Dynamic Pricing in the Airline Industry

R. Preston McAfee
and
Vera te Velde

California Institute of Technology

Abstract: Dynamic price discrimination adjusts prices based on the option value of future sales, which varies with time and units available. This paper surveys the theoretical literature on dynamic price discrimination, and confronts the theories with new data from airline pricing behavior.

Correspondence to: R. Preston McAfee, 100 Baxter Hall, California Institute of Technology, Pasadena, CA 91125, preston@mcafee.cc.

Computerized reservation systems were developed in the 1950s to keep track of airline seat booking and fare information. Initially these were internal systems, but were soon made available to travel agents. Deregulation of airline pricing in 1978 permitted much more extensive use of the systems for economic activity, especially pricing. The initial development of dynamically adjusted pricing is often credited to American Airlines' Robert Crandall, as a response to the rise of discount airline People's Express in the early 1980s. The complexity and opaqueness of airline pricing has grown over time. As a result, the "yield management" systems employed by airlines for pricing have become one of the most arcane and complex information systems on the planet, and one with a very large economic component. Airline pricing represents a great challenge for modern economic analysis because it is so distant from the "law of one price" level of analysis. This paper surveys the theoretical literature, which is mostly found in operations research journals, develops some new theory, assesses the holes in our knowledge, and describes some results from a new database of airline prices.

Dynamic pricing, which is also known as yield management or revenue management, is a set of pricing strategies aimed at increasing profits. The techniques are most useful when two product characteristics co-exist. First, the product expires at a point in time, like hotel rooms, airline flights, generated electricity, or time-dated ("sell before") products. Second, capacity is fixed well in advance and can be augmented only at a relatively high marginal cost. These characteristics create the potential for very large swings in the opportunity cost of sale, because the opportunity cost of sale is a potential foregone subsequent sale. The value of a unit in a shortage situation is the highest value of an unserved customer. Forecasting this value given current sales and available capacity represents dynamic pricing.

Yield management techniques are reportedly quite valuable. One estimate suggests that American Airlines made an extra $500 million per year based on its yield management techniques (Davis 1994). This number may be inflated for several reasons. First, it includes sales of yield management strategy to others, as opposed to American's own use of the techniques, although the value of American's internal use is put at just slightly less. Second, it incorporates "damaged good" considerations in the form of Saturday-night stayover restrictions, as well dynamic pricing. Such restrictions facilitate static price discrimination, and are reasonably well-understood in other contexts (Deneckere and McAfee 1996). Nevertheless, there is little doubt that dynamic price discrimination is economically important. The pricing systems used by most major airlines are remarkably opaque to the consumer, which is not surprising given one estimate that American Airlines changes half a million prices per day. The implied frequency of price changes seems especially large given that American carries around 50,000 passengers per day.

There is surprisingly little research in economics journals concerning yield management, given its prominence in pricing in significant industries and the economic importance attached to it. This paper contributes to our understanding of yield management in five ways. First, it provides an extensive survey of yield management research in operations research journals. Second, we explore an existing model of Gallego and van Ryzin (1994) that has a number of desirable properties, including closed form solutions and sharp predictions, to address dynamic pricing considerations. Third, most of the literature assumes demand takes a convenient but unlikely form. We consider the implications of constant elasticity of demand and demonstrate some new inequalities concerning this more standard case. We examine this case in the context of an efficient allocation, rather than the profit-maximizing allocation, and show that many of the conclusions attributed to profit-maximization are actually consequences of the dynamic efficiency. Fourth, we take a new look at dynamic pricing from the perspective of selling options. A problem airlines face is that late arrivals may have significantly higher value than early arrivals, suggesting the airline ought to sell two kinds of tickets: a guaranteed use ticket and a ticket that can be delayed at the airline's request. Fifth, we've collected airline pricing data and generated stylized facts about the determinants of pricing, facilitating the evaluation of models.

1. Airline Pricing

Airline pricing in the United States is opaque. It is not uncommon for one-way fares to exceed round-trip prices. The difference in price between refundable and non-refundable tickets is often a factor of four or five. Prices change frequently, with low fares on a particular flight being available, then not, and then available again. Average prices for round-trips between Phoenix and Los Angeles differ depending on whether they originate in Los Angeles or in Phoenix. This is particularly mysterious in that the same airlines fly these round-trips with the same set of offerings.

We collected data on fares for 1,260 flights from American Airlines, Orbitz and Travelocity. Initially we collected data on each flight eight times daily, but when American Airlines objected to the volume of searches, we stopped using American's site and scaled back the other two sites to once per day per flight. As we completed some of the searches, we scaled the frequency up. Nevertheless, this represents a more intensive look at dynamic pricing than is available from any other source to our knowledge. We will take up our findings after presenting the theory to put the findings in perspective.

We will use the following terminology. We will use price dispersion to refer to static randomization by firms in pricing, in which each customer pays an identical price, but that price is determined by a random process. We will use dynamic price discrimination to refer to charging different customers distinct markups over marginal cost based on the time of purchase; when such pricing is efficient (maximizes the expected present value of the gains of trade), we'll call it dynamic pricing rather than dynamic price discrimination. Restrictions like Saturday-night stayovers, that create less valuable products, involve *static* price discrimination.

2. Existing Literature

Readers seeking a general discussion of yield management are referred to Williams (1999), Brooks and Button (1994), and Kimes (1989).[1] Kimes (1989) discusses situations appropriate for yield management solutions, particular issues involved such as demand forecasting and overbooking, solution techniques, and managerial implications of yield management systems. Brooks and Button (1994) discuss the rise of yield management during and after deregulation of the 1970's and 1980's, using the shipping industry as a detailed example, and Williams (1999) discusses yield management in terms of the interactions between the firm, resources, products, customers, and competitors. Talluri and van Ryzin (2004) have created a thorough textbook for the student of yield management.

Yield management applications in the made-to-order (MTO) manufacturing industry include Harris and Pinder (1995), Sridharan (1998), and Barut and Sridharan (2004). Both MTO firms and service providers such as airlines face the problem of effectively utilizing a fixed capacity under uncertain or high demand in order to maximize revenue, and thus many yield management results are applicable to the MTO manufacturing industry. However, MTO manufacturing is different on the key points of finite time horizon and unchanging capacity. The horizon is infinite, since the factory never stops all operations at a specific time or sets a common deadline for all activity, and the capacity is not fixed, in that as orders are completed, capacity is replenished. Thus, the MTO problem is more of a "stock out" problem than a yield management problem. Harris and Pinder (1995) discuss the applicability of traditional yield management to MTO manufacturing and its managerial implications and develop a relevant theoretical framework using price classes based on unit-capacity rates. Sridharan (1998) describes the use of yield management in manufacturing situations with higher demand than capacity, discussing three methods of

[1] A thorough literature review is contained in McGill and Van Ryzin (1999), along with a complete glossary of terms applicable to yield management, particularly in the airline industry. This literature review is largely distinct from theirs and may be considered in tandem.

increasing efficiency and revenue: capacity rationing based on price classes, increased coordination between marketing and manufacturing, and subcontracting. Barut and Sridharan (2004) further explore capacity rationing by developing a dynamic capacity apportionment procedure based on discriminating in favor of projects with a higher expected profit margin.

Rather than a continuous stream of one-time manufacturing requests, Carr and Lovejoy (2000) consider a non-competitive firm that agrees to commitments of a normally distributed random annual demand. This "inverse newsvendor problem" matches a known capacity with a desired aggregate demand distribution. They also consider the effect of multiple price classes. Fen and Wang (1998) study a similar annual capacity management problem: developing an optimal harvesting policy of a renewable resource. The resource population is modeled as a time-dependent logistic equation with periodic coefficients, and the maximum annual-sustainable yield is determined along with the corresponding harvesting policy.

The most common setting for yield management research is motivated by the airline and hotel industries. The remainder of the literature review focuses on these applications, and the rest of the paper deals with the airline industry exclusively.

Botimer (1996) and Belobaba and Wilson (1997) investigate effects of yield management external to the firm using it. Botimer (1996) presents arguments for efficiency of yield management pricing in the airline industry, and Belobaba and Wilson (1997) investigate the impacts of yield management introduction in competitive airline markets.

Most yield management research, however, deals with how to actually maximize revenue. One approach is to assume that customers arrive to request a flight, state the price they will pay, and then the firm decides whether or not to serve them. Van Slyke and Young (2000) study this situation in terms of filling a knapsack of fixed capacity with objects of known weights (or vector weights, to fill multi-dimensional knapsacks) and value, each type which arrives as a time-dependent Poisson process. With the goal of maximizing value, each object is accepted or rejected at the time of arrival. The case of equal weights is applied to airline seat inventory control, since each customer uses one seat, and the multi-dimensional knapsack is applied to the problem of allocating seats in a multiple origin and destination flight network. Sawaki (2003) addresses a similar question: Customers arrive throughout the continuous time horizon and state their price and demand size, and the firm decides whether to accept the request. A semi-Markov decision process is used, and an optimal policy and its analytical properties are found when demand arrives as a semi-Markov process.

A more realistic way to treat customer price preferences with respect to the airline industry, however, is to assume that the customer's willingness to pay is unknown when they request a ticket. Gallego and van Ryzin (1994) use intensity control theory (controlling the demand intensity with price changes) to study dynamic pricing under imperfect competition (i.e, demand is price-sensitive) and stochastic demand modeled as a Poisson process. A closed-form solution is found for exponential demand functions and an upper-bound on revenue under general demand functions is found using a deterministic heuristic. Also, monotonicity of the optimal price policy is shown, as well as asymptotic optimality of a fixed-price heuristic with market size. These results are then extended to allow compound Poisson demand, discrete prices, time-dependent demand, overbooking, resupply and cancellations, holding costs and discounting, and variable initial capacity.

Feng and Xiao (2000a) consider the problem from the perspective of switching between a predetermined finite set of prices at calculated time thresholds depending on remaining time and stock. Demand is a Poisson process at each price level, and they find the optimal continuous time switching policy of an arbitrary number of either price mark-ups or mark-downs. Feng and Xiao (2000b) extend this to allow reversible price changes. They show that any subset of prices in the optimal solution is part of the maximum concave envelope contained in the full set of allowed prices. Feng and Gallego (2000) also address the discrete price time-threshold problem, but allow demand to be Markovian, i.e. fares and demand are not only dependent on remaining time but also on prior sales. Chatwin (2000) additionally allows demand (Poisson) to be time sensitive and provides the option of re-stocking at a unit cost. Zhao and Zheng (2000) find structural properties of the optimal price-switching policy (from a compact, but not necessarily finite, price set) when demand is a nonhomogenous Poisson process and investigate solution methods in the case of discrete prices. Badinelli (2000) considers product-differentiated market segments, each of which is allowed one advertised price at a time, and formulates an efficiently computed dynamic programming solution allowing general demand functions. In his model, customers request a hotel room or flight with a particular set of attributes and they are given a price quote based on remaining time and availability of all relevant commodities.

Rather than dynamically changing prices to maximize revenue, some authors ration capacity with price classes to ensure that high-paying customers are served, effectively implementing a mark-up policy based on remaining capacity and, if seat allocation between classes is dynamically controlled, remaining time. Ladany (1996), assuming deterministic non-linear demand and a fixed cost for creating each price class, develops a dynamic-programming solution for finding the optimal number of price classes, the optimal number of capacity units (specifically, hotel rooms) allocated to each price class, and the optimal price at each class. Bodily and Weatherford (1995) allow uncertain demand, and study when to curtail low-price sales based on the probability of spoilage and expected marginal revenue. They don't allow re-opening of price classes, but do not assume that all low-fare customers will arrive before high-fare customers (thereby forcing all high-fare customers to pay their full acceptable price). Li (2001) again considers deterministic demand, and studies the use of sales restrictions (such as advance purchasing, or minimum trip duration) for segmenting demand. He develops several key properties of the optimal restriction policies, and applies his results to airline pricing, where leisure travelers' relative price elasticity compared to business travelers permits the efficient use of restrictions.

Dana (1999a,b) studies aspects of market segmentation other than methods of optimal allocation. Dana (1999a) studies the phenomenon of equilibrium price dispersion, showing that the optimal pricing system under uncertain demand is price dispersed in monopolistic, imperfectly and perfectly competitive markets. He shows that the amount of price dispersion increases with the level of competition, using this to explain the observation that routes served by more airlines exhibit more price dispersion. Dana (1999b) demonstrates how setting multiple price levels on flights at different times can shift demand from the higher-demand departure time to the alternate flight, even when it is unknown which time is the peak. He addresses the competitive and monopoly cases and uses his model to show that high-fare customers may benefit from price dispersion as well as low-fare customers.

Moving beyond single flight-leg analysis, several authors consider origin-destination networks. Feng and Xiao (2001) consider a network of multiple origins, one hub, and one destination. Prices for each flight leg are distinct, and demand at each origin is a Poisson process. They use a nested fare structure for inventory control and find the optimal time thresholds for closing origin-hub flights. They then extend their results to allow multiple fares on each origin-destination flight and time-dependent demand. Kuyumcu and Garcia-Diaz (2000) and Garcia-Diaz and Kuyumcu (1997) use a graph-theory approach for allocating seats in a flight network. Garcia-Diaz and Kuyumcu (1997) assumes a non-nested seat allocation system, normally distributed random demand that is independent between price levels, and a fixed number of fare classes. They develop an algorithm utilizing cutting-planes for allocating seats throughout the origin-destination network and investigate computational times. Kuyumcu and Garcia-Diaz (2000) use the same assumptions except that demand on each day of the week may be different, and airline capacities are considered. They develop three models for the network pricing and seat allocation problem, the third being a polyhedral graph theory approach, and a solution procedure that they test computationally. The doctoral thesis of Williamson (1992) compares the use of several network-level seat inventory control heuristics in simulation. She concludes that considering network-wide effects may increase revenue by 2-4%, but only when the load factor is very high. de Boer, Freling and Piersma (2002) extend her research by disputing the finding that deterministic approximation methods outperform probabilistic heuristics which they claim is due to a difference in fare class structure in modeling and simulation.

In these studies of dynamic price discrimination, there is a tension between the practical and the insightful. Computation-based analyses in principle could be used to solve real-world problems of optimal pricing, while simpler theories elucidate potential principles for understanding pricing generally.

A common assumption in the literature posits a discrete price grid, or even just two prices, which simplifies the problem to one of deciding when to switch to another price (Feng and Xiao (2000a, 2000b), Feng and Gallego (2000), Sawaki (2003)). The assumption of a discrete price grid doesn't seem justified, either on theoretical grounds, since there is no economy of using few prices, or on practical grounds, since airlines and hotels in fact employ a large set of prices. Thus, the defense of the assumption relies on the simplification of the mathematical problem. The main advantage of the assumption is that proving existence of an optimal solution is trivial, but little else is gained. An exception is Chatwin (1999), which exploits the finite grid to develop a nice intuition for when prices rise and fall.

The use of a specific functional form for demand is common. The most common assumption is exponential demand, $q(p) = ae^{-bp}$, for constants a, and b. This form of demand is useful to assume because q'/q is a constant, and thus marginal revenue is price minus a constant. As a result, a monopolist would choose to charge a price which is marginal cost plus a constant. This feature of exponential demand makes the solution to the monopoly problem reduce to the problem of calculating marginal cost, a simpler though nontrivial problem. Gallego and van Ryzin (1994) is the best paper of the set of exponential demand papers, and we discuss it extensively below and indeed ask some additional questions of this theory.

A variety of mechanisms for bringing customers to the seller are considered in the literature. The most common assumption is a constant Poisson process but possibly with time-varying arrival rates (Zhao and Zheng 2000). An interesting variant on the arrival rate process is a markov switching model, which involves a signal extraction problem: detecting from the behavior of buyers whether demand conditions have changed (Feng and Gallego 2002). The connection between the classic peak-load pricing problem and dynamic pricing is explored in Dana (1999b). This paper concludes that prices should be increasing on average, because of a correlation between high demand and high prices. As in Dana's analysis, several authors posit multiple classes of customers, who may arrive via distinct processes.

While the most common objective function is to maximize expected revenue, Carr and Lovejoy (2000) consider the alternate assumption of pricing to sell out. Pricing to sell out is a bad strategy for several reasons. First, a profit maximizing firm would not price to sell out. However, more alarmingly, even the efficient solution that maximizes the gains from trade doesn't price to sell out.

Most of the literature focuses on the problem of a single flight, treating competition and alternate flights as fixed, but several authors have made headway into the problem of multiple flights and routes. Feng and Xiao (2001) examine a simple Y-shaped pattern through a hub. Network issues are also examined in de Boer, Freling and Piersma (2002).

Dana (1999a) is the only author to develop a full theory of competition between dynamically pricing sellers. The theory, involving two firms pricing over two periods, emphasizes that price dispersion may result from such competition. This result is also available in static, one-period problems.

An important oversight of the literature is the absence of discounting. Virtually the entire literature presumes no discounting. In the hotel context, zero discounting makes sense, because even if one books a hotel in advance, generally payment isn't made until the time one stays in the hotel, which implies the same discount factor whether one books the room early or later. With airline tickets, however, generally payment is made at the time of booking, not at the time of departure. This matters because the time intervals are long enough for discounting to be significant, given the tickets may be booked six months in advance.

3. Dynamic Price Discrimination with Price Commitment

The extent of price changes found in actual airline pricing is mysterious because a monopolist with commitment ability, in a standard framework, doesn't want to engage in it at all! To develop this conclusion first proved by Stokey (1979), we start with a simplified version of her analysis. This analysis was dramatically extended by Board (2004). The seller sells either a durable good or a one-time use good like an airplane trip. Time is composed of discrete periods t = 1, 2, … and buyers and the seller have a common discount factor δ. There are a continuum of potential buyers who have values represented by a demand q, so that q(p) gives the measure of buyers willing to pay p. We assume that a consumer's demand persists until the consumer purchases. The monopolist chooses a price sequence $p_1, p_2, \ldots$ which can be taken to be non-increasing without loss of generality. A consumer with a value v will prefer time t to time t+1 if

$$v - p_t > \delta(v - p_{t+1}) \qquad [1]$$

The equations producing indifference, $v - p_t = \delta(v - p_{t+1})$, define a sequence of critical values v_t that make the buyer indifferent between purchasing at t and purchasing at t+1. Note that the incentive constraint on buyers shows that, if a buyer with value v chooses to buy before time t, then all buyers with values exceeding v have also purchased by this time.

$$v_t - p_t = \delta(v_t - p_{t+1}) \qquad [2]$$

This set of equations can be solved for p_t in terms of the critical values:

$$p_t = (1-\delta)v_t + \delta p_{t+1} = (1-\delta)v_t + \delta((1-\delta)v_{t+1} + \delta p_{t+2}) = \ldots \qquad [3]$$
$$= (1-\delta)\sum_{j=0}^{\infty} \delta^j v_{t+j}.$$

The monopolist sells $q(v_t)$-$q(v_{t-1})$ in period t, where v_0 is defined so that $q(v_0)=0$. The monopolist's profits are

$$\pi = \sum_{t=1}^{\infty} \delta^{t-1} p_t q_t = \sum_{t=1}^{\infty} \delta^{t-1}(1-\delta)\left(\sum_{j=0}^{\infty} \delta^j v_{t+j}\right)\left(q(v_t) - q(v_{t-1})\right) \qquad [4]$$

$$=(1-\delta)\left[\sum_{t=1}^{\infty} q(v_t)\delta^{t-1}\left(\sum_{j=0}^{\infty}\delta^j v_{t+j}\right)-\sum_{t=1}^{\infty} q(v_{t-1})\delta^{t-1}\left(\sum_{j=0}^{\infty}\delta^j v_{t+j}\right)\right]$$

$$=(1-\delta)\left[\sum_{t=1}^{\infty} q(v_t)\delta^{t-1}\left(\sum_{j=0}^{\infty}\delta^j v_{t+j}\right)-\sum_{t=2}^{\infty} q(v_{t-1})\delta^{t-1}\left(\sum_{j=0}^{\infty}\delta^j v_{t+j}\right)\right]$$

$$=(1-\delta)\left[\sum_{t=1}^{\infty} q(v_t)\delta^{t-1}\left(\sum_{j=0}^{\infty}\delta^j v_{t+j}\right)-\sum_{t=1}^{\infty} q(v_t)\delta^{t}\left(\sum_{j=0}^{\infty}\delta^j v_{t+1+j}\right)\right]$$

$$=(1-\delta)\left[\sum_{t=1}^{\infty} q(v_t)\delta^{t-1}\left(\sum_{j=0}^{\infty}\delta^j v_{t+j}-\delta\sum_{j=1}^{\infty}\delta^{j-1} v_{t+j}\right)\right]$$

$$=(1-\delta)\sum_{t=1}^{\infty} q(v_t)\delta^{t-1}v_t=(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}q(v_t)v_t.$$

Thus, the optimum level of v_t is constant at the one-shot profit maximizing level, which returns the profits associated with a static monopoly. The ability to dynamically discriminate does not increase the ability of the monopolist to extract rents from the buyers.

There is an important lesson to be drawn from Stokey's theorem. If dynamic price discrimination is playing a role, it is because of new customers arriving, rather than an attempt to extract more profits from an existing set of customers by threatening low value customers with delayed purchases. That is, dynamic price discrimination is driven by customer dynamics rather than price discrimination over an existing set of customers.

4. Continuous Time Theory

Gallego and van Ryzin (1994) produced a closed form model of discrete time optimal pricing. This model is very useful for its tractability, and we reproduce some of their analysis here, as well as extend it by generating predictions about the average path of prices.

Let λ be the arrival probability of customers per unit of time, assumed constant. A constant arrival can be assumed without loss of generality by indexing time in terms of customer arrivals. Time will start at 0 and end at T. If not sold, the product perishes at T, which might occur because a flight takes off or the product is time-dated like a hotel room or time-share condominium, where what is for sale is the use of the product on date T. For technical reasons, no discounting is considered in this section. However, in some applications, no discounting is the right assumption. For example, hotels generally do not charge until the customer arrives, independently of the time the room is booked, a situation which corresponds to no discounting based on late booking. The marginal cost of the product is c, a value which might include normal marginal costs (cleaning a hotel room, a meal served on an airline) but could also include lost business – the customer that takes a particular flight is less likely to take an alternative flight by the same airline. Potential customers demand a single unit, and their willingness to pay is given by a cumulative distribution function F.

The value of having n items for sale at time t is denoted by $v_n(t)$. Clearly having nothing to sell conveys zero value. Moreover, if not sold by T, an inventory of items also has zero value, yielding:

$$v_0(t) = v_n(T) = 0. \qquad [5]$$

Consider a small increment of time, Δ, beyond a current time t. With probability 1-λΔ, no customer arrives, so that the current value becomes $v_n(t+\Delta)$. Alternatively, with probability λΔ, a customer arrives and the firm either makes a sale or does not. For price p, the sale occurs with probability 1-F(p). When a sale occurs, the value becomes $p - c + v_{n-1}(t+\Delta)$, because the inventory is decreased by one. Summarizing:

$$v_n(t) = \max_p \; (1-\lambda\Delta)v_n(t+\lambda\Delta) + \lambda\Delta\big((1-F(p))(p-c+v_{n-1}(t+\lambda\Delta)) + F(p)v_n(t+\lambda\Delta)\big) \qquad [6]$$

or

$$v_n(t) - v_n(t+\lambda\Delta) = \lambda\Delta \max_p \; (1-F(p))(p-c+v_{n-1}(t+\lambda\Delta)) - v_n(t+\lambda\Delta)) \; [7]$$

Therefore, dividing by Δ and sending Δ to zero,

$$-v_n'(t) = \lambda \max_p \;\; (1-F(p))(p-c+v_{n-1}(t) - v_n(t)) \qquad [8]$$

The expression for $v_n'(t)$ is composed of two terms. First, there are profits from a sale, p-c. Second, there is the lost option of selling the unit in the future, an option that has value $v_n(t) - v_{n-1}(t)$. It is not possible to solve this differential equation for an arbitrary demand function F. However, with a convenient choice of F, it is possible to provide an explicit solution. Let

$$F(p) = 1 - e^{-ap}. \qquad [9]$$

Note that $(1-e^{-ap})(p-mc)$ is maximized at $p^* = 1/a + mc.$[2] Then

---

[2] Let $h(p)= (1-e^{-ap})(p-mc)$. $h'(p) = e^{-ap}(1 - a(p - c))$ $h'(p) = 0$ implies $h''(p) \le 0$. Thus, every extreme point is a maximum, and so if there is an extreme point, it is the global maximum. Moreover, p* is an extreme point.

$$p_n^*(t) = \frac{1}{a} + c + v_n(t) - v_{n-1}(t), \qquad [10]$$

and

$$v_n'(t) = -\lambda e^{-a\left(\frac{1}{a}+c+v_n(t)-v_{n-1}(t)\right)} \frac{1}{a}. \qquad [11]$$

The multiplicative constant $\beta = \lambda e^{-1-ac}$ represents the arrival rate of buyers willing to pay the static monopoly price $1/a + c$. Thus, at time t, the expected number of buyers willing to pay the monopoly price is $\beta(T-t)$. This observation helps explain why

$$B_n(t) = \sum_{j=0}^{n} \frac{(\beta(T-t))^j}{j!} \qquad [12]$$

will appear in the solution. The first result characterizes the value function and prices in a closed-form manner.

Lemma 1 (Gallego & van Ryzin 1994): $v_n(t) = \frac{1}{a} Log(B_n(t))$ and $p_n^*(t) = \frac{1}{a} Log\left(\frac{\lambda B_n(t)}{\beta B_{n-1}(t)}\right)$.

At time zero, suppose there is an initial capacity k. Let $q_i(t)$ be the probability that there are i units left for sale at time t.

Theorem 2: $q_n(t) = \frac{(\beta t)^{k-n} B_n(t)}{(k-n)!\, B_k(0)}$. The expected number of seats sold is

$$E(k-n) = \sum_{n=0}^{k} (k-n) q_n(t) = \frac{\beta\, t\, B_{k-1}(0)}{B_k(0)}$$

Lemma 1 and Theorem 2 give a complete, closed form description of profits, prices and sales for the dynamic monopolist practicing yield management. For example, the probability that there is no available capacity at time T is

$$q_0(T) = \frac{(\beta T)^k / k!}{\sum_{j=0}^{k} (\beta T)^j / j!}. \qquad [13]$$

This formula insures that, with sufficient time and a given capacity k, the flight sells out, because

$$\lim_{T\to\infty} q_0(T) = 1. \qquad [14]$$

How does a thick market affect the outcome? To model this, consider increasing both the capacity k and the arrival rate of customers, λ, proportionally. Let $\gamma=\beta T/k$, so that

$$q_0(T) = \left(\sum_{j=0}^{k} \frac{k!\,(\beta T)^{j-k}}{j!}\right)^{-1} = \left(\sum_{j=0}^{k} \frac{k!\,(\gamma k)^{j-k}}{j!}\right)^{-1} \underset{k\to\infty}{\to} \begin{cases} \dfrac{\gamma-1}{\gamma} & \text{if } \gamma > 1 \\ 0 & \text{if } \gamma \le 1 \end{cases} \qquad [15]$$

What is interesting about this expression is that the probability of selling all the capacity, $q_0(T)$, converges to the same level as would arise if the price was just constant at the monopoly price $\frac{1}{a} + c$. Since the price exceeds this level always, because the option value is positive, the price must get very close to the static monopoly price most of the time in order for the limiting probabilities to coincide. Per unit profits of the dynamically discriminating firm are

$$\frac{v_k(0)}{k} = \frac{1}{ak} Log(B_k(0)) = \frac{1}{ak} Log\left(\sum_{j=0}^{k} \frac{(\gamma k)^j}{j!}\right) \underset{k\to\infty}{\to} \frac{1}{a} \begin{cases} \gamma & \text{if } \gamma < 1 \\ 1 + Log(\gamma) & \text{if } \gamma \ge 1 \end{cases} \qquad [16]$$

As an alternative to yield management and dynamic price discrimination, consider a firm that sets a price and leaves it fixed. Such a firm, a one-price monopolist, will earn lower profits than the yield-management firm. Is the profit reduction significant?

A monopolist who only offers one price p will have a flow-rate of sales of $\mu=\lambda(1-F(p))$. If $\eta_i(t)$ is the number of future sales given capacity i at time t, then $\eta_0(t)=0$ and

$$\eta_j(t) = \int_t^T \mu\, e^{-\mu(s-t)}(1+\eta_{j-1}(s))ds. \qquad [17]$$

Lemma 3: $\eta_j(t) = k - e^{-\mu(T-t)} \sum_{i=0}^{k-1} (k-j) \frac{(\mu(T-t))^j}{j!}$

The profits associated with a single price can be numerically maximized.

Figure 1 provides an example of dynamic price discrimination, starting with ten units for sale. The parameters include zero marginal cost, a 365 day period with an average of one customer every other day, and demand in the form $1 - F(p) = e^{-p}$. The prices associated with various numbers of units to be sold are illustrated, for example, p_3 shows the optimal price associated with three units to be sold. The average price, conditional on units remaining for sale, is denoted by Ep; this represents the average of posted prices. Note that units may not be available for sale, so that the expected price is a bit of a fiction, although the comparison to the "one price monopolist" is sensible since that price may also not be offered. The optimal constant price is denoted in p^1. Profits are 5.45% higher under dynamic price discrimination than they are with a constant price. In this figure, the monopoly price is the horizontal axis, at 1, and either scheme improves substantially on the static monopoly price.

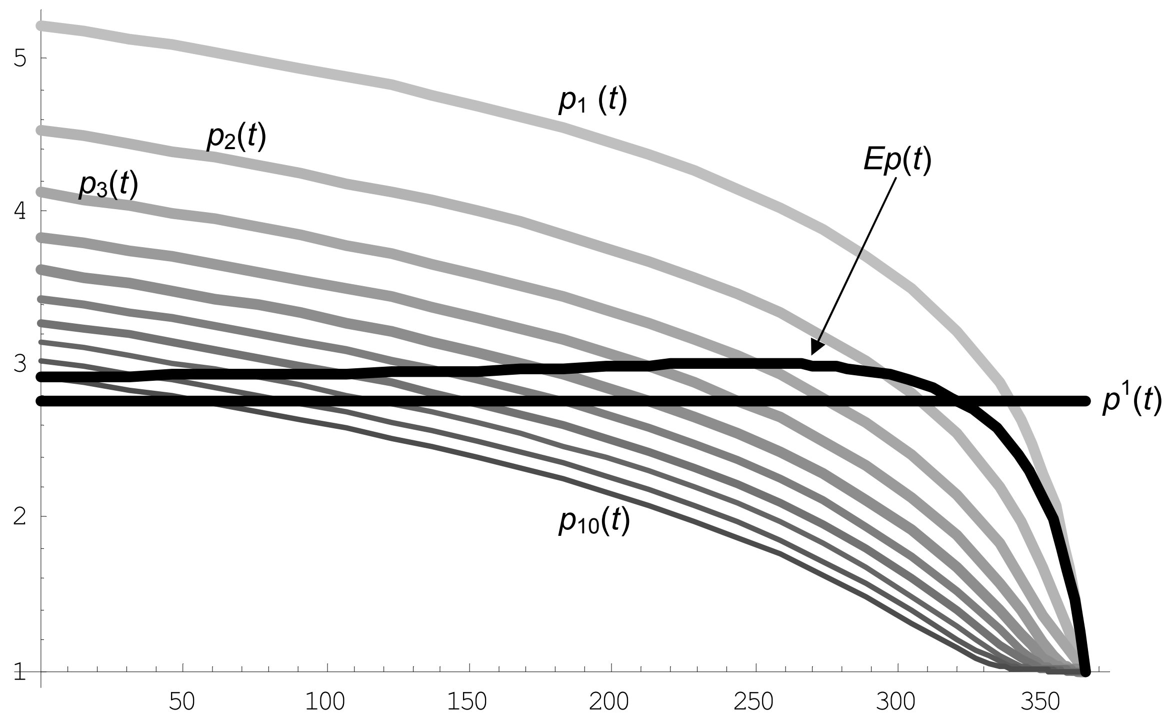

Figure 1: Prices, and expected price, *k*=10, λ=½, α=1, φ=0, *T*=365.

Figure 1 illustrates a common feature of simulations: the expected price rises, then falls. The forces involved are as follows. First, prices must eventually fall because there is a positive option value prior to time T, and this option value is zero at time T. Thus, prior to T, prices are strictly larger than the static monopoly price and converge to the static monopoly price at T. While it is possible that prices fall for the entire time interval, they may initially rise because early sales, by reducing available capacity, drive up the prices.

Figure 2 illustrates the probability that all units sell under dynamic price discrimination. This converges to slightly over 85.35% by the end of the period.

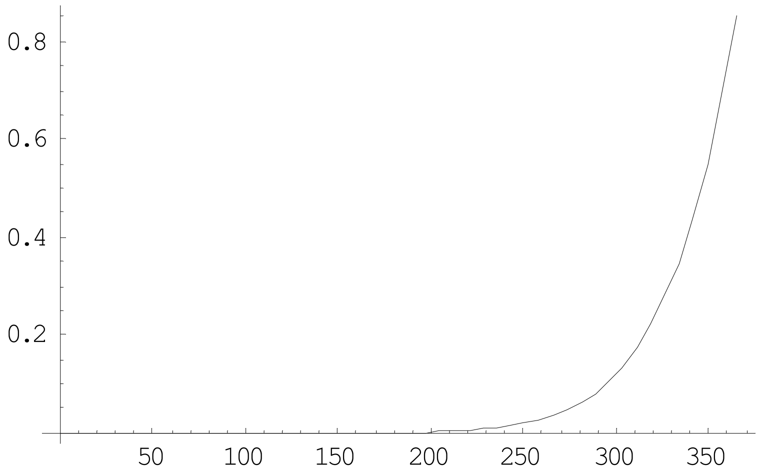

Figure 2: The Probability of Zero Capacity

Is dynamic price discrimination profitable when k and λ are both very large? Let θ represent λ/k. In this case, the solution for demand is

$$\frac{\eta_j(t)}{k} \xrightarrow[k\to\infty]{} \min\left\{\theta e^{-ap}T,\, 1\right\} \qquad [18]$$

This is maximized at the static monopoly price, $1/a + c$, provided $\gamma=\theta e^{-1-ac}T < 1$. Otherwise, the optimal price satisfies $\theta e^{-ap}T = 1$. It follows that profits are $\frac{1}{a}\begin{cases} \gamma & \text{if } \gamma < 1 \\ 1 + Log(\gamma) & \text{if } \gamma \geq 1 \end{cases}$ and agree with discriminatory profits in the limit for large k. That is, per unit gain in profits of dynamic price discrimination over an optimally chosen constant price converges to zero, although the total gain will still be positive. This happens because most sales take place at an approximately constant price; dynamic price discrimination is advantageous only as the probability of actually selling out changes, for a relatively small portion of the very large number of sales. One can reasonably interpret these results to say that dynamic price discrimination only matters on the last twenty or so sales, so when a large number of units are sold, dynamic price discrimination doesn't matter very much.

Dynamic price discrimination has a relatively modest effect when there are 100 or more seats available. The kinds of profits predicted, for reasonable parameter values, under dynamic price discrimination are not very large, less than 1%, when compared to an appropriately set constant price. An important aspect of this conclusion is that dynamic price discrimination does not appear to account for the kinds of value claimed for American Airlines' use of yield management.

5. Efficiency in the Gallego & van Ryzin Model

An efficient solution in this model has the property that the value function maximizes the gains from trade rather than the profit. The value function, then, satisfies

$$v_n(t) = \max_p \; (1-\lambda\Delta)v_n(t+\lambda\Delta) + \lambda\Delta\big((1-F(p))(C(p)+v_{n-1}(t+\lambda\Delta)) + F(p)v_n(t+\lambda\Delta)\big) \qquad [19]$$

where C(p) is the consumer's value, plus seller profit, conditional on the consumer's value exceeding p. In this model,

$$(1-F(p))C(p) = \int_p^\infty 1 - F(x)dx + (p-c)(1-F(p)) \qquad [20]$$

$$= \int_p^\infty e^{-ax}dx + (p-c)e^{-ap} = (p + 1/a - c)e^{-ap}.$$

Thus the efficient solution is the solution a monopoly whose costs are reduced by 1/a, the static monopoly profit, would choose. The assumed structure of demand insures that all the qualitative conclusions drawn apply equally to efficiency as to monopoly. In particular, the shape of the time path of prices and the conclusion that for large problems the gains of dynamic price discrimination are small apply equally to efficient solutions as they do to monopoly solutions.

In our view, the Gallego & van Ryzin model is not a very useful instrument for examining efficiency, because of the similarity of the monopoly price and efficient price. Reducing a monopolist's marginal costs by the monopolist's markup does not produce the efficient price, but a higher price under the standard regularity conditions like log concavity. However, it is worth emphasizing that the efficient solution and the monopoly solution lead to the same set of price paths, so that if costs are unobservable, the two models are observationally equivalent.

6. Efficiently Allocating Limited Capacity under Uncertainty

The closed form of Gallego – van Ryzin facilitates computation but obscures economic considerations because of the nature of the demand. In this section, we consider efficiently solving the allocation problem, rather than maximizing profit. Efficiency is not unreasonable, since airlines face competition and a perfectly competitive model would suggest the efficient allocation of the available quantity. It seems likely that neither the monopoly nor the fully efficient solution accurately represent the real world, which probably falls somewhere in between the two extremes. Borenstein and Rose (1994) is perhaps the foremost example of an attempt to assess airline price competition, but at the level of average fares. Developing a theory that permits assessing the degree of competition at the detailed level of the price path, in contrast, presents a formidable challenge. Observations of quantities are going to be central to the assessment of the degree of competition, because price observations alone are not sufficient to identify the model.

This kind of efficient allocation problem was first studied in the context of electricity generation. Costs of capacity induce a peak load problem (Boiteaux 1949, Williamson 1966), so named because the marginal cost of electricity generation is much higher when the plant capacity is reached than below that level, so that the peak load matters a lot to the economics of electricity generation. The standard assumption in peak load problems is that demands in each period are independent, and that the full supply is available in each period. The peak-load problem is really a capacity planning problem, where the needed capacity depends on prices. For a two period model, let q_i be the demand in period i. The firm's profits are given by

$$\pi = p_1 q_1 + p_2 q_2 - \beta \max\{q_1, q_2\} - mc(q_1 + q_2), \tag{21}$$

where β is the marginal cost of capacity and mc is the marginal cost of production.

In the airline and hotel context, the standard peak load model is poorly suited, because capacity sold in one period isn't available in the subsequent period. Thus, the standard model of the peak load problem asks how large a plane that flies round-trips between Dallas and Chicago should be, given variation in total demand from day to day or from week to week. While this problem represents an important logistical problem, it has little or nothing to do with dynamic pricing. In contrast, airlines also face the problem that seats on a particular flight can't be occupied by two passengers, and this problem of allocating limited capacity suggests a quite distinct peak load model. Moreover, the fact that future demand isn't known at the time of contracting in the first period is a crucial aspect of the peak load problem facing airlines and hotels. That is, airlines and hotels contract for initial sales, not knowing the realization of overall demand.

We introduce a new model of random arrival that has some attractive properties. The randomness comes in the form of random variables n_t in period t, with period t demand being, then, $n_t\, q(p)$ for price p. We also assume that n_t is observable at the beginning of period t, so that the firm can either set price or quantity in that period; what is uncertain is the future demand, not the current demand.[3] We will focus on the constant elasticity of demand case, $q(p)=\alpha p^{-\varepsilon}$, because this is a standard empirical model and assists in tractability. We will refer to q(p) as per capita sales, and $n_t\, q(p)$ as total sales, but this interpretation isn't exact in the airline context. In the airline context, if n_t is the number of customers, q(p) would have to be the probability of sale, in which case it should be no greater than one, which is inconsistent with the constant elasticity assumption. At best, the model represents an approximation for the airline context.

This section considers only the efficient allocation. The efficient allocation is interesting for several reasons. First, competition with other providers will generally push firms' product offerings toward efficiency and away from monopoly, so with vigorous competition, efficiency maximization is probably a better model than monopoly profit maximization. Second, the monopoly model has been better studied than the competitive, efficient model. Third, some of the interesting behavior in the monopoly model does not arise from price discrimination but because of the dictates of efficiency. The source of the behavior is challenging to see without studying the efficient allocation. That is, what appeared to be price discrimination is merely variations in marginal cost.

The seller has a capacity K. We will focus on the two-period case throughout this section, although this readily generalizes to more periods.

If per capita sales are s_1 of the n_1 first period demand, the number of seats available for sale in the second period is $K - s_1 n_1$. These are valued by consumers at a price p_2 satisfying

$$K - s_1\, n_1 = n_2\, q(p_2) \tag{22}$$

[3] Dana's 1998 model is a special case, with q taking on two positive values, interpreted as a willingness to pay for leisure and business travelers.

or

$$p_2 = q^{-1}\left(\frac{K - s_1 n_1}{n_2}\right) \tag{23}$$

The cost of first period sales rises quite dramatically in the number of sales. Suppose q has constant elasticity of demand, $q(p)=\alpha p^{-\varepsilon}$. We assume $\varepsilon>1$ so that the consumer surplus is finite. For any given n_2,

$$p_2 = \left(\frac{\alpha n_2}{K - s_1 n_1}\right)^{1/\varepsilon} \tag{24}$$

The price p_1 that clears the market in period 1 satisfies $s_1 = q(p_1)$. The customer's value of quantity q is $\frac{\alpha\varepsilon}{\varepsilon - 1} q^{\frac{\varepsilon-1}{\varepsilon}}$. Thus the overall gains from trade are

$$W = \frac{\alpha\varepsilon}{\varepsilon - 1}\left(n_1 s_1^{\varepsilon-1/\varepsilon} + E\left\{n_2\ q(p_2)^{\varepsilon-1/\varepsilon}\right\}\right) = \frac{\alpha\varepsilon}{\varepsilon - 1}\left(n_1 s_1^{\varepsilon-1/\varepsilon} + E\left\{n_2\left(\frac{K - s_1 n_1}{n_2}\right)^{\varepsilon-1/\varepsilon}\right\}\right) \tag{25}$$

$$= \frac{\alpha\varepsilon}{\varepsilon - 1}\left(n_1\left(s_1^{\varepsilon-1/\varepsilon}\right) + E\left\{n_2^{1/\varepsilon}\right\}(K - s_1 n_1)^{\varepsilon-1/\varepsilon}\right)$$

The gains from trade are maximized at

$$s_1 = \frac{K}{n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}. \tag{26}$$

The per capita second period sales are

$$Es_2 = E\left\{\frac{K - s_1 n_1}{n_2}\right\} = E\left\{\frac{K\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{n_2\left(n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}\right)}\right\} = s_1 E\left\{\frac{\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{n_2}\right\} \tag{27}$$

Because $p_i = \left(s_i \big/ \alpha \right)^{-1/\varepsilon}$, it is a rudimentary calculation to show that the expected price in the first period and second period coincide. Equality of the prices is a necessary condition for maximizing the gains from trade; otherwise it would increase the gains from trade to allocate more of the seats to the high-priced time.

Some seats are misallocated, because contracting with the early customers occurs prior to knowing the realization of number of later customers. Moreover, it is worth emphasizing that this misallocation is efficient – maximizes the gains from trade – and not a consequence of monopoly pricing. In our terminology, it is dynamic pricing but not dynamic price discrimination. How large is the effect? The average reduction in the share of period two customers satisfies

$$\frac{Es_2}{s_1} = E\left\{ \frac{\left(E\left\{ n_2^{1/\varepsilon} \right\} \right)^{\varepsilon}}{n_2} \right\}. \qquad [28]$$

Theorem 4: The second period share is larger on average than the first period share, that is,

$$\frac{Es_2}{s_1} = E\left\{ \frac{\left(E\left\{ n_2^{1/\varepsilon} \right\} \right)^{\varepsilon}}{n_2} \right\} \geq 1.$$

Proof in Appendix.

Theorem 4 shows that the share of customers served in the second period is at least as great as are served in the first period. This is a consequence of the possibility of small denominators leading to large shares; when few customers arrive, the price is set low enough to insure the plane sells out. The sales as a fraction of the average value of n_2 are less than in period 1. That is,

$$\frac{Es_2 n_2}{s_1 En_2} = \frac{\left(E\left\{ n_2^{1/\varepsilon} \right\} \right)^{\varepsilon}}{En_2} \leq 1.$$

Insight into the magnitude of this efficient misallocation is given by the next result. Let $CV = \sqrt{\frac{Var(n_2)}{(En_2)^2}}$ be the coefficient of variation of n_2.

Theorem 5: $1 \geq \frac{\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{En_2} \geq \left(1 + CV^2\right)^{1-\varepsilon}$

Proof in Appendix.

The share of seats allocated to the first period may deviate from the efficient equal share by as much as $\left(1 + CV^2\right)^{1-\varepsilon}$; for the case of ε=2 and CV = 1, this could be 50%. The bound in Theorem 5 is exact, in the sense that for a binomial random variable with one value equal to zero, the right inequality holds with equality, and at ε = 1, the left inequality holds with equality.

The gains from trade generated are

$$W = \frac{\alpha\varepsilon K}{\varepsilon - 1}\left(\frac{n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{K}\right)^{\frac{1}{\varepsilon}} = \frac{\alpha\varepsilon K^{\frac{\varepsilon-1}{\varepsilon}}}{\varepsilon - 1}\left(n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}\right)^{\frac{1}{\varepsilon}}.$$

The firm's revenue is

$$p_1 K = K\left(\frac{s_1}{\alpha}\right)^{\frac{-1}{\varepsilon}} = \alpha^{\frac{1}{\varepsilon}} K\left(\frac{n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{K}\right)^{\frac{1}{\varepsilon}} = \alpha^{\frac{1}{\varepsilon}} K^{\frac{\varepsilon-1}{\varepsilon}}\left(n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}\right)^{\frac{1}{\varepsilon}}$$

Thus, firm revenue is proportional to welfare under constant elasticity of demand, and the analysis of revenue maximization (under efficiency, of course) analyzes welfare maximization. To perform this analysis, we turn to the log normal case.

The inefficiency in this model arises because contracting with early customers occurs prior to the time that period two demand is realized. How large is this inefficiency? To answer that question, we consider the case where n_i is log-normally distributed. In the log normal distribution case, the welfare losses associated with the missing market can be very large.

7. *The Log Normal Case*

If the distribution of n_2 is log-normal, so that the $\log(n_2)$ is normally distributed with mean μ and variance σ^2, many of the expressions have closed forms, greatly facilitating computation. In particular,

$$En_2 = e^{\mu+\frac{1}{2}\sigma^2} \text{ and } \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon} = \left(e^{\mu/\varepsilon+\frac{\sigma^2}{2\varepsilon^2}}\right)^{\varepsilon} = e^{\mu+\frac{\sigma^2}{2\varepsilon}} \quad [29]$$

The coefficient of variation for n is

$$CV = \sqrt{\frac{Var(n)}{(En)^2}} = \sqrt{\frac{En^2-(En)^2}{(En)^2}} = \sqrt{\frac{En^2}{(En)^2}-1} = \sqrt{\frac{e^{2\mu+2\sigma^2}}{(e^{\mu+\frac{1}{2}\sigma^2})^2}-1} = \sqrt{e^{\sigma^2}-1}. \quad [30]$$

Thus,

$$\frac{\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{En_2} = e^{\frac{\sigma^2(1-\varepsilon)}{2\varepsilon}} = (1+CV^2)^{\frac{1-\varepsilon}{2\varepsilon}}, \text{ and} \quad [31]$$

$$E\left\{\frac{\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}}{n_2}\right\} = e^{\frac{\sigma^2(1+\varepsilon)}{2\varepsilon}} = (1+CV^2)^{\frac{1+\varepsilon}{2\varepsilon}}. \quad [32]$$

The expected gains from trade are

$$W = \frac{\alpha\varepsilon K^{\frac{\varepsilon-1}{\varepsilon}}}{\varepsilon-1}\left(n_1+\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}\right)^{\frac{1}{\varepsilon}} = \frac{\alpha\varepsilon K^{\frac{\varepsilon-1}{\varepsilon}}}{\varepsilon-1}\left(n_1+e^{\mu+\frac{\sigma^2}{2\varepsilon}}\right)^{\frac{1}{\varepsilon}} \quad [33]$$

How much does sequential contracting cost in terms of revenue or welfare? We treat n_1 at the expected value of a lognormal with the same variance but different mean than n_2. The proportion of the gains from trade preserved through contracting over two periods is denoted $\%\Delta W$. It has the value

$$\%\Delta W = \frac{\left(n_1 + \left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}\right)^{\frac{1}{\varepsilon}}}{E\left\{(n_1+n_2)^{1/\varepsilon}\right\}} \qquad [34]$$

We show by example that %ΔW can be as low as 64%, that is, the inability to contract in advance reduces welfare by a third relative to simultaneous contracting.[4]

This example has a much larger future population of demand than first period demand. However, the variance of demand is large too, and this causes the losses associated with uncertainty to be very large. In particular, $\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon}$ is a negligible fraction of $E\{n_2\}$ in this example, a difference which accounts for most of the efficiency loss. Even for more reasonable parameters, losses can be 15% or more. An example, with up to 20% losses, is illustrated in Figure 3. In these examples, the n_1 takes on the values 0.6, 20 and 1808, while the expected value of n_2 takes on the values 90, 2981 and 268337.

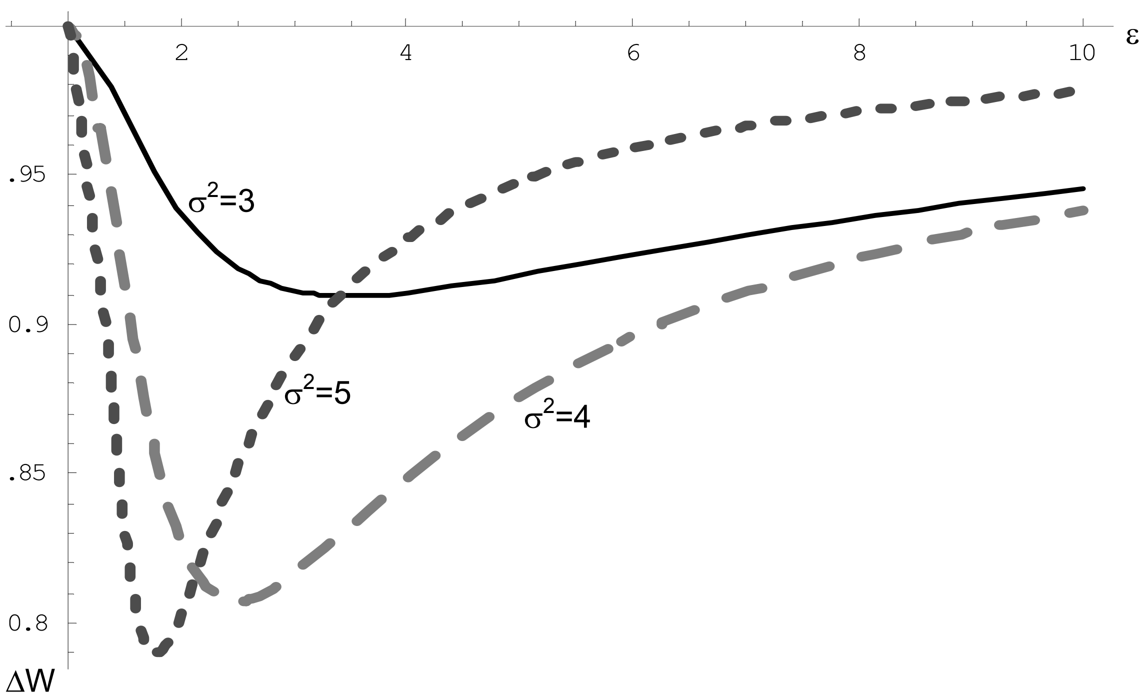

Figure 3: Efficiency loss relative to simultaneous contracting. Parameters: $\mu_1 = -5$, $\mu_2 = 0$.

[4] Specifically, one set of parameters is $\varepsilon = 3.474$, $\sigma^2 = 11.651$, $\mu = 45.153$, and $n_1 = 1.28 \times 10^{28}$. This set of parameters gives the following expected values for n_2. $E\{n_2\} = 1.22 \times 10^{49}$, and $\left(E\left\{n_2^{1/\varepsilon}\right\}\right)^{\varepsilon} = 1.24 \times 10^{28}$.

The main conclusion is that the important effects in dynamic pricing arise not from an attempt to extract more money from the consumer, but from incomplete markets, and in particular from impossibility of simultaneous contracting with all potential buyers. Dynamic pricing is used primarily to mitigate the costs of these missing markets. Moreover, welfare costs of the missing markets are potentially quite large. This conclusion suggests that options, which create markets for advance contracting, are an important aspect of both maximizing revenue and of efficiently allocating resources.

8. Options and Interruptible Sales

The welfare losses identified in the previous section arise because of incomplete markets. In particular, it is not possible to contract with the period 2 agents at the time that the period 1 market clears. This lack of contracting leads to inefficiency because sometimes too many seats are sold in the first period, when the demand in the second period is unusually high, while sometimes too few are sold, because demand was unexpectedly low. A solution to this problem generally is to sell contingent seats, that is, to sell an option.

In this case, the main value of the option is to permit the seller to sell the seat to a higher value buyer in the second period. This kind of option is little used in hotels and airlines, but is quite common in pipeline transport where it goes by the name of interruptible transmission rights, as opposed to firm (guaranteed) rights. Priceline.com sold a kind of interruptible service, where they sold the ticket well in advance but didn't specify the time of the flight until a day or so in advance.

To gain some insight into the sale of interruptible service, consider first selling two qualities of service, $\varphi_1 > \varphi_2$, where φ_i is the probability of service, and these are sold at prices $p_1 > p_2$. A consumer with type v values good i at $\varphi_i v - p_i$. A type v consumer prefers type 1 if

$$\varphi_1 v - p_1 \geq \varphi_2 v - p_2, \tag{35}$$

$$\text{or } v \geq \frac{p_1 - p_2}{\varphi_1 - \varphi_2}. \tag{36}$$

In addition, a consumer prefers good i to nothing at all if $\varphi_i v - p_i \geq 0$. If $\frac{p_1}{\varphi_1} \leq \frac{p_2}{\varphi_2}$, then no consumer ever buys good 2, so we impose the condition, without loss of generality, that $\frac{p_1}{\varphi_1} \geq \frac{p_2}{\varphi_2}$. Let F is the cumulative distribution of values. The demand for good 1 is $1 - F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right)$ and the demand for good 2 is $F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right) - F\left(\frac{p_2}{\varphi_2}\right)$.

Let c_i be the marginal cost of service type i. The seller's profits are

$$\pi = \underset{p_1, p_2}{Max} \; (p_1 - c_1)\left(1 - F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right)\right) + (p_2 - c_2)\left(F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right) - F\left(\frac{p_2}{\varphi_2}\right)\right). \qquad [37]$$

It is helpful to introduce the notation $R(x) = \underset{p}{Max} \; (p - x)(1 - F(p))$, which is the profit maximization associated with one good. Let $p^*(x) = \arg \underset{p}{Max} \; (p - x)(1 - F(p))$ be the solution to the one good maximization problem. Marginal revenue is decreasing if p* is an increasing function, so that an increase in cost reduces the quantity and increases the price charged by the monopolist. This is a standard assumption.

Theorem 6: Suppose marginal revenue is decreasing. A profit-maximizing monopolist sells the low quality good 2 if and only if $\frac{c_1}{\varphi_1} \geq \frac{c_2}{\varphi_2}$, in which case

$$\pi = (\varphi_1 - \varphi_2) R\left(\frac{c_1 - c_2}{\varphi_1 - \varphi_2}\right) + \varphi_2 R\left(\frac{c_2}{\varphi_2}\right),$$

$$p_1 = (\varphi_1 - \varphi_2) p^*\left(\frac{c_1 - c_2}{\varphi_1 - \varphi_2}\right) + \varphi_2 p^*\left(\frac{c_2}{\varphi_2}\right)$$

and $p_2 = \varphi_2 p^*\left(\frac{c_2}{\varphi_2}\right)$.

Otherwise, $\pi = \varphi_1 R\left(\frac{c_1}{\varphi_1}\right)$ and $p_1 = \varphi_1 p^*\left(\frac{c_1}{\varphi_1}\right)$.

The proof is in the appendix. The interruptible good problem breaks up into two separate maximization problems, one for the low quality good, and one for the difference of the low quality good and the high quality good. This type of result is distinct from the usual result where a monopolist selling to two types of consumers offers two qualities; here the monopolist is selling two goods to a continuum of consumers. The present result is reminiscent of Mussa and Rosen's 1978 analysis, however.

Consider a seller that didn't sell good 2, and then begins selling good 2. Does that seller's price rise, or fall? The price falls if

$$(\varphi_1 - \varphi_2) p^*\left(\frac{c_1 - c_2}{\varphi_1 - \varphi_2}\right) + \varphi_2 p^*\left(\frac{c_2}{\varphi_2}\right) < \varphi_1 p^*\left(\frac{c_1}{\varphi_1}\right) \qquad [38]$$

Thus, if p* is concave, the price falls when the second good is introduced, and the price rises when p* is convex. In the case of uniform distributions of values, or exponential distributions, or constant elasticity, p* is linear in cost, and thus introduction of the second good doesn't change the price p_1. More generally, it may increase or decrease.

In the airline context, the cost of a sale is generally a shadow cost – how much revenue is foregone because the seat is taken? Consequently, the condition $\frac{c_1}{\varphi_1} \geq \frac{c_2}{\varphi_2}$ is automatically satisfied. The cost of a probability of service φ_1 entails a loss in flexibility relative to the service level φ_2. That loss in flexibility means not rescheduling the service in circumstances in which is would be desirable to do so. Thus the cost c_1 should be an average of c_2 over the flights when the interruptible service is offered as scheduled, plus the cost in circumstances where it was profitable to interrupt. Since the latter should exceed the former given optimal interruptions, the cost of service for case 1 should exceed the cost in case 2, which means it is always optimal to offer the interruptible service.

As a practical matter, the "right" interruptible service is probably a ticket that lets the airline choose the time the passenger flies, but sets the flight in a day. The need to coordinate air travel with hotels and other destination activities restrict the ability of the most passengers to be flexible over multiple days. Nevertheless, flexibility with respect to the time of travel potentially produces significant gains in the expected value of air travel, both in the ability of the price system to identify and allocate seats to the high value passengers, and the ability to utilize more capacity.

The gains from trading options are larger when the firm posts a price that is good for an appreciable amount of time. This phenomenon was studied by Dana (1999a). Random demand will generally create a misallocation, which we modeled by allocating first period seats without yet knowing second period demand. However, under a posted price, seats will be misallocated because the posted price will generally not be the correct price to clear the market, either creating a surplus or a shortage. Consider the model developed in the previous section, with a random demand in the form nq(p) and constant elasticity $q(p)=\alpha p^{-\varepsilon}$. Let K be the capacity and suppose that when capacity binds, the customers are a random selection of those willing to purchase. The gains from trade under a constant price are

$$W = Min\{K, nq(p)\} \int_0^{q(p)} \left(\frac{x}{\alpha}\right)^{\frac{-1}{\varepsilon}} \frac{dx}{q(p)} = \frac{\varepsilon}{\varepsilon-1}\frac{1}{p} Min\{Kp^{\varepsilon}, \alpha n\} \qquad [39]$$

This equation arises because, when K<nq(p), the capacity constraint binds, and the gains from trade are the average of the value of seats over the set of consumers willing to buy those seats. One interesting feature of this equation is that efficiency requires a positive probability of empty seats. If there is a zero probability of empty seats, then the price is lower than the market clearing price. Low prices create random assignments when demand is large, which entails a loss of efficiency; this loss is balanced against the extra seats sold when demand is low. Indeed, the first order condition for maximizing W can be expressed as:

$$W = \frac{\varepsilon^2}{\varepsilon-1} Kp^{\varepsilon-1} \Pr\{K \leq nq(p)\} \qquad [40]$$

Thus, the probability that capacity binds, K≤nq(p), is positive at the price maximizing the gains from trade.

9. *Actual Airline Pricing Patterns*

In order to assess the extent of dynamic price discrimination, we collected data on four city pairs:

Los Angeles (LAX, BUR, SNA) to Las Vegas (LAS)
Bay Area (SFO, SJC, OAK) to Portland (PDX)
Dallas (DFW) to Oklahoma City (OKC)
Dallas (DFW) to Tulsa, OK (TUL)

We collected data on price offers from Orbitz, Travelocity and for part of the period, from AA.com. Unfortunately, Southwest Airlines is not featured on any of these sources. Of these nine airport pairs, four (LAX and BUR to LAS, SJC and OAK to PDX) are also served by Southwest. One would expect Southwest to exert a moderating influence, given its relatively low variance in prices, although Southwest has introduced a six fare structure which may result in more dispersion, rather than less.[5] In any case, with the data available, the effects of Southwest's presence will have to remain a topic for further research. Purchasing an American Airlines ticket on AA.com is generally $5 cheaper, but showed no other difference. The figures below only show fares quoted by Orbitz, which had the most extensive offerings.

We collected data on flights, all in the year 2004, departing on 8/26, 8/28 8/29, 9/09, 9/11, 9/12, 9/23, 9/25, 9/26, 10/7, 10/9, 10/10, 10/21 10/23 and 10/24. Six airlines represent over 99% of the data, with the following proportions of the flights. Average prices and the frequency of airlines are presented in Table 1.

Table 1: Average Prices in the Data

Airline	AA	Alaska	Delta	United	AmWest	USAir	NW
Proportion of Flights	25.5%	22.7%	6.1%	14.6%	8.9%	16.1%	6.1%
Average Price	$108.05	$126.59	$79.03	$44.06	$66.17	$239.01	$191.48

There are five major propositions that can be tested with the data.

1. Prices fall as takeoff approaches
2. Prices are rising initially
3. Competition reduces variance of prices
4. Prices change with each seat sold
5. Prices of substitutes are correlated
    a. Substitute times/days
    b. Substitute airports

The first proposition is that prices are falling as takeoff approaches. This is a remarkably robust prediction of theories that have identical customer types arriving over time. The robustness follows from the following observation. The cost of selling a seat is composed of three components. First, there is the cost of service. Second, there is the cost of not selling that seat on a substitute flight. This second component includes any psychological effects on consumer expectations – consumers may reason that current low prices are indicators of likely low future prices, which tends to make demand more elastic, to the airline's detriment. Third, there is the foregone option of selling the seat later on the same flight. The first two costs are roughly constant over time, and certainly approximately constant as the flight

[5] The six classes, with sample fares (BUR to LAS) are Refundable Anytime ($94), Restricted ($86), Advance Purchase ($72), Fun Fares ($49), Promotional ($39), and Internet One-way ($39).

approaches. The third cost, however, goes to zero as the flight approaches. Thus, most theories will suggest that prices fall in the last moments before a flight, not necessarily to zero but to some optimal price reflecting the market power of the airline and the effect of low-price sales on sales of future flights. But the price should still fall.

A more specialized prediction is that the average price rises initially. This is a feature of the Gallego-van Ryzin model already discussed, and others. It is more intuitive in models in which consumers can search than in the more common model where consumers who don't purchase disappear; if prices are falling on average, consumers will tend to wait to purchase. Late falling prices aren't so obviously exploitable because of the significant likelihood of the plane filling up. Thus, consumers must weigh the expected gains from delay against the costs of failing to acquire a seat at all; the latter cost is negligible with a sufficiently long time-horizon.

To assess these propositions, we ran the following regression to predict the price. We have dummy variables for days with one, two, three and four weeks remaining, airline-specific dummies (American omitted), hour of takeoff (overnight omitted), and city-pair identifiers (LAX to LAS omitted). We omitted Burbank, Long Beach and SNA airports based on few flights over the period. We only examined one-way economy fares, and have 138371 observations. The 38 variables account for 73.5% of the variation.

Prices rise $50 in the week before takeoff, which is on top of a rise of $28.20 the previous week. The penultimate fortnight before takeoff accounts for $16 in price increases. There is a slight, 1.9 cent per day increase prior to that. Thus, the main time-path prediction of the standard theory appears to be empirically false. These effects are quite significant, thanks to the size of the database. (We have not attempted to control for correlation in observations, so that the t-statistics, which treat observations as independent, likely overstate the significance.) The important conclusion is that changing demand is a salient feature of the data, and models that assume that the shape of demand is constant over time are empirically invalid.

A few other remarks follow from the regressions. Although American Airlines is a low priced airline in the data, it is high-priced adjusting for fixed effects, though not as expensive as US Air. Prices are highest mid-day and early evening. LAX is a relatively inexpensive airport, while San Francisco is high-priced. This suggests that competition and thick markets do not provide a good account of pricing, since San Francisco has a large number of flights. Southwest's presence at LAX, OAK and SJC might help explain the average pricing. Note, however, that airport dummies are also accounting for the length of flights, so comparisons should only be made among distinct airport pairs for a given city-pair.

Table 2: Price Regression

	Estimate	SE	TStat	PValue
Constant	54.1	0.9906	54.6	1.1×10^{-642}
Days	0.0192	0.0067	2.9	0.0042
WK1	50.0	0.6071	82.4	6.4×10^{-1441}
WK2	28.2	0.6091	46.3	1.7×10^{-464}
WK3	10.6	0.6511	16.3	0.
WK4	5.8	0.6364	9.1	0.
Alaska	- 4.6	0.5006	- 9.2	4.3×10^{-20}
Delta	- 24.1	0.6112	- 39.4	6.4×10^{-337}
United	- 6.3	0.5894	- 10.6	2.4×10^{-26}
Amwest	6.5	0.6725	9.6	0.
USAir	86.8	0.5801	149.6	3.1×10^{-4508}
NW	- 16.9	0.7070	- 23.9	8.6×10^{-126}
AM6	- 15.5	0.8189	- 18.9	2.0×10^{-79}
AM7	- 6.3	0.8380	- 7.5	5.5×10^{-14}
AM8	- 3.7	0.8348	- 4.4	0.
AM9	- 1.4	0.8806	- 1.6	0.1173
AM10	9.9	0.9115	10.8	0.
AM11	3.7	0.8221	4.5	6.1×10^{-6}
Noon	5.9	0.8504	6.9	4.7×10^{-12}
PM1	- 3.6	0.8696	- 4.2	0.
PM2	- 1.3	0.9844	- 1.3	0.1932
PM3	2.6	0.7849	3.3	0.0009
PM4	- 1.2	0.9832	- 1.3	0.2044
PM5	7.4	0.8073	9.2	0.
PM6	- 3.4	0.8424	- 4.1	0.
PM7	3.5	0.8061	4.3	0.
PM8	- 5.8	0.9118	- 6.4	2.0×10^{-10}
PM9	- 2.0	1.0	- 2.0	0.0496
TUL2DFW	41.9	0.7870	53.2	4.7×10^{-611}
OKC2DFW	41.9	0.8412	49.8	1.2×10^{-535}
DFW2TUL	41.5	0.8068	51.4	7.4×10^{-571}
DFW2OAK	31.8	0.8381	37.9	4.5×10^{-313}
SFO2PDX	149.6	0.5144	290.8	3.9×10^{-14331}
SJC2PDX	44.6	0.8246	54.1	3.1×10^{-632}
OAK2PDX	49.4	0.7120	69.4	8.7×10^{-1031}
PDX2SFO	152.1	0.5264	288.9	6.4×10^{-14184}
PDX2SJC	43.7	0.8253	53.0	1.3×10^{-605}
PDX2OAK	46.5	0.7352	63.2	1.8×10^{-858}
LAS2LAX	- 6.9	0.4877	- 14.1	2.9×10^{-45}

In another paper (Carbonneau, McAfee, Mialon & Mialon 2004), it is shown that the third proposition is not supported in a large dataset of airline prices. Indeed, more competition was weakly correlated with more dispersion, not less. This finding is consistent with Borenstein and Rose's 1994 findings.

Some theories posit a "two price" structure, others permit prices to vary continuously. In the data, some airlines clearly lean toward a two price structure. For example, American Airlines' prices on OAK to PDX show evidence of two main prices, with a third just prior to takeoff. This is illustrated in Figure 4.[6] Most of Northwest's fares show just two prices.

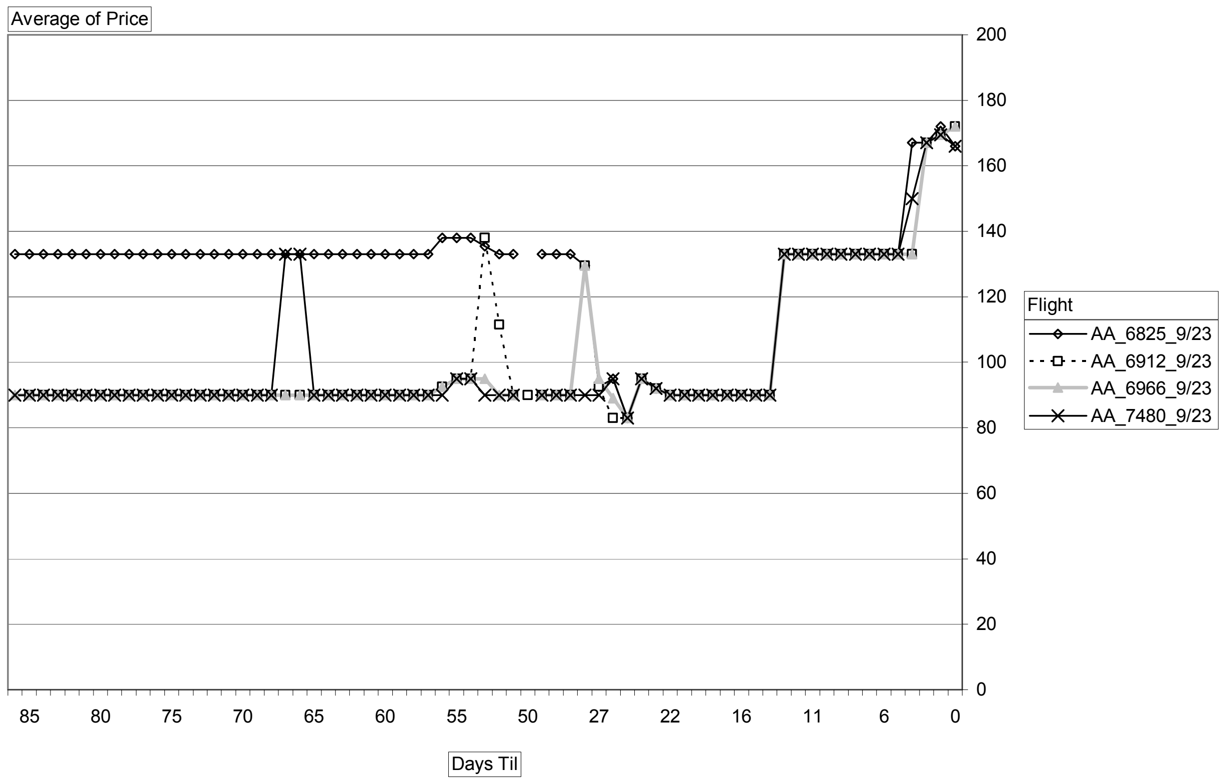

Figure 4: American OAK to PDX, September 23.

However, other flights show evidence of almost continuous adjustment, and indeed sometimes American appears almost to randomize its offers. Figure 5 shows an example of two American Airlines flights, AA 1038 on Sept 23 and Sept 25, with remarkable variation.

So the evidence in favor of this proposition – continuous adjustment of prices – is mixed. Why do airlines use two prices? A standard economic explanation is that there is a value in price commitment to assist in consumer planning. This is the reason given for restaurants, movies, and the like to maintain a constant price, or two prices, in the face of predictably fluctuating demand. However, prices don't seem nearly predictable enough for predictability to be the reason for using only two prices. A more plausible theory is that airlines use two prices because the theory is better developed for two price systems. This is plausible for a consistent airline like Northwest, but implausible for American Airlines, which has extraordinary price adjustments on other routes.

[6] The figures identify flights Airline_Flight Number_Departure Date. All data in the figures comes from Orbitz's website.

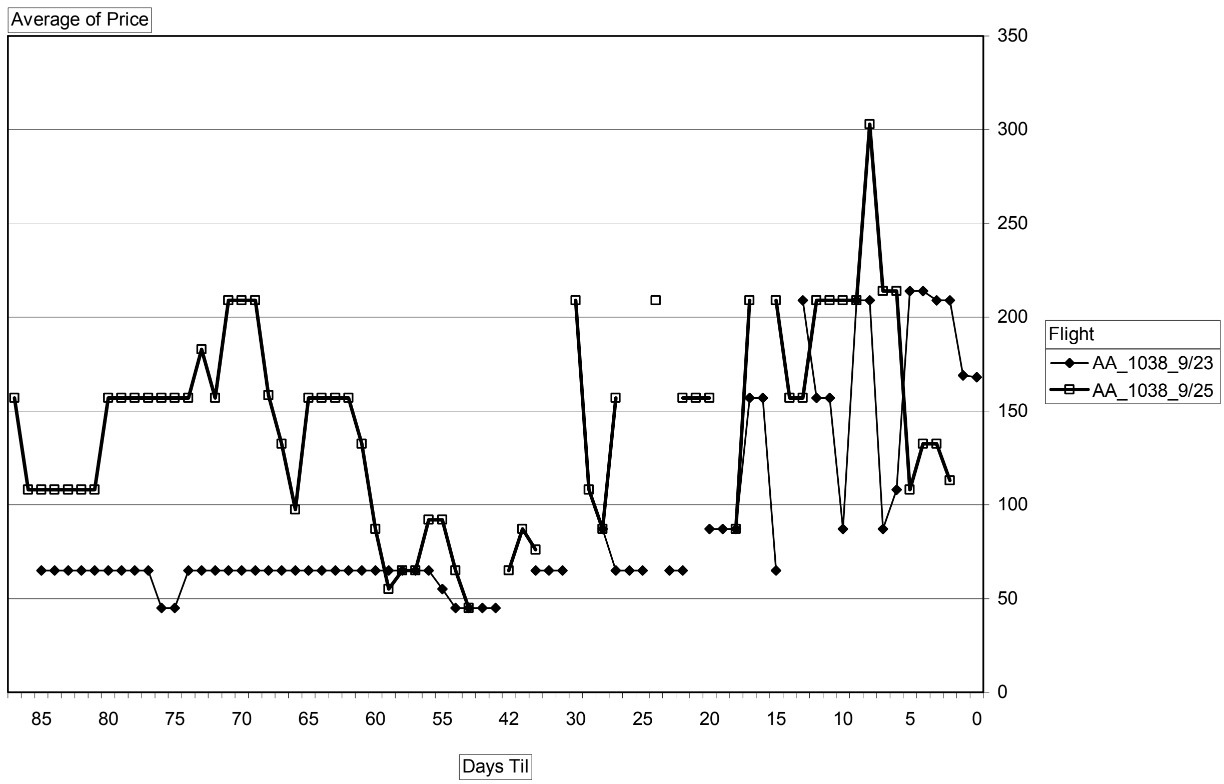

Figure 5: American 1038 prices plotted against days prior to 9/25

The final prediction of the theories is that the prices of substitutes should be correlated. This, again, is a robust prediction and requires only consumer substitution as a hypothesis. To assess this, we looked at the correlation between prices of flights from the bay area to Portland, and Portland to the bay area. Based on travel times, it seems that SFO and SJC should be the closest substitutes, followed by SFO and OAK, followed by SJC and OAK. In fact, however, the correlations are almost exactly the reverse.

Table 3: Correlation Coefficients

	Flight Pair	ρ	Flight Pair	ρ
1	SFO2PDX, SJC2PDX	0.167	PDX2SFO, PDX2SJC	0.596
2	SFO2PDX, OAK2PDX	0.190	PDX2SFO, PDX2OAK	0.568
3	SJC2PDX, OAK2PDX	0.867	PDX2SJC, PDX2OAK	0.954

Table 3 shows the correlation coefficient of average prices, with the correlation between mean price offers on each route as a function of the number of days prior to departure. The prediction is that the first row of the table should have the highest correlation, followed by the second, followed by the third. However, approximately the opposite arises. The presence of Southwest at San Jose and Oakland, but not at San Francisco, may be the explanatory factor, especially in light of the fact that the price level at SFO is much higher than at the other two airports.

It is more challenging to assess whether there is strong correlation between substitute flights on the same route. The following picture shows some correlation over time of the substitute flights between Portland

and Oakland at 6PM, 7PM and 8PM (± ½ hour). These flights have an average correlation coefficient of about 0.8 overall.

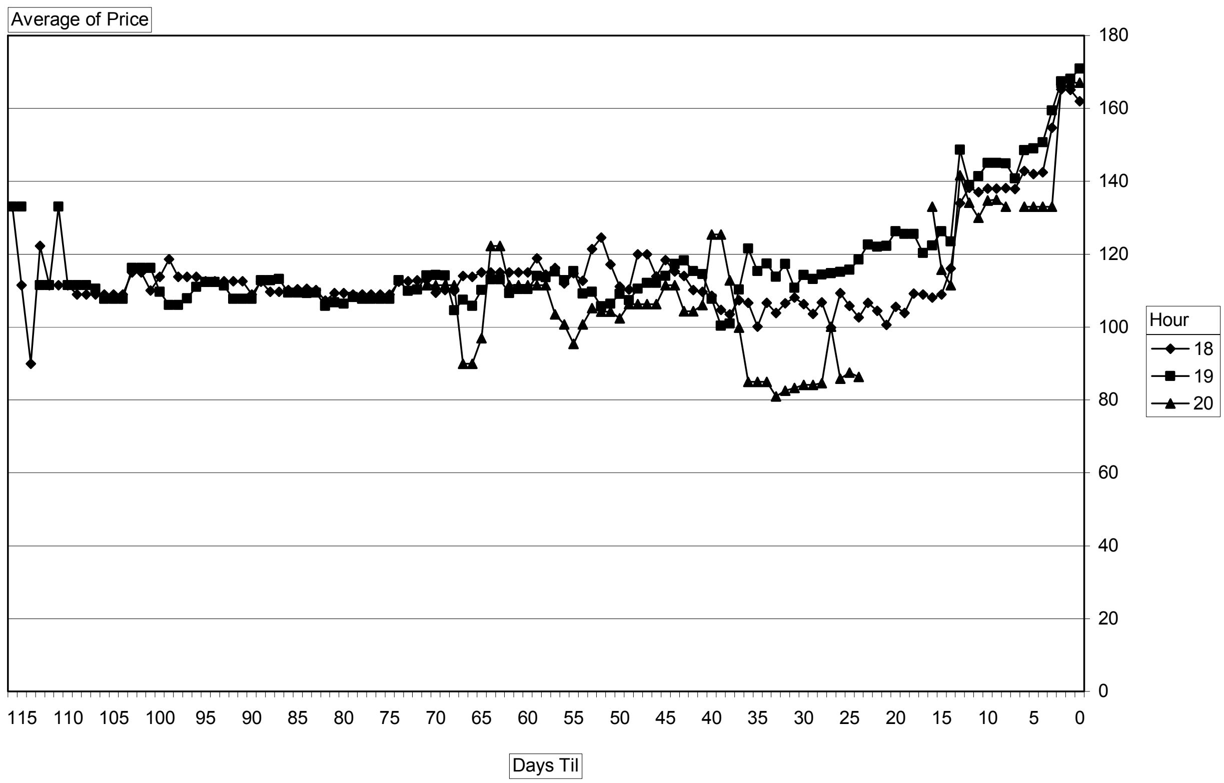

Figure 6: PDX to OAK, return, at various times

On the other hand, Figure 7 presents a similar examination of economy flights from LAX to LAS. No such correlation is readily apparent and these flights are weakly correlated. Afternoon flights, presented in Figure 8, are similar.

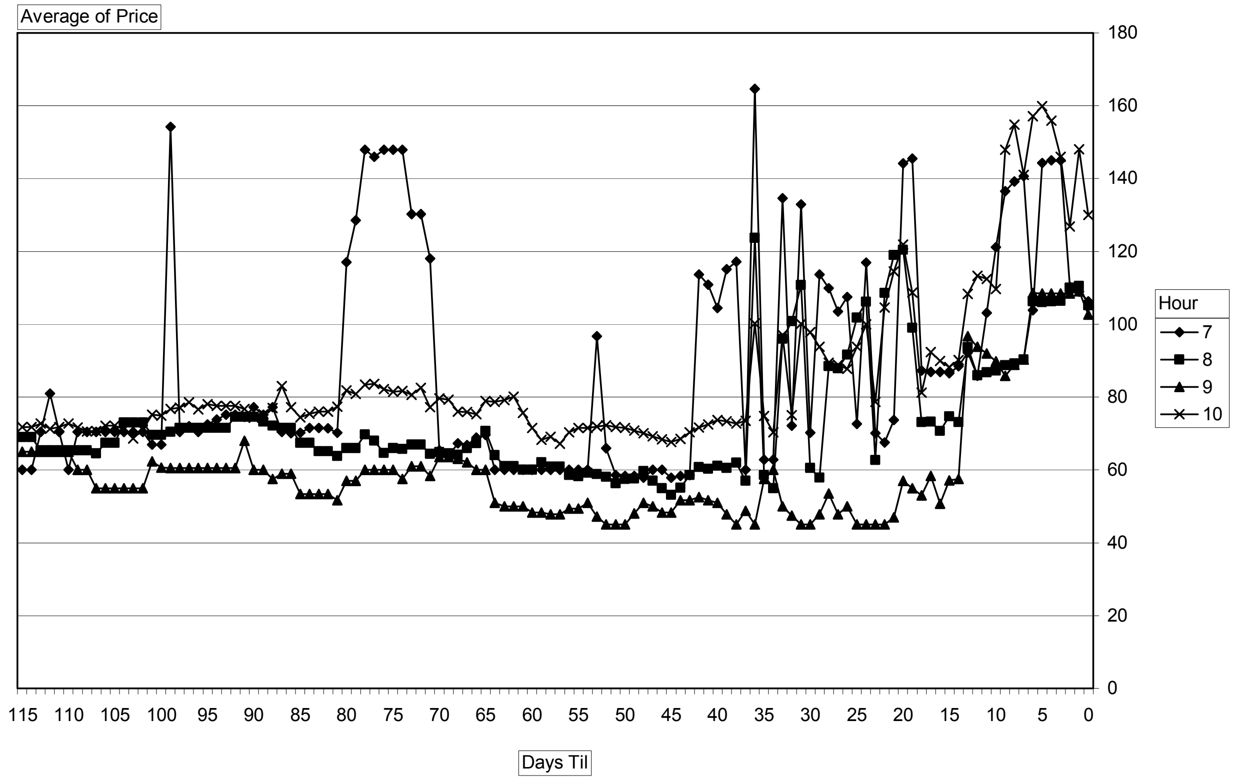

Figure 7: LAX to LAS Economy class, 6AM to 9AM

Overall, the major predictions – including the most robust predictions – of the theories appear to be violated in the data. Consequently, new theories, and probably a new approach to the analysis, are needed.

10. Research Projects and Mysteries

There are many research projects that flow from the analysis of this paper. First, much of the existing work is founded on a monopoly description. The analysis of this paper suggests that recasting the model as a competitive model may increase the tractability of the model without sacrificing either empirical relevance or complexity of behavior.

Second, the sale of options present the possibility of enhancing the efficiency of advance contracting, a possibility that has been little explored. Continuous time techniques from the finance literature may play an important role in developing a theory of options for perishable goods. This seems a fruitful approach, because options themselves are perishable, and continuous time stochastic calculus techniques have played an important role in simplifying similar problems.

The existing theory fared poorly when confronted with the data. In particular, the failure of prices to fall as takeoff approaches is devastating to theories, leaving standing only those theories in which late arriving potential passengers have a relatively high willingness-to-pay. While this is a reasonable hypothesis, it nonetheless needs further development, and it is challenging to think of natural assumptions to impose order on the problem, once demand can be almost anything.

The data present a variety of mysteries. The gains to searching are occasionally enormous. While prices rise as takeoff approaches, occasionally they bounce around considerably. This finding is mirrored in Etzioni et al, which empirically examined the gains to searching for flights. They find an average savings of 27% by using a search algorithm, relative to just booking the flight the first time it is checked.

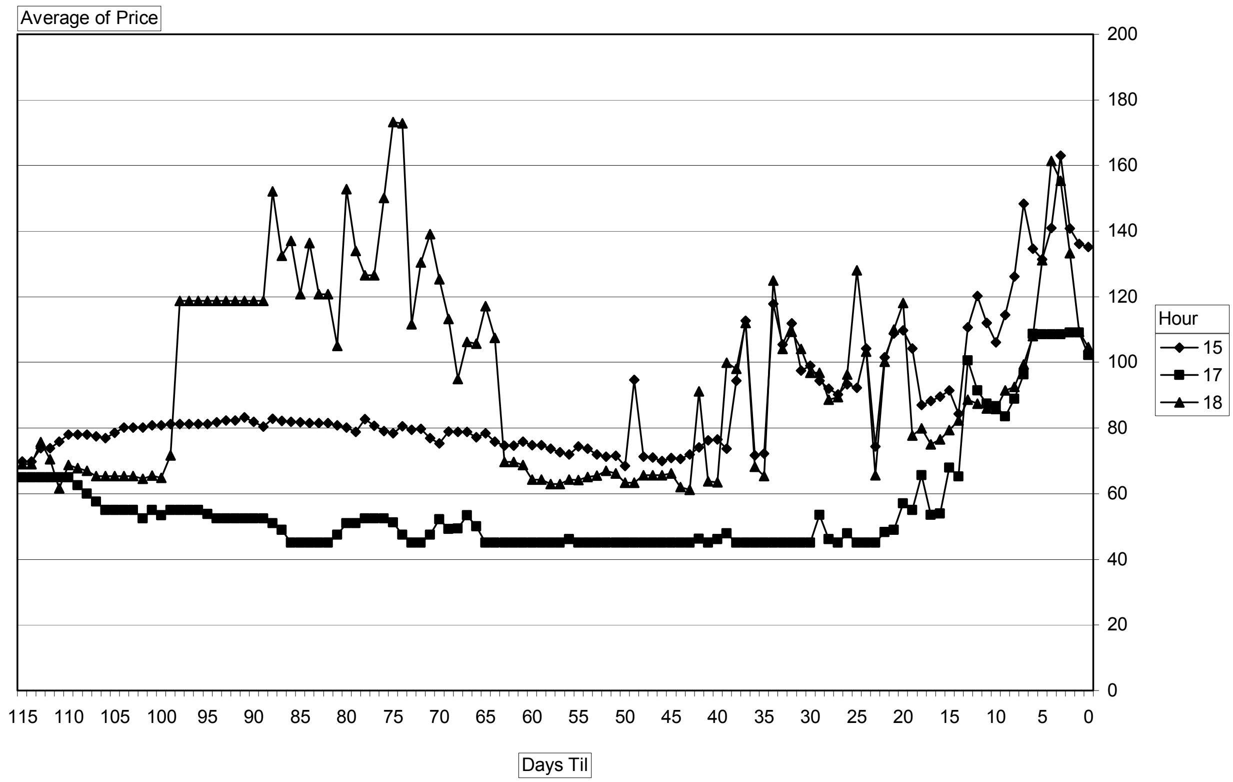

Figure 8: Afternoon LAX to LAS Flights by time of departure

What is perhaps even more mysterious, and illuminating, is an incident that occurred in mid-July on Alaska 101, a flight from OAK to PDX. For approximately a week, the price of AL101 departing 9/23 nearly doubled, while the prices of the 9/25 and 9/26 departures were approximately unchanged. The 9/26 departure had been more expensive. In fact, the price of the 9/26 departure then fell briefly. This is illustrated in Figure 9.

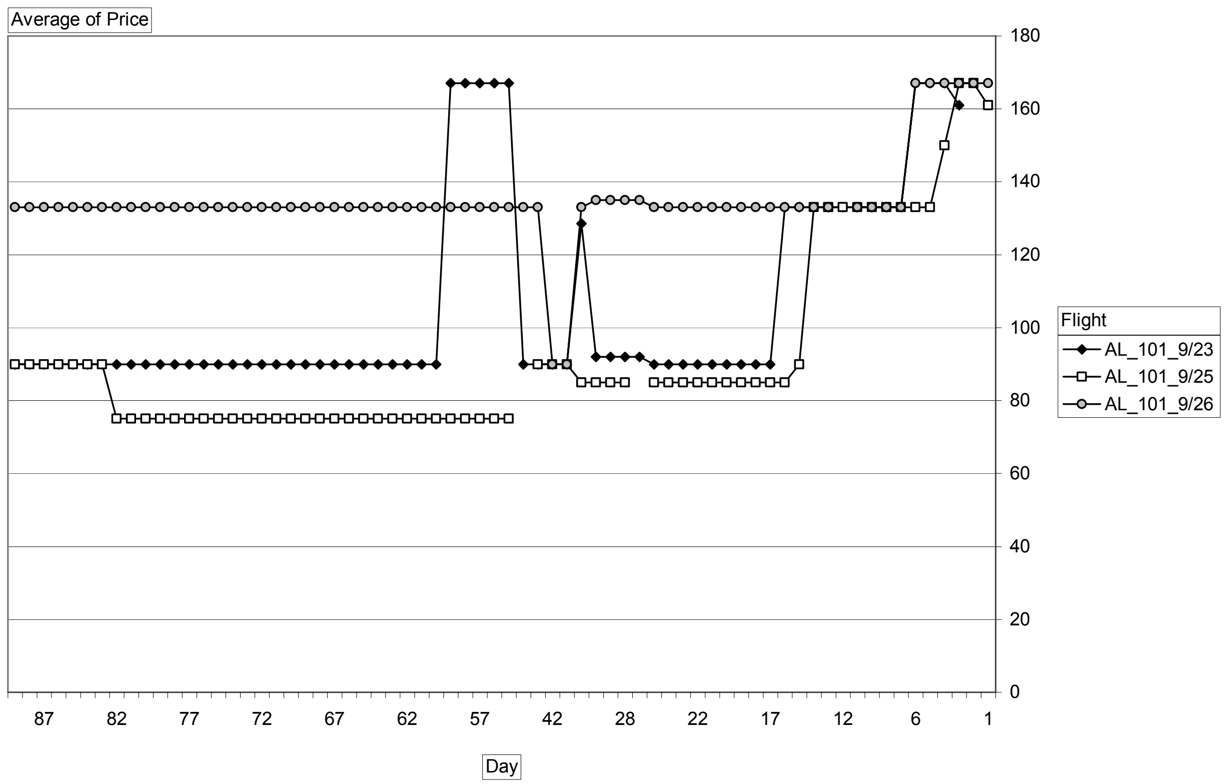

Figure 9: Prices of AL101, departing 9/23, 9/25 and 9/26

Given its departure time, American Airlines 6825 is a major substitute for AL101. During this time, AA 6825's flight experienced only a very modest increase in price, which is illustrated in Figure 10.

The price of AA6825 had been substantially higher than the price of AL 101, but when AL 101 went up, the price of AL 101 was much higher. The price of the AA 6825 went up slightly, then came back down to the previously prevailing levels. What makes this so mysterious is that AA 6825 and AL 101 represent the same airplane – AA 6825 is a code-share flight operated by Alaska airlines. That is, the real mystery is why these two identical fights are being marketed at such different prices, when the fact that the flights are code-shared is clearly indicated on the websites.

To make matters even more mysterious, the blip in prices of AL 101 coincides with a slight increase in the price of AA 6825 for flights departing on other days, but not the price of AL101 on other days. This is illustrated in Figure 11.

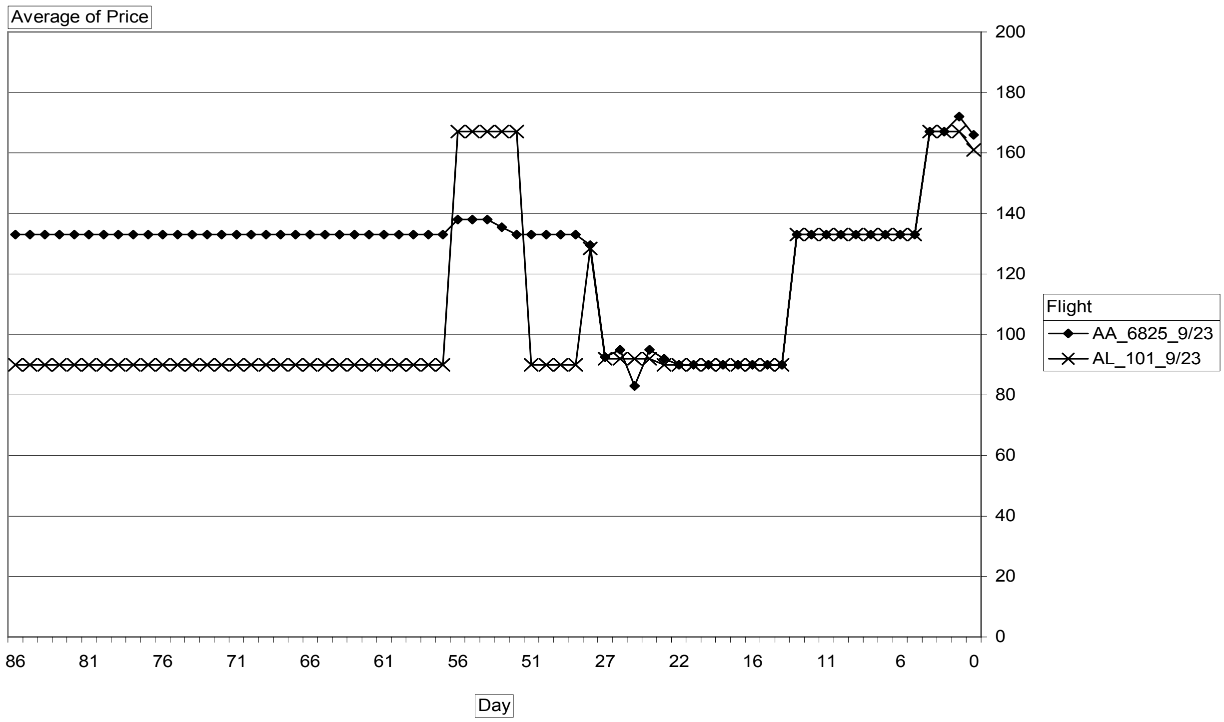

Figure 10: Prices of AL101, AA 6825

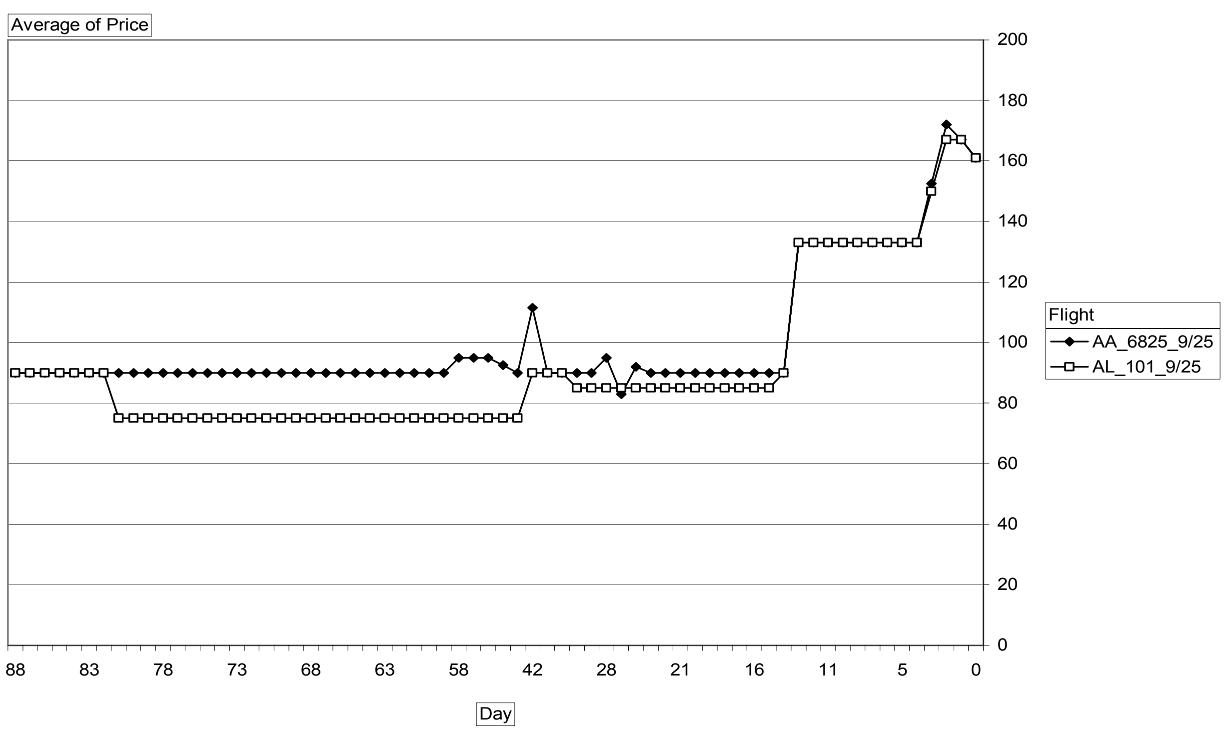

Figure 11: AL101, AA 6825 departing 9/25

It seems likely that American reacted to the price of its rival. The facts that (i) the identical flight was being offered for so much more when booked as a code-share on American rather than on Alaska, (ii) Alaska's price increase wasn't reflected in the prices of alternate days, (iii) the price of American's identical flight only rose slightly and also rose on alternative days, suggest that (1) consumers bring a substantial brand loyalty to the flights, (2) airlines use a fare program that reacts to pricing by others across a set of substitute flights, and (3) the process is poorly understood by researchers.

The ease of collecting data for non-commercial use suggests that reverse-engineering the pricing mechanisms of airlines is a feasible and potentially very interesting research project.

11. Conclusion

This paper has considered so many different theories and data that it is worth emphasizing a few highlights.

Dynamic price discrimination is primarily driven by customer dynamics rather than price discrimination over an existing set of customers.

With a large number of units to sell, the per unit gain in profits of dynamic price discrimination over a constant price is small, although the total gain will still be positive. Most sales take place at an approximately constant price; dynamic price discrimination is advantageous only as the probability of actually selling out changes, for a relatively small portion of a large number of sales. One way to summarize this conclusion is that dynamic price discrimination only matters significantly on the last twenty or so sales.

Monopoly and efficient dynamic pricing may be observationally equivalent, and are in the important model of Gallego and van Ryzin (1994). Directly solving for efficient solutions presents an alternative approach with potential empirical merit and relative tractability.

The most important effects in dynamic price discrimination arise not from an attempt to extract more money from the consumer, but from addressing incomplete markets, and in particular from the value and costs of advance contracting. Options, which create markets for advance contracting, are an important aspect of both maximizing revenue and of efficiently allocating resources.

The interruptible good problem breaks up into two separate maximization problems, one for the low quality good, and one for the difference of the low quality good and the high quality good. In the airline context, an interruptible good is one that provides the airline greater flexibility with respect to the time of flight. The cost of delivering a seat reserving greater airline flexibility is automatically lower, and thus is part of both profit maximization and efficient provision of services.

Efficiency requires a positive probability of empty seats. Pricing to sell out is inefficient.

Systematic changes in demand are a salient feature of the data, and models that assume that early and late arrivals are identical are empirically invalid.

References

Badinelli, Ralph D. 2000, 'An optimal, dynamic policy for hotel yield management', *European Journal of Operational Research*, vol. 121, no. 3, pp. 476-503.

Barut, M., and Sridharan, V. 2004, 'Design and evaluation of a dynamic capacity apportionment procedure', *European Journal of Operational Research*, vol. 155, no. 1, pp. 112-133.

Belobaba, Peter P., and Wilson, John L. 1997, 'Impacts of yield management in competitive airline markets', *Journal of Air Transport Management*, vol. 3, no. 1, pp. 3-9.

Board, Simon 'Dynamic Monopoly', Stanford University, unpublished manuscript.

Bodily, S.E., and Weatherford, L.R. 1995, 'Perishable-asset revenue management: generic and multiple-price yield management with diversion', *International Journal of Management Science*, vol. 23, no. 2, pp. 173-185.

Boiteaux, Marcel 1949, 'La tarification des demandes en point: application de la theorie de la vent au cout marginal', *Revue Generale de l'Electricite*, vol. 58 (August), pp. 321-340. (English translation available in *J Business* 33, April 1960.)

Borenstein, Severin, and Rose, Nancy 1994, 'Competition and price dispersion in the U.S. airline industry', *Journal of Political Economy*, vol. 103 (August), pp. 653-683.

Botimer, Theodore C. 1996, 'Efficiency considerations in airline pricing and yield management', *Transportation Research Part A*, vol. 30, no. 4 (July), pp. 307-317.

Brooks, M., and Button, K.J. 1994, 'Yield management: a phenomenon of the 1980s and 1990s?', *International Journal of Transport Economics*, vol. 21, no. 2 (June), pp. 177-196.

Carbonneau, Shane, McAfee, R. Preston, Mialon, Hugo and Mialon, Sue 2004, 'Price discrimination and market power', unpublished.

Carr, Scott, and Lovejoy, William 2000, 'The inverse newsvendor problem: choosing an optimal demand portfolio for capacitated resources', *Management Science*, vol. 46, no. 7 (July), pp. 912-927.

Chatwin, Richard E. 2000, 'Optimal dynamic pricing of perishable products with stochastic demand and a finite set of prices', *European Journal of Operational Research*, vol. 125, no. 1, pp. 149-174.

Dana, Jr., James D. 1999a, 'Equilibrium price dispersion under demand uncertainty: the roles of costly capacity and market structure', *RAND Journal of Economics*, vol. 30, no. 4 (Winter), pp. 632-660.

Dana, Jr., James D. 1999b, 'Using yield management to shift demand when the peak time is unknown', *RAND Journal of Economics*, vol. 30, no. 3 (Autumn), pp. 456-474.

Davis, Paul 1994, 'Airline ties profitability yield to management', *SIAM News*, vol. 27, no. 5 (May/June).

de Boer, Sanne V., Freling, Richard, and Piersma, Nanda 2002, 'Mathematical programming for network revenue management revisited', *European Journal of Operational Research*, vol. 137, no. 1, pp. 72-92.

deB. Harris, Frederick H., and Pinder, Jonathan P. 1995, ‘A revenue management approach to demand management and order booking in assemble-to-order manufacturing’, *Journal of Operations Management*, vol. 13, no. 4, pp. 299-309.

Deneckere, Raymond and McAfee, R. Preston 1996, ‘Damaged goods’, *Journal of Economics and Management Strategy*, vol. 5, no. 2 (Summer), pp. 149-174.

Etzioni, Oren, Knoblock, Craig A., Tuchinda, Rattapoom, and Yates, Alexander 2003, ‘To buy or not to buy: mining airline fare data to minimize ticket purchase price’, Proceeding of the *9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington DC.

Fan, Meng, and Wang, Ke 1998, ‘Optimal harvesting policy for single population with periodic coefficients’, *Mathematical Biosciences*, vol. 152, no. 2, pp. 165-177.

Feng, Youyi, and Gallego, Guillermo 2000, ‘Perishable asset revenue management with markovian time dependent demand intensities’, *Management Science*, vol. 46, no. 7 (July), pp. 941-956.

Feng, Youyi, and Xiao, Baichun 2000a, ‘Optimal policies of yield management with multiple predetermined prices’, *Operations Research*, vol. 48, no. 2 (March-April), pp. 332-343.

Feng, Youyi, and Xiao, Baichun 2000b. ‘A continuous-time yield management model with multiple prices and reversible price changes’, *Management Science*, vol. 46, no. 5 (May), pp. 644-657.

Feng, Youyi, and Xiao, Baichun 2001. ‘A dynamic airline seat inventory control model and its optimal policy’, *Operations Research*, vol. 49, no. 6 (November-December), pp. 938-949.

Gallego, Guillermo, and van Ryzin, Garrett 1994, ‘Optimal dynamic pricing of inventories with stochastic demand over finite horizons’, *Management Science*, vol. 40, no. 8 (August), pp. 999-1020.

Garcia-Diaz, Alberto, and Kuyumcu, Ahmet 1997, ‘A cutting-plane procedure for maximizing revenues in yield management’, *Computers and Industrial Engineering*, vol. 33, no. 1-2, pp. 51-54.

Kimes, Sheryl E. 1989, ‘Yield management: a tool for capacity-constrained service firms’, *Journal of Operations Management*, vol. 8, no. 4 (October), pp. 348-363.

Kuyumcu, A., and Garcia-Diaz, A. 2000, ‘A polyhedral graph theory approach to revenue management in the airline industry’, *Computers & Industrial Engineering*, vol. 38, no. 3, pp. 375-396.

Ladany, S.P. 1996, ‘Optimal market segmentation of hotel rooms-the non-linear case’, *International Journal of Management Science*, vol. 24, no. 1, pp. 29-36.

Lai, Kin-Keung, and Ng, Wan-Lung 2003, ‘A stochastic approach to hotel revenue optimization’, *Computers & Operations Research*, Article in Press.

Li, Michael Z.F. 2001, ‘Pricing non-storable perishable goods by using a purchase restriction with an application to airline fare pricing’, *European Journal of Operational Research*, vol. 134, no. 3, pp. 631-647.

McAfee, R. Preston, ‘Coarse matching’, *Econometrica*, vol. 70, no. 5, pp. 2025-2034.

McGill, Jeffrey I., and Van Ryzin, Garrett J. 1999, 'Revenue management: research overview and prospects', *Transportation Science*, vol. 33, no. 2 (May), pp. 233-256.

Mussa, Michael and Rosen, Sherwin 1978, 'Monopoly and product quality', *Journal of Economic Theory*, vol. 18, pp. 301-317.

Oren, Shmuel, Smith, Stephen, and Wilson, Robert 1985, 'Capacity pricing', *Econometrica*, vol. 53 (May), pp. 545-566.

Sawaki, K. 2003, 'Optimal policies in continuous time inventory control models with limited supply', *Computers & Mathematics with Applications*, vol. 46, no. 7, pp. 1139-1145.

Sridharan, 'Sri' V. 1998, 'Managing capacity in tightly constrained systems', *International Journal of Production Economics*, vol. 56-57, pp. 601-610.

Stokey, Nancy L. 1979, 'Intertemporal price discrimination', *The Quarterly Journal of Economics*, vol. 93, no. 3 (August), pp. 355-371.

Talluri, Kalyan T. and van Ryzin, Garrett J. 2004, *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers, Dordrecht, Netherlands.

Van Slyke, Richard, and Young, Yi 2000, 'Finite horizon stochastic knapsacks with applications to yield management', *Operations Research*, vol. 48, no. 1 (January-February), pp. 155-172.

Williams, Loren 1999, 'Revenue management: microeconomics and business modeling', *Business Economics* (April), pp. 39-45.

Williamson, Oliver 1966, 'Peak load pricing and optimal capacity under indivisibility constraints', *American Economic Review*, vol. 56 (September), pp. 810-827.

Wilson, Robert 1989, 'Efficient and competitive rationing', *Econometrica*, vol. 57 (January).

Zhao, Wen, and Zheng, Yu-Sheng 2000, 'Optimal dynamic pricing for perishable assets with nonhomogeneous demand', *Management Science*, vol. 46, no. 3 (March), pp. 375-388.

Appendix

Proof of Theorem 2:

The evolution of the probability that there are exactly i unsold seats is governed by the probability of selling a seat when there are i+1 unsold, and the probability of selling a seat when there are i unsold, so that

$$q_i'(t) = \lambda(1 - F(p_{i+1}(t))q_{i+1}(t) - \lambda(1 - F(p_i(t))q_i(t)$$

$$= \lambda e^{p_{i+1}(t)} q_{i+1}(t) - \lambda e^{p_i(t)} q_i(t)$$

$$= \frac{B_i'(t)}{B_i(t)} q_i(t) - \frac{B_{i+1}'(t)}{B_{i+1}(t)} q_{i+1}(t)$$

Given a capacity k at time 0, $q_k(0)=1$, $q_{k+1}(t)=0$, and

$$q_k(t) = \frac{B_k(t)}{B_k(0)}.$$

This is used as the base of an induction to establish the theorem. Suppose the theorem is true at $n \le k$. Then

$$q_{n-1}'(t) = \frac{B_{n-1}'(t)}{B_{n-1}(t)} q_{n-1}(t) - \frac{B_n'(t)}{B_n(t)} q_n(t) \qquad [41]$$

$$= \frac{B_{n-1}'(t)}{B_{n-1}(t)} q_{n-1}(t) - \frac{B_n'(t)(\beta t)^{k-n}}{B_k(0)(k-n)!}$$

This linear differential equation, with the boundary condition $q_{n-1}(0)=0$, gives

$$q_{n-1}(t) = B_{n-1}(t) \int_0^t \frac{1}{B_{n-1}(s)} \frac{B_n'(s)(\beta s)^{k-n}}{B_k(0)(k-n)!} ds \qquad [42]$$

$$q_{n-1}(t) = B_{n-1}(t) \int_0^t \frac{\beta(\beta s)^{k-n}}{B_k(0)(k-n)!} ds = \frac{(\beta t)^{k-n+1}}{B_k(0)(k-n+1)!} \qquad [43]$$

The expected number of unsold items, k-n, satisfies

$$E(k-n) = \sum_{n=0}^{k} (k-n) q_n(t) = \sum_{n=0}^{k-1} (k-n) \frac{(\beta t)^{k-n} B_n(t)}{(k-n)!\, B_k(0)} \qquad [44]$$

$$= \beta t \sum_{n=0}^{k-1} \frac{(\beta t)^{k-n-1} B_n(t)}{(k-n-1)!\, B_k(0)}$$

$$= \beta t \frac{B_{k-1}(0)}{B_k(0)} \sum_{n=0}^{k-1} \frac{(\beta t)^{k-n-1} B_n(t)}{(k-n-1)!\, B_{k-1}(0)}$$

$$= \beta t \frac{B_{k-1}(0)}{B_k(0)} \sum_{n=0}^{k-1} q_n^{k-1} = \beta t \frac{B_{k-1}(0)}{B_k(0)}. \qquad \text{Q.E.D.}$$

Lemma A1: $\displaystyle \sum_{j=0}^{k} \frac{k!(xk)^{j-k}}{j!} \xrightarrow[k \to \infty]{} \begin{cases} \dfrac{x}{x-1} & if\ x > 1 \\ \infty & if\ x < 1 \end{cases}$

Proof of Lemma A1: Note $\displaystyle \sum_{j=0}^{k} \frac{k!(xk)^{j-k}}{j!} = e^{xk} kx \int_1^{\infty} t^k e^{-xkt} dt = kx \int_1^{\infty} \left(te^{-x(t-1)}\right)^k dt$

Let $\psi(t) = te^{-x(t-1)}$. If x<1, ψ(t) ≥ 1 for t ∈ [1,1/x], so

$$kx \int_1^{\infty} \left(te^{-x(t-1)}\right)^k dt \xrightarrow[k \to \infty]{} \infty.$$

Now suppose x>1. For t≥1, ψ is decreasing. Locally, around 1, $\psi(t) \approx 1 + (1-x)(t-1)$. Thus,

$$kx \int_1^{\infty} \left(te^{-x(t-1)}\right)^k dt \approx kx \int_1^{1+\varepsilon} \left(1 + (1-x)(t-1)\right)^k dt$$

$$= \frac{x}{x-1} \frac{k}{k+1} \left(1 - (1+\varepsilon(1-x))^k\right) \xrightarrow[k \to \infty]{} \frac{x}{x-1}. \qquad [45]$$

The proof for x=1 can be handled by observing that $\sum_{j=0}^{k} \frac{k!(xk)^{j-k}}{j!}$ is non-increasing in x.

Q.E.D.

Lemma A2: $\frac{1}{k} Log\left(\sum_{j=0}^{k} \frac{(xk)^j}{j!} \right) \underset{k \to \infty}{\to} \begin{cases} x & \text{if } x < 1 \\ 1 + Log(x) & \text{if } x \geq 1 \end{cases}$

Proof of Lemma A2:

$$\frac{1}{k} Log\left(\sum_{j=0}^{k} \frac{(xk)^j}{j!} \right) = \frac{1}{k} Log\left(\frac{e^{xk}}{k!} \int_{xk}^{\infty} t^k e^{-t} dt \right) = \frac{1}{k} Log\left(\frac{e^{xk}}{k!} \int_{x}^{\infty} (kt)^k e^{-kt} kdt \right)$$

$$= \frac{1}{k} Log\left(\frac{e^{xk} k^k}{k!} \int_{x}^{\infty} \left(te^{-t}\right)^k kdt \right)$$

For x>1, and large k, the term te^{-t} inside the integral is decreasing in t and thus approximately equal to its first order taylor expansion $e^{-x}(x + (1-x)(y-x)) = e^{-x}((1-x)y + x^2))$. Therefore, for x>1,

$$= \frac{1}{k} Log\left(\frac{e^{xk} k^k}{k!} \int_{x}^{\infty} \left(te^{-t}\right)^k kdt \right) \approx \frac{1}{k} Log\left(\frac{e^{xk} k^k}{k!} \int_{x}^{\frac{x^2}{1-x}} \left(e^{-x}((1-x)y + x^2)\right)^k kdt \right) \quad [46]$$

$$= \frac{1}{k} Log\left(\frac{k^k}{k!} \int_{x}^{\frac{x^2}{1-x}} \left((1-x)y + x^2\right)^k kdt \right) = \frac{1}{k} Log\left(-\frac{k^k}{k!} \frac{\left((1-x)x + x^2\right)^k}{(1-x)(k+1)} k \right)$$

$$= \frac{1}{k} Log\left(-\frac{k^k}{k!} \frac{\left((1-x)x + x^2\right)^k}{(1-x)(k+1)} k \right) = \frac{1}{k} Log\left(\frac{k^k}{k!} \frac{x^k}{(x-1)(k+1)} k \right)$$

$$\approx \frac{1}{k} Log\left(\frac{e^k x^k k}{\sqrt{2\pi k}(x-1)(k+1)} \right) \xrightarrow[k \to \infty]{} 1 + Log(x).$$

The last approximation invokes Stirling's Formula, $k! \approx \sqrt{2\pi k} k^k e^{-k}$

For x<1, the proof is similar, applying Stirling's formula and obtaining

$$\frac{1}{k} Log\left(\frac{e^{xk} k^k}{k!} \int_x^{\infty} \left(te^{-t} \right)^k kdt \right) = x + \frac{1}{k} Log\left(\frac{e^k}{\sqrt{2\pi k}} \int_x^{\infty} \left(te^{-t} \right)^k kdt \right) \qquad [47]$$

$$\approx x + \frac{1}{k} Log\left(\frac{1}{\sqrt{2\pi k}} \int_x^{\infty} \left(te^{1-t} \right)^k kdt \right) = x + \frac{1}{k} Log\left(\frac{k}{\sqrt{2\pi k}} \int_{1-½\sqrt{2}}^{1+½\sqrt{2}} \left(1 - 2(1-t)^2 \right)^k dt \right)$$

$$= x + \frac{1}{k} Log\left(\frac{\sqrt{k}\,\Gamma(k+1)}{2\Gamma(k+1+½)} \right) \to x \qquad \text{Q.E.D.}$$

Justification for some assertions in the text:

If the price is constant at $1/a + c$, the probability of a sale given that a customer has arrived is e^{-1-ac}. Thus, let $p_i(t)$ be the probability of making at least i sales in the period [t, T]. $p_0(t)=1$, and $p_i(T)=0$ for i>0.

$$p_i(t) = \int_t^T \lambda e^{-\lambda(s-t)} \left(e^{-1-ac} p_{i-1}(s) + \left(1 - e^{-1-ac}\right) p_i(s) \right) ds \qquad [48]$$

$$p_i'(t) = \lambda p_i(t) - \lambda\left(e^{-1-ac} p_{i-1}(t) + \left(1 - e^{-1-ac}\right) p_i(t) \right)$$

$$= \beta\left(p_i(t) - p_{i-1}(t) \right) = \beta\left(1 - p_{i-1}(t) - (1 - p_i(t)) \right)$$

This solves for

$$p_i(t) = 1 - e^{\beta(t-T)} \sum_{j=0}^{i-1} \frac{(\beta(T-t))^j}{j!} \qquad [49]$$

Thus, the probability that the flight sells out is

$$p_k(0) = 1 - e^{-\beta T} \sum_{j=0}^{k-1} \frac{(\beta T)^j}{j!} = 1 - e^{-\gamma k} \sum_{j=0}^{k-1} \frac{(\gamma k)^j}{j!} \to \begin{cases} 1 & if\ \gamma > 1 \\ ½ & if\ \gamma = 1 \\ 0 & if\ \gamma < 1 \end{cases} \qquad [50]$$

Proof of Lemma 3:
First note that if capacity k is zero, then sales are zero, verifying the base of an induction. Now suppose Lemma 3 is true for k-1.

$$\eta_k(t) = \int_t^T \mu e^{-\mu(s-t)}(1+\eta_{k-1}(s))ds$$

$$= \int_t^T \mu e^{-\mu(s-t)}\left(1+k-1-e^{-\mu(T-s)}\sum_{j=0}^{k-2}(k-1-j)\frac{(\mu(T-s))^j}{j!}\right)ds$$

$$= k\left(1-e^{-\mu(T-t)}\right)-\int_t^T \mu e^{-\mu(T-t)}\left(\sum_{j=0}^{k-2}(k-1-j)\frac{(\mu(T-s))^j}{j!}\right)ds$$

$$= k\left(1-e^{-\mu(T-t)}\right)-e^{-\mu(T-t)}\sum_{j=0}^{k-2}(k-1-j)\frac{(\mu(T-s))^{j+1}}{j+1!}$$

$$= k\left(1-e^{-\mu(T-t)}\right)-e^{-\mu(T-t)}\sum_{j=1}^{k-1}(k-j)\frac{(\mu(T-s))^j}{j!}$$

$$= k-e^{-\mu(T-t)}\sum_{j=0}^{k-1}(k-j)\frac{(\mu(T-s))^j}{j!}$$ Q.E.D.

Proof of Theorem 4: Let x be a nonnegative random variable, and

$q=\frac{\varepsilon+1}{\varepsilon}$, $p=\varepsilon+1$. p and q are conjugate exponents. Let $f=x^{\frac{-1}{p}}$ and $g=x^{\frac{1}{\varepsilon q}}$. Then from the Holder Inequality,

$$\left(E\{f^p\}\right)^{1/p}\left(E\{g^q\}\right)^{1/q} \geq E\{fg\}$$

or

$$\left(E\{x^{-1}\}\right)^{\frac{1}{\varepsilon+1}}\left(E\left\{x^{\frac{1}{\varepsilon}}\right\}\right)^{\frac{\varepsilon}{\varepsilon+1}} \geq E\left\{x^{\frac{-1}{\varepsilon+1}+\frac{1}{\varepsilon+1}}\right\}=1.$$

or

$$E\left\{\frac{\left(E\left\{x^{\frac{1}{\varepsilon}}\right\}\right)^{\varepsilon}}{x}\right\} \geq 1 \qquad \text{Q.E.D.}$$

Proof of Theorem 5: Let x be a nonnegative random variable, and

$q = \dfrac{2\varepsilon - 1}{\varepsilon - 1}$, $p = \dfrac{2\varepsilon - 1}{\varepsilon}$. p and q are conjugate exponents. Let $f = x^{\frac{1}{\varepsilon p}}$ and $g = x^{\frac{2}{q}}$. Then from the Holder Inequality,

$$\left(E\left\{f^{p}\right\}\right)^{1/p}\left(E\left\{g^{q}\right\}\right)^{1/q} \geq E\{fg\}$$

or

$$\left(E\left\{x^{\frac{1}{\varepsilon}}\right\}\right)^{\frac{\varepsilon}{2\varepsilon-1}}\left(E\left\{x^{2}\right\}\right)^{\frac{\varepsilon-1}{2\varepsilon-1}} \geq E\left\{x^{\frac{1}{2\varepsilon-1}+\frac{2(\varepsilon-1)}{2\varepsilon-1}}\right\} = E\{x\}$$

or

$$\left(E\left\{x^{\frac{1}{\varepsilon}}\right\}\right)^{\varepsilon}\left(E\left\{x^{2}\right\}\right)^{\varepsilon-1} \geq \left(E\{x\}\right)^{2\varepsilon-1}$$

$$\frac{\left(E\left\{x^{\frac{1}{\varepsilon}}\right\}\right)^{\varepsilon}}{E\{x\}} \geq \left(\frac{E\left\{x^{2}\right\}}{\left(E\{x\}\right)^{2}}\right)^{1-\varepsilon} = \left(1 + CV^{2}\right)^{1-\varepsilon} \qquad \text{Q.E.D.}$$

Proof of Theorem 6:

$$\pi = \underset{p_1, p_2}{Max}\ (p_1 - c_1)\left(1 - F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right)\right) + (p_2 - c_2)\left(F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right) - F\left(\frac{p_2}{\varphi_2}\right)\right)$$

$$= \underset{p_1, p_2}{Max}\ (p_1 - c_1 - (p_2 - c_2))\left(1 - F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right)\right) + (p_2 - c_2)\left(1 - F\left(\frac{p_2}{\varphi_2}\right)\right)$$

$$= \underset{p_1, p_2}{Max}\ (\varphi_1 - \varphi_2)(\frac{p_1 - p_2}{\varphi_1 - \varphi_2} - \frac{c_1 - c_2}{\varphi_1 - \varphi_2})\left(1 - F\left(\frac{p_1 - p_2}{\varphi_1 - \varphi_2}\right)\right) + \varphi_2\left(\frac{p_2}{\varphi_2} - \frac{c_2}{\varphi_2}\right)\left(1 - F\left(\frac{p_2}{\varphi_2}\right)\right)$$

$$= (\varphi_1 - \varphi_2)R\left(\frac{c_1 - c_2}{\varphi_1 - \varphi_2}\right) + \varphi_2 R\left(\frac{c_2}{\varphi_2}\right)$$

We assume the standard condition that marginal revenue is decreasing, in which case there are sales of good 2 if and only if

$$\frac{p_1}{\varphi_1} \geq \frac{p_2}{\varphi_2}$$

if and only if

$$\frac{p_1 - p_2}{\varphi_1 - \varphi_2} \geq \frac{p_2}{\varphi_2}$$

if and only if

$$\frac{c_1 - c_2}{\varphi_1 - \varphi_2} \geq \frac{c_2}{\varphi_2}$$

if and only if

$$\frac{c_1}{\varphi_1} \geq \frac{c_2}{\varphi_2}.$$