OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0030-364X | EISSN 1526-5463 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Models of the Spiral-Down Effect in Revenue Management

William L. Cooper
Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55455, billcoop@me.umn.edu

Tito Homem-de-Mello
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, tito@northwestern.edu

Anton J. Kleywegt
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, anton@isye.gatech.edu

The spiral-down effect occurs when incorrect assumptions about customer behavior cause high-fare ticket sales, protection levels, and revenues to systematically decrease over time. If an airline decides how many seats to protect for sale at a high fare based on past high-fare sales, while neglecting to account for the fact that availability of low-fare tickets will reduce high-fare sales, then high-fare sales will decrease, resulting in lower future estimates of high-fare demand. This subsequently yields lower protection levels for high-fare tickets, greater availability of low-fare tickets, and even lower high-fare ticket sales. The pattern continues, resulting in a so-called spiral down. We develop a mathematical framework to analyze the process by which airlines forecast demand and optimize booking controls over a sequence of flights. Within the framework, we give conditions under which spiral down occurs.

Subject classifications: Pricing: revenue management. Forecasting: estimation and control. Probability: applications
Area of review: Manufacturing, Service, and Supply Chain Operations
History:

1. Introduction

Revenue management involves the application of quantitative techniques to improve profits by controlling the prices and availabilities of various products that are produced with scarce resources. The best known revenue management application occurs in the airline industry, where the products are tickets (for itineraries) and the resources are seats on flights. Over the past decade both practitioners and academics have helped to develop a considerable and rapidly growing literature on revenue management. Much of this work is reviewed in Talluri and van Ryzin (2004b), Bitran and Caldentey (2003), and Boyd and Bilegan (2003).

In almost every instance of published work, the starting point of the analysis is some set of assumptions regarding an underlying stochastic or deterministic demand process. With these assumptions in hand (and assumed to be correct), most papers proceed to analyze the model and derive policies that are good or optimal for the formulated model. In the airline context, such a policy usually prescribes which types of tickets should be available at which times, and under which circumstances.

However, the situation faced by revenue managers in practice is different in at least two key regards: assumptions may be incorrect, and model parameters are not known. There are a variety of reasons why a revenue manager may use a model with incorrect assumptions. Among these are (a) availability of intuitively-pleasing decision rules — such as the Littlewood rule considered herein, (b) simplification for analytical tractability, (c) availability of forecasting and optimization software, and (d) lack of understanding of the problem. Moreover, a revenue manager may be aware of a modeling error, but may not fully comprehend its consequences. We are specifically interested in the consequences of using incorrect models, especially if the parameters of such models are estimated with available data. Even if the data are good (say correctly untruncated demand data) and a good forecasting method is used, the problem remains that parameters are being estimated for an inappropriate model, and consequently there often do not exist parameter values that will make the revenue manager's model correct.

In revenue management practice, there is a process whereby controls (e.g., protection levels) are enacted, sales occur, flights depart, new data are observed, and parameter estimates are updated. The updated estimates are then used to choose new controls for the next set of flights, and so on. An important question is what can happen in such a forecasting and optimization process if the revenue manager uses a good forecasting method, but the chosen controls are based on erroneous assumptions.

As an example, suppose that there are two classes of tickets and that customers are flexible, that is, they are willing to buy either low-fare or high-fare tickets, but they will buy the low-fare tickets if both are available. Suppose also that the airline chooses how many seats to reserve for high-fare tickets (i.e., the protection level) based on past sales of high-fare tickets, while neglecting to account for the fact that availability of low-fare tickets will reduce sales for high-fare tickets. Then, if more low-fare tickets are made available, low-fare sales will increase and high-fare sales will decrease, resulting in lower future estimates of high-fare demand, and subsequently lower protection levels for high-fare tickets and greater availability of low-fare tickets. The pattern continues, resulting in a downward spiral of high-fare sales, protection levels, and revenues. It is of concern that the flawed model produces suboptimal controls (which is no surprise), but of even greater concern is the phenomenon that the controls can become systematically worse as the forecasting and optimization process continues. Boyd et al. (2001) have used simulation to demonstrate this spiral-down effect, which is known to some practitioners. However, to our knowledge, this phenomenon has not been studied in the literature, although Kuhlmann (2004) alluded to the underlying issue with his remark that "although airlines had spent considerable sums making forecasting, allocation, and other elements of revenue management more precise, they failed to deal with some of the inherent flawed assumptions of revenue management. For instance, if a carrier sold 50 B-class passengers on any given day, that was then established as the historical demand for B class, ignoring the fact that the absence of availability of other classes might have skewed the result."

In this paper we introduce a generic framework for the study of iterative data collection-forecasting-optimization processes. We begin by formalizing what it means for a forecasting method to be good, even if the method is used to estimate parameters for an incorrect model. Working within our framework, we study a process in which a revenue manager sets protection levels for a sequence of flights using the Littlewood rule, with its inputs estimated by various good forecasting methods applied to observed data. Underlying his use of the Littlewood rule is the revenue manager's reliance on a model in which one of these inputs, "the probability distribution of high-fare demand," is assumed to be exogenously determined, i.e., unaffected by the chosen protection levels. However, the observed historical data *do* depend on the past values of the protection levels, and this dependence is not captured properly by the revenue manager's model. That is, the revenue manager makes a *modeling error*.

We analyze the dynamic behavior of the Littlewood rule because it is widely used in practice and forms the basis of much revenue management software. The rule also allows for relatively tractable dynamics within our general framework. In addition, the observation that models founded on incorrect assumptions can lead to a systematic deterioration of performance remains relevant in a broader context, especially in light of the large number of models that do not accurately describe consumer behavior.

Our main results show that in many cases the protection levels converge to a value, in some cases to zero. The limit of the sequence of protection levels is a fixed point of a certain function. As a result, the data observed by the revenue manager seem consistent with his incorrect model, possibly reinforcing his belief in the model. Such a limit point is often suboptimal (in terms of the corresponding expected revenue), and in many cases it is much worse than the suboptimal decision that would have resulted if the revenue manager had not sought to improve the parameter estimates of the incorrect model with the observed data. This indicates that if one starts with a flawed model, then attempts to refine parameter estimates may be counterproductive.

Some additional insights are obtained from our analysis. The first one is that, contrary to what may be expected, the spiral-down phenomenon is not a consequence of data truncation. As we show, the phenomenon may occur even if all customers can be observed after all tickets have been sold. We also show that the problem is not forecast variability or the quality of the forecast method. These findings emphasize the variety of situations in which spiral down can occur, and demonstrate that the real issue is the modeling error. Another insight is that the relation between the distribution of "flexible" customers — i.e., customers who are willing to buy a high-fare ticket but prefer a low-fare one — and the ratio of the fares appears to be crucial to determine whether the protection levels spiral down or up or neither.

Processes that involve both estimation and control have been studied in various contexts. A large part of the literature on stochastic control addresses simultaneous parameter estimation and control. It is well known that the so-called parameter identifiability problem can lead to convergence of parameter estimates to incorrect values that are consistent with the observed data; for example, see Kumar and Varaiya (1986) and Bertsekas (2000). Van Ryzin and McGill (2000) model the process whereby an airline chooses revenue management controls for a sequence of flights with unknown model parameters. Similar to most published revenue management work, their model assumes that there is an exogenous demand for tickets of different fare classes, that is, they do not consider the possibility of misspecification in the revenue manager's model. In Section 5.3, we consider the consequences of modeling error in a similar setting. There are also publications that propose inventory control mechanisms for problems with unknown or partially-specified demand distributions. The basic focus is similar to that of van Ryzin and McGill (2000) insomuch as techniques are proposed to solve classes of problems, but the consequences of incorrect modeling assumptions are not investigated. For examples and references, see Burnetas and Smith (2000) and Carvalho and Puterman (2003).

Some of the literature on game theory and bounded rationality studies games in which the players learn about problem parameters and/or the actions of other players over multiple stages of the game; for example, see Fudenberg and Levine (1998). Although Kreps (1990), p.155, mentions the possibility of model misspecification by players in games, very little work in the literature incorporates such modeling error, that is, studies the consequences of parameter estimation and control with misspecified models for which there do not exist parameter values that will make the models correct.

Research that does consider effects of modeling error includes that of Cachon and Kok (2003), who study the issue in the context of a newsvendor problem. In their setting, the newsvendor uses an incorrect optimization model, namely the basic newsvendor model, with an input parameter called "the salvage value", to choose the initial inventory, and the newsvendor attempts to estimate the value of this input parameter with observed data. They study the newsvendor's estimates and chosen controls in the context of a model (the correct model) in which a clearance price is determined at the end of the primary selling season as a function of the remaining inventory at the end of the primary selling season. Other related work includes that of Balakrishnan et al. (2004), who study a deterministic inventory system where the demand rate depends on the inventory level. They consider a situation where orders are placed according to the standard EOQ formula with the demand rate estimated from data in a way that does not properly account for the dependence between demand rate and inventory level. Bertsekas and Tsitsiklis (1996) study the approximation of dynamic programming value functions with parameterized functions. In many cases, there are no parameter values that will make the approximate function equal to the value function. In their Section 8.3 they give an example of a controller for the game of tetris based on such a misspecified approximation, in which the performance of the controller deteriorates as it attempts to improve the estimates of the parameter values and the policies based on these parameter estimates.

This paper is organized as follows. Section 2 describes the framework that we use to study the revenue management forecasting and optimization process, including the data observed by the revenue manager, the quantities forecasted by the revenue manager, the method used by the revenue manager to choose controls, and how these aspects fit together in the process dynamics. Section 3 provides a simple deterministic example of spiral-down behavior. Section 4 defines what is regarded as a good forecasting method, and provides examples. Section 5 analyzes spiral-down behavior under three forecasting methods. Section 6 discusses how the results in previous sections can be extended. Section 7 establishes general results that relate the long-run behavior of forecasts and protection levels, giving additional conditions for spiral down. Section 8 briefly discusses what to do about the spiral-down effect, and resulting research questions. All proofs are given either in the Appendix at the end of the paper or in the Online Appendix.

2. The Framework

Consider a single flight with c seats, and suppose that there are class-1 and class-2 tickets for sale. The price of a class-i ticket is f_i, where $f_1 > f_2 > 0$. Suppose also that there is a revenue manager, whose job it is to control availability of the tickets to maximize the airline's expected revenue. Below, we describe a setup in which (a) the revenue manager determines booking policies using a model that is widely used in the airline industry that assumes that there is an exogenous random demand for each class of ticket, and (b) customers decide what to purchase based on their own preferences as well as the available alternatives, and hence there is actually no such thing as exogenous demand for each class of ticket.

2.1. Revenue Manager's Choice of Booking Control

Suppose that the revenue manager uses the well-known Littlewood rule (see, e.g., Littlewood 1972, Belobaba 1989, Wollmer 1992, Brumelle and McGill 1993, or van Ryzin and McGill 2000) to control the availability of class-1 and class-2 tickets. Specifically, the revenue manager chooses the protection level for class-1 tickets, and employs a model that takes as input the cumulative probability distribution H of the assumed exogenous demand for class-1 tickets. Given H, the revenue manager chooses a protection level ℓ that satisfies

$$\ell \in H^{-1}(\gamma) \tag{1}$$

where $\gamma := 1 - f_2/f_1$ and $H^{-1}(\gamma)$ denotes the set of γ-*quantiles* of H. That is, ℓ is chosen to satisfy

$$H(\ell) \geq \gamma \quad \text{and} \quad H(\ell-) \leq \gamma \tag{2}$$

where $H(\ell-) := \lim_{x\uparrow\ell} H(x)$ denotes the left limit of H at ℓ. For continuous demand distributions, condition (1) states that the protection level is chosen to satisfy

$$f_1 \times \text{Prob}\,[\text{Exogenous demand for class-1 tickets} \geq \ell] \;=\; f_2.$$

Similar interpretations are possible for integer-valued demand. The protection level $\ell \geq 0$ is used to control bookings as follows: the cheaper class-2 tickets are available as long as more than ℓ seats are available, that is, as long as fewer than $c - \ell$ tickets in total have been sold. Throughout the paper, we consider a setting in which the protection level for a flight is set just once during the time period over which bookings for the flight take place.

2.2. Revenue Manager's Observed Data

Once the revenue manager has decided to use the Littlewood rule, the assumed distribution H has to be estimated based on available data. In practice, these data typically include historical values of class-1 tickets sales, possibly after some so-called unconstraining to remove effects caused by censoring and/or truncation (see Section 4.2 of Boyd and Bilegan 2003). In our setting, these data consist of past values of what we call the *observed quantity* X, which may be censored or truncated or unconstrained data, and which the revenue manager believes to be observations of the exogenous class-1 demand.

By virtue of the assumption that the revenue manager is using the Littlewood rule (1), only the high-fare "demand distribution" has to be estimated. Upon using protection level ℓ to control the booking process, the revenue manager obtains a new value X of the observed quantity. Let $G(\ell,\cdot)$ denote the cumulative distribution function of X if the booking process is controlled with protection level ℓ. Note that the distribution of X depends on ℓ, whereas H does not, and the revenue manager's model does not contain a construct such as G. Later, we discuss various ways in which the revenue manager can use values of the observed quantity to estimate H.

We illustrate the above ideas with some examples that demonstrate how the distribution of X may depend on ℓ.

Example 2.1. The first example is the model that usually is associated with the Littlewood rule. There are only two types of customers, namely type-a and type-b customers. Type-a customers want only class-1 tickets and type-b customers want only class-2 tickets, and all type-b customers arrive before any type-a customers arrive. Let D_a denote the number of type-a customers that arrive, and let the observed quantity X be equal to D_a. Thus, here the revenue manager observes all the type-a customers who arrive, even customers who arrive after all c tickets have been sold. In addition, it is assumed that D_a does not depend on ℓ. Hence, $G(\ell,x) := \text{Prob}[X \le x] = \text{Prob}[D_a \le x]$, which is independent of ℓ. Under some additional independence assumptions, using (1) to choose the protection level is optimal in this example.

Example 2.2. In this example there are type-a, type-b, and type-ab customers, and no specific assumptions on the order of arrivals. Type-a customers buy class-1 tickets only, and type-b customers buy class-2 tickets only. Type-ab customers buy either class-1 or class-2 tickets. If class-2 tickets are available, then an arriving type-ab customer will purchase a class-2 ticket. If only class-1 tickets are available, then an arriving type-ab customer will purchase a class-1 ticket. The three types of customers arrive according to a marked point process that describes customer arrival times and customer types over the time interval between when tickets first become available and when the flight departs. The point process itself is independent of the chosen protection level ℓ.

We consider cases in which the observed quantity X is class-1 sales ("truncated class-1 demand") and "untruncated class-1 demand" separately. Let D_a and D_{ab} denote the number of type-a and type-ab customers respectively who arrive during the time interval. Let $D_a(\ell)$ and $D_{ab}(\ell)$ denote the number of type-a and type-ab customers respectively who arrive until $c-\ell$ tickets have been sold (if $\ell \ge c$, then $D_a(\ell) = D_{ab}(\ell) = 0$; if the total demand is less than $c-\ell$, then $D_a(\ell) = D_a$, $D_{ab}(\ell) = D_{ab}$). Note that $D_a(\ell)$ and $D_{ab}(\ell)$ both depend on the arrival processes of all three types of customers.

Case 2.2.A: Untruncated class-1 demand. In this case, X is equal to the number of type-a customers who arrive plus the number of type-ab customers who arrive when class-2 tickets are no longer available (that is, the number of type-ab customers who either purchase class-1 tickets or who arrive when no tickets are available). That is, $X = D_a + D_{ab} - D_{ab}(\ell)$, and therefore, $G(\ell,x) = \text{Prob}[D_a + D_{ab} - D_{ab}(\ell) \le x]$, which depends on ℓ. Note that in this example the revenue manager continues to observe customers even after c tickets have been sold.

Recall that the observed quantity X is what the revenue manager thinks is an observation of the supposed exogenous "class-1 demand," and thus it makes sense to the revenue manager to estimate the supposed distribution H using observed values of X. Note also that in this case we are eliminating the possibility of worry about truncated data by allowing the revenue manager to observe all arriving customers, even after all c tickets have been sold. In addition, we are giving the revenue manager the "benefit of the doubt" by including type-ab customers who are turned away after all tickets have been sold in the observed quantity X.

Case 2.2.B: Truncated class-1 demand. In this case, the observed quantity X is the number of class-1 tickets that are sold; that is, $X = D_a(\ell) + \min\{D_a - D_a(\ell) + D_{ab} - D_{ab}(\ell), c, \ell\}$, and therefore, $G(\ell, x) = \text{Prob}\,[D_a(\ell) + \min\{D_a - D_a(\ell) + D_{ab} - D_{ab}(\ell), c, \ell\} \leq x]$, which also depends on ℓ. In this example the revenue manager does not continue to observe customers after c tickets have been sold.

We use the above cases in some of the examples that we consider later in this paper. In other cases we do not describe the details of how the distribution of X arises from the interaction of customers' behavior and the choice of ℓ, but rather we directly work with $G(\ell, \cdot)$; see, e.g., Sections 5.1 and 5.3. We emphasize that the general results developed in this paper do not hinge on the particular examples of X above. Rather, we require only that the following general setup prevails.

2.3. Dynamics of the Forecasting and Optimization Process

We consider a sequence indexed $k = 1, 2, 3, \ldots$, of a particular type of flight, for example, an 8 AM Monday flight from New York to Los Angeles. The revenue manager selects a protection level L^0 for flight 1, and subsequently sees observed quantity X^1. The distribution of the observed quantity X^1 for flight 1 is $G(L^0, \cdot)$. Based on what is observed, the revenue manager selects a new protection level L^1 for flight 2. The distribution of the observed quantity X^2 for flight 2 is $G(L^1, \cdot)$, and given L^1, X^2 is conditionally independent of the past. The revenue manager continues in this fashion, yielding sequences $\{L^k\}$ and $\{X^k\}$, where L^{k-1} and X^k denote the protection level and observed quantity respectively for flight k. After each flight k, the revenue manager uses the data $X^1, \ldots, X^k$ that have been observed so far to construct an estimate $\hat{H}^k$ of the probability distribution of the assumed exogenous demand for class-1 tickets. Then, the revenue manager chooses the protection level $L^k \in (\hat{H}^k)^{-1}(\gamma)$ for flight $k+1$ using the Littlewood rule (1).

To precisely describe the iterative forecasting and booking control process, we introduce some more notation. Let $\mathcal{P}(\mathbb{R})$ denote the space of probability distribution functions on $\mathbb{R}$. For each $k \in \mathbb{N}$, let $\phi^k : \mathcal{P}(\mathbb{R}) \times \mathbb{R}^k \mapsto \mathcal{P}(\mathbb{R})$ denote a generic update function that maps the initial estimate $\hat{H}^0 \in \mathcal{P}(\mathbb{R})$, and the data $(X^1, \ldots, X^k) \in \mathbb{R}^k$ observed so far, to a new estimate $\hat{H}^k$. The process evolves on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}^k\}$. Expectation with respect to $\mathbb{P}$ is denoted by $\mathbb{E}$. The initial estimate $\hat{H}^0$ for class-1 demand is specified and a protection level $L^0 \in (\hat{H}^0)^{-1}(\gamma)$ is chosen. The sequence $\{(X^k, \hat{H}^k, L^k) : k \in \mathbb{N}\}$ is adapted to $\{\mathcal{F}^k\}$. We assume that with probability 1 (w.p.1), for each $k \in \mathbb{N}$,

$$\mathbb{P}\left[X^{k+1} \leq x \,|\, \mathcal{F}^k\right] \;=\; G(L^k, x) \quad \text{for all } x \in \mathbb{R}, \tag{3}$$

that is, the conditional distribution of X^{k+1}, given the history of the process up to flight k, depends only on L^k. Forecasts and protection levels are updated according to

$$\hat{H}^k := \phi^k(\hat{H}^0, X^1, \ldots, X^k) \tag{4}$$

$$L^k \in (\hat{H}^k)^{-1}(\gamma) \tag{5}$$

for each $k \in \mathbb{N}$. The revenue manager's chosen forecasting method determines each ϕ^k. We will mostly be interested in forecasting methods that are good in a certain sense. Loosely speaking, "good" will mean that if the distributions $G(L^k,\cdot)$ settle down to a limit as k gets large, then the forecasts $\hat{H}^k$ will approach the same limit. Before delving into these details, we discuss an example in the next section.

3. A Deterministic Example

In order to motivate some of the issues concerning the spiral-down phenomenon within the framework described above, we discuss a simple deterministic example. Consider the setting described in Example 2.2.A. Suppose that there are d customers who want a ticket, where $d \ge 0$ is a fixed constant, and that all customers are of type ab. That is, we have $D_a = 0$, $D_{ab} = d$, and $D_{ab}(\ell) = \min\{d, (c-\ell)^+\}$. Thus, if the protection level is ℓ, then the observed quantity is deterministically equal to $X = d - \min\{d, (c-\ell)^+\} = [d-(c-\ell)^+]^+$, and hence

$$G(\ell,x) \quad = \quad \begin{cases} 1 & \text{if } x \ \ge \ [d-(c-\ell)^+]^+ \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

It is easy to see that the total class-2 sales is equal to $\min\{d,(c-\ell)^+\} = d - X$, and the total class-1 sales is equal to $\min\{c,d\} -$ (class-2 sales). Because $f_1 > f_2$, the best thing to do is to set the protection level at c or higher so that the number of high-fare tickets sold will be $\min\{c,d\}$ and the number of low-fare tickets sold will be 0. The worst thing to do is to set the protection level at $\max\{0, c-d\}$ or lower so that the number of low-fare tickets sold will be $\min\{c,d\}$ and the number of high-fare tickets sold will be 0.

Suppose that we start with an arbitrary $L^0 \ge 0$, and that for each $k \in \mathbb{N}$, the forecast $\hat{H}^k$ is the empirical distribution function of the observed quantity X; i.e.,

$$\hat{H}^k(x) \quad := \quad \frac{1}{k}\sum_{j=1}^{k} \mathbb{I}_{\{X^j \le x\}}. \tag{7}$$

We discuss three cases separately: (i) $d = c$, (ii) $d < c$, and (iii) $d > c$.

Case (i): $d = c$: In this case, for any $\ell \in [0,c]$, $X = \ell$, and thus $G(\ell,x) = \mathbb{I}_{\{\ell \le x\}}$. Hence, for any $L^0 \in [0,c]$, $X^1 = L^0$ and $\hat{H}^1(x) = \mathbb{I}_{\{L^0 \le x\}} = G(L^0,x)$, and thus it follows that $L^k = L^0$ and $\hat{H}^k = \hat{H}^1$ for all $k \ge 1$. If $L^0 > c$, then $X^1 = d = c$, and $\hat{H}^1(x) = \mathbb{I}_{\{d \le x\}}$. It follows that $L^k = d$ and $\hat{H}^k(x) = \mathbb{I}_{\{d \le x\}}$ for all $k \ge 1$.

Case (ii): $d < c$: In this case, $X^1 = [d-(c-L^0)^+]^+$, and $\hat{H}^1(x) = \mathbb{I}_{\{[d-(c-L^0)^+]^+ \le x\}}$. Since $[d-(c-L^0)^+]^+ \le [d-(c-L^0)]^+ \le L^0$, it follows that $L^1 = (\hat{H}^1)^{-1}(\gamma) = [d-(c-L^0)^+]^+ \le L^0$. Furthermore, we have a strict inequality (i.e., $L^1 < L^0$) unless $L^1 = L^0 = 0$. In general, we have the following result.

PROPOSITION 1. *Suppose that the probability distribution of the observed quantity is given by* (6) *with $d < c$, and that forecasts are made according to* (7)*. Then $L^{k+1} \le L^k$ for all k. Furthermore, there exists a k^* such that $L^j = 0$ and $X^j = 0$ for all $j \ge k^*$.*

Observe that Proposition 1 gives a situation in which the protection levels spiral down to the worst possible value. Also note that $0 = G^{-1}(0,\gamma)$.

It is also interesting that the revenue manager's estimates $\hat{H}^k$ converge to the point mass at zero, which is indeed consistent with what the revenue manager observes — namely, that $X^j = 0$ for all j large enough. Hence, (a) the use of an incorrect model, and (b) the application of a forecasting method that is "good" insomuch as it agrees with the observations, together combine to produce the worst possible protection levels. Also, observe that the cause of the problem is not censored or truncated sales data, because all customers are observed by the revenue manager.

Table 1 shows the spiral-down effect described by Proposition 1. The values of c, d, f_1, f_2 and L^0 were chosen to have an example that can be followed in a step-by-step manner, using only manual calculations. Note that the initial protection level is optimal. In spite of that, the revenue manager's incorrect assumptions lead the protection levels to settle on the worst possible value.

Table 1 Spiral down with $c = 10$, $d = 8$, $f_1 = 500$, $f_2 = 200$, and $L^0 = 10$.

k	Protection Level L^{k-1}	Obs. Qty. X^k	Class-1 Sales	Class-2 Sales	Revenue
1	10	8	8	0	\$4000
2	8	6	6	2	\$3400
3	8	6	6	2	\$3400
4	6	4	4	4	\$2800
⋮	⋮	⋮	⋮	⋮	⋮
8	6	4	4	4	\$2800
9	4	2	2	6	\$2200
⋮	⋮	⋮	⋮	⋮	⋮
20	4	2	2	6	\$2200
21	2	0	0	8	\$1600
⋮	⋮	⋮	⋮	⋮	⋮
50	2	0	0	8	\$1600
51	0	0	0	8	\$1600

Case (iii): $d > c$: If $L^0 > c$, then $X^k = d$, $\hat{H}^k(x) = \mathbb{I}_{\{d \le x\}}$, and $L^k = d$ for all $k \ge 1$. If $L^0 \in [0, c]$, then the behavior of L^k is described in the proposition below.

PROPOSITION 2. *Suppose that the probability distribution of the observed quantity is given by* (6) *with $d > c$, and that forecasts are made according to* (7). *Suppose that $L^0 \in [0, c]$. Then $L^{k+1} \ge L^k$ for all k. Furthermore, there exists a k° such that $L^j = d$ and $X^j = d$ for all $j \ge k^\circ$.*

Proposition 2 illustrates a situation in which protection levels drift upward to the best value, even though the revenue manager is using a model based on the wrong assumptions. Observe also that we have $d = G^{-1}(d, \gamma)$.

Although in this example spiral down occurs only when there is spare capacity (i.e. $d < c$), in general spare capacity is not necessary for spiral down to occur, as illustrated by the results in Sections 5, 6, and 7.

In each of the cases described above, all the protection levels are eventually equal to a fixed point ℓ^* of $G^{-1}(\cdot, \gamma)$, that is, for all k large enough it holds that $L^k = G^{-1}(L^k, \gamma)$. Hence, when the revenue manager's forecast $\hat{H}^k$ is such that $\ell^* = (\hat{H}^k)^{-1}(\gamma)$, then $L^k = \ell^*$, and the next observed quantity X^{k+1} is also equal to ℓ^*, which to the revenue manager appears to be consistent with the forecast $\hat{H}^k$. This causes forecast $\hat{H}^{k+1}$ not to differ much from $\hat{H}^k$, and $L^{k+1} = L^k = \ell^*$. In Section 7 we discuss this characteristic in greater generality.

4. Good Forecasting Methods

The behavior of the forecast and optimization process depends on the forecasting method being used. We are particularly interested in a certain class of forecasting methods, which we call "good" ones. Our definition of good will formalize the notion that the estimates $\{\hat{H}^k\}$ are consistent with the distribution of the quantities $\{X^k\}$ over the long run (see the discussion after Proposition 1).

In order to specify precisely what is meant by a good forecasting method, we need some definitions. Let $\xrightarrow{w}$ denote weak convergence; recall that a sequence of distribution functions $\{F^k\}$ on $\mathbb{R}$ converges weakly to distribution function F on $\mathbb{R}$ (written $F^k \xrightarrow{w} F$) if $F^k(x) \to F(x)$ for all x at which F is continuous. If we want to emphasize that a distribution function (say F) depends on the sample path $\omega \in \Omega$ (i.e., depends on the evolution of the iterative process), then we write $F(\omega, \cdot)$. In this case, we say that F is a random distribution function. More formally, by a random distribution function F we mean a random element $F : \Omega \mapsto \mathcal{P}(\mathbb{R})$, measurable with respect to $\mathcal{F}$ and the Borel σ-algebra $\mathcal{B}$ corresponding to the topology of weak convergence on $\mathcal{P}(\mathbb{R})$. Proposition 17 in the Appendix establishes that, for each $x \in \mathbb{R}$, $F(\omega, x)$ is a well-defined random variable. It also establishes that sets of the form $\{\omega \in \Omega : F^k(\omega,\cdot) \text{ converges weakly as } k \to \infty\}$ and $\left\{\omega \in \Omega : F^k(\omega,\cdot) \xrightarrow{w} F(\omega,\cdot)\right\}$ are in $\mathcal{F}$.

DEFINITION 1. Consider a sequence $\{\hat{H}^k\}$ of random distribution functions and a sequence $\{Y^k\}$ of real-valued random variables, both defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to filtration $\{\mathcal{F}^k\}$. Let $F^k : \Omega \mapsto \mathcal{P}(\mathbb{R})$ be given by $F^k(\omega, x) := \mathbb{P}[Y^{k+1} \le x \,|\, \mathcal{F}^k]$, that is, F^k is the conditional distribution of Y^{k+1}. Let $\Omega^* := \{\omega \in \Omega : F^k(\omega,\cdot) \text{ converges weakly as } k \to \infty\}$, and for all $\omega \in \Omega^*$, let $F^*(\omega,\cdot)$ denote the weak limit of $\{F^k(\omega,\cdot)\}$.

We say that $\{\hat{H}^k\}$ is a *good* forecasting method for $\{Y^k\}$ if there exists a set $\Omega' \subset \Omega$ such that $\mathbb{P}[\Omega'] = 0$, and

$$\hat{H}^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot) \tag{8}$$

for all $\omega \in \Omega^* \setminus \Omega'$.

The definition calls $\{\hat{H}^k\}$ a good forecasting method for $\{Y^k\}$ if for $\mathbb{P}$-almost all sample paths $\omega \in \Omega$ such that the sequence of conditional distribution functions of $\{Y^k\}$ converges weakly to some distribution function $F^*(\omega,\cdot)$, the forecast distributions $\{\hat{H}^k(\omega,\cdot)\}$ converge weakly to the same $F^*(\omega,\cdot)$. In particular, if $\{Y^k\}$ is an i.i.d. sequence with common distribution F^*, then $\{\hat{H}^k\}$ is called a good forecasting method for $\{Y^k\}$ if $\{\hat{H}^k\}$ converges weakly to F^* w.p.1 — a natural requirement for a reliable forecasting method.

Note that the conditions in the definition do not require the sequences $\{F^k(\omega,\cdot)\}$ to have limits, and the limits, if they exist, do not have to be the same for all ω. Also note that Definition 1 allows for any behavior of the forecasts $\{\hat{H}^k\}$ on the set of $\omega \in \Omega$ on which the sequence $\{F^k\}$ does not converge weakly. Finally, note that (8) is equivalent to the following:

$$\mathbb{P}\left[\hat{H}^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot),\ F^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot)\right] \;=\; \mathbb{P}\left[F^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot)\right]; \tag{9}$$

i.e., $\mathbb{P}[\hat{H}^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot) \,|\, F^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot)] = 1$ whenever $\mathbb{P}[F^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot)] > 0$.

The idea of defining a "reliable" forecast as one that approaches the true distribution is not new. Blackwell and Dubins (1962) introduced the concept of *merging* as a way to formulate that property mathematically, and Dawid (1982) proposed the idea of *calibration*, which means that the observed empirical distributions converge to the forecasted ones. Kalai, Lehrer, and Smorodinsky (1999) expanded on those ideas and not only proposed alternative definitions for merging but also showed the equivalence between appropriately defined concepts of merging and calibration. A general definition of merging is the following. Let $\hat{\mu}(\cdot|\mathcal{F}^k)$ and $\mu(\cdot|\mathcal{F}^k)$ be respectively the forecasted and true distributions given the history of a process $\{Y^k\}$ up to step k. Suppose that Y^k takes values in state space S, and let $\mathcal{A}$ be a collection of subsets of a set such as S or S^∞ (the choice of set and collection $\mathcal{A}$ is different for different notions of merging). Suppose that

$$\sup_{A \in \mathcal{A}} \left|\hat{\mu}\left(A \,|\, \mathcal{F}^k\right) - \mu\left(A \,|\, \mathcal{F}^k\right)\right| \;\to\; 0 \quad \text{w.p.1} \tag{10}$$

as $k \to \infty$. The collection $\mathcal{A}$ determines how strong condition (10) is. For example, in Blackwell and Dubins (1962), $\mathcal{A}$ is the σ-algebra on S^∞ generated by all histories of the process $\{Y^k\}$, and Kalai et al. (1999) say that $\hat{\mu}$ *strongly merges* to μ if (10) holds for such a choice of $\mathcal{A}$. Thus, for $\hat{\mu}$ to strongly merge to μ, convergence of $\hat{\mu}$ to μ is required not only for the 1-step forecasts but also for the n-step forecasts for all n. Also, if $\mathcal{A}$ is the σ-algebra on S representing the collection of events in one step, and (10) holds for such a choice of $\mathcal{A}$ (so that convergence is required only for the 1-step forecast), then Kalai et al. (1999) say that $\hat{\mu}$ *merges* to μ.

We can relate the above definitions to our definition by choosing the probability measures $\hat{\mu}$ and μ to be the measures induced by the involved random variables. That is, we have

$$\hat{\mu}\left[(-\infty,x]\,|\,\mathcal{F}^k\right] := \hat{H}^k(\omega,x), \qquad \mu[(-\infty,x]\,|\,\mathcal{F}^k] := \mathbb{P}[Y^{k+1}\le x\,|\,\mathcal{F}^k] = F^k(\omega,x).$$

We see from (10) that merging, in the terminology of Kalai et al. (1999), implies that

$$\sup_{x\in\mathbb{R}}\left|\hat{H}^k(\omega,x)-F^k(\omega,x)\right| \to 0 \quad \text{w.p.1} \tag{11}$$

as $k\to\infty$. Condition (11) implies that $\hat{H}^k(\omega,\cdot)\xrightarrow{w}F^*(\omega,\cdot)$ almost everywhere on $\{\omega : F^k(\omega,\cdot)\xrightarrow{w}F^*(\omega,\cdot)\}$, which is condition (8) in Definition 1. Hence, if $\{\hat{H}^k\}$ merges to $\{F^k\}$, then $\{\hat{H}^k\}$ is a good forecasting method for $\{Y^k\}$, i.e., merging implies goodness of the forecasting method. Moreover, Definition 1 applies to processes with state space $\mathbb{R}$, whereas Kalai et al. (1999) assume that the state space is finite. The two definitions coincide when the state space is finite and $\{F^k\}$ converges weakly w.p.1.

An interesting question that arises from the above conclusion concerns the relationship between good forecasts and the notion of calibration referred to earlier. In Kalai et al. (1999), a forecast is said to be calibrated if it passes a certain set of *checking rules*; based on that, they proceed to prove that (strong) merging is equivalent to (strong) calibration. It is natural to ask whether Definition 1 corresponds to some weaker notion of calibration. The fact that merging in the terminology of Kalai et al. (1999) implies goodness of the forecasting method suggests that being a good forecasting method may be equivalent to passing a smaller set of checking rules than the set required for the standard notion of calibration. A more thorough examination of these ideas falls outside the scope of this paper.

Also note that if the conditional distributions converge weakly to F^* w.p.1, then by Proposition 3 below, the empirical distributions also converge weakly to F^* w.p.1, and hence the values of $\{Y^k\}$ will be "in agreement" with their limiting conditional distribution. In such a case, if a good forecasting method is used, then the forecaster will perceive the values of $\{Y^k\}$ to be in agreement with the forecasts, which again suggests that good forecasts are, in a sense, calibrated with F^*.

In subsequent sections we will apply the results of this section with $\{Y^k\}=\{X^k\}$. The results of the present section do not depend on this particular choice; they are valid independent of the setup of the rest of the paper. Note also that the definition of a good forecasting method is not affected by whether or not the sequence $\{F^k\}$ is dependent on $\{\hat{H}^k\}$. In the case where $\{Y^k\}=\{X^k\}$, F^k indeed depends on $\hat{H}^k$, which may cause the revenues to spiral down. Hence, a good forecasting method can be an ingredient of a poorly behaving revenue management process.

Next we discuss some examples of forecasting methods and check whether they satisfy Definition 1.

Empirical Distribution Function. The empirical distribution function is defined as

$$\hat{H}^k(x) := \frac{1}{k}\sum_{j=1}^{k}\mathbb{I}_{\{Y^j\le x\}} \tag{12}$$

One can use $\hat{H}^k$ to forecast the distribution of Y^{k+1}. When the sequence $\{Y^k\}$ is i.i.d. with common distribution F, then a stronger property than being good holds. Namely, for $\mathbb{P}$-almost all ω, $\hat{H}^k$ converges to F uniformly in x by the Glivenko-Cantelli Theorem — see Theorem 5.5.1 of Chung (1974). For the general (i.e., possibly non-i.i.d.) case, the next proposition shows that the empirical distribution provides a good forecasting method.

Proposition 3. *The sequence $\{\hat{H}^k\}$ defined in (12) is a good forecasting method for $\{Y^k\}$.*

Empirical Moving Average (EMA) Model. Suppose that the distribution function $\hat{H}^k$ defined in (4) can be written as

$$\hat{H}^k(\omega,\cdot) := H(M^k(\omega),\cdot) \tag{13}$$
$$M^k(\omega) := \frac{k-1}{k}M^{k-1}(\omega) + \frac{1}{k}Y^k(\omega) \quad = \quad \frac{1}{k}\sum_{j=1}^{k} Y^j(\omega). \tag{14}$$

In the above equations, $H(m,\cdot)$ is a distribution with unknown parameter m that the forecaster estimates with the sample average of $\{Y^k\}$. This setting is applicable when $\hat{H}^k$ and the limiting distribution F^* belong to a parametric family with a single unknown parameter given by the mean, such as the exponential and Poisson distributions. In Section 5.1 we examine two such cases in more detail. The proposition below shows that under some assumptions, (13) and (14) specify a good forecasting method for $\{Y^k\}$.

Proposition 4. *Consider a family of distributions $\{H(m,\cdot) : m \in \mathbb{M} \subset \mathbb{R}\}$, where $m = \int xH(m,dx)$ is the mean of $H(m,\cdot)$, $\mathbb{M}$ is closed, and $H(m,\cdot)$ is continuous in m with respect to the topology of weak convergence. Suppose that $\{Y^k\}$ and $\{F^k\}$ as in Definition 1 satisfy $F^k(\omega,\cdot) = H(U^k(\omega),\cdot)$ w.p.1, where $U^k := \mathbb{E}[Y^{k+1} \,|\, \mathcal{F}^k]$. Also suppose that $\sup_{k\geq 0} \mathbb{E}[(Y^{k+1})^2 \,|\, \mathcal{F}^k] < Z$ w.p.1, for some integrable random variable Z. Then $\{\hat{H}^k\}$ in (13)–(14) is a good forecasting method for $\{Y^k\}$.*

If for each m, the distribution $H(m,\cdot)$ has a density $h(m,\cdot)$ with respect to a (common for all k) measure μ, then by Scheffé's Theorem (see, e.g., Billingsley 1968), a sufficient condition for continuity of $H(m,\cdot)$ in m is continuity of $h(m,x)$ in m for μ-almost all x. Hence, it can be readily seen that many widely-used distributions satisfy the continuity assumption in Proposition 4.

It is worthwhile commenting on the relationship between the EMA model and the non-negative exponentially weighted moving average (EWMA) forecasting method. In EWMA, the distribution function $\hat{H}^k$ is defined as in (13), but M^k is defined as $M^k = \rho M^{k-1} + (1-\rho)Y^k$ for some constant $\rho \in [0,1)$. It is easy to check that, in general, EWMA is not a good forecasting method in the sense of Definition 1.

5. Some Specific Cases

In this section we study three forecasting methods and the resulting sequence of protection levels. The first one is based on affine updates, which includes the EMA model described in Section 4, with the underlying distribution being normal or exponential. The second one is the empirical distribution. Finally, the third case involves stochastic approximation updates of the protection levels as proposed by van Ryzin and McGill (2000). The latter method directly updates the protection levels and is not meant to be a forecasting method, but as shown later it fits in the framework of Section 2.3.

5.1. Affine Updates

Suppose that the sequence of protection levels satisfy the following inductive equation:

$$\mathbb{E}\left[L^k \,|\, \mathcal{F}^{k-1}\right] \quad = \quad L^{k-1} + \frac{1}{k}\left[\alpha - (1-\beta)L^{k-1}\right] \quad = \quad \frac{k-1+\beta}{k}L^{k-1} + \frac{\alpha}{k} \tag{15}$$

Before establishing the results, we give some examples in which the induction (15) occurs. Both examples use the EMA forecasting model described in Section 4. The first one assumes that the true distribution is normal with known variance, and aims to estimate the mean; the second one assumes that the true distribution is exponential, and again aims to estimate the mean. In both cases, the observed quantity has a continuous distribution, and in the normal case the observed quantity can assume negative values. Of course, these are not realistic distributions for the demand for airline tickets. Shlifer and Vardi (1975) claim that, based on a study of the data collected by the OR team at El-Al, it was observed that the number of passengers on a flight is approximately normally distributed. According to Belobaba (1989), past analyses generally have assumed that demand is normally distributed. Curry (1990) refers to the truncated normal distribution as typical, and Wollmer (1992) states that demand is often assumed to approximate a continuous distribution such as the normal, and he also presents expressions specifically for the case with normally distributed demand. Brumelle and McGill (1993) point out that the normal distribution is often used with methods such as EMSR. We have learned from conversations with revenue management professionals that to this day the normal distribution is often used in demand models in practice. Thus, although a model with normally distributed observed quantity is unrealistic, such models have been used many times, and therefore we think that it is of interest to take a closer look at the dynamic behavior of such a model. Nevertheless, the settings in this section and Section 5.2 are clearly restrictive.

In this section we assume that the mean of the observed quantity X^k (conditional on $\mathcal{F}^{k-1}$) is equal to the protection level L^{k-1}, and in Section 5.2 we assume that all customers prefer the low fare but are willing to pay the high fare if the low fare is not available. These assumptions are made to facilitate analysis, and in Section 6 we show how the results in this section and Section 5.2 can be used to obtain results regarding the dynamic behavior of more complicated models. For example, we generalize the spiral-down results to the case where the mean of X^k is $\mu(L^{k-1})$, where μ is an arbitrary function such that $\mu(\ell) \le \ell$ for all ℓ.

We use the notation $N(\mu,\sigma^2)$ to denote the normal distribution with mean μ and variance σ^2; with a slight abuse of the notation, $N(\mu,\sigma^2)$ also denotes the distribution function of the normal distribution. That is, if A is a random variable and B is a distribution function, then $A \sim N(\mu,\sigma^2)$ means that A is normally distributed with mean μ and variance σ^2, whereas $B = N(\mu,\sigma^2)$ means that $B(x) = \int_{-\infty}^{x} (\sigma\sqrt{2\pi})^{-1} e^{-(t-\mu)^2/(2\sigma^2)} dt$ for all $x \in \mathbb{R}$. Similarly, we use $\exp(1/\lambda)$ to denote both the exponential distribution with mean λ and its distribution function. The correct interpretation should be clear from the context.

The Normal Case. Suppose that, given any protection level ℓ, X is normally distributed with mean ℓ and variance σ^2, that is, $G(\ell,\cdot) = N(\ell,\sigma^2)$. Suppose that the revenue manager knows that the distribution is normal and also knows σ^2, and assume that he uses the sample average of observed X values to estimate the supposed mean of X. Let M^k denote the sample average of the first k observed X values. Then $\hat{H}^k = N(M^k,\sigma^2)$. The process is started with some protection level L^0, and then, inductively,

$$X^k \sim N(L^{k-1},\sigma^2) \tag{16}$$
$$M^k = \frac{k-1}{k} M^{k-1} + \frac{1}{k} X^k \tag{17}$$
$$\hat{H}^k = N(M^k,\sigma^2) \tag{18}$$
$$L^k = (\hat{H}^k)^{-1}(\gamma) \quad = \quad M^k + \alpha \tag{19}$$

where $\alpha = \sigma\Phi^{-1}(\gamma)$ and Φ denotes the standard normal cdf. It follows that

$$\mathbb{E}\left[L^k \mid \mathcal{F}^{k-1}\right] \quad = \quad L^{k-1} + \frac{\alpha}{k} \tag{20}$$

Thus, in this example, $\beta = 1$.

The Exponential Case. Suppose that, given any protection level ℓ, X is exponentially distributed with mean ℓ, that is, $G(\ell,\cdot) = \exp(1/\ell)$. Suppose that the revenue manager knows that the distribution is exponential, and suppose that he uses the sample average of observed X values to estimate the supposed mean of X. Let M^k denote the sample average of the first k observed X values. Then $\hat{H}^k = \exp(1/M^k)$. The process is started with some protection level L^0, and then, inductively,

$$X^k \sim \exp(1/L^{k-1}) \tag{21}$$
$$M^k = \frac{k-1}{k} M^{k-1} + \frac{1}{k} X^k \tag{22}$$
$$\hat{H}^k = \exp(1/M^k) \tag{23}$$
$$L^k = (\hat{H}^k)^{-1}(\gamma) \quad = \quad \ln\left(\frac{1}{1-\gamma}\right) M^k. \tag{24}$$

It follows that

$$\mathbb{E}\left[L^k \mid \mathcal{F}^{k-1}\right] \quad = \quad \frac{k-1+\ln\left(\frac{1}{1-\gamma}\right)}{k} L^{k-1} \tag{25}$$

Thus, in this example, $\beta = \ln(1/(1-\gamma))$ and $\alpha = 0$.

Next we consider sequences $\{L^k\}$ generated by (15). Let m be a positive integer such that $m-1+\beta > 0$. Let $\{(f^k, g^k)\}_{k=m}^{\infty}$ be a sequence defined inductively by

$$f^{m-1} \;:=\; 1, \quad f^k \;:=\; f^{k-1}\tfrac{k}{k-1+\beta} \quad \text{for all } k \ge m \tag{26}$$
$$g^{m-1} \;:=\; 0, \quad g^k \;:=\; g^{k-1} + \tfrac{\alpha}{k} f^k \quad \text{for all } k \ge m. \tag{27}$$

Note that it follows from (15) that

$$\mathbb{E}\left[f^k L^k - g^k \mid \mathcal{F}^{k-1}\right] \quad = \quad f^{k-1}L^{k-1} - g^{k-1}. \tag{28}$$

In what follows we need the following assumption for $\{f^k L^k - g^k\}$ to be a convergent martingale. Later we give examples of cases where this condition is satisfied.

Assumption (A). The sequence $\{(f^k, g^k, L^k)\}_{k=m}^{\infty}$ satisfies $\sup_{k \ge m} \mathbb{E}|f^k L^k - g^k| < \infty$.

The proposition below gives the behavior of the sequence $\{L^k\}$ in terms of α and β.

Proposition 5. *Suppose that Assumption* (A) *holds. Then, the sequence* $\{f^k L^k - g^k\}$ *forms a convergent martingale, that is, there exists a finite random variable* A *such that* $f^k L^k - g^k \to A$ *w.p.1 as* $k \to \infty$*. In addition,*

1. *If* $\beta < 1$ *and* $\alpha = 0$*, then* $L^k \to 0$ *w.p.1.*
2. *If* $\beta = 1$ *and* $\alpha = 0$*, then* $\{L^k\}$ *is a martingale and* $L^k \to A$ *w.p.1.*
3. *If* $\beta = 1$ *and* $\alpha \neq 0$*, then* $L^k \to sgn(\alpha)\infty$ *w.p.1.*

If $\beta > 1$, then we are not sure how L^k behaves, except that there are many cases in which $L^k \to \pm\infty$, as explained later. If $\beta < 1$ and $\alpha \neq 0$, then without additional information we do not know how L^k behaves. For more discussion, see the Online Appendix. Next we return to the examples.

The Normal Case (continued). Recall that in this case, $\alpha = \sigma\Phi^{-1}(\gamma)$, and $\beta = 1$. First we establish that Assumption (A) holds.

Lemma 1. *If the system evolves according to* (16)–(19)*, then Assumption* (A) *holds.*

The following result follows from Proposition 5 and Lemma 1:

Proposition 6. *If the system evolves according to* (16)–(19)*, then there exists a finite random variable* A *such that* $f^k L^k - g^k \to A$ *w.p.1 as* $k \to \infty$*. In addition, the following holds:*

(i) If $\gamma := 1 - f_2/f_1 < 1/2$, i.e., if $\alpha < 0$, then $L^k \to -\infty$ w.p.1 as $k \to \infty$;
(ii) If $\gamma > 1/2$, i.e., if $\alpha > 0$, then $L^k \to +\infty$ w.p.1 as $k \to \infty$;
(iii) If $\gamma = 1/2$, i.e., if $\alpha = 0$, then $L^k \to A$ w.p.1 as $k \to \infty$. Moreover, in this case $A \sim N(L^0, \sigma^2 \sum_{i=1}^{\infty} 1/i^2) = N(L^0, \sigma^2\pi^2/6)$.

In applications, it is often the case that $f_2/f_1 > 1/2$, and thus $L^k \to -\infty$ as $k \to \infty$, which is the spiral-down effect.

The Exponential Case (continued). Recall that in this case, $\alpha = 0$ and $\beta = \ln(1/(1-\gamma))$. Note that $\beta > 0$. First we establish that Assumption (A) holds.

LEMMA 2. *If the system evolves according to* (21)–(24)*, then Assumption* (A) *holds.*

The following result follows from Proposition 5 and Lemma 2:

PROPOSITION 7. *If the system evolves according to* (21)–(24)*, then there exists a finite random variable A such that $f^k L^k \to A$ w.p.1 as $k \to \infty$. In addition, the following holds:*
(i) If $1 - \gamma = f_2/f_1 > 1/e$, i.e., if $\beta \in (0,1)$, then $L^k \to 0$ w.p.1 as $k \to \infty$;
(ii) If $1 - \gamma = f_2/f_1 = 1/e$, i.e., if $\beta = 1$, then $L^k \to A$ w.p.1 as $k \to \infty$;
(iii) If $1 - \gamma = f_2/f_1 < 1/e$, i.e., if $\beta > 1$, then $f^k \to 0$ as $k \to \infty$. Thus, for those $\omega \in \Omega$ such that $f^k L^k(\omega) \to A(\omega) > 0$, it holds that $L^k(\omega) \to \infty$. However, if $A(\omega) = 0$, then we need more information to determine the asymptotic behavior of L^k.

In applications, it is often the case that $f_2/f_1 > 1/e$, and thus $L^k \to 0$ as $k \to \infty$, which again gives the spiral-down effect.

5.2. Empirical Distribution

In this section we study the protection levels resulting from forecasting with the empirical distribution (7). Consider the same setting of the deterministic example in Section 3, but assume that the total demand D is random. Thus, if the protection level is ℓ, then the observed quantity is given by $X := [D - (c - \ell)^+]^+$.

Consider a sequence of outcomes in the situation described above. Assume that $\{D^k : k \geq 1\}$ is an i.i.d. sequence, and let $X^k := [D^k - (c - L^{k-1})^+]^+$. Given $X^1, \ldots, X^k$, the revenue manager constructs the corresponding empirical distribution $\hat{H}^k$ and chooses L^k to be a γ-quantile of $\hat{H}^k$. The next observed quantity X^{k+1} is then given by $X^{k+1} := [D^{k+1} - (c - L^k)^+]^+$.

Next we compute L^k explicitly in terms of $X^1, \ldots, X^k$. Let $X^{1:k}, \ldots, X^{k:k}$ denote the order statistics of $X^1, \ldots, X^k$. Note that if $k\gamma$ is an integer, then the set $(\hat{H}^k)^{-1}(\gamma)$ of γ-quantiles of $\hat{H}^k$ may not be a singleton. However, for any $\gamma \in (0,1]$ and any k, we have that $X^{\lceil k\gamma \rceil:k}$ is an element of $(\hat{H}^k)^{-1}(\gamma)$. Therefore, in this section we follow the convention that

$$L^k \;:=\; X^{\lceil k\gamma \rceil:k} \;\in\; (\hat{H}^k)^{-1}(\gamma)$$

Now we can compare $L^k := X^{\lceil k\gamma \rceil:k}$ and $L^{k+1} := X^{\lceil (k+1)\gamma \rceil:k+1}$. There are two cases:
Case 1: $\lceil (k+1)\gamma \rceil = \lceil k\gamma \rceil$. Then,

$$X^{k+1} \geq L^k \iff L^{k+1} = L^k \tag{29}$$
$$X^{\lceil k\gamma \rceil - 1:k} \leq X^{k+1} \leq L^k \iff L^{k+1} = X^{k+1} \tag{30}$$
$$X^{k+1} \leq X^{\lceil k\gamma \rceil - 1:k} \iff L^{k+1} = X^{\lceil k\gamma \rceil - 1:k}. \tag{31}$$

Case 2: $\lceil (k+1)\gamma \rceil = \lceil k\gamma \rceil + 1$. Then,

$$X^{k+1} \leq L^k \iff L^{k+1} = L^k \tag{32}$$
$$L^k \leq X^{k+1} \leq X^{\lceil k\gamma \rceil + 1:k} \iff L^{k+1} = X^{k+1} \tag{33}$$
$$X^{k+1} \geq X^{\lceil k\gamma \rceil + 1:k} \iff L^{k+1} = X^{\lceil k\gamma \rceil + 1:k}. \tag{34}$$

Using the fact that $X^{k+1} := [D^{k+1} - (c - L^k)^+]^+$, we can compute the probabilities of the events on the left hand side of the above implications. For any $a \geq 0$,

$$\mathbb{P}\left[X^{k+1} > a | \mathcal{F}^k\right] = \mathbb{P}\left[[D^{k+1} - (c - L^k)^+]^+ > a | \mathcal{F}^k\right] = \mathbb{P}\left[D^{k+1} > a + (c - L^k)^+ | \mathcal{F}^k\right].$$

In particular, if $L^k \geq 0$ then

$$\mathbb{P}\left[X^{k+1} > L^k | \mathcal{F}^k\right] = \mathbb{P}\left[D^{k+1} > L^k + (c - L^k)^+ | \mathcal{F}^k\right] = \mathbb{P}\left[D^{k+1} > \max\{c, L^k\} | \mathcal{F}^k\right]. \tag{35}$$

Based on the above relations we can reach some conclusions regarding the behavior of $\{L^k\}$. For instance, if $D^k \leq c$ w.p.1, then we see from (35) that $X^{k+1} \leq L^k$ w.p.1, and so from (29)–(34) we have that $L^{k+1} \leq L^k$ w.p.1. Since $\{L^k\}$ is bounded below by zero, it is a bounded monotone sequence, so there exists a random variable L such that $L^k \to L$ w.p.1 as $k \to \infty$. On the other extreme, if $D^k \geq c$ w.p.1, then it is easy to see from (35) that $X^k = D^k$ for k large enough, i.e., after a transient period, $\{X^k\}$ is an i.i.d. sequence. In that case, the distance between L^k and the set of γ-quantiles of the distribution of D converges to 0 (cf. Lemma 4).

Similar results are obtained if the revenue manager does not continue to observe customers after c tickets have been sold. In this case the observed quantity X is equal to the observed sales of class-1 tickets. This situation corresponds to Example 2.2.B with $D_{ab} = D^k$, and we have $X^{k+1} = \min\{[D^{k+1} - (c - L^k)^+]^+, c, L^k\}$. It follows that $X^{k+1} \leq L^k$ w.p.1, and so as before it follows that there exists a random variable L such that $L^k \to L$ w.p.1 as $k \to \infty$. Unlike the case in the previous paragraph, in this case the protection levels spiral down irrespective of the relation between D^k and c.

5.3. Stochastic Approximation Updates

In this section we discuss a stochastic approximation algorithm for updating protection levels, as proposed by van Ryzin and McGill (2000). Their approach does not require that the demand distributions be known, but it does require that the demand for tickets of different fare classes be exogenous. In this section we consider what happens if the demand for tickets of different fare classes depends on the chosen protection levels and the revenue manager uses a stochastic approximation algorithm.

Similar to most published revenue management work, and translated into our notation for two ticket classes (van Ryzin and McGill consider an n-class problem), their model assumes that the observed quantity X is the demand for fare class 1, and that the distribution $G(\cdot)$ of the demand for fare class 1 is *independent of the protection level* ℓ; that is, $G(\ell, \cdot) = G(\cdot)$ for all ℓ. Hence, for continuous G, the goal is to find a protection level L^* that satisfies $G(L^*) = \gamma$. It is interesting to observe that such a protection level L^* is given by $G^{-1}(\gamma)$, which is the limiting point established by Propositions 15 and 16 in Section 7.

The proposed method updates the protection levels L^k according to the equation

$$L^{k+1} := L^k + \xi_k \left[\gamma - \mathbb{I}_{\{X^{k+1} \leq L^k\}}\right] \tag{36}$$

where $\{\xi_k\}$ is a sequence of nonnegative step sizes satisfying $\sum_k \xi_k = \infty$ and $\sum_k \xi_k^2 < \infty$. Note that this scheme corresponds to a degenerate forecasting distribution $\hat{H}^k$ that is updated as follows:

$$\hat{H}^{k+1}(x) := \begin{cases} \mathbb{I}_{\{L^k - \xi_k(1-\gamma) \leq x\}} & \text{if } X^{k+1} \leq L^k \\ \mathbb{I}_{\{L^k + \xi_k \gamma \leq x\}} & \text{otherwise} \end{cases}$$

Their primary result shows that under some assumptions, if protection levels are updated using the stochastic approximation algorithm above, then the sequence of protection levels converges to a protection level L^* that satisfies $G(L^*) = \gamma$.

Below, we show under some conditions that if the distribution of the observed quantity depends on the protection level and if L^k is updated according to (36), then L^k converges to a deterministic limit ℓ^*. Under an additional continuity assumption it follows that $G(L^k, L^k) \to G(\ell^*, \ell^*) = \gamma$ w.p.1. We use the following result from Section II.5.1 of Benveniste et al. (1990).

PROPOSITION 8. *Consider the random sequences $\{X^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ and $\{L^k\}_{k=0}^{\infty} \subset \mathbb{R}^m$ that satisfy $L^{k+1} = L^k + \xi_k S(L^k, X^{k+1})$, where $\{\xi_k\}_{k=0}^{\infty}$ is a deterministic nonnegative step size sequence that satisfies $\sum_{k=0}^{\infty} \xi_k = \infty$ and $\sum_{k=0}^{\infty} \xi_k^2 < \infty$. Let $\mathcal{F}^k$ denote the σ-algebra generated by $X^1, \ldots, X^k, L^0, \ldots, L^k$. Suppose that the following assumptions hold:*

1. *For each $\ell \in \mathbb{R}^m$, there exists a probability distribution $G(\ell, \cdot)$ on $\mathbb{R}^n$ such that for any Borel function $g : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}_+$,*

$$\mathbb{E}\left[g(L^k, X^{k+1}) \,|\, \mathcal{F}^k\right] \;=\; \int_{\mathbb{R}^n} g(L^k, x) G(L^k, dx)$$

w.p.1. That is, the conditional distribution of X^{k+1}, given the history of the process up to iteration k, depends only on L^k.

2. *There exist constants $K_1, K_2 > 0$ such that*

$$\int_{\mathbb{R}^n} \|S(\ell, x)\|^2 G(\ell, dx) \;\leq\; K_1 + K_2 \|\ell\|^2$$

for all $\ell \in \mathbb{R}^m$.

3. *Let $s : \mathbb{R}^m \mapsto \mathbb{R}^m$ be given by $s(\ell) := \int_{\mathbb{R}^n} S(\ell, x) G(\ell, dx)$. There exists an $\ell^* \in \mathbb{R}^m$ such that*

$$\inf\left\{(\ell^* - \ell)^T s(\ell) \,:\, \varepsilon \leq \|\ell^* - \ell\| \leq 1/\varepsilon\right\} \;>\; 0$$

for all $\varepsilon > 0$. That is, the conditional expected direction vector $s(\ell)$ points to the interior of the halfspace at ℓ that contains ℓ^*.

Then $L^k \to \ell^*$ w.p.1.

Assumptions 1 and 2 of Proposition 8 imply that, w.p.1, $\mathbb{E}[\|S(L^k, X^{k+1})\|^2 \,|\, \mathcal{F}^k] \leq K_1 + K_2 \|L^k\|^2$ and the conditional expected direction vector $s(L^k) = \mathbb{E}[S(L^k, X^{k+1}) \,|\, \mathcal{F}^k]$ is finite for all k.

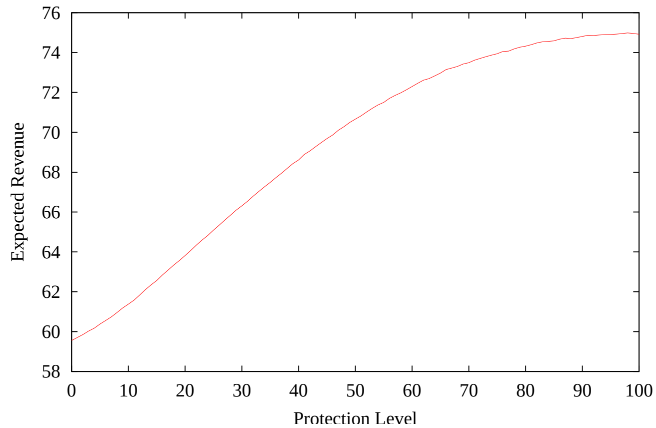
Let $F : \mathbb{R} \mapsto [0,1]$ be given by $F(\ell) := G(\ell, \ell)$. In our problem, $S(\ell, x) = \gamma - \mathbb{I}_{\{x \leq \ell\}}$, and thus $s(\ell) = \gamma - G(\ell, \ell) = \gamma - F(\ell)$. It is easy to see that assumption 3 of Proposition 8 is satisfied if and only if Assumption (B1) holds.

ASSUMPTION (B1). There exists an $\ell^* \in \mathbb{R}$ such that the following holds. For any $\varepsilon > 0$, there exists a $\delta > 0$ such that $F(\ell) \leq \gamma - \delta$ for all $\ell \in [\ell^* - 1/\varepsilon, \ell^* - \varepsilon]$ and $F(\ell) \geq \gamma + \delta$ for all $\ell \in [\ell^* + \varepsilon, \ell^* + 1/\varepsilon]$.

Under the assumptions of van Ryzin and McGill (2000), Assumption (B1) holds. In their setting, G does not depend on the protection level, and thus $F(\ell) = G(\ell)$. They assume that there is a constant $c > 0$ such that $-(\ell - \ell^*)[\gamma - F(\ell)] \geq c|\ell - \ell^*|^2$ for all ℓ. That is, $F(\ell) \leq \gamma + c(\ell - \ell^*)$ for all $\ell \leq \ell^*$, and $F(\ell) \geq \gamma + c(\ell - \ell^*)$ for all $\ell \geq \ell^*$, which imply that Assumption (B1) holds with $\delta = c\varepsilon$. (In fact, their assumption cannot hold for all ℓ, because $F(\ell) \in [0,1]$ for all ℓ. Nevertheless, since F is nondecreasing when G does not depend on ℓ, if $-(\ell - \ell^*)[\gamma - F(\ell)] \geq c|\ell - \ell^*|^2$ for all ℓ in some neighborhood of ℓ^*, then Assumption (B1) is satisfied, and their line of argument works.)

Next we verify that the conditions in Proposition 8 are satisfied in our problem if Assumption (B1) holds. It follows from Proposition 5 on p. 451 of Fristedt and Gray (1997) that the assumption in (3) and assumption 1 of Proposition 8 are equivalent. Next, recall that $S(\ell, x) = \gamma - \mathbb{I}_{\{x \leq \ell\}} \in (-1, 1)$ for all ℓ and x, and thus $\int_{\mathbb{R}} |S(\ell, x)|^2 G(\ell, dx) < 1$ for all $\ell \in \mathbb{R}$, that is, assumption 2 of Proposition 8 holds. Finally, assumption 3 of Proposition 8 and Assumption (B1) are equivalent. Proposition 9 follows from these observations.

Figure 1 Example 5.1: Expected revenue as a function of the protection levels.



Proposition 9. *Suppose that Assumption* (B1) *holds and that the protection levels are updated according to* (36)*. Then $L^k \to \ell^*$ w.p.1.*

Note that the assumptions do not require F to be continuous at ℓ^*. If in addition it is assumed that F is continuous at ℓ^*, then $F(L^k) \to F(\ell^*) = \gamma$ w.p.1, that is, $G(L^k, L^k) \to G(\ell^*, \ell^*) = \gamma$ w.p.1, and $\ell^* \in G^{-1}(\ell^*, \gamma)$. Alternatively, when ℓ takes on integer values only and the family of distributions $\{G(\ell, \cdot)\}$ is stochastically increasing in ℓ, then $G(\ell^*, \ell^* - 1) \le G(\ell^* - 1, \ell^* - 1) < \gamma < G(\ell^*, \ell^*)$, i.e., $\ell^* \in G^{-1}(\ell^*, \gamma)$.

The Online Appendix gives sufficient conditions for the sequence $\{L^k\}$ generated by the stochastic approximation method to satisfy $G(L^k, L^k) \to \gamma$ w.p.1, even if L^k does not converge to a deterministic or random limit. In addition, it is shown that every limit point L^* of $\{L^k\}$ satisfies $G(L^*, L^*) = \gamma$, that is, $L^* \in G^{-1}(L^*, \gamma)$.

Next we consider two examples with stochastic approximation updates, which correspond to Examples 2.2.A and 2.2.B. We start with the following result.

Proposition 10. *In Examples 2.2.A and 2.2.B, the function $F(\ell) := G(\ell, \ell)$ is nondecreasing.*

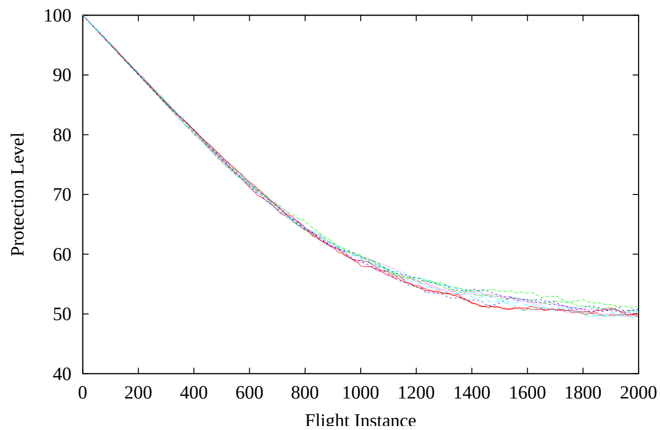
It follows that F is a nondecreasing step function, and thus F satisfies Assumption (B1) unless γ happens to be equal to one of the values of F. That is, $\{L^k\}$ converges w.p.1 to some ℓ^*. Notice also that in those examples the family of distributions $\{G(\ell, \cdot)\}$ is stochastically increasing in ℓ, so that $\ell^* \in G^{-1}(\ell^*, \gamma)$.

In the following examples, class-1 tickets have price 1, class-2 tickets have price 0.5 (so $\gamma = 0.5$), the time horizon is 100, the capacity is $c = 100$, and the step size parameter is $\xi_k = 10^4/(10^5 + k)$.

Example 5.1. This example corresponds to Example 2.2.A, in which the observed quantity X is the "untruncated class-1 demand". Type-a customers arrive according to a nonhomogeneous Poisson process with rate $0.005t$, type-b customers arrive according to a nonhomogeneous Poisson process with rate $0.5 - 0.005t$, and type-ab customers arrive according to a homogeneous Poisson process with rate 0.5. Notice that the expected total number of arrivals over the booking period is 100, of which 50% are type-ab customers, 25% are type-a and 25% are type-b. Figure 1 below shows the expected revenue (estimated via simulation) as a function of the protection level for this system. The optimal protection level is 98, which corresponds to an expected revenue of about 75. We did calculate confidence intervals for the estimated quantities, but since they were negligible due to the large sample size used we chose not to display them.

Figure 2 shows 10 sample paths of the protection levels produced by the stochastic approximation method. In this example, when $\ell^* = 49$ we have $G(\ell^* - 1, \ell^* - 1) < \gamma$ and $G(\ell^*, \ell^*) > \gamma$ (these

Figure 2 Example 5.1: Spiral down of protection levels from the optimal protection level of 100, shown for 10 sample paths.



quantities were estimated via simulation, again with a large sample size). Thus, by Proposition 10, Assumption (B1) holds. Notice that the protection levels do spiral down from the nearly-optimal protection level of $L^0 = 100$ to $\ell^* = 49$, as predicted by Proposition 9.

Example 5.2. The parameters for this example are the same as for Example 5.1, except that the observed quantity X is as in Example 2.2.B, that is, X is equal to the number of class-1 tickets sold. As it turns out, for this example we again have $G(\ell^* - 1, \ell^* - 1) < \gamma$ and $G(\ell^*, \ell^*) > \gamma$ for $\ell^* = 49$, and thus it follows from Proposition 10 that Assumption (B1) is satisfied. The graph for this system is very similar to that in Figure 2, so we do not display it.

A few comments about the above examples are in order. First, notice that the two examples above — which correspond to untruncated and truncated observations — behave very similarly. In general, this is not the case; in fact, the differences become starker as the ratio of expected demand by capacity increases (in these examples the ratio is equal to 1). Also, from Figure 1 we see that the expected revenue corresponding to $\ell^* = 49$ is about 70. That is, the loss in revenue resulting from the modeling error is above 6% (compared to the revenue corresponding to the optimal protection level of 98), which is a significant amount in terms of airline revenues. Finally, it is worth mentioning that the limit of the spiral-down is directly related to the percentage of flexible (i.e., type-ab) customers, as given in the following result:

Proposition 11. *Consider Examples 2.2.A and 2.2.B, and suppose that all customers are type-ab. Suppose that the protection levels are updated according to* (36). *In Example 2.2.A, if* $\mathbb{P}[D_{ab} \le c] > \gamma$ *then* $L^k \to 0$ *w.p.1, whereas if* $\mathbb{P}[D_{ab} \le c] < \gamma$*, and there is no* $\ell > c$ *such that* $\mathbb{P}[D_{ab} \le \ell] = \gamma$*, then* $L^k \to \ell^*$ *w.p.1 for some* $\ell^* > c$*. In Example 2.2.B,* $L^k \to 0$ *w.p.1 regardless of anything else.*

6. Extensions

In the previous sections we considered the dynamic behavior of sequences of forecasts $\{\hat{H}^k\}$ and protection levels $\{L^k\}$ for various families of distributions $G(\ell, \cdot)$ of X and various forecasting methods. Next we show how to extend these results to other settings. In Sections 6.1 and 6.2 we discuss extensions through stochastic comparisons and pathwise comparisons respectively, of random variables in the other settings with random variables considered in the previous sections.

6.1. Stochastic Comparisons

Let $\le_{\rm st}$ denote the usual stochastic order for distributions on the real line; i.e., for real-valued random variables $Z_1 \sim H_1, Z_2 \sim H_2$, we write $Z_1 \le_{\rm st} Z_2$ or $H_1 \le_{\rm st} H_2$ if for all bounded nondecreasing measurable functions $f:\mathbb{R}\mapsto\mathbb{R}$, $\int_{\mathbb{R}} f dH_1 \le \int_{\mathbb{R}} f dH_2$, i.e., $\mathbb{E}[f(Z_1)]\le\mathbb{E}[f(Z_2)]$. It can be shown that $H_1 \le_{\text{st}} H_2$ if and only if $H_1(x)\ge H_2(x)$ for all $x\in\mathbb{R}$ (Müller and Stoyan 2002). Consider the Polish space $\mathcal{P}(\mathbb{R})$ of distributions on the real line endowed with a metric that induces weak convergence, such as the Lévy metric, and the closed partial order $\le_{\text{st}}$ defined above (Kamae et al. 1977). Let $\preceq_{\text{st}}$ denote the usual stochastic order on the space $\mathcal{M}(\mathcal{P}(\mathbb{R}))$ of probability measures on $\mathcal{P}(\mathbb{R})$, that is, for $P_1,P_2\in\mathcal{M}(\mathcal{P}(\mathbb{R}))$, and two $\mathcal{P}(\mathbb{R})$-valued random elements $H_1\sim P_1$ and $H_2\sim P_2$, we write $H_1\preceq_{\text{st}} H_2$ or $P_1\preceq_{\text{st}} P_2$ if for all bounded nondecreasing measurable functions $f:\mathcal{P}(\mathbb{R})\mapsto\mathbb{R}$, $\int_{\mathcal{P}(\mathbb{R})} f dP_1 \le \int_{\mathcal{P}(\mathbb{R})} f dP_2$, i.e., $\mathbb{E}_{P_1}[f(H_1)]\le\mathbb{E}_{P_2}[f(H_2)]$. Note that $f:\mathcal{P}(\mathbb{R})\mapsto\mathbb{R}$ nondecreasing means that for any $h_1,h_2\in\mathcal{P}(\mathbb{R})$ with $h_1\le_{\text{st}} h_2$, it holds that $f(h_1)\le f(h_2)$.

Suppose that we want to consider a setting with a family $\{\underline{G}(\ell,\cdot)\}_\ell$ of distributions of observed quantity $\underline{X}$ and a particular forecasting method, producing sequences of forecasts $\{\underline{\hat{H}}^k\}$ and protection levels $\{\underline{L}^k\}$. Suppose that for a setting with distributions $G(\ell,\cdot)$ of X, forecasts $\{\hat{H}^k\}$, and protection levels $\{L^k\}$ — for example, a setting considered in one of the previous sections — one can establish that $\underline{\hat{H}}^k \preceq_{\text{st}} \hat{H}^k$ for all k. Then it follows that

$$\underline{L}^k \;\equiv\; \min\left\{x\in\mathbb{R} \,:\, \underline{\hat{H}}^k(x)\ge\gamma\right\} \;=\; \min\left\{x\in\left(\underline{\hat{H}}^k\right)^{-1}(\gamma)\right\} \;\le_{\text{st}}\; L^k \;\in\; \left(\hat{H}^k\right)^{-1}(\gamma)$$

for all k. In such a case, a result that implies spiral-down of $\{L^k\}$, in distribution or almost surely, also implies spiral-down of $\{\underline{L}^k\}$ in distribution. Next we give examples of how the results established in the previous sections can be extended as described above.

Proposition 12 **(Stochastic comparison with empirical distributions)**. *Suppose* $\underline{G}(\underline{\ell},\cdot)\le_{\text{st}} G(\ell,\cdot)$ *for all* $\underline{\ell}\le\ell$, *and the empirical distribution is used for both* $\hat{H}$ *and* $\underline{\hat{H}}$, *that is,* $\hat{H}^k(x):=k^{-1}\sum_{j=1}^k \mathbb{I}_{\{X^j\le x\}}$ *and* $\underline{\hat{H}}^k(x):=k^{-1}\sum_{j=1}^k \mathbb{I}_{\{\underline{X}^j\le x\}}$. *If* $\underline{L}^0\le_{\text{st}} L^0$, *then*

$$\begin{aligned}\underline{G}(\underline{L}^k,\cdot) &\preceq_{\text{st}} G(L^k,\cdot)\\ \underline{X}^{k+1} &\le_{\text{st}} X^{k+1}\\ \underline{\hat{H}}^{k+1} &\preceq_{\text{st}} \hat{H}^{k+1}\\ \underline{L}^{k+1} &\le_{\text{st}} L^{k+1}\end{aligned}$$

for all $k=0,1,\ldots$.

Proposition 12 can be used to stochastically bound a sequence $\{\underline{X}^k,\underline{\hat{H}}^k,\underline{L}^k,\underline{G}(\underline{L}^k,\cdot)\}$ with another sequence $\{X^k,\hat{H}^k,L^k,G(L^k,\cdot)\}$, such as the sequence considered in Section 3, as follows: Suppose that, for some deterministic constant d, it holds that given the protection level $\underline{\ell}$, the observed quantity $\underline{X}$ is almost surely less than or equal to $[d-(c-\underline{\ell})^+]^+$, and that the empirical distribution is used for $\underline{\hat{H}}$. If $\underline{L}^0\le_{\text{st}} L^0$, then $\underline{G}(\underline{L}^k,\cdot)\preceq_{\text{st}} G(L^k,\cdot)$, $\underline{X}^{k+1}\le_{\text{st}} X^{k+1}$, $\underline{\hat{H}}^{k+1}\preceq_{\text{st}}\hat{H}^{k+1}$, and $\underline{L}^{k+1}\le_{\text{st}} L^{k+1}$ for all $k=0,1,\ldots$. Note that if L^0 is a deterministic constant, then $\underline{L}^0\le_{\text{st}} L^0$ implies that w.p.1, $\underline{L}^k\le L^k$, $\underline{G}(\underline{L}^k,\cdot)\le_{\text{st}} G(L^k,\cdot)$, $\underline{X}^{k+1}\le X^{k+1}$, $\underline{\hat{H}}^{k+1}\le_{\text{st}}\hat{H}^{k+1}$ for all $k=0,1,\ldots$. In addition, if $d<c$, then $\{\underline{L}^k\}$ spirals down to zero w.p.1.

Proposition 13 **(Stochastic comparison with affine updates)**. *Suppose that* $\mu:\mathbb{R}\mapsto\mathbb{R}$ *satisfies* $\mu(\ell)\le\ell$ *for all* ℓ. *Suppose that* $\underline{G}(\underline{\ell},\cdot)=G(\mu(\underline{\ell}),\cdot)$, *and that* $G(\underline{\ell},\cdot)\le_{\text{st}} G(\ell,\cdot)$ *for all* $\underline{\ell}\le\ell$. *Suppose that* $\hat{H}^k=G(M^k,\cdot)$ *and* $\underline{\hat{H}}^k=G(\underline{M}^k,\cdot)$, *where* $M^k=k^{-1}\sum_{j=1}^k X^j$ *and* $\underline{M}^k=k^{-1}\sum_{j=1}^k\underline{X}^j$. *If* $\underline{L}^0\le_{\text{st}} L^0$, *then*

$$\begin{aligned}\underline{G}(\underline{L}^k,\cdot) &\preceq_{\text{st}} G(L^k,\cdot)\\ \underline{X}^{k+1} &\le_{\text{st}} X^{k+1}\\ \underline{M}^{k+1} &\le_{\text{st}} M^{k+1}\\ \underline{\hat{H}}^{k+1} &\preceq_{\text{st}} \hat{H}^{k+1}\\ \underline{L}^{k+1} &\le_{\text{st}} L^{k+1}\end{aligned}$$

for all $k = 0, 1, \ldots$.

Observe that Proposition 13 in conjunction with Proposition 6(i) or 7(i) allows us to identify many situations in which spiral down occurs. For instance, if the conditions of Proposition 13 and Proposition 7(i) hold, then the distribution of L^k converges weakly to the point mass at zero. Note that the assumption $\mu(\ell) \le \ell$ is natural, since larger values of the protection level ℓ tend to produce larger values of the observed quantity X for small values of ℓ only.

6.2. Pathwise Comparisons

The stochastic comparisons in the previous section do not require the sequences $\{X^k, \hat{H}^k, L^k, G(L^k, \cdot)\}$ and $\{\underline{X}^k, \underline{\hat{H}}^k, \underline{L}^k, \underline{G}(\underline{L}^k, \cdot)\}$ to be constructed on the same probability space. If the sequences are defined on the same probability space, then stronger results such as $\underline{L}^k \le L^k$ w.p.1 can sometimes be obtained. For example, suppose that one wants to study the behavior of a sequence $\{\underline{X}^k, \underline{\hat{H}}^k, \underline{L}^k, \underline{G}(\underline{L}^k, \cdot)\}$, but that complications hinder direct analysis of the sequence. One may construct another sequence $\{X^k, \hat{H}^k, L^k, G(L^k, \cdot)\}$ of which the behavior is already understood on the same probability space, and then derive insight regarding the behavior of $\{\underline{X}^k, \underline{\hat{H}}^k, \underline{L}^k, \underline{G}(\underline{L}^k, \cdot)\}$ through pathwise comparisons with $\{X^k, \hat{H}^k, L^k, G(L^k, \cdot)\}$.

Another setting in which pathwise comparisons are natural is as follows. Suppose that $\underline{X}^k$ is the number of high-fare tickets sold (truncated demand), and that X^k is the untruncated demand estimate corresponding to the same underlying customer demand. Then it is natural to model the two sequences on the same probability space, and to assume that if $\underline{L}^k \le L^k$ w.p.1, then $\underline{X}^{k+1} \le X^{k+1}$ w.p.1. We obtain the following pathwise comparison result.

Proposition 14 **(Pathwise comparison)**. *Consider any* $\omega \in \Omega$ *such that, for any* k, $\underline{L}^k(\omega) \le L^k(\omega)$ *implies that* $\underline{X}^{k+1}(\omega) \le X^{k+1}(\omega)$. *Suppose that the forecasting method used in both sequences satisfies the following condition for all* k*: If* $(\underline{X}^1(\omega), \ldots, \underline{X}^k(\omega)) \le (X^1(\omega), \ldots, X^k(\omega))$, *then* $\underline{\hat{H}}^k(\omega, \cdot) \le_{\text{st}} \hat{H}^k(\omega, \cdot)$. *If* $\underline{L}^0(\omega) \le L^0(\omega)$, *then*

$$\begin{aligned} \underline{X}^k(\omega) &\le X^k(\omega) \\ \underline{\hat{H}}^k(\omega, \cdot) &\le_{\text{st}} \hat{H}^k(\omega, \cdot) \\ \underline{L}^k(\omega) &\le L^k(\omega) \end{aligned}$$

for all $k = 1, 2, \ldots$.

Note that the result above applies to individual sample paths ω. If the assumptions of the proposition hold w.p.1, then the conclusions hold w.p.1.

6.3. Batching of Observations

In this section we consider a variation of the methods in Sections 5.1–5.3 in which the protection level is not updated after every observation of X but rather after a batch consisting of n observations (recall one observation corresponds to one flight instance). Under this approach, if the underlying forecast method is good — in the sense of Section 4 — then for any fixed value of L^{k-1} the batch forecasts converge to the true distribution $G(L^{k-1}, \cdot)$ as n becomes large. Thus, with larger values of n, the choice of the next protection level L^k will be based on a forecast that tends to be closer to the distribution $G(L^{k-1}, \cdot)$ than with smaller values of n.

Let $\hat{H}_n^k$ denote the forecast and let L_n^k denote the protection level after batch k has been observed, with each batch consisting of n observations. Thus, L_n^1 is based on observations $X^1, \ldots, X^n$, and L_n^{k+1} is updated using new observations $X^{kn+1}, \ldots, X^{(k+1)n}$. Protection levels remain constant between updates; that is, $L^{nk+j} = L_n^k$ for $j = 0, \ldots, n-1$, and

$$\mathbb{P}\left[X^{kn+j+1} \le x \,|\, \mathcal{F}^{kn+j}\right] \;=\; G(L^{kn+j}, x) \;=\; G(L_n^k, x) \tag{37}$$

for all $j = 0, \ldots, n-1$ and all $x \in \mathbb{R}$.

Consider initially the case where the same batch size n is used for all k. It is not difficult to check that the results of Sections 5.1-5.3 are readily extended to this situation. For example, consider the analysis for the normal distribution in Section 5.1. With batching, (16) is replaced by $X^{(k-1)n+1}, \ldots, X^{kn} \sim N(L_n^{k-1}, \sigma^2)$. Define $\bar{X}_n^k := \sum_{j=(k-1)n+1}^{kn} X^j/n$, and replace (17)–(19) by $M_n^k = [(k-1)M_n^{k-1} + \bar{X}_n^k]/k$, $\hat{H}_n^k = N(M_n^k, \sigma^2)$, and $L_n^k = (\hat{H}_n^k)^{-1}(\gamma) = M_n^k + \alpha$. It then follows that $\mathbb{E}[L_n^k \mid \mathcal{F}_n^{k-1}] = L_n^{k-1} + \alpha/k$, where $\mathcal{F}_n^{k-1} := \mathcal{F}^{(k-1)n}$. The proof of Lemma 1 carries over with X^k, M^k, L^k, and σ^2 replaced everywhere by $\bar{X}_n^k$, M_n^k, L_n^k, and $\sigma_n^2 := \sigma^2/n$ respectively. Therefore, Proposition 6 (with L^k and σ^2 replaced by L_n^k and σ_n^2) applies to the situation where updates occur every n observations. Similar conclusions are obtained for the remaining cases in Sections 5.1-5.3.

Consider now the limiting case (as $n \to \infty$) in which $\hat{H}^k = G(L^k, \cdot)$. In the context of the normal distribution discussed in Section 5.1, we have

$$L^k \;=\; G^{-1}(L^{k-1}, \gamma) \;=\; L^{k-1} + \alpha. \tag{38}$$

Clearly, if $\alpha \neq 0$ then $\alpha L^k \to \infty$ as $k \to \infty$, whereas if $\alpha = 0$ then $L^k = L^0$. Note the similarity between this conclusion and the results of Proposition 6.

In the context of the empirical distribution approach described in Section 5.2, stronger conclusions for the limiting case can be obtained, compared to the case of finite n. We have

$$L^k \;=\; G^{-1}(L^{k-1}, \gamma) \tag{39}$$

which is the same as in (38), except that in the former case L^k can be computed explicitly because of the particular form of G. Additional assumptions are needed to establish the convergence of L^k generated by (39). For example, if $f_\gamma : \mathbb{R} \mapsto \mathbb{R}$ given by $f_\gamma(\ell) := G^{-1}(\ell, \gamma)$ is a contraction mapping, then L^k converges to the fixed point of f_γ, which can be a sub-optimal point.

Finally, for the stochastic approximation procedure described in Section 5.3, in the limiting case $(n \to \infty)$ (36) becomes $L^{k+1} := L^k + \xi_k [\gamma - G(L^k, L^k)] = L^k + \xi_k s(L^k)$, and thus we can apply a simpler version of Proposition 8 where Assumptions 1 and 2 are not needed.

In summary, in all cases discussed in this section we have seen that reducing (or even eliminating) the forecast variability does not prevent spiral down. This emphasizes the observation that the spiral-down effect is a consequence of modeling error, and not forecast variability.

7. Relating the Convergence of Forecasts and Protection Levels

We turn now to an analysis of the protection level process $\{L^k\}$ when a general good forecast method is used. As the results in Section 5 illustrate, *ad hoc* tools are typically required to calculate the limiting values (if any) of $\{L^k\}$. Thus, the results we establish in this section do not give explicit conditions for convergence of $\{L^k\}$; rather, the results relate the asymptotic behaviors of $\{L^k\}$ and of the forecasting sequence $\{H^k\}$. Roughly speaking, we show that $\{L^k\}$ converges if and only if $\{H^k\}$ converges weakly to some distribution. That is, the forecasts "stabilize" if and only if the protection levels do. Although intuitive, such a result is valid only under appropriate assumptions; indeed, we present an example in which the sequence of protection levels $\{L^k\}$ does not converge even though $\{H^k\}$ does. We will be interested in the following set of assumptions.

Assumption (C1). $\{\hat{H}^k\}$ is a good forecasting method for $\{X^k\}$.
Assumption (C2). The distribution function $G(\ell, \cdot)$ is continuous in ℓ in the topology corresponding to weak convergence.

Observe that if the protection level ℓ takes on only integer values, then Assumption (C2) automatically holds, because then $G(\ell', \cdot) = G(\ell, \cdot)$ for all ℓ' and ℓ such that $|\ell' - \ell| < 1$.

The following are the main results of this section. Note that Proposition 16 applies to the deterministic example discussed in Section 3. For $b \in \mathbb{R}$ and $A \subset \mathbb{R}$, let $d(b,A) := \inf\{|b-a| : a \in A\}$ denote the distance between the point b and the set A.

Proposition 15. *Consider the stochastic process described by* (3)–(5)*, and suppose that Assumptions* (C1) and (C2) *hold. If there exists a random distribution function $\hat{H}$ such that $\mathbb{P}[\hat{H}^k \xrightarrow{w} \hat{H}] = 1$, then $\mathbb{P}[d(L^k, \hat{H}^{-1}(\gamma)) \to 0] = 1$. In addition, if $\mathbb{P}[\hat{H}^{-1}(\gamma)$ is a singleton$] = 1$, then there exists a random variable L such that $\mathbb{P}[L^k \to L] = 1$ and $\mathbb{P}[L = G^{-1}(L,\gamma)] = 1$.*

Proposition 16. *Consider the stochastic process described by* (3)–(5)*, and suppose that Assumptions* (C1) and (C2) *hold. If there exists a random variable L such that $\mathbb{P}[L^k \to L] = 1$, then $\mathbb{P}[\hat{H}^k \xrightarrow{w} G(L,\cdot)] = 1$, and $\mathbb{P}[L \in G^{-1}(L,\gamma)] = 1$.*

Example 7.1. The following example shows that, without additional assumptions, the sequence of protection levels $\{L^k\}$ may not converge. Let $0 \le a < b$. Let $G(\ell,x) = \mathbb{I}_{\{a+b-\ell \le x\}}$. Specifically, $G(a,x) = \mathbb{I}_{\{b \le x\}}$, and $G(b,x) = \mathbb{I}_{\{a \le x\}}$. Note that Assumption (C2) holds: if $\ell^k \to \ell$, then $G(\ell^k,x) \to G(\ell,x)$ for all continuity points x of $G(\ell,\cdot)$, that is, for all $x \ne \ell$. Also note that $G(\ell,\cdot)$ is stochastically decreasing in ℓ, instead of stochastically increasing in ℓ as would have been more appealing for the application.

Suppose that we use the empirical distribution (7) as forecast distribution for X. In case $(\hat{H}^k)^{-1}(\gamma)$ is not a singleton, choose $L^k = \min\{x : \hat{H}^k(x) \ge \gamma\} = \min\{x \in (\hat{H}^k)^{-1}(\gamma)\}$. Let $\gamma = 1/2$. Let $L^0 = a$, or equivalently, let $\hat{H}^0(x) = \mathbb{I}_{\{a \le x\}}/2 + \mathbb{I}_{\{b \le x\}}/2$. Then it is easy to show by induction on k that

$$\begin{aligned} L^k \;=\; X^k &= \begin{cases} a \text{ if } k \text{ even} \\ b \text{ if } k \text{ odd} \end{cases} \\ G(L^k,x) &= \begin{cases} G(a,x) = \mathbb{I}_{\{b \le x\}} \text{ if } k \text{ even} \\ G(b,x) = \mathbb{I}_{\{a \le x\}} \text{ if } k \text{ odd} \end{cases} \\ \hat{H}^k(x) &= \begin{cases} \frac{1}{2}\mathbb{I}_{\{a \le x\}} + \frac{1}{2}\mathbb{I}_{\{b \le x\}} & \text{if } k \text{ even} \\ \frac{k-1}{2k}\mathbb{I}_{\{a \le x\}} + \frac{k+1}{2k}\mathbb{I}_{\{b \le x\}} & \text{if } k \text{ odd.} \end{cases} \end{aligned}$$

Let $\hat{H}(x) := \mathbb{I}_{\{a \le x\}}/2 + \mathbb{I}_{\{b \le x\}}/2 = \hat{H}^0(x)$. Then $\|\hat{H}^k - \hat{H}\|_\infty \to 0$ as $k \to \infty$, but the sequence $\{L^k\}$ does not converge. Of course, we have to violate some assumptions of Proposition 15 here. Specifically, we violate the assumption that $\mathbb{P}[\hat{H}^{-1}(\gamma)$ is a singleton$] = 1$.

8. Conclusions

In this paper, we introduced a framework for analyzing the dynamics of forecasting and optimization in revenue management. We considered a model that has been studied widely in the revenue management literature and that has been used widely in revenue management practice, combined with a number of forecasting methods that have been proposed for revenue management. We gave conditions under which the spiral-down phenomenon occurs.

8.1. What to Do about the Spiral-Down Effect

The results in this paper suggest a number of interesting and important research questions. Before discussing some of these questions, we reiterate the motivation for our work. Most papers on revenue management specify a model, suppose that the model is correct, and then propose a method to obtain decisions based on the model. While this approach has led to some useful results, so far it has ignored the question of what happens when (a) the assumptions of the models do not hold, and (b) the models are repeatedly updated and used. The questionable nature of some of the assumptions of many widely used revenue management models provides additional emphasis to the importance of this question. The results in this paper illustrate how an error in such a model can lead to a systematic deterioration of the controls if the model is updated and used repeatedly. Such systematic deterioration is different in nature, and potentially of greater concern, than the suboptimality of solutions obtained if a model with error is used only once. Hence, the main contribution of this paper is to address an important question that has long been ignored.

A question that naturally follows from the results in this paper is how to avoid the spiral-down effect. In general, this is a difficult question, because more accurate, but still not entirely correct, models are not guaranteed to have better dynamic behavior than less accurate models. In addition, merely preventing the controls or the objective from spiraling down may not give satisfactory results. As an example, consider a model that produces the same control irrespective of the observed data. Such a model does not suffer from the spiral-down effect; however, a model that makes no use of observed data may be undesirable. In a comparison of models, one model may make better use of observed data and produce better controls when its assumptions are satisfied, but may be worse in terms of dynamic behavior when its assumptions are not satisfied. Nevertheless, it seems to be a good starting point to develop more accurate models, and then to study their dynamic behavior. In what follows, we make a few brief comments about attempts to accomplish these goals.

Brumelle et al. (1990), Bodily and Weatherford (1995), and Belobaba and Weatherford (1996) consider models in which a customer who requests a ticket in a particular lower fare class may be willing to purchase a ticket in a higher fare class if no more tickets in the lower fare class are available. These models assume that there is exogenous demand for each fare class, while allowing that some customers may buy tickets in a different class if their preferred class is not available. In general, the assumption that there is such a thing as exogenous demand for a particular fare class is becoming more difficult to justify as customers are able to easily obtain information about large sets of alternatives and their attributes. (In addition, this also raises the need to include competition in models.) Hence, it may be desirable to move toward formulations that do not rely on the notion of exogenous demand for a fare class, such as those in which each customer chooses among the presented set of alternatives with choice probabilities that depend on that set (Talluri and van Ryzin 2004a). Nevertheless, for the problem considered in this paper, the models referred to above may have better dynamic behavior than the model associated with Littlewood's rule. A more detailed investigation of the dynamics of these models is beyond the scope of this paper.

Taking into account that models necessarily have error, in this paper we illustrate why it may be important to understand the dynamic behavior of models with errors. In general, it seems prudent for practitioners to study their models, including the associated forecasting methods, in situations where the assumptions of the model do not hold. If the study reveals that the model is not robust to violations of the assumptions, then the practitioner may want to consider a different model or forecasting method.

Another way to potentially avoid phenomena such as spiral down is to follow a model-free approach and to directly track how the control (the protection levels) affects the quantity of interest (revenue). For example, one may attempt to use a response surface approach to maximize the expected revenue as a function of the controls, without using the notion of demand. For such an approach to work, the airline would need to use different values of the controls to estimate how the revenue responds as a function of the controls, at least in a neighborhood of control values. Such an approach has shortcomings, including the need to use potentially bad values of the controls to estimate the response of the expected revenue, and the need to obtain many observations, especially when the control is high dimensional, to accurately approximate the response surface.

8.2. Future Directions

There is a clear need for more work on how to avoid or mitigate the spiral-down effect. In addition, there remain some open technical questions related to the analysis of the spiral down effect. For example, we have considered forecasting techniques for which, w.p.1, the forecast distributions converge weakly, and hence the associated quantiles also converge. Although this notion is less restrictive than related ones studied in the literature (see the discussion in Section 4), some forecasting methods used in revenue management practice, such as the empirical distribution, have this property, whereas some others, such as exponential smoothing, do not. Nevertheless, some of the forecasting methods that do not have this property have other desirable characteristics. For example, methods such as exponential smoothing use smoothing constants or weights for the new observations of X that remain bounded away from zero, as opposed to the forecasting methods considered in this paper that use weights of $1/k$ for the new observations X^k. Such weights that remain large prevent these methods from having the property above, but at the same time these weights allow the forecasts to adjust faster to possible changes in the underlying random processes, such as seasonality, trends, "shifts" in demand, and user interventions. Thus it seems desirable to develop a theory, possibly based on other modes of convergence of the forecast distributions (or at least of their quantiles), that would include other forecasting methods used in practice.

The results in this paper describe, under various assumptions on the relationship between the conditional distributions $G(\ell,\cdot)$ of the observed quantity X and the protection level ℓ, conditions that result in the spiral-down effect when certain forecasting systems are used. In each case we have relied on arguments tailored to the specific forecasting system under consideration to establish convergence of the protection levels $\{L^k\}$, and in some cases, to identify the limits. It would be of interest to find more general, easily checkable conditions on the behavior of $G(\ell,\cdot)$ as a function of ℓ and the forecasting update functions ϕ^k in (4) that would allow one to show when spiral down does (or does not) occur. Our results in Section 7 indicate that the existence of an ℓ that satisfies the fixed-point condition $G^{-1}(\ell,\gamma)=\ell$ will likely be important in such a theory. Other conditions that appear to be potentially useful include: (a) $f_\gamma(\ell) := G^{-1}(\ell,\gamma)$ is a contraction mapping; (b) $f_\gamma(\ell)\le \ell$; (c) $f_\gamma(\ell)-\ell$ is monotone; (d) the family $\{G(\ell,\cdot)\}$ is stochastically increasing in ℓ; (e) $X^k \le L^{k-1}$ implies that $L^k \le L^{k-1}$. The particular cases that we have covered in this paper have satisfied some of these conditions. Additional work is required to establish such general, easily checkable conditions that characterize spiral-down behavior.

Appendix. Auxiliary Results and Proofs

PROPOSITION 17. *Let B denote the Borel σ-algebra on $\mathbb{R}$. Consider the space $(\mathcal{P}(\mathbb{R}),\mathcal{B})$ of probability distributions on $\mathbb{R}$, endowed with the Borel σ-algebra $\mathcal{B}$ corresponding to the topology of weak convergence on $\mathcal{P}(\mathbb{R})$. Consider a measurable space $(\Omega,\mathcal{F})$. Let $\{H^k:\Omega\mapsto\mathcal{P}(\mathbb{R})\}$ be a sequence of $(\mathcal{F},\mathcal{B})$-measurable functions.*

(i) Consider a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a filtration $\{\mathcal{F}^k\}$. Consider a random sequence $\{Y^k\}$ adapted to filtration $\{\mathcal{F}^k\}$, where $Y^k:\Omega\mapsto\mathbb{R}$. Let $F^k:\Omega\mapsto\mathcal{P}(\mathbb{R})$ be given by $F^k(\omega,x) := \mathbb{P}[Y^{k+1}\le x\,|\,\mathcal{F}^k]$, that is, F^k is the conditional distribution of Y^{k+1}. Then F^k is $(\mathcal{F}^k,\mathcal{B})$-measurable.

(ii) The set $\Omega^* := \{\omega\in\Omega : H^k(\omega,\cdot)$ converges weakly as $k\to\infty\}$ is in $\mathcal{F}$.

(iii) Let $\Omega^* := \{\omega\in\Omega : H^k(\omega,\cdot)$ converges weakly as $k\to\infty\}$, and let $\mathcal{F}^* := \{A\in\mathcal{F} : A\subset\Omega^*\}$. For each $\omega\in\Omega^*$, let $H^*(\omega,\cdot)$ denote the weak limit of $\{H^k(\omega,\cdot)\}$. Then $\mathcal{F}^*$ is a σ-algebra on Ω^*. In addition, H^* is $(\mathcal{F}^*,\mathcal{B})$-measurable, and thus H^* is also $(\mathcal{F},\mathcal{B})$-measurable.

(iv) For any $(\mathcal{F},\mathcal{B})$-measurable $F:\Omega\mapsto\mathcal{P}(\mathbb{R})$, the set $\left\{\omega\in\Omega : H^k(\omega,\cdot)\xrightarrow{w}F(\omega,\cdot)\right\}$ is in $\mathcal{F}$.

(v) Let $F:\Omega\mapsto\mathcal{P}(\mathbb{R})$ be an $(\mathcal{F},\mathcal{B})$-measurable function. For any $x\in\mathbb{R}$, let $f_x:\Omega\mapsto\mathbb{R}$ be defined as $f_x(\omega) := F(\omega,x)$. Then, f_x is $(\mathcal{F},B)$-measurable. That is, f_x is a real-valued random variable.

The proof of Proposition 17 is given in the Online Appendix.

PROPOSITION 18. *Suppose that $\{Y^k:\Omega\mapsto\mathbb{R}\}$ are $(\mathcal{F},B)$-measurable random variables. Then $\hat{H}^k$ defined in (12) is $(\mathcal{F},\mathcal{B})$-measurable for all k.*

The proof of Proposition 18 is given in the Online Appendix.

LEMMA 3. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the space $(\mathcal{P}(\mathbb{R}), \mathcal{B})$ of probability distributions on $\mathbb{R}$ endowed with the Borel σ-algebra $\mathcal{B}$ corresponding to the topology of weak convergence on $\mathcal{P}(\mathbb{R})$. Let $F : \Omega \mapsto \mathcal{P}(\mathbb{R})\}$ be a $(\mathcal{F}, \mathcal{B})$-measurable function. For each $\omega \in \Omega$, let $D(\omega) := \{x \in \mathbb{R} : F(\omega, x) > F(\omega, x-)\}$ denote the set of jump points of $F(\omega, \cdot)$. Then the set $\{x \in \mathbb{R} : \mathbb{P}[x \in D(\omega)] > 0\}$ is countable.*

The proof of Lemma 3 is given in the Online Appendix.

Proof of Proposition 3. Proposition 18 establishes that $\hat{H}^k$ is $(\mathcal{F}, \mathcal{B})$-measurable for all k, i.e., $\hat{H}^k$ is a random distribution function.

Let $\{F^k\}$, Ω^*, and F^* be as in Definition 1. For any $\omega \in \Omega$ and any $x \in \mathbb{R}$, let

$$S^n(\omega, x) \;\; := \;\; \sum_{k=1}^{n} \left(\mathbb{I}_{\{Y^k(\omega) \le x\}} - F^{k-1}(\omega, x) \right).$$

Note that $\{S^n(\omega, x)\}$ is a martingale with respect to $\{\mathcal{F}^n\}$, since $\mathbb{E}[S^n(\omega, x) \,|\, \mathcal{F}^{n-1}] = \mathbb{E}[\mathbb{I}_{\{Y^n(\omega) \le x\}} - F^{n-1}(\omega, x) \,|\, \mathcal{F}^{n-1}] + \mathbb{E}[S^{n-1}(\omega, x) \,|\, \mathcal{F}^{n-1}] = \mathbb{P}[Y^n(\omega) \le x \,|\, \mathcal{F}^{n-1}] - F^{n-1}(\omega, x) + S^{n-1}(\omega, x) = S^{n-1}(\omega, x)$. In addition,

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}[(\mathbb{I}_{\{Y^k(\omega) \le x\}} - F^{k-1}(\omega, x))^2]}{k^2} \;\; \le \;\; \sum_{k=1}^{\infty} \frac{1}{k^2} \;\; < \;\; \infty$$

It follows from a strong law of large numbers for martingales (Chow 1967) that $\lim_{n \to \infty} S^n(\omega, x)/n = 0$ w.p.1, that is, there is a set $\Omega'(x) \subset \Omega$ such that $\mathbb{P}[\Omega'(x)] = 0$ and $\lim_{n \to \infty} S^n(\omega, x)/n = 0$ for all $\omega \in \Omega \setminus \Omega'(x)$. Since $\hat{H}^n(\omega, x) := (1/n) \sum_{k=1}^{n} \mathbb{I}_{\{Y^k(\omega) \le x\}}$, it follows that $\hat{H}^n(\omega, x) - (1/n) \sum_{k=1}^{n} F^{k-1}(\omega, x) \to 0$ for all $\omega \in \Omega \setminus \Omega'(x)$.

For each $\omega \in \Omega^*$, let $D(\omega) := \{x \in \mathbb{R} : F^*(\omega, x) > F^*(\omega, x-)\}$ denote the set of jump points of $F^*(\omega, \cdot)$. It is shown in Lemma 3 that the set $\{x \in \mathbb{R} : \mathbb{P}[x \in D(\omega)] > 0\}$ is countable. Let S be a countable dense set in $\mathbb{R}$ such that $\mathbb{P}[x \in D(\omega)] = 0$ for all $x \in S$. Let $\Omega''(x) := \{\omega \in \Omega^* : x \in D(\omega)\}$. Then, for any $x \in S$, x is a continuity point of $F^*(\omega, \cdot)$ for all $\omega \in \Omega^* \setminus \Omega''(x)$, and $\mathbb{P}[\Omega''(x)] = 0$. Then $F^{k-1}(\omega, x) \to F^*(\omega, x)$, and thus $\frac{1}{n} \sum_{k=1}^{n} F^{k-1}(\omega, x) \to F^*(\omega, x)$ for all $\omega \in \Omega^* \setminus \Omega''(x)$.

Let $\Omega' := \cup_{x \in S} \Omega'(x) \cup \Omega''(x)$. Then $\mathbb{P}[\Omega'] = 0$ and $\hat{H}^n(\omega, x) \to F^*(\omega, x)$ for all $x \in S$ and all $\omega \in \Omega^* \setminus \Omega'$. This implies that $\hat{H}^n(\omega, \cdot) \xrightarrow{w} F^*(\omega, \cdot)$ for all $\omega \in \Omega^* \setminus \Omega'$ (Fristedt and Gray 1997, p.245, Proposition 2). □

Proof of Proposition 5. It follows from (26) and (27) that

$$f^k \;\; = \;\; \prod_{i=m}^{k} \frac{i}{i - 1 + \beta} \qquad \text{and} \qquad g^k \;\; = \;\; \sum_{j=m}^{k} \frac{\alpha}{j} f^j.$$

Recall from (28) that $\mathbb{E}\left[f^k L^k - g^k \mid \mathcal{F}^{k-1}\right] = f^{k-1} L^{k-1} - g^{k-1}$. By Assumption (A) we have $\mathbb{E}|f^k L^k - g^k| \le \sup_{j \ge m} \mathbb{E}|f^j L^j - g^j| < \infty$. Hence, $\{f^k L^k - g^k\}$ is a martingale.

First consider the case with $\beta < 1$ as in part 1 of the proposition. Consider f^k. For sufficiently large values of i, $i/(i - 1 + \beta) \in (1, \infty)$, and $i/(i - 1 + \beta) \downarrow 1$, i.e., $(i - 1 + \beta)/i \in (0, 1)$ and $(i - 1 + \beta)/i \uparrow 1$, as $i \to \infty$. First, we want to determine $\lim_{k \to \infty} f^k$. Recall the following result: Consider a sequence $\{a_i\}_{i=m}^{\infty}$, with $a_i \in (0, 1)$ for all i. Then $\prod_{i=m}^{\infty} a_i = 0$ if and only if $\sum_{i=m}^{\infty} (1 - a_i) = \infty$. Let $a_i := (i - 1 + \beta)/i$. Then $\sum_{i=m}^{\infty} (1 - a_i) = \sum_{i=m}^{\infty} (1 - \beta)/i = \infty$. Thus $1/f^k = \prod_{i=m}^{k} a_i \to 0$, and hence $f^k \to \infty$ as $k \to \infty$.

Next consider $g^k = \sum_{j=m}^{k} f^j \alpha / j$. If $\alpha = 0$, then $g^k = 0$ for all k. By Assumption (A) we have $\sup_k \mathbb{E}|f^k L^k - g^k| = \sup_k \mathbb{E}|f^k L^k| < \infty$, and consequently $\{f^k L^k\}$ is an L^1-bounded martingale. Therefore, the martingale convergence theorem (see, e.g., Theorem 9.4.4 of Chung 1974) implies that w.p.1, $f^k L^k \to A$ as $k \to \infty$, where A is a finite random variable. Then it follows from the observation that $f^k \to \infty$ as $k \to \infty$ that w.p.1, $L^k \to 0$ as $k \to \infty$.

Suppose now that $\beta = 1$ as in parts 2 and 3 of the proposition. In this case $f^k = 1$ for all k. Also, $g^k = \sum_{j=1}^k \alpha/j$, and thus if $\alpha < 0$, then $g^k \to -\infty$; and if $\alpha > 0$, then $g^k \to \infty$ as $k \to \infty$; if $\alpha = 0$, then $g^k = 0$ for all k. Observe that $\sup_k \mathbb{E}|f^k L^k - g^k| = \sup_k \mathbb{E}|L^k - g^k| < \infty$ by Assumption (A). Hence, we can again apply the martingale convergence theorem to conclude that w.p.1, $L^k - g^k \to A$ as $k \to \infty$, where A is a finite random variable. Thus, if $\alpha < 0$, then $L^k \to -\infty$; if $\alpha > 0$, then $L^k \to \infty$; and if $\alpha = 0$, then $L^k \to A$ as $k \to \infty$. $\square$

Proof of Lemma 1. First, we show by induction that

$$X^{k+1} \sim N\left(L^0 + \alpha \sum_{i=1}^{k} \frac{1}{i},\, \sigma^2 \left(1 + \sum_{i=1}^{k} \frac{1}{i^2}\right)\right) \tag{40}$$

$$M^{k+1} \sim N\left(L^0 + \alpha \sum_{i=2}^{k+1} \frac{1}{i},\, \sigma^2 \sum_{i=1}^{k+1} \frac{1}{i^2}\right) \tag{41}$$

$$L^{k+1} \sim N\left(L^0 + \alpha \sum_{i=1}^{k+1} \frac{1}{i},\, \sigma^2 \sum_{i=1}^{k+1} \frac{1}{i^2}\right) \tag{42}$$

for all $k = 0, 1, \ldots$. It follows from (16)–(19) that $X^1 \sim N(L^0,\, \sigma^2)$, $M^1 = X^1 \sim N(L^0,\, \sigma^2)$, and $L^1 = M^1 + \alpha \sim N(L^0 + \alpha,\, \sigma^2)$, so (40)–(42) holds for $k = 0$. Suppose that the inductive hypothesis holds for $1, \ldots, k-1$.

The conditional distribution of X^{k+1}, given $\mathcal{F}^k$, is $N(L^k, \sigma^2)$. Using the inductive hypothesis, it follows from a result often used in Bayesian statistics that the (unconditional) distribution of X^{k+1} is $N(L^0 + \alpha \sum_{i=1}^k 1/i, \sigma^2(1 + \sum_{i=1}^k 1/i^2))$, so (40) holds for k. Moreover, $M^{k+1} = \sum_{i=1}^{k+1} X^i/(k+1)$ is normally distributed with mean

$$\mathbb{E}[M^{k+1}] \;=\; \mathbb{E}[\mathbb{E}[M^{k+1} \,|\, \mathcal{F}^k]] \;=\; \mathbb{E}\left[M^k + \frac{\alpha}{k+1}\right] \;=\; L^0 + \alpha \sum_{i=2}^{k+1} \frac{1}{i}$$

and variance

$$\begin{aligned}
\mathrm{Var}[M^{k+1}] &= \mathrm{Var}[\mathbb{E}[M^{k+1} \,|\, \mathcal{F}^k]] + \mathbb{E}[\mathrm{Var}[M^{k+1} \,|\, \mathcal{F}^k]] \\
&= \mathrm{Var}\left[M^k + \frac{\alpha}{k+1}\right] + \mathbb{E}\left[\mathrm{Var}\left[\frac{k}{k+1} M^k + \frac{1}{k+1} X^{k+1} \,\middle|\, \mathcal{F}^k\right]\right] \\
&= \mathrm{Var}[M^k] + \mathbb{E}\left[\mathrm{Var}\left[\frac{1}{k+1} X^{k+1} \,\middle|\, \mathcal{F}^k\right]\right] \\
&= \sigma^2 \sum_{i=1}^{k} \frac{1}{i^2} + \sigma^2 \frac{1}{(k+1)^2},
\end{aligned}$$

which gives (41). Finally, since $L^{k+1} = M^{k+1} + \alpha$, (42) follows.

Next, notice that since $\beta = 1$ in this case, it follows that $f^k = 1$, so in order to show that Assumption (A) holds we must show that $\sup_k \mathbb{E}|L^k - g^k| < \infty$. It follows from (42) that $\mathbb{E}[(L^k - g^k)^2] = (\mathbb{E}[L^k - g^k])^2 + \mathrm{Var}[L^k - g^k] = (L^0)^2 + \sigma^2 \sum_{i=1}^k 1/i^2$ and thus $\sup_k \mathbb{E}[(L^k - g^k)^2] = (L^0)^2 + \sigma^2 \sum_{i=1}^\infty 1/i^2 < \infty$. It follows from the Cauchy-Schwartz inequality that $\mathbb{E}|L^k - g^k| \le \mathbb{E}[(L^k - g^k)^2]^{1/2}$, and hence $\sup_k \mathbb{E}|L^k - g^k| < \infty$ as stated. $\square$

Proof of Lemma 2. It is easy to see from (25) that

$$\mathbb{E}[L^k] = L^0 \prod_{i=1}^{k} \frac{i - 1 + \beta}{i} \;=\; \frac{L^0}{f^k}$$

for all k, thus $\mathbb{E}[f^k L^k] = L^0$, and hence $\sup_k \mathbb{E}|f^k L^k| = L^0 < \infty$. $\square$

Proof of Proposition 10. Recall that the marked point process that describes customer arrival times and customer types is independent of the chosen value of the protection level ℓ. Thus we can compare what happens along each sample path of the point process with different choices of ℓ.

Consider any sample path, and let N denote the total number of arrivals for that sample path. Consider protection levels ℓ and $\ell+1$. Let the corresponding values of the observed quantity up to and including arrival n be denoted by $X(\ell,n)$ and $X(\ell+1,n)$ respectively, $n=0,\ldots,N$. Note that, for a given sample path, $X(\ell,N)$ denotes the final observed quantity X with protection level ℓ. We show that along any sample path, the observed quantity $X(\ell+1,N)$ exceeds the observed quantity $X(\ell,N)$ by at most 1. Then it follows that

$$F(\ell+1) \;=\; G(\ell+1,\ell+1) \;=\; \mathbb{P}[X(\ell+1,N)\le \ell+1] \;\ge\; \mathbb{P}[X(\ell,N)\le \ell] \;=\; F(\ell)$$

that is, F is nondecreasing as claimed.

Case 2.2.A: Untruncated class-1 demand: Recall that in this case the observed quantity is the number of customers who would purchase a class-1 ticket if a class-1 ticket were available. Thus $X(\ell,n)$ denotes the number of customers up to and including arrival n who would purchase a class-1 ticket if a class-1 ticket were available upon arrival. Clearly, if $N\ge c-\ell$ and arrival number $c-\ell$ is type ab, then $X(\ell+1,N)=X(\ell,N)+1$. Otherwise, $X(\ell+1,N)=X(\ell,N)$. Thus, for any sample path such that $X(\ell,N)\le \ell$, it holds that $X(\ell+1,N)\le \ell+1$, and the result follows.

Case 2.2.B: Truncated class-1 demand: Here the observed quantity is the number of class-1 tickets sold. Thus $X(\ell,n)$ denotes the number of class-1 tickets sold up to and including arrival n. Let $Y(\ell,n)$ be the total number of tickets sold up to and including arrival n. As in the previous case, if $N<c-\ell$ then $X(\ell+1,N)=X(\ell,N)$. Otherwise, $N\ge c-\ell$ and the following cases hold:

- If arrival number $c-\ell$ is type a, then $X(\ell+1,n)=X(\ell,n)$ for all n.
- If arrival number $c-\ell$ is type b, then $Y(\ell,n)=Y(\ell+1,n)+1$ for all $n\ge c-\ell$ until capacity is reached with protection level ℓ. Thus, we have the following cases:

  —If capacity c is not reached with protection level ℓ, then capacity also is not reached with protection level $\ell+1$, and $X(\ell+1,N)=X(\ell,N)$.

  —If capacity c is reached with protection level ℓ, then at that time there still is a remaining space with protection level $\ell+1$.

  ∗ If that space is filled (it can only be filled with a class-1 ticket), then $X(\ell+1,N)=X(\ell,N)+1$.

  ∗ Otherwise, $X(\ell+1,N)=X(\ell,N)$.
- If arrival number $c-\ell$ is type ab, then $Y(\ell,n)=Y(\ell+1,n)$ for all n, and $X(\ell+1,n)=X(\ell,n)+1$ for all $n\ge c-\ell$, thus $X(\ell+1,N)=X(\ell,N)+1$.

As before, for any sample path such that $X(\ell,N)\le \ell$, it also holds that $X(\ell+1,N)\le \ell+1$, and the result follows. □

Proof of Proposition 11. First consider the untruncated case, i.e. the setting of Example 2.2.A. As before, let $D_{ab}(\ell)$ be the number of customers arriving until $c-\ell$ tickets are sold. Then, for any $\ell\ge 0$ we have that $D_{ab}(\ell)=\min\{D_{ab},[c-\ell]^+\}$ and thus the observed quantity X is given by $D_{ab}-D_{ab}(\ell)=[D_{ab}-[c-\ell]^+]^+$. Hence, the distribution of X (for $x\ge 0$) is

$$\begin{aligned}
G(\ell,x) &= \mathbb{P}([D_{ab}-[c-\ell]^+]^+\le x)\\
&= \mathbb{P}\left([D_{ab}-[c-\ell]^+]^+\le x,\ D_{ab}\le [c-\ell]^+\right)+\mathbb{P}\left([D_{ab}-[c-\ell]^+]^+\le x,\ D_{ab}>[c-\ell]^+\right)\\
&= \mathbb{P}(D_{ab}\le [c-\ell]^+)+\mathbb{P}\left([c-\ell]^+<D_{ab}\le x+[c-\ell]^+\right)\\
&= \mathbb{P}\left(D_{ab}\le x+[c-\ell]^+\right).
\end{aligned} \tag{43}$$

It follows from (43) that, if $\ell\in[0,c]$, then $G(\ell,\ell)=\mathbb{P}(D_{ab}\le c)$, which is a constant. For $\ell>c$, $G(\ell,\ell)=\mathbb{P}(D_{ab}\le \ell)$, which goes to 1 as $\ell\to\infty$. Finally, it is clear that $G(\ell,\ell)=0$ for $\ell<0$. It follows that if $\mathbb{P}(D_{ab}\le c)>\gamma$ then Assumption (B1) holds for $\ell^*=0$, whereas if $\mathbb{P}(D_{ab}\le c)<\gamma$

and there is no $\ell > c$ such that $\mathbb{P}(D_{ab} \le \ell) = \gamma$ then Assumption (B1) holds for some $\ell^* > c$. The result then follows from Proposition 9. If $\mathbb{P}(D_{ab} \le c) = \gamma$ then Assumption (B1) does not hold, so we cannot determine the limiting behavior of the sequence $\{L^k\}$.

Consider now the truncated case, i.e. the setting of Example 2.2.B. Now the observed quantity X is given by $\min\{D_{ab} - D_{ab}(\ell), c, \ell\}$, i.e., $X \le \ell$ w.p.1 and hence $G(\ell,\ell) = 1$ for all $\ell \ge 0$. Since $G(\ell,\ell) = 0$ for all $\ell < 0$ and $\gamma < 1$, we see that Assumption (B1) holds for $\ell^* = 0$. Again, the result follows from Proposition 9. □

Proof of Proposition 15. It follows from Lemma 4 below that $\mathbb{P}[d(L^k, \hat{H}^{-1}(\gamma)) \to 0] = 1$. If $\mathbb{P}[\hat{H}^{-1}(\gamma)$ is a singleton$] = 1$, then let $L := \hat{H}^{-1}(\gamma)$, and it follows that $\mathbb{P}[L^k \to L] = 1$. By assumption (C2), $\{\omega : L^k(\omega) \to L(\omega)\} \subset \{\omega : G(L^k(\omega), \cdot) \xrightarrow{w} G(L(\omega), \cdot)\}$, and thus it follows from $\mathbb{P}[L^k \to L] = 1$ that $\mathbb{P}[G(L^k, \cdot) \xrightarrow{w} G(L, \cdot)] = 1$. Then, since $G(L^k, \cdot)$ is the conditional distribution of X^{k+1} given $\mathcal{F}^k$, it follows from assumption (C1) that $\mathbb{P}[\hat{H}^k \xrightarrow{w} G(L, \cdot)] = 1$. Therefore, $\mathbb{P}[\hat{H} = G(L, \cdot)] = 1$, and hence $L = G^{-1}(L, \gamma)$, w.p.1. □

Lemma 4. *Consider a sequence of distribution functions $\{F^k\} \subset \mathcal{P}(\mathbb{R})$ such that $F^k \xrightarrow{w} F \in \mathcal{P}(\mathbb{R})$. For $\gamma \in (0,1)$, let $[q^k, Q^k] := (F^k)^{-1}(\gamma)$, that is, $[q^k, Q^k]$ denotes the set of γ-quantiles of F^k [cf. (2)], and let $[q, Q] := F^{-1}(\gamma)$. Then, $q \le \liminf_{k\to\infty} q^k \le \limsup_{k\to\infty} Q^k \le Q$. That is, for any sequence $\{\xi^k\}$ of γ-quantiles of F^k, $d(\xi^k, F^{-1}(\gamma)) \to 0$ as $k \to \infty$.*

The proof of Lemma 4 is given in the Online Appendix.

Proof of Proposition 16. $\mathbb{P}(L^k \to L) = 1$ and assumption (C2) imply that $\mathbb{P}(G(L^k, \cdot) \xrightarrow{w} G(L, \cdot)) = 1$. This, combined with assumption (C1), imply that $\mathbb{P}(\hat{H}^k \xrightarrow{w} G(L, \cdot)) = 1$.

For the second part, note that $L^k \in (\hat{H}^k)^{-1}(\gamma)$ if and only if $\hat{H}^k(L^k) \ge \gamma$ and $\hat{H}^k(x) \le \gamma$ for all $x < L^k$, and similarly $L \in G^{-1}(L, \gamma)$ if and only if $G(L, L) \ge \gamma$ and $G(L, x) \le \gamma$ for all $x < L$. First we show by contradiction that, w.p.1, $G(L, L) \ge \gamma$. Suppose that $G(L, L) < \gamma$. It follows from $G(L, \cdot)$ being right-continuous that there exists a $\delta > 0$ such that $G(L, x) < \gamma$ for all $x < L + \delta$. $G(L, \cdot)$ must have a continuity point $y \in (L, L + \delta)$. Note that, w.p.1, for all sufficiently large k, $L^k < y$. Thus $\gamma \le \hat{H}^k(L^k) \le \hat{H}^k(y) \to G(L, y) < \gamma$, whereupon a contradiction is reached. Hence, w.p.1, $G(L, L) \ge \gamma$.

Next we show by contradiction that, w.p.1, $G(L, x) \le \gamma$ for all $x < L$. Suppose that there exists an $x < L$ such that $G(L, x) > \gamma$. $G(L, \cdot)$ must have a continuity point $y \in (x, L)$. Note that $G(L, y) \ge G(L, x) > \gamma$. Also, w.p.1, $\hat{H}^k(y) \to G(L, y)$, and thus, for all sufficiently large k, $\hat{H}^k(y) > \gamma$ and $L^k > y$. But that contradicts $L^k \in (\hat{H}^k)^{-1}(\gamma)$. Hence, w.p.1, $G(L, x) \le \gamma$ for all $x < L$, and therefore $L \in G^{-1}(L, \gamma)$. □

Acknowledgments
William Cooper and Tito Homem-de-Mello were supported by the National Science Foundation under grant DMI-0115385. Anton Kleywegt was supported by the National Science Foundation under grant DMI-9875400.

References

Balakrishnan, A., M. S. Pangburn, E. Stavrulaki. 2004. "Stack them high, let 'em fly": Lot-sizing policies when inventories stimulate demand. *Management Sci.* **50**(5) 630–644.

Belobaba, P. P., L. R. Weatherford. 1996. Comparing decision rules that incorporate customer diversion in perishable asset revenue management situations. *Decision Sciences* **27**(2) 343–363.

Belobaba, P. P. 1989. Application of a probabilistic decision model to airline seat inventory control. *Oper. Res.* **37**(2) 183–197.

Benveniste, A., M. Métivier, P. Priouret. 1990. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, Berlin, Germany.

Bertsekas, D. P. 2000. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA.

Bertsekas, D. P., J. N. Tsitsiklis. 1996. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA.

Billingsley, P. 1968. *Convergence of Probability Measures*. John Wiley & Sons, New York.

Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* **5**(3) 202–229.

Blackwell, D., L. Dubins. 1962. Merging of opinions with increasing information. *Ann. Math. Statist.* **33**(3) 882–886.

Bodily, S. E., L. R. Weatherford. 1995. Perishable-asset revenue management: Generic and multi-price yield management with diversion. *Omega* **23**(2) 173–185.

Boyd, E. A., I. C. Bilegan. 2003. Revenue management and E-commerce. *Management Sci.* **49**(10) 1363–1386.

Boyd, E. A., E. Kambour, J. Tama. 2001. The impact of buy down on sell up, unconstraining, and spiral down. Presented at *1st Annual INFORMS Revenue Management Section Conference*, New York. Slides available at `http://209.238.188.14/html_files/roi/roi_agenda.htm`.

Brumelle, S. L., J. I. McGill. 1993. Airline seat allocation with multiple nested fare classes. *Oper. Res.* **41**(1) 127–137.

Brumelle, S. L., J. I. McGill, T. H. Oum, K. Sawaki, M. W. Tretheway. 1990. Allocation of airline seats between stochastically dependent demands. *Transportation Sci.* **24**(3) 183–192.

Burnetas, A. N., C. E. Smith. 2000. Adaptive ordering and pricing for perishable products. *Oper. Res.* **48**(3) 436–443.

Cachon, G. P., A. G. Kok. 2003. Heuristic equilibrium and the estimation of the salvage value in the newsvendor model with clearance pricing. Working paper, Univ. of Pennsylvania.

Carvalho, A. X., M. L. Puterman. 2003. Dynamic pricing and learning over short time horizons. Working paper, Univ. of British Columbia.

Chow, Y. S. 1967. On a strong law of large numbers for martingales. *Ann. Math. Statist.* **38**(2) 610.

Chung, K. L. 1974. *A Course in Probability Theory*. 2nd ed. Academic Press, New York.

Curry, R. E. 1990. Optimal airline seat allocation with fare classes nested by origins and destinations. *Transportation Sci.* **24**(3) 193–204.

Dawid, A. P. 1982. The well-calibrated Bayesian. *J. Amer. Statist. Association* **77**(379) 605–613.

Fristedt, B., L. Gray. 1997. *A Modern Approach to Probability Theory*. Birkhäuser, Boston.

Fudenberg, D., D. K. Levine. 1998. *The Theory of Learning in Games*. MIT Press, Cambridge, MA.

Kalai, E., E. Lehrer, R. Smorodinsky. 1999. Calibrated forecasting and merging. *Games and Econom. Behavior* **29**(1-2) 151–169.

Kamae, T., U. Krengel, G. L. O'Brien. 1977. Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5**(6) 899–912.

Kreps, D. M. 1990. *Game Theory and Economic Modelling*. Oxford Univ. Press, Oxford, UK.

Kuhlmann, R. 2004. Why is revenue management not working? *J. of Revenue and Pricing Management* **2**(4) 378–387.

Kumar, P. R., P. Varaiya. 1986. *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice-Hall, Englewood Cliffs, N.J.

Littlewood, K. 1972. Forecasting and control of passenger bookings. *AGIFORS Sympos. Proc.*, vol. 12. 95–117.

Müller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester, UK.

Shlifer, E., Y. Vardi. 1975. An airline overbooking policy. *Transportation Sci.* **9**(2) 101–114.

Talluri, K., G. van Ryzin. 2004a. Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* **50**(1) 15–33.

Talluri, K., G. van Ryzin. 2004b. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Dordrecht, Netherlands.

van Ryzin, G., J. McGill. 2000. Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Sci.* **46**(6) 760–775.

Wollmer, R. D. 1992. An airline seat management model for a single leg route when lower fare classes book first. *Oper. Res.* **40**(1) 26–37.

Online Appendix

OA–1. Proofs for the Deterministic Example

PROPOSITION 1 *Suppose that the probability distribution of the observed quantity is given by (6) with $d < c$, and that forecasts are made according to (7). Then $L^{k+1} \le L^k$ for all k. Furthermore, there exists a k^* such that $L^j = 0$ and $X^j = 0$ for all $j \ge k^*$.*

Proof. Note that

$$X^{k+1} \;=\; [d-(c-L^k)^+]^+ \;\le\; [d-(c-L^k)]^+ \;\le\; L^k. \tag{OA–1}$$

In view of (7), we see that $\hat{H}^{k+1}(x) > \hat{H}^k(x)$ for all $x \ge X^{k+1}$ such that $\hat{H}^k(x) < 1$, and $\hat{H}^{k+1}(x) = 1$ for all $x \ge X^{k+1}$ such that $\hat{H}^k(x) = 1$. Therefore, $L^{k+1} \le L^k$ by (5), so the first part of the proposition is proved.

Let $\varepsilon := c - d > 0$. Notice that if $L^j \ge \varepsilon$ then $X^{j+1} \le L^j - \varepsilon$ by (OA–1). Moreover, (OA–1) also implies that if $0 \le L^j < \varepsilon$, then $X^{j+1} = 0$. Since we have already shown that the sequence of protection levels is non-increasing, it follows that if k is such that $L^k \ge \varepsilon$, then $X^{j+1} \le L^k - \varepsilon$ for all $j \ge k$.

Define

$$k' \;:=\; \min\left\{ j > k : \frac{k}{j}\hat{H}^k(L^k-\varepsilon) + \frac{j-k}{j} > \gamma \right\}. \tag{OA–2}$$

Observe that $k' < \infty$, because $\gamma < 1$. By (OA–2), we have that $\hat{H}^{k'}(L^k - \varepsilon) > \gamma$. Therefore, if $x \in (\hat{H}^{k'})^{-1}(\gamma)$ then $x \le L^k - \varepsilon$. Since $L^{k'} \in (\hat{H}^{k'})^{-1}(\gamma)$, it follows that $L^{k'} \le L^k - \varepsilon$.

Suppose now that $0 \le L^k < \varepsilon$. Then, (OA–1) implies that $X^{k+1} = 0$. An argument similar to that used above shows that there exists a $k^* > k$ such that $L^{k^*} = 0$. Since the sequence of protection levels is non-increasing, the second part of the proposition follows. □

PROPOSITION 2 *Suppose that the probability distribution of the observed quantity is given by (6) with $d > c$, and that forecasts are made according to (7). Suppose that $L^0 \in [0, c]$. Then $L^{k+1} \ge L^k$ for all k. Furthermore, there exists a k° such that $L^j = d$ and $X^j = d$ for all $j \ge k^\circ$.*

Proof. For the first part of the proposition, suppose that $L^k \in [0, c]$. Note that

$$X^{k+1} \;=\; d-(c-L^k) \;=\; L^k + \varepsilon. \tag{OA–3}$$

In view of (7), we see that $\hat{H}^{k+1}(x) \le \hat{H}^k(x)$ for all $x < X^{k+1}$; in addition, $\hat{H}^{k+1}(x) < \hat{H}^k(x)$ for all $x < X^{k+1}$ such that $\hat{H}^k(x) > 0$. Therefore, $L^{k+1} \ge L^k$ by (5).

Recall that $\hat{H}^k(X^{k+1}-) := \lim_{x \uparrow X^{k+1}} \hat{H}^k(x)$ denotes the left limit of $\hat{H}^k$ at X^{k+1}. Consider any integer $j > k\hat{H}^k(X^{k+1}-)/\gamma$. Then one of two cases must hold: either there is an integer $k' \le j$ such that $L^{k'} > c$, or $L^i \in [0, c]$ for all $i \le j$. In the latter case, choose $k' = j$, and note that $\hat{H}^j(X^{k+1}-) = k\hat{H}^k(X^{k+1}-)/j < \gamma$, and thus $L^j := (\hat{H}^j)^{-1}(\gamma) \ge X^{k+1} = L^k + \varepsilon$. In summary, k' is such that $L^{k'} > c$ or $L^{k'} \ge L^k + \varepsilon$.

Next, note that if $L^k > c$, then $X^{k+1} = d$. An argument similar to that used above shows that there exists a $k^\circ \ge k$ such that $L^{k^\circ} = d$. Note that at the first time k' such that $L^{k'} > c$, it still holds that $L^{k'} \le d$, because $X^k \le d$ and thus $\hat{H}^k(d) = 1$ for all k, and hence $L^k \le L^{k+1}$ also when $L^k > c$. For the same reason, given that $L^{k^\circ} = d$ then $L^k = d$ for all $k \ge k^\circ$, which is the second assertion of the proposition. □

OA–2. Proof of Proposition 17

LEMMA OA–1. *Consider the metric space $(\mathcal{P}(\mathbb{R}), \lambda)$ of probability distributions on $\mathbb{R}$ endowed with the Lévy metric λ, defined as follows for $F, H \in \mathcal{P}(\mathbb{R})$:*

$$\lambda(F,H) \quad := \quad \inf\{\varepsilon > 0 : F(x-\varepsilon) - \varepsilon \le H(x) \le F(x+\varepsilon) + \varepsilon \ \forall\, x \in \mathbb{R}\}.$$

Let $\mathbb{N}$ denote the natural numbers, and let $\mathbb{Q}$ denote the rational numbers. Then for any $F, H \in \mathcal{P}(\mathbb{R})$ and any $r > 0$, $\lambda(F,H) < r$ if and only if there exists $m \in \mathbb{N}$ such that

$$F(x - r + 1/m) - r + 1/m \quad < \quad H(x) \quad < \quad F(x + r - 1/m) + r - 1/m$$

for all $x \in \mathbb{Q}$.

Proof. First, suppose that $\lambda(F,H) < r$. Then there exists $m \in \mathbb{N}$ such that $\lambda(F,H) < r - 1/m$, and it follows from F being nondecreasing that $F(x - r + 1/m) - r + 1/m < H(x) < F(x + r - 1/m) + r - 1/m$ for all $x \in \mathbb{R}$, and thus for all $x \in \mathbb{Q}$.

Next, suppose that there exists an $m \in \mathbb{N}$ such that $F(x - r + 1/m) - r + 1/m < H(x) < F(x + r - 1/m) + r - 1/m$ for all $x \in \mathbb{Q}$. Consider any $x \in \mathbb{R}$, and a sequence $\{x^n\} \subset \mathbb{Q}$ such that $x^n \downarrow x$. Then $F(x^n - r + 1/m) - r + 1/m < H(x^n) < F(x^n + r - 1/m) + r - 1/m$ for all n. It follows from the right continuity of F and H that $F(x - r + 1/m) - r + 1/m \le H(x) \le F(x + r - 1/m) + r - 1/m$. Hence $\lambda(F,H) := \inf\{\varepsilon > 0 : F(x-\varepsilon) - \varepsilon \le H(x) \le F(x+\varepsilon) + \varepsilon \ \forall\, x \in \mathbb{R}\} \le r - 1/m < r$. □

PROPOSITION 17 *Let B denote the Borel σ-algebra on $\mathbb{R}$. Consider the space $(\mathcal{P}(\mathbb{R}), \mathcal{B})$ of probability distributions on $\mathbb{R}$, endowed with the Borel σ-algebra $\mathcal{B}$ corresponding to the topology of weak convergence on $\mathcal{P}(\mathbb{R})$. Consider a measurable space $(\Omega, \mathcal{F})$. Let $\{H^k : \Omega \mapsto \mathcal{P}(\mathbb{R})\}$ be a sequence of $(\mathcal{F}, \mathcal{B})$-measurable functions.*

(i) Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\{\mathcal{F}^k\}$. Consider a random sequence $\{Y^k\}$ adapted to filtration $\{\mathcal{F}^k\}$, where $Y^k : \Omega \mapsto \mathbb{R}$. Let $F^k : \Omega \mapsto \mathcal{P}(\mathbb{R})$ be given by $F^k(\omega, x) := \mathbb{P}[Y^{k+1} \le x \,|\, \mathcal{F}^k]$, that is, F^k is the conditional distribution of Y^{k+1}. Then F^k is $(\mathcal{F}^k, \mathcal{B})$-measurable.

(ii) The set $\Omega^* := \{\omega \in \Omega : H^k(\omega, \cdot)$ converges weakly as $k \to \infty\}$ is in $\mathcal{F}$.

(iii) Let $\Omega^* := \{\omega \in \Omega : H^k(\omega, \cdot)$ converges weakly as $k \to \infty\}$, and let $\mathcal{F}^* := \{A \in \mathcal{F} : A \subset \Omega^*\}$. For each $\omega \in \Omega^*$, let $H^*(\omega, \cdot)$ denote the weak limit of $\{H^k(\omega, \cdot)\}$. Then $\mathcal{F}^*$ is a σ-algebra on Ω^*. In addition, H^* is $(\mathcal{F}^*, \mathcal{B})$-measurable, and thus H^* is also $(\mathcal{F}, \mathcal{B})$-measurable.

(iv) For any $(\mathcal{F}, \mathcal{B})$-measurable $F : \Omega \mapsto \mathcal{P}(\mathbb{R})$, the set $\left\{\omega \in \Omega : H^k(\omega, \cdot) \xrightarrow{w} F(\omega, \cdot)\right\}$ is in $\mathcal{F}$.

(v) Let $F : \Omega \mapsto \mathcal{P}(\mathbb{R})$ be an $(\mathcal{F}, \mathcal{B})$-measurable function. For any $x \in \mathbb{R}$, let $f_x : \Omega \mapsto \mathbb{R}$ be defined as $f_x(\omega) := F(\omega, x)$. Then, f_x is $(\mathcal{F}, B)$-measurable. That is, f_x is a real-valued random variable.

Proof. (i) Fix k. For each $x \in \mathbb{R}$, define the function $\pi_x : \mathcal{P}(\mathbb{R}) \mapsto \mathbb{R}$ by $\pi_x(F) := F(x)$. Consider $\pi_x \circ F^k : \Omega \mapsto \mathbb{R}$. Note that $\pi_x(F^k(\omega, \cdot)) = F^k(\omega, x) := \mathbb{P}[Y^{k+1} \le x \,|\, \mathcal{F}^k]$, and thus $\pi_x \circ F^k$ is $(\mathcal{F}^k, B)$-measurable.

Convergence in the Lévy metric λ, defined in Lemma OA–1, is equivalent to weak convergence of elements of $\mathcal{P}(\mathbb{R})$. Moreover, the space $\mathcal{P}(\mathbb{R})$, endowed with the Lévy metric λ, is complete and separable. For any $F \in \mathcal{P}(\mathbb{R})$ and $r > 0$, let $B(F, r) := \{H \in \mathcal{P}(\mathbb{R}) : \lambda(F, H) < r\}$ denote the ball with center F and radius r in $(\mathcal{P}(\mathbb{R}), \lambda)$. Since $(\mathcal{P}(\mathbb{R}), \lambda)$ is separable, its Borel sigma algebra $\mathcal{B}$ is generated by the countable collection of open balls $\{B(F, 1/m) : F \in D, m \in \mathbb{N}\}$, where D is a countable, dense subset of $\mathcal{P}(\mathbb{R})$. Therefore, to prove that F^k is $(\mathcal{F}^k, \mathcal{B})$-measurable, it suffices to show that $(F^k)^{-1}(B(F, r)) \in \mathcal{F}^k$ for all $F \in \mathcal{P}(\mathbb{R})$ and $r > 0$.

Consider any $F \in \mathcal{P}(\mathbb{R})$ and $r > 0$. For any $m \in \mathbb{N}$ and $x \in \mathbb{R}$, let $A_{m,x} := (F(x - r + 1/m) - r + 1/m, F(x + r - 1/m) + r - 1/m)$. It follows from Lemma OA–1 that $B(F, r) = \cup_{m \in \mathbb{N}} \cap_{x \in \mathbb{Q}} \pi_x^{-1}(A_{m,x})$.

Thus, for any $B(F,r)$,

$$\begin{aligned}(F^k)^{-1}\big(B(F,r)\big) &= (F^k)^{-1}\left(\bigcup_{m\in\mathbb{N}}\bigcap_{x\in\mathbb{Q}}\pi_x^{-1}(A_{m,x})\right)\\ &= \bigcup_{m\in\mathbb{N}}\bigcap_{x\in\mathbb{Q}}(F^k)^{-1}\big(\pi_x^{-1}(A_{m,x})\big)\\ &= \bigcup_{m\in\mathbb{N}}\bigcap_{x\in\mathbb{Q}}(\pi_x\circ F^k)^{-1}(A_{m,x})\end{aligned}$$

Recall that $\pi_x \circ F^k$ is $(\mathcal{F}^k, B)$-measurable. Thus $(\pi_x \circ F^k)^{-1}(A_{m,x}) \in \mathcal{F}^k$, and hence $(F^k)^{-1}\big(B(F,r)\big) \in \mathcal{F}^k$.

(ii) Completeness of $(\mathcal{P}(\mathbb{R}),\lambda)$ implies that

$$\big\{\omega\in\Omega : \lim_{k\to\infty} F^k(\omega,\cdot) \text{ exists}\big\} \quad = \quad \big\{\omega\in\Omega : \{F^k(\omega,\cdot)\} \text{ is Cauchy}\big\}. \tag{OA–4}$$

The event on the right above can be expressed as

$$\bigcap_{m\geq 1}\bigcup_{n\geq 1}\bigcap_{\{i\,:\,i\geq n\}}\bigcap_{\{j\,:\,j\geq n\}}\big\{\omega\in\Omega : \lambda(F^i(\omega,\cdot),F^j(\omega,\cdot))<1/m\big\} \tag{OA–5}$$

Separability of $(\mathcal{P}(\mathbb{R}),\lambda)$ implies that the mappings $\Lambda^{ij} : \Omega \mapsto \mathbb{R}$ defined by $\Lambda^{ij}(\omega) := \lambda(F^i(\omega,\cdot),F^j(\omega,\cdot))$ are all $(\mathcal{F},B)$-measurable (Billingsley 1968, p.25). Hence $\big\{\omega \in \Omega : \lambda(F^i(\omega,\cdot),F^j(\omega,\cdot)) < 1/m\big\} \in \mathcal{F}$ for all i,j,m, and therefore the set in (OA–5) is in $\mathcal{F}$.

(iii) It is easy to verify that $\mathcal{F}^*$ is a σ-algebra on Ω^*. It follows from Dudley (2002), Theorem 4.2.2, that F^* is $(\mathcal{F}^*,\mathcal{B})$-measurable. It follows immediately that F^* is also $(\mathcal{F},\mathcal{B})$-measurable.

(iv) As before, separability of $(\mathcal{P}(\mathbb{R}),\lambda)$ implies that the mappings $\Lambda^k : \Omega \mapsto \mathbb{R}$ defined by $\Lambda^k(\omega) := \lambda(F^k(\omega,\cdot),F(\omega,\cdot))$ are all $(\mathcal{F},B)$-measurable, and therefore so is $\limsup_{k\to\infty}\Lambda^k$. Since

$$\Big\{\omega\in\Omega : F^k(\omega,\cdot)\xrightarrow{w}F(\omega,\cdot)\Big\} \quad = \quad \Big\{\omega\in\Omega : \limsup_{k\to\infty}\Lambda^k(\omega)=0\Big\},$$

and since the set on the right is in $\mathcal{F}$, it follows that the set on the left is in $\mathcal{F}$ as well.

(v) We follow closely an argument in Billingsley (1968), p.121. For each $x \in \mathbb{R}$, define the function $\pi_x : \mathcal{P}(\mathbb{R}) \mapsto \mathbb{R}$ by $\pi_x(F) := F(x)$. For each $\varepsilon > 0$, define the function $\pi_x^\varepsilon : \mathcal{P}(\mathbb{R}) \mapsto \mathbb{R}$ by $\pi_x^\varepsilon(F) := \varepsilon^{-1}\int_x^{x+\varepsilon} F(u)du$. We first show that π_x^ε is continuous on $\mathcal{P}(\mathbb{R})$ and thus measurable. Consider any sequence $\{H^k\} \subset \mathcal{P}(\mathbb{R})$ such that $H^k \xrightarrow{w} H$. It follows from a characterization of weak convergence of distribution functions on $\mathbb{R}$ that $H^k(u) \to H(u)$ for all u except on a countable set (the set of discontinuities of H). Since $H^k(u) \leq 1$ for all u, it follows by the bounded convergence theorem that $\int_x^{x+\varepsilon} H^k(u)du \to \int_x^{x+\varepsilon} H(u)du$, and thus $\pi_x^\varepsilon(H^k) \to \pi_x^\varepsilon(H)$. Hence, π_x^ε is continuous and thus measurable. Next, since H is right-continuous, it follows that $\pi_x(H) = \lim_{\varepsilon\downarrow 0}\pi_x^\varepsilon(H) = \lim_{m\to\infty}\pi_x^{1/m}(H)$. Thus π_x is the limit of a sequence of measurable functions and therefore is $(\mathcal{B},B)$-measurable. It follows from the definition of π_x that $f_x(\omega) = \pi_x(F(\omega,\cdot))$. Therefore, f_x is $(\mathcal{F},B)$-measurable. □

OA–3. Supporting Material for Proposition 3

Below, we use some notation from Section OA–2: λ is the Lévy metric on $\mathcal{P}(\mathbb{R})$, B is the Borel σ-algebra on $\mathbb{R}$, and $B(h,r)$ is the ball of radius r about $h \in \mathcal{P}(\mathbb{R})$.

LEMMA OA–2. *Let $\psi : \mathbb{R} \mapsto \mathcal{P}(\mathbb{R})$ be given by*

$$\psi(y) \quad := \quad \mathbb{I}_{\{\,\cdot\,\geq y\}}\,. \tag{OA–6}$$

Then the mapping ψ is $(B,\mathcal{B})$-measurable.

Proof. For any $h \in \mathcal{P}(\mathbb{R})$ and $y \in \mathbb{R}$ we have

$$\begin{aligned}
\lambda(h,\psi(y)) &= \inf\left\{\varepsilon>0 : \begin{array}{l} h(x-\varepsilon)-\varepsilon \le 0 \le h(x+\varepsilon)+\varepsilon \text{ for all } x : x<y \text{ and} \\ h(x-\varepsilon)-\varepsilon \le 1 \le h(x+\varepsilon)+\varepsilon \text{ for all } x : x \ge y \end{array}\right\} \\
&= \inf\left\{\varepsilon>0 : \begin{array}{l} h(x-\varepsilon) \le \varepsilon \text{ for all } x : x<y \text{ and} \\ h(x+\varepsilon) \ge 1-\varepsilon \text{ for all } x : x \ge y \end{array}\right\} \\
&= \inf\{\varepsilon>0 : \lim_{x\uparrow y} h(x-\varepsilon) \le \varepsilon \text{ and } h(y+\varepsilon) \ge 1-\varepsilon\} \qquad \text{(OA–7)}
\end{aligned}$$

The space $(\mathcal{P}(\mathbb{R}),\lambda)$ is separable with countable base given by $\{B(h,r) : h \in D, r \in \mathbb{Q}\}$, where D is a countable dense subset of $\mathcal{P}(\mathbb{R})$. Hence, to show the $(B,\mathcal{B})$-measurability of ψ, it suffices to show that $\psi^{-1}(B(h,r)) \in B$ for all h and r.

To this end, for $h \in \mathcal{P}(\mathbb{R})$ and $\varepsilon > 0$, define $\psi^1_{h,\varepsilon}, \psi^2_{h,\varepsilon}, \psi_{h,\varepsilon} : \mathbb{R} \mapsto \mathbb{R}$ by

$$\psi^1_{h,\varepsilon}(y) := \varepsilon - \lim_{x\uparrow y} h(x-\varepsilon), \qquad \text{(OA–8)}$$

$$\psi^2_{h,\varepsilon}(y) := h(y+\varepsilon) - 1 + \varepsilon, \qquad \text{(OA–9)}$$

$$\psi_{h,\varepsilon}(y) := \min\{\psi^1_{h,\varepsilon}(y), \psi^2_{h,\varepsilon}(y)\} \qquad \text{(OA–10)}$$

The functions in (OA–8)–(OA–10) above are (B,B)-measurable because $h \in \mathcal{P}(\mathbb{R})$. Moreover, by (OA–7) and (OA–8)–(OA–10), we have

$$\begin{aligned}
\psi^{-1}(B(h,r)) &= \{y \in \mathbb{R} : \lambda(h,\psi(y)) < r\} \\
&= \left\{y \in \mathbb{R} : \inf\{\varepsilon>0 : \lim_{x\uparrow y} h(x-\varepsilon) \le \varepsilon \text{ and } h(y+\varepsilon) \ge 1-\varepsilon\} < r\right\} \\
&= \{y \in \mathbb{R} : \inf\{\varepsilon>0 : \psi_{h,\varepsilon}(y) \ge 0\} < r\} \\
&= \bigcup_{n : n^{-1} < r} \{y \in \mathbb{R} : \psi_{h,r-1/n}(y) \ge 0\}.
\end{aligned}$$

In view of the measurability of $\psi_{h,\varepsilon}(\cdot)$, all the sets in the union in the final expression are in B, and hence the proof is complete. □

LEMMA OA–3. *Suppose that $H_1, H_2 : \Omega \mapsto \mathcal{P}(\mathbb{R})$ are both $(\mathcal{F},\mathcal{B})$-measurable mappings and that $\alpha \in [0,1]$. The mapping $\xi_{\alpha,H_1,H_2} : \Omega \mapsto \mathcal{P}(\mathbb{R})$ given by*

$$\xi_{\alpha,H_1,H_2}(\omega,x) := \alpha H_1(\omega,x) + (1-\alpha) H_2(\omega,x), \quad x \in \mathbb{R} \qquad \text{(OA–11)}$$

is $(\mathcal{F},\mathcal{B})$-measurable.

Proof. Note that ξ_{α,H_1,H_2} can be expressed as $\theta_\alpha \circ J_{H_1,H_2}$ where $J_{H_1,H_2} : \Omega \mapsto \mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R})$ is defined by $J_{H_1,H_2}(\omega) := (H_1(\omega), H_2(\omega))$, and $\theta_\alpha : \mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R}) \mapsto \mathcal{P}(\mathbb{R})$ is defined by

$$\theta_\alpha(h_1,h_2)(x) := \alpha h_1(x) + (1-\alpha) h_2(x), \quad x \in \mathbb{R}.$$

The mapping J_{H_1,H_1} is $(\mathcal{F},\mathcal{B}\times\mathcal{B})$-measurable, where $\mathcal{B}\times\mathcal{B}$ is defined as the σ-algebra generated by sets of the form $A_1 \times A_2$ with $A_1, A_2 \in \mathcal{B}$. So the lemma will be proved if we can show that θ_α is $(\mathcal{B}\times\mathcal{B},\mathcal{B})$-measurable.

For this, consider the metric space $\mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R})$ with metric λ^* given by

$$\lambda^*((h_1,h_2),(h'_1,h'_2)) := \max\{\lambda(h_1,h'_1), \lambda(h_2,h'_2)\};$$

see Billingsley (1968), p.225. From the definitions of λ, λ^*, and θ_α, it follows that $\lambda^*((h_1,h_2),(h'_1,h'_2)) \ge \lambda(\theta_\alpha(h_1,h_2), \theta_\alpha(h'_1,h'_2))$. Therefore, θ_α is continuous. That is, for any open (in the topology metrized by λ) set $O \subset \mathcal{P}(\mathbb{R})$, $\theta_\alpha^{-1}(O)$ is an open set in the topology metrized by λ^*. The Borel sigma algebra on $(\mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R}), \lambda^*)$ is precisely $\mathcal{B}\times\mathcal{B}$ (Billingsley 1968, p.225), so the open sets metrized by λ^* are in $\mathcal{B}\times\mathcal{B}$. Summarizing, the open sets metrized by λ generate $\mathcal{B}$, and the inverse image of any such open set under θ_α is in $\mathcal{B}\times\mathcal{B}$. Hence, θ_α is $(\mathcal{B}\times\mathcal{B},\mathcal{B})$-measurable, which completes the proof. □

PROPOSITION 18 *Suppose that $\{Y^k : \Omega \mapsto \mathbb{R}\}$ are $(\mathcal{F}, B)$-measurable random variables. Then $\hat{H}^k$ defined in (12) is $(\mathcal{F}, \mathcal{B})$-measurable for all k.*

Proof. The proof is by induction. Let ψ be as defined in (OA–6). Note that

$$\hat{H}^k(\omega,\cdot) \;=\; \frac{1}{k}\sum_{n=1}^{k}(\psi\circ Y^n)(\omega,\cdot) \;=\; \frac{1}{k}(\psi\circ Y^k)(\omega,\cdot)+\frac{k-1}{k}\,\hat{H}^{k-1}(\omega,\cdot)$$

Lemma OA–2 and the assumptions on Y^n imply that $\psi\circ Y^n$ is $(\mathcal{F}, \mathcal{B})$-measurable for each n. Hence we immediately see that $\hat{H}^1 := \psi\circ Y^1$ is $(\mathcal{F}, \mathcal{B})$-measurable. Suppose that the result holds for $k-1$. With $\alpha = 1/k$, $H_1 = \psi\circ Y^k$, and $H_2 = \hat{H}^{k-1}$, we see that

$$\hat{H}^k(\omega,\cdot) \;=\; \xi_{1/k\,,\,\psi\circ Y^k\,,\,\hat{H}^{k-1}}(\omega,\cdot), \tag{OA–12}$$

where ξ_{α,H_1,H_2} is defined in (OA–11). The desired result now follows from (OA–12), the induction hypothesis, and Lemma OA–3. □

LEMMA OA–4. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a collection $\{A_i : i\in I\}\subset\mathcal{F}$ of events, where I is a countable index set. Suppose that $\mathbb{P}[A_i]\geq\varepsilon>0$ for all $i\in I$, and that for any $n+1$ distinct indices $i_1,\ldots,i_{n+1}\in I$, it holds that $A_{i_1}\cap\cdots\cap A_{i_{n+1}}=\varnothing$. Then $|I|\leq n/\varepsilon$.*

Proof. Let $\{S_j : j\in J\}\subset 2^I$ denote the collection of all subsets of I such that $1\leq|S_j|\leq n$ for all $j\in J$. Note that J is countable. For all $i\in I$,

$$A_i \;=\; \bigcup_{\{j\in J\,:\,i\in S_j\}}\;\bigcap_{i'\in S_j}A_{i'}\bigcap_{i'\in S_j^c}A_{i'}^c$$

and the sets $\{\cap_{i'\in S_j}A_{i'}\cap_{i'\in S_j^c}A_{i'}^c : j\in J\}$ are disjoint. Thus,

$$\mathbb{P}[A_i] \;=\; \sum_{\{j\in J\,:\,i\in S_j\}}\mathbb{P}\left[\bigcap_{i'\in S_j}A_{i'}\bigcap_{i'\in S_j^c}A_{i'}^c\right] \;\geq\; \varepsilon \;>\; 0$$

Also,

$$\bigcup_{i\in I}A_i \;=\; \bigcup_{j\in J}\bigcap_{i\in S_j}A_i\bigcap_{i\in S_j^c}A_i^c$$

and, as before, the sets $\{\cap_{i\in S_j}A_i\cap_{i\in S_j^c}A_i^c : j\in J\}$ are disjoint. Thus,

$$\sum_{j\in J}\mathbb{P}\left[\bigcap_{i\in S_j}A_i\bigcap_{i\in S_j^c}A_i^c\right] \;=\; \mathbb{P}\left[\bigcup_{j\in J}\bigcap_{i\in S_j}A_i\bigcap_{i\in S_j^c}A_i^c\right] \;=\; \mathbb{P}\left[\bigcup_{i\in I}A_i\right] \;\leq\; 1$$

Also,

$$\begin{aligned}&\sum_{j\in J}\mathbb{P}\left[\bigcap_{i\in S_j}A_i\bigcap_{i\in S_j^c}A_i^c\right]\\ &\quad\geq\;\inf\left\{\sum_{j\in J}x_j\;:\;\sum_{\{j\in J\,:\,i\in S_j\}}x_j=\mathbb{P}[A_i]\;\forall\,i\in I,\;x_j\geq 0\;\forall\,j\in J\right\}\end{aligned}$$

$$
\begin{aligned}
&\geq \inf\left\{\sum_{j\in J} x_j \;:\; \sum_{\{j\in J\,:\,i\in S_j\}} x_j \geq \varepsilon \;\forall\, i\in I,\; x_j\geq 0 \;\forall\, j\in J\right\} \\
&= \inf\left\{\sup\left\{\sum_{j\in J} x_j + \sum_{i\in I} y_i\left(\varepsilon - \sum_{\{j\in J\,:\,i\in S_j\}} x_j\right) \;:\; y_i\geq 0\;\forall\, i\in I\right\} \;:\; x_j\geq 0\;\forall\, j\in J\right\} \\
&\geq \sup\left\{\inf\left\{\sum_{j\in J} x_j + \sum_{i\in I} y_i\left(\varepsilon - \sum_{\{j\in J\,:\,i\in S_j\}} x_j\right) \;:\; x_j\geq 0\;\forall\, j\in J\right\} \;:\; y_i\geq 0\;\forall\, i\in I\right\} \\
&= \sup\left\{\inf\left\{\sum_{i\in I} \varepsilon y_i + \sum_{j\in J} x_j\left(1 - \sum_{i\in S_j} y_i\right) \;:\; x_j\geq 0\;\forall\, j\in J\right\} \;:\; y_i\geq 0\;\forall\, i\in I\right\} \\
&= \sup\left\{\sum_{i\in I} \varepsilon y_i \;:\; \sum_{i\in S_j} y_i \leq 1\;\forall\, j\in J, y_i\geq 0\;\forall\, i\in I\right\} \\
&\geq |I|\varepsilon/n
\end{aligned}
$$

where the last inequality follows from the observation that $y_i = 1/n$ for all $i\in I$ satisfies $\sum_{i\in S_j} y_i \leq 1$ for all $j\in J$, because $|S_j|\leq n$ for all $j\in J$. Combining the results above, it follows that $|I|\varepsilon/n\leq 1$, and thus $|I|\leq n/\varepsilon$. $\square$

LEMMA 3 *Consider a probability space $(\Omega,\mathcal{F},\mathbb{P})$, and the space $(\mathcal{P}(\mathbb{R}),\mathcal{B})$ of probability distributions on $\mathbb{R}$ endowed with the Borel σ-algebra $\mathcal{B}$ corresponding to the topology of weak convergence on $\mathcal{P}(\mathbb{R})$. Let $F:\Omega\mapsto\mathcal{P}(\mathbb{R})\}$ be a $(\mathcal{F},\mathcal{B})$-measurable function. For each $\omega\in\Omega$, let $D(\omega):=\{x\in\mathbb{R} : F(\omega,x)>F(\omega,x-)\}$ denote the set of jump points of $F(\omega,\cdot)$. Then the set $\{x\in\mathbb{R} : \mathbb{P}[x\in D(\omega)]>0\}$ is countable.*

Proof. For each $n\in\mathbb{N}$ and $x\in\mathbb{R}$, let $\Omega_x^n := \{\omega\in\Omega : F(\omega,x)-F(\omega,x-)>1/(n+1)\}$. Then $\{\omega\in\Omega : x\in D(\omega)\} = \cup_{n\in\mathbb{N}}\Omega_x^n$. Thus $\mathbb{P}[x\in D(\omega)] = \mathbb{P}\left[\cup_{n\in\mathbb{N}}\Omega_x^n\right] \leq \sum_{n\in\mathbb{N}}\mathbb{P}\left[\Omega_x^n\right]$.

Consider any $n+1$ distinct points $x_1,\ldots,x_{n+1}\in\mathbb{R}$. Suppose that $\omega\in\cap_{i=1}^{n+1}\Omega_{x_i}^n$. Then $\sum_{i=1}^{n+1}[F(\omega,x_i)-F(\omega,x_i-)] > (n+1)/(n+1) = 1$. However, $\sum_{i=1}^{n+1}[F(\omega,x_i)-F(\omega,x_i-)] \leq \sum_{x\in D(\omega)}[F(\omega,x_i)-F(\omega,x_i-)]\leq 1$, and thus $\cap_{i=1}^{n+1}\Omega_{x_i}^n=\varnothing$.

For each $m,n\in\mathbb{N}$, let $D^{m,n} := \{x\in\mathbb{R} : \mathbb{P}\left[\Omega_x^n\right]\geq 1/m\}$. Then $\{x\in\mathbb{R} : \mathbb{P}[x\in D(\omega)]>0\} = \cup_{m,n\in\mathbb{N}}D^{m,n}$. We show by contradiction that each set $D^{m,n}$ is finite. Suppose that $D^{m,n}$ is infinite; if $D^{m,n}$ is uncountable, choose a countably infinite subset of $D^{m,n}$ and denote the subset with $D^{m,n}$ as well. Consider the countably infinite collection of events $\{\Omega_x^n : x\in D^{m,n}\}$. Recall that for any $n+1$ distinct points $x_1,\ldots,x_{n+1}\in\mathbb{R}$, $\cap_{i=1}^{n+1}\Omega_{x_i}^n=\varnothing$. Also recall that $\mathbb{P}\left[\Omega_x^n\right]\geq 1/m$ for all $x\in D^{m,n}$. Thus it follows from Lemma OA–4 that $|D^{m,n}|\leq mn$. Hence each set $D^{m,n}$ is finite, and therefore $\{x\in\mathbb{R} : \mathbb{P}[x\in D(\omega)]>0\} = \cup_{m,n\in\mathbb{N}}D^{m,n}$ is countable. $\square$

OA–4. Proof of Proposition 4

PROPOSITION 4 *Consider a family of distributions $\{H(m,\cdot) : m\in\mathbb{M}\subset\mathbb{R}\}$, where $m=\int xH(m,dx)$ is the mean of $H(m,\cdot)$, $\mathbb{M}$ is closed, and $H(m,\cdot)$ is continuous in m with respect to the topology of weak convergence. Suppose that $\{Y^k\}$ and $\{F^k\}$ as in Definition 1 satisfy $F^k(\omega,\cdot)=H(U^k(\omega),\cdot)$ w.p.1, where $U^k := \mathbb{E}[Y^{k+1}\,|\,\mathcal{F}^k]$. Also suppose that $\sup_{k\geq 0}\mathbb{E}[(Y^{k+1})^2\,|\,\mathcal{F}^k] < Z$ w.p.1, for some integrable random variable Z. Then $\{\hat{H}^k\}$ in (13)–(14) is a good forecasting method for $\{Y^k\}$.*

Proof. Note initially that, since H is continuous in the first argument and M^k is $(\mathcal{F},B)$-measurable, it follows that $\hat{H}^k$ is $(\mathcal{F},\mathcal{B})$-measurable for all k, i.e., $\hat{H}^k$ is a random distribution function.

Let

$$S^n \;:=\; \sum_{k=1}^{n}(Y^k - U^{k-1}).$$

Note that $\{S^n\}$ is a martingale with respect to $\{\mathcal{F}^n\}$, because $E|S^n| < \infty$ and $\mathbb{E}[S^n \,|\, \mathcal{F}^{n-1}] = \mathbb{E}[Y^n - U^{n-1} \,|\, \mathcal{F}^{n-1}] + \mathbb{E}[S^{n-1} \,|\, \mathcal{F}^{n-1}] = \mathbb{E}[Y^n \,|\, \mathcal{F}^{n-1}] - U^{n-1} + S^{n-1} = S^{n-1}$. In addition, $\mathbb{E}[(Y^k - U^{k-1})^2] = \mathbb{E}[(Y^k)^2] - \mathbb{E}[(U^{k-1})^2] \le \mathbb{E}[(Y^k)^2] \le \mathbb{E}[Z]$, and consequently

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}[(Y^k - U^{k-1})^2]}{k^2} \;\le\; \sum_{k=1}^{\infty} \frac{\mathbb{E}[Z]}{k^2} \;<\; \infty.$$

It follows from a strong law of large numbers for martingales (Chow 1967) that $\lim_{n\to\infty} S^n/n = 0$ w.p.1, that is, there is a set $\Omega'' \subset \Omega$ such that $\mathbb{P}[\Omega''] = 0$ and $\lim_{n\to\infty} S^n(\omega)/n = 0$ for all $\omega \in \Omega \setminus \Omega''$. Since $M^n := (1/n)\sum_{k=1}^{n} Y^k$, it follows that

$$M^n(\omega) - \frac{1}{n}\sum_{k=1}^{n} U^{k-1}(\omega) \;\to\; 0 \quad \text{for all } \omega \in \Omega \setminus \Omega''. \tag{OA–13}$$

Let $\Omega''' := \cup_{k\ge 1}\{\omega \in \Omega \,:\, F^k(\omega,\cdot) \ne H(U^k(\omega),\cdot)\} \cup \{\omega \in \Omega \,:\, \sup_{k\ge 0} \mathbb{E}[(Y^{k+1})^2 \,|\, \mathcal{F}^k](\omega) \ge Z(\omega)\}$, and observe that $\mathbb{P}[\Omega'''] = 0$. Then, for all $\omega \in \Omega^* \setminus \Omega'''$, $H(U^k(\omega),\cdot) = F^k(\omega,\cdot) \xrightarrow{w} F^*(\omega,\cdot)$. In addition, for such ω, $\sup_{k\ge 0}\int x^2 F^k(\omega,dx) < Z(\omega)$, and hence by Theorem 4.5.2 of Chung (1974),

$$U^k(\omega) \;=\; \int x H(U^k(\omega),dx) \;=\; \int x F^k(\omega,dx) \;\to\; \int x F^*(\omega,dx) \;=:\; U(\omega) \quad \text{for all } \omega \in \Omega^* \setminus \Omega'''. \tag{OA–14}$$

Therefore, for all $\omega \in \Omega^* \setminus \Omega'''$, it holds that $F^*(\omega,\cdot) = H(U(\omega),\cdot)$, because $H(m,\cdot)$ is continuous in m.

Let $\Omega' = \Omega'' \cup \Omega'''$, and observe that $\mathbb{P}[\Omega'] = 0$. Then $M^k(\omega) \to U(\omega)$ for all $\omega \in \Omega^* \setminus \Omega'$ by (OA–13) and (OA–14). Again using the continuity of $H(m,\cdot)$ in m, it follows that $\hat{H}^k(\omega,\cdot) := H(M^k(\omega),\cdot) \xrightarrow{w} H(U(\omega),\cdot) = F^*(\omega,\cdot)$ for all $\omega \in \Omega^* \setminus \Omega'$, which proves that $\{\hat{H}^k\}$ is a good forecasting method for $\{Y^k\}$. □

OA–5. Remark Regarding Proposition 5

We briefly explain the difficulties in obtaining results for cases not covered by the proposition.

In the $\beta < 1$ case, note that $f^j > 1$ for all j. Thus, if $\alpha > 0$, then $g^k > \sum_{j=m}^{k} \alpha/j$ and hence $g^k \to \infty$ as $k \to \infty$. If $\alpha < 0$, then $g^k < \sum_{j=m}^{k} \alpha/j$ and hence $g^k \to -\infty$ as $k \to \infty$. Thus, if $\alpha \ne 0$, then even if we use the martingale convergence theorem to establish that, w.p.1, $f^k L^k - g^k \to A$, where A is a finite random variable, it does not establish the asymptotic behavior of L^k.

Next consider the case with $\beta > 1$. Note that $i/(i-1+\beta) \in (0,1)$ for all i, so $f^k \in (0,1)$ for all k. Let $a_i := i/(i-1+\beta)$. Then $\sum_{i=1}^{\infty}(1-a_i) = \sum_{i=1}^{\infty}(\beta-1)/(i-1+\beta) = \infty$. Thus $f^k = \prod_{i=1}^{k} a_i \to 0$ as $k \to \infty$. Next, consider

$$\begin{aligned}
\log(f^k) &= \sum_{i=1}^{k} \log\left(\frac{i}{i-1+\beta}\right) \;=\; \sum_{i=1}^{k} \log\left(1 + \frac{1-\beta}{i-1+\beta}\right) \\
&\le \sum_{i=1}^{k} \frac{1-\beta}{i-1+\beta} \;\le\; -(\beta-1)\int_1^{k+1} \frac{1}{x-1+\beta}\,dx \\
&= -(\beta-1)\left[\log(k+\beta) - \log(\beta)\right] \\
&= \log\left[\left(\frac{\beta}{k+\beta}\right)^{\beta-1}\right].
\end{aligned}$$

It follows that $f^k \le \left(\frac{\beta}{k+\beta}\right)^{\beta-1}$ and hence

$$\frac{g^k}{\alpha} \;=\; \sum_{j=1}^{k}\frac{1}{j}f^j \;\le\; \sum_{j=1}^{k}\frac{1}{j}\left(\frac{\beta}{j+\beta}\right)^{\beta-1} \;\le\; \sum_{j=1}^{\infty}\frac{1}{j}\left(\frac{\beta}{j+\beta}\right)^{\beta-1} \;<\; \infty.$$

In addition $\{g^k\}$ is non-decreasing, and thus $g^k \to \bar{g}$ as $k \to \infty$, where $|\bar{g}| < \infty$. Therefore, if $\sup_k \mathbb{E}|f^k L^k - g^k| < \infty$, then w.p.1, $f^k L^k - g^k \to A$ as $k \to \infty$, where A is a finite random variable. Then $f^k L^k \to B$ as $k \to \infty$, where B is a finite random variable. Recall that $f^k \in (0,1)$ for all k, and $f^k \to 0$ as $k \to \infty$. Thus, if $B(\omega) < 0$, then $L^k(\omega) \to -\infty$; and if $B(\omega) > 0$, then $L^k(\omega) \to \infty$. However, if $B(\omega) = 0$, then we need more information to determine the asymptotic behavior of L^k.

OA–6. Proof of Lemma 4

Lemma 4 *Consider a sequence of distribution functions $\{F^k\} \subset \mathcal{P}(\mathbb{R})$ such that $F^k \xrightarrow{w} F \in \mathcal{P}(\mathbb{R})$. For $\gamma \in (0,1)$, let $[q^k, Q^k] := (F^k)^{-1}(\gamma)$, that is, $[q^k, Q^k]$ denotes the set of γ-quantiles of F^k [cf. (2)], and let $[q, Q] := F^{-1}(\gamma)$. Then, $q \le \liminf_{k\to\infty} q^k \le \limsup_{k\to\infty} Q^k \le Q$. That is, for any sequence $\{\xi^k\}$ of γ-quantiles of F^k, $d(\xi^k, F^{-1}(\gamma)) \to 0$ as $k \to \infty$.*

Proof. Consider any $q' < q$. We show that for all k sufficiently large, $q^k > q'$. Let $q^* \in (q', q)$ be a continuity point of F. Then $F(q^*) < \gamma$, and $F^k(q^*) \to F(q^*)$ as $k \to \infty$, and thus for all k sufficiently large, $F^k(q') \le F^k(q^*) < \gamma$. Hence $q' < q^k$ for all k sufficiently large, and thus $q \le \liminf_{k\to\infty} q^k$. It follows by a similar argument that $\limsup_{k\to\infty} Q^k \le Q$. □

OA–7. More on Stochastic Approximation

In this section we show that, under appropriate assumptions, if the distribution of the observed quantity depends on the protection level and if L^k is updated according to (36), then $G(L^k, L^k)$ converges to γ. It follows that if L^k converges then it converges to a random variable L^* that satisfies $\mathbb{P}(L^* \in G^{-1}(L^*, \gamma)) = 1$. In this section we assume that $G(\ell, x) = 0$ for all $x < 0$ and all $\ell \in \mathbb{R}$, and therefore $X^k \ge 0$ w.p.1.

The following result on the convergence of stochastic approximation iterations is given in Proposition 4.1 of Bertsekas and Tsitsiklis (1996).

Proposition OA–1. *Consider the random sequences $\{S^k\}_{k=1}^{\infty}$ and $\{L^k\}_{k=0}^{\infty}$ in $\mathbb{R}^n$ that satisfy $L^{k+1} = L^k + \xi_k S^{k+1}$, where $\{\xi_k\}_{k=0}^{\infty}$ is a deterministic nonnegative step size sequence that satisfies $\sum_{k=0}^{\infty} \xi_k = \infty$ and $\sum_{k=0}^{\infty} \xi_k^2 < \infty$. Let $\mathcal{F}^k$ denote the σ-algebra generated by $S^1, \ldots, S^k, L^0, \ldots, L^k$. Consider a function $V : \mathbb{R}^n \mapsto \mathbb{R}_+$ with the following properties:*

1. *∇V is Lipschitz continuous on $\mathbb{R}^n$.*
2. *There is a constant $c > 0$ such that, w.p.1,*

$$-\nabla V(L^k)^T \mathbb{E}[S^{k+1} \,|\, \mathcal{F}^k] \;\;\ge\;\; c\|\nabla V(L^k)\|^2$$

for all k.

3. *There exist constants $K_1, K_2 > 0$ such that, w.p.1,*

$$\mathbb{E}[\|S^{k+1}\|^2 \,|\, \mathcal{F}^k] \;\;\le\;\; K_1 + K_2\|\nabla V(L^k)\|^2$$

for all k.

Then the following hold w.p.1:

1. *$V(L^k)$ converges to a random variable V^* as $k \to \infty$.*
2. *$\nabla V(L^k) \to 0$ as $k \to \infty$.*

3. *Every limit point L^* of $\{L^k\}$ satisfies $\nabla V(L^*) = 0$.*

Next we construct a potential function V to study the convergence of (36). Note that by the assumptions we make on G in this section, we have that $F(\ell) = G(\ell, \ell) = 0$ if $\ell < 0$. We also make the following assumption:

Assumption (B2) The function F is Lipschitz continuous, i.e., there exists an $M > 0$ such that $|F(\ell_1) - F(\ell_2)| \leq M|\ell_1 - \ell_2|$ for all $\ell_1, \ell_2 \in \mathbb{R}$.

This essentially says that the rate of change of $G(\ell, \ell)$ with respect to ℓ is bounded for all ℓ. Assumption (B2) is satisfied, for instance, if

$$G(\ell, x) \;=\; 1 - e^{-x/m(\ell)}, \qquad x \geq 0, \tag{OA–15}$$

for $\ell \geq 0$, and $G(\ell, \cdot) = G(0, \cdot)$ for $\ell < 0$, i.e., negative protection levels have the same effect as $\ell = 0$. Here $m(\ell) > 0$ for all $\ell \geq 0$ and $r(\ell) := \ell/m(\ell)$ is Lipschitz continuous on $[0, \infty)$. Indeed, note that if $\ell_1, \ell_2 < 0$ then $|F(\ell_1) - F(\ell_2)| = 0$, and if $\ell_1 < 0 \leq \ell_2$ then $|F(\ell_1) - F(\ell_2)| = |F(0) - F(\ell_2)|$, so it suffices to check that F is Lipschitz continuous on $[0, \infty)$, which is indeed the case, because for $\ell_1, \ell_2 \geq 0$, we have $|F(\ell_1) - F(\ell_2)| = |e^{-r(\ell_2)} - e^{-r(\ell_1)}| \leq |r(\ell_2) - r(\ell_1)|$ since $r(\ell_1), r(\ell_2) \geq 0$.

One choice for $m(\ell)$ that satisfies the above conditions is

$$m(\ell) \;:=\; a_1 - a_2 e^{-a_3 \ell} \tag{OA–16}$$

where $a_1 > a_2 \geq 0$ and $a_3 \geq 0$. If the observed quantity X has distribution specified by (OA–15)–(OA–16), then it has properties that so-called "unconstrained demand" for high-price tickets could reasonably be expected to have (it is immaterial how this unconstraining is done — it only matters that it results in X). For instance, $m(\ell)$ increases in ℓ and approaches a constant as $\ell \to \infty$, which is an appealing property since one would not expect the mean demand to grow unboundedly with increasing protection levels.

To see that (OA–16) makes r Lipschitz continuous, note that

$$\begin{aligned} |r'(\ell)| = \; & \left| \frac{a_1 - a_2 e^{-a_3\ell} - \ell(a_2 a_3 e^{-a_3\ell})}{(a_1 - a_2 e^{-a_3\ell})^2} \right| \;\leq\; \left| \frac{1}{a_1 - a_2 e^{-a_3\ell}} \right| + \left| \frac{\ell(a_2 a_3 e^{-a_3\ell})}{(a_1 - a_2 e^{-a_3\ell})^2} \right| \\ \leq \; & \left| \frac{1}{a_1 - a_2} \right| + \left| \frac{\ell(a_2 a_3 e^{-a_3\ell})}{(a_1 - a_2)^2} \right| \;\leq\; \frac{1}{a_1 - a_2} + \frac{a_2 e^{-1}}{(a_1 - a_2)^2}. \end{aligned}$$

The final expression follows from the fact that $\ell e^{-a_3\ell}$ is maximized over $[0, \infty)$ at $\ell = 1/a_3$.

At this point we need the following assumption:

Assumption (B3) The quantity $\nu := \min_{\ell \in \mathbb{R}} \int_0^\ell [F(s) - \gamma]\, ds$ is finite.

When $\ell < 0$, we interpret the integral in the above expression for ν as $-\int_\ell^0$. Thus, for any $\ell < 0$, $\int_0^\ell [F(s) - \gamma] ds = -\int_\ell^0 [F(s) - \gamma] ds = -\int_\ell^0 [0 - \gamma] ds = -\ell\gamma > 0$. Hence, Assumption (B3) holds, for example, if there exists an $\ell_0 > 0$ such that $F(\ell) \geq \gamma$ for all $\ell \geq \ell_0$. For instance, this is the case when (OA–15)–(OA–16) specify the distribution of the observed quantity, since $F(\ell) \geq \gamma \Leftrightarrow \ln(1 - \gamma) \geq -r(\ell) \Leftrightarrow -m(\ell)\ln(1 - \gamma) \leq \ell$, which does indeed hold for ℓ sufficiently large. Under the assumptions of van Ryzin and McGill (2000), Assumptions (B2) and (B3) hold. Specifically, Assumption (B3) holds since it is always the case that $F(\ell) \geq \gamma$ for all ℓ large enough when G does not depend on ℓ.

Consider the function $V : \mathbb{R} \mapsto \mathbb{R}_+$ defined by

$$V(\ell) \;:=\; \int_0^\ell [F(s) - \gamma]\, ds \;-\; \nu. \tag{OA–17}$$

Next we verify that V satisfies the conditions in Proposition OA–1. Note that $V'(\ell) = F(\ell) - \gamma$.

1. V' is Lipschitz continuous, since by Assumption (B2) F is Lipschitz continuous.
2. Note from (36) that $S^{k+1} = \gamma - \mathbb{I}_{\{X^{k+1} \leq L^k\}}$. Thus

$$\mathbb{E}[S^{k+1} \,|\, \mathcal{F}^k] \;=\; \gamma - \mathbb{P}[X^{k+1} \leq L^k \,|\, L^k] \;=\; \gamma - G(L^k, L^k) \;=\; \gamma - F(L^k) \;=\; -V'(L^k).$$

3. Note that $S^{k+1} \in (-1, 1)$ w.p.1, and thus there exist constants $K_1, K_2 > 0$ such that

$$\mathbb{E}[(S^{k+1})^2 \,|\, \mathcal{F}^k] \;\leq\; K_1 + K_2[V'(L^k)]^2$$

Recall that the stepsizes ξ_k satisfy $\sum_k \xi_k = \infty$ and $\sum_k \xi_k^2 < \infty$, and thus we obtain the conclusions of Proposition OA–1. Specifically, we have the following.

Proposition OA–2. *Suppose that Assumptions* (B2) *and* (B3) *hold and that the protection levels are updated according to* (36)*. Then $G(L^k, L^k) \to \gamma$ w.p.1, and every limit point L^* of $\{L^k\}$ satisfies $G(L^*, L^*) = \gamma$, that is, $L^* \in G^{-1}(L^*, \gamma)$.*

Note that Propositions 8 and 9 require the existence of a deterministic quantity ℓ^* that satisfies assumption 3 in Proposition 8 or Assumption (B1) respectively, and that convergence of L^k to this deterministic quantity ℓ^* is then established. In contrast, Propositions OA–1 and OA–2 do not require the existence of such a deterministic quantity, and do not establish convergence of L^k.

OA–8. Proofs for Stochastic Comparisons and Pathwise Comparisons

Lemma OA–5. *For any two $\mathcal{P}(\mathbb{R})$-valued random elements $H_1 \sim P_1$ and $H_2 \sim P_2$, $H_1 \preceq_{\text{st}} H_2$ implies that $P_1[H_1(x) \geq \alpha] \geq P_2[H_2(x) \geq \alpha]$ for all $x, \alpha \in \mathbb{R}$.*

Proof. Fix any $x, \alpha \in \mathbb{R}$, and let $f : \mathcal{P}(\mathbb{R}) \mapsto \mathbb{R}$ be given by $f(h) := -\mathbb{I}_{\{h(x) \geq \alpha\}}$. Clearly f is bounded, and it follows from the characterization of $\leq_{\text{st}}$ that f is nondecreasing. Moreover, by the argument in the proof of Proposition 17(v) we have that f is measurable.

Consider any two $\mathcal{P}(\mathbb{R})$-valued random elements $H_1 \preceq_{\text{st}} H_2$. Then it follows that

$$P_1[H_1(x) \geq \alpha] \;=\; -\mathbb{E}_{P_1}[f(H_1)] \;\geq\; -\mathbb{E}_{P_2}[f(H_2)] \;=\; P_2[H_2(x) \geq \alpha].$$

□

To simplify the exposition below, suppose that L^k and $\underline{L}^k$ are chosen to be the smallest elements of the set of γ-quantiles of $\hat{H}^k$ and $\underline{\hat{H}}^k$ respectively, that is, $L^k \equiv \min\left\{x \in \mathbb{R} : \hat{H}^k(x) \geq \gamma\right\}$ and $\underline{L}^k \equiv \min\left\{x \in \mathbb{R} : \underline{\hat{H}}^k(x) \geq \gamma\right\}$.

Lemma OA–6. *Suppose that $\underline{G}(\underline{\ell}, \cdot) \leq_{\text{st}} G(\ell, \cdot)$ for all $\underline{\ell} \leq \ell$, and that the empirical distribution is used for both $\hat{H}$ and $\underline{\hat{H}}$, that is $\hat{H}^k(x) := k^{-1} \sum_{j=1}^k \mathbb{I}_{\{X^j \leq x\}}$ and $\underline{\hat{H}}^k(x) := k^{-1} \sum_{j=1}^k \mathbb{I}_{\{\underline{X}^j \leq x\}}$. If $\underline{\hat{H}}^k \preceq_{\text{st}} \hat{H}^k$, then*

$$\begin{aligned} \underline{L}^k &\leq_{\text{st}} L^k \\ \underline{G}(\underline{L}^k, \cdot) &\preceq_{\text{st}} G(L^k, \cdot) \\ \underline{X}^{k+1} &\leq_{\text{st}} X^{k+1} \\ \underline{\hat{H}}^{k+1} &\preceq_{\text{st}} \hat{H}^{k+1} \end{aligned}$$

Proof. Suppose $\{\underline{\hat{H}}^k, \underline{L}^k, \underline{X}^k\}$ is defined on probability space $(\underline{\Omega}, \underline{\mathcal{F}}, \underline{\mathbb{P}})$, and let $\underline{\mathbb{E}}$ denote expectation with respect to $\underline{\mathbb{P}}$. Suppose $\underline{\hat{H}}^k \preceq_{\text{st}} \hat{H}^k$. Then it follows from Lemma OA–5 that for all $x \in \mathbb{R}$,

$$\underline{\mathbb{P}}[\underline{L}^k \leq x] \;=\; \underline{\mathbb{P}}[\underline{\hat{H}}^k(x) \geq \gamma] \;\geq\; \mathbb{P}[\hat{H}^k(x) \geq \gamma] \;=\; \mathbb{P}[L^k \leq x].$$

That is, $\underline{L}^k \leq_{\rm st} L^k$. By assumption, $\underline{G}(\underline{\ell},\cdot) \leq_{\rm st} G(\ell,\cdot)$ for all $\underline{\ell} \leq \ell$, and thus it follows easily from Kamae et al. (1977), Theorem 1 [in particular, the equivalence of (i) and (iv)], that $\underline{G}(\underline{L}^k,\cdot) \preceq_{\rm st} G(L^k,\cdot)$. For $h \in \mathcal{P}(\mathbb{R})$, define $\ell(h) = \min\{x \in \mathbb{R} : h(x) \geq \gamma\}$. Then $\ell(\underline{h}) \leq \ell(h)$ for all $\underline{h} \leq_{\rm st} h$. Hence, for $\underline{h} \leq_{\rm st} h$ it holds that

$$\mathbb{P}[\underline{X}^{k+1} \leq x | \underline{\hat{H}}^k = \underline{h}] \;=\; \underline{G}(\ell(\underline{h}),x) \;\geq\; G(\ell(h),x) \;=\; \mathbb{P}[X^{k+1} \leq x|\hat{H}^k = h].$$

Since $\underline{\hat{H}}^k \preceq_{\rm st} \hat{H}^k$, it now follows from Proposition 1 of Kamae et al. (1977) that $\underline{X}^{k+1} \leq_{\rm st} X^{k+1}$ and $(\underline{X}^{k+1},\underline{\hat{H}}^k) \prec (X^{k+1},\hat{H}^k)$ where $\prec$ denotes the usual stochastic order with the coordinate-wise partial ordering on $\mathbb{R}\times\mathcal{P}(\mathbb{R})$ — see page 901 of Kamae et al. (1977). Note that $\underline{\hat{H}}^{k+1} = \eta_k(\underline{X}^{k+1},\underline{\hat{H}}^k)$ and $\hat{H}^{k+1} = \eta_k(X^{k+1},\hat{H}^k)$ where $\eta_k : \mathbb{R}\times\mathcal{P}(\mathbb{R}) \mapsto \mathcal{P}(\mathbb{R})$ is defined by

$$\eta_k(x,h) \;=\; \frac{k}{k+1}h + \frac{1}{k+1}\mathbb{I}_{\{x\leq\ \cdot\}}$$

and observe that η_k is increasing on $\mathbb{R}\times\mathcal{P}(\mathbb{R})$; i.e., $\eta_k(\underline{x},\underline{h}) \leq_{\rm st} \eta(x,h)$ when $\underline{x} \leq x$ and $\underline{h} \leq_{\rm st} h$. It follows that for bounded increasing $f : \mathcal{P}(\mathbb{R}) \mapsto \mathbb{R}$,

$$\mathbb{E}[f(\underline{\hat{H}}^{k+1})] \;=\; \mathbb{E}[(f\circ\eta_k)(\underline{X}^{k+1},\underline{\hat{H}}^k)] \;\leq\; \mathbb{E}[(f\circ\eta_k)(X^{k+1},\hat{H}^k)] \;=\; \mathbb{E}[f(\hat{H}^{k+1})],$$

where the inequality follows from the fact that $f\circ\eta_k$ is bounded and increasing on $\mathbb{R}\times\mathcal{P}(\mathbb{R})$ and $(\underline{X}^{k+1},\underline{\hat{H}}^k) \prec (X^{k+1},\hat{H}^k)$. Hence, $\underline{\hat{H}}^{k+1} \preceq_{\rm st} \hat{H}^{k+1}$. □

Proposition 12 follows from Lemma OA–6.

Proposition 12 **(Stochastic comparison with empirical distributions)** *Suppose* $\underline{G}(\underline{\ell},\cdot) \leq_{\rm st} G(\ell,\cdot)$ *for all* $\underline{\ell} \leq \ell$*, and the empirical distribution is used for both* $\hat{H}$ *and* $\underline{\hat{H}}$*, that is,* $\hat{H}^k(x) := k^{-1}\sum_{j=1}^k \mathbb{I}_{\{X^j \leq x\}}$ *and* $\underline{\hat{H}}^k(x) := k^{-1}\sum_{j=1}^k \mathbb{I}_{\{\underline{X}^j \leq x\}}$*. If* $\underline{L}^0 \leq_{\rm st} L^0$*, then*

$$\begin{aligned}\underline{G}(\underline{L}^k,\cdot) &\preceq_{\rm st} G(L^k,\cdot)\\ \underline{X}^{k+1} &\leq_{\rm st} X^{k+1}\\ \underline{\hat{H}}^{k+1} &\preceq_{\rm st} \hat{H}^{k+1}\\ \underline{L}^{k+1} &\leq_{\rm st} L^{k+1}\end{aligned}$$

for all $k = 0,1,\ldots$*.*

Proposition 13 **(Stochastic comparison with affine updates)** *Suppose that* $\mu : \mathbb{R} \mapsto \mathbb{R}$ *satisfies* $\mu(\ell) \leq \ell$ *for all* ℓ*. Suppose that* $\underline{G}(\underline{\ell},\cdot) = G(\mu(\underline{\ell}),\cdot)$*, and that* $G(\underline{\ell},\cdot) \leq_{\rm st} G(\ell,\cdot)$ *for all* $\underline{\ell} \leq \ell$*. Suppose that* $\hat{H}^k = G(M^k,\cdot)$ *and* $\underline{\hat{H}}^k = G(\underline{M}^k,\cdot)$*, where* $M^k = k^{-1}\sum_{j=1}^k X^j$ *and* $\underline{M}^k = k^{-1}\sum_{j=1}^k \underline{X}^j$*. If* $\underline{L}^0 \leq_{\rm st} L^0$*, then*

$$\underline{G}(\underline{L}^k,\cdot) \preceq_{\rm st} G(L^k,\cdot) \tag{OA–18}$$
$$\underline{X}^{k+1} \leq_{\rm st} X^{k+1} \tag{OA–19}$$
$$\underline{M}^{k+1} \leq_{\rm st} M^{k+1} \tag{OA–20}$$
$$\underline{\hat{H}}^{k+1} \preceq_{\rm st} \hat{H}^{k+1} \tag{OA–21}$$
$$\underline{L}^{k+1} \leq_{\rm st} L^{k+1} \tag{OA–22}$$

for all $k = 0,1,\ldots$*.*

Proof. The proof is by induction; (OA–18)–(OA–22) hold for $k=0$. For the inductive step, suppose that (OA–18)–(OA–22) hold for $k-1$ and consider a general k. Since $\underline{L}^k \leq_{\rm st} L^k$, Theorem 1 of Kamae et al. (1977) implies that $\mu(\underline{L}^k) \leq_{\rm st} L^k$ and $G(\mu(\underline{L}^k),\cdot) \preceq_{\rm st} G(L^k,\cdot)$. Hence, $\underline{G}(\underline{L}^k,\cdot) \preceq_{\rm st} G(L^k,\cdot)$. For $\underline{m} \leq m$, we have

$$\underline{\mathbb{P}}(\underline{X}^{k+1} \leq x | \underline{M}^k = \underline{m}) \;=\; G(\,\ell(G(\underline{m},\cdot))\,,x\,) \;\geq\; G(\,\ell(G(m,\cdot))\,,x\,) \;=\; \mathbb{P}(X^{k+1} \leq x | M^k = m),$$

where $\ell(h) = \min\{x \in \mathbb{R} : h(x) \geq \gamma\}$ for $h \in \mathcal{P}(\mathbb{R})$. Proposition 1 of Kamae et al. (1977) implies that $\underline{X}^{k+1} \leq_{\rm st} X^{k+1}$ and $(\underline{X}^{k+1}, \underline{M}^k) \prec (X^{k+1}, M^k)$, where $\prec$ here denotes the usual stochastic order on $\mathbb{R}^2$. Observe that $M^{k+1} = \varphi_k(X^{k+1}, M^k)$ and $\underline{M}^{k+1} = \varphi_k(\underline{X}^{k+1}, \underline{M}^k)$ where

$$\varphi_k(x,m) \;=\; \frac{k}{k+1} m + \frac{1}{k+1} x.$$

It follows that $\underline{M}^{k+1} \leq_{\rm st} M^{k+1}$, and hence $\underline{\hat{H}}^{k+1} \preceq_{\rm st} \hat{H}^{k+1}$. Finally, $\underline{\mathbb{P}}[\underline{L}^{k+1} \leq x] = \underline{\mathbb{P}}[\underline{\hat{H}}^{k+1}(x) \geq \gamma] \geq \mathbb{P}[\hat{H}^{k+1}(x) \geq \gamma] = \mathbb{P}[L^{k+1} \leq x]$, so $\underline{L}^{k+1} \leq_{\rm st} L^{k+1}$. □

PROPOSITION 14 **(Pathwise comparison)** *Consider any $\omega \in \Omega$ such that, for any k, $\underline{L}^k(\omega) \leq L^k(\omega)$ implies that $\underline{X}^{k+1}(\omega) \leq X^{k+1}(\omega)$. Suppose that the forecasting method used in both sequences satisfies the following condition for all k: If $(\underline{X}^1(\omega), \ldots, \underline{X}^k(\omega)) \leq (X^1(\omega), \ldots, X^k(\omega))$, then $\underline{\hat{H}}^k(\omega,\cdot) \leq_{\rm st} \hat{H}^k(\omega,\cdot)$. If $\underline{L}^0(\omega) \leq L^0(\omega)$, then*

$$\begin{aligned} \underline{X}^k(\omega) &\leq X^k(\omega) \\ \underline{\hat{H}}^k(\omega,\cdot) &\leq_{\rm st} \hat{H}^k(\omega,\cdot) \\ \underline{L}^k(\omega) &\leq L^k(\omega) \end{aligned}$$

for all $k = 1, 2, \ldots$.

Proof. The result follows from induction on k. □

References

Bertsekas, D. P., J. N. Tsitsiklis. 1996. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA.

Billingsley, P. 1968. *Convergence of Probability Measures*. John Wiley & Sons, New York.

Chow, Y. S. 1967. On a strong law of large numbers for martingales. *Ann. Math. Statist.* **38**(2) 610.

Chung, K. L. 1974. *A Course in Probability Theory*. 2nd ed. Academic Press, New York.

Dudley, R. M. 2002. *Real Analysis and Probability*. Cambridge Univ. Press, Cambridge, UK.

Kamae, T., U. Krengel, G. L. O'Brien. 1977. Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5**(6) 899–912.

van Ryzin, G., J. McGill. 2000. Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Sci.* **46**(6) 760–775.