2012-03-10

Quadratic Spline Approximation of the Newsvendor Problem Optimal Cost Function

Christina Marie Burton
Brigham Young University - Provo

Quadratic Spline Approximation of the Newsvendor Problem Optimal Cost Function

Christina M. Burton

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Robin Roundy
Jeffrey Humpherys
Shue-Sum Chow

Department of Mathematics

Brigham Young University

April 2012

ABSTRACT

Quadratic Spline Approximation of the Newsvendor Problem Optimal Cost Function

Christina M. Burton
Department of Mathematics, BYU
Master of Science

We consider a single-product dynamic inventory problem where the demand distributions in each period are known and independent but with density. We assume the lead time and the fixed cost for ordering are zero and that there are no capacity constraints. There is a holding cost and a backorder cost for unfulfilled demand, which is backlogged until it is filled by another order. The problem may be nonstationary, and in fact our approximation of the optimal cost function using splines is most advantageous when demand falls suddenly. In this case the myopic policy, which is most often used in practice to calculate optimal inventory level, would be very costly. Our algorithm uses quadratic splines to approximate the optimal cost function for this dynamic inventory problem and calculates the optimal inventory level and optimal cost.

Keywords: newsvendor problem, single-product, nonstationary, discrete-time, multi-period, finite horizon, independent demands, no capacity constraints, backordering, optimal inventory level, near optimal policy, value function approximation, quadratic splines

Contents

List of Figures

1. INTRODUCTION

The private sector of the United States is currently stocking inventories worth hundreds of billions of dollars. Increasingly, these inventories are being managed by computer software, not by humans. A very small percentage improvement in the management of these inventories can add up to massive savings. We propose to improve a number of the fundamental computations that are performed many times per day in managing inventories that experience random demands.

In most algorithm-driven, real-world inventory management systems, the algorithms do not attempt the computation of an optimal policy. Instead, most commercial software uses the myopic policy. In other words, they optimize the costs that will be incurred in a single time period and ignore the future. That often works well, but when demand levels fall rapidly it leads to very expensive errors. As life cycles for manufactured products continue to shrink and demand volatility continues to grow, this problem is becoming more prevalent.

We will start by studying a single item that experiences random demands, whose inventory is replenished once in every time period. The dominant point of view among both researchers and practitioners is that solving this dynamic program is inappropriate for large, real-world inventory systems because the optimal cost function $\nu_t(x)$ does not have an analytic representation, and approximating it effectively is too time-consuming to solve. We intend to change that perception by creating an algorithm that models the optimal cost function using quadratic splines, which makes it easy to quickly calculate the optimal inventory level in every period.

Amazingly, the published literature on inventory systems has not attempted to use spline functions to approximate the optimal cost function $\nu_t(x)$. The first accepted discussion of the newsvendor problem was by Edgeworth (1888) who used the normal distribution to find the probability that a certain cash reserve level was sufficient to cover all customer demands at a bank. Morse and Kimball (1951) coined the term "newsboy" when formulating the problem while Arrow et al. (1951) pioneered the discrete-time dynamic programming of the single-product problem with independent demands. Karlin (1958) investigated the linear order cost model and Veinott (1965) formulated the myopic policy and its properties. In the 1960's and 1970's, the problem garnered a lot of attention, evolving in name from the Christmas Tree Problem to the Newsboy Problem and finally to the less gender specific name "newsvendor problem", suggested by Matt Sobel in the 1980's [Porteus (2002)]. Since then the study of algorithms that have fast performance guarantees for this fundamental problem has been very limited.

There are two books that give a thorough derivation of some of the most common splines. Shikin and Plis (1995) and Kvasov (2000) use linear algebra to compute the parameters for one dimensional lagrange and cubic splines and prove that these methods converge. The processes described can be modified to change whether none or a few of the spline's derivatives are continuous at the knots. We used quadratic splines to approximate $\nu_t(x)$ in order to have fewer parameters to compute per spline in order to reduce the running time of our program. This lowered the number of degrees of freedom that we could use so ultimately our spline was not smooth like $\nu_t(x)$, though still continuous and convex.

We have created software that calculates the optimal inventory level over many time periods and whose running time will be attractive for companies like Amazon, which stocks millions of different items. In addition, the quadratic spline functions we use to model the optimal cost $\nu_t(x)$ are provably a close approximation. In Section 2 we prove this after defining the dynamic program that we will approximate and after discussing the error terms. We prove that the optimal cost function and its derivatives are bounded in Section 3 and we define the spline function parameters and properties in Section 4. In Section 5 we describe the algorithm that calculates the spline parameters. We summarize the process of creating the spline in Section 6 while Section 7 analyzes the tradeoff between running time and accuracy of our computer program. We leave the reader with some concluding remarks in Section 8. An Appendix is included for the more lengthy proofs and derivations.

2. APPROACH

In this section we define the dynamic program we want to solve and introduce the spline and its properties. We prove the important result that the spline is a close approximation to the dynamic program.

The dynamic program that defines an optimal solution is

$$\nu_t(x) = \min_{y \geq x}\{g_t(y) + \mathbb{E}[\nu_{t+1}(y - D_t)]\} \tag{2.1}$$

In this equation x is the inventory level before ordering in period t, y is the inventory level after ordering, D_t is random demand in period t with finite mean, $\mathbb{E}$ represents the expectation, and $\nu_t(x)$ is the optimal cost of managing the inventory system in period t. Let $x^+ = \max(0, x)$. Then

$$g_t(y) = h_t\mathbb{E}[(y - D_t)^+] + \pi_t\mathbb{E}[(D_t - y)^+] \tag{2.2}$$

is the single-period cost function with holding cost h_t and backorder cost π_t. Let T be the last period to order inventory. The salvage value function is $g_T(x)$, i.e., $\nu_{T+1}(x) = 0$. Since the term in braces is a convex function of y that diverges as

$|y| \to \infty$ [see Zipkin (2000) Theorem 9.4.1], there exists an $R_t < \infty$ that minimizes the term in braces, and $\nu_t(x)$ is constant for $x \leq R_t$.

The true cost function is complicated because the retailer will order new inventory in every period t if the inventory level is less than the cost minimizing level R_t. Its formula is a function of the maximum of the amount of inventory on hand and the cost minimizing level in every time period. For example, let

$$H_t(x) = g_t(x) + \mathbb{E}[\nu_{t+1}(x - D_t)]. \tag{2.3}$$

Define $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$. Then (2.1) can be equivalently written

$$\nu_t(x) = H_t(x \vee R_t), \tag{2.4}$$

and $\nu_t(x)$ is non-decreasing.

The function $\nu_T(x)$ and its derivative are known; however, in periods $t < T$, $\nu_t(x)$ must be computed. We create a function $\mathbb{S}_t(x)$ that is a spline approximation of $\nu_t(x)$, which has some of the same properties–it is convex, continuous and asymptotically linear, but not smooth. We define the equivalents of (2.3) and (2.4) by

$$\mathcal{H}_t(x) = g_t(x) + \mathbb{E}[\mathbb{S}_{t+1}(x - D_t)] \text{ and } v_t(x) = \min_{y \geq x} \mathcal{H}_t(y) = \mathcal{H}_t(x \vee \mathcal{R}_t) \tag{2.5}$$

where $\mathcal{R}_t$ minimizes $\mathcal{H}_t(x)$, and $\mathbb{S}_{T+1}(x) = 0$.

Our algorithm creates the function $\mathbb{S}_t(x)$ so that is at most ϵ_t above or below $v_t(x)$, where $v_t(x)$ is within ε_{t+1} of the optimal cost function $\nu_t(x)$. Hence $\mathbb{S}_t(x)$ is at most

$\varepsilon_{t+1} + \epsilon_t$ above or below $\nu_t(x)$. Thus the error is cumulative, and we have

$$\varepsilon_t = \sum_{j=t}^{T} \epsilon_j. \tag{2.6}$$

Theorem 2.1. *If we create $\mathbb{S}_t(x)$ such that $|\mathbb{S}_t(x) - v_t(x)| \le \epsilon_t$ for all $x \in (-\infty, x_t^{N_t}]$, then $|\mathbb{S}_t(x) - \nu_t(x)| \le \varepsilon_t$ for all x and all t. Furthermore $|\mathcal{H}_t(x) - H_t(x)| \le \varepsilon_{t+1}$.*

Proof. The general idea of this proof was described above. The formal proof is in the Appendix. □

The definition of $x_t^{N_t}$ is not needed in this section. It is given in Section 5.1.

3. BOUNDS ON THE TRUE VALUE FUNCTION

We construct upper and lower bounds for $\nu_t(x)$ and its derivatives. If we expand our definition of $H_t(x)$ we get

$$\begin{aligned} H_t(x) &= g_t(x) + \mathbb{E}[H_{t+1}((x - D_t) \vee R_{t+1})] \\ &= g_t(x) + \mathbb{E}[g_{t+1}((x - D_t) \vee R_{t+1}) + \mathbb{E}[\nu_{t+2}(((x - D_t) \vee R_{t+1}) - D_{t+1})]] \\ &= g_t(x) + \mathbb{E}[g_{t+1}((x - D_t) \vee R_{t+1}) + \\ &\qquad \mathbb{E}[g_{t+2}((((x - D_t) \vee R_{t+1}) - D_{t+1}) \vee R_{t+2}) + \mathbb{E}[\ldots]]]. \end{aligned}$$

Let $D_{[t,j]} = \sum_{i=t}^{j} D_i$ and $D_{[t,j)} = \sum_{i=t}^{j-1} D_i$. Define $D_{(t,j]}$ similarly and set $D_{[t,t)} = 0$. If $x \ge R_j + D_{[t,j)}$ for all $t \le j \le T$, then we have so much inventory that we never have to order again. In the limit as x approaches infinity, the probability that there is enough inventory to meet all future demand converges to 1 and the inequalities in the derivation below become equalities. By the monotonicity of $H_t(x)$,

$$\begin{aligned}
H_t(x) &= g_t(x) + \mathbb{E}[H_{t+1}((x-D_t) \vee R_{t+1})] \\
&\le g_t(x) + \mathbb{E}[H_{t+1}(x-D_t)] = g_t(x) + \mathbb{E}[g_{t+1}(x-D_t) + \mathbb{E}[\nu_{t+2}(x-D_{[t,t+1]})]] \\
&= g_t(x) + \mathbb{E}[g_{t+1}(x-D_t) + \mathbb{E}[H_{t+2}((x-D_{[t,t+1]}) \vee R_{t+2})]] \\
&\le g_t(x) + \mathbb{E}[g_{t+1}(x-D_t) + \mathbb{E}[H_{t+2}(x-D_{[t,t+1]})]] \\
&= \ldots = \sum_{j=t}^{T} \mathbb{E}[g_j(x-D_{[t,j)}] = \widetilde{H}_t(x). \qquad (3.1)
\end{aligned}$$

Note that

$$\widetilde{H}_t(x) = g_t(x) + \mathbb{E}[\widetilde{H}_{t+1}(x-D_t)]. \qquad (3.2)$$

We will bound $H_t(x)$ in the following theorem.

Theorem 3.1. *Let* $\mu_{[t,j]} = \mathbb{E}[D_{[t,j]}]$. *Then*

$$H_t(x) \le \sum_{j=t}^{T} \big(h_j(x-\mu_{[t,j]}) + (h_j+\pi_j)\mathbb{E}[(D_{[t,j]}-x)^+]\big) = \widetilde{H}_t(x) \textit{ for all } x. \qquad (3.3)$$

Proof. The idea of the proof appears above. The formal proof is in the Appendix. □

To find a simple lower bound on $H_t(x)$, notice that for all t

$$g_t(x) \ge h_t\mathbb{E}[(x-D_t)^+] \ge h_t\mathbb{E}[x-D_t] = h_t(x-\mu_t). \qquad (3.4)$$

Theorem 3.2.

$$H_t(x) \ge \sum_{j=t}^{T} h_j(x-\mu_{[t,j]}) = \overline{H}_t(x). \qquad (3.5)$$

Proof. See the Appendix for the proof. □

To summarize,

$$\overline{H}_t(x) \leq \nu_t(x) \leq \widetilde{H}_t(x) \,\forall\, x. \tag{3.6}$$

Note that

$$\begin{aligned}\widetilde{H}_t(x) - \overline{H}_t(x) &= \sum_{j=t}^{T} \big(h_j(x - \mu_{[t,j]}) + (h_j + \pi_j)\mathbb{E}[(D_{[t,j]} - x)^+]\big) - \sum_{j=t}^{T} h_j(x - \mu_{[t,j]}) \\ &= \sum_{j=t}^{T} (h_j + \pi_j)\mathbb{E}[(D_{[t,j]} - x)^+]. \end{aligned} \tag{3.7}$$

Clearly, $\widetilde{H}_t(x) - \overline{H}_t(x)$ is a non-negative, decreasing function in x.

We can prove more about the relationships between the derivatives of $\nu_t(x)$ and the derivatives of $\widetilde{H}_t(x)$ and $\overline{H}_t(x)$. We note that convex functions are differentiable almost anywhere. In general, we use $f'(x)$ to refer to the left-handed derivative of f at x. Since $\nu_t(x)$ is constant to the left of R_t and $\nu_t(x) = H_t(x)$ if $x > R_t$,

$$\nu_t'(x) = \left\{ \begin{array}{cc} 0 & \text{if } x \leq R_t \\ H_t'(x) & \text{if } x > R_t \end{array} \right\} = H_t'(x \vee R_t) \geq H_t'(x) \tag{3.8}$$

where $H_t'(R_t) = 0$ by the definition of R_t. Similarly,

$$\nu_t''(x) = \left\{ \begin{array}{cc} 0 & \text{if } x \leq R_t \\ H_t''(x) & \text{if } x > R_t \end{array} \right\} = H_t''(x \vee R_t) \leq H_t''(x). \tag{3.9}$$

Theorem 3.3. *If the density of D_t is bounded then $\widetilde{H}_t'(x) \leq \nu_t'(x) \,\forall\, x$.*

Proof. By (2.1), (2.2), (2.3) and (3.1), $\nu_t(x)$, $H_t(x)$, and $\widetilde{H}_t(x)$ are uniformly Lipschitz continuous. By (3.2), $\widetilde{H}_t'(x) = g_t(x) + \frac{d}{dx}\mathbb{E}[\widetilde{H}_{t+1}(x - D_t)]$. By way of induction, note that $\widetilde{H}_{T+1}(x) = \nu_{T+1}(x) = 0$ and assume $\widetilde{H}_{t+1}'(x) \leq \nu_{t+1}'(x) \,\forall\, x$. By uniform Lipschitz continuity, we can invert the derivative and expectation. From (3.8) and (2.3),

$$\nu_t'(x) \geq H_t'(x) = g_t'(x) + \mathbb{E}[\nu_{t+1}'(x - D_t)] \geq g_t'(x) + \mathbb{E}\big[\widetilde{H}_{t+1}'(x - D_t)\big] = \widetilde{H}_t'(x).$$

$\square$

Theorem 3.4. $\nu_t'(x) \leq \overline{H}_t'(x)\,\forall\, x$.

Proof. Note that $\nu_T'(x) = g_T'(x) \leq h_T = \overline{H}_T'(x)$. Assume $\nu_{t+1}'(x) \leq \overline{H}_{t+1}'(x)\,\forall\, x$. By (3.8) and the uniform Lipschitz continuity of $\nu_{t+1}(x)$,

$$\begin{aligned} \nu_t'(x) = H_t'(x \vee R_t) &= g_t'(x \vee R_t) + \mathbb{E}\big[\nu_{t+1}'((x \vee R_t) - D_t)\big] \\ &\leq g_t'(\infty) + \mathbb{E}\big[\overline{H}_{t+1}'((x \vee R_t) - D_t)\big] = h_t + \sum_{j=t+1}^{T} h_j = \overline{H}_t'(x). \end{aligned}$$

$\square$

Remark 3.5. *Theorems 3.3 and 3.4 say that both $\widetilde{H}_t(x) - \nu_t(x)$ and $\nu_t(x) - \overline{H}_t(x)$ decrease monotonically in x.*

Theorem 3.6. *If the density of D_t is bounded then $\nu_t''(x) \leq \widetilde{H}_t''(x)\,\forall\, x$.*

Proof. Assume $\nu_{t+1}''(x) \leq \widetilde{H}_{t+1}''(x)\,\forall\, x$. Because D_t has a bounded density, $g_t'(x)$ is uniformly Lipschitz continuous. By (2.1) and (3.2), the same can be said for $\nu_{t+1}(x)$. Therefore we can invert expectations and second derivatives. By (3.9)

$$\nu_t''(x) \leq H_t''(x) = g_t''(x) + \mathbb{E}\left[\nu_{t+1}''(x - D_t)\right] \leq g_t''(x) + \mathbb{E}\left[\widetilde{H}_{t+1}''(x - D_t)\right] = \widetilde{H}_t''(x).$$

$\square$

The second derivative bound is important in calculating the maximum error of our approximation.

We construct $v_t(x)$ so that, like $\nu_t(x)$, it has the following properties:

Theorem 3.7. *$\overline{H}_t(x) \leq v_t(x) \leq \widetilde{H}_t(x)$ for all x.*

Proof. In period T, $v_T(x) = \nu_T(x)$ so $\overline{H}_T(x) \leq v_T(x) \leq \widetilde{H}_T(x)$ by (3.6).

By (3.2), Lemma 5.2 and (9.5),

$$\begin{aligned}
\widetilde{H}_t(x) \geq \widetilde{H}_t(x \vee \mathcal{R}_t) &= g_t(x \vee \mathcal{R}_t) + \mathbb{E}[\widetilde{H}_{t+1}((x \vee \mathcal{R}_t) - D_t)] \\
&\geq g_t(x \vee \mathcal{R}_t) + \mathbb{E}[\mathbb{S}_{t+1}((x \vee \mathcal{R}_t) - D_t)] \\
&\geq g_t(x \vee \mathcal{R}_t) + \mathbb{E}[\overline{H}_{t+1}((x \vee \mathcal{R}_t) - D_t)] = \overline{H}_t(x \vee \mathcal{R}_t) \geq \overline{H}_t(x)
\end{aligned}$$

where $v_t(x) = g_t(x \vee \mathcal{R}_t) + \mathbb{E}[\mathbb{S}_{t+1}((x \vee \mathcal{R}_t) - D_t)]$ by (2.5). $\square$

Theorem 3.8. *If the density of D_t is bounded, then $v_t(x)$ is convex and non-decreasing.*

Proof. By convexity of $\nu_T(x)$ [Zipkin (2000) Theorem 9.4.1], we have $v_T(x) = \nu_T(x)$ is convex.

For $t < T$, by (2.5) we have $v_t''(x) = g_t''(x \vee \mathcal{R}_t) + \frac{d^2}{dx^2}\mathbb{E}[\mathbb{S}_{t+1}((x \vee \mathcal{R}_t) - D_t)]$. Due to the bounded density of D_t, we can invert the derivative and expectation to get $v_t''(x) = g_t''(x \vee \mathcal{R}_t) + \mathbb{E}[\mathbb{S}_{t+1}''((x \vee \mathcal{R}_t) - D_t)]$. $g_t(x)$ is convex and $\mathbb{S}_{t+1}(x)$ is convex by Theorem 5.1 so $g_t''(x) \geq 0$ and $\mathbb{S}_t''(x) \geq 0$, which implies $v_t''(x) \geq 0$ so $v_t(x)$ is convex. By (2.5), $v_t'(x) = 0$ to the left of $\mathcal{R}_t$ so monotonicity follows. $\square$

4. THE SPLINE FUNCTIONS

4.1. **Spline Properties.** In period t our task is to approximate $v_t(x)$ with a function $\mathbb{S}_t(x)$. $\mathbb{S}_t(x)$ has some useful properties. We will prove in this section that $\mathbb{S}_t(x)$ is convex, asymptotically linear and bounded from below. If x_t^i and x_t^{i+1} are successive knots of $\mathbb{S}_t(x)$ then the function values of $\mathbb{S}_t(x)$ on the interval $[x_t^i, x_t^{i+1}]$ will depend only on the function values and derivatives of $v_t(x)$ at x_t^i and x_t^{i+1}, i.e., the dependence of $\mathbb{S}_t(x)$ on $v_t(x)$ is localized. We use quadratic splines as opposed to cubic splines because the simple algebraic forms of quadratic splines make convexity easier to retain and reduce the computational effort per knot.

A knot is the x value at the end of one quadratic section of the spline and the beginning of another. Let the first knot be $x_t^1 = \mathcal{R}_t$. As we remarked after (2.1), $\mathbb{S}_t(x)$ will be constant to the left of x_t^1 so let $C_t = v_t(\mathcal{R}_t)$. Following (2.5), let $\mathbb{S}_t(x) = \mathcal{H}_t(\mathcal{R}_t) = C_t$ for $x \leq x_t^1 = \mathcal{R}_t$. We will discuss in depth how to create the knots $\{x_t^i\}_{i=2}^{M_t}$ in Section 5. The formula of the spline is

$$\mathbb{S}_t(x) = C_t + \sum_{i=1}^{M_t} (c_t^i (x - x_t^i)^+ + \frac{d_t^i}{2}((x - x_t^i)^+)^2) \tag{4.1}$$

where c_t^i and d_t^i are defined in Section 4.4.

4.2. **Upper Bound on the Second Derivative.** In order to create a quadratic spline that is within ϵ_t of $v_t(x)$, it is helpful to know the maximum of $v_t''(x)$ on an interval. We will show how $v_t''(x)$ can be evaluated in Section 4.4. We use the open source chebfun package created by a team from the Oxford University Mathematical Institute to find the coordinates of the local minimums and maximums of $v_t''(x)$ on $[x_t^i, x_t^{i+1}]$. First we find the set of local maxima and the set of local minima that have x values between x_t^i and x_t^{i+1}. If the set of local maxima is empty, we evaluate $v_t''(x_t^i)$ and $v_t''(x_t^{i+1})$. Otherwise the x values of the local maxima and minima tell us whether x_t^i and x_t^{i+1} are local maximizers of $v_t''(x)$ on $[x_t^i, x_t^{i+1}]$ or not, and we check all local maximizers of $v_t''(x)$. Let B^i be half of the maximum of the elements of the set of local maxima and $v_t''(x_t^i)$ and $v_t''(x_t^{i+1})$ if necessary.

4.3. **Error Analysis For a Single Interval.** Our goal is to create a spline $\mathbb{S}_t(x)$ such that $|\mathbb{S}_t(x) - v_t(x)| \leq \epsilon_t$ for $x_t^i \leq x \leq x_t^{i+1}$. Once we have fixed a knot x_t^i, we guess at a knot x_t^{i+1}, define our quadratic spline $\mathbb{S}_t(x)$ on $[x_t^i, x_t^{i+1}]$, and get an upper bound on $|\mathbb{S}_t(x) - v_t(x)|$ over $[x_t^i, x_t^{i+1}]$. The bound depends on B^i and the values and derivatives of $v_t(x)$ at x_t^i and x_t^{i+1}. If the error bound is satisfactory we accept x_t^{i+1}; otherwise we adjust it (see Section 5.1).

To standardize our error analysis, we transform $v_t(x)$ by mapping $(x_t^i, v_t(x_t^i))$ into $(0,0)$ and $(x_t^{i+1}, v_t(x_t^{i+1}))$ into $(0,1)$. Let $\overline{v}_t^i(x')$ be the image of $v_t(x)$ on $[x_t^i, x_t^{i+1}]$ under this transformation. The formula for the transformation is given in the Appendix. Note that $(\overline{v}_t^i)'(0)$ and $(\overline{v}_t^i)'(1)$ are determined by $v_t'(x_t^i)$ and $v_t'(x_t^{i+1})$, and that the upper bound on $\frac{1}{2}(\overline{v}_t^i)''(x')$ becomes

$$B_i = B^i(x_t^{i+1} - x_t^i)^2 = B^i \delta_i^2. \tag{4.2}$$

Let

$$\lambda = -(\overline{v}_t^i)'(0) \wedge (\overline{v}_t^i)'(1) \text{ and } \Lambda = -(\overline{v}_t^i)'(0) \vee (\overline{v}_t^i)'(1). \tag{4.3}$$

Since $v_t(x)$ is convex, λ and Λ are positive. We create our spline by defining $(S_t^i)''(x') = 2\lambda$. Then $S_t^i(x') = \lambda x'(x'-1)$ for $x' \in [0,1]$. $\mathbb{S}_t(x)$ will be convex at x_t^i and x_t^{i+1} because $\lambda = (S_t^i)'(1-\epsilon) \leq (\overline{v}_t^i)'(1)$ and $-\lambda = (S_t^i)'(\epsilon) \geq (\overline{v}_t^i)'(0)$.

The function $\overline{v}_t^i(x')$ must stay convex between $(0,0)$ and $(1,0)$ and cannot curve too quickly because $(\overline{v}_t^i)''(x') \leq 2B_i$. We use this information to create upper and lower bound functions for $\overline{v}_t^i(x')$ called $V_i(x')$, $i = 0,1,2$. In our derivations of $V_i(x')$, $i = 0,1,2$ and in Figure 1 we assume that $\Lambda = -(\overline{v}_t^i)'(0)$ and $\lambda = (\overline{v}_t^i)'(1)$. Otherwise Figure 2 represents the upper and lower bounds. Our formulas for the maximum error apply in either case.

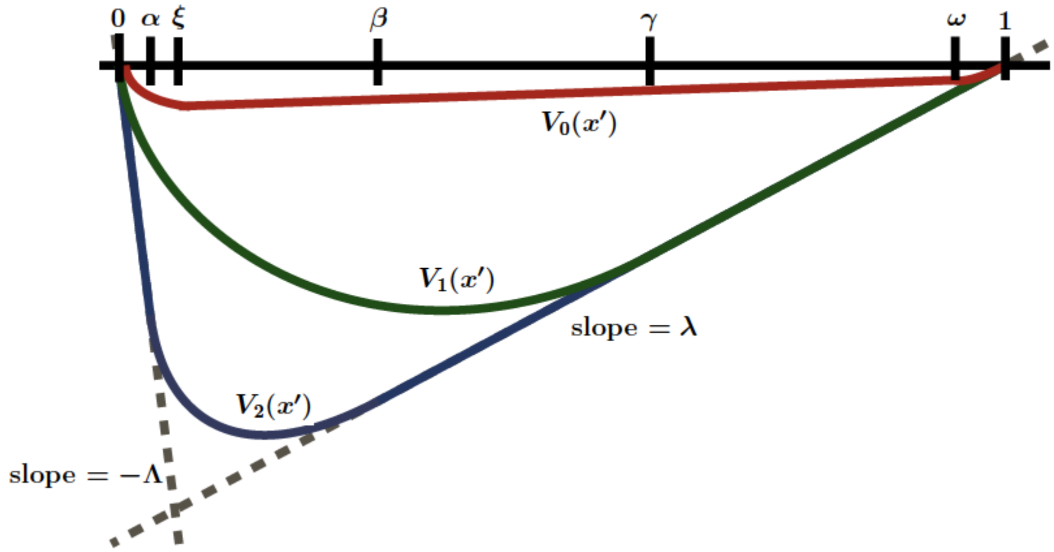

FIGURE 1. Upper and lower bounds for $\overline{v}_t^i(x')$ with $\lambda = (\overline{v}_t^i)'(1)$ and $\Lambda = -(\overline{v}_t^i)'(0)$.

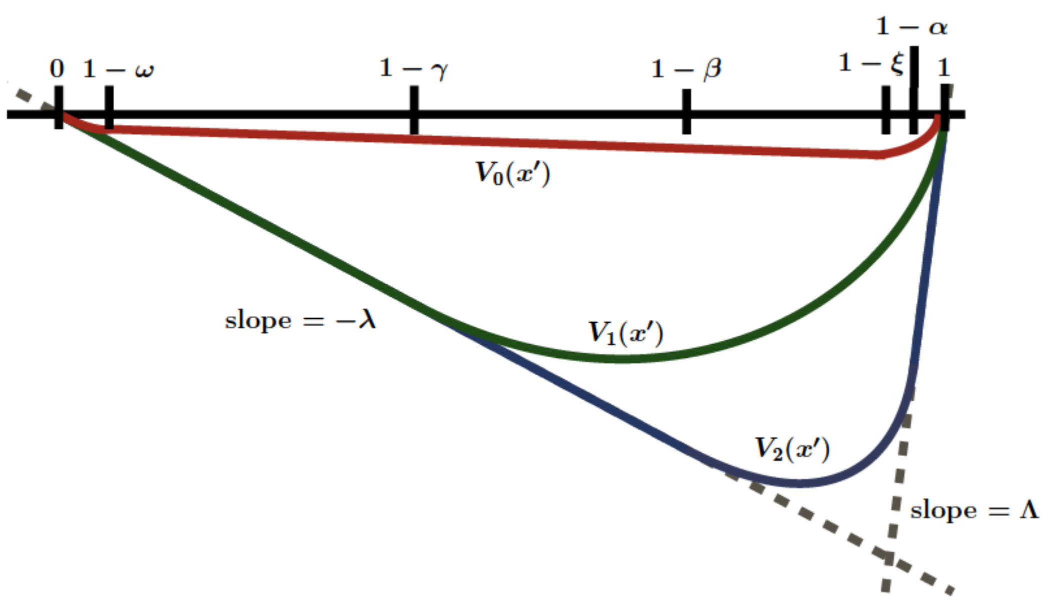

FIGURE 2. Upper and lower bounds for $\overline{v}_t^i(x')$ on the unit domain with $\lambda = -\overline{v}_t^i(0)$ and $\Lambda = \overline{v}_t^i(1)$.

Since $(\overline{v}_t^i)'(0)$ and $(\overline{v}_t^i)'(1)$ are fixed, $(\overline{v}_t^i)''(x') \leq 2B_i$ and $\overline{v}_t^i(x')$ is convex, $\overline{v}_t^i(x')$ must be below the graph of

$$V_0(x') = \begin{cases} B_i(x')^2 - \Lambda x' & \text{if } 0 \leq x' \leq \xi, \\ (2B_i(\omega - 1) + \lambda)x' + B_i(1 - \xi^2) - \lambda & \text{if } \xi \leq x' \leq \omega, \\ (x' - 1)(B_i(x' - 1) + \lambda) & \text{if } \omega \leq x' \leq 1, \end{cases}$$

where

$$\xi = (\Lambda - (\Lambda + \lambda)^2/(4B_i))/(2B_i - (\Lambda + \lambda)), \text{ and}$$
$$\omega = 1 - (\lambda - (\Lambda + \lambda)^2/(4B_i))/(2B_i - (\Lambda + \lambda)).$$

Note that $V_0(x')$ is continuous and smooth on $[0, 1]$. $V_0(x') - S_t^i(x')$ is largest where their derivatives are equal, at $x_* = 1 + \frac{B_i(\omega-1)}{\lambda}$. The maximum error is

$$\overline{\epsilon} = V_0\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right) - S_t^i\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right) = B_i(\omega - 1)^2\left(\frac{B_i}{\lambda} - 1\right). \quad (4.4)$$

We obtain two bounds on the maximum amount by which $S_t^i(x')$ can be above $\overline{v}_t^i(x')$. The applicable bound depends on the relative sizes of λ, Λ and B_i. The functions that bound $\overline{v}_t^i(x')$ from below, shown in Figure 1, are

$$V_1(x') = \begin{cases} B_i(x')^2 - \Theta x' & \text{if } 0 \leq x' \leq \gamma, \\ \lambda(x' - 1) & \text{if } \gamma \leq x' \leq 1, \end{cases} \quad (4.5a)$$

$$V_2(x') = \begin{cases} -\Lambda x' & \text{if } 0 \leq x' \leq \alpha, \\ B_i(x' - \alpha)^2 - \Lambda x' & \text{if } \alpha \leq x' \leq \beta, \\ \lambda(x' - 1) & \text{if } \beta \leq x' \leq 1, \end{cases} \quad (4.5b)$$

where $\Theta = -\lambda + 2\sqrt{B_i\lambda}$, $\gamma = (\lambda+\Theta)/(2B_i)$, $\alpha = \frac{\lambda-\Lambda}{2(\lambda+\Lambda)} + \left(\frac{\lambda+\Lambda}{2B_i}\right)\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)$ and $\beta = \frac{\lambda}{\lambda+\Lambda} + \frac{\lambda+\Lambda}{4B_i}$.

The largest error $S_t^i(x')$ can be above $V_1(x')$ and $V_2(x')$ occurs at $x_* = \frac{\Theta-\lambda}{2(B_i-\lambda)}$ and $x_* = \frac{(\Lambda-\lambda+2B_i\alpha)}{2(B_i-\lambda)}$ respectively, and is equal to

$$\underline{\epsilon} = \begin{cases} \frac{\lambda(\sqrt{B_i}-\sqrt{\lambda})}{\sqrt{B_i}+\sqrt{\lambda}} & \text{if } B_i \leq \frac{(\lambda+\Lambda)^2}{4\lambda} \\ \frac{1}{4}\Big[B_i - \lambda - \frac{B_i(\Lambda-\lambda)^2}{(\lambda+\Lambda)^2} & \\ \quad + \left(\frac{(\Lambda-\lambda)^2}{B_i-\lambda} - \frac{(\lambda+\Lambda)^2}{B_i}\right)\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)^2\Big] & \text{if } \frac{(\lambda+\Lambda)^2}{4\lambda} \leq B_i. \end{cases} \tag{4.6}$$

For the derivation of the parameters of $V_i(x')$, $i = 0, 1$ and 2, see the Appendix. Let

$$\epsilon_t^i = \max\{\bar{\epsilon}, \underline{\epsilon}\}. \tag{4.7}$$

Lemma 4.1. *If* $(\lambda+\Lambda)^2 \geq 4B_i\lambda$ *then* $V_0(x') \geq \overline{v}_t^i(x') \geq V_1(x')$ *for* $x' \in [0,1]$. *Otherwise* $V_0(x') \geq \overline{v}_t^i(x') \geq V_2(x')$.

Proof. See the Appendix. □

Lemma 4.2. $|S_t^i(x') - \overline{v}_t^i(x')| \leq \epsilon_t^i$ *for* $0 \leq x' \leq 1$.

Proof. This was proved in the development above. □

Having defined $S_t^i(x')$, we find $\mathbb{S}_t(x)$ for $x_t^i \leq x \leq x_t^{i+1}$ by inverting the transformation we applied to $v_t(x)$ on $[x_t^i, x_t^{i+1}]$. The details are in the Appendix.

Remark 4.3. *As the transformation* (9.16) *does not scale the units of measure along the* y *axis,* $|\mathbb{S}_t(x) - v_t(x)| \leq \epsilon_t^i$ *for* $x_t^i \leq x \leq x_t^{i+1}$.

4.4. **The Algebra of the Spline and Minimizing $\mathcal{H}_t(x)$.** In the previous subsection we defined $\mathbb{S}_t(x)$ on each interval. The result is equation (9.20a), which says

$\mathbb{S}_t(x) = v_t(x_t^i) + m_t^i(x - x_t^i) + A^i(x - x_t^i)(x - x_t^{i+1})$ for $x_t^i \le x \le x_t^{i+1}$ where A^i, defined in (9.19), is a function of λ, x_t^i, and x_t^{i+1}. $\mathbb{S}_t(x)$ is convex and is within ϵ_t of $v_t(x)$ for $x \le x_t^{N_t}$. Now we need to represent $\mathbb{S}_t(x)$ in the format of (4.1). Making these definitions agree yields formulas for the coefficients of (4.1), namely

$$c_t^i = m_t^i + A^i(x_t^i - x_t^{i+1}) - \sum_{k=1}^{i-1}(c_t^k + d_t^k(x_t^i - x_t^k)) \text{ and } d_t^i = 2A^i - \sum_{k=1}^{i-1} d_t^k. \tag{4.8}$$

To minimize $\mathcal{H}_t(x)$ and find $\mathcal{R}_t$ we use the MATLAB function *fminunc*, which accepts the first and second derivatives of $v_t(x)$ for greater accuracy. When we evaluate $v_t(x)$ and its derivatives, we use the representation of $\mathbb{S}_{t+1}(x)$ given by (4.1). We now describe how this is done.

Define $n_{D_u}(x) = \mathbb{E}[(D_u - x)^+]$ and $n_{D_u}^2(x) = \mathbb{E}[((D_u - x)^+)^2]$. Using (4.1) we can compute $\mathcal{H}_t(x) = g_t(x) + \mathbb{E}[\mathbb{S}_{t+1}(x - D_t)]$ easily as follows:

$$\begin{aligned}
\mathcal{H}_t(x) &= h_t(x - \mu_t) + (h_t + \pi_t)\mathbb{E}[(D_t - x)^+] + C_{t+1} \\
&\qquad + \sum_{i=1}^{M_{t+1}} (c_{t+1}^i \mathbb{E}[(x - x_{t+1}^i - D_t)^+] + \frac{d_{t+1}^i}{2}\mathbb{E}[((x - x_{t+1}^i - D_t)^+)^2]) \\
&= h_t(x - \mu_t) + (h_t + \pi_t) n_{D_t}(x) + \sum_{i=1}^{M_{t+1}} c_{t+1}^i \big(x - x_{t+1}^i - \mu_t + n_{D_t}(x - x_{t+1}^i)\big) \\
&\qquad + C_{t+1} + \sum_{i=1}^{M_{t+1}} \frac{d_{t+1}^i}{2}\big(\mathbb{E}[D_t^2] - 2\mu_t(x - x_{t+1}^i) + (x - x_{t+1}^i)^2 - n_{D_t}^2(x - x_{t+1}^i)\big).
\end{aligned} \tag{4.9}$$

Let $\overline{F}_{D_t}(x) = 1 - F_{D_t}(x)$ where $F_{D_t}(x)$ and $f_{D_t}(x)$ are the cumulative and probability distribution functions for D_t respectively. By (4.9) we have

$$\begin{aligned}\mathcal{H}_t'(x) &= h_t - (h_t + \pi_t)\overline{F}_{D_t}(x) \\ &\quad + \sum_{i=1}^{M_{t+1}} \left[c_{t+1}^i F_{D_t}(x - x_{t+1}^i) + d_t^i\big(x - x_{t+1}^i - \mu_t + n_{D_t}(x - x_{t+1}^i)\big)\right],\end{aligned} \tag{4.10}$$

and

$$\mathcal{H}_t''(x) = (h_t + \pi_t) f_{D_t}(x) + \sum_{i=1}^{M_{t+1}} c_{t+1}^i f_{D_t}(x - x_{t+1}^i) + d_t^i F_{D_t}(x - x_{t+1}^i). \tag{4.11}$$

When finding $\mathcal{R}_t$ by numerically minimizing $\mathcal{H}_t(x)$ we can restrict the search for $\mathcal{R}_t$ to a small interval, $[r_t \wedge \mathcal{R}_{t+1}, r_t]$, where $r_t = \arg\min g_t(x)$. This is advantageous because r_t is easily evaluated; it is the solution to the classical newsvendor problem. To see why r_t is an upper bound on $\mathcal{R}_t$, note that $g_t(x)$ and $\mathbb{E}[\mathbb{S}_{t+1}(x - D_t)]$ are convex functions, and the latter is non-decreasing. Hence the minimizer of their sum cannot be greater than the minimizer of $g_t(x)$. Also, note that if $x \le \mathcal{R}_{t+1}$ then $x - D_t \le x \le \mathcal{R}_{t+1}$, so $\mathbb{S}_{t+1}(x) = C_{t+1}$. Therefore if $x \le r_t \wedge R_{t+1}$, then $g_t(x) + \mathbb{E}[\mathbb{S}_{t+1}(x - D_t)]$ is a decreasing function.

5. THE KNOT CREATION ALGORITHM

5.1. **Knot Creation Algorithm for $x \le x_t^{N_t}$.** In Section 4 we gave the formula for the spline as a function of the knots and the values of $v_t(x)$ and of $v_t'(x)$ at the knots. We also calculated the maximum approximation error between two consecutive knots. In this section we describe the algorithm we use to choose knots that are few in number, but which keep the spline close to $v_t(x)$.

The basic idea of the knot creation algorithm is that given x_t^i we want to find an x_t^{i+1} and a quadratic spline between x_t^i and x_t^{i+1} that is within ϵ_t of $v_t(x)$. The error

ϵ_t^i, defined in (4.7), is a function of x_t^i and x_t^{i+1}, $v_t(x_t^i)$, $v_t(x_t^{i+1})$, $v_t'(x_t^i)$, and $v_t'(x_t^{i+1})$ as well as the maximum of the second derivative of $v_t(x)$ on $[x_t^i, x_t^{i+1}]$. Because of the way our computations are sequenced, it is most importantly a function of x_t^{i+1}. The larger x_t^{i+1} is, the larger ϵ_t^i will be.

The number of knots would be minimized if we found the value of x_t^{i+1} that made ϵ_t^i exactly equal to ϵ_t. However, to avoid excessive computation in finding x_t^{i+1}, we will make an initial conjecture on where x_t^{i+1} should be, then test and, if necessary, refine our conjecture. Our initial conjecture is based on $\underline{\epsilon}$ (see (4.6)). If we make the simplifying assumption that $\lambda = (\overline{v}_t^i)'(1)$, we can set $\epsilon_t = \frac{\lambda(\sqrt{B_i}-\sqrt{\lambda})}{\sqrt{B_i}+\sqrt{\lambda}}$ and substitute into (4.2), solve for δ_i, and get

$$\delta_i = \frac{\sqrt{\lambda}+\epsilon_t}{\sqrt{\lambda}-\epsilon_t} \cdot \frac{\kappa\sqrt{\lambda}}{\sqrt{B^i}}. \tag{5.1}$$

The parameters κ and Δ give the user freedom to tune the algorithm. We use $\kappa = \Delta = 1$.

We will terminate the knot creation algorithm if the spline is below $\widetilde{H}_t(x)$ or above $\overline{H}_t(x)$ by less than ε_t. Define the distance functions

$$ud_t(x) = \widetilde{H}_t(x) - \mathbb{S}_t(x) \quad \text{and} \quad dd_t(x) = \mathbb{S}_t(x) - \overline{H}_t(x). \tag{5.2}$$

Initially we guess that $x_t^{i+1} = \delta_i + x_t^i$, but if $\epsilon_t^i > \epsilon_t$ we cut the distance to our next knot in half and try again. If the error ϵ_t^i is less than ϵ_t, but ϵ_t^i is too small, we double the distance to our next knot and try again. If $\frac{\epsilon_t}{4} \le \epsilon_t^i \le \epsilon_t$ then we accept x_t^{i+1} and check whether $ud_t(x_t^{i+1}) \le \varepsilon_t$ or $dd_t(x_t^{i+1}) \le \varepsilon_t$. If not, the knot algorithm continues. If so we set $N_t = i+1$, change our approach and start using linear splines. After the last knot $x_t^{M_t}$ we define the spline to be $\overline{H}_t(x)$ because it maintains a simple linear form and sandwiches $\mathbb{S}_t(x)$ between $\widetilde{H}_t(x)$ and $\overline{H}_t(x)$ for all $x \geq x_t^{N_t}$ as proven in Theorem 3.7.

The following are the steps for the first knot creation algorithm:

<u>Knot Creation Algorithm for $x \leq x_t^{N_t}$:</u>

(1) $x_t^1 = \mathcal{R}_t$, $i = 1$, $k = 2$, $\kappa = 1$, $\Delta = 1$, $n = 0$.

(2) Set $B^i = \frac{1}{2} v_t''(x_t^i)$ using (4.11). Define δ_i using (5.1) and let $x_t^{i+1} = x_t^i + \delta_i$.

(3) Set $B^i = \frac{1}{2} \max_{[x_t^i, x_t^{i+1}]} v_t''(x)$. Use (4.4), (4.6) and (4.7) to find ϵ_t^i, which is the maximum of the top and bottom error.

(4) Set $b = n$. Redefine

$$n = \begin{cases} 1 & \text{if } \epsilon_t^i < \frac{\epsilon_t}{4\Delta} \\ -1 & \text{if } \epsilon_t^i > \epsilon_t \\ 0 & \text{otherwise.} \end{cases}$$

If $n = 0$ or 1 and in addition either $ud_t(x_t^{i+1}) \leq \varepsilon_t$ or $dd_t(x_t^{i+1}) \leq \varepsilon_t$, set $N_t = i + 1$, define c_t^i and d_t^i as in (4.8), and go to 6. Otherwise go to 5.

(5) If $n = -1$ and $b = 1$, set $\delta_i = \frac{\delta_i}{k}$ and $x_t^{i+1} = x_t^i + \delta_i$.

If $nb = -1$ or $n = 0$, define c_t^i and d_t^i as in (4.8), set $i = i + 1$, and set $n = 0$.

If $n \neq 0$, and either $nb = 1$ or $b = 0$, let $\delta_i = \delta_i * k^n$ and $x_t^{i+1} = x_t^i + \delta_i$.

Go to 3.

(6) Define $\mathbb{S}_t(x)$ for all x greater than or equal to $x_t^{N_t}$ using the "Knot Creation Algorithm for $x \geq x_t^{N_t}$".

5.2. **Knot Creation Algorithm for $x \geq x_t^{N_t}$.** When $0 \leq ud_t(x_t^i) \leq \varepsilon_t$ or $0 \leq dd_t(x_t^i) \leq \varepsilon_t$, we have different criteria for creating the knots because we have more information about $\nu_t(x)$. Recall that $\nu_t(x)$ is bounded between $\widetilde{H}_t(x)$ and $\overline{H}_t(x)$, and that both $\widetilde{H}_t(x) - \nu_t(x)$ and $\nu_t(x) - \overline{H}_t(x)$ are non-increasing as proven in Theorems 3.1, 3.2, 3.3 and 3.4. These are the steps for the knot creation algorithm after $0 \leq ud_t(x_t^i) \leq \varepsilon_t$ or $0 \leq dd_t(x_t^i) \leq \varepsilon_t$:

(1) If $0 \leq ud_t(x_t^i) < \epsilon_t$ (case 1a) or $dd_t(x_t^i) \leq \varepsilon_t$ (case 1b), find the linear function $V_i(x) = m_t^i x + b_t^i$ that passes through the point $(x_t^i, \mathbb{S}_t(x_t^i))$ and is tangent to $\widetilde{H}_t(x)$ at some $\widetilde{x} > x_t^i$. Go to 3 if $0 \leq ud_t(x_t^i) < \epsilon_t$ and go to 4 otherwise.

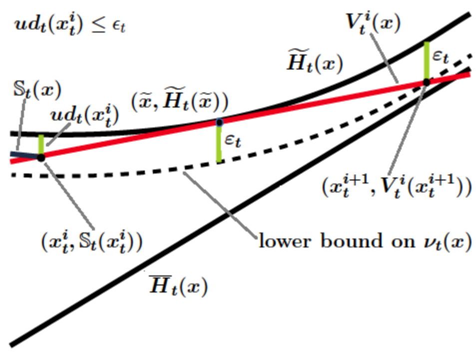

Figure 3. Linear spline in case 1a.

(2) If $\epsilon_t \leq ud_t(x_t^i) \leq \varepsilon_t$, find the line $V_t^i(x) = m_t^i x + b_t^i$ that passes through the point $(x_t^i, \mathbb{S}_t(x_t^i))$ and is tangent to $\widetilde{H}_t(x) - (ud_t(x_t^i) - \epsilon_t)$ at some $\widetilde{x} > x_t^i$. If $\mathbb{S}_t'(x_t^i) > m_t^i$ (case 2b), let $m_t^i = \mathbb{S}_t'(x_t^i)$ and define b_t^i such that $V_t^i(x) = m_t^i x + b_t^i$ goes through $(x_t^i, \mathbb{S}_t(x_t^i))$ (otherwise case 2a applies). Go to 3.

(3) Find the point $(x_t^{i+1}, V_t^i(x_t^{i+1}))$ so that $\widetilde{H}_t(x_t^{i+1}) - V_t^i(x_t^{i+1}) = \varepsilon_t$ and $x_t^{i+1} > x_t^i$. If $V_t^i(x_t^{i+1}) < \overline{H}_t(x_t^{i+1})$ go to 4. Otherwise let $i = i + 1$ and go to 2.

(4) Find the point $\overline{x}$ where $V_t^i(x)$ and $\overline{H}_t(x)$ intersect (case 3). Let $\mathbb{S}_t(x) = V_t^i(x)$ for $x_t^i \leq x \leq \overline{x}$. Let $M_t = i + 1$ and define $x_t^{M_t} = \overline{x}$. Let $\mathbb{S}_t(x) = \overline{H}_t(x)$ for $x \geq x_t^{M_t}$. Stop.

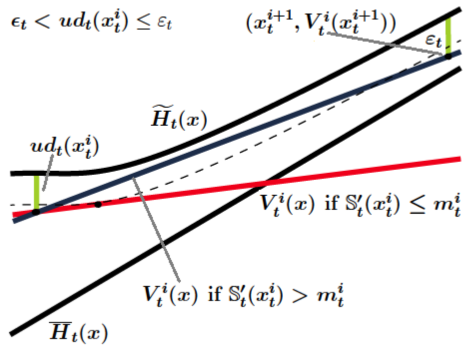

FIGURE 4. Linear spline in cases 2a, 2b.

Lemma 5.1. $\mathbb{S}_t(x)$ *is convex and non-decreasing.*

Proof. Convexity follows from the algorithm above. The key points in the argument are the paragraph following (4.3) and step (2) of the "Knot Creation Algorithm for $x \ge x_t^{N_t}$". Since $\mathbb{S}_t'(x) = 0$ for $x \le x_t^1$ (see (4.1)), the convexity implies monotonicity.

□

Lemma 5.2. *Assuming that* $\overline{H}_t(x) \le v_t(x) \le \widetilde{H}_t(x)$ *for all* x, $\overline{H}_t(x) \le \mathbb{S}_t(x) \le \widetilde{H}_t(x)$ *for all* x.

Proof. In the "Knot Creation Algorithm for $x \ge x_t^{N_t}$", $ud_t(x) > \varepsilon_t$, $dd_t(x) > \varepsilon_t$ and $|\mathbb{S}_t(x) - v_t(x)| \le \epsilon_t$ by Lemma 4.2 so

$$\overline{H}_t(x) < v_t(x) - \varepsilon_t \le \mathbb{S}_t(x) + \epsilon_t - \varepsilon_t \le \mathbb{S}_t(x)$$

and similarly,

$$\mathbb{S}_t(x) \le \mathbb{S}_t(x) - \epsilon_t + \varepsilon_t < v_t(x) + \varepsilon_t < \widetilde{H}_t(x).$$

for $x \le x_t^{N_t}$.

For $x \geq x_t^{N_t}$ in cases 1a, 1b, 2a, and 4, $\mathbb{S}_t(x) \leq \widetilde{H}_t(x)$ since $\mathbb{S}_t(x)$ is tangent to $\widetilde{H}_t(x)$ or some translate down. In case 2b, $\mathbb{S}_t(x_t^i) = v_t(x_t^i)$ and $v_t'(x) \geq v_t'(x_t^i) = \mathbb{S}_t'(x)$ by convexity of $v_t(x)$ so $\mathbb{S}_t(x) \leq v_t(x) \leq \widetilde{H}_t(x)$ by Theorem 3.7 and by linearity of $\mathbb{S}_t(x)$. As in step (2), if $\mathbb{S}_t(x) < \overline{H}_t(x)$, we redefine $\mathbb{S}_t(x)$ so $\mathbb{S}_t(x) \geq \overline{H}_t(x)$ for $x \geq x_t^{M_t}$.

□

Theorem 5.3. $|\mathbb{S}_t(x) - \nu_t(x)| \leq \varepsilon_t$ *for all* $x \geq x_t^{N_t}$.

Proof. We begin with case 1a, in which $\widetilde{H}_t(x_t^i) - \mathbb{S}_t(x_t^i) \leq \epsilon_t$, and $\mathbb{S}_t(x_t^i) = v_t(x_t^i)$. Let $x_t^i \leq x \leq x_t^{i+1}$. By (9.2) and by the triangle inequality

$$|\widetilde{H}_t(x_t^i) - \nu_t(x_t^i)| \leq |\widetilde{H}_t(x_t^i) - v_t(x_t^i)| + |v_t(x_t^i) - \nu_t(x_t^i)| \leq \epsilon_t + \varepsilon_{t+1} = \varepsilon_t. \qquad (5.3)$$

By Theorem 3.3 $\widetilde{H}_t'(x) \leq \nu_t'(x)$ so the distance between $\widetilde{H}_t(x)$ and $\nu_t(x)$ will decrease as x grows. Since $\widetilde{H}_t(x_t^i) - \varepsilon_t \leq v_t(x_t^i) - \varepsilon_{t+1} \leq \nu_t(x_t^i)$, $\widetilde{H}_t(x) - \varepsilon_t$ is a lower bound on $\nu_t(x)$ as shown in Figure 3. Because $\widetilde{H}_t(x)$ is an upper bound on $\nu_t(x)$ and because $\mathbb{S}_t(x) = V_t^i(x)$ is between $\widetilde{H}_t(x)$ and $\widetilde{H}_t(x) - \varepsilon_t$, $|\mathbb{S}_t(x) - \nu_t(x)| \leq \varepsilon_t$ for $x_t^i \leq x \leq x_t^{i+1}$.

In case 2a we have $\epsilon_t < ud_t(x_t^i) \leq \varepsilon_t$. Let $x_t^i \leq x \leq x_t^{i+1}$. By an argument similar to the one used in case 1a, $\widetilde{H}_t(x) - ud_t(x_t^i) - \varepsilon_{t+1} \leq \nu_t(x) \leq \widetilde{H}_t(x)$. By construction, $\widetilde{H}_t(\widetilde{x}) - V_t^i(\widetilde{x}) = ud_t(x_t^i) - \epsilon_t = \widetilde{H}_t(x_t^i) - V_t^i(x_t^i) - \epsilon_t$. Since $\widetilde{H}_t(x)$ is convex and $x_t^i < \widetilde{x}$, we must have $\widetilde{H}_t'(x_t^i) \leq (V_t^i)'(x_t^i) = m_t^i$. Hence $\widetilde{H}_t(x) - V_t^i(x)$ decreases until the point $(\widetilde{x}, V_t^i(\widetilde{x}))$. Thereafter $\widetilde{H}_t(x) - V_t^i(x)$ increases, but this iteration of our knot creation algorithm stops when $\widetilde{H}_t(x) - V_t^i(x) = \varepsilon_t$, so $\widetilde{H}_t(x) - V_t^i(x) \leq \varepsilon_t$ for $x_t^i \leq x \leq x_t^{i+1}$. On this interval, the gap between $V_t^i(x)$ and the lower bound $\widetilde{H}_t(x) - ud_t(x_t^i) - \varepsilon_{t+1}$ on $\nu_t(x)$ expands from ε_{t+1} at x_t^i to ε_t at $\widetilde{x}$ and then decreases. Therefore $\mathbb{S}_t(x)$ is within ε_t of the bounds of $\nu_t(x)$ (see Figure 5(a)), so it follows that $|\mathbb{S}_t(x) - \nu_t(x)| \leq \varepsilon_t$ for $x_t^i \leq x \leq x_t^{i+1}$.

In case 2b we have $\epsilon_t < ud_t(x_t^i) \leq \varepsilon_t$ and $(V_t^i)'(x_t^i) = v_t'(x_t^i)$. As before, let $x_t^i \leq x \leq x_t^{i+1}$. We now have a different lower bound on $\nu_t(x)$. Because $v_t(x)$ is convex $v_t'(x) \geq v_t'(x_t^i)$. Combined with the fact that $|v_t(x) - \nu_t(x)| \leq \varepsilon_{t+1}$, we know that the line with slope $v_t'(x_t^i)$ that passes through the point $(x_t^i, v_t(x_t^i) - \varepsilon_{t+1})$ as shown in Figure 5(b) is a lower bound on $\nu_t(x)$. Keeping $V_t^i(x)$ parallel to this lower bound as we do in case 2b, and because $\widetilde{H}_t(x_t^{i+1}) - V_t^i(x_t^{i+1}) = \varepsilon_t$, we have $|\mathbb{S}_t(x) - \nu_t(x)| = |V_t^i(x) - \nu_t(x)| \leq \varepsilon_t$.

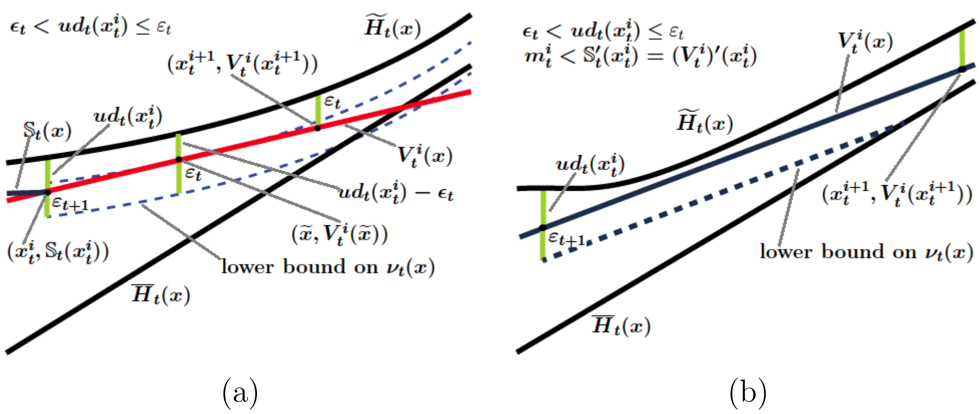

FIGURE 5. Linear spline in (a) case 2a and (b) case 2b.

Consider case 1b in which $\mathbb{S}_t(x_t^i) - \overline{H}_t(x_t^i) \leq \varepsilon_t$ (Figure 6). If $\mathbb{S}_t(x_t^i) + \varepsilon_{t+1} \geq \widetilde{H}_t(x_t^i)$ set $x' = x_t^i$. Otherwise there is a line that passes through $(x_t^i, \mathbb{S}_t(x_t^i) + \varepsilon_{t+1})$ and is tangent to $\widetilde{H}_t(x)$ at $(x', \widetilde{H}_t(x'))$, where $x_t^i \leq x' \leq \widetilde{x}$. We know by Theorem 3.1 that $\nu_t(x) \leq \widetilde{H}_t(x)$, that $\nu_t(x)$ is convex and that $\nu_t(x) - \mathbb{S}_t(x) \leq \varepsilon_{t+1}$. Hence the line segment that connects $(x_t^i, \mathbb{S}_t(x_t^i) + \varepsilon_{t+1})$ and $(x', \widetilde{H}_t(x'))$ is an upper bound on $\nu_t(x)$. For $x \geq x'$, $\widetilde{H}_t(x)$ is an upper bound on $\nu_t(x)$. The slope of the line segment is less than m_t^i in step (1), so the distance between the upper bound on $\nu_t(x)$ and $V_t^i(x)$ is decreasing for $x_t^i \leq x \leq \widetilde{x}$. Hence $V_t^i(x)$ stays within ε_t of the upper bound on $\nu_t(x)$. By Theorem 3.2 $\nu_t(x) \geq \overline{H}_t(x)$, so $|\mathbb{S}_t(x) - \nu_t(x)| = |V_t^i(x) - \nu_t(x)| \leq \varepsilon_t$ for $x_t^i \leq x < x_t^{i+1}$.

We consider case 3 and assume that $x \geq \overline{x} = x_t^{M_t}$. We claim that $\varepsilon_t \geq \widetilde{H}(\overline{x}) - \overline{H}(\overline{x})$. This is automatic if we reach case 3 from any of cases 1a, 2a, 2b. If we reach case 3 from case 1b then if $x_t^{i+1} = x_t^{M_t}$, Theorem 3.7 says $m_t^i = \widetilde{H}_t'(\widetilde{x}) \leq \overline{H}_t'(\widetilde{x}) = \overline{H}_t'(x)$ for all $x \in [x_t^i, \overline{x}]$, so by Remark 3.5

$$\varepsilon_t \geq V_t^i(x_t^i) - \overline{H}_t(x_t^i) \geq V_t^i(\widetilde{x}) - \overline{H}_t(\widetilde{x}) = \widetilde{H}_t(\widetilde{x}) - \overline{H}_t(\widetilde{x}) \geq \widetilde{H}_t(\overline{x}) - V_t^i(\overline{x}).$$

We define $\mathbb{S}_t(x) = \overline{H}_t(x)$ for all $x \geq x_t^{M_t}$. Because $\widetilde{H}_t(x) \geq \nu_t(x) \geq \overline{H}_t(x)$, $\mathbb{S}_t(x)$ stays within ε_t of $\nu_t(x)$.

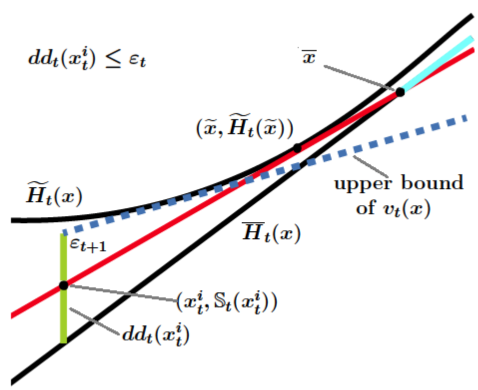

FIGURE 6. Linear spline in case 1b.

□

6. SUMMARY

Often times the details of the proofs tend to obfuscate the simple elegance of the method. We intend to summarize the algorithm for creating the spline to clarify the proceedure and refer the interested reader to the proofs in previous sections and the Appendix of the paper.

Assuming that we know the demand distribution in each period, we know the analytic representation of the optimal cost function $\nu_T(x)$ in period T, whose minimum

R_T is the solution to the myopic policy and can be computed. Thus the approximate optimal cost function $v_T(x)$ is equal to $\nu_T(x)$ and the first knot is $x_T^1 = R_T$. From here we make an educated guess about the location of the next knot x_T^2. We then compute the maximum possible error $|\mathbb{S}_T(x) - \nu_T(x)|$ that can occur for $x \in [x_T^1, x_T^2]$. This computation is facilitated by transforming the interval $[x_T^1, x_T^2]$ to the unit domain so we can use a standardized quadratic spline with fixed endpoints at $(0, 0)$ and $(1, 0)$.

If the computed error bound is greater than the set limit ϵ_T, then the knot x_T^2 is not close enough to x_T^1. We try again with a closer knot. If the bounds are so close together that the maximum possible error is less than one fourth of ϵ_T, then we try to move the knot x_T^2 farther away from x_T^1 so that in the end we use fewer knots, which reduces the number of terms in the formula for the spline. If x_T^2 has an acceptable maximum error, then we fix the knot x_T^2 and repeat the process for the knot x_T^3, and so on. We let the spline be equal to $\nu_T(R_t)$ for all x less than R_T and for large x we let the spline be equal to the lower bound $\overline{H}_T(x)$ when the lower bound and the spline intersect. We choose our quadratic splines so that the spline is continuous, convex and has the same upper and lower bound functions as $\nu_T(x)$.

After approximating $\nu_T(x)$ with the spline function $\mathbb{S}_T(x)$, we can evaluate $\mathcal{H}_{T-1}(x)$, $\mathcal{H}'_{T-1}(x)$, and $\mathcal{H}''_{T-1}(x)$ via (4.9)-(4.11). To $\mathbb{E}[\mathbb{S}_T(x - D_{T-1})]$ we add the single period cost function $g_{T-1}(x)$ to obtain $\mathcal{H}_{T-1}(x)$. We minimize this function to obtain $\mathcal{R}_{T-1}$ and define $v_{T-1}(x)$. Then we repeat the above proceedure by defining the spline approximation $\mathbb{S}_{T-1}(x)$ to $v_{T-1}(x)$.

The error bounds accumulate from one time period to the next. Reducing the error limits counterbalances the error accumulation, but forces us to use more knots and increases the run time. This algorithm for creating spline approximations is still fast.

The computations tell us what the optimal inventory level is in every time period and what the optimal costs are.

7. COMPUTATIONS

In this section, we describe the Standard Case that we use to test our software. Then we compare the run time and accuracy of our results.

We let the mean demand be the following function of $t \in [1, 7/2]$ (see Figure 7):

$$f(t) = 3 - (\eta - 3)(3\sin(\pi t) - 2t + 2)/8. \tag{7.1}$$

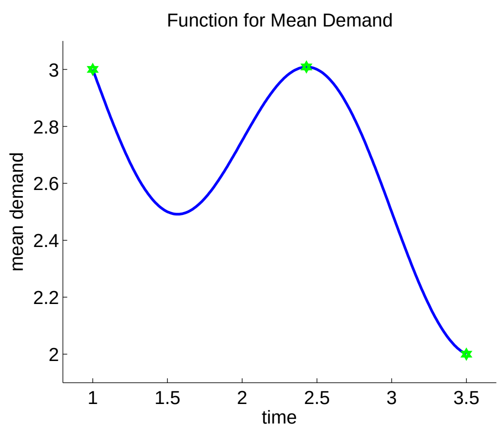

FIGURE 7. Graph of the mean demand function $f(t)$ where $\eta = 2$.

This function has maximum value of 3 at $t = 1$ and decays to $\eta < 3$ by time $t = 7/2$. We obtain mean demands for discrete time periods by uniformly scaling the time interval $1 \leq t \leq 7/2$. This function models the mean demand, so in all cases $\mathbb{E}[D_1] = 3$, demand then dips slightly before recovering, and demand finally crashes to $\mathbb{E}[D_T] = \eta$.

In our Standard Case $T = 10$, $h_t = 1$, $\pi_t = 10$, $\epsilon_t = .5$, and $\eta = 2$. We also let $\sigma_t^2/\mu_t = 3/4$ for every time period t. We vary one of these parameters at a time. The table headings are $v_t(\mathcal{R}_1)$, which is our approximation of the optimal cost over all time periods of each dynamic program with the specific parameters; $\mathcal{R}_1$, the optimal inventory in period 1; and "knots" – the number of knots used in the first time period.

We did our computations in MATLAB on a 32-bit Windows 7 PC with an AMD Athlon(tm) 64 X2 Dual Core Processor 4600+ 2.40 GHz CPU.

π_t/h_t	$v_t(\mathcal{R}_t)$	$\mathcal{R}_t$	knots	time (sec)
5	21.10	4.4296	38	7.5095
10	25.32	4.9821	40	7.4958
15	27.69	5.2809	40	7.4597
20	29.31	5.4825	40	7.4326

Cutting the error limit by ninety nine percent to $\epsilon_t = .01$ in the table below increased our run time by twenty eight percent. It also increased the number of knots to more than nine times as many.

ϵ_t	$v_t(\mathcal{R}_t)$	$\mathcal{R}_t$	knots	time (sec)
.9	25.34	4.9793	9	6.9293
.5	25.33	4.9798	12	6.8789
.1	25.32	4.9820	26	7.2503
.01	25.32	4.9822	86	8.8372

We maintained a constant ratio for the variance over the mean, but as this value grew, the absolute size of the standard deviation also grew. When the standard deviation grew larger than 1 over all time periods t ($\sigma_t^2/\mu_t = 3$ with $2 \leq \mu_t \leq 3$), the expected cost quickly rose to \$52.58 as demand became more volatile.

σ_t^2/μ_t	$v_t(\mathcal{R}_t)$	$\mathcal{R}_t$	knots	time (sec)
3	52.58	6.6979	42	4.0157
3/4	25.32	4.9821	40	7.4155
1/3	16.82	4.3340	37	12.2966
3/16	12.61	4.0013	38	18.6921

In this table, smaller values of η mean average demand is lower in all time periods except the first. The optimal inventory level remains constant over the change in demand, but the cost decreases as demand decreases because backorders are less likely and the backorder cost π_t is very large compared to our holding cost h_t.

η	$v_t(\mathcal{R}_t)$	$\mathcal{R}_t$	knots	time (sec)
.5	22.55	4.9355	33	9.4734
1	23.46	4.9568	36	8.1355
1.5	24.41	4.9718	38	7.7085
2	25.32	4.9821	40	7.5257

The number of time periods and the time are directly correlated as shown in the next table.

T	$v_t(\mathcal{R}_t)$	$\mathcal{R}_t$	knots	time (sec)
5	12.66	4.9769	20	3.5368
10	25.32	4.9821	40	7.8773
15	34.00	4.9866	56	16.1837
20	50.68	4.9888	72	19.1496

As the reader can clearly see, our program is predictable and robust. We hope to soon use a MATLAB compiler for even faster execution times.

8. CONCLUSION

Our algorithm for approximating the optimal cost function of the single-product dynamic inventory problem is both fast and accurate. Approximation by quadratic splines is simple and computationally effective and bypasses having to compute the analytic representation of $\nu_t(x)$. With this software, we hope companies will have better means of regulating their inventory optimally.

9. APPENDIX

We begin by proving some of the Theorems in Sections 2 and 3. We derive the parameters for V_0, V_1 and V_2 in Sections 9.1 - 9.3. In Section 9.4 we prove Lemma

4.1. The formulas for the transformation and inverse transformation are in Section 9.5. In Section 9.6 we prove a well known result that simplifies our computations.

Theorem 2.1. *If we create* $\mathbb{S}_t(x)$ *such that* $|\mathbb{S}_t(x) - \upsilon_t(x)| \leq \epsilon_t$ *for all* $x \in (-\infty, x_t^{N_t}]$, *then* $|\mathbb{S}_t(x) - \nu_t(x)| \leq \varepsilon_t$ *for all* x *and all* t. *Furthermore* $|\mathcal{H}_t(x) - H_t(x)| \leq \varepsilon_{t+1}$.

Proof. We proceed by induction. Note that $H_T(x) = \mathcal{H}_T(x)$ so $\nu_T(x) = \upsilon_T(x)$. Thus if $|\mathbb{S}_T(x) - \upsilon_T(x)| \leq \epsilon_T$ for all $x \in (-\infty, x_t^{N_t}]$ and by Theorem 5.3 $|\mathbb{S}_T(x) - \nu_T(x)| \leq \epsilon_T$ for $x \geq x_t^{N_t}$, then $|\mathbb{S}_T(x) - \nu_T(x)| \leq \epsilon_T$ for all x.

In period $t < T$ assume that $|\mathbb{S}_{t+1}(x) - \nu_{t+1}(x)| \leq \varepsilon_{t+1}$ for all x by the induction hypothesis. Then

$$\begin{aligned} |\mathcal{H}_t(x) - H_t(x)| &= |g_t(x) + \mathbb{E}[\mathbb{S}_{t+1}(x - D_t)] - (g_t(x) + \mathbb{E}[\nu_{t+1}(x - D_t)])| \\ &= |\mathbb{E}[\mathbb{S}_{t+1}(x - D_t) - \nu_{t+1}(x - D_t)]| \leq |\mathbb{E}[\varepsilon_{t+1}]| = \varepsilon_{t+1}. \end{aligned} \tag{9.1}$$

We now show that $|\upsilon_t(x) - \nu_t(x)| \leq \varepsilon_{t+1}$. Because we use properties that apply to both $(\upsilon_t(x), \mathcal{H}_t(x), \mathcal{R}_t)$ and to $(\nu_t(x), H_t(x), R_t)$, we assume WLOG that $R_t \geq \mathcal{R}_t$. We have for all $x \geq \max\{\mathcal{R}_t, R_t\}$, $|\upsilon_t(x) - \nu_t(x)| = |\mathcal{H}_t(x) - H_t(x)| \leq \varepsilon_{t+1}$ using (2.4) and (2.5). Then for all $\mathcal{R}_t \leq x \leq R_t$,

$$\begin{aligned} \nu_t(x) + \varepsilon_{t+1} &= \nu_t(R_t) + \varepsilon_{t+1} = H_t(R_t) + \varepsilon_{t+1} \geq \mathcal{H}_t(R_t) = \upsilon_t(R_t) \geq \upsilon_t(x) \\ &\geq \upsilon_t(\mathcal{R}_t) = \mathcal{H}_t(\mathcal{R}_t) \geq H_t(\mathcal{R}_t) - \varepsilon_{t+1} \geq H_t(R_t) - \varepsilon_{t+1} = \nu_t(x) - \varepsilon_{t+1}. \end{aligned}$$

We have proved

$$|\upsilon_t(x) - \nu_t(x)| \leq \varepsilon_{t+1} \text{ for all } x. \tag{9.2}$$

By the discussion in Section 4.1, $\mathbb{S}_t(x) = \upsilon_t(x)$ for $x \in (-\infty, \mathcal{R}_t]$. In Lemma 4.2 we will prove that $|\mathbb{S}_t(x) - \upsilon_t(x)| \leq \epsilon_t$ for all $x \in (\mathcal{R}_t, x_t^{N_t}]$. Then by (9.2) and by the triangle inequality

$$|\mathbb{S}_t(x) - \nu_t(x)| \leq |\mathbb{S}_t(x) - v_t(x)| + |v_t(x) - \nu_t(x)| \leq \epsilon_t + \varepsilon_{t+1} = \varepsilon_t$$

for all $x \in (-\infty, x_t^{N_t}]$. Combining this with Theorem 5.3, we have $|\mathbb{S}_t(x) - \nu_t(x)| \leq \varepsilon_t$ for all x. □

Theorem 3.1. *Let* $\mu_{[t,j]} = \mathbb{E}[D_{[t,j]}]$. *Then*

$$H_t(x) \leq \sum_{j=t}^{T} \left(h_j(x - \mu_{[t,j]}) + (h_j + \pi_j)\mathbb{E}[(D_{[t,j]} - x)^+] \right) = \widetilde{H}_t(x) \text{ for all } x. \tag{9.3}$$

Proof. Formally, we proceed by induction. Let $t = T$. Since the salvage value is zero, $\nu_{T+1}(x - D_t) = v_{T+1}(x - D_t) = 0$. So

$$\begin{aligned} H_T(x) &= g_T(x) + \mathbb{E}[\nu_{T+1}(x - D_T)] \\ &= h_T(x - \mu_T) + (h_T + \pi_T)\mathbb{E}[(D_T - x)^+] = \widetilde{H}_T(x). \end{aligned}$$

Assume (9.3) for time $t + 1$. Then by (2.3), (2.4), the fact that R_{t+1} minimizes $H_{t+1}(x)$ and by (3.2) we have

$$\begin{aligned} H_t(x) &= g_t(x) + \mathbb{E}[\nu_{t+1}(x - D_t)] = g_t(x) + \mathbb{E}[H_{t+1}((x - D_t) \vee R_{t+1})] \\ &\leq g_t(x) + \mathbb{E}[H_{t+1}(x - D_t)] \leq g_t(x) + \mathbb{E}[\widetilde{H}_{t+1}(x - D_t)] = \widetilde{H}_t(x). \end{aligned}$$

□

Theorem 3.2.

$$H_t(x) \geq \sum_{j=t}^{T} h_j(x - \mu_{[t,j]}) = \overline{H}_t(x). \tag{9.4}$$

Proof. For $t = T$, the result follows from (3.4) since $H_T(x) = g_T(x)$. Assume (9.4) for time $t+1$. Since $\overline{H}_{t+1}(x)$ is increasing, in period t

$$\begin{aligned} H_t(x) &= g_t(x) + \mathbb{E}[\nu_{t+1}(x - D_t)] = g_t(x) + \mathbb{E}[H_{t+1}((x - D_t) \vee R_{t+1})] \\ &\geq g_t(x) + \mathbb{E}[\overline{H}_{t+1}((x - D_t) \vee R_{t+1})] \geq g_t(x) + \mathbb{E}[\overline{H}_{t+1}(x - D_t)] \qquad (9.5) \\ &\geq h_t(x - \mu_t) + \sum_{j=t+1}^{T} h_j(x - \mu_{[t,j]}) = \overline{H}_t(x). \end{aligned}$$

Thus $\overline{H}_t(x)$ is a lower bound on $H_t(x)$ in every period t. $\square$

9.1. V_0 **Parameter Derivation and Error Analysis.** We maximize $V_0(x') - S_t^i(x')$. Second order conditions require that $V_0''(x'_*) \leq (S_t^i)''(x'_*) = 2\lambda$. By the Mean Value Theorem $(\overline{v}_t^i)'(1) - (\overline{v}_t^i)'(0) = (\overline{v}_t^i)''(x'_\star)$ for some $x'_\star \in (0, 1)$ so

$$2\lambda = 2(-(\overline{v}_t^i)'(0) \wedge (\overline{v}_t^i)'(1)) \leq (\overline{v}_t^i)'(1) + (-(\overline{v}_t^i)'(0)) = (\overline{v}_t^i)''(x'_\star) \leq 2B_i. \qquad (9.6)$$

Thus $V_0''(x') = 2B_i \geq 2\lambda$ for $x' < \xi$ and $x' > \omega$, so WLOG we assume that $\xi \leq x' \leq \omega$. We equate $V_0'(x')$ and $(S_t^i)'(x')$ to find the largest error. For $\xi \leq x' \leq \omega$,

$$V_0'(x'_*) - (S_t^i)'(x'_*) = 2B_i(\omega - 1) + \lambda - 2\lambda x'_* + \lambda \qquad \Rightarrow x'_* = 1 + \frac{B_i(\omega - 1)}{\lambda}.$$

Setting $x' = 1 + \frac{B_i(\omega-1)}{\lambda}$ we obtain

$$\begin{aligned} V_0\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right) - S_t^i\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right) &= (2B_i(\omega - 1) + \lambda)\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right) \\ &+ B_i(1 - \omega^2) - \lambda - \lambda\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right)^2 + \lambda\left(1 + \frac{B_i(\omega - 1)}{\lambda}\right) \\ = \frac{2B_i^2(\omega - 1)^2}{\lambda} - B_i(\omega - 1)^2 &= B_i(\omega - 1)^2\left(\frac{B_i}{\lambda} - 1\right). \end{aligned}$$

The error is nonnegative by (9.6).

9.2. V_1 **Parameter Derivation and Error Analysis.** We now derive the parameters of $V_1(x')$. We note that

$$\lambda = V_1'(\gamma) = -\Theta + 2B_i\gamma \Rightarrow \gamma = (\lambda + \Theta)/(2B_i). \tag{9.7}$$

Clearly, $S_t^i(x') - V_1(x')$ is biggest when $0 = S_i'(x_*') - V_1'(x_*') = 2\lambda x_*' - \lambda - [2B_i x_*' - \Theta]$ or where $x_*' = \frac{\Theta-\lambda}{2(B_i-\lambda)}$. So

$$\begin{aligned} S_t^i(x_*') - V_1(x_*') &= \lambda x_*'(x_*' - 1) - [B_i(x_*')^2 - \Theta x_*'] = (\Theta - \lambda)x_*' - (B_i - \lambda)(x_*')^2 \\ &= \frac{(\Theta - \lambda)^2}{4(B_i - \lambda)} = \frac{(2\sqrt{B_i\lambda} - 2\lambda)^2}{4(B_i - \lambda)} = \frac{\lambda(\sqrt{B_i} - \sqrt{\lambda})^2}{B_i - \lambda} = \frac{\lambda(\sqrt{B_i} - \sqrt{\lambda})}{\sqrt{B_i} + \sqrt{\lambda}}. \end{aligned} \tag{9.8}$$

The error applies if $V_1'(0) = -\Theta \geq -\Lambda = V_2'(0)$, which is true if and only if $\big(\lambda+\Lambda\big)^2 \geq 4B_i\lambda$.

9.3. V_2 **Parameter Derivation and Error Analysis.** The parameters for $V_2(x')$ and the corresponding error are derived in a manner that is conceptually similar. Since $\overline{v}_t^i(x')$ and $S_t^i(x')$ are smooth in $[0, 1]$ and are equal at the endpoints,

$$-\Lambda + 2(\beta - \alpha)B_i = \lambda \iff \beta - \alpha = \frac{\lambda + \Lambda}{2B_i}. \tag{9.9}$$

Because $V_2(x')$ is continuous at β, $-\Lambda\beta + B_i(\beta - \alpha)^2 = \lambda(\beta - 1)$, so

$$\beta = \frac{[B_i(\beta - \alpha)^2 + \lambda]}{\lambda + \Lambda} = \frac{\lambda}{\lambda + \Lambda} + \frac{\lambda + \Lambda}{4B_i}, \text{ and} \tag{9.10}$$

$$\begin{aligned} \alpha &= \beta - \frac{\lambda + \Lambda}{2B_i} = \frac{\lambda}{\lambda + \Lambda} - \frac{\lambda + \Lambda}{4B_i} \\ &= \frac{\big(\lambda + \Lambda\big) + \big(\lambda - \Lambda\big)}{2(\lambda + \Lambda)} - \frac{\lambda + \Lambda}{4B_i} = \frac{\lambda - \Lambda}{2(\lambda + \Lambda)} + \frac{1}{2} - \frac{\lambda + \Lambda}{4B_i} \qquad (9.11) \\ &= \frac{\lambda - \Lambda}{2(\lambda + \Lambda)} + \left(\frac{\lambda + \Lambda}{2B_i}\right)\left(\frac{B_i}{\lambda + \Lambda} - \frac{1}{2}\right). \qquad (9.12) \end{aligned}$$

Substituting $\frac{\lambda+\Lambda}{4B_i} = \frac{1}{4B_i}(\lambda - \Lambda + 2\Lambda)$ into (9.11) we get

$$\alpha = \frac{1}{2}\left(1 - \frac{\Lambda}{B_i} + (\lambda - \Lambda)\left(\frac{1}{\lambda+\Lambda} - \frac{1}{2B_i}\right)\right). \tag{9.13}$$

The maximum error between $S_t^i(x')$ and $V_2(x')$ occurs at x'_* if the derivatives are equal at x'_*. Thus $S_i'(x'_*) = V_2'(x'_*) \Rightarrow 2\lambda x'_* - \lambda = 2B_i(x'_* - \alpha) - \Lambda$ which is true if and only if $x'_* = \frac{(\Lambda-\lambda+2B_i\alpha)}{2(B_i-\lambda)}$. Then

$$\begin{aligned} S_t^i(x'_*) - V_2(x'_*) &= \lambda((x'_*)^2 - x'_*) - (B_i(x'_* - \alpha)^2 - \Lambda x'_*) \\ &= (\lambda - B_i)(x'_*)^2 + (\Lambda + 2B_i\alpha - \lambda)x'_* - B_i\alpha^2 = \frac{(\Lambda - \lambda + 2B_i\alpha)^2}{4(B_i - \lambda)} - B_i\alpha^2. \end{aligned}$$

Since $V_2'(\alpha) = V_2'(0) = S_i'(0) < S_i'(1) = V_2'(1) = V_2'(\beta)$ and $V_2(0) = S_t^i(0) = S_t^i(1) = V_2(1) = 0$, there exists an $x' \in [\alpha, \beta]$ such that $V_2'(x') = S_i'(x')$. This error is valid if $0 \le \alpha, \beta \le 1$, which is true if and only if $4B_i\lambda \ge (\lambda + \Lambda)^2$.

$$\begin{aligned} S_t^i(x'_*) - V_2(x'_*) &= \frac{1}{4(B_i - \lambda)}\left(\Lambda - \lambda + B_i\left[1 - \frac{\Lambda}{B_i} + (\lambda - \Lambda)\left(\frac{1}{\lambda+\Lambda} - \frac{1}{2B_i}\right)\right]\right)^2 \\ &\quad - B_i\left(\frac{\lambda - \Lambda}{2(\lambda+\Lambda)} + \left(\frac{\lambda+\Lambda}{2B_i}\right)\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)\right)^2 \\ &= \frac{1}{4(B_i - \lambda)}\left[B_i - \lambda + (\lambda - \Lambda)\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)\right]^2 - \frac{B_i(\Lambda - \lambda)^2}{4(\lambda+\Lambda)^2} \\ &\quad - \frac{\lambda - \Lambda}{2}\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right) - \frac{(\lambda+\Lambda)^2}{4B_i}\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)^2 \\ &= \frac{B_i - \lambda}{4} + \frac{\lambda - \Lambda}{2}\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right) + \frac{(\Lambda - \lambda)^2}{4(B_i - \lambda)}\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)^2 \\ &\quad - \frac{B_i(\Lambda - \lambda)^2}{4(\lambda+\Lambda)^2} - \frac{\lambda - \Lambda}{2}\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right) - \frac{(\lambda+\Lambda)^2}{4B_i}\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)^2. \end{aligned}$$

The resulting error bound is

$$\frac{1}{4}\left[B_i - \lambda - \frac{B_i(\Lambda-\lambda)^2}{(\lambda+\Lambda)^2} + \left(\frac{(\Lambda-\lambda)^2}{B_i-\lambda} - \frac{(\lambda+\Lambda)^2}{B_i}\right)\left(\frac{B_i}{\lambda+\Lambda} - \frac{1}{2}\right)^2\right]. \tag{9.14}$$

9.4. **Proof of Lemma 4.1.**

Lemma 4.1. *If* $(\lambda+\Lambda)^2 \geq 4B_i\lambda$ *then* $V_0(x') \geq \overline{v}_t^i(x') \geq V_1(x')$ *for* $x' \in [0,1]$. *Otherwise* $V_0(x') \geq \overline{v}_t^i(x') \geq V_2(x')$.

Proof. We will prove that $V_0(x')$ is the best upper bound and $V_1(x')$ and $V_2(x')$ are the best lower bounds of any function $f(x')$ with the properties

$$f(0) = f(1) = 0, f'(0) = -\Lambda, f'(1) = \lambda, f(x') \text{ is convex, and } f''(x') \leq 2B_i. \tag{9.15}$$

$V_0'(0) = -\Lambda = f'(0)$, $V_0'(1) = \lambda = f'(1)$ and $V_0(x')$ has the maximum second derivative value $2B_i$ for x' near the endpoints 0 and 1 so $V_0(x)$ forms two parabolas that are upper bounds on $f(x)$ for $0 \leq x' \leq \xi$ and $\omega \leq x' \leq 1$. The line between $V_0(\xi)$ and $V_0(\omega)$ is an upper bound on $f(x')$ because $f(x')$ is convex. Thus $V_0(x')$ is an upper bound of $f(x')$.

By convexity $0 \leq f''(x')$ and $V_2''(x') = 0$ for $1 \leq x'\alpha, \beta \leq x' \leq 1$. Also since $f(x')$ is convex, $V_2'(x') = V_2'(0) = -\Lambda = f'(0)$ for $0 \leq x' \leq \alpha$ and $V_2'(x') = V_2'(1) = \lambda = f'(1)$ for $\beta \leq x' \leq 1$, $V_2(x')$ is a lower bound of $f(x')$ for $0 \leq x' \leq \alpha$, $\beta \leq x' \leq 1$. Suppose $f(x_*') < V_2(x_*')$, $\alpha < x_*' < \beta$. If $f'(x_*') < V_2(x_*')$, then $f'(\beta) < V_2'(\beta) = \lambda$ and $f(\beta) < V_2(\beta)$ since $f''(x') \leq V_2(x') = 2B_i$ for $\alpha < x' < \beta$. Thus $f(x')$ has slope strictly less than λ when it intersects the line with slope λ that goes through $(1,0)$. Because of the finite bound on the second derivative of $f(x')$, this contradicts the fact that $f(1) = 0$. If $f'(x_*') > V_2(x_*')$, then $f'(\alpha) < V_2'(\alpha) = -\Lambda$ and $f(\alpha) > V_2(\alpha)$ since $f''(x') \leq V_2(x') = 2B_i$ for $\alpha < x' < \beta$. Thus $f(x')$ has slope strictly greater than $-\Lambda$ when it intersects the line with slope $-\Lambda$ that goes through $(0,0)$. Because of the finite bound on the second derivative of $f(x')$, but we need $f(0)=0$, we have a contradiction. Thus $f(x') \geq V_2(x')$ for $\alpha < x' < \beta$ so $V_2(x')$ is a lower bound of $f(x')$ for $0 \leq x' \leq 1$.

If the second derivative bound is constrictive, then $V_1(x')$ is a lower bound of $f(x')$ by the same argument as in the previous paragraph.

Since $\overline{v}_t^i(x')$ satisfies properties (9.15), the Lemma is proved. $\square$

9.5. **Transformation and Inverse Transformation.** Assume $v_t(x)$ passes through the points $(x_t^i, v_t(x_t^i))$ and $(x_t^{i+1}, v_t(x_t^{i+1}))$ and that $v_t''(x) \leq 2B^i$ for $x \in [x_t^i, x_t^{i+1}]$ (see Section 4.2). In the original domain we view $v_t(x_t^i)$ as a function of x_t^i. To facilitate the error analysis for a given t and i we define the unit domain by mapping $(x_t^i, v_t(x_t^i))$ and $(x_t^{i+1}, v_t(x_t^{i+1}))$ into $(0,0)$ and $(1,0)$ respectively using the transformation

$$x' = \frac{x - x_t^i}{x_t^{i+1} - x_t^i}, \qquad m_t^i = \frac{v_t(x_t^{i+1}) - v_t(x_t^i)}{x_t^{i+1} - x_t^i}, \qquad \text{and } y' = y - m_t^i x - (v_t(x_t^i) - m_t^i x_t^i). \tag{9.16}$$

In the unit domain we view y' as a function of x'. The functions $v_t(x)$ and $\mathbb{S}_t(x)$ from the original domain map into $S_t^i(x)$ and $\overline{v}_t^i(x)$ in the unit domain. Thus

$$v_t(x) = \overline{v}_t^i\left(\frac{x - x_t^i}{x_t^{i+1} - x_t^i}\right) + v_t(x_t^i) + m_t^i(x - x_t^i), \tag{9.17}$$

$$\overline{v}_t^i(x') = v_t(x_t^i + x'(x_t^{i+1} - x_t^i)) - v_t(x_t^i) - m_t^i x'(x_t^{i+1} - x_t^i), \tag{9.18}$$

and similar expressions hold for $\mathbb{S}_t(x)$ and $S_t^i(x')$.

We need to do the inverse of the transformation of (9.16). Using (9.16) to map $y' = S_t^i(x') = \lambda x'(x'-1)$ into the initial domain we get

$$\mathbb{S}_t(x) - m_t^i x - (v_t(x_t^i) - m_t^i x_t^i) = \lambda \left(\frac{x - x_t^i}{x_t^{i+1} - x_t^i} \right) \left(\frac{x - x_t^i}{x_t^{i+1} - x_t^i} - 1 \right)$$

for $x \in [x_t^i, x_t^{i+1}]$. Let

$$A^i = \frac{\lambda}{(x_t^{i+1} - x_t^i)^2}. \tag{9.19}$$

Hence,

$$\mathbb{S}_t(x) = v_t(x_t^i) + m_t^i(x - x_t^i) + A^i(x - x_t^i)(x - x_t^{i+1}), \tag{9.20a}$$

$$\mathbb{S}_t'(x) = m_t^i + A^i(2x - x_t^i - x_t^{i+1}), \tag{9.20b}$$

$$\mathbb{S}_t''(x) = 2A^i. \tag{9.20c}$$

We rectify the expressions for $\mathbb{S}_t(x)$ in (4.1) and (9.20a) by choosing c_t^i and d_t^i appropriately. Let $\mathbf{1}\{x \geq x_t^k\}$ be one for all $k \leq i$ and zero for all $k > i$. The derivatives of (4.1) are

$$\mathbb{S}_t'(x) = \sum_{i=1}^{M_t} (c_t^i \mathbf{1}\{x \geq x_t^i\} + d_t^i (x - x_t^i)\mathbf{1}\{x \geq x_t^i\}), \tag{9.21a}$$

$$\mathbb{S}_t''(x) = \sum_{i=1}^{M_t} d_t^i \mathbf{1}\{x \geq x_t^i\}. \tag{9.21b}$$

If $x \in [x_t^i, x_t^{i+1}]$ and $i \leq M_t - 1$, then we get $\mathbb{S}_t'(x) = \sum_{k=1}^{i}(c_t^k + d_t^k(x - x_t^k))$ and $\mathbb{S}_t''(x) = \sum_{k=1}^{i} d_t^k$. Consequently, the definitions of c_t^i and d_t^i are (4.8). When we define a spline by a straight line, $A^i = 0$.

9.6. **Identities.** To simplify our computations, we repeat the following well-known result.

Proposition 9.1. *Assume that the distribution of the demands is either normal or gamma, and that the demands are independent. If the demands are normally distributed let* $D_t \sim \mathcal{N}(\mu_t, \sigma_t)$, $\mu_{[t,j]} = \sum_{s=t}^{j} \mu_s$, $\sigma_{[t,j]} = \sqrt{\sum_{s=t}^{j} \sigma_s^2}$, *and* $k_{[t,j]} = \frac{x-\mu_{[t,j]}}{\sigma_{[t,j]}}$. *We have*

$$n_{D_{[t,j]}}(x) = \sigma_{[t,j]}(\phi(k_{[t,j]}) - k_{[t,j]}\overline{\Phi}(k_{[t,j]})) \qquad \text{and} \tag{9.22a}$$

$$n^2_{D_t}(x) = \sigma_t^2((k_t^2+1)\overline{\Phi}(k_t) - k_t\phi(k_t)). \tag{9.22b}$$

If D_t *follows a gamma distribution then* $D_t \sim Gamma(\alpha_t, \theta)$, *all demands have the same shape parameter* θ, *and* $\alpha_{[t,j]} = \sum_{s=t}^{j} \alpha_s$. *We have*

$$n_{\alpha_{[t,j]},\theta}(x) = \frac{\alpha_{[t,j]}}{\theta}\overline{F}_{\alpha_{[t,j]}+1,\theta}(x) - x\overline{F}_{\alpha_{[t,j]},\theta}(x) \qquad \text{and} \tag{9.23a}$$

$$n^2_{\alpha_t,\theta}(x) = \frac{\alpha_t(\alpha_t+1)}{\theta^2}\overline{F}_{\alpha_t+2,\theta}(x) - \frac{2\alpha_t x}{\theta}\overline{F}_{\alpha_t+1,\theta}(x) + x^2\overline{F}_{\alpha_t,\theta}(x). \tag{9.23b}$$

Proof. Let $D \sim \mathcal{N}(\mu_t, \sigma_t)$ and let $z = \frac{s-\mu_t}{\sigma_t}$ so $dz = \frac{ds}{\sigma_t}$. Then

$$n_{D_t}(x) = \int_x^\infty (s-x) f_{D_t}(s)ds = \int_{\frac{x-\mu_t}{\sigma_t}}^\infty (\sigma_t z + \mu_t - x)\phi(z)dz$$

$$= \sigma_t \int_{k_t}^\infty (z-k_t)\phi(z)dz = \sigma_t(\phi(k_t) - k_t\overline{\Phi}(k_t)) \tag{9.24}$$

because $z\phi(z) = -\frac{d}{dz}\phi(z)$, and since $z^2\phi(z) = \phi(z) + \frac{d^2}{dz^2}\phi(z)$,

$$n^2_{D_t}(x) = \int_x^\infty (s-x)^2 f_{D_t}(s)ds = \int_{\frac{x-\mu_t}{\sigma_t}}^\infty (\sigma_t z + \mu_t - x)^2\phi(z)dz$$

$$= \sigma_t^2 \int_{k_t}^\infty (z-k_t)^2\phi(z)dz = \sigma_t^2((k_t^2+1)\overline{\Phi}(k_t) - k_t\phi(k_t)). \tag{9.25}$$

If $D_t \sim Gamma(\alpha_t, \theta)$ then for $n \in \mathbb{Z}^+$

$$\begin{aligned}\int_x^\infty s^n g_{\alpha_t,\theta}(s)ds &= \int_x^\infty s^n \frac{\theta^{\alpha_t} s^{\alpha_t-1} e^{-\theta s}}{\Gamma(\alpha_t)} ds \\ &= \int_x^\infty \frac{\alpha_t(\alpha_t+1)\dots(\alpha_t+n-1)}{\theta^n} \frac{\theta^{\alpha_t+n} s^{\alpha_t+n-1} e^{-\theta s}}{\Gamma(\alpha_t+n)} ds \\ &= \frac{\alpha_t(\alpha_t+1)\dots(\alpha_t+n-1)}{\theta^n} \int_x^\infty g_{\alpha_t+n,\theta}(s)ds \\ &= \frac{\alpha_t(\alpha_t+1)\dots(\alpha_t+n-1)}{\theta^n} \overline{F}_{\alpha_t+n,\theta}(x).\end{aligned}$$

Furthermore,

$$n_{\alpha_t,\theta}(x) = \int_x^\infty (s-x) g_{\alpha_t,\theta}(s)ds = \frac{\alpha_t}{\theta}\overline{F}_{\alpha_t+1,\theta}(x) - x\overline{F}_{\alpha_t,\theta}(x) \tag{9.26}$$

$$\begin{aligned}n^2_{\alpha_t,\theta}(x) &= \int_x^\infty (s-x)^2 g_{\alpha_t,\theta}(s)ds \\ &= \frac{\alpha_t(\alpha_t+1)}{\theta^2}\overline{F}_{\alpha_t+2,\theta}(x) - \frac{2\alpha_t x}{\theta}\overline{F}_{\alpha_t+1,\theta}(x) + x^2\overline{F}_{\alpha_t,\theta}(x).\end{aligned} \tag{9.27}$$

If the D_t are normally distributed then $\sum_{t=j}^k D_t \sim \mathcal{N}(\sum_{t=j}^k \mu_t, \sum_{t=j}^k \sigma_t^2)$ and if the D_t follow gamma distributions then $\sum_{t=j}^k D_t \sim \text{Gamma}(\sum_{t=j}^k \alpha_t, \theta)$. Thus the expectation of the sum of these random variables is again a random variable of the same distribution type with the parameters described above. $\square$

References

K. Arrow, T. Harris, and J. Marschak. Optimal inventory policy. *Econometrica*, 19: 250–272, 1951.

F. Edgeworth. The mathematical theory of banking. *Journal of Royal Statististical Society*, pages 113–127, 1888.

S. Karlin. One stage inventory models with uncertainty. *Studies in the Mathematical Theory of Inventory and Production*, pages 109–134, 1958.

B.I. Kvasov. *Methods of shape-preserving spline approximation*. World Scientific, 2000. ISBN 9789810240103. URL http://books.google.com/books?id=_So7vgJTEZYC.

P. Morse and G. Kimball. *Methods of Operations Research*. Technology Press of MIT, Cambridge, 1951.

Evan L. Porteus. *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, 2002.

E.V. Shikin and A.I. Plis. *Handbook on splines for the user*. CRC Press, 1995. ISBN 9780849394041. URL http://books.google.com/books?id=DL88KouJCQkC.

A. Veinott, Jr. Optimal policy for a multi-product, dynamic, nonstationary, inventory problem. *Management Science*, 12:206–222, 1965.

Paul H. Zipkin. *Foundations of Inventory Management*. McGraw-Hill, Cambridge, 2000. ISBN 0-521-77043-2.