

POLICIES FOR THE STOCHASTIC INVENTORY
PROBLEM WITH FORECASTING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Gavin J. Hurley

May 2008

© 2008 Gavin J. Hurley
ALL RIGHTS RESERVED

POLICIES FOR THE STOCHASTIC INVENTORY PROBLEM WITH
FORECASTING

Gavin J. Hurley, Ph.D.

Cornell University 2008

The design of effective inventory control policies for models with stochastic demands and forecast updates that evolve dynamically over time is a fundamental problem in supply chain management. In particular, this has been a very challenging theoretical and practical problem, even for models with a very simple forecast update mechanism. In this work, we present new algorithms for this problem and present extensive computational results that demonstrate their empirical performance.

Our primary contribution to the study of this problem is a new policy iteration algorithm that yields a well-performing, computationally tractable approximation to the solution. In addition, we build on work of Levi *et al.* [39] and extend their new Minimizing and Balancing policies for the problem. Furthermore, we perform an extensive computational investigation of all our new policies and compare their performance to the Myopic policy.

BIOGRAPHICAL SKETCH

Gavin J. Hurley was born on December 9, 1977 in Cork, Ireland. He attended high school at Coláiste an Spioraid Naoimh in Bishopstown, Cork, including a year each at Clague Middle School in Ann Arbor, Michigan and Gymnasium Ohlstedt in Hamburg, Germany.

In 1996 he began studies at Trinity College of the University of Cambridge where he earned a B.A. in Mathematics in June of 1999. He spent another year at Cambridge, earning a Certificate of Advanced Study in Mathematics (Part III) in June of 2000.

In August of 2000 he entered the School of Operations Research and Industrial Engineering at Cornell University where he completed a Ph.D. in Applied Operations Research.

Do ciaróg uimhir a h-aon

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 The Stochastic Inventory Problem	2
1.1.1 Notation	2
1.1.2 Dynamic programming formulation	6
1.2 A New Policy Iteration Scheme	10
1.2.1 Existing Approximations	11
1.2.2 New Approximation: Reversion to Myopic	14
1.3 Extensions to the Minimizing and Balancing Policies	17
1.4 Organization	18
2 Literature Review	19
2.1 Martingale Model of Forecast Evolution	19
2.1.1 Information Sharing	20
2.1.2 Other MMFE-related papers	21
2.1.3 Extensions of the MMFE	21
2.2 Other forecasting methods	22
2.2.1 Time-series methods	22
2.2.2 Bayesian	23
2.3 Myopic Policy	23
2.4 Computation	25
2.4.1 Demand Approximation	25
2.4.2 Infinitesimal Perturbation Analysis	26
3 A Policy Improvement Algorithm	27
3.1 Theoretical Background	27
3.2 Computational Implementation	37
3.2.1 Infinitesimal Perturbation Analysis	39
3.2.2 Value of $\frac{\partial}{\partial y} \mathbb{E}[C_t^\pi(t, x, y)]$ on sample paths	41
4 Extensions to the Minimizing and Balancing Policies	51
4.1 Parameterization	51
4.2 Bounding	56
4.3 Surplus balancing	58

5	Computation using the Martingale Model of Forecast Evolution	61
5.1	The Martingale Model of Forecast Evolution (MMFE)	62
5.1.1	Interpreting the Variance-Covariance Matrix	65
5.2	Sample Path Computation of Demand and Base-Policy Target Inventory Levels	70
5.3	Experimental Design	78
5.3.1	Performance Measures Used	87
6	Experimental results	89
6.1	First Order Effects	91
6.2	Second Order Effects	96
6.3	Other Aspects	100
	Bibliography	109

LIST OF TABLES

3.1	Sample path values of $C'(y)$: Hand-worked evolution of inventory levels when ordering up to y	45
3.2	Sample path values of $(\Delta^\pi)'$: Hand-worked evolution of inventory levels when ordering up to $y + \delta$	46
3.3	Values of $\rho(y), \mu(y)$ and $C'(y)$	46
4.1	Minimizing and Balancing Policies and Extensions	60
5.1	Scenario codes	86
6.1	Product Launch: $AR(\pi)$, for certain Product Launch scenarios	93
6.2	End Of Life: $AR(\pi)$, for certain End Of Life scenarios	94
6.3	Demand Crash: policy performance in the Demand Crash scenario (Both AT and AR measures are presented)	95
6.4	Seasonality: $AR(\pi)$, for certain Seasonality scenarios	97
6.5	Coefficient of Variation: $AR(\pi)$, for certain Coefficient of Variation scenarios ($L = 0$)	98
6.6	Time of Learning: Average of $AR(\pi)$ over all policies, for the Time of Learning scenarios ($L = 0$)	99
6.7	Correlation: $AR(\pi)$ for the Correlation scenarios ($L = 0$)	99
6.8	Number of scenarios in which each policy performs best (under the $AR(\pi)$ measure)	100
6.9	Robustness statistics: % above the cost of the best heuristic, across the 76 scenarios	102
6.10	Improvement in $AR(\pi)$ due to bounding	103

LIST OF FIGURES

3.1	Plot of $(C'(y))$ against y for our example.	47
5.1	p-p plot of Empirical versus Approximating Distribution of Cumulative Demand.	71
5.2	Initial forecast vectors used in Product Launch Scenarios.	82
5.3	Cumulative sum of the diagonal elements of Σ for constant, early, late and mid-horizon learning.	85
6.1	Scatter plot of the cost on each of the 1000 runs in the Demand Crash scenario with $L = 4$. Myopic policy (Y-axis) against the SB policy (X-axis. Highest 200 highest ratios are marked with an 'x'.	104
6.2	Graph of the inventory on-hand for the SB policy for each period in the Demand Crash scenario ($L = 4$) for the five runs with the highest ratio of SB 's cost to Myopic's.	105
6.3	Difference between the order-up-to levels of the following policies and Minimizing, before bounding: 1) Myopic (thick solid line), 2) $B(0.5)$ (dashed and dotted line), 3) $B(2)$ (dashed line) and 4) SB (solid line).	107

Chapter 1

Introduction

The design of effective inventory control policies for models with stochastic demands and forecast updates that evolve dynamically over time is a fundamental problem in supply chain management. In particular, this has been a very challenging theoretical and practical problem, even for models with a very simple forecast update mechanism. In this work, we present new algorithms for this problem and present extensive computational results that demonstrate their excellent empirical performance.

Most of the existing literature has focused on characterizing the structure of optimal policies. For many of these inventory models, it is well known that there exists an optimal *state-dependent base-stock policy* [68]. In contrast, there has been relatively little progress on how to compute good inventory policies within complex environments. In particular, finding an optimal base-stock policy is usually computationally intractable. As a result, in most practical situations the default policy has been to use a *Myopic Policy*, which computes its decision at the beginning of each period by minimizing the expected cost for the current period, and ignores all future costs. The Myopic Policy is attractive since it can be computed efficiently even in complex environments with forecast updates. However, it performs very poorly in many important scenarios, such as in settings in which the demand is highly variable.

Our primary contribution to the study of this problem is a new policy iteration algorithm that yields a well-performing, computationally tractable approximation to the solution. In addition, we build on work of Levi *et al.* [39] and extend their

new Minimizing and Balancing policies for the problem. Furthermore, we perform an extensive computational investigation of all our new policies and compare their performance to the Myopic Policy.

In this Chapter we first describe the finite horizon, linear cost, stochastic inventory problem. We outline briefly the standard dynamic programming approach to this problem and perform a cost transformation that allows us to ignore ordering costs. In the second section of this Chapter we introduce our new approaches and place them in the context of recent work on the problem.

1.1 The Stochastic Inventory Problem

1.1.1 Notation

We consider a finite planning horizon consisting of T discrete periods, in each of which we make an ordering decision. Demand arrives at a single location and is for a single item. We denote the demand in period t by D_t ; we assume that its density function is continuous. Demand is fully backlogged; that is, when the demand exceeds the amount of stock on hand, the excess demand is put on backorder and satisfied as soon as more units arrive. This is in contrast with a lost sales model where demand that cannot be satisfied immediately is lost.

Often we will require a shorthand notation for the cumulative demand over periods s to t inclusive, i.e. $\sum_{j=s}^t D_j$. We denote this by $D_{[s,t]}$. We will use this convention with other quantities, for example the order quantity Q_t . To exclude the upper endpoint value, we let $D_{(s,t)} = \sum_{j=s}^{t-1} D_j$.

There are three types of cost. We are charged a cost of \hat{p}_t per unit that we order at the beginning of period t . We can treat the value of $-\hat{p}_{T+1}$ as the salvage

value for unsold inventory at the end of the horizon. At the end of period t we are charged \hat{h}_t per unit of unsold inventory and \hat{b}_t per unit of demand on backorder. In Section 1.1.2 we give a cost transformation that will allow us to ignore ordering and salvage costs, i.e. take $\hat{p}_t = 0$ for all t .

At the beginning of each period, we can place an order for a quantity $q_t \geq 0$. There is no fixed cost associated with placing an order; the only cost is that of \hat{p}_t per unit. There is a fixed and known leadtime of L time periods, meaning that an order for q_t units placed at time t is available to satisfy demand L periods later, that is, in period $t + L$.

Costs are discounted by a discount factor of $\alpha \leq 1$. However, as the horizon is finite and the costs are time-dependent, without loss of generality, we take $\alpha = 1$. The only constraints on the cost values are such as to ensure there is no speculative motivation to hold excess inventory or to have backorders. These conditions are

$$\hat{p}_t + \hat{h}_{t+L} \geq \hat{p}_{t+1}; \quad (1.1)$$

$$\hat{p}_t \leq \hat{p}_{t+1} + \hat{b}_{t+L}. \quad (1.2)$$

To see where the first of these comes from, suppose instead that $\hat{p}_t + \hat{h}_{t+L} < \hat{p}_{t+1}$. In this case, we would be better off catering for the demand in period $t + L + 1$ by ordering extra units in period t and holding them over at the end of period $t + L$ (at a cost of $\hat{p}_t + \hat{h}_{t+L}$ per unit) than by ordering them in period $t + 1$ (at a cost of \hat{p}_{t+1} per unit). This is the sort of speculative buying we wish to remove from consideration.

We refer to the amount of stock on hand less backorders at the end of period t as the *net inventory* and denote this by NI_t . By contrast, the *inventory position* in period t before ordering is the net inventory plus the orders in the pipeline.

Denoted by x_t this is given by

$$x_t = NI_{t-1} + q_{[t-L, t-1]},$$

where $q_{[t-L, t-1]} = \sum_{j=t-L}^{t-1} q_j$. Let y_t denote the inventory position after ordering, i.e. $y_t = x_t + q_t$. We refer to this as the *post-order inventory level*.

The precise sequence of events in each period, t , is as follows:

- The order placed in period $t - L$ of q_{t-L} units arrives;
- We decide on q_t , the amount to order and incur a cost of $\hat{p}_t q_t$;
- D_t , the demand in period t is realized;
- We compute our net inventory position; it is the net inventory position from the previous period augmented by the q_{t-L} units that arrived and decremented by the demand D_t , i.e. $NI_t = NI_{t-1} + q_{t-L} - D_t$;
- We incur holding or backorder costs given by $\hat{h}_t(NI_t)^+ + \hat{b}_t(NI_t)^-$, where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$.

We turn now to the model of demand. There are various properties that we would like this model to satisfy which reflect scenarios experienced by inventory managers. Demand may be non-stationary across periods. Many products have highly seasonal demand patterns such as, for example, bread stuffing mixes which have demand spikes around Thanksgiving and Easter as observed by Berns *et al.* [8]. In addition, we wish to allow for correlation between the demand in different periods. This correlation could be positive, as for example when a product enters the beginning or end of its life-cycle, or it could be negative, as for example when an advertising or promotional offer increases demand in one period but decreases it in the subsequent period.

Another crucial motivation for this work is the fact that through forecasting we can expect to learn more about the distribution of demand in a period as that period comes nearer. For example it may be realistic to expect that our knowledge of the distribution of D_t will be quite uncertain in period $s \ll t$, but then be quite accurate in period $s = t - 1$.

We require notation to describe this idea of evolving information about the level of demand. We use the notation \mathcal{F}_t to capture all information available at time t ; in probability theory this is known as a filtration. This can be information about past demand, information about the conditional distribution of future demands, a combination thereof, or other relevant facts. We will often take expectations of the form

$$\mathbb{E}[D_t|\mathcal{F}_s];$$

this is the expected value of D_t , the demand in period t , given all we know in period s (i.e. given \mathcal{F}_s). For $s \leq t$, the distribution of D_t given \mathcal{F}_s may still be random, whereas if $s \geq t + 1$, given \mathcal{F}_s , then D_t is fixed and known.

In Chapter 5 we introduce the specific model we will use in our computational studies, namely the Martingale Model of Forecast Evolution, as developed independently by Heath and Jackson [25] and Graves *et al.* [22]. This model can accommodate seasonality, correlation, and evolving forecasts and it fits nicely into the \mathcal{F} -notation framework described here.

However, in what follows, the only assumption that we require on demands is that the quantity $\mathbb{E}[D_t|\mathcal{F}_s]$ is always well-defined and finite and that D_t is a continuous random variable. This allows for non-stationarity, correlation, and the evolution of information about future demands. The method we develop is quite general.

We have described the costs in the model, the decisions available (the order quantity q_t at time t) and the underlying randomness of demand. The objective in our problem is to choose the quantity to order in every period t , that is q_t , so as to minimize the total expected holding and backorder costs over the horizon.

Recall that $y_t = x_t + q_t$. Thus, in the remainder of the work, we express the problem of finding an optimal order quantity $q_t \geq 0$ equivalently as finding the optimal post-order inventory level y_t such that $y_t \geq x_t$.

1.1.2 Dynamic programming formulation

The traditional approach to our problem is via dynamic programming. We describe briefly the form of the dynamic program and then prove a structural result.

We discussed in Section 1.1.1 how an order at time t only arrives at time $t + L$. Thus, when seeking the optimal quantity to order at time t it is natural to consider only the costs for periods that we can actually affect, namely those of periods $t + L$ up to the end of the horizon.

Therefore, we define the sum of holding and backorder costs charged in period t as

$$\hat{c}_t(y) := \hat{h}_{t+L}(y - D_{[t,t+L]})^+ + \hat{b}_{t+L}(y - D_{[t,t+L]})^-.$$

Thus, period $(t + L)$ -costs are actually being accounted for in period t . Note that this quantity is a random variable.

Let $\hat{F}_t(x)$ be the optimal expected total cost from period t through T , given that the inventory position before ordering in period t is x . Then, the relevant dynamic program is given by

$$\hat{F}_t(x) = \min_{y \geq x} \left\{ \hat{p}_t(y - x) + \mathbb{E} \left[\hat{c}_t(y) + \hat{F}_{t+1}(y - D_t) \mid \mathcal{F}_t \right] \right\} \quad (1.3)$$

with terminal condition

$$\hat{F}_{T+1}(x) = -\hat{p}_{T+1}x$$

where in Equation 1.3, the first term is the ordering cost, the second is the expected holding and backorder cost in period $t + L$ and the final is the recursive term representing the “cost-to-go”.

It may happen that there are multiple y values that minimize Equation 1.3. In such a case, we break ties by choosing the smallest minimizing y . This ensures that our solution to the dynamic program is unique.

Although it makes it a more realistic model, including ordering costs can distract from the fact that the problem we are studying is that of minimizing inventory costs as opposed to the costs of production. If, as is likely to be the case, the ordering cost \hat{p}_t is very large compared to the holding and backorder costs \hat{h}_t and \hat{b}_t , then a significant improvement in the inventory policy may appear to have only a small effect on the overall cost. Thus we would like to eliminate the ordering cost \hat{p}_t . An informal argument for doing so is to assume that by the end of the horizon we will end up satisfying all demand and therefore the only question of interest is when to time the orders, allowing us to ignore the ordering cost \hat{p}_t .

It is well-known that this argument can be made rigorous, and also account for differing \hat{p}_t values by folding the ordering cost into the holding and backorder costs. (See, for example, the work of Janakiraman and Muckstadt [32] who consider a more general setting.) Specifically, define $h_{t+L} := \hat{h}_{t+L} + (\hat{p}_t - \hat{p}_{t+1})$ and $b_{t+L} := \hat{b}_{t+L} - (\hat{p}_t - \hat{p}_{t+1})$. Both h_{t+L} and b_{t+L} are non-negative due to our previous assumptions in Equations 1.1 and 1.2. Now let

$$\begin{aligned} c_t(y) &:= h_{t+L} (y - D_{[t,t+L]})^+ + b_{t+L} (y - D_{[t,t+L]})^- \\ F_t(x) &:= \hat{p}_t x + \hat{F}_t(x) \end{aligned}$$

and thus we have $c_t(y) = \hat{c}_t(y) + (\hat{p}_t - \hat{p}_{t+1})(y - D_{[t,t+L]})$, giving us the formulation

$$F_t(x) = \min_{y \geq x} \mathbb{E}[c_t(y) + F_{t+1}(y - D_t) + (\hat{p}_t - \hat{p}_{t+1})D_{[t,t+L]} + \hat{p}_{t+1}D_t | \mathcal{F}_t] \quad (1.4)$$

for $t = 0, 1, 2, \dots, T$ with boundary condition

$$F_{T+1}(x) = 0, \quad \forall x$$

The purpose of solving Equation 1.5 is to obtain an optimal decision rule in period t , which we denote by the function $y_t^*(\cdot)$. As y appears only in the first two terms, we can write

$$F_t(x) = \min_{y \geq x} \mathbb{E}[c_t(y) + F_{t+1}(y - D_t) | \mathcal{F}_t], \quad (1.5)$$

and thus the optimal decision for y is given by:

$$y_t^*(x) = \arg \min_{y \geq x} \mathbb{E}[c_t(y) + F_{t+1}(y - D_t) | \mathcal{F}_t], \quad x \in \mathbb{R}. \quad (1.6)$$

Optimality of Base-Stock Policies We now review the well-known proof (see Zipkin [68] for example) that the optimal policy for this problem is a base-stock policy.

Definition 1.1.1 *We say that a policy π is a base-stock policy if, for each t , there is a threshold, or base-stock level \bar{y}_t^π such that*

$$y_t^\pi(x) = x \vee \bar{y}_t^\pi$$

where $x \vee y = \max(x, y)$.

Lemma 1.1.2 *In each period t , the optimal policy $y_t^*(x)$ is a base-stock policy.*

Proof : The proof is by induction. We show jointly that

1. $F_t(x)$ is convex in x
2. There exists \bar{y}_t^* such that $y_t^*(x) = x \vee \bar{y}_t^*$

Define

$$G_t(y) = \mathbb{E}[c_t(y) + F_{t+1}(y - D_t) | \mathcal{F}_t] \quad (1.7)$$

As $F_{T+1}(x) \equiv 0$, in period T , $G_T(y) = \mathbb{E}[c_T(y) | \mathcal{F}_T]$. This is clearly a convex function and in addition $G_T(y) \rightarrow \infty$ as $y \rightarrow \pm\infty$. Let \bar{y}_T be its global minimizer.

Then we see that

$$\arg \min_{y \geq x} G_T(y) = \bar{y}_T \vee x$$

The function

$$F_T(x) = G_T(x \vee \bar{y}_T)$$

is convex in x . To see this, note that $F_T(x) \equiv G_T(\bar{y}_T)$ for $x \leq \bar{y}_T$ and $F_T(x) = G_T(x)$ for $x \geq \bar{y}_T$ and, as $G_T(\cdot)$ is convex, thus $F_T(x)$ is convex on each of these regions separately. It remains to consider $x_1 \leq \bar{y}_T \leq x_2$ and to show that, for any $\lambda \in [0, 1]$, we have $F(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda F_T(x_1) + (1 - \lambda)F_T(x_2)$.

Let $\bar{\lambda}$ satisfy $\bar{y}_T = \bar{\lambda}x_1 + (1 - \bar{\lambda})x_2$. Then, for $\lambda \leq \bar{\lambda}$, we have

$$\begin{aligned} F_T(\lambda x_1 + (1 - \lambda)x_2) &= G_T(\bar{y}_T) \\ &= \lambda G_T(\bar{y}_T) + (1 - \lambda)G_T(\bar{y}_T) \\ &\leq \lambda F_T(x_1) + (1 - \lambda)F_T(x_2) \end{aligned}$$

as $F_T(x_2) = G_T(x_2) \geq G_T(\bar{y}_T) = F_T(x_1)$ by the definition of \bar{y}_T .

In particular, let $\beta = \bar{\lambda}F_T(x_1) + (1 - \bar{\lambda})F_T(x_2)$ and note that $\beta \geq F_T(\bar{y}_T)$.

For, $\lambda \geq \bar{\lambda}$, note that there exists $\mu \in [0, 1]$ such that $\lambda x_1 + (1 - \lambda)x_2 = \mu \bar{y}_t + (1 - \mu)x_2$. Thus,

$$\begin{aligned} \lambda F_T(x_1) + (1 - \lambda)F_T(x_2) &= \mu\beta + (1 - \mu)F_T(x_2) \\ &\geq \mu F_T(\bar{y}_T) + (1 - \mu)F_T(x_2) \\ &\geq F_T(\mu \bar{y}_T + (1 - \mu)x_2) \\ &= F_T(\lambda x_1 + (1 - \lambda)x_2) \end{aligned}$$

where we have used the inequality from above and the convexity of $F_T(\cdot)$ on $x \geq \bar{y}_T$.

This establishes the inductive hypothesis for $t = T$.

Now, assuming the inductive hypothesis holds for $t = s + 1, s + 2, \dots, T$ we see that $G_s(y) = \mathbb{E}[c_s(y) + F_{s+1}(y - D_s) | \mathcal{F}_s]$ is the sum of two convex functions and hence is itself convex. Thus, applying the same logic as above, the hypothesis holds for $t = s$ and the proof follows. \blacksquare

1.2 A New Policy Iteration Scheme

Since the optimal value function, $F_{t+1}(\cdot)$, is available only through recursive calculation over an uncountable state space, the approach of using Equation 1.6 to obtain optimal decision rules in each period is computationally intractable. There are numerous approximation techniques but they all require numerous function evaluations over very large state spaces. For many practical applications it is desirable to find computationally efficient approximations to $F_{t+1}(\cdot)$ as input to Equation 1.6 so that good policies can be computed quickly. Several approximations have been proposed in the literature. We will briefly present three of them and introduce a fourth approximation that is a logical extension of the developmental trend. This new approximation, in turn, suggests a policy improvement

approach to dynamic programs of this form that we believe to be novel. While we do not propose to implement more than one policy iteration in practice, the approach gives us confidence that our choice of approximation has a good theoretical justification. Computational experience demonstrates that the approach is both efficient and effective.

1.2.1 Existing Approximations

Myopic Objective The first approximation suggested in the literature is simply to ignore the optimal value function in Equation 1.6 and use a myopic decision rule $y_t^M(\cdot)$:

$$y_t^M(x) = \arg \min_{y \geq x} \mathbb{E}[c_t(y) | \mathcal{F}_t]$$

Note that due to the convexity of $\mathbb{E}[c_t(y) | \mathcal{F}_t]$, the policy M is a base stock policy.

In stationary demand settings, this rule tends to perform well but it can result in overstocking in situations where a drop-off in demand is forecast. It has been shown (see, for example Zipkin [68]) that this is an *upper bounding policy*, in the sense that

$$y_t^*(x) \leq y_t^M(x) \quad \forall x \in \mathbb{R}$$

To understand this result intuitively, consider the fact that the Myopic Policy only considers the most immediate period in making its ordering decision. On a simple level, it may get things wrong in later period either by having too much inventory (overstocking) or too little (understocking). In the latter case, it can order more at a later period to correct itself, however, in the case of overstocking, inventory cannot be made to disappear. The optimal policy will choose the optimal

amount to order for period $t+L$ but tempered by this danger of future overstocking. Thus, the optimal policy will never order more than the Myopic Policy.

Of all the approximations considered in this work, the myopic objective is the easiest to compute.

Conservative Objective Muckstadt and Sox [49] propose to avoid the overstocking discussed above by adding a penalty function to the myopic objective:

$$y_t^s(x) = \arg \min_{y \geq x} \mathbb{E} \left[c_t(y) + \sum_{j=t+L+1}^{T+L} h_j(y - D_{[t,j]})^+ | \mathcal{F}_t \right]$$

In this objective, inventory in excess of what is needed to cover lead time demand is penalized at the rate h_j per unit for every period j that the stock remains in the system.

Note that, due to the convexity of $\mathbb{E} \left[c_t(y) + \sum_{j=t+L+1}^{T+L} h_j(y - D_{[t,j]})^+ | \mathcal{F}_t \right]$, the policy s is a base-stock policy.

This is a conservative objective because it neglects the contribution of excess inventory towards reducing backorder costs in periods beyond the lead time. It can be shown that this objective leads to a *lower bounding policy* in the sense that

$$y_t^s(x) \leq y_t^*(x) \quad \forall x \in \mathbb{R}$$

Define the runout time $\tilde{\tau}$ as $\tilde{\tau} = \min\{j : D_{[t,j]} \geq y\}$. Note that this quantity could be considerably less than the time horizon $T + L$ so the computation of the sum in the objective function for any value of y could be truncated to some fixed value τ_α such that $\mathbb{P}(\tilde{\tau} > \tau_\alpha) < \alpha$ for some appropriately small probability α . That is, the objective function could be further approximated as

$$\mathbb{E} \left[c_t(y) + \sum_{j=t+L+1}^{\tau_\alpha} h_j(y - D_{[t,j]})^+ | \mathcal{F}_t \right]$$

Compensated Conservative Objective (Minimizing Policy) Levi *et al.* [39] also propose a compensated objective. It is essentially the same as the Muckstadt and Sox objective but it is adjusted to give credit for the initial inventory:

$$y_t^m(x) = \arg \min_{y \geq x} \mathbb{E} \left[c_t(y) + \sum_{j=t+L+1}^{T+L} h_j [(y - D_{[t,j]})^+ - (x - D_{[t,j]})^+] | \mathcal{F}_t \right]$$

It is clear that because the additional terms depend only upon x , they have no effect on the optimization. That is

$$y_t^m(x) = y_t^s(x) \quad \forall t, x$$

We refer to this policy as the Minimizing Policy.

We can split this Compensated Conservative objective function into two components. The first is the incremental holding cost, the holding cost that can be attributed to the $y - x$ units ordered in period t ; it is given by

$$h_t(x, y) = \mathbb{E} \left[\sum_{j=t+L}^{T+L} h_j [(y - D_{[t,j]})^+ - (x - D_{[t,j]})^+] | \mathcal{F}_t \right] \quad (1.8)$$

and the second is the single period backorder cost, given by

$$b_t(y) = \mathbb{E}[b_t (y - D_{[t,t+L]})^- | \mathcal{F}_t].$$

Then

$$y_t^m(x) = \arg \min_{y \geq x} [h_t(x, y) + b_t(y)]$$

By similar arguments to before, we see that $h_t(x, y) + b_t(y)$ is convex and so both the Conservative and Minimizing Policies are base-stock policies; that is, there is a target inventory level given by

$$\bar{y}_t^m = \arg \min_y [h_t(x, y) + b_t(y)], \quad (1.9)$$

and, from this,

$$y_t^m(x) = x \vee \bar{y}_t^m$$

Balancing Objective Levi *et al.* [39] propose a different decision rule, one that is based upon equating single period backorder costs with incremental holding costs. That is, the post-order inventory level, $y_t^B(x)$, chosen by the *Balancing Policy*, $y_t^B(x)$ is such that

$$h_t(x, y_t^B(x)) = b_t(y_t^B(x))$$

The corresponding objective would be to minimize the maximum of incremental holding and backorder costs. For any $x \in \mathbb{R}$, the functions $h_t(x, y)$ and $b_t(y)$ are convex and continuous in y . In addition, $h_t(x, 0) = 0$ and $h_t(x, \cdot)$ is increasing while $b_t(y) \geq 0$ and $b_t(y) \rightarrow 0$ as $y \rightarrow \infty$. Thus $y_t^B(x)$ exists.

Note, in addition, that unless $b_t(y) \equiv 0$, then $y_t^B(x) > x$; that is, the Balancing Policy always places an order. Thus it is the first policy we have encountered which is not a base-stock policy; in a base stock policy whether or not we order depends on whether our inventory position x is less than the threshold (base stock) level or not.

Levi *et al.* [39] show that the Balancing Policy achieves a performance guarantee of at most twice the optimal cost. This is the first known performance guarantee for the stochastic inventory problem. However, our focus in this work is not on policies that yield performance guarantees. It is, instead, on ways to improve the approximation of $F_{t+1}(\cdot)$ in Equation 1.6.

1.2.2 New Approximation: Reversion to Myopic

As noted, the myopic objective leads to an upper bounding policy and the conservative objective leads to a lower bounding policy (the Minimizing Policy); that is, we have

$$y_t^m(x) \leq y_t^*(x) \leq y_t^M(x). \tag{1.10}$$

We seek an objective function that would yield an intermediate result with the hope that it would lead to improved results without greatly increasing the computation time.

A natural idea to investigate in this context is that instead of charging holding costs in the objective until the excess inventory is consumed (at the runout time $\tilde{\tau}$) we instead only charge until another order would be placed under any reasonable policy, such as the Myopic Policy. Carrying this idea further, we could compare the cost of two policies: ordering up to y in this period and then following the myopic decision rule in all subsequent periods versus simply following the myopic rule in this and all subsequent periods. The cost differential between these two policies could then be used to form the objective function for choosing y .

Suppose that we start period t with x units of inventory. Let y_s^M denote the inventory position in period s , $s \geq t$, assuming that we follow the myopic decision in each period:

$$y_t^M = y_t^M(x),$$

and, for all $s > t$,

$$y_{s+1}^M = y_{s+1}^M(y_s^M - D_s).$$

If, instead of following the myopic rule in period t , we order up to inventory level y , then we would order again, under the myopic rule, as soon as $y - D_{[t,s]} < \bar{y}_s^M$, where \bar{y}_s^M is the base-stock level of the Myopic Policy at time s .

Let $\rho(t, x, y)$ denote the earliest such time:

$$\rho(t, x, y) = \min_{s>t} \{s : y - D_{[t,s]} < \bar{y}_s^M\}$$

We refer to this hitting time as the *time of reversion* or, more simply, the *reversion time*. It is the time at which our temporary diversion from the Myopic Policy

reverts back to following the decision rules of the Myopic Policy.

Using this notation, we can now define a new objective function more precisely.

$$y_t^R(x) = \arg \min_{y \geq x} \mathbb{E} \left[c_t(y) + \sum_{s=t+1}^{\rho(t,x,y)} \{c_s(y - D_{[t,s]}) - c_s(y_s^M - D_{[t,s]})\} | \mathcal{F}_t \right]$$

This approach is somewhat more computationally intensive than the conservative approach. However, the time to reversion for many values of y is likely to be quite short and this fact can be exploited in the computational procedures.

Observe that this objective captures both the benefits and costs of excess inventory after the lead time. For example, if $y > y_t^M$, then holding costs will be greater than under the myopic decision rule until the time of reversion. These additional holding costs are not actually incurred until a lead time into the future but they are traceable to differences in decision rules from period t until the time of reversion.

Likewise, the objective function will also capture the benefit, in terms of potentially lower backorder costs a lead time hence, of entering the period of reversion with more inventory than a straight application of the myopic decision rule would have obtained.

Because the reversion-to-myopic objective captures more interesting tradeoffs than either the myopic or the conservative approaches we would expect that the decision rule based on minimizing this function would lead to superior performance.

Beyond heuristic justification, however, this approach suggests that a more formal, theoretical justification is possible. We note that the use of the Myopic Policy in defining the reversion-to-myopic objective function is purely for computational convenience. Computational issues aside, reversion time could be computed with respect to any base-stock policy. If the policy that results from optimizing such an

objective function is itself a base stock policy, then we can imagine a policy improvement algorithm constructed along these lines. Hence, our formal justification for this approach will come from the policy improvement machinery of Markov decision processes.

1.3 Extensions to the Minimizing and Balancing Policies

Our new approximation is motivated by the desire to achieve an intermediate result between the upper and lower bounds of the myopic and conservative objectives. The second theme in this work is to generate extensions to the Minimizing and Balancing policies with the same goal in mind.

We take three approaches in this:

Parameterization: Here we develop a method of parameterizing the space between $[y^m(x), y^M(x)]$. In addition, we describe dynamic methods of choosing the relevant parameter.

Bounding: If a policy suggests an order that falls outside $[y^m(x), y^M(x)]$ we can bound or truncate the order at the appropriate endpoint.

Surplus Balancing: While the original Balancing Policy balances the backorder cost with the marginal holding cost for all units ordered, we instead consider the marginal holding cost for only those units ordered above the order suggested by the Minimizing Policy.

1.4 Organization

The remainder of this work is organized as follows. In Chapter 2 we give an overview of the related literature. Chapter 3 contains our primary contribution which uses the reversion to myopic idea described above. It provides the theoretical justification for the method, as well as a computational methodology for implementing it. We present our extensions to the Minimizing and Balancing Policies in Chapter 4. Chapter 5 introduces the Martingale Model of Forecast Evolution, which is our model for demand. It describes how the new policies from Chapter 3 are implemented under this model. It also describes the set of scenarios we use when performing our computational experiments, the results of which are presented in Chapter 6.

Chapter 2

Literature Review

2.1 Martingale Model of Forecast Evolution

In this work, we use the Martingale Model of Forecast Evolution (MMFE) as our model for demand and forecasting. This was introduced independently by Heath and Jackson [25] and Graves, Meal, Dasu and Qin [22]. We describe the MMFE in detail in Chapter 5. In their work, Heath and Jackson apply the MMFE to an industrial problem involving a company that produces multiple products with highly correlated and highly seasonal demand. They use the MMFE to compare two forecasting models in use at the company. In addition they incorporate it into a linear programming model of production and distribution. By means of simulation, they demonstrate that the company could use a lower safety stock level.

Graves, Kletter and Hetzel [21] incorporate the MMFE into a Materials Requirements Planning (MRP) problem, which traditionally assumes static forecasts. For a single-stage production system, they optimize analytically the trade off between production variance and inventory variance and use this as a building block for a multi-stage system. Toktay and Wein [62] study the same problem but with capacity constraints. They model the production stage as a single-server queue and apply heavy traffic theory to determine a forecast-corrected base stock policy.

In two papers, Güllü ([23] and [24]) demonstrates that both inventory levels and system costs are lower when forecast information (modelled by the MMFE) is used. The first of these ([23]) addresses a single location model; the second ([24]) considers a two-echelon model where supplies are received at a single location and

then distributed among retailers.

Chen, Ruppert and Shoemaker study the problem of approximating the future value function in stochastic dynamic programming and apply both experimental design and regression splines. They apply their method to an inventory problem using forecasting from the MMFE. Dong and Lee study a serial multiechelon inventory system. They show that Clark and Scarf's classical result that the optimal inventory policy for the entire system is to follow a base-stock policy at each echelon continues to hold when demand is generated by means of a MMFE process. In this case, the base-stock levels are forecast-dependent.

Gallego and Özer [17] consider a model of advance demand information. They show that state dependent base-stock and (s, S) policies are optimal for the stochastic inventory problem without and with fixed costs respectively. Here the state is the information already revealed about future demand. Özer [50] extends this work to a distribution system and proposes a two-stage approximation that firstly follows a state-dependent (the state again depends on advance demand information) base-stock policy for the centralized warehouse and the a myopic allocation policy that uses part of the advance demand information. Dellaert and Melo [13] also consider an advance demand model for a production system with setup costs. They investigate two heuristic policies which exploit the advance demand information and demonstrate their superior performance in simulations.

2.1.1 Information Sharing

In a series of papers ([3], [4] and [5]), Aviv studies the value of collaborative forecasting between agents in a supply chain. In one of these papers [3], the forecasting model is a special case of the MMFE; here Aviv compares the cases

where inventory management ignores forecasting, where all forecasting is done locally by each agent and where forecasting is coordinated across the supply chain. Zhu and Thonemann [67] consider a similar problem where a retailer can, at a cost, receive a demand forecast from his customers. This forecasted demand and the actual realized demand evolve as a two-stage additive MMFE. They compute the optimal number of customers to ask, and the subsequent optimal inventory level. Krajewski and Wei [38] is another study of integrating scheduling and inventory decisions in the presence of forecast information.

2.1.2 Other MMFE-related papers

Sethi, Yan and Zhang *et al.* have a series of papers ([57], [58] and [59]) on inventory models with successively improving forecast updates and varying production speeds. Bradley and Glynn [10] jointly optimize inventory and capacity decisions under a queueing theory model, which they extend to a model with forecast updates driven by the MMFE. Kaminsky and Swaminathan [34] study optimal inventory policies in the presence of forecasts which fall between deterministic and decreasing bands.

2.1.3 Extensions of the MMFE

Catanyildirim and Roundy [11] study forecasting in the semiconductor industry and provide an extensive scheme for studying the variance and covariance of forecast errors and use the MMFE to describe forecast evolution. Their approach is validated against industrial data. Zhou, Jackson, Roundy and Zhang [66] consider a situation where forecasts are available for all products a certain number of periods in advance, but only at an aggregated level thereafter. They model this evolution

using an extension of the MMFE and propose and compare various methods of extracting individual forecasts from the aggregated ones. Zhang *et al.* [65] study the problem of combining forecasts from various sources into an aggregated forecast, proposing and testing various schemes.

2.2 Other forecasting methods

2.2.1 Time-series methods

Johnson and Thompson [33] consider a model where demand is driven by an autoregressive moving average (ARMA) process and demonstrate that a myopic policy is optimal in this case as it can always order up to a critical or base-stock level. Erkip, Hausman and Nahmias [14] consider a multi-echelon model and use an ARMA model to allow for correlation between demand across locations and over time.

Miller [45] develops a model whereby demand is a combination of a random variable and exponentially weighted past demand; like in the MMFE this implies that further information about demand in period t is revealed as one gets closer to t . He provides an efficient dynamic programming formulation for this model. Kim and Ryan [37] consider a supply chain with a manufacturer and retailer where the retailer sees demand driven by an AR(1) process. They quantify the benefit of the retailer using an exponential smoothing forecasting method like that of Miller [45] in addition to the benefit of sharing demand information between the retailer and supplier.

Snyder, Koehler, Hyndman and Ord [60] and Snyder, Koehler and Ord [61] discuss a variety of exponential smoothing methods which can be used to forecast the

average and variance of lead-time demand when that demand is driven by ARMA models that include trend and seasonality effects. They provide a computational study of how these can be used to optimize the fill rate.

Goto, Lewis and Puterman [20] consider the provisioning of meals for airline flights. They model the evolving information about passenger load with a Markov chain and formulate a Markov decision process model for the minimization of costs, which they then investigate computationally.

2.2.2 Bayesian

Scarf ([54] and [55]) introduces a Bayesian model, whereby the demand in a period is a function of some parameter and that parameter has a prior distribution which depends on realized demand. Azoury [6] extends this work and demonstrates that the multi-dimensional state space dynamic program arising from such a model can be reduced to a model with a one-dimensional state space.

2.3 Myopic Policy

Two papers by Karlin ([35] and [36]) consider the case where the distribution of demand varies with time. He shows that a base-stock policy is optimal for the case of backlogging with no fixed ordering costs. He also defines a stochastic ordering on the demand distributions in future periods and shows that the order-up-to levels are increasing in this measure and that hence a Myopic Policy is optimal. Veinott ([63] and [64]) continues this work and considers the case of multiple products. The paper of Johnson and Thompson [33] mentioned above also considers the Myopic Policy.

In the papers above, the future demand distributions are known. Lovejoy [41]

considers the case where future demand is a function of previously observed demand and provides conditions on this function such that the Myopic Policy continues to be optimal. For the case when these assumptions and the demand update function do not hold, Lovejoy [42] provides a method for bounding the cost increase due to following a Myopic Policy as opposed to the optimal policy.

Morton and Pentico [48] give a stronger bound for the finite horizon case. They present a range of heuristics inspired by the Myopic Policy and conduct a computational investigation. Anupindi, Morton and Pentico [2] extends this to the infinite horizon case.

Zipkin [68] has a detailed overview of work on the Myopic Policy.

Two papers that are especially close to our work are those of Iida and Zipkin [31] and Lu, Song and Regan [43]. Both study the MMFE and prove structural results about the optimal policy. In addition they consider the performance of the myopic policy in this setting.

Iida and Zipkin [31] show that a forecast-level dependent base-stock policy is optimal for the finite-horizon stochastic inventory problem. They provide upper and lower bounds on the optimal inventory level and show that under certain conditions the difference between them tends to zero. They establish a sufficient condition for the Myopic Policy to be optimal; one such case is in the additive MMFE when costs are stationary. In addition, they develop an approximation method which involves generating a piecewise-linear approximation to the (convex) cost function and computing expectations via sampling. They use this to perform a computational investigation.

Lu, Song and Regan [43] provide a simple necessary and sufficient condition for the Myopic Policy to be optimal. They also provide upper and lower bounds on the

optimal base-stock level for the stochastic inventory problem, but their approach is to use a path-wise argument. They consider the heuristic policy that takes a fixed weighted average of the lower and upper bound and bound the difference between its cost and that of the optimal policy. They also perform a numerical investigation.

Our approach of jolted policies described in Chapter 3 is similar to that of Morton ([46] and [47]) who allows for negative ordering in each period in his study of the non-stationary infinite horizon inventory problem. Cheevaprawatdomrong and Smith [12] extend this and note that in a system with time varying demand, the optimal policy can be computed using a shorter forecast horizon; this relates to our idea of reversion time.

2.4 Computation

2.4.1 Demand Approximation

Wilkinson's approximation for the distribution of the sum of correlated lognormal random variables is described in Schwartz and Yeh [56]. Abu-Dayya and Beaulieu [1] compare three approximations (Wilkinson's approximation, an extension of the work of Schwartz and Yeh and a cumulants matching approach) and conclude that Wilkinson's is the best due to its accuracy and computational tractability. The approximation is applied variously in finance by Milevsky and Posner [44] and in mobile telephony by Beaulieu, Abu-Dayya and McLane [7], Safak [53] and Ligeti [40].

2.4.2 Infinitesimal Perturbation Analysis

Ho and various co-authors ([26], [27] and [28] did the early work on the method of Infinitesimal Perturbation Analysis and Glasserman [18] gives an overview.

For inventory applications, Glasserman and Tayur [19] apply the method to capacitated inventory system to compute the derivatives of cost with respect to base-stock levels and demonstrate that these derivatives converge in given settings. Fu [15] and Fu and Healy [16] study systems with fixed costs and describe how to compute sample path derivatives. Bhaskaran [9] is a more recent application that attempts to reduce inventory fluctuations to create a more stable supply chain.

Chapter 3

A Policy Improvement Algorithm

3.1 Theoretical Background

In this section, we prove a restatement of results for policy improvement algorithms in Markov Decision Processes. The restatement is in terms of a special formulation of the single period objective function. Our formulation here is more general than that in Chapter 1 and hence our results hold more generally. In Section 3.2 we will return to the specific problem of the linear cost stochastic inventory problem. The primary goal of this chapter is to ground heuristic approaches to this problem in a solid theoretical framework.

Consider the Markov Decision Process (MDP) expressed by the following dynamic program:

$$F_t(x) = \min_{y \in R_t(x)} \mathbb{E}[c_t(x, y) + F_{t+1}(T_t(y)) | \mathcal{F}_t], \quad (3.1)$$

for all $x \in X$, and for $t = 1, 2, \dots, T$, subject to the boundary condition

$$F_{T+1}(x) = 0,$$

for all $x \in X$. Note that this MDP is expressed in terms of *pre-states*, x , and *post-states*, y , rather than states and actions, as is more typical. We assume that X is a subset of \mathcal{R}^n . The state-dependent set $R_t(x) \subseteq X$ describes the post-states, y , that are reachable from pre-state x . The transition function $T_t(y)$, possibly random, maps the post-state in this period to a pre-state in the next period. \mathcal{F}_t is the information set as known at the beginning of period t . The function $c_t(x, y)$, possibly random, is the single period cost function for period t expressed in terms

of the pre-state and the post-state. The optimal value function in period t is $F_t(\cdot)$. Observe that this formulation includes a wide variety of inventory control problems, including backordered demand, lost sales, fixed charge, capacitated production, multi-item, and multi-location problems.

A feasible policy, $\pi = \{\pi_t : t = 1, 2, \dots, T\}$, is a series of functions, possibly random, mapping pre-states into reachable post-states:

$$\pi_t(x) \in R_t(x),$$

for $t = 1, 2, \dots, T$. Starting from any pre-state x in period t , let $\{(x_s^\pi(t, x), y_s^\pi(t, x)) : s = t, t+1, \dots, T\}$ denote the sample path that we would obtain by following policy π from that period onward. That is,

$$x_t^\pi(t, x) = x,$$

and the dynamics are described by the recursive equations:

$$y_s^\pi(t, x) = \pi_s(x_s^\pi(t, x))$$

and

$$x_{s+1}^\pi(t, x) = T_s(y_s^\pi(t, x)),$$

for $s = t, t+1, \dots, T$.

Let π now denote a particular feasible policy, the so-called *reference policy*. Suppose we begin period t in pre-state x , but instead of choosing the post-state $\pi_t(x)$, as proposed by the reference policy, we instead choose a value y not necessarily even in $R_t(x)$. Assume we continue in all future periods to follow policy π . Let $\{(x_s^\pi(t, x, y), y_s^\pi(t, x, y)) : s = t, t+1, \dots, T\}$ denote the sample path that we would obtain from period t onward under this so-called *jolted reference policy*. To

be precise:

$$x_t^\pi(t, x, y) = x,$$

$$y_t^\pi(t, x, y) = y,$$

and the dynamics are described by the recursive equations:

$$y_{s+1}^\pi(t, x, y) = \pi_{s+1}(x_s^\pi(t, x, y))$$

and

$$x_{s+1}^\pi(t, x, y) = T_s(y_s^\pi(t, x, y)),$$

for $s = t, t + 1, \dots, T - 1$.

For any $s \in \{t, t + 1, \dots, T\}$, let $C_s^\pi(t, x)$ denote the total cost over periods s to T of following the reference policy sample path starting from pre-state x in period t :

$$C_s^\pi(t, x) = \sum_{u=s}^T c_u(x_u^\pi(t, x), y_u^\pi(t, x)).$$

Similarly, let $C_s^\pi(t, x, y)$ denote the total cost over periods s to T of following the jolted reference policy sample path starting from pre-state x jolted to post-state y in period t :

$$C_s^\pi(t, x, y) = \sum_{u=s}^T c_u(x_u^\pi(t, x, y), y_u^\pi(t, x, y)).$$

We are interested in a new single period cost objective, $\Delta_t^\pi(x, y)$, defined as the difference in cost from period t onward between the jolted reference policy sample path and the reference policy sample path:

$$\Delta_t^\pi(x, y) = C_t^\pi(t, x, y) - C_t^\pi(t, x). \quad (3.2)$$

We refer to this new single period cost objective as the Delta- π objective. It is central to this chapter. We will first establish some basic properties of this

objective and then show how it can be used to construct a dynamic programming formulation that is equivalent to Equation 3.1.

Let $\hat{\pi}$ denote any feasible policy and let $\{(w_s, z_s) = (x_s^{\hat{\pi}}(t, x), y_s^{\hat{\pi}}(t, x)) : s = t, t + 1, \dots, T\}$ denote the sample path that would obtain starting from pre-state x in period t and following policy $\hat{\pi}$ thereafter.

Lemma 3.1.1 *Given a reference policy π and any other feasible policy $\hat{\pi}$, then, starting from pre-state x in period t , we have*

$$C_s^\pi(s, w_s, z_s) = c_s(w_s, z_s) + C_{s+1}^\pi(s + 1, w_{s+1}), \quad (3.3)$$

for all $s = t, t + 1, \dots, T - 1$.

Proof : By definition,

$$\begin{aligned} C_s^\pi(s, w_s, z_s) &= \sum_{u=s}^T c_u(x_u^\pi(s, w_s, z_s), y_u^\pi(s, w_s, z_s)) \\ &= c_s(x_s^\pi(s, w_s, z_s), y_s^\pi(s, w_s, z_s)) \\ &\quad + \sum_{u=s+1}^T c_u(x_u^\pi(s, w_s, z_s), y_u^\pi(s, w_s, z_s)) \\ &= c_s(w_s, z_s) + \sum_{u=s+1}^T c_u(x_u^\pi(s, w_s, z_s), y_u^\pi(s, w_s, z_s)) \\ &= c_s(w_s, z_s) + \sum_{u=s+1}^T c_u(x_u^\pi(s + 1, w_{s+1}), y_u^\pi(s + 1, w_{s+1})) \end{aligned}$$

since $w_{s+1} = T_s(z_s)$ and policy π is followed from periods $s + 1$ onward. The result follows from the definition of $C_{s+1}^\pi(s + 1, w_{s+1})$. ■

Proposition 3.1.2 *Given a reference policy π and any other feasible policy $\hat{\pi}$, then, starting from pre-state x in period t , we have*

$$\sum_{s=t}^T \Delta_s^\pi(w_s, z_s) = C_t^{\hat{\pi}}(t, x) - C_t^\pi(t, x).$$

Proof : Starting from Equation 3.2, the definition of the new cost single period cost objective,

$$\begin{aligned}
\sum_{s=t}^T \Delta_s^\pi(w_s, z_s) &= \sum_{s=t}^T \{C_s^\pi(s, w_s, z_s) - C_s^\pi(s, w_s)\} \\
&= \sum_{s=t}^{T-1} \{C_s^\pi(s, w_s, z_s) - C_s^\pi(s, w_s)\} \\
&\quad + C_T^\pi(T, w_T, z_T) - C_T^\pi(T, w_T) \\
&= \sum_{s=t}^{T-1} \{c_s(w_s, z_s) + C_{s+1}^\pi(s+1, w_{s+1}) - C_s^\pi(s, w_s)\} \\
&\quad + C_T^\pi(T, w_T, z_T) - C_T^\pi(T, w_T)
\end{aligned}$$

by Lemma 3.1.1. Continuing,

$$\begin{aligned}
\sum_{s=t}^T \Delta_s^\pi(w_s, z_s) &= \sum_{s=t}^{T-1} c_s(w_s, z_s) + \sum_{s=t}^{T-1} \{C_{s+1}^\pi(s+1, w_{s+1}) - C_s^\pi(s, w_s)\} \\
&\quad + C_T^\pi(T, w_T, z_T) - C_T^\pi(T, w_T) \\
&= \sum_{s=t}^{T-1} c_s(w_s, z_s) \\
&\quad + C_T^\pi(T, w_T) - C_t^\pi(t, w_t) + C_T^\pi(T, w_T, z_T) - C_T^\pi(T, w_T) \\
&= \sum_{s=t}^{T-1} c_s(w_s, z_s) - C_t^\pi(t, w_t) + C_T^\pi(T, w_T, z_T) \\
&= \sum_{s=t}^T c_s(w_s, z_s) - C_t^\pi(t, w_t) \\
&= C_t^{\hat{\pi}}(t, x) - C_t^\pi(t, x).
\end{aligned}$$

■

Remark 3.1.3 *The point of Proposition 3.1.2 is that the difference of cost between a reference policy π and any other policy $\hat{\pi}$ can be computed as the sum of Delta- π objective functions evaluated along the $\hat{\pi}$ policy sample path.*

Let

$$f_t^\pi(x) \equiv F_t(x) - \mathbb{E}[C_t^\pi(t, x) | \mathcal{F}_t], \forall x \in X, \forall t \in \{1, 2, \dots, T\}. \quad (3.4)$$

Proposition 3.1.4 *Let π^* be the optimal policy for the original dynamic program given in Equation 3.1. Then, for all $x \in X$, and for all $t \in \{1, 2, \dots, T\}$,*

$$f_t^\pi(x) = \mathbb{E}[\Delta_t^\pi(x, \pi_t^*(x)) + f_{t+1}^\pi(T_t(\pi_t^*(x))) | \mathcal{F}_t].$$

Proof : We have

$$\begin{aligned} F_t(x) &= \mathbb{E} [C_t^{\pi^*}(t, x) | \mathcal{F}_t] \\ &= \mathbb{E} \left[\sum_{s=t}^T \Delta_s^\pi(w_s^*, z_s^*) + C_t^\pi(t, x) | \mathcal{F}_t \right] \\ &= \mathbb{E} \left[C_t^\pi(t, x) + \Delta_t^\pi(w_t^*, z_t^*) + \sum_{s=t+1}^T \Delta_s^\pi(w_s^*, z_s^*) | \mathcal{F}_t \right] \\ &= \mathbb{E} [C_t^\pi(t, x) + \Delta_t^\pi(x, \pi_t^*(x)) + C_{t+1}^{\pi^*}(t+1, w_{s+1}^*) - C_{t+1}^\pi(t+1, w_{s+1}^*) | \mathcal{F}_t] \\ &= \mathbb{E} [C_t^\pi(t, x) + \Delta_t^\pi(x, \pi_t^*(x)) + F_{t+1}(T_t(z_s^*)) - C_{t+1}^\pi(t+1, T_t(z_s^*)) | \mathcal{F}_t]. \end{aligned}$$

Therefore,

$$\begin{aligned} F_t(x) - \mathbb{E} [C_t^\pi(t, x) | \mathcal{F}_t] &= \mathbb{E} [\Delta_t^\pi(x, \pi_t^*(x)) + F_{t+1}(T_t(z_s^*)) \\ &\quad - C_{t+1}^\pi(t+1, T_t(z_s^*)) | \mathcal{F}_t] \\ &= \mathbb{E} [\Delta_t^\pi(x, \pi_t^*(x)) + F_{t+1}(T_t(z_s^*)) | \mathcal{F}_t] \\ &\quad - \mathbb{E} [C_{t+1}^\pi(t+1, T_t(z_s^*)) | \mathcal{F}_t] \\ &= \mathbb{E} [\Delta_t^\pi(x, \pi_t^*(x)) + F_{t+1}(T_t(z_s^*)) | \mathcal{F}_t] \\ &\quad - \mathbb{E} [\mathbb{E} [C_{t+1}^\pi(t+1, T_t(z_s^*)) | \mathcal{F}_{t+1}] | \mathcal{F}_t] \\ &= \mathbb{E} [\Delta_t^\pi(x, \pi_t^*(x)) + f_{t+1}^\pi(T_t(z_s^*)) | \mathcal{F}_t] \end{aligned}$$

from Equation 3.4. The result follows. ■

Remark 3.1.5 *We have shown that π_t^* gives rise to a new value function $f_t^\pi(\cdot)$ when evaluated using the Delta- π objective. We next show that this new value*

function is actually an optimal value function. The optimal policy π^* from 3.1 is an optimal policy of a new dynamic program.

Theorem 3.1.6 For all $x \in X$, and for all $t \in \{1, 2, \dots, T\}$,

$$f_t^\pi(x) = \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) + f_{t+1}^\pi(T_t(y)) | \mathcal{F}_t]$$

and

$$\pi_t^*(x) = \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) + f_{t+1}^\pi(T_t(y)) | \mathcal{F}_t].$$

Proof : Proposition 3.1.4 established that $y = \pi_t^*(x)$ yields $f_t^\pi(x)$ when substituted into the right hand side objective. It remains to show that this value yields a minimum to the right hand side objective. Suppose there is a feasible value of y such that

$$\mathbb{E}[\Delta_t^\pi(x, y) + f_{t+1}^\pi(T_t(y)) | \mathcal{F}_t] < f_t^\pi(x).$$

Applying the definitions and lemma 3.1.1,

$$\begin{aligned} & \mathbb{E}[\Delta_t^\pi(x, y) + f_{t+1}^\pi(T_t(y)) | \mathcal{F}_t] \\ &= \mathbb{E}[\Delta_t^\pi(x, y) + F_{t+1}(T_t(y)) - \mathbb{E}[C_{t+1}^\pi(t+1, T_t(y)) | \mathcal{F}_{t+1}] | \mathcal{F}_t] \\ &= \mathbb{E}[C_t^\pi(t, x, y) - C_t^\pi(t, x) + F_{t+1}(T_t(y)) - \mathbb{E}[C_{t+1}^\pi(t+1, T_t(y)) | \mathcal{F}_{t+1}] | \mathcal{F}_t] \\ &= \mathbb{E}[c_t(x, y) + C_{t+1}^\pi(t+1, T_t(y)) - C_t^\pi(t, x) + F_{t+1}(T_t(y)) \\ & \quad - C_{t+1}^\pi(t+1, T_t(y)) | \mathcal{F}_t] \\ &= \mathbb{E}[c_t(x, y) + F_{t+1}(T_t(y)) | \mathcal{F}_t] - \mathbb{E}[C_t^\pi(t, x) | \mathcal{F}_t] \\ &\geq F_t(x) - \mathbb{E}[C_t^\pi(t, x) | \mathcal{F}_t] \\ &= f_t^\pi(x), \end{aligned}$$

which leads to a contradiction. ■

We have established that reformulating the single period objective function in this way leads to an equivalent expression of the dynamic program.

We define a new policy $\delta(\pi) = \{\delta_t(\pi) : t = 1, 2, \dots, T\}$, based on the reference policy, as follows:

$$\delta_t(\pi)(x) = \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t]. \quad (3.5)$$

That is, $\delta(\pi)$ is the ‘‘Myopic Policy’’ for the equivalent dynamic program relative to reference policy π . We refer to $\delta(\pi)$ as the Delta- π Myopic Policy.

Remark 3.1.7 *If there are ties for the minimum in (3.5), then our convention is to choose the minimizing y that minimizes the distance from $\pi_t(x)$, that is, that minimizes $|y - \pi_t(x)|$.*

Proposition 3.1.8 *If π is a fixed point of the mapping $\delta : \Pi \rightarrow \Pi$, then π is an optimal policy. Furthermore, $f_t^\pi(x) = 0$, $\forall x \in X$, and $\forall t \in \{1, 2, \dots, T\}$.*

Proof : The proof is by induction. By definition,

$$f_{T+1}^\pi(x) = 0, \quad \forall x \in X.$$

Assume $\pi_s^*(x) = \pi_s(x)$ and $f_s^\pi(x) = 0$, $\forall x \in X$, and $\forall s \in \{t+1, 2, \dots, T\}$. Then, by theorem 3.1.6,

$$f_t^\pi(x) = \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) + f_{t+1}^\pi(T_t(y)) | \mathcal{F}_t]$$

and

$$\pi_t^*(x) = \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) + f_{t+1}^\pi(T_t(y)) | \mathcal{F}_t].$$

By the induction hypothesis, these statements reduce to

$$f_t^\pi(x) = \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t]$$

and

$$\pi_t^*(x) = \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t].$$

By assumption, π is a fixed point, therefore:

$$\pi_t^*(x) = \pi_t(x).$$

It then follows that, $\forall x \in X$,

$$\begin{aligned} f_t^\pi(x) &= \mathbb{E}[\Delta_t^\pi(x, \pi_t(x)) | \mathcal{F}_t] \\ &= 0. \end{aligned}$$

■

Proposition 3.1.9 *If π is any feasible policy, then $\delta(\pi)$ is an improved policy in the following sense:*

$$\mathbb{E}[C_t^{\delta(\pi)}(t, x) | \mathcal{F}_t] \leq \mathbb{E}[C_t^\pi(t, x) | \mathcal{F}_t]$$

$\forall x \in X, \forall t \in \{1, 2, \dots, T\}$. Furthermore, if we have equality throughout (that is, if

$$\mathbb{E}[C_t^{\delta(\pi)}(t, x) | \mathcal{F}_t] = \mathbb{E}[C_t^\pi(t, x) | \mathcal{F}_t]$$

$\forall x \in X, \forall t \in \{1, 2, \dots, T\}$), then π is an optimal policy.

Proof : The proof is by induction. It is trivially true for $t = T + 1$. Assume $\mathbb{E}[C_s^{\delta(\pi)}(s, x) | \mathcal{F}_s] \leq \mathbb{E}[C_s^\pi(s, x) | \mathcal{F}_s]$, $\forall x \in X$ and $\forall s \in \{t + 1, 2, \dots, T\}$. By definition,

$$\begin{aligned} \delta_t(\pi)(x) &= \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t] \\ &= \arg \min_{y \in R_t(x)} \mathbb{E}[C_t^\pi(t, x, y) - C_t^\pi(t, x) | \mathcal{F}_t] \\ &= \arg \min_{y \in R_t(x)} \mathbb{E}[C_t^\pi(t, x, y) | \mathcal{F}_t] \end{aligned}$$

since the term $C_t^\pi(t, x)$ is unaffected by the optimization. It follows that

$$\begin{aligned}\mathbb{E}[C_t^\pi(t, x, \delta_t(\pi)(x))|\mathcal{F}_t] &\leq \mathbb{E}[C_t^\pi(t, x, \pi_t(x))|\mathcal{F}_t] \\ &= \mathbb{E}[C_t^\pi(t, x)|\mathcal{F}_t].\end{aligned}$$

By definition and by the induction hypothesis,

$$\begin{aligned}\mathbb{E}[C_t^{\delta(\pi)}(t, x)|\mathcal{F}_t] &= \mathbb{E}[c_t(x, \delta_t(\pi)(x)) + C_{t+1}^{\delta(\pi)}(t+1, \delta_t(\pi)(x))|\mathcal{F}_t] \\ &= \mathbb{E}[c_t(x, \delta_t(\pi)(x)) + \mathbb{E}[C_{t+1}^{\delta(\pi)}(t+1, \delta_t(\pi)(x))|\mathcal{F}_{t+1}]|\mathcal{F}_t] \\ &\leq \mathbb{E}[c_t(x, \delta_t(\pi)(x)) + \mathbb{E}[C_{t+1}^\pi(t+1, \delta_t(\pi)(x))|\mathcal{F}_{t+1}]|\mathcal{F}_t].\end{aligned}$$

From this, it follows that

$$\begin{aligned}\mathbb{E}[C_t^{\delta(\pi)}(t, x)|\mathcal{F}_t] &\leq \mathbb{E}[c_t(x, \delta_t(\pi)(x)) + C_{t+1}^\pi(t+1, \delta_t(\pi)(x))|\mathcal{F}_t] \\ &= \mathbb{E}[C_t^\pi(t, x, \delta_t(\pi)(x))|\mathcal{F}_t] \\ &\leq \mathbb{E}[C_t^\pi(t, x)|\mathcal{F}_t]\end{aligned}$$

showing that the induction hypothesis is also true for $s = t$. This last set of inequalities shows that if $\mathbb{E}[C_t^{\delta(\pi)}(t, x)|\mathcal{F}_t] = \mathbb{E}[C_t^\pi(t, x)|\mathcal{F}_t] \forall x \in X, \forall t \in \{1, 2, \dots, T\}$, then

$$\begin{aligned}\mathbb{E}[C_t^\pi(t, x, \delta_t(\pi)(x))|\mathcal{F}_t] &= \mathbb{E}[C_t^\pi(t, x)|\mathcal{F}_t] \\ &= \mathbb{E}[C_t^\pi(t, x, \pi_t(x))|\mathcal{F}_t].\end{aligned}$$

By our convention that ties in (3.5) are resolved in favor of $\pi_t(x)$, this implies that $\pi_t(x)$ satisfies

$$\pi_t(x) = \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y)|\mathcal{F}_t]$$

and, hence, π is a fixed point of the mapping $\delta : \Pi \rightarrow \Pi$. Proposition 3.1.8 establishes that such a policy is optimal. \blacksquare

Corollary 3.1.10 *If X is finite, then a policy iteration algorithm of the form*

$$\pi^{n+1} = \delta(\pi^n),$$

beginning from any feasible policy $\pi^0 \in \Pi$, will converge to an optimal policy in a finite number of iterations.

Proof : After each iteration, either $\mathbb{E}[C_t^{\pi^{n+1}}(t, x)|\mathcal{F}_t] < \mathbb{E}[C_t^{\pi^n}(t, x)|\mathcal{F}_t]$ for some $x \in X$ and some $t \in \{1, 2, \dots, T\}$ or π^n is a fixed point. It is easily seen from this that no policy that is not a fixed point can be repeated. Consequently, if there are a finite number of feasible policies, then a fixed point must be found. Such a fixed point policy is optimal. A sufficient condition for Π to be finite is to have a finite state space X . ■

We have established that the Delta- π Myopic Policy is a one-step application of the classical policy improvement iteration in Markov Decision Processes.

3.2 Computational Implementation

We turn now to the question of performing the minimization in Equation 3.5, that is of finding

$$\delta_t(\pi)(x) = \arg \min_{y \in R_t(x)} \mathbb{E}[\Delta_t^\pi(x, y)|\mathcal{F}_t].$$

We restrict our attention to the finite horizon, linear cost, stochastic inventory problem with backordering as outlined in Chapter 1.

Theorem 3.2.1 *For the stochastic inventory problem with backordering, the function $\Delta_t^\pi(x, y)$ is convex in y*

Proof : Recall from Equation 3.2 that $\Delta_t^\pi(x, y) = C_t^\pi(t, x, y) - C_t^\pi(t, x)$. The term $C_t^\pi(t, x)$ is independent of y and thus it suffices to show that $C_t^\pi(t, x, y)$ is

convex in y . Recall that

$$C_t^\pi(t, x, y) = \sum_{s=t}^T c_s(x_s^\pi(t, x, y), y_s^\pi(t, x, y)).$$

As the transition function $T_t(y)$ for this problem is the mapping $T_t(y) = y - D_t$, we can see that $y_s^\pi(t, x, y)$ is an increasing function of y . We know also that $c_s(x, y)$ is convex in y for each s . Thus $C_t^\pi(t, x, y)$ is a sum of convex functions in y and, hence, is itself convex in y . ■

In addition, we assume that the reference policy π is a base-stock policy. In this case we have the following result

Theorem 3.2.2 *If the reference policy π is a Base-Stock Policy, then so too is the Delta- π Myopic Policy, $\delta_t(\pi)$.*

Proof : For the Delta- π Myopic Policy we seek

$$y_t^{\delta_t(\pi)}(x) = \arg \min_{y \geq x} \mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t]$$

Applying the same logic as that in Lemma 1.1.2 (which showed that the optimal policy for our problem is a base-stock policy) we know from the fact that $\Delta_t^\pi(x, y)$ is convex that if

$$\bar{y}_t^{\delta_t(\pi)}(x) = \arg \min_{y \in \mathbb{R}} \mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t]$$

is the unconstrained minimizer of $\mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t]$, then the decision of the policy $\delta_t(\pi)$ is

$$y_t^{\delta_t(\pi)}(x) = x \vee \bar{y}_t^{\delta_t(\pi)}(x),$$

that is, the Delta- π Myopic Policy is a base-stock policy. ■

The differentiability of $C_t^\pi(t, x, y)$ is a consequence of convexity (see Rockafellar [52], for example). It follows that in making our decision under the Delta- π Myopic

Policy, we must find $\bar{y}_t^{\delta_t(\pi)}(x)$, the base-stock level y that solves

$$\frac{\partial}{\partial y} \mathbb{E}[C_t^\pi(t, x, y) | \mathcal{F}_t] = 0. \quad (3.6)$$

We will use the idea of a reversion time ρ introduced in Section 1.2.2 to allow us to perform this calculation more efficiently. The general method for the calculation is the Monte Carlo approach of *Infinitesimal Perturbation Analysis*.

3.2.1 Infinitesimal Perturbation Analysis

The technique of Infinitesimal Perturbation Analysis (IPA) was pioneered by Ho and his co-authors ([26], [27] and [28]). Glasserman ([18]) is a detailed study of the methodology and Glasserman and Tayur ([19]) provide an application of IPA to an inventory problem. We follow the latter two sources in this exposition.

Suppose that we have a random variable X , the value of which depends both on the random quantity $\omega \in \Omega$ and a parameter $\theta \in \Theta$. We wish to know the value of

$$\mathbb{E}[X(\theta, \omega)]' = \frac{\partial}{\partial \theta} \mathbb{E}[X(\theta, \omega)].$$

IPA provides a simulation based method of evaluating this quantity. Specifically, we draw N i.i.d. samples from Ω , denoted $\{\omega^i\}_{i=1}^N$ and approximate

$$\mathbb{E}[X(\theta, \omega)]' \approx \frac{1}{N} \sum_{i=1}^N X'(\theta, \omega^i).$$

The power of IPA lies in the fact that given knowledge of ω_i , the quantity $X'(\theta, \omega^i)$ is very easy to compute. We shall see shortly that it certainly is for our specific example.

The only technical assumption to be checked in this approximation is the unbiasedness of the random derivative, that is whether it is true that

$$\mathbb{E}[X(\theta, \omega)]' = \mathbb{E}[X'(\theta, \omega)].$$

This is satisfied provided that a few technical conditions are met, one of which is that, as a function of θ , we have that $X(\theta, \omega)$ is almost surely *Lipschitz*.

Definition 3.2.3 *A function f , mapping Θ to \mathbb{R} , is Lipschitz if there exists a constant K_f , called the Lipschitz modulus, such that*

$$|f(x) - f(y)| \leq K_f |x - y|$$

for every $x, y \in \Theta$.

The full conditions are given in the following lemma (which is a rephrasing of Lemma 3.2 in Glasserman and Tayur ([19])).

Lemma 3.2.4 *Let $\{X(\theta, \omega), \theta \in \Theta, \omega \in \Omega\}$ be a random function with Θ an open subset of \mathbb{R} . Suppose that the following conditions hold:*

1. $\mathbb{E}[X(\theta, \omega)] < \infty$ for all $\theta \in \Theta$.
2. X is differentiable at $\theta_0 \in \Theta$ with probability one.
3. X is almost surely Lipschitz with modulus K_X satisfying $\mathbb{E}[K_X] < \infty$.

Then $\mathbb{E}[X(\theta_0, \omega)]'$ exists and equals $\mathbb{E}[X'(\theta_0, \omega)]$.

Proof : From the Lipschitz property we know that

$$\frac{|X(\theta_0 + \delta, \omega) - X(\theta_0, \omega)|}{\delta} \leq K_X$$

for all $\theta_0 + \delta \in \Theta$ and $\omega \in \Omega$ (except for sets of measure zero). As $\mathbb{E}[K_X] < \infty$, by the dominated convergence theorem,

$$\mathbb{E}[X(\theta_0, \omega)]' = \lim_{\delta \rightarrow 0} \frac{\mathbb{E}[X(\theta_0 + \delta, \omega)] - \mathbb{E}[X(\theta_0, \omega)]}{\delta}$$

exists and equals

$$\mathbb{E} \left(\lim_{\delta \rightarrow 0} \frac{|X(\theta_0 + \delta, \omega) - X(\theta_0, \omega)|}{\delta} \right) = \mathbb{E}[X'(\theta_0, \omega)].$$

■

While discussing the IPA method, we will make explicit the dependence on ω in our random variables by writing, for example, $C_t^\pi(t, x, y, \omega)$. In order for the method to be valid, we must verify that $C_t^\pi(t, x, y, \omega)$ satisfies the conditions of Lemma 3.2.4. We will defer this verification until after we have investigated the form of $C_t^\pi(t, x, y, \omega^i)$ for a particular sample ω^i .

3.2.2 Value of $\frac{\partial}{\partial y} \mathbb{E}[C_t^\pi(t, x, y)]$ on sample paths

From Section 3.2.1 we know that we can compute $\frac{\partial}{\partial y} \mathbb{E}[C_t^\pi(t, x, y) | \mathcal{F}_t]$ by generating N random samples of the random variable $(C_t^\pi)'(t, x, y)$, given the information \mathcal{F}_t . The N samples are used to form the following approximation:

$$\frac{\partial}{\partial y} \mathbb{E}[C_t^\pi(t, x, y) | \mathcal{F}_t] \approx \frac{1}{N} \sum_{i=1}^N (C_t^\pi)'(t, x, y, \omega^i)$$

All that is necessary for us to compute $(C_t^\pi)'(t, x, y, \omega^i)$ is the availability of an independent sampling of the demand up to the end of the horizon, denoted by $\{D_t(\omega^i), D_{t+1}(\omega^i), \dots, D_{T+L}(\omega^i)\}$ and also the target inventory levels of the policy π , denoted by $\{(\bar{y}_t^\pi(\omega^i), \bar{y}_{t+1}^\pi(\omega^i), \dots, \bar{y}_T^\pi(\omega^i))\}$. In Section 5.2 we will show how to compute these for a specific model of demand and three choices of reference policy π . For the moment, we assume that these are available.

To compute the value of $(C_t^\pi)'(t, x, y, \omega^i)$, we compute two *threshold values*. The first of these is the *reversion time* which is the time at which our jolted policy reverts to following the decisions of the reference policy. Specifically, we define the

reversion time, $\rho^\pi(t, x, y, \omega_i)$, as

$$\rho^\pi(t, x, y, \omega_i) = \min \{s : s > t, y - D_{[t,s]}(\omega_i) < \bar{y}_s^\pi \text{ or } s = T + 1\} \quad (3.7)$$

where \bar{y}_s^π is the target inventory level of our base-stock reference policy π in period s as seen from period t .

Remark 3.2.5 *This definition of reversion is what allows us to compute $(C_t^\pi)'(t, x, y, \omega^i)$ efficiently. It relies crucially on the fact that π is a base-stock policy, as this means that at reversion, the evolution is effectively reset and all memory of the jolted level, y , have dissipated. Thus, changing y has no effect on the subsequent cost, and we can ignore periods after reversion. We use this fact in Proposition 3.2.6.*

For the second threshold, we imagine a policy where we order up to level y and then place no further orders. The threshold is the first period in which we would stop incurring holding costs and instead incur backorder costs under this ordering policy. We call this second threshold the *stock-out time*. We denote it by $\mu(t, y, \omega^i)$:

$$\mu(t, y, \omega^i) = \min \{s : s \geq t, y < D_{[t,s+L]}(\omega^i)\}$$

with $\mu(t, y, \omega^i) = t$ if $y < D_{[t,t+L]}(\omega^i)$.

We assume that we know the full sample path of demands and target inventory levels and thus we can compute both $\rho(t, y, \omega^i)$ and $\mu(t, y, \omega^i)$.

With these thresholds defined, we can state the result that allows us to perform the computation.

Proposition 3.2.6 *Except for y taking values in the set $\mathcal{D} = \{D_{[t,s+L]}(\omega^i) : s \geq$*

$t\}$, we have

$$\frac{\partial}{\partial y}(C_t^\pi)(t, x, y, \omega^i) = \left(\sum_{j=t}^{(\mu(t,y,\omega^i)-1) \wedge (\rho^\pi(t,x,y,\omega^i)-1)} h_{j+L} - \sum_{j=\mu(t,y,\omega^i)}^{\rho^\pi(t,x,y,\omega^i)-1} b_{j+L} \right).$$

Proof :

Let $\Gamma_t(y, \omega^i)$ be the cost of ordering up to y in period t and placing no further orders until the end of the horizon T . From the definition of $\mu(t, y, \omega^i)$ we know that

$$\Gamma_t(y) = \sum_{j=t}^{\mu(t,y,\omega^i)-1} h_{j+L}[y - D_{[t,j+L]}(\omega^i)] + \sum_{j=\mu(t,y,\omega^i)}^T b_{j+L}[D_{[t,j+L]}(\omega^i) - y].$$

Consider dy small enough so that $\mu(t, y + dy, \omega^i) = \mu(t, y, \omega^i)$. This is possible because $y \notin \mathcal{D}$ and hence there exists s such that $D_{[t,s+L]} < y < D_{[t,s+1+L]} = D_{[t,s+L]} + D_{s+1+L}$. Thus

$$\begin{aligned} \Gamma_t(y + dy) &= \sum_{j=t}^{\mu(t,y,\omega^i)-1} h_{j+L}[(y + dy) - D_{[t,j+L]}(\omega^i)] \\ &+ \sum_{j=\mu(t,y,\omega^i)}^T b_{j+L}[D_{[t,j+L]}(\omega^i) - (y + dy)]. \end{aligned}$$

Thus

$$\Gamma(y + dy) - \Gamma(y) = dy \left(\sum_{j=t}^{\mu(t,y,\omega^i)-1} h_{j+L} - \sum_{j=\mu(t,y,\omega^i)}^T b_{j+L} \right). \quad (3.8)$$

We are interested in the value of

$$\frac{\partial}{\partial y}(C_t^\pi)(t, x, y, \omega^i) = \lim_{dy \rightarrow 0} \frac{C_t^\pi(t, x, y + dy, \omega^i) - C_t^\pi(t, x, y, \omega^i)}{dy}$$

But the value of $C_t^\pi(t, x, y + dy, \omega^i) - C_t^\pi(t, x, y, \omega^i)$ is simply the expression on the right hand side of Equation 3.8, truncated at $\rho^\pi(t, x, y, \omega^i) - 1$, as that is the point at which we revert to the policy π and changes in y no longer affect the cost.

Thus, we have $\frac{1}{dy}[C_t^\pi(t, x, y + dy, \omega^i) - C_t^\pi(t, x, y, \omega^i)]$ equals

$$\left(\sum_{j=t}^{\mu(t,y,\omega^i) \vee (\rho^\pi(t,x,y,\omega^i)-1)} h_{j+L} - \sum_{j=\mu(t,y,\omega^i)+1}^{\rho^\pi(t,x,y,\omega^i)-1} b_{j+L} \right)$$

and the result follows. ■

Definition 3.2.7 For $y \in \mathcal{D}$, we define $\frac{\partial}{\partial y}(C_t^\pi)(t, x, y, \omega^i)$ to be the right hand derivative, that is

$$\lim_{\epsilon \searrow 0} \left[\frac{\partial}{\partial y}(C_t^\pi)(t, x, y + \epsilon, \omega^i) \right]$$

Example An example will help to illustrate this proposition. Consider the case where $L = 0$, $T = 4$ and we are in period $t = 1$ with $x_0 = 0$. Suppose that the holding cost per unit per period is stationary with value $h_t = 1$ and the backorder cost per unit per period is also stationary with value $b_t = 10$. Suppose that a specific sample ω^i gives demand values $\{D_j\}_{j=1}^4 = \{40, 40, 40, 40\}$ and values for the target inventory level for π of $\{y_j^\pi(t, x)\}_{j=1}^4 = \{20, 20, 20, 20\}$.

To simplify the notation, we suppress the dependence on all parameters other than y , that is we write $C(y)$ for $C_1^\pi(1, 0, y, \omega_i)$, $\rho(y)$ for $\rho_\pi(1, 0, y, \omega_i)$ and $\mu(y)$ for $\mu^\pi(1, y, \omega_i)$. By $C'(y)$ we mean $\frac{\partial}{\partial y}C(y)$.

We will compute explicitly the value of $C(y)$, the total cost to the end of the horizon, for various values of y . Then by altering the value to $y + \delta$ for a sufficiently small $|\delta|$ (for the examples given, $|\delta| \leq 1$ is sufficient) we can compute the value of $C'(y)$.

Table 3.1 summarizes the detailed computation for various values of y . To understand the table, let us go through in detail what happens when $y = 45$ (this is Case 2). We order 45 units in period 1. These arrive immediately ($L = 0$ here). Then a demand of 40 is realized and so the inventory position at the end of the

Table 3.1: Sample path values of $C'(y)$: Hand-worked evolution of inventory levels when ordering up to y

Case		t=1	t=2	t=3	t=4	Cost
Case 1: $y = 25$	y	25	20	20	20	
	x	-15	-20	-20	-20	750
Case 2: $y = 45$	y	45	20	20	20	
	x	5	-20	-20	-20	605
Case 3: $y = 65$	y	65	25	20	20	
	x	25	-15	-20	-20	575
Case 4: $y = 85$	y	85	45	20	20	
	x	45	5	-20	-20	450

period is 5. This value is given in the row labeled x and column $t = 1$). Then in period 2, as 5 is below 20, the policy π 's target inventory level in period 2, we order up to 20. This value is given in the row labeled y and column $t = 2$. The demand of $D_2 = 40$ is realized, giving an inventory position of -20 . We continue in this fashion to complete the evolution of x and y values. Finally, the cost is computed by evaluating $\sum_{t=1}^4 \{(x_t)^+ + 10(x_t)^-\}$.

Table 3.2 describe the same computation but for $y + \delta$. Reading the costs from these two tables we see that, for small enough $|\delta|$, $C(45 + \delta) - C(45) = 1(\delta)$ and hence $C'(45) = 1$.

Let us now check that this is the same value we get from Equation 3.9. When $y = 45$ reversion has occurred by period 2 and hence $\rho(45) = 2$. Also, if we ordered 45 units in period 1 and placed no further orders, there would be positive inventory in the system in period 1 and negative inventory in periods 2, 3 and 4.

Table 3.2: Sample path values of $(\Delta^\pi)'$: Hand-worked evolution of inventory levels when ordering up to $y + \delta$

Case		t=1	t=2	t=3	t=4	Cost
Case 1: $y = 25$	y	$25 + \delta$	20	20	20	
	x	$-15 - \delta$	-20	-20	-20	$750 - 10(\delta)$
Case 2: $y = 45$	y	$45 + \delta$	20	20	20	
	x	$5 + \delta$	-20	-20	-20	$605 + 1(\delta)$
Case 3: $y = 65$	y	$65 + \delta$	$25 + \delta$	20	20	
	x	$25 + \delta$	$-15 - \delta$	-20	-20	$575 + 1(\delta) - 10(\delta)$
Case 4: $y = 85$	y	$85 + \delta$	$45 + \delta$	20	20	
	x	$45 + \delta$	$5 + \delta$	-20	-20	$450 + 1(\delta) + 1(\delta)$

Hence $\mu(45) = 1$ and thus

$$C'(45) = \sum_{j=1}^{\mu(45) \vee (\rho(45)-1)} h_j - \sum_{j=\mu(45)+1}^{\rho(45)-1} b_j = \sum_{j=1}^1 (1) = 1 \quad (3.9)$$

Table 3.3 gives the value of $\rho(y)$, $\mu(y)$ and $C'(y)$ for our four cases.

As well as helping to illustrate Equation 3.9, this example is important in that it shows that, path-wise, $C'(y)$ is not increasing in y . Thus, although we noted that $\mathbb{E}[\Delta_t^\pi(x, y) | \mathcal{F}_t]$ is convex in y , this is not necessarily true *path-wise*.

Table 3.3: Values of $\rho(y)$, $\mu(y)$ and $C'(y)$.

y	$\rho(y)$	$\mu(y)$	$C'(y)$
25	1	0	-10
45	1	1	1
65	2	1	-9
85	2	2	2

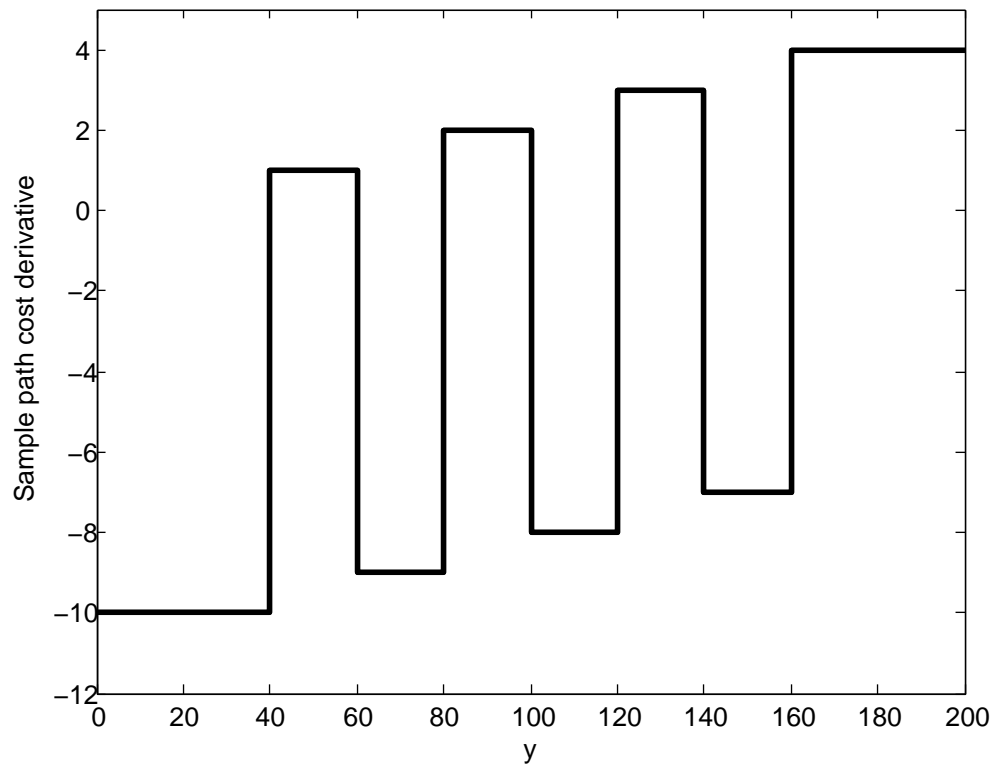


Figure 3.1: Plot of $(C'(y))$ against y for our example.

Figure 3.1 illustrates this; it plots the value of $C'(y)$ against y for our example value of ω^i . It is clearly not increasing.

MATLAB implementation The code for our computational experiments is implemented in MATLAB. MATLAB is optimized for greatest efficiency when code is vectorized. Thus, at each stage in computing $\bar{y}_t^{\delta(\pi)}$ we generate our independent samples of demand and base stock levels \bar{y}_t^π . For each sample, we compute the vector of breakpoints at which the values of $\rho(y)$ and $\mu(y)$ change. By placing these in a matrix, we can efficiently compute $\rho(y)$ and $\mu(y)$ for any given y . Combining this with a mapping of ρ and μ to $C'(\cdot)$ gives a very efficient implementation. In particular, it is significantly faster than the more usual method of using a `for`-loop type implementation.

Verification of Infinitesimal Perturbation Analysis We turn now to verifying the IPA method is valid for our problem. For this purpose we will assume that we select y from some open subset of \mathbb{R} , denoted \mathbb{Y} . That is, we assume that $|y| < M$ for some large M . We do not lose anything by thus restricting our choice of y because the problem we are studying has a finite horizon and the demand variables have finite means, meaning that sample paths for which we would have $y > M$ have close to measure zero.

Lemma 3.2.8 *The function $C_t^\pi(t, x, y, \omega_t)$ satisfies the conditions of Lemma 3.2.4.*

That is, for any $y_0 \in \mathbb{R}$ we have

1. $\mathbb{E}[C_t^\pi(t, x, y, \omega^i) | \mathcal{F}_t] < \infty$ for all $y \in \mathbb{R}$.
2. $C_t^\pi(t, x, y, \omega^i)$ is differentiable at $y_0 \in \mathbb{R}$ with probability one.

3. $C_t^\pi(t, x, y, \omega^i)$ is almost surely Lipschitz with modulus K_C satisfying $\mathbb{E}[K_C] < \infty$.

Proof :

1. The finiteness of $\mathbb{E}[C_t^\pi(t, x, y, \omega^i)|\mathcal{F}_t]$ follows from the finiteness of demand, costs (both holding and backorder) and the fact that we study the finite-horizon problem.
2. From our discussion above, we know that on a sample path basis (that is, given ω^i), $(C_t^\pi)'(t, x, y, \omega^i)$ is piecewise-constant with a finite set of breakpoints, \mathcal{D} . This set of breakpoints has measure zero thus and $C_t^\pi(t, x, y, \omega^i)$ is differentiable at $y_0 \in \mathbb{Y}$ with probability one.
3. We can (loosely) bound the sample-path derivative, $(C_t^\pi)'(t, x, y, \omega^i)$ by

$$K_C = \sum_{j=t}^T (h_{j+L} + b_{j+L}) < \infty.$$

The result follows. ■

Summary In this Chapter, we introduced our Delta- π Myopic policy, showing how it leads to the formulation of an equivalent dynamic program. We provided justification for its usage by proving it has desirable theoretical properties, namely that it converges to the optimal policy and is related to classical policy improvement schemes.

We then considered how the policy could be calculated in the context of the stochastic inventory problem by using Infinitesimal Perturbation Analysis. In the

case where the reference policies are base-stock policies, this is accomplished efficiently by the use of two thresholds in time, the reversion time and the stock-out time.

Chapter 4

Extensions to the Minimizing and Balancing Policies

Overview In this Chapter we present our extensions to the Minimizing and Balancing policies and prove various theoretical results about them.

In Equation 1.10, we give an upper and lower bound on the optimal post-order inventory level level, that is

$$y_t^m(x) \leq y_t^*(x) \leq y_t^M(x)$$

. This section exploits that bound in three ways:

1. *Parameterization:* Here we develop a method of parameterizing the space between $[y^m(x), y^M(x)]$. In addition, we describe dynamic methods of choosing the relevant parameter.
2. *Bounding* If a policy suggests an order that falls outside $[y^m(x), y^M(x)]$ we can bound or truncate the order at the appropriate endpoint.
3. *Surplus Balancing* While the original Balancing Policy balances the back-order cost with the marginal holding cost for all units ordered, we instead consider the marginal holding cost only for those units ordered above the order suggested by the Minimizing Policy.

4.1 Parameterization

Our idea of parameterization seeks to parameterize the space between $y_t^m(x)$ and $y_t^M(x)$.

Recall that the $h_t(x, y)$ function captures the expected marginal holding cost over the period $[t + L, T + L]$ from ordering up to the level y . Suppose that instead of looking to the end of the horizon we look ahead k periods (or to the end of the horizon, whichever comes first), i.e. we consider the expected holding cost over the periods $[t + L, (t + L - 1 + k) \wedge T]$.

This value k need not be restricted to integer values. To count the marginal holding cost over k periods into the future we define

$$\begin{aligned} h_t^k(x, y) &= \mathbb{E} \left[\sum_{k=t}^{\lfloor k \rfloor + t} h_{j+L} [(y - D_{[t, j+L]})^+ - (x - D_{[t, j+L]})^+] \right. \\ &\quad \left. + (k - \lfloor k \rfloor) \{ h_{\lceil k \rceil + t + L} [(y - D_{[t, \lceil k \rceil + t + L]})^+ - (x - D_{[t, \lceil k \rceil + t + L]})^+] \} \mid \mathcal{F}_t \right] \end{aligned}$$

where the floor function $\lfloor k \rfloor$ is the greatest integer less than or equal to k and the ceiling function $\lceil k \rceil$ is the smallest integer greater than or equal to k .

We now generalize the Minimizing Policy introduced in Section 1.2.1 and define the *Minimizing(k)* family of policies. Denoted by $m(k)$, these choose the post-order inventory level analogously to Equation 1.9, namely by setting:

$$y_t^{m(k)}(x) = \arg \min_{y \geq x} \{ h_t^k(x, y) + b_t(y) \} \quad (4.1)$$

Just as the Minimizing Policy is a base-stock policy, each of the Minimizing(k)-Policies is a base-stock policy with base-stock level given by

$$\bar{y}_t^{m(k)}(x) = \arg \min_y \{ h_t^k(x, y) + b_t(y) \}.$$

Moreover, the family of policies $m(k)$ will give a set of base stock levels that are decreasing in k . Specifically, we have the following lemma:

Lemma 4.1.1 *The longer a look-ahead period k we use in $m(k)$, the lower the post-order inventory level, i.e. for $k_1 \geq k_2$ we have $\bar{y}_t^{m(k_1)}(x) \leq \bar{y}_t^{m(k_2)}(x)$.*

Proof : Define $l_t^k(x, y) = h_t^k(x, y) + b_t(y)$ and note that this is a convex function. The value $\bar{y}_t^{m(k)}(x)$ minimizes this function. We assume that it always chooses the smallest minimizer, i.e. $\bar{y}_t^{m(k)}(x) = \min\{\arg \min l_t^k(x, y)\}$.

Suppose that the lemma is false, i.e. $\bar{y}_t^{m(k_1)}(x) > \bar{y}_t^{m(k_2)}(x)$.

Define $d_t^{k_1}(x, y) = l_t^{k_1}(x, y) - l_t^{k_2}(x, y)$. It is easy to see that this is an increasing convex function of y .

Thus, when we evaluate the partial derivative of $l_t^{k_1}(x, \cdot)$ at $\bar{y}_t^{m(k_2)}(x)$, we see

$$\begin{aligned} \left[l_t^{k_1} \left(x, \bar{y}_t^{m(k_2)}(x) \right) \right]' &= \left[l_t^{k_2} \left(x, \bar{y}_t^{m(k_2)}(x) \right) \right]' + \left[d_t^{k_1} \left(x, \bar{y}_t^{m(k_2)}(x) \right) \right]' \\ &> 0 \end{aligned}$$

where the inequality is strict because if $\left[l_t^{k_1} \left(x, \bar{y}_t^{m(k_2)}(x) \right) \right]' = 0$ then $\bar{y}_t^{m(k_1)}(x) \leq \bar{y}_t^{m(k_2)}(x)$ as we choose the smallest minimizer; this gives a contradiction.

Note that $\left[l_t^{k_1} \left(x, y_t^{m(k_1)}(x) \right) \right]' = 0$. Also $l_t^{k_1}(x, y)$ is convex. As we assumed $y_t^{k_1}(x) > y_t^{k_2}(x)$, this gives a contradiction. \blacksquare

Note that $m(1)$ is the Myopic Policy, i.e. $M = m(1)$. Also, by definition $m = m(T - t)$, where m is the original Minimizing Policy. Thus we have

$$y_t^m(x) = y_t^{M(T-t)}(x) \leq y_t^{M(T-(t-1))}(x) \leq \dots \leq y_t^{m(2)}(x) \leq y_t^{m(1)}(x) = y_t^M(x) \quad (4.2)$$

Thus, we have developed a method of parameterizing the space between $[y_t^m(x), y_t^M(x)]$.

While choosing a static k to use in all periods may give a good policy, it is natural to try to think of dynamic methods of choosing k .

Run-Out Recall in Equation 3.8 we defined the stock-out time to be

$$\mu_t(y) = \min \{ s : s \geq t, y < D_{[t, s+L]} \}$$

with $\mu_t(y) = t$ if $y < D_{[t,t+L]}$. (We have simplified the notation slightly.)

The stock-out time gives the point in time that all units in our post-order inventory level, y , have been consumed. Thus, the time to stock-out, or *run-out time* of the units is $\mu_t(y) - t$. Then the expected post lead-time run-out of the quantity y (where $y > 0$) is given by:

$$r_t(y) = \mathbb{E} \{ (\mu_t(y) - (t + L))^+ | \mathcal{F}_t \}.$$

If we assume that we consume our orders on a *first-ordered, first-consumed* basis, then $r_t(y)$ is the expected time that the final unit among the y is consumed. We can extend this interpretation of run-out time from applying to a single unit ordered to the average post lead-time run-out time of a range of units. Specifically, define $r_t([y_1, y_2]) = r_t(y_2) - r_t(y_1)$ for $0 \leq y_1 \leq y_2$. This quantity gives us the average post-lead-time run-out of the quantity $y_2 - y_1$.

Dynamic method of choosing k We now describe three methods of dynamically choosing k . The basic idea is to choose k , the number of periods we look ahead, to equal some measure of expected run-out, given that we choose our post-order inventory level to be $y_t^{m(k)}$. We take three measures of run-out.

- (i) *Final unit run-out:* Here we use the measure that corresponds to the run-out of the final unit of inventory ordered by $m(k)$. That is, we solve for k^* in:

$$k = r_t(y_t^{m(k)}).$$

- (ii) *Average marginal units run-out:* Here we choose k to be the average post-lead-time run-out time of the marginal quantity ordered at time t by $m(k)$, that is the quantity $y_t^{m(k)} - x_t$. That is, we solve for k^* in:

$$k = \frac{r([x_t^+, y_t^{m(k)}])}{y_t^{m(k)} - x_t^+}.$$

- (iii) *Average total units run-out:* Here we choose k to be the average post-lead-time run-out time of the inventory that was on-hand or on-order after ordering at time t , according to $m(k)$. That is, we solve for k^* in:

$$k = \frac{r([0, y_t^{m(k)}])}{y_t^{m(k)}}.$$

It remains to show that such k 's exists.

Lemma 4.1.2 *The equation*

$$k = r_t(y_t^{m(k)})$$

has a solution.

Proof : We seek a zero of the function $f(k) = r_t(y_t^{m(k)}) - k$. Lemma 4.1.1 demonstrates that $y_t^{m(k)}$ is decreasing in k ; hence $r_t(y_t^{m(k)})$ is also and so $f(k)$ is decreasing.

For $k = 0$, the function $h_t^0(x, y) \equiv 0$ and we have $y_t^{m(0)}(x) = \arg \min_{y \geq x} b_t(y)$. The solution to this is \bar{D} , the highest attainable demand, or ∞ if there is no such upper bound. In this case, $r_t(y_t^{m(0)}) > 0$ and hence $f(0) > 0$.

Again from Lemma 4.1.1, we know that for $k \geq T - t$, $m(k) = m$ and thus $f(k) \rightarrow -\infty$.

In addition, $f(k)$ is continuous and hence there is a k^* such that $f(k^*) = 0$. ■

The proof for the other two cases is similar. We need the additional assumption that $\mathbb{E}[D_t] \geq 1$ for each t . This ensures that $r_t([0, y])$ increases at a faster rate than y and hence that the equivalent $f(k)$ function is decreasing in k .

In practice, we find k^* by means of a bisection method.

Denote by k -fin, k -mar and k -tot the final values of k for the *final*, *marginal* and *total* methods, respectively. This gives us three new policies that choose k dynamically, denoted $m(k\text{-fin})$, $m(k\text{-mar})$ and $m(k\text{-tot})$.

4.2 Bounding

The motivation for this idea is simple. As before, we know that the optimal policy should order so as the post-order inventory level is in the range $[y^m(x), y^M(x)]$ and so if our policy suggests an ordering level that falls outside this range, we bound or truncate the order at the appropriate endpoint.

Specifically, given any policy π , define the bounded version of π , denoted $\hat{\pi}$, as the policy that orders to a post-order inventory level of $y_t^{\hat{\pi}}(x) = y_t^m(x)$ if $y_t^\pi(x) \leq y_t^m(x)$, orders up to $y_t^{\hat{\pi}}(x) = y_t^M(x)$ if $y_t^\pi(x) \geq y_t^M(x)$ and orders up to $y_t^{\hat{\pi}}(x) = y_t^\pi(x)$ otherwise.

Clearly, this bounding will not affect any of the Minimizing Policies. It is important to note that the effect of bounding in a particular period will be felt in future periods as well. As an example, if the original Balancing Policy gave an order of $y_t^B(x) > y_t^M(x)$ then changing the order to $y_t^{\hat{B}}(x) = y_t^M(x)$ will affect the decision in period $t + 1$ as the inventory position $x_{t+1}^{\hat{B}}$ will be reduced.

Our bounding procedure will always improve the performance of a policy, as the following results demonstrate.

Proposition 4.2.1 *Let π be some feasible policy. Suppose that for some possible state (t, x_t, \mathcal{F}_t) the policy π orders above the myopic base-stock level $\bar{y}_t^M(x)$. Consider the policy $\hat{\pi}$ that for the state $(t, x_t^\pi, \mathcal{F}_t)$ orders only up to $\bar{y}_t^M(x)$ (and orders nothing if $x_t \geq \bar{y}_t^M(x)$), and for any other state (s, x_s, \mathcal{F}_s) aims to order up y_s^π (if*

y_s^π is reachable). Then $\hat{\pi}$ has expected cost no larger than the expected cost of π .

Proof : If the state (x_t, \mathcal{F}_t) does not occur then it is clear that the two policies have the same cost. Suppose now that (x_t, \mathcal{F}_t) does occur. Recall that we assume $c_s = 0$ and $h_s \geq 0$ for each $s = 1, \dots, T$. It is clear that over the interval $[1, t)$ the two policies π and $\hat{\pi}$ are identical and hence incur exactly the same cost. Since, by the definition of the Myopic Policy, $\hat{\pi}$ orders to minimize the expected overall holding and backlogging cost in period t , we conclude that the expected holding and backlogging cost of $\hat{\pi}$ in period t is at most the expected holding and backlogging of $\hat{\pi}$ in period t . Moreover, $x_{t+1}^{\hat{\pi}} \leq x_{t+1}^\pi \leq y_{t+1}^\pi$. Recall that $\hat{\pi}$ aims to order up to y_s^π in any other state (s, x_s, \mathcal{F}_s) and, as $x_{t+1}^{\hat{\pi}} \leq y_{t+1}^\pi$ we have that $y_{t+1}^\pi = y_{t+1}^{\hat{\pi}}$ and hence π and $\hat{\pi}$ incur the same cost over the interval $(t, T]$. The proof then follows. \blacksquare

Proposition 4.2.2 *Let π be some feasible policy. Suppose that for some possible state (t, x_t, \mathcal{F}_t) the policy π orders below the minimizing base-stock level $\bar{y}_t^m(x)$. Consider the policy $\hat{\pi}$ that for the state $(t, x_t^\pi, \mathcal{F}_t)$ orders to $\bar{y}_t^m(x)$ (and orders nothing if $x_t \geq \bar{y}_t^m(x)$), and for any other state (x_s, \mathcal{F}_s) aims to order up to y_s^π (if y_s^π is reachable). Then $\hat{\pi}$ has expected cost no larger than the expected cost of π .*

Proof : Consider the marginal cost accounting scheme. If the state (t, x_t, \mathcal{F}_t) does not occur then it is clear that the two policies have the same cost. Assume that (t, x_t, \mathcal{F}_t) does occur. Again it is clear that over $[1, t)$ the two policies π and $\hat{\pi}$ incur the same cost and that $x_t^\pi = x_t^{\hat{\pi}} = x_t$. By the definition of $\bar{y}_t^m(x_t)$ we conclude that,

$$E[h_t(x_t, y_t^{\hat{\pi}}(x_t)) + b_t(y_t^{\hat{\pi}}(x_t))] \leq E[h_t(x_t, y_t^\pi(x_t)) + b_t(y_t^\pi(x_t))].$$

Moreover, $x_{t+1}^{\hat{\pi}} > x_{t+1}^{\pi}$. Since after period t the policy $\hat{\pi}$ imitates π it is clear that $y_s^{\hat{\pi}} \geq y_s^{\pi}$ for each $s > t$. This means that the backorder cost $b_t(y_s)$ in each such period s will be lower under policy $\hat{\pi}$. In addition, for each $s > t$, as we aim to order up to y_s^{π} and in period $t+1$ we have a higher pre-order inventory level ($x_{t+1}^{\hat{\pi}} > x_{t+1}^{\pi}$), we know that we will order no more than π , that is $y_s^{\hat{\pi}} - x_s^{\hat{\pi}} \leq y_s^{\pi} - x_s^{\pi}$. This implies that the marginal holding cost $h_s(x_s, y_s)$ will also be lower under policy $\hat{\pi}$. The result follows. ■

Corollary 4.2.3 *The bounded version of the Dual-Balancing Policy provides a worst-case performance guarantee of 2.*

Proof : By repeated application of Theorems 4.2.1 and 4.2.2 we see that the bounding procedure always improves the performance of a policy. We already know from Levi *et al.* [39] that the Dual-Balancing Policy has a worst-case performance guarantee of 2 and thus this must also hold for the bounded version. ■

4.3 Surplus balancing

Under the original Balancing Policy the function $h_t(x, y)$ defined in Equation 1.8 measures the total marginal holding cost associated with the entire quantity ordered, that is $y - x$. However, from the bound in Equation 1.10 we know that we should always order at least enough to have a post-order inventory level of $y_t^m(x)$. Thus, we consider instead the total marginal holding cost associated with a surplus order in excess of that amount. That is, we take $x = y_t^m(x)$ and plug it into $h_t(x, y)$, as defined in Equation 1.8 giving

$$h_t(y_t^m(x), y) = \mathbb{E}\left[\sum_{j=t+L}^{T+L} h_j[(y - D_{[t,j]})^+ - (y_t^m(x) - D_{[t,j]})^+]\middle|\mathcal{F}_t\right] \quad (4.3)$$

This allows us to define a new policy, the *Surplus Balancing Policy*, denoted SB , which seeks the order quantity $y_t^{SB}(x)$ to solve the equation

$$h_t(y^m(x), y) = b_t(y)$$

where $b_t(y)$ is the total expected backorder cost in period in period $t + L$, as before.

One of the strengths of the Balancing Policy is its guarantee to have overall expected cost no greater than twice that of the optimal policy. In Hurley *et al.* [30] there is a proof of the following theorem.

Theorem 4.3.1 *The surplus dual-Balancing Policy provides a worst-case performance guarantee of 2.*

Parameterizing the Balancing Policy The Balancing Policy chooses $y_t^B(x) = y_t^B$ where $y_t^B(x)$ satisfies $h_t(x, y_t^B) = b_t(y_t^B)$. If, for some $\alpha > 0$ we instead choose y_t^B , which satisfies $h_t(x, y_t^B) = \alpha b_t(y_t^B)$, this gives us a parameterized family of policies, denoted $B(\alpha)$, giving post-order inventory levels of $y_t^{B(\alpha)}(x)$.

As with the $m(k)$ family of policies, we consider policies based on both fixed values of α and on a dynamic method of choosing α . For the dynamic choice, we compute the ratio of the total expected holding and backorder costs incurred in period $t + L$ by the Myopic Policy. That is, we set

$$\alpha\text{-myo} = \frac{h_t^1(0, y_t^M)}{b_t(y_t^M)}.$$

We denote this policy by $B(\alpha\text{-myo})$

We summarize the extensions to the Minimizing and Balancing Policies and their short-hand names in Table 4.1

Table 4.1: Minimizing and Balancing Policies and Extensions

Policy Name	Description
M	Myopic Policy
B	Balancing Policy
SB	Surplus Balancing Policy
$B(\alpha)$	Balancing Policy that seeks $y_t^B(x)$ such that $h_t(x, y) = \alpha b_t(y)$, for fixed α
$B(\alpha\text{-myo})$	$B(\alpha)$ with α equal to ratio of expected holding and backorder cost for Myopic
m	Minimizing Policy
$m(k)$	Minimizing Policy with holding cost look-ahead of k , for fixed k
$m(k\text{-fin})$	$m(k)$ with $k = k\text{-fin}$, the post lead-time run out of the final unit ordered
$m(k\text{-mar})$	$m(k)$ with $k = k\text{-mar}$, the average post lead-time run out of marginal units ordered
$m(k\text{-tot})$	$m(k)$ with $k = k\text{-tot}$, the average post lead-time run out of current inventory position

Chapter 5

Computation using the Martingale

Model of Forecast Evolution

In Chapter 3 we defined the Delta- π myopic improvement scheme. We described how to use Infinitesimal Perturbation Analysis to compute the post-order inventory level for the Delta- π Myopic Policy. This gives us a powerful tool which combines the computationally tractability of methods such as the Myopic Policy with a method that considers the effect over a sensible number of periods into the future of ordering up to y now.

This work, in Section 3.2, made no assumptions about our model of demand, or the nature of the reference policy, beyond the fact that it is a base-stock policy. In this Chapter, we consider how to apply this method to a specific model of demand provided by the Martingale Model of Forecast Evolution of Heath and Jackson [25] and Graves *et al.* [22] which is in itself a very general model of evolving demand forecasts. In Chapter 4 we considered several computationally tractable policies. In this Chapter, we consider applying Delta- π myopic improvement to three of these policies: the Myopic, Minimizing, and Balancing policies. We have noted that the Balancing Policy is not a base-stock policy; we will remark on our additional assumptions with reference to it later.

There are three parts to this chapter. Firstly, we outline the Martingale Model of Forecast Evolution itself and describe the approximation method we use to compute cumulative demand under it. Secondly, we describe the details of applying our scheme of Section 3.2 to this model of demand for the three reference policies named above. Finally we describe the set of scenarios we used to test both the

Delta- π Myopic scheme and those introduced in Chapter 4. These comprise a rich set of realistic situations an inventory manager may face.

5.1 The Martingale Model of Forecast Evolution (MMFE)

The Martingale Model of Forecast Evolution (MMFE) was developed independently by Heath and Jackson [25] and Graves *et al.* [22]. The process of forecasting is an area of intense interest but our concern here is not specific methods of forecasting. Rather, we assume that some forecasting method is in place and we ask how best to use the evolving forecast information to plan inventory ordering. The MMFE was developed specifically for this purpose, namely that of building models of inventory and production, not for developing forecasting techniques. When used in practice, it requires a phase of data collection and analysis so as to describe how variable forecasts are.

The MMFE neatly captures the idea of a forecast evolving. A forecast for the demand we expect in twelve months time may have a lot of variability but when we are a month away from that demand, the variability may be considerably reduced. The MMFE allows us to model this, as well as the time at which this reduction in variability occurs.

Turning to the specifics, recall from Section 1.1.1 that we denote demand in period t by D_t . We also defined the filtration \mathcal{F}_t which captures all the information we know at time t .

Let $D_{s,t}$ be the forecast of D_t , the demand in period t , that is made at the beginning of period s . For $s > t$, this is no longer a forecast, as D_t is known in periods $t + 1$ and beyond. However, $D_{t,t}$ is still a forecast, as we assume that forecasts occur at the start of the period and demand is realized at the end of the

period.

Denote the vector of current forecasts by ϕ_t , this is given by

$$\phi_t = (D_{t,t}, D_{t,t+1}, D_{t,t+2}, \dots, D_{t,T}). \quad (5.1)$$

The MMFE assumes that the process $\{D_{s,t}\}_{s=1}^t$ is a martingale. Specifically, we assume that

$$D_{s,t} = \mathbb{E}[D_t | \mathcal{F}_s]. \quad (5.2)$$

Saying that the process $\{D_{s,t}\}_{s=1}^t$ is a martingale is equivalent to saying that it stays constant in expectation, as the following computation shows:

$$\begin{aligned} \mathbb{E}[D_{s+1,t} | \mathcal{F}_s] &= \mathbb{E}\{\mathbb{E}[D_t | \mathcal{F}_{s+1}] | \mathcal{F}_s\} \\ &= \mathbb{E}[D_t | \mathcal{F}_s] \\ &= D_{s,t} \end{aligned}$$

In terms of forecasts, this says that our forecasts are unbiased. This is a very natural property for a model of forecasts to have. Suppose instead that

$$\mathbb{E}[D_{s+1,t} | \mathcal{F}_s] > D_{s,t}$$

which says that we expect our period- $s + 1$ forecast of D_t to be larger than our current forecast. But if the forecasting method were biased in this fashion, it is only rational to remove the bias.

Naturally, there are a great many specific forecast processes which would satisfy the MMFE assumption in Equation 5.2. Two specific instances of the MMFE are the *additive* and the *multiplicative* MMFE.

The Additive MMFE In the additive MMFE, we start with an initial forecast vector $\phi_1 = (D_{1,1}, D_{1,2}, \dots, D_{1,T+L})$. Define the *shift operator* on an n -dimensional

vector, $s(\cdot)$, by

$$s((x_1, x_2, x_3, \dots, x_n)) = (x_2, x_3, \dots, x_n)$$

At the start of period t , we have the forecast vector ϕ_t and use it to generate the demand in this period D_t and the new forecast vector ϕ_{t+1} . To do so we generate an update vector $\epsilon_t = (\epsilon_{t,t}, \epsilon_{t,t+1}, \dots, \epsilon_{t,T+L})$ which is multivariate normal with mean $\mathbf{0} = (0, 0, \dots, 0)$ and variance-covariance matrix Σ_t . The demand is given by

$$D_t = D_{t,t} + \epsilon_{t,t} \quad (5.3)$$

and the new forecast vector ϕ_{t+1} is given by

$$\phi_{t+1} = s(\phi_t) + s(\epsilon_t) \quad (5.4)$$

The Multiplicative MMFE As before, at the start of period t , we have the forecast vector $\phi_t = (D_{t,t}, D_{t,t+1}, \dots, D_{t,T+L})$. Now, however, the updates are multiplicative, that is

$$D_t = \gamma_{t,t} D_{t,t} \quad (5.5)$$

and the new forecast vector ϕ_{t+1} is given by

$$\phi_{t+1} = s(\gamma_t) s(\phi_t) \quad (5.6)$$

where the multiplication is component-wise.

The update vector, γ_t is given by $\gamma_t = e^{\epsilon_t}$, where ϵ_t is a multivariate normal random variable with variance-covariance matrix Σ_t and mean $-\frac{\text{diag}(\Sigma_t)}{2}$, where $\text{diag}(\Sigma_t)$ is the vector of diagonal elements of Σ_t . Thus γ_t is a vector of multivariate lognormal random variables of mean $\mathbf{1} = (1, 1, \dots, 1)$. Writing it component-wise, we have $\gamma_t = (\gamma_{t,t}, \gamma_{t,t+1}, \gamma_{t,T+L}) = (e^{\epsilon_{t,t}}, e^{\epsilon_{t,t+1}}, \dots, e^{\epsilon_{t,T+L}})$.

In this work we use the multiplicative MMFE. Apart from a study done by Heath and Jackson [25], to the best of our knowledge, all other computational studies have used the additive MMFE.

We choose the multiplicative MMFE over the additive version for three main reasons. Firstly, for any sensible choice of parameters, there is a significant probability that the additive MMFE will give negative demand values; the multiplicative MMFE never does. Secondly, industry forecasts tend to be updated in a relative sense (as done by the multiplicative MMFE) rather than an absolute sense (as done by the additive MMFE).

The third reason is to study situations in which the Myopic Policy is not known to be optimal. Iida and Zipkin [31] show that under the additive MMFE with stationary forecast updates and stationary costs then when demand is guaranteed to be nonnegative, the Myopic Policy is optimal and the problem we study thus loses much of its richness. No such results hold under the multiplicative MMFE.

5.1.1 Interpreting the Variance-Covariance Matrix

The MMFE has great strength and versatility in allowing one to generate complicated demand processes, including those that exhibit seasonality and correlation.

The variance-covariance matrix used to generate the update vectors, Σ_t , can be time dependent. To interpret the matrix, however, we temporarily assume it is stationary: $\Sigma_t = \Sigma$. At the end of period s , the current forecast of demand in period t is a function of the initial forecast $D_{0,t}$ and the updates that have already occurred, i.e.,

$$D_{s,t} = \left(\prod_{j=1}^s \gamma_{j,t} \right) D_{0,t} = \exp \left(\sum_{j=1}^s \epsilon_{j,t} \right) D_{0,t}.$$

Since we are at the end of period s , all of these quantities are deterministic. Sim-

ilarly, the actual demand that will occur in period t is a function of the current (period- s) forecast and the future updates:

$$D_t = \left(\prod_{j=s+1}^t \gamma_{j,t} \right) D_{s,t} = \exp \left(\sum_{j=s+1}^t \epsilon_{j,t} \right) D_{s,t}.$$

Currently (the end of period s), $D_{s,t}$ is deterministic, and the other quantities are all random. Thus, the set of values $\{\epsilon_{j,t}\}$ can be viewed as the uncertainty about the true demand in period t that will be resolved in time periods $j = s + 1, \dots, t$. Note that the random variables $\{\epsilon_{j,t} : s + 1 \leq j \leq t\}$ are independent because they will be observed in different time periods. The variance of $\epsilon_{j,t}$ is given by the $t - j + 1^{\text{st}}$ diagonal entry of Σ . The sum of the diagonal entries of Σ is a measure of the total uncertainty in the demand. The ratio of the cumulative sum of the first $t - s$ diagonal entries of Σ , to the sum of all the diagonal entries, is the fraction of the variability (i.e., variance) in $\ln(D_t)$ that is still *unresolved* at the end of period s . If all diagonal elements are identical, then the uncertainty is resolved in a linear fashion. We say the process exhibits *constant learning* in this case. We say the process exhibits *late learning* (resp. *early learning*) if proportionally more of the total variance is unresolved in the later (resp. earlier) periods. Equivalently, late learning (resp. early learning) is associated with proportionally higher variances in the first (resp. last) diagonal entries of Σ .

The covariance of $\epsilon_{s,t}$ with $\epsilon_{s,t+1}$ is given by $\sigma_{t-s+1, t-s+2}$, an off-diagonal element of Σ . If this element is positive, then before time s , the forecasts $D_{s,t}$ and $D_{s,t+1}$ will be positively correlated, and the demands D_t and D_{t+1} will be positively correlated. Such correlations exist when, for example, good news causes forecasts for demand in several periods to be revised upwards, or bad news causes forecasts for demand in several periods to be revised downward. A negative correlation arises in practice, for example, when a large forecasted demand is shifted earlier

or later in time.

Cumulative Demand Distribution under the MMFE Given the forecast vector $\phi_s = (D_{s,s}, D_{s,s+1}, \dots, D_{s,T+L})$ we have

$$\begin{aligned}
 D_{[s,t]} &= \sum_{j=s}^t D_j \\
 &= \gamma_{s,s} D_{s,s} + \gamma_{s,s+1} \gamma_{s+1,s+1} D_{s,s+1} + \dots + \gamma_{s,t} \gamma_{s+1,t} \dots \gamma_{t,t} D_{s,t} \\
 &= \sum_{j=s}^t (\prod_{k=s}^j \gamma_{k,j}) D_{s,j} \\
 &= \sum_{j=s}^t \left(\exp \left\{ \log D_{s,j} + \sum_{k=s}^j \epsilon_{k,j} \right\} \right)
 \end{aligned}$$

which is a sum of correlated log-normal random variables. The correlation comes from the correlation between terms in $\epsilon_k = (\epsilon_{k,k}, \epsilon_{k,k+1}, \dots, \epsilon_{k,t}, \dots)$ each of which appears once in the expression for $(D_k, D_{k+1}, \dots, D_t)$ respectively for $s \leq k \leq t$.

There is no closed form expression for the distribution of the sum of lognormal random variables, let alone the sum of correlated lognormal random variables. However, the problem arises in finance (Milevsky and Posner [44]) and wireless technology (Abu-Dayya and Beaulieu [1] and Beaulieu et al. [7]). Abu-Dayya and Beaulieu [1] consider three approximations and demonstrate that among these, an approach due to Wilkinson [56] is the best.

To the best of our knowledge, this is the first time the Wilkinson approximation has been applied in the inventory literature.

Wilkinson's approach As described by Abu-Dayya and Beaulieu [1], the key ideas in Wilkinson's method are firstly that the sum of lognormal random variables $L = e^{Y_1} + e^{Y_2} + \dots + e^{Y_n}$ is well approximated by a single lognormal random variable ($L \approx e^Z$, where Z is a normal random variable) and secondly that we match the first and second moments of the sum to obtain the parameters for e^Z .

The details are as follows. Suppose that we wish to approximate the distribution of L , where

$$\begin{aligned} L &= L_1 + L_2 + \cdots + L_n \\ &= e^{Y_1} + e^{Y_2} + \cdots + e^{Y_n} \end{aligned}$$

where the $\{Y_i\}$'s are a vector of multivariate normal random variables with mean $(\mu_1, \mu_2, \dots, \mu_n)$ and variance-covariance matrix Σ . Thus, if the entries of Σ are denoted by $\sigma_{i,j}$, the variance of Y_i is $\sigma_{i,i}$ and $\sigma_{i,j} = \mathbb{E}[Y_i - \mu_i]\mathbb{E}[Y_j - \mu_j]$.

Thus each L_i is a lognormal variable and L is therefore a sum of correlation lognormal random variables.

We make the approximation that L is itself a lognormal random variable, that is $L \approx e^Z$ for some normal random variable Z . We then compute the first and second moments of L and fit the distribution of Z to these.

For the first moment we have

$$\begin{aligned} m := \mathbb{E}[L] &= \mathbb{E}[e^{Y_1} + e^{Y_2} + \cdots + e^{Y_n}] \\ &= \sum_{i=1}^n \exp\left\{\mu_i + \frac{\sigma_{i,i}}{2}\right\} \end{aligned}$$

whereas

$$\begin{aligned} s := \mathbb{E}[L^2] &= \mathbb{E}[(e^{Y_1} + e^{Y_2} + \cdots + e^{Y_n})^2] \\ &= \sum_{i=1}^n \mathbb{E}[(e^{Y_i})^2] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}[e^{Y_i+Y_j}] \\ &= \sum_{i=1}^n \exp\{2\mu_i + 2\sigma_{i,i}\} \\ &\quad + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \exp\left\{\mu_i + \mu_j + \frac{1}{2}(\sigma_{i,i} + 2\sigma_{i,j} + \sigma_{j,j})\right\}. \end{aligned}$$

Assuming that $L \approx e^Z$ is lognormal, we let the parameters of the normal

random variable Z be (μ, σ^2) . Then

$$\begin{aligned}\mathbb{E}[e^Z] &= \exp\left\{\mu + \frac{\sigma^2}{2}\right\}, \\ \mathbb{E}[(e^Z)^2] &= \exp\{2\mu + 2\sigma^2\}.\end{aligned}$$

Setting these equal to m and s , respectively, we get

$$\begin{aligned}\mu &= 2 \log m - \frac{1}{2} \log s, \\ \sigma^2 &= \log s - 2 \log m.\end{aligned}$$

Tests of approximation To test the validity of the approximation we considered the demand model with $\phi_1 = (400, 400, \dots, 400)$ and the variance-covariance matrix, Σ , is as described in the seasonal demand part of Section 5.3. We first sampled 10,000 realizations of (D_1, D_2, \dots, D_8) , computed $D_{[1,8]}$, and sorted these. This gives the empirical quartiles of the distribution of $D_{[1,8]}$. We then generated these same quartiles from the approximation scheme.

In Figure 5.1 we give a scatter plot (a p-p plot) of the quartiles of the approximating and empirical distribution. The approximating distribution corresponds to readings on the X-axis, the empirical distribution to those on the Y-axis. If the approximating distribution were perfectly accurate, this plot would be a straight line.

In fact, there is a slight trend above the 45-degree line for higher value points. This indicates that the approximating distribution underestimates the probability of very high demands. However this behavior occurs only for 50 or so points (out of 10,000) and so is not very significant.

Our standard implementation of the policies we tested uses the approximating distribution. However, as a further test of that approximation's validity, we also

implemented the policies using a Monte Carlo scheme to compute the distribution of cumulative demand. and then ran both of these implementations on 1,000 demand paths. There was an average difference of only 0.042% in expected cost between the Monte Carlo and Wilkinson Approximation methods.

5.2 Sample Path Computation of Demand and Base-Policy Target Inventory Levels

In Section 3.2.2, we assumed that for each of N samples of ω^i we had available the realized demands, denoted $\{D_s(\omega^i)\}_{s=t}^{T+L}$, and target inventory levels of the policy π , denoted $\{\bar{y}_s^\pi(\omega^i)\}_{s=t}^{T+L}$. In this section we discuss how to compute these values using the multiplicative MMFE as our model of demand and for three reference policies π : the Myopic, Minimizing and Balancing Polices. Altering the procedure for the additive MMFE, or another demand model is straightforward.

Under the multiplicative MMFE, the forecast vector, ϕ_t defined in Equation 5.1 contains all the information we need about the past in order to compute the current distribution of future demand. Thus, our problem is to take $\phi_t = (D_{t,t}, D_{t,t+1}, D_{t,t+2}, \dots, D_{t,T+L})$ and use it to generate N independent samples of $\{D_s\}_{s=t}^{T+L}$ and $\{\bar{y}_s^\pi\}_{s=t}^T$.

Demand Values As the computation of $\{\bar{y}_s^\pi\}_{s=t}^T$ will rely on having N samples of the evolution of the entire forecast vector, this is in fact what we compute. By denoting $D_{t+1,t} = \gamma_{t,t} D_{t,t}$ and recalling from Equation 5.5 that this equals the realized demand D_t we can do this with one set of notation. That is, given $\phi_t = (D_{t,t}, D_{t,t+1}, D_{t,t+2}, \dots, D_{t,T+L})$, for each sample ω_i we compute $D_{j,s}(\omega_i)$ for $s \leq T + L$ and $j \leq s + 1$ and then take $\{D_{s+1,s}(\omega_i)\}_{s=t}^{T+L}$ for our sample paths of

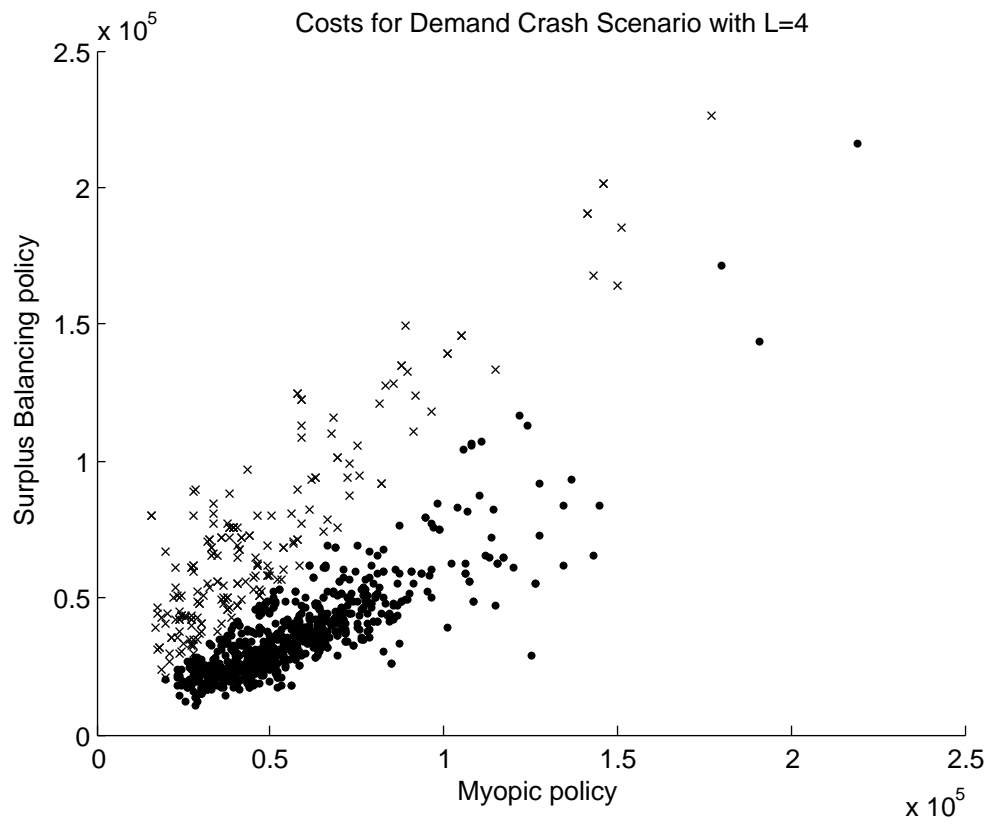


Figure 5.1: p-p plot of Empirical versus Approximating Distribution of Cumulative Demand.

demand.

We assume that the variance-covariance matrix of our updates Σ_t is effectively constant (Section 5.3). This allows us to increase efficiency by sampling the update vectors only once. That is, for each $\{\omega_i\}_{i=1}^N$ we generate once a set of vectors $\{\gamma_1(\omega_i), \gamma_2(\omega_i), \dots, \gamma_T(\omega_i)\}$ where $\gamma_j(\omega_i) = \exp(\epsilon_j(\omega_i))$ and $\epsilon_j(\omega_i)$ is a multivariate normal random variable with variance-covariance matrix Σ and mean $-\frac{\text{diag}\Sigma}{2}$. Then, given $\phi_t = (D_{t,t}, D_{t,t+1}, D_{t,t+2}, \dots, D_{t,T+L})$ we compute $D_{j,s}(\omega_i)$ via repeated applications of Equation 5.6 giving

$$D_{j,s}(\omega_i) = \prod_{k=1}^{s-t} \gamma_{k,j-t+1}(\omega_i) D_{t,s}.$$

Note that this re-use of the sampled $\{\gamma_s\}_{s=0}^{T+L}$ vectors applies only during the computation of $\bar{y}_t^{\delta(\pi)}$; when generating the true realized values of $\{\phi_s\}_{s=1}^{T+L}$ and hence $\{D_s\}_{s=1}^{T+L}$, we generate fresh update vectors.

We turn now to the computation of $\{\bar{y}_s^\pi(\omega_i)\}_{s=t}^T$ for our three choices of the reference policy π .

Reference Policy: Myopic Here we take $\pi = M$, the Myopic Policy. We assume that we have generated the sample forecast vectors $\{\phi_j(\omega_i)\}_{j=t}^T$. The base stock level \bar{y}_t^M is the unconstrained minimizer of $\mathbb{E}[c_t(y)|\mathcal{F}_t]$.

We have noted that this is a convex function and thus we solve

$$\begin{aligned} 0 &= \frac{\partial}{\partial y} \mathbb{E}[c_t(y)|\mathcal{F}_t] \\ &= \frac{\partial}{\partial y} \left\{ h_{t+L} \mathbb{E} \left[(y - D_{[t,t+L]})^+ | \mathcal{F}_t \right] + b_{t+L} \mathbb{E} \left[(y - D_{[t,t+L]})^- | \mathcal{F}_t \right] \right\} \\ &= \frac{\partial}{\partial y} \left\{ h_{t+L} \int_0^y (y-d) dF_{\phi_t}^{[t,t+L]}(d) + b_{t+L} \int_y^\infty (d-y) dF_{\phi_t}^{[t,t+L]}(d) \right\} \\ &= h_{t+L} F_{\phi_t}^{[t,t+L]}(y) - b_{t+L} + b_{t+L} F_{\phi_t}^{[t,t+L]}(y) \end{aligned}$$

where $F_{\phi_t}^{[t,t+L]}(\cdot)$ is the distribution of the cumulative demand between t and $t+L$, given the forecast vector ϕ_t . We have noted above that the forecast vector ϕ_t is all the information from the information set \mathcal{F}_t that we require. For $L > 0$ this distribution of cumulative demand is computed using the approximation described in Section 5.1. Setting the expression in Equation 5.7 equal to zero we find that

$$\bar{y}_t^M = \left(F_{\phi_t}^{[t,t+L]} \right)^{-1} \left(\frac{b_{t+L}}{b_{t+L} + h_{t+L}} \right). \quad (5.7)$$

Thus, when $\pi = M$, the Myopic Policy, the computation of sample values of the target inventory level is very simple; it requires only a single evaluation of the inverse cumulative distribution function of a log-normal random variable.

Zero-finding Algorithm When we take π to be the Minimizing or Balancing Policy, there is no closed-form expression such as that in Equation 5.7 and, instead, we must perform the minimization numerically. In both cases, the objective functions to be minimized are convex and so the problem reduces to finding the zero of the derivative.

For this purpose, we have adapted two well known zero-finding routines: the Newton-Raphson and Bisection Search methods. These two methods are outlined in Press *et al.* [51]; we describe here how we have combined them in general terms.

Suppose that we wish to find x_0 such that $f(x_0) = 0$. We assume that we are given an interval $[x_l^0, x_u^0]$ such that $f(\cdot)$ is increasing on $[x_l^0, x_u^0]$ and $f(x_l^0) \leq 0$ and $f(x_u^0) \geq 0$. Also, $f'(\cdot)$ exists at every point in $[x_l^0, x_u^0]$. We are given x_0^0 , an initial guess for x_0 such that $x_0^0 \in [x_l^0, x_u^0]$.

At the $(n+1)$ -st step, the algorithm computes

$$x_{NR}^{n+1} = x_0^n - \frac{f(x_0^n)}{f'(x_0^n)},$$

the Newton-Raphson step. The value x_{NR}^{n+1} is simply the zero of the first-order Taylor expansion of $f(\cdot)$ about x_0^n .

Then, if $x_{NR}^{n+1} \in [x_l^n, x_u^n]$, we let $x_0^{n+1} = x_{NR}^{n+1}$, otherwise we let $x_0^{n+1} = \frac{1}{2}(x_u^n + x_l^n)$.

Finally, we evaluate $f(x_0^{n+1})$. If its value is negative, then we set $x_l^{n+1} = x_0^{n+1}$ and $x_u^{n+1} = x_u^n$, otherwise set $x_l^{n+1} = x_l^n$ and $x_u^{n+1} = x_0^{n+1}$.

We continue applying this until we reach a termination condition. This could be either that the candidate x_0 values are sufficiently close (i.e. $x_u^n - x_l^n < \epsilon$ for some ϵ) or that the function value is sufficiently close to 0 (i.e. $|f(x_0^n)| < \epsilon$ for some ϵ).

The advantage of the Newton-Raphson method lies in the fact that its convergence rate is *quadratic*; if $e^n = x_0^n - x_0$ is the error term at the n -th iteration, then, for sufficiently large n , we have

$$e^{n+1} = A(e^n)^2$$

for some constant A . For the times when we instead take a bisection step, that is set $x_0^{n+1} = \frac{1}{2}(x_u^n + x_l^n)$, the convergence is *linear*.

The faster convergence rate of the Newton-Raphson method is the reason that it is the default method in generating the next value of x_n . However, the pure Newton-Raphson method can be unstable and this is why we use a hybrid of the Newton-Raphson and Bisection Search methods.

To see why this hybrid is stable, note that our upper and lower bounds on x_0 clearly preserve the two properties that $x_0 \in [x_l^n, x_u^n]$ and that $f(x_l^n) \leq 0$ and $f(x_u^n) \geq 0$. Thus, in cases where the value from the Newton-Raphson step, x_{NR}^{n+1} , falls outside these bounds, we simply take the midpoint of the bounds as our next step instead.

Reference Policy: Minimizing We turn now to the application of this method to finding the target inventory level \bar{y}_t^m for the Minimizing Policy defined in Section 1.2.1

Recall from Equation 1.9 that the target inventory level of the Minimizing Policy is given by

$$\bar{y}_t^m = \arg \min_y \{h_t(x, y) + b_t(y)\}.$$

We have noted that the function being minimized is convex and thus the base-stock level is the value of y that solves

$$\begin{aligned} 0 &= \frac{\partial}{\partial y} \{h_t(x, y) + b_t(y)\} & (5.8) \\ &= \frac{\partial}{\partial y} \left(\sum_{s=t}^T h_{s+L} \mathbb{E} \left[(y - D_{[t,s+L]})^+ \mid \mathcal{F}_t \right] \right) + b_{t+L} \mathbb{E} \left[(y - D_{[t,t+L]})^- \mid \mathcal{F}_t \right] \\ &= \frac{\partial}{\partial y} \left\{ \left(\sum_{s=t}^T h_{s+L} \int_0^y (y-d) dF_{\phi_t}^{[t,s+L]}(d) \right) + b_{t+L} \int_y^\infty (d-y) dF_{\phi_t}^{[t,t+L]}(d) \right\} \\ &= \left(\sum_{s=t}^T h_{s+L} F_{\phi_t}^{[t,s+L]}(y) \right) - b_{t+L} + b_{t+L} F_{\phi_t}^{[t,t+L]}(y), \end{aligned}$$

where, as before, $F_{\phi_t}^{[t,s]}$ is the cumulative distribution function of the sum of demand between periods t and s , given the forecast vector ϕ_t at time t .

Taking derivatives again, we see that

$$\frac{\partial^2}{\partial y^2} \{h_t(x, y) + b_t(y)\} = \left(\sum_{s=t}^T h_{s+L} f_{\phi_t}^{[t,s+L]}(y) \right) + b_{t+L} f_{\phi_t}^{[t,t+L]}(y),$$

where $f_{\phi_t}^{[t,t+L]}(\cdot)$ is the probability distribution function corresponding to the cumulative distribution function $F_{\phi_t}^{[t,t+L]}$.

We find the value of y that solves Equation 5.8 by applying the zero-finding algorithm described above, with $\frac{\partial}{\partial y} \{h_t(x, y) + b_t(y)\}$ taking the place of $f(x)$ and $\frac{\partial^2}{\partial y^2} \{h_t(x, y) + b_t(y)\}$, its derivative $f'(x)$.

For the upper and lower bounds on the zero of $\frac{\partial}{\partial y}\{h_t(x, y) + b_t(y)\}$, recall Equation 1.10 which states

$$\bar{y}_t^m \leq \bar{y}_t^* \leq \bar{y}_t^M,$$

where \bar{y}_t^* is the optimal target inventory level. Thus, we can take \bar{y}_t^M as the upper bound and 0 as a trivial lower bound. The starting point can be any value in $[0, \bar{y}_t^M]$; in our computational experiments we took it to be $\alpha\bar{y}_t^M$ with $\alpha = 0.8$.

Reference Policy: Balancing The final reference policy, π , to which we apply our Delta- π Myopic methodology is the Balancing Policy, denoted B . Our definition of reversion in Equation 3.7 uses base-stock target inventory levels. However, we showed that the Balancing Policy is not a base-stock policy and hence it does not have a target inventory level. Instead we use the *actual* post-order inventory level of the Balancing Policy, defining the reversion time as

$$\rho^\pi(t, x, y) = \min \{s : s > t, y - D_{[t,s]} < \bar{y}_s^B(x_s^B) \text{ or } s = T + 1\}.$$

Recall Remark 3.2.5 where we noted that when our reference policy is a base-stock policy, changing y has no effect on cost beyond the reversion time. This will not be true here. To see this, note that, in the reversion period — which we will denote ρ for brevity — the Balancing Policy will have a pre-order inventory level, x_ρ^B , whereas the jolted Balancing Policy will have a potentially different pre-order inventory level $x_\rho^B(t, x_t, y)$.

As the post-order inventory level the Balancing Policy depends on the pre-order inventory level in a more complicated way than simply a threshold (as with base-stock policies), the post-order inventory level for the jolted Balancing Policy will not necessarily be equal to $y_s^B(x_s^B)$. Indeed the ordering level in all future periods may be different and thus there is a continued dependence on y . This dependence

will be small, however, and thus for experimental purposes, we proceed with the definition of reversion given.

It follows that we must compute these values in stages, as we need to know not only the sampled value of ϕ_s but also that of x_s^B to compute $y_s^B(x_s^B)$. To achieve this, we compute $y_t^B(x_t)$ using our actual pre-order inventory level x_t and then, for each sample path, we compute $x_{t+1}^B = y_t^B(x_t) - D_t$ and use this value in computing $y_{t+1}^B(x_{t+1})$ and so on.

Recall from Section 1.2.1 that the value $y_t^B(x)$ is the value of y that solves the equation

$$h_t(x, y) - b_t(y) = 0.$$

We apply our zero-finding algorithm with $h_t(x, y) - b_t(y)$ taking the place of $f(\cdot)$.

We know that this function is increasing and that its derivative is

$$\frac{\partial}{\partial y} h_t(x, y) - b_t(y) = \left(\sum_{s=t}^T h_{s+L} F_{\phi_t}^{[t, s+L]}(y) \right) + b_{t+L} - b_{t+L} F_{\phi_t}^{[t, t+L]}(y).$$

As the Balancing Policy always places an order, we know that we can use the pre-order inventory position x as the lower bound on the zero of $h_t(x, y) - b_t(y)$. There is no readily available upper bound and so we use a naïve method and test the myopic target inventory level, y_t^M multiplied by powers of two. That is, we search for the smallest $i \geq 0$ such that $h_t(x, 2^i y_t^M) - b_t(2^i y_t^M) \geq 0$

As part of computing $h_t(x, y) - b_t(y)$, we have to evaluate expressions of the form $\mathbb{E}[(y - Z)^+]$, where Z is a log-normal random variable, denoted $Z \sim LN(\mu, \sigma^2)$. That is, we have $Z = e^X$, where X is a normal random variable, $X \sim N(\mu, \sigma^2)$. Thus, we seek to compute

$$\mathbb{E}[(y - Z)^+] = \int_0^\infty (y - z)^+ dF_Z(z) = \int_0^y (y - z) dF_Z(z)$$

where $F_Z(\cdot)$ is the cumulative distribution function of Z . Substituting $x = \log(z)$ and denoting the cumulative and probability distribution functions of a $N(\mu, \sigma^2)$ random variable by $\phi_{\mu, \sigma^2}(\cdot)$ and $\Phi_{\mu, \sigma^2}(\cdot)$, respectively, we have

$$\begin{aligned} \int_0^y (y - z) dF_Z(z) &= \int_{-\infty}^{\log y} (y - e^x) \phi_{\mu, \sigma^2}(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\log y} (y - e^x) \exp\left\{\frac{-(x - \mu)^2}{2\sigma^2}\right\} \\ &= y\Phi_{\mu, \sigma^2}(\log(y)) - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\log y} \exp\left\{x + \frac{-(x - \mu)^2}{2\sigma^2}\right\}. \end{aligned}$$

Completing the square for the exponent in the second term, we see that

$$\begin{aligned} x + \frac{-(x - \mu)^2}{2\sigma^2} &= \frac{-1}{2\sigma^2} (x^2 - 2(\mu + \sigma^2)x + \mu^2) \\ &= \frac{-1}{2\sigma^2} \left((x - (\mu + \sigma^2))^2 - 2\mu\sigma^2 - \sigma^4 \right) \\ &= \frac{-1}{2\sigma^2} \left((x - (\mu + \sigma^2))^2 \right) + \mu + \frac{\sigma^2}{2}. \end{aligned}$$

Therefore

$$\mathbb{E}[(y - Z)^+] = y\Phi_{\mu, \sigma^2}(\log(y)) - e^{\mu + \sigma^2} \Phi_{\mu + \sigma^2, \sigma^2}(\log(y)).$$

5.3 Experimental Design

We now describe the specific computational experiments we used to test our new policies. The space of potential parameter settings for this study is very large. In addition to parameters describing the inventory system, there are many parameters that describe the manner in which forecasts of demand evolve over time. A fully comprehensive study is beyond the scope of this work. Our goal is to study a broad range of potential application settings, with emphasis on the demand and forecasting processes. The experimental design is oriented around a Base Case and six sets of *scenarios*, each of which expands the Base Case in an interesting

dimension. In each set of scenarios we vary specific input parameters. The first three of these scenario sets study *first-order effects*; here, it is the initial forecast ϕ_0 that varies. The final three scenario sets study *second order effects* by varying the variance-covariance matrix Σ in different ways.

We begin this section by discussing the parameters of the Base Case. After that we describe the manner in which the parameters of the Base Case are varied in each of the six scenario sets.

The Base Case In all of the experiments, we let our holding and backorder costs per unit per period be stationary and assume values $h_t = 1$ and $b_t = 10$ for all t . As noted in Section 1.1.2, we take $p_t = 0$ for all t without loss of generality.

We consider a horizon of length $T = 40$. All experiments are conducted for two different values of the lead-time: $L = 0$ and $L = 4$. Note that when $L = 4$, in the first four time periods the costs incurred are determined by decisions made in the past, and are not influenced by our choice of policy. Therefore, to facilitate comparison between results, in the case $L = 0$ costs are not counted during the first four time periods.

The initial demand forecast is flat, i.e., $\phi_0 = (400, 400, \dots, 400)$. The horizon over which the user generates forecasts is of length 12. This implies that we learn nothing about the period- t demand until we are within 12 periods of period t , i.e., $D_{t-12,t} = d_{0,t} = 400$ for $t > 12$. Algebraically, recall from Section 5.1 that the standard multiplicative model updates forecasts using the formula $D_{st} = \gamma_{st}D_{s-1,t}$. The assumption is that for $t > s + 11$ we have $\gamma_{st} = 1$. This implies that at all times s the first 12 elements of the forecast vector ϕ_s will be different from each other, but the 13-th element and every subsequent element will be equal to 400.

Recall from Section 5.1 that in the multiplicative MMFE, the period- t update vector is $\gamma_t = e^{\epsilon_t}$, where ϵ_t is a $T-t+1$ - dimensional random vector with variance-covariance matrix Σ_t and mean $-\frac{1}{2}\text{diag}(\Sigma_t)$. We obtain the $(T-t+1) \times (T-t+1)$ matrix Σ_t from Σ_{t-1} , by dropping the last row and column. The previous paragraph implies that in our experiments, forecast evolution and demand are driven by a 12×12 covariance matrix Σ . We obtain the $T \times T$ matrix Σ_1 from Σ by appending $T-12$ extra rows and columns to Σ , with 1's on the diagonal and 0's elsewhere. Therefore, for $t > 12$, ϵ_t is a degenerate random variable with mean 0 and variance 0.

In the Base Case we have *constant learning*, meaning that all of the entries on the diagonal of Σ are equal. The diagonal elements are selected so that for $t \geq 12$, the coefficient of variation of the demand D_t , seen from the beginning of time period 1, is 0.75. A formula for the coefficient of variation is provided in the description of the Coefficient of Variation Scenarios, below.

The off-diagonal entries of the covariance matrix Σ determine the degree of correlation between the updates that are observed in a given time period, say, time period s . The Base Case assumes that there is some correlation between these updates, modeled by having non-zero, positive values in the first off-diagonal of Σ . Consequently, in the Base Case, if the forecast for the demand in period t goes up in period s (i.e., if $D_{st} > D_{s-1,t}$), then the forecast for demand in period $t+1$ is likely to increase in period s as well (if $t+1 \leq s+11$), but this does not tell us anything about the forecast for demand in period $t+2$. The values of the non-zero off-diagonal elements are chosen to give a correlation coefficient of 0.5 for each pair of adjacent forecast updates. That is, for each s and each t , $s \leq t \leq s+10$, the update factors γ_{st} and $\gamma_{s,t+1}$ observed in period s have correlation coefficient

0.5, but γ_{st} and $\gamma_{s,t+2}$ are stochastically independent.

Product Launch Scenarios In this set of scenarios we study the effect of rising demand, as might be encountered at a product launch. Only the initial forecast vector ϕ_0 is varied. For comparison with the base case, we ensure that the mean of the values in ϕ_0 is 400. We consider upward demand trends of +5, +10 and +20 per period. In addition, we consider two examples in which the demand rises in a steeper, non-linear manner, mid-way through the horizon; these are generated using an appropriately scaled normal CDF curve. The five initial forecast vectors are plotted in Figure 5.2.

End-of-Life Scenarios Here we study scenarios associated with products that are in an end-of-life situation, namely those with decreasing initial forecast vectors. Essentially, these are the reverse of the Product Launch scenarios; we have initial forecast vectors with forecasted demand decreasing by 5, 10 and 20 per period. We also consider two examples whose demands have steeper drop-off curves, generated using the normal complementary CDF curve. In addition, we study a total demand crash, in which the demand is forecast to crash to 10 midway through the time horizon.

Seasonality Scenarios In the seasonality study, we use the common base-values described above for all parameters except for the initial forecast vector ϕ_0 . In the base case the demand is flat, with $\phi_0 = (400, 400, \dots, 400)$. In addition to the base case we conduct experiments with two forms of seasonality, one defined via a sinusoidal function and the other via a step function. In both cases, the maximum value attained is 700 and the minimum is 100. This allows us to compare results

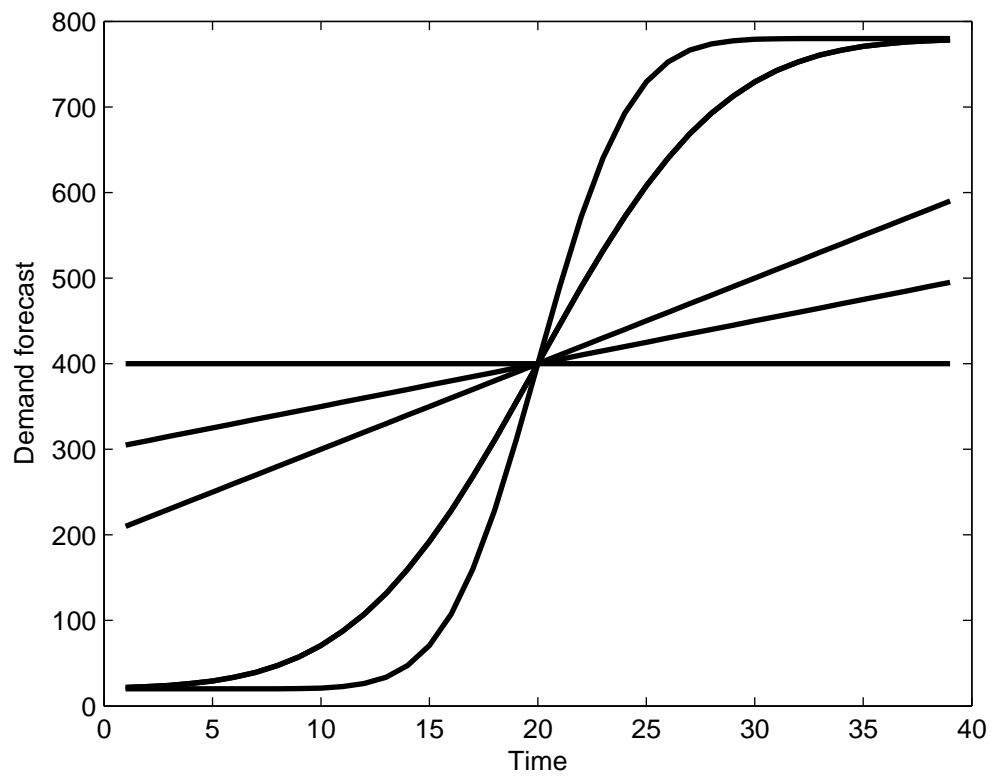


Figure 5.2: Initial forecast vectors used in Product Launch Scenarios.

more easily with the base case, because the mean of the entries in the initial forecast vector is 400 in all cases.

By the *cycle length* we mean the number of time periods between two consecutive high-points. We consider cycle lengths with values 2, 4 and 8. For example, for the step-function with period 4, we have

$$\phi_0 = (700, 700, 100, 100, 700, 700, 100, 100, \dots)$$

The above scenario sets test the effect of varying ϕ_0 , the initial forecast vector. In the final three scenario sets we focus instead on varying Σ . In all of these we take $\phi_0 = (400, 400, \dots, 400)$.

Coefficient of Variation Scenarios In this scenario set, we study the effect of varying the magnitude of the variance in the demands and the forecasts. Note that for $t \geq 12$, at the end of time period $t - 12$, we have $D_{t,t} = \Gamma_t d_{t-12,t}$, where Γ_t is random and has the same distribution as

$$\Gamma = \prod_{i=1}^{12} \gamma_i = \exp \left(\sum_{i=1}^{12} \epsilon_i \right).$$

The ϵ_i 's are independent normal random variables, with mean such that $E[e^{\epsilon_i}] = 1$, and with variance σ_{ii} , which is the i -th diagonal element of Σ , our forecast update matrix (note that σ_{ii} is a variance, not a standard deviation). Thus, the mean of Γ is one and the variance is $\exp(\sum_{i=1}^{12} \sigma_{ii}) - 1$. The coefficient of variation of Γ is given by $(\exp(\sum_{i=1}^{12} \sigma_{ii}) - 1)^{1/2}$, and is set equal to 0.75 in the Base Case. In the scenarios where we investigate the effect of variance, we scale the entries of Σ such that the coefficient of variation of this series of twelve updates takes specific values, namely 0.5, 0.7, 1, 2, 4, and 8. This corresponds to different levels of variability in the demands.

Time of Learning Scenarios As mentioned in Section 5.1.1, the ratio of the sum of the first j diagonal entries of Σ , to the sum of all the diagonal entries, is the fraction of variability in D_t that is *unresolved* in period $s = t - j$. In Figure 5.3, we plot four different possibilities for the way in which variability is resolved. When all the entries in the diagonal are identical (a straight line plot) then the variance of each update is the same. This corresponds to what we call *constant learning*. When the values in $\text{diag}(\Sigma)$ are weighted towards the end of the vector (a convex plot), then the unresolved uncertainty is low when j is small (s is close to t). This corresponds to *early learning*. Conversely, when the values in $\text{diag}(\Sigma)$ are weighted towards the beginning of the vector (a concave plot), then this corresponds to *late learning*: most of the uncertainty about the true value of D_t is only resolved in periods s that are close to t . We also consider the setting in which there is more weight in the center of $\text{diag}(\Sigma)$ than at the endpoints: we learn most in the middle of the forecast horizon.

We construct variance-covariance matrices Σ to correspond with these four cases: constant, early, late and mid-horizon learning. In all cases, the values of Σ are scaled to ensure that the coefficient of variation of Γ , and of D_t for $t \geq 12$, remains constant at 0.75.

Correlation Scenarios In this scenario set we test the effect of different types of correlation between the updates. We vary correlation in two ways. First, we set the number of non-zero off-diagonals of our 12x12 matrix, Σ , to 0 (which corresponds to no correlation), 1, 4 and 8. Secondly, the sign of the off-diagonal elements can be all positive, all negative, or entries alternating between positive and negative. (The base case corresponds to 1 off-diagonal with non-zero elements which are

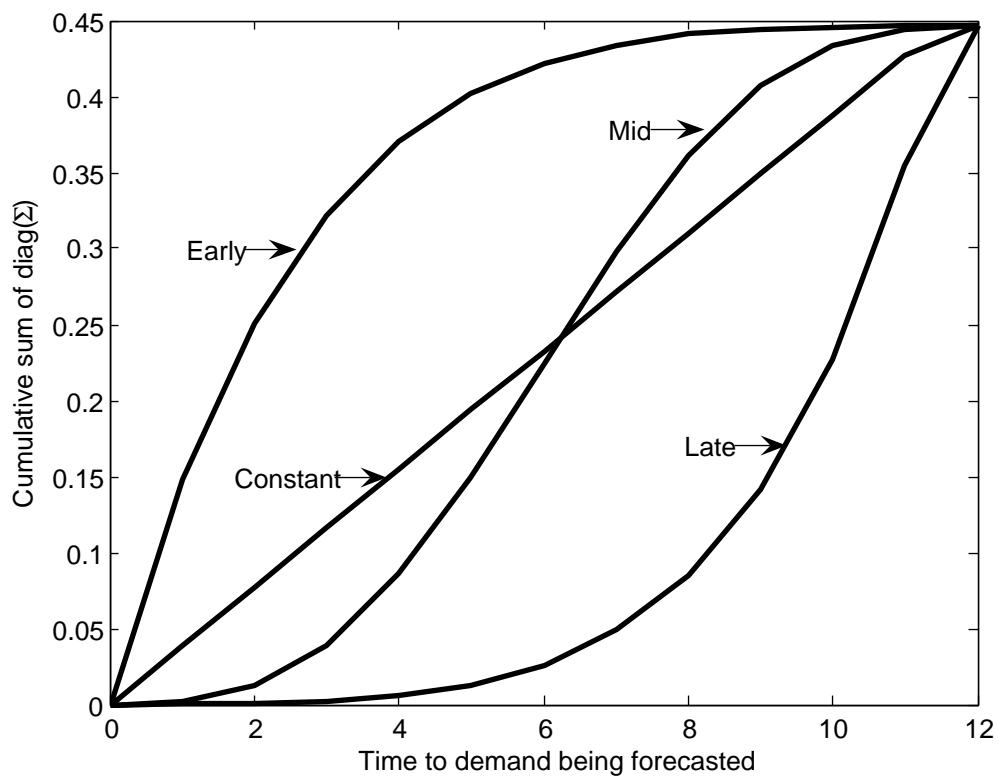


Figure 5.3: Cumulative sum of the diagonal elements of Σ for constant, early, late and mid-horizon learning.

all positive.) As in the base case, the diagonal of Σ corresponds to the constant learning case, and the coefficient of variation of Γ is 0.75.

Table 5.1: Scenario codes

Topic	Code	Description
Product Launch (5)	$+I$	Increment I per period, $I \in \{5, 10, 20\}$
	Curve	Increasing scaled normal CDF curve
	S. Curve	Steeper scaled normal CDF curve
End-of-Life (6)	$-I$	Decrement I per period, $I \in \{5, 10, 20\}$
	Curve	Decreasing scaled normal CDF curve
	S. Curve	Steeper scaled normal CDF curve
Seasonal (7)	Crash	Demand crash
	Base Case	Initial forecast vector is flat
	$\text{Sin}(n)$	Sinusoidal periodicity with cycle length n , $n \in \{2, 4, 8\}$
Coeff. of Var. (6)	$\text{Step}(n)$	Step-function periodicity with cycle length n
	$\text{CV} = \beta$	Coefficient of variation of β $\beta \in \{0.5, 0.7, 1, 2, 4, 8\}$
Learning Rate (4)	Const	Constant learning
	Late	Late learning
	Early	Early learning
	Mid	Mid-horizon learning
Correlation (10)	None	All off-diagonal elements of Σ are 0
	$\text{Pos}(n)$	First n off-diagonals of Σ have positive entries, $n \in \{1, 4, 8\}$
	$\text{Neg}(n)$	First n off-diagonals of Σ have negative entries
	$\text{Mix}(n)$	First n off-diagonals of Σ have entries alternating positive and negative

Table 5.1 summarizes the scenarios we study. The number of scenarios for each group is given in brackets after the group name; we see that there are 38 in total. We run each of these with $L = 0$ and $L = 4$ for an overall total of 76 scenario - lead time pairs.

For each of the scenarios, we ran $N = 1,000$ independent trials for a horizon of length $T = 40$. For the scenarios with a lead time of $L = 4$, our decisions only

influenced costs from periods 5 through 40. Therefore, in order to compare costs on an even footing, we consistently computed the total holding and backorder costs excluding the first 4 periods for both $L = 0$ and $L = 4$.

5.3.1 Performance Measures Used

Finally, we describe how we measure the performance of our new policies. We take the Myopic Policy as our benchmark and measure how the new policies perform relative to it.

Specifically, for a fixed policy π , we let $\mathcal{C}_i(\pi)$ denote the cost of the i -th run ($i = 1, \dots, 1000$), excluding the first 4 periods. Note that since we consider a complex environment and a relatively long horizon ($T = 40$), it is not tractable to find the optimal policy or even evaluate the optimal expected cost. Instead, we use two performance measures of a policy's effectiveness. Both measures are computed relative to the performance of our benchmark, the Myopic Policy, M . The first is the *relative total cost*, given by

$$AT(\pi) = \left(1 - \left(\frac{\sum_{i=1}^N \mathcal{C}_i(\pi)}{\sum_{i=1}^N \mathcal{C}_i(M)} \right) \right) * 100\%$$

whereas the second is the *average relative cost per run*, which is given by

$$AR(\pi) = \left(1 - \left(\frac{1}{N} \sum_{i=1}^N \frac{\mathcal{C}_i(\pi)}{\mathcal{C}_i(M)} \right) \right) * 100\%.$$

Note that both $AT(\pi)$ and $AR(\pi)$ can be positive or negative. If they are positive this implies that they improve upon the Myopic Policy, and a the higher value indicates higher improvement. Conversely, if they are negative, this implies that myopic is doing better. (Thus, in the tables given in Chapter 6 to come, positive numbers indicate good relative performance with respect to the Myopic Policy.)

For each run, we also compute a lower bound on the costs to provide an additional reference. Recall that the order-up-to level of the Minimizing Policy is always below that of the optimal, whereas that of the Myopic is always above the optimal. Thus, the sum of the holding cost from the Minimizing Policy and the backorder cost from the Myopic Policy gives a lower bound on the cost of the optimal policy. If we imagine that this is the cost of a policy π , we then compute $AT(\pi)$ and $AR(\pi)$ in the manner shown above, but with respect to this lower bound 'policy'. This statistic is denoted LB and is an *upper* bound on the potential relative improvements over the Myopic Policy that can be further achieved.

Chapter 6

Experimental results

Introduction In this Chapter we present the results of the computational investigation of the average performance of the policies described in Chapters 3 and 4. We find especially good performance exhibited by the three versions of the Delta- π Myopic policy that we introduced in Section 5.2, that is the application of the Delta- π Myopic scheme to the Myopic, Minimizing and Balancing policies. These are denoted $\Delta(M)$, $\Delta(m)$ and $\Delta(B)$ respectively. In addition, we demonstrate that, of the policies extending the Minimizing and Balancing policies introduced in Chapter 4, two in particular, the Surplus-Balancing policy (denoted *SB*), and the Minimizing policy, *m(k-tot)*, exhibit superior performance. Recall that the *m(k-tot)* policy is the member of the Minimizing(*k*) family of policies which chooses *k* equal to the average run-out time of all the units present in the system. Collectively these policies achieve an average cost that is up to 30% lower than that of the Myopic policy (our benchmark); they out-perform the Myopic policy in almost every scenario; and they are never much worse than the best performing policy in any scenario (see Table 6.9). In all of the tables of this section, the results for these policies are highlighted using boldface type.

The greatest improvements over the Myopic policy occur in contexts where steep demand drops can occur. These contexts include end-of-life scenarios (Tables 6.2 and 6.3), seasonality (Table 6.4), and systems with highly variable demands and forecasts (Table 6.5). Long lead times make the improvements more dramatic. Moreover, the average performance of the variants of the Balancing policy seems to be significantly better than the worst-case guarantee of 2.

We note that the standard Balancing and Minimizing policies are greatly improved by the various refinements introduced in this work. The most universally applicable of these refinements is the Bounding concept presented in Chapter 4. At the end of this section we discuss bounding and its effect; otherwise, all of results presented in this section include the improvements due to bounding.

Near the end of Section 5.3 we defined two performance measures, the relative total cost $AT(\pi)$ (which places more weight on randomly generated problem instances in which the total costs are higher), and the average relative cost per run $AR(\pi)$. We prefer $AR(\pi)$ because it weights all problem instances equally. As the accompanying technical report [29] indicates, the two measures usually tell similar stories. However $AR(\pi)$ is usually 0-2% higher than $AT(\pi)$. This is because our back-order cost is ten times the level of our holding cost. Thus the randomly generated scenarios in which the total costs are highest, are usually ones in which the demand grows unexpectedly. In these scenarios the Myopic policy performs somewhat better relative to other policies. The strongest exception to the 0-2% rule is the demand crash scenario shown in Table 6.3 below, for which we report both measures (see, also, the robustness study below (Table 6.9)). In the demand-crash scenario, the problem instances with the greatest costs are the ones in which the Myopic policy dramatically over-stocks. These problem instances have a disproportionate impact on $AT(\pi)$, and favor policies that are not myopic.

In the technical report [29] we report the performance of all the policies in Table 4.1 under each scenario in Table 5.1, measured by both $AR(\pi)$ and $AT(\pi)$. In this work, we report a subset of those results, focusing on the most interesting 14 policies (eleven of those listed in Table 4.1 and the three variants on Delta- π Myopic).

This Chapter is organized as follows. First we present computational results for the scenario sets defined in Section 5.3 in the following sequence: Product Launch, End of Life, Demand Crash, Seasonality, Coefficient of Variation, Learning, Correlation. We then study the robustness of the different heuristics over the 76 scenarios tested. Finally, we examine the effect of Bounding.

6.1 First Order Effects

Product Launch In the Product Launch scenarios, the results for which are given in Table 6.1, the demand is trending upwards strongly. There is little risk of overstocking when this is the case, and hence, the performance of the Myopic policy should be at its peak. The Myopic policy is close to the lower bound when the lead time is short which we see from the fact that in Table 6.1 there is at most a 4.01% improvement possible compared to the LB when $L = 0$. Each of the new policies improves relative to the Myopic policy even in these scenarios, but the improvement is slight: less than 0.6% for $L = 0$.

As the lead time increases, so does the gap between the Myopic policy and the lower bound. Over half of the new policies, especially the recommended policies ($\Delta(M)$, $\Delta(m)$, $\Delta(B)$, SB , and $m(k\text{-tot})$), show noticeable improvement (as high as 2.06%) over Myopic in these scenarios. A majority of the policies are markedly worse compared to the Myopic policy with longer lead times. This is a general pattern that is apparent in all scenarios. The most likely explanation for this pattern is that increased lead times magnify errors. For example, in the Base Case, all policies except $m(2)$, $B(2)$, $B(\beta\text{-myo})$ and the recommended policies, under-order on average when $L = 0$, and do so more strongly when $L = 4$. These are the policies whose performance deteriorates as L increases from 0 to 4. There

are examples in other scenarios where under-ordering becomes more prevalent as the lead time increases.

End of Life Scenarios The End of Life scenarios are like the Product Launch scenarios, except that the trend is for decreasing demand. In our experiments, the risk of overstocking when using the Myopic policy is low as long as the lead time is short. Table 6.2 demonstrates that the Myopic policy is close to the lower bound (within 5.19%) for all scenarios with $L = 0$.

The results for long lead times ($L = 4$) reveal a weakness in the Myopic policy. When the lead time is long and the demand decline is steep, then the new policies perform as much as 10.94% better than myopic. This can be seen in Table 6.2.

The Demand Crash Scenario As could be expected, we see in Table 6.3 that the new policies perform significantly better than Myopic under the demand crash scenario. In this scenario, improvements over the Myopic policy range between 10% and 20%. This improvement is due to the fact that the policies are much better at avoiding overstocking in the periods after the crash, namely periods 21 through to 40. We measure this by computing the total holding cost incurred by each policy in these periods over all 1000 runs, expressed as a percentage of the same cost incurred by the myopic policy. We denote this measure of performance by HC . In Table 6.3 we report the values of $AT(\pi)$ and $AR(\pi)$ and HC for all policies. The fact that the new policies outperform the Myopic policy by up to 80% in periods after the crash is what makes them better overall.

Seasonality Table 6.4 summarizes the results from the Seasonality scenarios. As the LB row of the table indicates, there is opportunity for improvement over

Table 6.1: Product Launch: $AR(\pi)$, for certain Product Launch scenarios
 ($AR(\pi)$ is the average percent improvement over Myopic, per run.)

Scenario	$L = 0$			
	Base	+20	Curve	S. Curve
$\Delta(\mathbf{M})$	0.39%	0.40%	0.28%	0.48%
$\Delta(\mathbf{m})$	0.55%	0.53%	0.41%	0.55%
$\Delta(\mathbf{B})$	0.55%	0.53%	0.41%	0.55%
B	0.37%	0.34%	0.21%	0.47%
SB	0.13%	0.11%	0.03%	0.10%
$B(0.5)$	0.36%	0.34%	0.21%	0.47%
$B(2)$	0.46%	0.43%	0.31%	0.53%
$B(\alpha\text{-myo})$	0.39%	0.37%	0.26%	0.41%
m	0.36%	0.34%	0.21%	0.47%
$m(2)$	0.46%	0.42%	0.33%	0.51%
$m(3)$	0.44%	0.41%	0.26%	0.51%
$m(k\text{-fin})$	0.44%	0.40%	0.29%	0.51%
$m(k\text{-mar})$	0.29%	0.26%	0.26%	0.29%
$\mathbf{m(k\text{-tot})}$	0.29%	0.26%	0.22%	0.29%
LB	4.01%	3.71%	3.67%	3.79%
	$L = 4$			
$\Delta(\mathbf{M})$	1.97%	1.45%	1.73%	2.06%
$\Delta(\mathbf{m})$	1.43%	1.43%	1.61%	1.88%
$\Delta(\mathbf{B})$	1.51%	1.42%	1.62%	1.90%
B	-2.58%	-3.05%	-2.67%	-2.26%
SB	1.91%	1.24%	1.52%	1.79%
$B(0.5)$	-4.15%	-4.15%	-3.75%	-3.44%
$B(2)$	1.52%	0.93%	1.25%	1.61%
$B(\alpha\text{-myo})$	1.47%	0.98%	1.13%	1.32%
m	-4.32%	-4.24%	-3.86%	-3.54%
$m(2)$	0.82%	0.04%	0.36%	0.75%
$m(3)$	-1.71%	-2.39%	-2.02%	-1.57%
$m(k\text{-fin})$	-1.89%	-2.36%	-2.00%	-1.60%
$m(k\text{-mar})$	-0.51%	-1.03%	-0.70%	-0.33%
$\mathbf{m(k\text{-tot})}$	1.90%	1.32%	1.57%	1.82%
LB	25.92%	22.91%	23.18%	23.75%

Table 6.2: End Of Life: $AR(\pi)$, for certain End Of Life scenarios

	$L = 0$			
Scenario	Base	+20	Curve	S. Curve
$\Delta(\mathbf{M})$	0.39%	0.56%	0.74%	0.76%
$\Delta(\mathbf{m})$	0.55%	0.72%	0.73%	0.98%
$\Delta(\mathbf{B})$	0.55%	0.72%	0.73%	0.98%
B	0.37%	0.53%	0.83%	0.66%
SB	0.13%	0.16%	0.19%	0.35%
$B(0.5)$	0.36%	0.52%	0.82%	0.58%
$B(2)$	0.46%	0.59%	0.75%	0.87%
$B(\alpha\text{-myo})$	0.39%	0.48%	0.53%	0.68%
m	0.36%	0.52%	0.82%	0.58%
$m(2)$	0.46%	0.62%	0.76%	0.90%
$m(3)$	0.44%	0.62%	0.81%	0.91%
$m(k\text{-fin})$	0.44%	0.60%	0.83%	0.83%
$m(k\text{-mar})$	0.29%	0.37%	0.48%	0.60%
m(k-tot)	0.29%	0.36%	0.40%	0.56%
LB	4.01%	4.63%	4.55%	5.19%
	$L = 4$			
$\Delta(\mathbf{M})$	1.97%	3.22%	5.27%	10.62%
$\Delta(\mathbf{m})$	1.43%	3.11%	3.88%	10.86%
$\Delta(\mathbf{B})$	1.51%	3.18%	3.99%	10.94%
B	-2.58%	-1.65%	0.91%	6.12%
SB	1.91%	3.14%	5.02%	9.93%
$B(0.5)$	-4.15%	-4.24%	-2.59%	1.26%
$B(2)$	1.52%	3.00%	5.25%	10.31%
$B(\alpha\text{-myo})$	1.47%	2.50%	4.47%	8.59%
m	-4.32%	-4.58%	-3.07%	0.36%
$m(2)$	0.82%	2.63%	4.99%	8.53%
$m(3)$	-1.71%	0.05%	3.10%	8.65%
$m(k\text{-fin})$	-1.89%	-0.93%	1.57%	7.28%
$m(k\text{-mar})$	-0.51%	0.70%	3.17%	8.85%
m(k-tot)	1.90%	3.15%	5.17%	9.36%
LB	25.92%	31.49%	36.14%	44.12%

Table 6.3: Demand Crash: policy performance in the Demand Crash scenario (Both AT and AR measures are presented)

policy	$L = 0$			$L = 4$		
	$AT(\pi)$	$AR(\pi)$	HC	$AT(\pi)$	$AR(\pi)$	HC
$\Delta(\mathbf{M})$	21.43%	13.19%	71.45%	18.15%	16.62%	45.72%
$\Delta(\mathbf{m})$	21.54%	13.24%	71.86%	18.79%	15.32%	52.85%
$\Delta(\mathbf{B})$	21.54%	13.24%	71.85%	18.30%	15.60%	49.59%
B	22.10%	13.12%	75.57%	18.31%	12.29%	64.37%
SB	19.87%	12.39%	64.21%	20.93%	15.60%	65.32%
SBB	20.69%	13.01%	67.41%	20.43%	14.82%	67.49%
$SBBM$	20.69%	13.01%	67.36%	20.42%	14.85%	67.26%
$B(0.5)$	22.66%	12.66%	80.23%	14.93%	5.81%	74.85%
$B(2)$	20.19%	12.98%	65.27%	20.20%	16.43%	50.88%
$B(\alpha\text{-myo})$	19.08%	12.31%	61.14%	15.82%	13.90%	46.74%
m	22.66%	12.65%	80.33%	14.13%	3.92%	81.51%
$m(2)$	10.65%	8.36%	32.37%	11.18%	11.39%	26.20%
$m(3)$	15.31%	11.19%	47.80%	15.02%	14.10%	40.10%
$m(k\text{-fin})$	21.44%	13.25%	71.14%	19.74%	13.59%	69.90%
$m(k\text{-mar})$	14.53%	10.23%	44.92%	20.30%	15.15%	64.57%
$\mathbf{m(k-tot)}$	13.33%	9.39%	40.95%	16.93%	15.37%	41.96%
LB	31.35%	23.94%	-	55.33%	56.03%	-

the Myopic policy, particularly as the lead time increases.

On average, all of the new policies do better relative to the Myopic policy with longer cycle lengths (because the effects of over-stocking last longer, hurting the Myopic policy). A closer look reveals that for $L = 4$, all of the new policies do much better than Myopic with a cycle length of 8 than they do with shorter cycle lengths. This suggests that the Myopic policy is less heavily affected by seasonality when the lead time is long enough to include at least one full cycle.

With regard to lead times, the recommended policies ($\Delta(M)$, $\Delta(m)$, $\Delta(B)$, SB and $m(k\text{-tot})$) exhibit a mixed, but fairly stable performance as the lead time grows from $L = 0$ to 4. In marked contrast, the policies that are not recommended all suffer as the lead time increases and sometimes perform worse than Myopic. This is consistent with the general pattern discussed in the "Product Launch" scenarios above.

6.2 Second Order Effects

Finally, we consider briefly the remaining three sets of scenarios. The initial forecast vector in these scenarios is flat because the focus is on the investigation of second order effects (that is, the form of the variance-covariance matrix of the updates, Σ).

Coefficient of Variation In Table 6.5 we report the average value of $AR(\pi)$ for all policies under the Coefficient of Variation scenarios. They demonstrate clearly that the new policies' performance improvement increases as the coefficient of variation increases. For highly variable forecast change (C.V.=8), the improvement over myopic can be as high as 30%. This is caused by the fact that for larger

Table 6.4: Seasonality: $AR(\pi)$, for certain Seasonality scenarios

	$L = 0$			
policy	Base	Step(2)	Step(4)	Step(8)
$\Delta(\mathbf{M})$	0.39%	5.65%	5.07%	7.20%
$\Delta(\mathbf{m})$	0.55%	4.15%	5.51%	8.21%
$\Delta(\mathbf{B})$	0.55%	4.15%	5.51%	8.21%
B	0.37%	5.52%	4.83%	7.20%
SB	0.13%	2.22%	3.39%	4.80%
$B(0.5)$	0.36%	5.49%	4.82%	7.18%
$B(2)$	0.46%	5.89%	5.22%	6.69%
$B(\alpha\text{-myo})$	0.39%	5.20%	4.81%	5.89%
m	0.36%	5.50%	4.82%	7.18%
$m(2)$	0.46%	6.01%	5.15%	5.93%
$m(3)$	0.44%	5.90%	5.15%	7.16%
$m(k\text{-fin})$	0.44%	5.76%	5.14%	7.22%
$m(k\text{-mar})$	0.29%	4.38%	3.98%	4.61%
m(k-tot)	0.29%	4.19%	3.61%	4.02%
LB	4.01%	22.93%	20.29%	18.22%
	$L = 4$			
$\Delta(\mathbf{M})$	1.97%	3.20%	3.69%	7.68%
$\Delta(\mathbf{m})$	1.43%	2.61%	3.26%	7.00%
$\Delta(\mathbf{B})$	1.51%	2.64%	3.31%	7.32%
B	-2.58%	-1.97%	-2.52%	2.71%
SB	1.91%	3.01%	3.50%	7.44%
$B(0.5)$	-4.15%	-3.67%	-5.16%	-0.96%
$B(2)$	1.52%	2.39%	2.83%	6.99%
$B(\alpha\text{-myo})$	1.47%	2.21%	2.57%	5.12%
m	-4.32%	-3.85%	-5.36%	-1.18%
$m(2)$	0.82%	2.20%	3.19%	7.20%
$m(3)$	-1.71%	-0.17%	-0.47%	5.99%
$m(k\text{-fin})$	-1.89%	-1.05%	-2.03%	2.91%
$m(k\text{-mar})$	-0.51%	0.88%	0.41%	5.05%
m(k-tot)	1.90%	3.08%	3.79%	7.14%
LB	25.92%	30.48%	34.38%	40.82%

Table 6.5: Coefficient of Variation: $AR(\pi)$, for certain Coefficient of Variation scenarios ($L = 0$)

C.V.	0.5	0.7	1	2	4	8
$\Delta(\mathbf{M})$	0.02%	0.28%	1.83%	10.03%	22.68%	28.97%
$\Delta(\mathbf{m})$	0.02%	0.41%	1.86%	10.13%	23.09%	29.71%
$\Delta(\mathbf{B})$	0.02%	0.41%	1.86%	10.13%	23.09%	29.71%
B	0.01%	0.23%	1.59%	9.74%	22.22%	26.84%
SB	0.00%	0.06%	1.02%	9.65%	22.21%	29.15%
$B(0.5)$	0.01%	0.23%	1.58%	8.63%	19.21%	20.46%
$B(2)$	0.01%	0.29%	1.94%	10.74%	22.26%	29.36%
$B(\alpha\text{-myo})$	0.01%	0.24%	1.78%	10.77%	22.87%	29.01%
m	0.01%	0.23%	1.58%	8.62%	18.98%	18.81%
$m(2)$	0.01%	0.28%	1.92%	10.09%	19.32%	24.54%
$m(3)$	0.01%	0.27%	1.86%	11.03%	22.96%	29.81%
$m(k\text{-fin})$	0.01%	0.27%	1.87%	10.42%	22.74%	27.73%
$m(k\text{-mar})$	0.01%	0.15%	1.68%	10.45%	22.98%	30.23%
m(k-tot)	0.01%	0.16%	1.48%	10.07%	22.43%	30.17%
LB	0.39%	2.99%	10.27%	35.60%	57.68%	72.40%

coefficients of variation the demand can fall quickly, resulting in over-stocking by the Myopic policy.

Learning For $L = 0$, Table 6.6 gives the average value of $AR(\pi)$, averaged over all policies under the various learning scenarios. Since the forecast horizon (12 months) is much longer than the lead time (0-4 months), we would expect that under early learning the Myopic policy would see approximately deterministic demand. For the same reason, late learning should favor the new policies. This is born out by Table 6.6, but the gain is small. Otherwise, the patterns that we have seen in other scenario sets with respect to the performance of different policies and the effect of different lead times, hold here as well.

Correlation Table 6.7 shows the results for the Correlation scenarios. These show that under the scenarios where updates are positively correlated, the new

Table 6.6: Time of Learning: Average of $AR(\pi)$ over all policies, for the Time of Learning scenarios ($L = 0$)

Scenario	Early	Mid	Const	Late
Average	0.00%	0.04%	0.39%	1.21%

Table 6.7: Correlation: $AR(\pi)$ for the Correlation scenarios ($L = 0$)

Scenario	None	Pos(4)	Neg(4)	Mix(4)	Pos(8)	Neg(8)	Mix(8)
$\Delta(\mathbf{M})$	1.05%	0.18%	1.06%	2.19%	0.10%	1.65%	2.19%
$\Delta(\mathbf{m})$	1.04%	0.22%	1.25%	2.39%	0.09%	1.44%	2.41%
$\Delta(\mathbf{B})$	1.04%	0.22%	1.25%	2.39%	0.09%	1.44%	2.41%
B	1.00%	0.17%	0.93%	2.15%	0.07%	1.69%	2.17%
SB	0.35%	0.02%	0.34%	0.49%	0.01%	0.54%	0.49%
$B(0.5)$	1.00%	0.17%	0.94%	2.15%	0.07%	1.69%	2.17%
$B(2)$	1.04%	0.21%	1.07%	2.05%	0.15%	1.61%	2.07%
$B(\alpha\text{-myo})$	0.89%	0.15%	0.94%	1.69%	0.12%	1.33%	1.68%
m	1.00%	0.17%	0.93%	2.15%	0.07%	1.69%	2.17%
$m(2)$	0.99%	0.21%	1.00%	2.10%	0.16%	1.58%	2.15%
$m(3)$	1.02%	0.19%	0.97%	2.16%	0.12%	1.64%	2.20%
$m(k\text{-fin})$	1.02%	0.20%	0.98%	2.15%	0.13%	1.68%	2.18%
$m(k\text{-mar})$	0.71%	0.17%	0.83%	1.44%	0.14%	1.14%	1.46%
m(k-tot)	0.68%	0.13%	0.78%	1.37%	0.11%	1.07%	1.39%
LB	0.98%	2.73%	7.27%	8.85%	2.72%	7.45%	8.90%

policies are less of an improvement over the Myopic policy than in the Base Case (where there is no correlation). The improvement over the Base Case is about the same or slightly greater when the off-diagonal elements of Σ are all negative, and is greatest when the signs are mixed. None of the improvements is greater than about 2.5%. This is due to the fact that with positive correlation the demand is approximately constant. However, in the negative and mixed scenarios the demands are choppier, and when both large and small demands are present the Myopic policy is more likely to overstock.

Table 6.8: Number of scenarios in which each policy performs best (under the $AR(\pi)$ measure)

	$L = 0$	$L = 4$
$\Delta(\mathbf{M})$	2	17
$\Delta(\mathbf{m})$	14	1
$\Delta(\mathbf{B})$	11	11
B	1	-
SB	-	-
$B(0.5)$	1	-
$B(2)$	3	3
$B(\alpha\text{-myo})$	-	-
m	-	-
$m(2)$	3	1
$m(3)$	1	-
$m(k\text{-fin})$	1	-
$m(k\text{-mar})$	1	-
$\mathbf{m(k-tot)}$	-	5

6.3 Other Aspects

Robustness It is useful to summarize the performance of the heuristics over the different scenarios. We approach this in two different ways. First, for all of the 38 scenarios and for each of the lead time cases ($L = 0$ and $L = 4$), we compute the number of times each policy is the best, and report this in Table 6.8. The Delta- π Myopic set of policies clearly dominate the others; the best policy comes from among these three in 56 out of the 76 scenario/lead-time pairs. Another of our five recommended policies, $m(k\text{-tot})$ is the best under five scenarios when $L = 4$. The final policy we recommend, SB , is never the best; however it is robust in another sense which we now define.

To arrive at this other measure of robustness, for each of the 76 scenarios, we compute the percentage by which $AR(\pi)$, the average relative cost of each policy π , exceeds that of the best performing policy for that scenario. Then, for each

policy, we compute the mean of these 76 values, along with the median, and the 70th, 75rd and overall (76th) highest values. This information is reported in Table 6.9. Note that in this table all the entries are positive and that small numbers indicate a better and more robust performance.

In this study it matters whether we use the average total cost $AT(\pi)$ or our preferred measure, $AR(\pi)$. For the best policies both measures are reported. Again, we see that our Delta- π Myopic policies perform very well; they are never more than about 3.5% worse than the best policy in any scenario.

We see from this table that the Surplus-Balancing policy SB is also very robust. Specifically, it never exceeds the cost of best policy by more than 4.1 percent and on average is within 1 percent of the lowest expected cost. The $m(k-tot)$ policy performs well also, as does $B(\beta-my)$.

We provide a further study of robustness in Figure 6.1. This shows a scatter plot of the cost on each of the 1000 runs in the Demand Crash scenario with $L = 4$ for the Myopic policy (Y-axis) against the SB policy. The 200 runs with the highest ratio are marked with an 'x'. We note that most of the runs clump below the 45-degree line. In this region, the SB policy out-performs the Myopic policy. Were it not for the 'extreme' runs marked with an 'x', the SB policy would out-perform the Myopic by even more.

The new policies outperform the Myopic policy in this Demand Crash scenario because they correctly anticipate the crash and temper their orders accordingly. However, in the case that there is an unexpected large demand just before the crash, they will incur a larger backorder cost than that incurred by Myopic because of their lower order-up-to levels. This is what happens in the extreme cases considered above. In Figure 6.2 we graph the evolution of the inventory on-hand for the SB

Table 6.9: Robustness statistics: % above the cost of the best heuristic, across the 76 scenarios

Results are for $AR(\pi)$ unless otherwise noted

policy	Mean	Median	70th of 76	75rd of 76	Highest
$\Delta(\mathbf{M})$	0.24%	0.10%	0.81%	2.67%	3.42%
$\Delta(\mathbf{M}) - \mathbf{AT}$	0.22%	0.08%	0.48%	2.80%	3.52%
$\Delta(\mathbf{m})$	0.19%	0.05%	0.67%	1.77%	1.98%
$\Delta(\mathbf{m}) - \mathbf{AT}$	0.31%	0.18%	0.95%	2.08%	2.71%
$\Delta(\mathbf{B})$	0.16%	0.00%	0.54%	1.67%	1.98%
$\Delta(\mathbf{B}) - \mathbf{AT}$	0.30%	0.15%	0.82%	2.39%	3.34%
B	2.32%	0.89%	5.34%	6.83%	7.11%
SB	0.76%	0.40%	2.10%	5.31%	6.30%
SB - AT	0.53%	0.14%	2.16%	3.84%	4.05%
$B(0.5)$	3.68%	1.02%	10.99%	13.48%	14.00%
$B(2)$	0.70%	0.25%	1.04%	12.56%	15.89%
$B(\alpha\text{-myo})$	0.75%	0.38%	2.58%	6.18%	7.01%
m	4.65%	1.00%	12.37%	37.99%	43.15%
$m(2)$	1.23%	0.40%	4.45%	11.58%	14.95%
$m(3)$	1.72%	0.63%	3.97%	6.35%	7.01%
$m(k\text{-fin})$	2.35%	0.58%	5.61%	16.15%	18.32%
$m(k\text{-mar})$	1.73%	0.82%	3.71%	12.60%	14.59%
m(k-tot)	0.61%	0.25%	1.95%	5.61%	6.67%
m(k-tot) - AT	0.79%	0.09%	3.43%	9.78%	12.06%

Table 6.10: Improvement in $AR(\pi)$ due to bounding

policy	Mean (Min, Max)	[< Min, > Myo]
B	3.43% (0.04%, 11.54%)	[70.21%, 10.76%]
SB	1.96% (0.03%, 5.00%)	[0.00%, 62.60%]
$B(0.5)$	17.51% (0.39% 33.21%)	[86.02% 7.05%]
B(2)	0.27% (0.00%, 5.26%)	[38.75%, 19.12%]
$B(\beta\text{-myo})$	1.95% (0.00%, 5.06%)	[0.00%, 40.07%]

policy for the five most extreme of these scenarios. We see that there clearly was a large and unexpected demand just before the crash, as hypothesized.

The Effect of Bounding We now turn to the effect of the Bounding improvement scheme that we discussed in Chapter 4. Note that this does not apply to the Minimizing policies as they fall within the bounds by definition. Also note that the Surplus-Balancing policy SB might exceed the upper bound, but it cannot order less than the lower bound. Table 6.10 lists the average improvement in $AR(\pi)$ generated by bounding for these policies, as well as the minimum and maximum improvement. Note that the improvement is always positive, as we would expect from Theorems 4.2.1 and 4.2.2.

The two policies B and $B(0.5)$ are dramatically improved by bounding. The reason for this is that more often than not they fall outside of the known limits on the optimal order-up-to levels, provided by the Minimizing and Myopic policies. To demonstrate this, Table 6.10 also includes the percentage of all order-up-to levels that fall either below the Minimizing level or above the Myopic level. The policies B and $B(0.5)$ fall outside this range 80.97% and 93.07% of the time respectively.

We demonstrate this graphically in Figure 6.3, which plots the evolution of order-up-to level of the Myopic, Minimizing, and three Balancing policies in a single run, for one of the Coefficient of Variation scenarios. To highlight the difference,

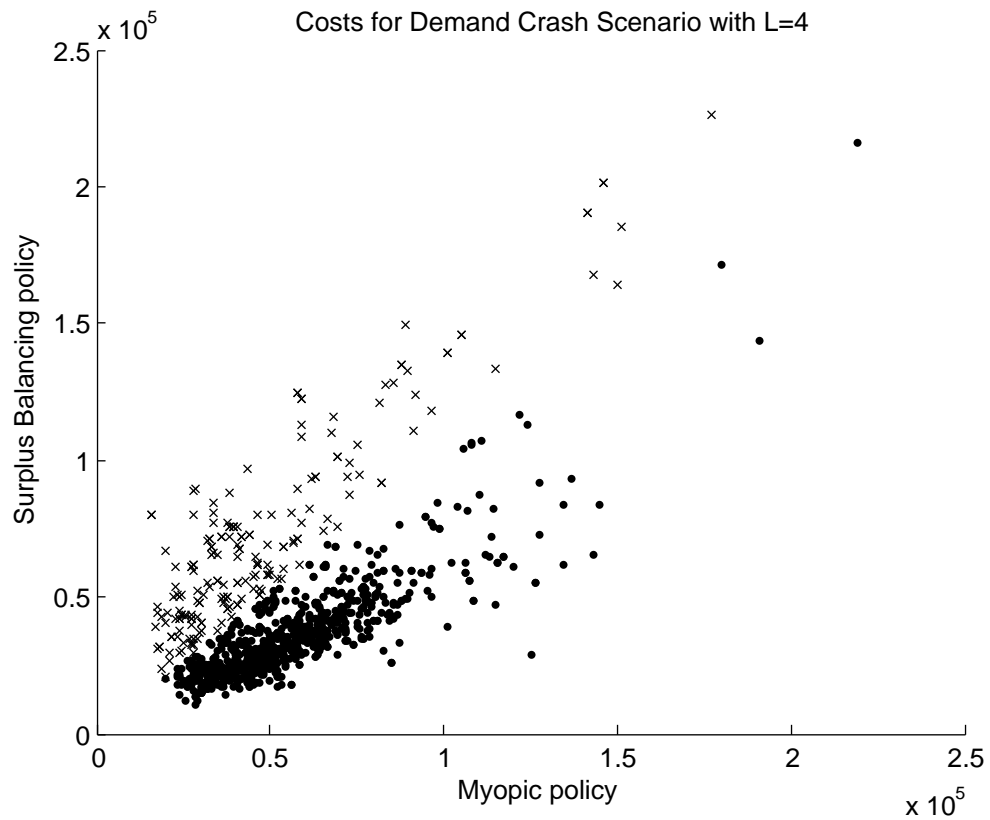


Figure 6.1: Scatter plot of the cost on each of the 1000 runs in the Demand Crash scenario with $L = 4$. Myopic policy (Y-axis) against the *SB* policy (X-axis). Highest 200 highest ratios are marked with an 'x'.

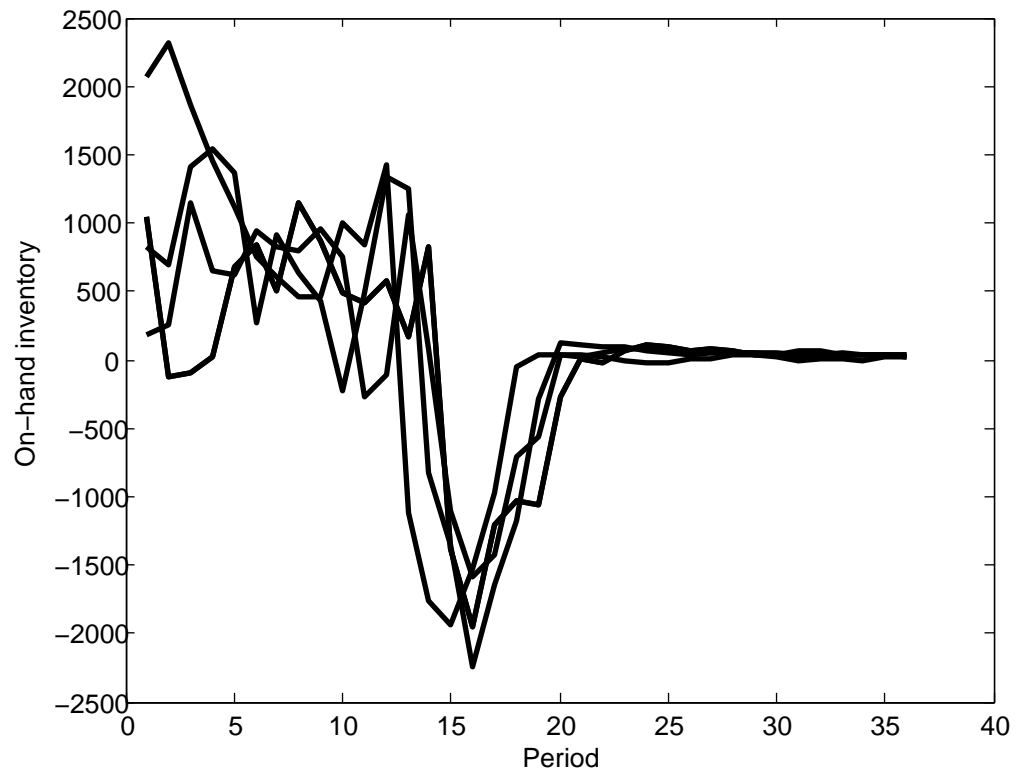


Figure 6.2: Graph of the inventory on-hand for the SB policy for each period in the Demand Crash scenario ($L = 4$) for the five runs with the highest ratio of SB 's cost to Myopic's.

we subtract from each order-up-to level, the order-up-to level of the Minimizing policy (which corresponds to the heavy, horizontal line at 0). The other heavy line corresponds to the Myopic policy. It is interesting to note that $B(2)$ (the dashed line) closely tracks the optimal range, while $B(0.5)$ (dashed and dotted line) is always below it. The Surplus-Balancing policy (SB , the lighter solid line) is often above the range - Table 6.10 indicates that this occurs 62.60% of the time, on average.

Comparison between the variants of Delta- π Myopic policies This chapter has demonstrated that the Delta- π Myopic improvement scheme is very successful. We considered applying the scheme to three base policies - the Myopic, Minimizing and Balancing policies. Even though the Minimizing and Balancing policies each perform better than the Myopic policy, we do not see such a noticeable advantage to applying the Delta- π scheme to the Minimizing or Balancing policies as opposed to the Myopic policy. For example, in the Steep Curve scenario from the End of Life group, when $L = 4$, the improvement over Myopic for $\Delta(MY)$ is 10.62%, whereas for $\Delta(m)$ and $\Delta(B)$ this increases only to 10.86% and 10.94% respectively.

Given the extra computational effort required to compute the post-order inventory levels for the Minimizing and Balancing policies, this small improvement is not worth it. Thus, while it was interesting to consider the three variants, we recommend only using $\Delta(MY)$.

Final Remarks Our three Delta- π Myopic Policies perform very well, both in terms of improvement over the Myopic Policy and in terms of their robustness. Additionally, they are computationally tractable and there is a strong mathematical

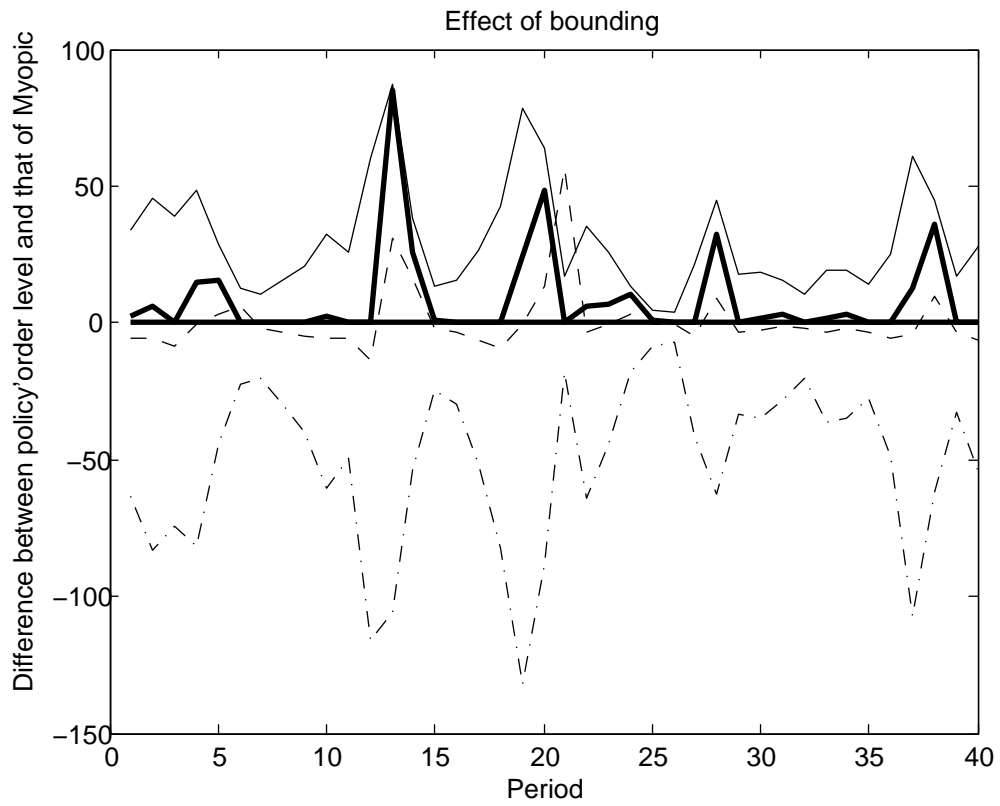


Figure 6.3: Difference between the order-up-to levels of the following policies and Minimizing, before bounding: 1) Myopic (thick solid line), 2) $B(0.5)$ (dashed and dotted line), 3) $B(2)$ (dashed line) and 4) SB (solid line).

justification underpinning them. These facts mean that they can be recommended for use on the stochastic inventory problem for both practical and theoretical reasons.

Among the three Delta- π Myopic policies we studied, we saw that using either the Balancing or Minimizing Policies as reference policies did lead to slightly better performance, especially in terms of robustness. The improvement was slight, though, and must be balanced against the extra computational effort required to compute the base-stock levels of the Minimizing Policy or the realized post-order inventory levels of the Balancing Policy. Thus, among this group of three, using the Delta- π scheme with the normal Myopic Policy as the reference policy seems to capture most of the improvement.

This trade-off between computational effort and performance results extends to the comparison of the Delta- π Myopic policies with the extensions to the Minimizing and Balancing Policies that we present. The Surplus Balancing policy, *SB* and Minimizing Policy with dynamic choice of k , $m(k\text{-tot})$ perform quite well and, unlike the Delta- π Myopic policies, do not require a Monte Carlo scheme to evaluate.

BIBLIOGRAPHY

- [1] A. Abu-Dayya and N. Beaulieu. Outage probability in the presence of correlated lognormal interferers. *IEEE Transactions on Vehicular Technology*, 43(1):164–173, 1994.
- [2] R. Anupindi, T. Morton, and D. Pentico. The nonstationary stochastic lead-time inventory problem: Near-myopic bounds, heuristics, and testing. *Management Science*, 42(1):124–129, 1996.
- [3] Y. Aviv. The effect of collaborative forecasting on supply chain performance. *Management Science*, 47(10):1326–1343, 2001.
- [4] Y. Aviv. Gaining benefits from joint forecasting and replenishment processes: The case of auto-correlated demand. *Manufacturing & Service Operations Management*, 4(1):55–75, 2002.
- [5] Y. Aviv. A time-series framework for supply-chain inventory management. *Operations Research*, 51(2):210–227, 2003.
- [6] K. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31(9):1150–1160, 1985.
- [7] N. Beaulieu, A. Abu-Dayya, and P. McLane. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Transactions on Communications*, 43(12):2869–2873, 1995.
- [8] A. Berns, J. Heppolette, M. Sadowski, and T. Siegel. Demand forecasting and production simulation at kraft general foods. Master’s thesis.
- [9] S. Bhaskaran. Simulation analysis of a manufacturing supply chain. *Decision Sciences*, 29(3):633–657, 1998.
- [10] J. Bradley and P. Glynn. Managing capacity and inventory jointly in manufacturing systems. *Management Science*, 48(2):273–288, 2002.
- [11] M. Cakanyildirim and R. Roundy. Sedfam: semiconductor demand forecast accuracy model. *IIE Transactions*, 34(5):449–465, 2002.
- [12] T. Cheevaprawatdomrong and R. Smith. Infinite horizon production scheduling in time-varying systems under stochastic demand. *Operations Research*, 52(1):105–115, 2004.
- [13] N. Dellaert and M. Melo. Approximate solutions for a stochastic lot-sizing problem with partial customer-order information. *European Journal of Operational Research*, 150(1):163–180, 2003.

- [14] N. Erkip, W. Hausman, and S. Nahmias. Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands. *Management Science*, 36(3):381–392, 1990.
- [15] M. Fu. Sample path derivatives for (s,s) inventory systems. *Operations Research*, 42(2):351–364, 1994.
- [16] M. Fu and K. Healy. Techniques for optimization via simulation: An experimental study on an (s,s) inventory system. *IIE Transactions*, 29(3):191–199, 1997.
- [17] G. Gallego and Ö. Özer. Integrating replenishment decisions with advance demand information. *Management Science*, 47(10):1344–1360, 2001.
- [18] P. Glasserman. *Gradient Estimation via Perturbation Analysis*. Kluwer Academic Publishers, 1991.
- [19] P. Glasserman and S. Tayur. Sensitivity analysis for base-stock levels in multi-echelon production-inventory systems. *Management Science*, 41(2):263–281, 1995.
- [20] J. Goto, M. Lewis, and M. Puterman. Coffee, tea, or ...?: A markov decision process model for airline meal provisioning. *Transportation Science*, 38(1):107–118, 2004.
- [21] S. C. Graves, D. B. Kletter, and W. B. Hetzel. A dynamic model for requirements planning with application to supply chain operations. *Operations Research*, 46:S35–S49, 1998.
- [22] S. C. Graves, H. Meal, S. Dasu, and Y. Qin. Two-stage production planning in a dynamic environment. In S. Axsater, C. Schneeweiss, and E. Silver, editors, *Multi-Stage Production Planning and Control*, Lecture Notes in Economics and Mathematical Systems, pages 9–43. Springer-Verlag, 1986.
- [23] R. Güllü. On the value of information in dynamic production inventory problems under forecast evolution. *Naval Research Logistics*, 43(2):289–303, 1996.
- [24] R. Güllü. A two-echelon allocation model and the value of information under correlated forecasts and demands. *European Journal of Operational Research*, 99(2):386–400, 1997.
- [25] D. C. Heath and P. L. Jackson. Modeling the evolution of demand forecasts with application to safety stock analysis in production distribution-systems. *IIE Transactions*, 26(3):17–30, 1994.
- [26] Y. Ho and X. Cao. Optimization and perturbation analysis of queueing networks. *Journal of Optimization Theory and Applications*, 40:559–582, 1983.

- [27] Y. Ho, M. Eyster, and T. Chien. A gradient technique for general buffer storage design in a serial production line. *International Journal of Production Research*, 17:557–580, 1979.
- [28] Y. Ho and J. Hu. An infinitesimal perturbation analysis algorithm for a multi-class g/g/1 queue. *Operations Research Letters*, 9:35–44, 1990.
- [29] G. Hurley, P. Jackson, R. Levi, R. Roundy, and D. Shmoys. New policies for the stochastic inventory control problem: Theoretical and computational results. Technical report, School of Operations Research and Industrial Engineering, Cornell University, 2006.
- [30] G. J. Hurley, P. L. Jackson, R. Levi, R. O. Roundy, and D. B. Shmoys. A theoretical and computational study of new algorithms for the stochastic inventory control problem. Technical report, Department of Operations Research and Industrial Engineering, Cornell University.
- [31] T. Iida and P. Zipkin. Approximate solutions of a dynamic forecast-inventory model. Technical report, Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2003.
- [32] G. Janakiraman and J. A. Muckstadt. Inventory control in directed networks: A note on linear costs. *Operations Research*, 52(3):491–495, 2004.
- [33] G. Johnson and H. Thompson. Optimality of myopic inventory policies for certain dependent demand processes. *Management Science*, 21(11):1303–1307, 1975.
- [34] P. Kaminsky and J. Swaminathan. Utilizing forecast band refinement for capacitated production planning. *Manufacturing & Service Operations Management*, 3(1):68–81, 2001.
- [35] S. Karlin. Dynamic inventory policy with varying stochastic demands. *Management Science*, 6:231–258, 1960.
- [36] S. Karlin. Optimal policy for dynamic inventory process with stochastic demands subject to seasonal variations. *Journal of the Society of Industrial and Applied Mathematics*, 8:611–629, 1960.
- [37] H. Kim and J. Ryan. The cost impact of using simple forecasting techniques in a supply chain. *Naval Research Logistics*, 50(5):388–411, 2003.
- [38] L. Krajewski and J. Wei. The value of production schedule integration in supply chains. *Decision Sciences*, 32(4):601–634, 2001.
- [39] R. Levi, M. Pal, R. Roundy, and D. Shmoys. Approximation algorithms for stochastic inventory control models. Technical report, Department of Operations Research and Industrial Engineering, Cornell University, 2004.

- [40] A. Ligeti. Outage probability in the presence of correlated lognormal useful and interfering components. *IEEE Communications Letters*, 4(1):15–17, 2000.
- [41] W. S. Lovejoy. Myopic policies for some inventory models with uncertain demand distributions. *Management Science*, 36(6):724–738, 1990.
- [42] W. S. Lovejoy. Stopped myopic policies in some inventory models with generalized demand processes. *Management Science*, 38(5):688–707, 1992.
- [43] X. Lu, J. Song, and A. Regan. Inventory planning with forecast updates: Approximate solutions and cost error bounds. Technical report, Graduate School of Management, University of California, Irvine, 2003.
- [44] M. Milevsky and S. Posner. Asian options, the sum of lognormals and the reciprocal gamma distribution. *Journal of Financial and Quantitative Analysis*, 33(3):409–422, 1998.
- [45] B. L. Miller. Scarf’s state reduction method, flexibility, and a dependent demand inventory model. *Operations Research*, 34:83–90, 1986.
- [46] T. Morton. The nonstationary infinite horizon inventory problem. *Management Science*, 24(14):1474–1482, 1978.
- [47] T. Morton. Infinite-horizon dynamic programming models - a planning-horizon formulation. *Operations Research*, 27(4):730–742, 1979.
- [48] T. Morton and D. Pentico. The finite horizon nonstationary stochastic inventory problem: Near-myopic bounds, heuristics, testing. *Management Science*, 41(2):334–343, 1995.
- [49] J. A. Muckstadt and C. R. Sox. Multi-item, multi-period production planning with uncertain demand.
- [50] Ö. Özer. Replenishment strategies for distribution systems under advance demand information. *Management Science*, 49(3):255–272, 2003.
- [51] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
- [52] R. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [53] A. Safak. Statistical analysis of the power sum of multiple correlation lognormal components. *IEEE Transactions on Vehicular Technology*, 42(1):58–61, 1993.
- [54] H. Scarf. Bayes solutions of the statistical inventory problem. *Annals of Mathematical Statistics*, 30(2):490–508, 1959.

- [55] H. Scarf. Some remarks on bayes solution to the inventory problem. *Naval Research Logistics Quarterly*, 7:591–596, 1960.
- [56] S. Schwartz and Y. Yeh. On the distribution function and moments of power sums with lognormal components. *Bell Syst. Tech. J.*, 61:1441–1462, 1982.
- [57] S. Sethi, H. Yan, and H. Zhang. Peeling layers of an onion: Inventory model with multiple delivery modes and forecast updates. *Journal of Optimization Theory and Applications*, 108(2):253–281, 2001.
- [58] S. Sethi, H. Yan, and H. Zhang. Inventory models with fixed costs, forecast updates and two delivery modes. *Operations Research*, 51(2):321–328, 2003.
- [59] S. Sethi, H. Yan, and H. Zhang. Quantity flexibility contracts: Optimal decisions with information updates. *Decision Sciences*, 35(4):691–712, 2004.
- [60] R. Snyder, A. Koehler, R. Hyndman, and J. Ord. Exponential smoothing models: Means and variances for lead-time demand. *European Journal of Operational Research*, 158(2):444–455, 2004.
- [61] R. Snyder, A. Koehler, and J. Ord. Forecasting for inventory control with exponential smoothing. *International Journal of Forecasting*, 18(1):5–18, 2002.
- [62] L. Toktay and L. Wein. Analysis of a forecasting-production-inventory system with stationary demand. *Management Science*, 47(9):1268–1281, 2001.
- [63] A. Veinott. Optimal stockage policies with non-stationary stochastic demands. In H. Scarf, D. Gilford, and M. Shelly, editors, *Multistage Inventory Models and Techniques*, pages 85–115. Stanford University Press, Stanford, CA., 1963.
- [64] A. Veinott. Optimal policy for a multi-product, dynamic, non-stationary inventory problem. *Management Science*, 12(3):206–222, 1965.
- [65] F. Zhang, A. Heching, S. Hood, J. Hosking, Y. Leung, R. Roundy, and J. Wong. Industrial data experiments on demand forecasts combination. Technical report, Department of Operations Research and Industrial Engineering, Cornell University, 2004.
- [66] V. Zhou, P. Jackson, R. Roundy, and R. Zhang. The evolution of family level sales forecasts into product level forecasts. Technical report, Johnson Graduate School of Management, Cornell University, 2004.
- [67] K. Zhu and U. Thonemann. Modeling the benefits of sharing future demand information. *Operations Research*, 52(1):136–147, 2004.
- [68] P. Zipkin. *Foundations of Inventory Management*. The McGraw-Hill Companies, Inc., 2000.