to supplement or replace the most recently observed value for each included feature. The time-delayed observations are called *lagged features* in the text.

These novel aspects are applicable to many real-world multivariate domains, but this paper demonstrates the power of this technique on the domain of ticket price prediction. When comparing to a deployed commercial system, the proposed model is more informative than the output of the Bing Travel "Fare Predictor", the best commercial system currently available for airline fare prediction. This paper extends upon our previous work ([Groves and Gini 2013]) by considering situations in which a customer has specific preferences: for example, a customer may only want to consider purchasing non-stop flights from a specific airline, and would thus need a decision model targeted at a specific subset of available flights. The model here is extended to accommodate these preferences. This prediction task is more difficult than predicting only the lowest cost ticket but it is more useful for actual airline passengers. An additional benefit is that the model can provide insights into the domain: the importance of individual variables can be assessed by their presence or absence in the computed optimal model.

2. BACKGROUND AND RELATED WORK

Airlines determine the prices to offer for each flight through a process called yield management which is designed to maximize revenue given constraints such as capacity and future demand estimates (for an overview, see [Belobaba 1987; Smith et al. 1992; Elmaghraby and Keskinocak 2003]). Mismatches between airplane size and passenger demand are equalized through pricing, which can adjust demand. Choosing optimal pricing on an entire airline network is complex because there are instances (in hub-and-spoke networks) when sacrificing revenue on a particular flight can increase overall revenue of the entire network.

The current state of yield management and competition in the airline industry is a direct result of historical decisions made about regulation in the industry [Smith et al. 1992]. The techniques evolved beginning with a simple overbooking of flights. Due to regulatory changes, airlines became free to adjust the airfare for each seat without restriction. This allowed airlines to divide the seats for each flight into different “fare classes” and charge different prices for effectively the same service. The development of fare classes was critical in maximizing passenger throughput in hub-and-spoke air networks because a passenger taking a single non-stop flight will accrue a different amount of revenue than a passenger taking a longer multi-stop flight. To maximize revenue, an airline needs to be able to offer competitive fares to both types of passengers and yield management is a way to amortize these differences within the company.

In traditional yield management, the lowest air ticket prices quoted to customers are based on the available seats in each fare class for a particular flight (or origin-destination pair in the case of multi-stop itineraries). An airline can adjust the rate-of-fill for a particular flight by moving seats between fare classes (i.e. by moving high cost seats into lower cost fare classes). These decisions are traditionally made by humans who take into account previous demand, current sales, and competitive market conditions.

Yield management can be applied to other industries with properties such as the need to handle advance reservations, a range of customer values for the same product, the ability for customers to cancel, a non-negligible probability of no-shows, or stock perishable inventory [Elmaghraby and Keskinocak 2003]. Industries with these properties include hotel booking, railroad transport (linear networks, many origin-destination pairs along a shared linear route), car rental, electric utility, and broadcasting industries.

Dynamic pricing can also be beneficial in industries that can store inventory but these techniques have traditionally not been applied because of the high cost of changing prices. Instead, these industries have concentrated on aggressively tracking inventories to reduce overall inventory size and cost.

The key features enabling dynamic pricing are availability of demand data, ability to inexpensively change prices, and the use of decision support tools.

Additional market studies have addressed how the airline market has changed with the introduction of low cost airlines (LCAs). An overview of the competitive considerations in pricing strategies developed by LCAs in the European air travel market is in [Piga and Filippi 2002]. A general econometric model is developed to assess the most significant factors determining ticket prices from LCAs. The authors find that tickets purchased between 30 and 8 days prior to departure are more expensive than tickets bought in other periods. Tickets bought in the few days prior to departure can be significantly cheaper but are not always available due to demand. The econometric results show that the LCAs do not compete against conventional airlines on price alone. They also use horizontal product differentiation to minimize the necessity to compete on price. Specifically, LCAs use secondary airports (not significantly served by conventional airlines) and fly on schedules that are maximally distant from existing players. Preferences about schedule convenience and location also play a significant role in customers' purchasing decisions and ought to be considered in any predictive model in this domain.

In a later investigation ([Bachis and Piga 2011]) on measurements of market power of LCAs in the European airline market, airlines that have a significant share of the traffic at an individual airport tend to have higher prices than other carriers at the same airport. Also, an airline having a large portion of traffic between two pairs of airports (one direct route) tends to have greater market power than an airline having a large portion of traffic between two airports without a direct route. There is greater substitutability on routes with one or more stops, so market power is lower.

Some work has been done for determining optimal purchase timing for airline tickets. Our work has been inspired by [Etzioni et al. 2003], where several purchasing agents attempt to predict the optimal purchase time of an airline ticket for a particular flight. The agents determine the optimal purchase time within the last 21 days prior to departure for specific flights in their collected data set. The authors compute the purchasing policy (a sequence of wait/buy signals) for many unique simulated passengers with a specific target airline, target flight, and date of departure to satisfy. The optimal policy (the sequence of buy/wait signals that leads to the lowest possible ticket price) is used as a benchmark for each simulated passenger and the cost of each alternative purchasing agent is computed. The aggregate result shows that, given these purchasing criteria, it is possible to save a significant amount when purchasing. We understand that Bing Travel's "Fare Predictor" is a commercialized version of the models in [Etzioni et al. 2003], so real-world results from this form a benchmark for our results.

Our work extends the state-of-the-art because we directly compute a policy for finding the minimum cost ticket of *any* flight from *any* airline given a route and departure date. This is a more difficult problem because the aggregate minimum price varies less than the price of an individual flight from an individual airline. Our paper goes beyond their work in several ways: the model is not limited to a single flight, the purchases are up to 60 days before departure (instead of 21 in the existing work), the model is compared against realistic financial benchmarks (including buying as early as possible), and the model provides a regression estimate for the expected best price between the current date and departure.

There are several efforts in the game theory community to model aspects of the airline ticket domain, usually for the purpose of understanding competitive market dynamics of the oligopoly of sellers. In [Subramanian et al. 1999], a dynamic programming model is presented for determining optimal fare class allocation (of four fare classes) on a single flight. The major contribution of this model lies in incorporating fare class-dependent and time-dependent cancellation, overbooking, and no-show probabilities. Valuable insights provided by this study are that booking limits need not change monotonically over time (may increase or decrease), it may be optimal to accept a lower fare class while simultaneously rejecting a higher fare class (due to differences in cancellation characteristics), and cancellations cause the optimal policy to depend on both total capacity and remaining capacity.

A one-shot game theory-based simulation of pricing competition in the airline ticket price domain is presented in [Isler and Imhof 2008]. When two airlines with significant capacity compete with each other and their products are not sufficiently differentiated, the equilibrium price falls to a minimum price threshold, referred to as the "spiral down" price. This result may shed some light on the long term decisions airlines make about airplane size and flight frequency. The authors also note that the airline pricing domain is more similar to a repeated game than a one-shot game. Other equilibria can be enforced in repeated games that are significantly above the spiral down price found in the non-repeated game. This work also shows that a completely automatic pricing mechanism can be potentially ruinous for an airline. There must be supervisory mechanisms that take into account other aspects into pricing beyond price competition.

A game theory model of dynamic pricing that incorporates an oligopoly facing strategic customers, buyers who will delay purchase until a future time period if there is a high likelihood of price decreases lower price later, is presented in [Levin et al. 2009]. If even a portion of the population of customers is strategic, revenue is reduced for the sellers and any strategic defenses in such a transparent market cannot fully amelio-

Table I. Airline price quote specifications for all airlines from specific 5-day round trips. The exact dates and cities shown are for illustration purposes only.

	Example 1	Example 2
Quote Date:	13 May 2011	13 May 2011
Origin City:	SEA	NYC
Destination City:	IAD	LAX
Departure Date:	20 May 2011	20 May 2011
Return Date:	25 May 2011	25 May 2011
No. of itineraries returned:	1135	1304
No. of airlines quoting itineraries:	9	13

rate this effect. The critical conclusion of this work is that the most effective method to inhibit the impact of strategic consumers is to reduce the amount of information available to consumers. This may explain in part why, in spite of the technical feasibility, few predictive tools are available to individual purchasers of airline tickets.

From a technical perspective, this prediction problem can be phrased as a machine learning problem involving both many features (possible variables which are relevant to prediction) and temporal trends. In the literature, there are many exclusively data-driven feature selection techniques applicable to this domain [Molina et al. 2002]. [Hall 2000] presents the CFS algorithm (Correlation-based Feature Selection) to perform a filter-based feature selection using a merit heuristic (normalized Pearson's Correlation). The algorithm starts with an empty feature set and uses forward best-first search to incrementally add features. Wrapper-based methods employing search (such as best first search (BFS)) using an underlying machine learning algorithm have also been employed [Kohavi and John 1997]. Both techniques are included in the results for benchmarking.

3. DATA SOURCES

The data for our analysis was collected as daily price quotes from a major travel search web site between Feb. 22 and Jun. 10, 2011 (109 observation days). A web spider was used to query for each route and departure date pair in our study, so the results are representative of what a customer could observe in the market.[1] This set is split sequentially into 3 data sets with the following lengths: 48, 20, and 41 days. The three periods are utilized as the training set, calibration set, and test set, respectively. Each query returned, on average, 1,200 unique round-trip itineraries from all airlines; most queries returned results from more than 10 airlines. Example queries are in Table I. Feature values are aggregate statistics computed from each day's itineraries. For consistency, the queries were initiated at the same time each day.

Bing Travel, a popular travel search web site, has a "Fare Predictor" tool that provides a daily buy/wait policy recommendation for many routes and departure dates. These recommendations were obtained daily from the site for the test set period and are directly compared with our agent's policy.

3.1. Pricing Patterns in Historical Data

There are strong cyclic patterns in the time series of prices. For example, Figure 1 shows the mean lowest price quoted by all airlines for a specific origin-destination pair for 2 months of 5-day round trip itineraries departing on (a) Thursdays and (b) Mondays. The Thursday to Tuesday itinerary time series shows a regular drop in

[1] All data sets are available, upon request, from the authors.

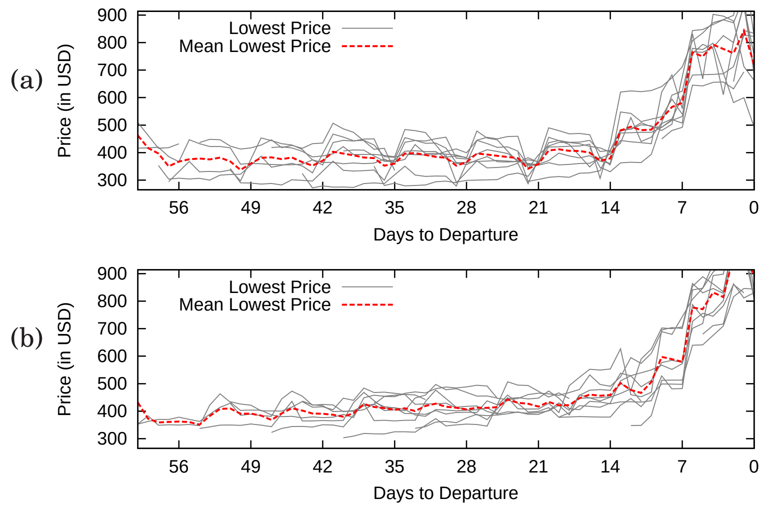


Fig. 1. Mean lowest price from all airlines (a) for New York City (NYC) to Minneapolis, MN, USA (MSP) 5-day round trip flights having Thursday departure and Tuesday return (Th-Tu) or (b) for NYC to MSP M-F itineraries. Each solid line series indicates quotes for a different departure (8 departure dates in each graph). A dotted series indicates the mean.

prices for Tuesday, Wednesday, and Thursday purchases ((*days-to-departure* modulo $7) \in \{0, 1, 2\}$), while the (b) series shows significant increases for Thursday, Friday, and Saturday purchases. As expected, both exhibit price increases in the last few days before departure (*days-to-departure* ≤ 7) but series (b) exhibits this increase earlier in the time series. We posit that the majority of business flights would be Monday to Friday itineraries, and thus demand for series (b) flights would be less sensitive to price increases than leisure flights. The weekly depression in costs in (a) may be due to market segmentation: customers buying mid-week are more price sensitive than weekend purchasers.

The pricing behaviors exhibited for other origin-destination pairs also differ. A high traffic origin-destination pair such as the New York City to Los Angeles route (shown in Figure 2) exhibits much weaker cyclic patterns. We conjecture that strategic pricing is likely to have a much greater observed effect for routes that have relatively few (2 or 3) competing airlines than for routes with a large number of competitors (> 3).

3.2. Observed Pricing Relationships

In this section, we characterize the pricing relationships between airlines that are observed in the collected pricing data.

The prices quoted each day for a specific query (such as the two examples in Table 1) often vary significantly by airline, but the prices observed have structural relationships which can be leveraged for prediction. Figure 3 plots the minimum price time series for each airline from four weeks of data for a specific itinerary (depart MSP on May 5, 2011, return NYC on May 10, 2011). Pricing patterns for competing airlines have been covered empirically in the literature ([Bachis and Piga 2011]) and this example illustrates these relationships. Airlines can be divided into two categories: low cost airlines (LCAs) and ”legacy” airlines (MCAs and HCAs). LCAs use their primary advantage, the ability to offer lower ticket prices due to lower internal costs, to compete aggressively against legacy airlines. Legacy airlines use other benefits to compete

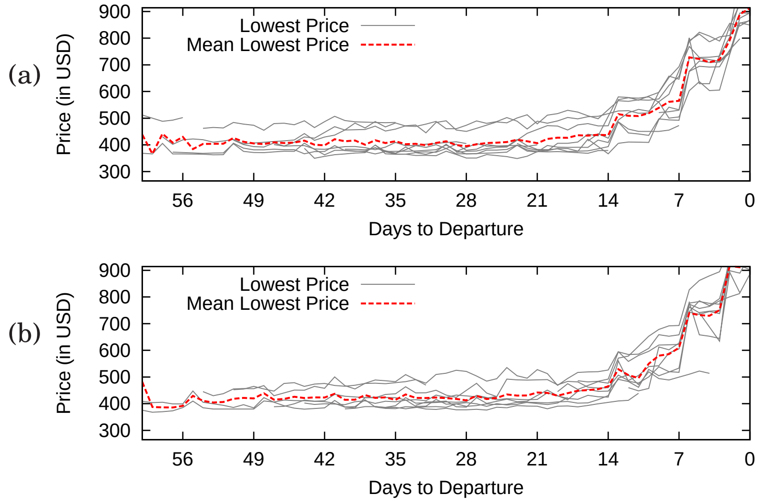


Fig. 2. Mean lowest price offered by all airlines for New York City (NYC) to Los Angeles (LAX) 5-day round trip flights having (a) Thursday departure and Tuesday return, or (b) Monday departure and Friday return itineraries.

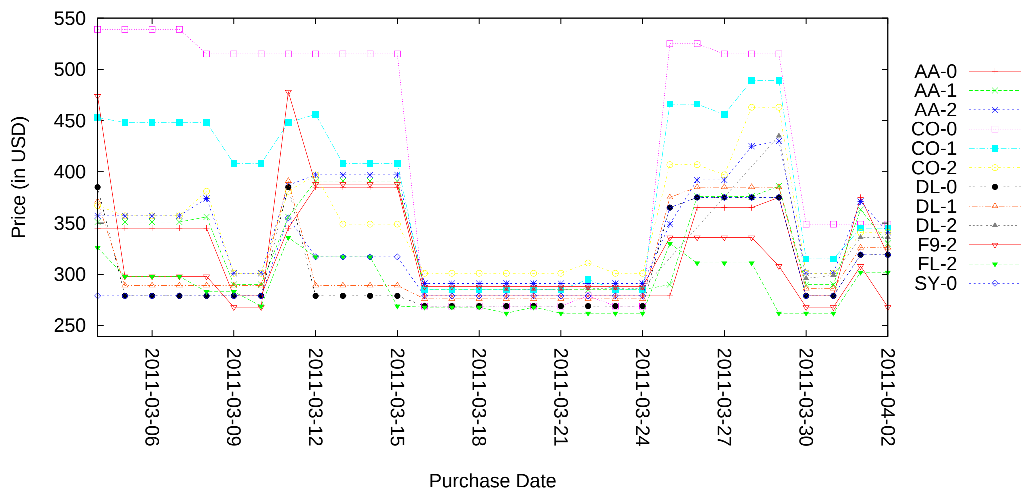


Fig. 3. A price time series for many airlines quoting prices for the NYC-MSP route for Thursday departure (May 05, 2011) with Tuesday return. Each series represents the minimum price of the day's quotes for a specific airline and number of stops pair: for example, **DL-0** refers to all Delta Airlines, non-stop flights and **AA-2** refers to all American Airlines, 2 intermediate stop flights.

against LCAs including: greater availability of departure times, a larger network of connecting flights, and loyalty rewards programs.

For a specific route's prices (shown in Figure 3), the competing airlines can be divided into one of three categories based on the pricing behavior: LCA, MCA, and HCA. Low cost airlines (LCAs) tend to always compete on price and will consistently offer prices at or below all other types. A LCA may even compete against other LCAs by lowering prices further in order to increase demand for their product. MCAs, or medium cost airlines, are legacy airlines that tend to price aggressively for the route but will rarely

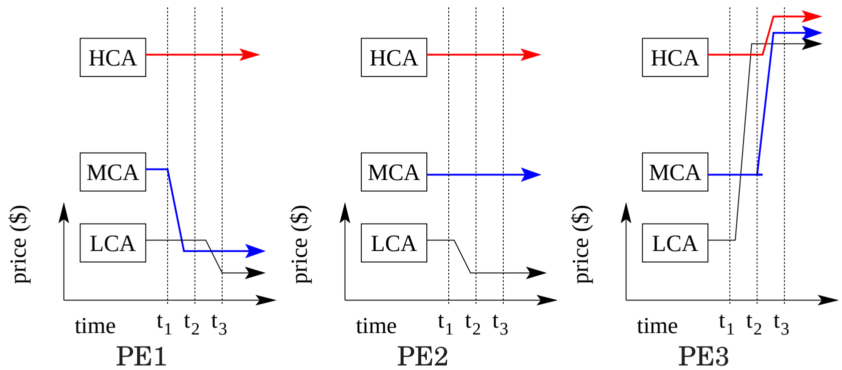


Fig. 4. Stylized diagrams of changes in price equilibria. Three price behavior types are shown: LCA refers to low cost airlines, MCA refers to medium cost airlines, and HCA refers to high cost airlines.

set prices below the best LCA price. HCAs, or high cost airlines, are legacy airlines that do not compete aggressively based on price but will still quote (usually high) prices for the route. Customers may still buy these higher priced offers (from MCAs or HCAs) because of the other benefits of the specific airline or itinerary.

In this route, the airlines can be categorized based on behavior as:

— LCA: F9, FL, SY
— MCA: DL
— HCA: AA, CO

Some airlines will quote itineraries with different number so intermediate stops. For example, Delta airlines (DL) quotes itineraries with no intermediate stops (non-stop, as DL-0), 1 intermediate stop (DL-1), or 2 intermediate stops (DL-2). For the purposes of the analysis, we treat each itinerary type separately as shown in Figure 3.

The price time series shows repeated patterns among the airlines for this departure. The LCAs are consistently at the bottom of the price range and there is a cluster of MCAs that have similar prices during periods of price stability. To characterize some of the price shifts, Figure 4 provides some examples. In PE1, the MCA lowers its price below the prevailing LCA price; on the next day, the LCAs adjust their prices to match or beat the best price. In PE2, an LCA lowers its price, but no other airline follows. In PE3, the LCA raises its price and the MCAs and HCAs adjust their prices upward as well.

From a statistical analysis of the data, it is possible to make these categorizations. These persistent relationships can be leveraged using a regression model because, within a route, the relationships between airlines will be persistent. A machine learning model, such as those described later in Section 4.3, can leverage these to make predictions about future prices.

4. PROPOSED MODEL

When constructing prediction models for real-world domains, practical complexities must be addressed to achieve good prediction results. Typically, there are too many sources of data (features). Limiting the set of features in the prediction model is essential for good performance, but prediction accuracy can be lost if relevant inputs are pruned. This is even more acute in situations where the number of observations is limited, often a feature of real-world domains. To meet this challenge, we construct a prediction model that involves the following distinct steps, which we then describe in detail:

(1) Feature Extraction – The raw data observed in the market are aggregated into a fixed length feature set.
(2) Lagged Feature Computation – A lag scheme is computed using a hierarchy of the features that incorporates some domain knowledge.
(3) Regression Model Construction – Using the augmented feature set generated from the lag scheme, a regression model is generated using partial least squares (PLS) regression.
(4) Policy Computation – A search of decision threshold parameters is done to minimize calibration set cost.
(5) Optimal Model Selection – For each candidate model computed using the previous steps, the one which performs best on the calibration set is chosen. The final performance is estimated on the test set.

Table II. Raw Features

Class	Size	Variable List
Det	8 vars	Days-to-departure, Quote DoW is Mon, Quote DoW is Tue, ..., Quote DoW is Sun
All-A	3 vars	ALL-min-A, ALL-mean-A, ALL-count-A
All-S	9 vars	ALL-min-0, ALL-mean-0, ALL-count-0, ALL-min-1, ALL-mean-1, ALL-count-1, ALL-min-2, ALL-mean-2, ALL-count-2
Each-A	18 vars	DL-min-A, DL-mean-A, DL-count-A, ..., OTHER-min-A, OTHER-mean-A, OTHER-count-A
Each-S	54 vars	DL-min-0, DL-mean-0, DL-count-0, DL-min-1, DL-mean-1, DL-count-1, DL-min-2, DL-mean-2, ...

Note: Raw features by feature class for each quote day on a specific departure day and route. The number of features in some classes (EACH-A, EACH-S) will vary based on the number of airlines quoting the route. The counts given are specific to the NYC-MSP route (92 total raw features). Features are named as "airline-statistic-#ofStops": i.e. ALL-min-A = minimum price quoted by any airline, ALL-min-0 = minimum price quoted by any airline for non-stop flights only, DL-min-A = minimum price quoted by a specific airline (DL = Delta Airlines), and Quote DoW is Mon = Boolean variable (1 if quote is retrieved on Monday).

4.1. Feature Extraction

The large number of itineraries (>1000) in each daily query made some data aggregation necessary. The features extracted for prediction are aggregated variables computed from the (large) list of quotes observed on individual query days. For each query day, there are possibly many airlines quoting flights for a specific origin-destination and date combination. This is possibly due to strategic decisions of the airline or due to lack of available capacity. We limit the number of airlines used for distinct features by focusing on airlines that quote for a specific route more than 40% of the query days. Also, each airline may present itineraries that contain non-stop segments or segments with one or more stop. We divide the quotes by number of stops into three bins: non-stop round trips, round trips with a maximum of one stop in each direction, and round trips with 2 or more stops in either direction. For each bin, three features are computed: the minimum price, mean price, and the number of quotes. Additionally, these three features are computed for the union of all three bins. So for each airline, 12 features are computed on each quote day. For airlines not exceeding the 40% criteria, their itineraries are combined into a separate "OTHER" category placeholder. Finally, these same 12 aggregates are generated for all itineraries and are placed in the "ALL" airlines category. Boolean variables indicating the query's weekday are added. A *days-to-departure* (number of days between the query and departure dates) value is computed based on the departure date.

A listing of the features for each query day is shown in Table II. Each of the 92 features is in a class based on its specificity using the feature class hierarchy in Figure 5.

4.2. Lagged Feature Computation

Using only the most recent values (92 features for the NYC $\rightarrow$ MSP route) as the entire feature set may provide reasonable prediction results in some domains, but such a model cannot predict trends or temporal relationships present in the data. This simple scheme is shown in Table III(b) and in the results as *Minimal Lag Scheme*. The need to represent temporally-offset relationships (such as weekly cycles or trends) motivates adding time-delayed observations to the feature set as well. We refer to this as the addition of *lagged features*. For instance, if the cost of a route on day $t-7$ is representative of the price available on day $t+1$, the 7 day delayed observation should have a high weight in the model. A regression model which includes all time-delayed instances up to a depth of n days of all features can produce good results, but the inclusion of many variables into the model can result in poor performance. A diagram of this model is in Table III(a). The performance of a model of this construction ($n = 7$) is shown in Table VI as the model *Full Lag Scheme*.

Table III. Basic lag schemes used for benchmarking of feature selection methods. The dots "•" indicate the feature class at the corresponding time lag is included in the feature set provided to the learning model.

(a) maximal scheme

Class	Lagged Offsets 0	1	2	3	4	5	6	7
DET	•							
ALL-A	•	•	•	•	•	•	•	•
ALL-S	•	•	•	•	•	•	•	•
EACH-A	•	•	•	•	•	•	•	•
EACH-S	•	•	•	•	•	•	•	•

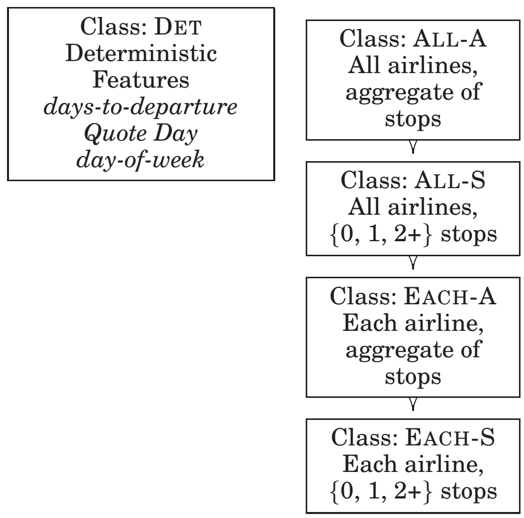
(b) minimal scheme

Class	Lagged Offsets 0	1	2	3	4	5	6	7
DET	•							
ALL-A	•							
ALL-S	•							
EACH-A	•							
EACH-S	•							

Our technique uses the assumption that more recent observations are likely to have high informational value for price prediction, but time-delayed features may hold informational value as well (i.e. the environment is not completely stochastic). The time between a change in the market and its effect on the target variable may be longer than one day and lagged variables can leverage those delayed relationships.

By examining all combinations (a small number) of feature classes it is possible to automatically tune the feature vector to achieve better results. Another constraint is added: more specific feature classes will not have more time delayed instances than more general classes. We posit that time-delayed observations from the target variable (such as the all airline minimum price in *Class* ALL-A) are likely to be most predictive. Time-delayed observations from other more-specific feature classes *may also be* but

Fig. 5. Lag scheme class hierarchy for product price prediction. Arrow denotes a *subset* relationship (i.e. class ALL-A should have an equal or greater set size than class ALL-S).



are less likely to be predictive. It is by this principle that the hierarchy and strict ordering of lagged data additions are based. By constraining the classes so that the less information-dense classes contribute fewer features, we prevent the inclusion of extraneous, irrelevant features.

Next, time lagged data is reformatted to form the augmented feature set, which is called a *lag scheme expansion*. A search of possible configurations is performed to find the best performing configuration for a target (the optimal choice may be different for each route).

The inclusion of a little domain knowledge using feature classes and a class hierarchy to constrain the search for high-performance feature sets allows for significant reductions in the number of possible feature set configurations. The number of lag schemes as formulated with the hierarchy in Figure 5 for a maximum time delay of 7 days is 8517. Without the constraints between classes, there are configurations of the 5 feature classes if constrained to possible time delays of $\{\varnothing, 0, 1, 2, \ldots, 7\}$, but many configurations will be uninteresting variants. Finally, without the hierarchy and constraints between classes, there are $\approx 10^{82}$ configurations of the 92 original features.[2] Using both the feature classification and the constraint hierarchy allows for a greater variety of "interesting" lag schemes to be tested for the same search effort.

While a domain expert could design a high-performance feature set, the automated lag scheme search should contain a configuration similar to what a domain expert can build. Also, the results of the optimal lag scheme search can elicit some surprising relationships in the data. Table IV shows optimal lag schemes for several targets. It is interesting to note that non-stop targets in Table IV(b, d) benefit from a larger feature set (both in temporal depth and feature class breadth).

[2] 84 price features, and 8 deterministic features (days-to-departure and quote weekday) $= 2 * (2^7) * (9^{84})$

Table IV. Optimal lag schemes of a domestic route and an international route for a 5-day trip with Monday departure. The dots "•" indicate the feature class at the corresponding time lag is included in the best performing feature set.

(a) New York → Minneapolis

Class	Lagged Offsets 0	1	2	3	4	5	6	7
DET	•							
ALL-A	•	•	•	•	•	•	•	•
ALL-S		•	•	•	•	•	•	
EACH-A				•				
EACH-S								

(MC: \$280, EP: \$317, 75.4%)

(b) New York → Minneapolis (non-stop only)

Class	Lagged Offsets 0	1	2	3	4	5	6	7
DET	•							
ALL-A	•	•	•	•	•			
ALL-S				•				
EACH-A				•				
EACH-S				•				

(MC: \$365, EP: \$414, 66.8%)

(c) New York → Hong Kong

Class	Lagged Offsets 0	1	2	3	4	5	6	7
DET	•							
ALL-A	•	•	•	•	•	•	•	•
ALL-S								
EACH-A								
EACH-S								

(MC: \$1190, EP: \$1207, 28%)

(d) New York → Hong Kong (non-stop only)

Class	Lagged Offsets 0	1	2	3	4	5	6	7
DET	•							
ALL-A	•	•	•	•	•	•	•	•
ALL-S	•	•	•	•	•	•	•	
EACH-A	•	•	•	•				
EACH-S	•	•	•	•				

(MC: \$1404, EP: \$1416, 13%)

4.3. Regression Model Construction

Mathematically, PLS regression deterministically computes a linear function that maps a vector of the input features x_i into the output variable y_i (the label) using a vector of weights $\bar{w}$. Several implementations of PLS exist [de Jong 1993; Martens and Næs 1992]; each with its own performance characteristics. We use the orthogonalized PLS, Non-Integer PLS (NIPALS), implementation in [Wold et al. 1983]. PLS was chosen over similar multivariate techniques including multiple linear regression, ridge regression [Hoerl and Kennard 2000], and principal component regression (PCR) [Jolliffe 1982] because it produces better performance than the others and has the ability to adjust model complexity.

This algorithm has multiple advantages. First, PLS regression is able to handle very high-dimensionality inputs because it implicitly performs dimensionality reduction from the number of inputs to the number of PLS factors. Second, the model complexity can be adjusted by changing the number of PLS factors to use in computing the regression result. This value is adjusted in our experiments to determine the optimal model complexity in each prediction class. Third, the algorithm is generally robust to highly collinear or irrelevant features. Fourth, the structure of a trained model can be examined for knowledge about the domain. For these reasons, this algorithm was chosen.

PLS regression allows users to adjust the model complexity by selecting the number of PLS factors to generate when training. These factors are analogous to the principal component vectors used in principal component regression. The number of PLS factors determines the dimensionality of the intermediate variable space that the data is mapped to (a limit of 20 factors is used in these results). The computation time does not significantly increase for a larger number of factors but the choice can effect prediction performance: too few factors can cause the model to be unable to represent relationships in the data.

Other machine learning algorithms can also be used in place of PLS. Experiments using support vector regression (nuSVR, [Schölkopf et al. 2000]), ridge regression, and decision trees (REPTree, [Witten and Frank 2005]) are also shown for comparison.

4.4. Policy Computation and Evaluation

An obvious approach to choosing a good regression model (lag scheme and trained machine learning model) is to use the model with the highest prediction accuracy, but it may not be the model that generates the lowest average cost policy. Instead, we propose to grade the models by measuring the cost that results from following the computed policy recommendation. To use the regression output (an expected future price) to compute an action policy, we introduce the concept of a decision threshold function. Given $\hat{e}_t$, the model estimate future price at time t, the current observed price p_t and the current number of days-to-departure d_{dtd} (an integer), the current action policy $r_t \in \{\text{BUY}, \text{WAIT}\}$ is computed by Equation 1.

$$r_t = \begin{cases} \text{BUY} & : \hat{e}_t > p_t * (c + (1/30) * s * d_{dtd}) \\ \text{WAIT} & : \text{otherwise} \end{cases} \tag{1}$$

The two parameters c and s are expressed as decimals. Intuitively, c can be thought of as an adjustment in the likelihood of a buy signal. Values of $c > 1.0$ correspond to a policy that is only likely to emit BUY when the current price is far below the expected future price. This situation indicates the current price is a bargain for the customer. The parameter s corresponds to the percent change in the threshold per 30 days of advance purchase (0.02 corresponds to a 2% change in the threshold at $d_{dtd} = 30$). Values of $s > 0.0$ generate a policy more likely to WAIT when far from departure. When a departure is far in the future and $s > 0.0$, the agent is more likely to wait until a highly favorable (low) price appears before deciding to purchase. Adjusting these two parameters can be thought of as determining the optimal level of risk depending on the current price and the degree of advance purchase. The range of c and s values searched was $[0.7, 1.3]$ and $[-0.1, 0.1]$, respectively, in increments of 0.01.

We use this two-parameter approach to make decisions, because it is simple, works well, and provides an intuitive understanding of the policy computation. This is not to rule out more sophisticated approaches, such as reinforcement learning. We leave exploration of this aspect for future work.

4.5. Optimal Model Selection

The proposed search of lag schemes is exhaustive, but due to the feature class hierarchy, the number of configurations is relatively small and can be fully explored. A model is constructed for each potential lag scheme: first, for each lag scheme a pricing model is generated using the training set data, then the decision threshold parameters (c and

s) are calibrated on the out-of-sample calibration set that results in the lowest average ticket price. Performance is measured by scoring the model on the test set.

5. EXPERIMENTAL RESULTS

The experiments were designed to estimate real-world costs of using various prediction models to develop a purchase policy. A survey of the literature revealed that: airlines assume a relatively fixed rate of purchases until a flight is full, and most tickets for a flight are sold within 60 days of departure [Belobaba 1987]. Using these facts, we measure performance as the cost of following the purchase recommendations for a specific departure once for purchases between 1 and 60 days before departure ($dtd \in \{1, 2, \ldots, 60\}$). This measure involves hypothetically purchasing an itinerary precisely 60 times for each purchase algorithm under test (but some purchases may be deferred for a few days based on the model output). Each of the 60 purchases is called a *purchase episode*. Table V shows examples of purchasing signals generated by four different policy generators (Earliest Purchase, Our Model, Optima, and Latest Purchase) for a specific city pair and departure date. The performance of each policy generator is the mean of the cost values across all purchase episodes.

Table V. Examples of policies computed by various models. The ◯ symbol indicates a "WAIT" signal for that day, and the • symbol indicates a "BUY" signal for that day. The models are compared by considering the mean value of the cost vector and represents the average procurement price of all simulated purchases.

Days to departure		...	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Earliest Purchase	action		•	•	•	•	•	•	•	•	•	•	•	•	•	•
	cost (in $)		258	257	257	257	257	282	292	330	298	330	330	222	469	453
Our Model	action		•	◯	◯	•	•	◯	◯	◯	•	◯	◯	•	◯	•
	cost (in $)		258	257	257	257	257	298	298	298	298	222	222	222	453	453
Optimal	action		◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	•	◯	•
	cost (in $)		222	222	222	222	222	222	222	222	222	222	222	222	453	453
Latest Purchase	action		◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	•
	cost (in $)		453	453	453	453	453	453	453	453	453	453	453	453	453	453

5.1. Performance Benchmarks

The naïve purchase algorithm, called *earliest purchase*, is to purchase a ticket once for each day in the α day range. Its purchase episodes terminate with a purchase event on the first day of the episode and mean cost would be equal to the mean of prices across the α day period. The lowest achievable cost is called the *optimal cost* and is based on purchasing for each of the α episodes at the lowest price between the beginning of the episode and its departure date. The comparison methodology involving simulated purchases is similar to that used in [Etzioni et al. 2003].

In Table VI, the results of estimated costs for several purchasing policies based on purchasing 5-day (Monday or Thursday departure) round trip itineraries from NYC to MSP (265 simulated purchases per column) are shown. The results show how costs vary based on preferences such as a customer requiring non-stop itinerary as well. We also compare our policy result against the cost of following the buy/wait recommendation from Bing Travel's "Fare Predictor." For different lag scheme based approaches, there are several benchmarks: 1) the *most recent observation* model contains only the most recent value from each feature, 2) the *all possible lags* is a model containing all lags $\{0, 1, \ldots, 7\}$ from every feature, and 3) the *Time Series* model contains all lags from only the target variable.

Table VI. Model results comparison on a single route data set for the lowest cost itinerary on any airline. All itineraries are 5 days (Monday to Friday, or Thursday to Tuesday). Cities are NYC (New York City) and MSP (Minneapolis, MN, USA).

Feature Selection Type	Learning Method	Learning Method Output	NYC-MSP Mon-Fri	NYC-MSP M-F nonstop	NYC-MSP Tu-Th	NYC-MSP Tu-Th nonstop
			(mean cost (in $), efficiency (as % of optimal savings))			
No Feature Selection	Earliest Purchase	Buy/Wait	(317, 0.00%)	(414, 0.00%)	(309, 0.00%)	(374, 0.00%)
	Optimal	Buy/Wait	(268, 100%)	(341, 100%)	(263, 100%)	(301, 100%)
	Bing Travel	Buy/Wait	(308, 2.56%)	N/A	(306, 0.903%)	N/A
	PLS w/ Minimal Lag Sch.	Regression	(314, 6.87%)	(384, 41.0%)	(294, 34.0%)	(354, 26.7%)
	PLS w/ Full Lag Scheme	Regression	(300, 34.1%)	(398, 22.4%)	(316, −13.4%)	(345, 38.7%)
Off-the-shelf Methods	PLS w/ CFS	Regression	(313, 7.93%)	(413, 0.223%)	(308, 0.331%)	(371, 2.98%)
	PLS w/ BFS	Regression	(317, −1.21%)	(416, −3.56%)	(310, −2.32%)	(369, 5.70%)
Lag Scheme Feature Selection	Decision Tree	Buy/Wait	(288, 58.8%)	(388, 35.3%)	(289, 42.9%)	(382, 56.8%)
	nu-SVR	Regression	(295, 45.1%)	(396, 24.5%)	(289, 42.9%)	(338, 48.3%)
	Ridge Regression	Regression	(293, 49.9%)	(383, 42.0%)	(316, −13.5%)	(372, 2.71%)
	Decision Tree	Regression	(284, 65.5%)	(375, 52.0%)	(280, 62.0%)	(334, 54.4%)
	PLS Regression	Regression	(280, 75.3%)	(365, 66.8%)	(276, 72.5%)	(330, 59.9%)

Table VII shows there is, on average, a possible 11% savings to be achieved over *earliest purchase*. We denote this percentage as the *savings margin*. Our method of a lag scheme search coupled with PLS Regression and a decision threshold achieves consistently closer to the optimal action sequence than any of the other methods compared. The 74% efficiency achieved by PLS (in Table VI) represents a savings of 8% less than the *earliest purchase* strategy for the NYC→MSP route.

5.2. Bing Travel Performance Comparison

It is surprising however that Bing Travel is not able to achieve a greater savings margin on the Any Airline target. We posit that this is due to a risk averse approach taken by the algorithm: it is more likely than our method to advise immediate purchase.

This assertion can be validated by looking at the distribution of buy and wait signals computed for each day by the various policy generators: in the NYC→MSP M-F route, the optimal policy has only a 15% proportion of buy signals. It is noteworthy that the best models constructed with our method emits a similar proportion of wait signals: in the NYC→MSP M-F route, the model with the lowest average cost ($280) only emits a buy signal on 34% of the days. Bing's model has a much higher proportion of buy signals: in the same route, the Bing model emits a buy signal 83% of the days. The results are similar for all routes and dates in our survey: Bing emits buy signals for at least 70% of the days. While the precise reasons for the Bing Travel model bias towards buy signals is unknown, we posit that the model may be more averse to possible future price increases than our tuned minimum cost approach.

Table VII. Mean percentage-based performance comparison of various decision theoretic approaches across a combination of 7 (domestic and international) routes by 3-digit airport code for both 5-day round trip Monday and Thursday departures. All values are in %. Savings Margin computed as % of earliest purchase cost.

	Model					
	Optimal	Our Model	Linear (LR)	Ripper	Earliest Purch.	Latest Purch.
Method Origin:	baseline	this paper	[Etzioni et al. 2003]		baseline	baseline
HOU → NYC	100.0	70.9	7.12	-26.8	0.00	-261
MSP → NYC	100.0	73.9	3.09	-45.2	0.00	-227
NYC → CDG∗	100.0	63.2	3.00	-74.8	0.00	-295
NYC → CHI	100.0	54.6	-4.73	-222	0.00	-626
NYC → HKG∗	100.0	56.9	10.5	-141	0.00	-161
NYC → MSP	100.0	64.9	5.74	-121	0.00	-289
SEA → IAD	100.0	69.4	4.91	-25.7	0.00	-190
Mean Efficiency	100.0	64.8	4.21	-93.8	0.00	-293
Savings Margin	11.0	7.25	0.514	-10.4	0.00	-32.3

∗ denotes an international route.

5.3. Multi-route Comparison

To show that this technique is generalizable to other routes (including international routes), we provide performance statistics of 7 routes in Table VII. The proposed method achieves an average of 69% of the optimal savings which represents an average cost savings of 7.25% when compared to the earliest purchase strategy. Given the high cost of airline tickets, this represents a significant savings. For the purposes of comparison with existing approaches,we provide results of two decision theoretic methods from [Etzioni et al. 2003]: Ripper and LR. Those models use a smaller number of features compared to our model and do not leverage the competitive relationships between airlines when making predictions. We believe prediction approaches should consider price competition between airlines.

5.4. Specific Preference Models

The proposed model is used to predict the future expected minimum price of all available flights on a specific route and date based on a corpus of historical price quotes. Also, we apply our model to predict prices of flights with specific desirable properties such as flights from a specific airline, non-stop only flights, or multi-segment flights. Buyers are likely to have preferences about airline tickets beyond price such as loyalty to a specific airline or the desire for overall minimum travel time. By comparing models with different target properties, buyers can determine the likely cost of their preferences.

The effect of more specific preferences can be observed in the prediction models that are generated for each preference. In Table VIII, the legacy airlines (DL and CO) offering multiple types including non-stop and multi-stop flights may have different lag schemes based on the preference. The general pattern observed is that more specific preferences (such as non-stop only flights) will require more information in terms of more specific feature classes and more time lags than less specific preferences (such as any flight from any airline). Also, more desirable flights such as non-stop only flights will tend to be more expensive than multi-stop ones. This is also observed in the lag scheme plots: our model is able to obtain an average price of $452 for Continental (CO) non-stop flights, but can obtain an average price of $345 for flights possibly with multiple stops.

For the LCAs, the contents of the lag schemes suggest that there is are delays in each airline's response to changes in prices. For example, the LCA models use little (Airtran)

Table VIII. Optimal lag schemes for specific preferences of airline and number of stops. Experiments are for 5-day round trips departing on Thursdays. The statistics are as follows: “MC” refers to average model cost, “EP” refers to average earliest purchase cost, and the percentage is the savings efficiency of the model.

Legacy Airlines

Continental Airlines (CO) — Non-stop

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A	●	●	●	●	●	●	●	●
ALL-S	●	●	●	●	●	●		
EACH-A		●	●	●	●	●		
EACH-S				●				

(MC: $452.7, EP: $483.6, 50.9%)

Continental Airlines (CO) — 0, 1, or 2-stops

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A	●	●	●	●	●	●		
ALL-S	●	●	●	●	●	●		
EACH-A								
EACH-S								

(MC: $344.8, EP: $386.3, 74.3%)

Continental Airlines (CO) — 1-stop

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A		●	●					
ALL-S		●	●					
EACH-A		●						
EACH-S								

(MC: $422.7, EP: $456.1, 71.1%)

Delta Airlines (DL) — Non-stop

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A			●	●	●	●	●	●
ALL-S			●	●	●	●	●	●
EACH-A			●					
EACH-S			●					

(MC: $389.5, EP: $411.6, 45.5%)

Delta Airlines (DL) — 0, 1, or 2-stops

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A	●	●	●	●	●	●	●	●
ALL-S	●	●	●					
EACH-A		●						
EACH-S		●						

(MC: $366.8, EP: $393.8, 53.7%)

Low-Cost Airlines (LCAs)

Frontier (F9) — 0, 1, or 2-stops

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A			●	●	●	●	●	●
ALL-S				●	●	●	●	●
EACH-A						●	●	
EACH-S						●	●	

(MC: $319.0, EP: $357.2, 78.0%)

Airtran (FL) — 0, 1, or 2-stops

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A	●	●	●	●	●	●	●	●
ALL-S			●	●	●	●	●	●
EACH-A								●
EACH-S								

(MC: $315.0, EP: $335.7, 67.0%)

Sun Country (SY) — Non-stop

Class	Lagged Offsets							
	0	-1	-2	-3	-4	-5	-6	-7
DET	●							
ALL-A		●	●	●	●	●	●	
ALL-S		●						
EACH-A		●						
EACH-S								

(MC: $352.6, EP: $399.1, 91.2%)

or no (Frontier and Sun Country) information from the current day (lag offset 0) and previous day (lag offset -1) for prediction. An alternative explanation for no current day information in the model may be in the airline’s role as a price leader. If an airline is setting prices aggressively, the airline will not make rapid changes in response to the actions of other companies in the market.

Also, Table VIII finds that some LCA models (Sun Country and AirTran) do not use information at the EACH-S level. This suggests that these airlines do not use detailed information from other airlines to determine their responses. The prices for these air-

Table IX. Comparison of the two airline routes under investigation.

	Significant Competitors Route	Mean Passengers per Day	Std. dev. of Lowest Cost Ticket by Departure (for $14 \leq$ *days-to-departure* ≤ 60)	
			Thu.	Mon.
MSP–NYC	3	1012	0.058	0.065
NYC–LAX	6	4031	0.035	0.043

lines generally are set equal to or just below the prices of the legacy airlines, so detailed information about the legacy airlines' pricing for non-stop and multi-stop flights may not be needed for prediction. Also, because the pricing relationships for the LCAs is less sophisticated than the relationships of the legacy airlines, the prediction models should be more effective for predicting LCA pricing than for legacy airline pricing. This can be observed in the efficiency values observed for LCAs compared to legacy airlines.

Using these statistics, it is also possible to reason about the relative costs of various preferences: for example, what is the expected price difference of a non-stop Delta flight and a flight on any available airline? Such information could help customers to quantify the expected costs of their preferences.

5.5. Significant Competitors

Comparing the effectiveness of our policy construction approach on two different airline routes has led to some insights about differences in the market structure of the two routes. We define a *significant competitor* on a route as an airline serving more than 1% of all passengers on a route. By comparing the results of NYC-MSP models against the NYC-LAX models in Table VI, reveals that there is a smaller savings margin in the NYC-LAX data. The price quotes reveal that there is significantly more competition and passenger volume in the NYC-LAX route. A comparison[3] of the two routes across several measures is provided in Table IX. The lower variance of daily minimum prices shown in the NYC-LAX route is likely due to the large number of competitive carriers along the route. In contrast, the NYC-MSP route has fewer "significant competitors", so individual airlines can assert greater pricing power.

6. CONCLUSIONS AND FUTURE WORK

To our knowledge, these results represent the state-of-the-art in ticket price prediction using consumer-accessible data.

This investigation shows that, given publicly-observable information, it is possible to predict airline ticket prices to systematically reduce costs. We believe that there is a significant market for this these kinds of models in the hands of consumers. In particular, reliable price models can assist buyers in determining the range of expected prices for an itinerary.

While it is the most obvious purchase policy, buying at the earliest opportunity is not the best policy in most cases. First, the long lead-time price may not be the lowest price available for that flight before departure. Also, there is an opportunity cost associated with early commitment: a customer risks being locked into a specific schedule that may need to be changed (for a fee). Because there is sufficient structured price volatility on many airline routes, there are significant opportunities for savings when using guidance of a predictive model. In addition to the results of this work, we believe

[3]The raw data source for computing these values was the US Department of Transportation BTS Origin-Destination Survey (see www.bts.gov).

there are additional cost reductions that can be found to obtain results closer to the optimal policy.

This feature selection technique also has wide applicability to other multivariate domains where basic domain knowledge is common but not utilized. Building the feature class hierarchy requires only basic domain knowledge to be successful; but, greater expertise in hierarchy construction will improve efficiency by focusing the feature selection search. The constraints also contribute by preventing overfitting (evident from the poor performance of off-the-shelf feature selection approaches) which can occur with many features and few training instances. The ability to incorporate a little domain knowledge into the model construction facilitates improved prediction performance and enables the discovery of meaningful relationships. Ticket price prediction is just one potential application of this technique. A novelty of this work also rests in a formulation appropriate for domains that have significant intra-variable and inter-variable temporal relationships. The resulting lag scheme model can be examined for domain understanding. The inclusion of lagged features in the model captures temporal relationships among features and improves the predictions. This method also contributes in facilitating domain understanding: by examining the relative performance of candidate lag schemes, domain knowledge can be extracted: the significance of individual features can be determined by observing their presence in the lag scheme.

REFERENCES

Enrico Bachis and Claudio A Piga. 2011. Low-cost airlines and online price dispersion. *International Journal of Industrial Organization* 29, 6 (2011), 655–667.

Peter P. Belobaba. 1987. Airline Yield Management. An Overview of Seat Inventory Control. *Transportation Science* 21, 2 (1987), 63–73.

Sijmen de Jong. 1993. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* 18, 3 (1993), 251–263.

Wedad Elmaghraby and Pinar Keskinocak. 2003. Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions. *Management Science* 49, 10 (2003), 1287–1309.

Oren Etzioni, Rattapoom Tuchinda, Craig Knoblock, and Alexander Yates. 2003. To buy or not to buy: mining airfare data to minimize ticket purchase price. In *SIGKDD Conf. on Knowledge Discovery and Data Mining*. ACM, 119–128.

William Groves and Maria Gini. 2013. Optimal Airline Ticket Purchasing Using Automated User-Guided Feature Selection. In *IJCAI '13: Proc. 23rd Int'l Joint Conf. on Artificial Intelligence*. AAAI Press, 150–156.

Mark A. Hall. 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In *Int'l Conf. on Machine Learning (ICML)*. 359–366.

Arthur E. Hoerl and Robert W. Kennard. 2000. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 42, 1 (2000), 80–86.

K. Isler and H. Imhof. 2008. A game theoretic model for airline revenue management and competitive pricing. *Journal of Revenue & Pricing Management* 7, 4 (2008), 384–396.

Ian T. Jolliffe. 1982. A Note on the Use of Principal Components in Regression. *J. of Royal Statistical Society. (Applied Statistics)* 31, 3 (1982), 300–303.

Ron Kohavi and George H. John. 1997. Wrappers for Feature Subset Selection. *Artif. Intell.* 97, 1-2 (1997), 273–324.

Y. Levin, J. McGill, and M. Nediak. 2009. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management Science* 55, 1 (2009), 32–46.

Harald Martens and Tormod Næs. 1992. *Multivariate Calibration*. John Wiley & Sons.

Luis Carlos Molina, Lluís Belanche, and Àngela Nebot. 2002. Feature Selection Algorithms: A Survey and Experimental Evaluation. In *IEEE Int'l Conf. on Data Mining (ICDM)*. IEEE Computer Society, 306–313.

C.A. Piga and N. Filippi. 2002. Booking and flying with low-cost airlines. *International Journal of Tourism Research* 4, 3 (2002), 237–249.

Bernhard Schölkopf, Alex J Smola, Robert C Williamson, and Peter L Bartlett. 2000. New support vector algorithms. *Neural computation* 12, 5 (2000), 1207–1245.

Barry C. Smith, John F. Leimkuhler, and Ross M. Darrow. 1992. Yield Management at American Airlines. *Interfaces* 22, 1 (1992), 8–31.

Janakiram Subramanian, Jr. Stidham, Shaler, and Conrad J. Lautenbacher. 1999. Airline Yield Management with Overbooking, Cancellations, and No-Shows. *Transportation Science* 33, 2 (1999), 147–167.

Ian H Witten and Eibe Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

S. Wold, H. Martens, and H. Wold. 1983. Multivariate calibration problem in chemistry solved by the PLS method. *Matrix Pencils* 973, 18 (1983), 286–293.