Statistical Estimation Problems in Inventory Management

Alp Eren Akcay

A Dissertation
Presented to the Faculty
of Carnegie Mellon University
for the Degree of
Doctor of Philosophy

Pittsburgh, Pennsylvania
May 2013

Abstract

Most of the literature on inventory management assumes that demand distributions and the parameters that characterize these distributions are known with certainty. However, this is not the case in practice and the unknown parameters must be estimated using only a finite (and sometimes very limited) amount of historical demand data. The sequential process of first estimating the parameters and then optimizing the inventory based on these estimates does not perform well especially when there is limited amount of data for estimation. The discrepancy between the performance of an optimal inventory target and the performance of its estimate from a short demand history is a serious, but often ignored, operational problem.

The first study in this dissertation aims to solve this problem by considering a demand history with distributional characteristics that are hard to capture using standard distributions, and uses a flexible system of distributions that can capture a wide variety of distributional shapes with asymmetry, peakedness, and tail weight. The second study, on the other hand, considers an intermittent demand history which includes many zero values because demand does not arrive every inventory-review period. In both of these studies, the objective is to develop inventory-target estimation methods that account for the operational costs of incorrectly estimating the unknown parameters in the demand model. In particular, we combine inventory management and parameter estimation into a single task to balance the costs of under- and overestimation of the optimal inventory target. In the third study, we focus on finding a probabilistic guarantee on the near-optimality of an inventory-target estimator in the presence of temporally dependent demand data. Our findings shed light on how the autocorrelation and tail dependence in a demand process affect the number of demand observations required to achieve a performance arbitrarily close to the performance of the optimal inventory target, which has been only investigated for independent and identically distributed demand in the inventory management literature.

Acknowledgements

In my last five years at Carnegie Mellon, I have closely worked with Prof. Bahar Biller and Prof. Sridhar Tayur. Bahar was a wonderful advisor, and she was always there for me throughout my time at Tepper. I would like to express my special thanks to her. I greatly benefited from Sridhar's wisdom and experience, and I would like to thank him for his support, guidance, and mentorship over the years. My thanks also go to Prof. Alan Scheller-Wolf, who provided very helpful feedback and insightful comments to improve this dissertation. I would also like to thank Prof. Fatma Kilinc-Karzan. She kindly served in my dissertation committee and provided many valuable suggestions and insights.

I have been very fortunate to be a part of the Operations Management group at Carnegie Mellon. I would like to thank its faculty and doctoral students for creating an inspirational and challenging environment. I was one of the three students admitted to Operations Management program five years ago. Xin Fang and Elvin Coban were my classmates. I would like to thank them for their friendship and collaboration in my early years at Tepper. I shared my office with Emre Nadar, Tinglong Dai, Canan Corlu, Michele Dufalla, Vince Slaugh, Xin Wang, Gaoqing Zhang, and Erkut Sonmez. It was a pleasure working alongside them throughout the years. I would also like to thank Tepper's doctoral program coordinator Lawrence Rapp for all his help. I gratefully acknowledge the financial support provided by the William Larimer Mellon fellowship.

Finally, I am thankful to my parents and my brother for their continuous support, encouragement, and love. My deepest thanks go to Tugce Gizem Martagan, who never ceased to believe in and support me. I feel very lucky to have Gizem in my life. This dissertation is dedicated to her.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A fundamental issue in the theory and practice of inventory management has been the modeling of random demand since the inception of stochastic inventory models in the 1950s. Wagner (2002) distinguishes two approaches in his assessment of the relation between inventory management and the use of historical demand data during the fifty-year period from early 1950s to 2000s. The first approach builds on the premise of using point forecasts and observed variation in forecast error to answer the fundamental questions of inventory management such as 'is it now time to replenish inventory?' and 'what should be the order quantity?'. Pioneered by Brown (1959), this approach aims to answer these questions with the current demand data on hand and is motivated by the requirements of practice. An example is Brown's renowned exponential smoothing method: The inventory manager uses historical data to forecast the demand over lead time plus a review period and accounts for the accompanying forecast error so as to hedge against demand uncertainty. The forecast-error distribution purportedly reveals inherent uncertainty about future demand, and typically, is assumed to be normally distributed. Wagner (2002) designates the approach of using point forecasts and observed variation in forecast error by the term 'PFErr'. The second approach, on the other hand, is based on applying an inventory replenishment formula obtained from a stochastic inventory model by using repeatedly updated statistical estimates of the parameter values in the assumed demand distribution of the inventory model. Wagner (2002) names this alternative approach 'OREst'. We refer the reader to Wagner (2002) and the references therein for the underlying assumptions of the 'PFErr' approach. These assumptions typically are seriously violated in practice, and thus contribute to its failure to deliver what it promises in terms of the target service levels and inventory costs.

In this dissertation, we focus on the so-called 'OREst' approach. Most of the literature on inventory management assumes that the demand (and other uncertain quantities) is described completely by probability distributions with precisely specified parameter values. However, an explicit consideration of the impact of statistical errors arising from having only a limited amount of data (for example, due to a nonstationary environment) on the performance of the stochastic inventory models is conspicuously lacking from the current literature. Wagner (2002) states that dealing with this impact is one of the major challenges facing a practitioner. For example, when the inventory manager uses historical data to estimate the mean and standard deviation of the postulated demand distribution, the inventory target – which is known to be optimal for the true values of the mean and standard deviation – may be either overestimated or underestimated. The costs of underestimation and overestimation are often not the same, and thus they must be taken into account in inventory-target estimation.

We start our study of inventory-target estimation by analyzing industrial demand data; we identify a large number of demand histograms with significant levels of asymmetry and tail weight. This naturally casts doubt on the performance of inventory-target estimates obtained from limited amounts of demand data under the assumption of a normal demand distribution. Our industrial data analysis also points out the need for an adequate representation of the shape of demand distributions, especially with respect to the right tail of the demand, which is the most critical for inventory management. In Chapter 3, we address this issue in a newsvendor setting by modeling the demand with a flexible system of distributions known as the Johnson Translation System, which captures a wide variety of distributional shapes with asymmetry, peakedness, and tail weight. Using the Johnson translation system for demand modeling permits the development of procedures that require no restrictive assumptions about the first four moments of the demand random variable. In this first study we consider a newsvendor problem and quantify the expected cost of parameter uncertainty (also called the inaccuracy in inventory-target estimation) as a function of the length of the historical demand data, the critical fractile, and the shape parameters of the demand distribution. We also develop an inventory-target estimation procedure that minimizes the expected total operating cost, which is the sum of the expect cost due the stochastic demand uncertainty and the expected cost due to parameter uncertainty. That is, the resulting inventory-target estimate accounts for the operational costs of incorrectly estimating the unknown parameters in the demand model.

In Chapter 4, we consider the problem of estimating inventory targets from limited amount of demand data when the stationary demand process is intermittent. In every period, either a positive demand or zero demand is observed with an unknown probability. We represent the size of positive demand by a location-scale family of distributions with unknown mean and variance. Similar to our first study, we first quantify the expected cost of treating the estimates of the unknown parameters as if they were the true values. We then minimize this expected cost due to parameter uncertainty with respect to a threshold variable that factors the estimation error into inventory decision. In certain cases, the use of an optimized threshold leads to significant reductions in the expected cost of parameter uncertainty.

Motivated by industrial data of intermittent demand, we also introduce a copula-based demand model to capture the correlation between demand size and the number of zero-demand periods preceding the demand. In this general setting, we propose two finite-sample hypothesis tests to investigate the existence of correlation in an intermittent demand history. We show that a statistical test which accounts for the expected cost of parameter uncertainty tends to reject the independence assumption less frequently than a statistical test which only considers the sampling distribution of the copula-parameter estimator. We find that such situation arises especially when the intermittent demand history is short and the percentage of nonzero demand realizations in the demand history is small.

In Chapter 5, we focus on a demand process with temporal dependence where the demand realization in a time period depends on the past realizations of demand in the previous periods. The AutoRegressive (AR) process with normally distributed random shocks is one model with these characteristics that is widely used in inventory management. It is well known that the linearity of this process implies that the marginal demand distribution is normal. The distinguishing feature of our study is the use of a copula-based time series model and its semi-parametric estimation that allow the decision maker to avoid any parametric assumptions about the marginal demand distribution. The objective of our study is to identify a probabilistic guarantee on the near-optimality of the resulting inventory-target estimator. More specifically, we are interested in finding a lower bound to the probability of the expected cost of the inventory-target estimator being at most $1 + \varepsilon$ of the minimum expected cost of the optimal inventory target, where ε is a positive constant. This lower bound serves as a confidence level for the decision maker to assure a performance guarantee to the inventory-target estimate obtained from a limited demand history.

Conditions that ensure this so-called ε-optimality of an inventory target estimate have been studied in inventory management literature in the presence of independent and identically distributed demand observations. To the best of our knowledge, we are the first to investigate an ε-optimality guarantee for temporally dependent demand data. Our findings shed light on how the autocorrelation and tail dependence (i.e., the dependence at very high or very low demand realizations) in a demand process affect the number of demand observations required to achieve a performance arbitrarily close to the performance of the optimal inventory target. We also provide insights on when it is safe to ignore the temporal dependence in a demand history in light of the additional cost imposed by the statistical estimation error around the temporal-dependence parameters, and when it is necessary to account for the temporal dependence in the demand process despite the limited history of temporally dependent demand.

We next present a general overview of the inventory management literature under incomplete demand information in Chapter 2. We will provide a more detailed review specialized to each of the own studies in the relevant chapters of the dissertation. Likewise, the necessary mathematical notation is defined in each chapter. Finally, we conclude with a summary and potential future research directions in Chapter 6.

Chapter 2

Related Work

In this chapter, we present an overview of the inventory management literature under incomplete demand information. A more detailed discussion of the past research related to each of the three studies will be presented in the subsequent chapters.

Inventory management literature under incomplete demand information falls in two main categories: Parametric and nonparametric. Parametric methods assume that the true demand distribution belongs to a known parametric family of distributions, and use parameters estimated from historical demand data. This assumption is relaxed by nonparametric methods, which use the empirical distribution for demand modeling. A parametric method can be further classified into two streams: Bayesian methods which treat unknown parameters as random variables and update their distributions over time, and the frequentist methods which assume a fixed true value for each unknown parameter. Our work in Chapters 3 and 4 is in "parametric and frequentist" stream. In Chapter 5, we position our work as "semiparametric and frequentist" as we adopt a nonparametric method to estimate the marginal demand distribution and a parametric method to estimate the temporal dependence structure.

The Bayesian methods choose a prior distribution for an unknown parameter and update its distribution with the observations collected over time; see Scarf (1959, 1960), and Iglehart (1964) for early studies. This approach builds on the assumption that the prior and posterior distributions have the same form. The main premise of this method is reducing an inventory control problem with a two-dimensional state space (i.e., one state for the inventory level and another for a sufficient statistic of the historical demand data) to a revised dynamic program with a single state variable, and then scaling the optimal inventory target of the revised problem up to the solution of the original problem. This procedure works well when the demand is restricted to specific distributions such as exponential, gamma, and range families. Azoury (1985) extends the method to the case where demand is uniform or Weibull. Lovejoy (1990) uses Bayesian analysis to show that a myopic inventory policy based on a critical fractile can be optimal or near optimal in the incomplete information setting. In contrast, we assume that an unknown parameter is not random but has a fixed value which is only known by nature. We capture the uncertainty around the point estimate of an unknown parameter by its sampling distribution, as in papers in the frequentist stream.

Hayes (1969) pioneered the frequentist approach in inventory management to estimate the newsvendor quantity that hedges against both the stochastic demand uncertainty and the uncertainty around the parameter estimates, when the demand distribution is either exponential or normal. Hayes achieves this by minimizing a frequentist risk measure called Expected Total Operating Cost (ETOC). Katircioglu (1996) revisits Hayes' method and shows that the method applies not only to exponential and normal distributions but to all distributions known up to a location and scale parameter. Akcay et al. (2011) extend this approach to a flexible, transformation-based system of distributions to capture a wider variety of distributional shapes. In parallel to Hayes (1969), Liyanage and Shanthikumar (2005) introduce the concept of operational statistics to maximize the a priori expected profit; i.e., the expected newsvendor profit where the expectation is taken with respect to demand history and the demand in the forthcoming period, when the demand distribution is known up to a scale parameter. Chu et al. (2008) obtain the optimal operational statistic for the location-scale family of distributions. Ramamurthy et al. (2012) also consider a newsvendor model and propose a heuristic based on operational statistics to obtain improved inventory targets in the presence of an unknown shape parameter.

The objective of all studies above is to integrate inventory optimization and parameter estimation into a single task. This is also the goal in Chapters 3 and 4 in this dissertation. In Chapter 3, we use a flexible family of distributions named Johnson Translation System (JTS, Johnson 1949) as in Akcay et al. (2011) to capture a wide variety of distributional shapes for the demand size. In Chapter 4, we represent the demand size with location-scale family of distributions as in Katircioglu (1996) and Chu et al. (2008), but we also consider intermittency of demand which is quite common in practice. Furthermore, we explicitly model the correlation between the demand size and the number of inter-demand periods, and minimize the a priori expected cost of not acting optimally (i.e., the expected regret) due to incorrect estimation of intermittent demand parameters.

In Chapter 5, our objective is not to integrate parameter estimation and inventory optimization. Instead, we focus on finding a probabilistic guarantee on the ε-optimality of a practical inventory-target estimation method in the presence of temporally dependent demand data. The inventory-target estimate in Chapter 5 corresponds to the Sample Average Approximation (SAA) solution of Levi et al. (2007) under the assumption of independent and identically distributed demand data. To the best of our knowledge, we are the first in investigating an ε-optimality guarantee for temporally dependent demand data.

Robust inventory models address the ambiguity in the demand distribution by considering a family of distributions characterized by their descriptive statistics such as mean, variance, mode, and range. Scarf (1958) and Gallego and Moon (1993) identify the newsvendor quantity that maximizes the minimum profit that would occur when only the demand mean and variance are specified. Moon and Gallego (1994) and Gallego et al. (2001) apply this approach to multi-period models. Perakis and Roels (2008) derive the order quantity that minimizes the newsvendor's maximum regret by also considering factors such as asymmetry and unimodality of the demand distribution.

Since we minimize the expected total operating cost in Chapter 3 and the expected cost of parameter uncertainty associated with incorrectly estimating the unknown parameters in Chapter 4, our work is related to this research stream. But, in contrast, we make direct use of historical demand data instead of assuming partial knowledge about the demand distribution. Bertsimas and Thiele (2006) address demand ambiguity in a multiperiod inventory control problem by specifying an uncertainty set parameterized by a budget of uncertainty based on the framework of robust optimization. See and Sim (2010) introduce a factor-based demand model that can capture autocorrelation as well. More recently, Klabjan et al. (2013) introduce a model that only requires historical data and minimize the worst case expected cost over a set of demand distributions, which is defined as all the possible distributions satisfying a chi-square goodness-of-fit test. In this dissertation, we directly account for the random nature of demand parameter estimators instead of building uncertainty sets for unknown demand data.

We conclude this section with a review of the nonparametric methods which make no assumptions regarding the parametric form of the demand distribution; Bookbinder and Lordahl (1989), Lordahl and Bookbinder (1994), and Godfrey and Powell (2001) are some early examples. More recently, Huh and Rusmevichientong (2009), Huh et al. (2011), and Besbes and Muharremoglu (2013) focus on a repeated newsvendor problem to develop adaptive data-driven algorithms for setting inventory targets and study the implications of demand censoring. Considering the newsvendor problem and its multi-period extension, Levi et al. (2007) identify the minimum number of independent demand samples drawn from the true demand distribution to guarantee that the difference between the expected cost of the policy using the empirical demand distribution and the expected cost of the optimal policy with full access to the demand distribution is not greater than a prespecified error with a certain level of confidence. Levi et al. (2012) characterize the properties of the demand distribution that impact the quality of an inventory target obtained from the empirical demand distribution. Huh et al. (2011) propose nonparametric adaptive inventory control policies that converge almost surely to the optimal solution over time in the presence of historical sales data. We focus on parametric methods of inventory-target estimation in Chapters 3 and 4. In Chapter 5, the copula-based representation of temporally dependent demand allows the decision maker to construct an estimate of the multivariate distribution, which characterizes the stationary time-series data, by using the empirical demand distribution function. That is, the decision maker follows a nonparametric approach for the estimation of the marginal demand distribution and a parametric approach for the estimation of the temporal dependence in the demand process. The so-called semiparametric approach leads to an estimate of the critical fractile solution as a function of the empirical demand distribution and the estimated values of the temporal dependence parameters.

Chapter 3

Improved Inventory Targets in the Presence of Limited Historical Demand Data

3.1 Introduction

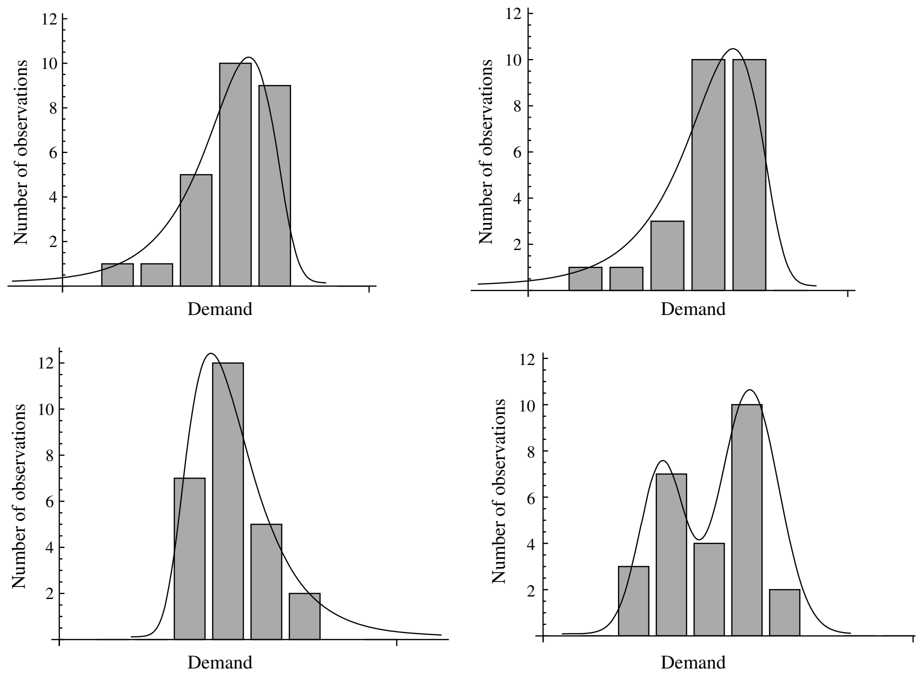
Traditionally, the literature on inventory management assumes that the demand distribution and the values of its parameters are known with certainty. However, this is not the case in practice, and inventory targets must be estimated using only a finite (and sometimes, very limited) amount of historical demand data. In this chapter, we consider this practical situation and describe the repeated newsvendor setting of interest as follows: (i) Historical demand data $\{x_t;\ t = 1, 2, \ldots, n\}$ of length n are available with no forecasting process in place. (ii) The cumulative distribution function (cdf) of the demand random variable is assumed to be $F(\cdot; \boldsymbol{\theta})$ with the stationary (but unknown) parameter vector $\boldsymbol{\theta}$. (iii) Maximum likelihood estimation (MLE) is used for obtaining the estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ from the historical demand data. That is, $\hat{\boldsymbol{\theta}}$ is chosen as the vector of parameter values that maximize the likelihood function of the observed data $\{x_t;\ t = 1, 2, \ldots, n\}$. Under fairly general regularity conditions, the MLEs are consistent and their asymptotical joint distribution approaches the multivariate normal distribution with the minimum attainable covariance matrix (Rohatgi and Saleh, 2000). (iv) The decision maker sets the inventory target to $\hat{Q} = F^{-1}(\varphi; \hat{\boldsymbol{\theta}})$, where φ is the critical fractile defined as the ratio of unit shortage cost to the sum of unit shortage and inventory holding costs. The inventory policy that determines the inventory target in this way is called the *maximum likelihood policy* (MLP) (Scarf 1959; Fukuda 1960; Gupta 1960).

The decision maker is, however, rarely fortunate enough to be in an asymptotic situation. Since a small sample size is generally the rule when estimating an inventory target, the asymptotical properties of the MLEs may not hold in practical cases with limited historical demand data. Nevertheless, the MLP builds on the implicit assumption that the estimated demand cdf $F(\cdot; \hat{\boldsymbol{\theta}})$ is identical to the true demand cdf. Consequently, the MLP ignores the uncertainty around $\hat{\boldsymbol{\theta}}$ and hence, the uncertainty around the estimated inventory target $\hat{Q}$. Hayes (1969) quantifies the inaccuracy in the estimation of this particular inventory target using the concept of Expected Total Operating Cost (ETOC); i.e., the expected one-period cost associated with operating under an estimated inventory policy. Assuming exponentially and normally distributed demands, Hayes identifies the inventory targets that minimize the ETOC in the presence of limited historical demand data. The distinguishing feature of the resulting inventory targets is that they are *biased estimators* in a framework that combines statistical estimation with inventory optimization. While it is the first study to show that a statistically good job of estimation does not necessarily result in an inventory target that minimizes the ETOC, Hayes' results are limited to demand random variables that are either exponentially or normally distributed.

In this chapter, we quantify the inaccuracy in inventory-target estimation and identify the ETOC-minimizing inventory target for a demand random variable that does not necessarily have an exponential or normal distribution. For the exponentially distributed demand, the coefficient of variation is 1, the coefficient of skewness $\sqrt{\beta_1}$ (i.e., the measure of the symmetry of the demand's density function) is 2, and the coefficient of kurtosis β_2 (i.e., the measure of the peakedness and the tail weight of the demand's density function) is 9 for any value of mean, while $\sqrt{\beta_1}$ is 0 and β_2 is 3 for the normally distributed demand with any pair of mean and variance. Our study differs from Hayes (1969) in that we allow the inventory manager to avoid any assumptions about the first four moments (i.e., mean, variance, $\sqrt{\beta_1}$, and β_2) of the demand random variable. We do this by representing the demand with the Johnson translation system (JTS); i.e., a parameterized family of distributions that has the ability of matching any finite first four moments of a random variable (Johnson, 1949). The use of JTS for demand modeling provides the flexibility of capturing (unimodal and bimodal) distributional shapes with different levels of symmetry, peakedness, and tail weights.

Both historical demand data of brick-and-mortar companies from the Fortune 1000 list (Cornacchia and Shamir, 2009) and data collected by SmartOps Corporation from Fortune 500 companies in manufacturing, consumer packaged goods, chemicals, technology, and distribution/retail industries (SmartOps Corporation, 2009) indicate that the flexibility provided by the JTS is, in fact, necessary. Specifically, Cornacchia and Shamir (2009) find that demands of products with highly-skewed, long-tailed distributions often constitute a large portion of a company's total demand. For example, 98% of the products of an automotive aftermarket parts company have highly skewed demand and these products have a share of 62% in the total revenue. The percentage of products with similar demand characteristics and the share of these products in total revenue are 86% and 46% for a consumer packaged goods company, and 44% and 36% for a food and beverage company. The data collected by SmartOps Corporation from a consumer packaged goods company and illustrated

Figure 3.1: Histograms of the demand data collected by SmartOps Corporation for products from a consumer packaged goods company



in Figure 3.1 exhibit similar demand characteristics as well as bimodality; so these features of demand cannot be ignored.

A close look at the existing literature reveals that the impact of the shape of the demand's density function (and thus, $\sqrt{\beta_1}$ and β_2) on the inventory-target estimation has often been overlooked. For example, Naddor (1978) and Fortuin (1980) discuss that inventory decisions are best described by the mean and the variance of the demand, and the coefficient of skewness $\sqrt{\beta_1}$ and the coefficient of kurtosis β_2 may not be important. However, this discussion is based on modeling the demand with one- and two-parameter distributions, which can only represent limited values of $\sqrt{\beta_1}$ and β_2. Actually, the demand distribution function often must have at least four parameters for the adequate representation of the mean, dispersion, skewness, and long tails (Lau et al., 1998). Indeed, Heuts et al. (1986) revisit the examples in Naddor (1978) using the four-parameter Schmeiser-Deutsch distribution (Schmeiser and Deutsch, 1977), and conclude that the optimal inventory target depends on the shape of the demand's density function. Kottas and Lau (1980) also point out the insufficiency of two-parameter distributions in representing demands with skewed distributional shapes; the authors use four-parameter distributions of the Pearson's system (Pearson, 1895) and the Schmeiser-Deutsch's system. This is one of the early papers noting that the use of standard distributions for demand modeling may be overly restrictive and unrealistic, while flexible systems of distributions such as the Pearson's system and the Schmeiser-Deutsch's system are versatile and realistic; see Lau and Zaki (1982), Kumaran and Achary (1996), Bartezzaghi et al. (1999), and Tang and Grubbstrom (2006) for further discussion on capturing the distributional shape of stochastic demand with a flexible distribution system.

In this study, we also use a flexible distribution system – the Johnson Translation System (JTS) – for representing the demand distribution. The ease in the application of the ETOC concept of Hayes (1969) for setting inventory targets depends on the availability of well-defined and easy-to-compute sampling distributions for the MLEs of the unknown demand parameters. In this aspect, JTS is a good candidate for demand modeling in comparison to the flexible distribution systems used in the past inventory research. For example, the curves of the JTS generally agree with Pearson curves having (nearly) the same first four moments. However, the cdf of the JTS allows us to obtain the sampling distributions of the Johnson shape parameter estimates by invoking well-established normal distribution theory.

By using the ETOC concept together with the JTS for flexible demand modeling, we set the inventory target of a product by accounting for the uncertainty around the demand parameters estimated from finite historical data. We summarize our contributions as follows:

1. We use the JTS for demand modeling within the ETOC framework for the first time in inventory management.

2. We quantify the inaccuracy in the inventory-target estimation as a function of the length of the historical data, the critical fractile, and the shape parameters of the demand distribution.

3. In the presence of this inaccuracy, we identify the inventory target with the minimum ETOC, accounting for the uncertainty around the demand parameters estimated from limited historical data. We do that by seeking the ETOC-minimizing inventory target estimate from a class of estimators implied by the JTS. We name the inventory policy that determines the inventory target in this way as Hayes Inventory Policy (HIP).

4. We extend HIP to set inventory targets subject to Type 1 and Type 2 service-level constraints.

The remainder of this chapter is organized as follows. In Section 3.2, we present the modeling framework. We quantify the inaccuracy in the inventory-target estimation in Section 3.3, and use HIP for identifying the ETOC-minimizing inventory targets in Section 3.4. In Section 3.5, we provide insights on how the shape of the demand's density function influences the effectiveness of HIP in search of an improved inventory targets in the presence of limited historical demand data. We discuss the implementation details in Section 3.6 and the extension of HIP for service-level constraints in Section 3.7.

3.2 The Model

In Section 3.2.1, we present the assumptions of the inventory model and introduce the JTS for demand modeling. We present the maximum likelihood policy with the Johnson translation system (MLP with JTS) in Section 3.2.2 and the Hayes Inventory Policy (HIP) in Section 3.2.3. In Section 3.2.4, we discuss the computation of the ETOC function.

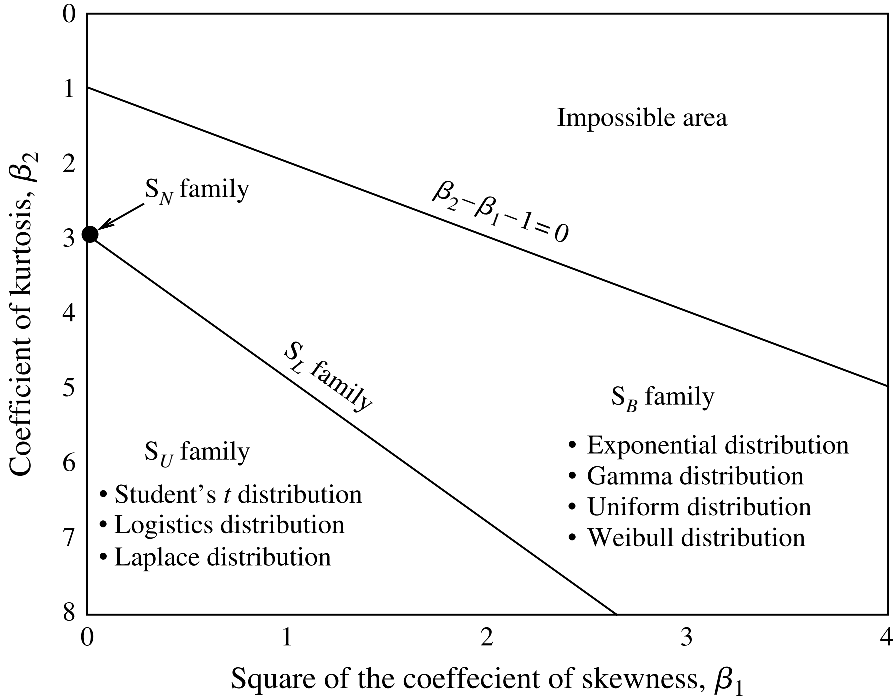
3.2.1 Demand Model and Newsvendor Problem

In a repeated newsvendor setting, we study the problem of determining inventory targets in the presence of limited historical demand data. Specifically, the decision maker approaches the ordering decision of each period with zero stock on hand, and sets the inventory target to q units that arrive instantaneously at the beginning of each period. The inventory target is determined using only a historical data set comprised of independent demand observations. This reflects the situation where the decision maker can keep track of all of the orders, and the records regarding the lost sales can be retained even if some of the orders cannot be met, e.g., when the orders arrive through Internet, fax, or phone (Cachon and Terwiesch, 2009). The stock of q units is then depleted by the random demand X. We assume that all remaining units, if any, are disposed and the shortages are written off at the end of each period.

A common assumption in modeling the random demand X is that the true demand distribution is a standard distribution such as exponential, gamma, or normal. However, the shapes represented by the standard families of distributions are limited, and therefore, they might fail to capture the distributional characteristics of the historical demand data. It is also possible that a goodness-of-fit test might reject or accept all candidate demand distributions, depending on the number of available demand observations. Furthermore, the inventory manager might not be familiar with all the standard distributions on a lengthy list that can be used for modeling the demand. We overcome these challenges of demand modeling by using a highly flexible system of distributions known as the Johnson Translation System (JTS). Specifically, the JTS for random demand X is defined by a cdf of the form

$$F(x; \gamma, \delta, \xi, \lambda) = \Phi\left(\gamma + \delta r\left(\frac{x-\xi}{\lambda}\right)\right),$$

Figure 3.2: The (β_1, β_2)-space covered by the JTS



where γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, $\Phi(\cdot)$ is the cdf of the standard normal random variable, and $r(\cdot)$ is one of the following transformations:

$$r(y) = \begin{cases} \log(y) & \text{for the } S_L \text{ (lognormal) family} \\ \log\left(y + \sqrt{y^2+1}\right) & \text{for the } S_U \text{ (unbounded) family} \\ \log(y/(1-y)) & \text{for the } S_B \text{ (bounded) family} \\ y & \text{for the } S_N \text{ (normal) family} \end{cases}$$

There is a unique family (choice of r) for each feasible combination of the coefficient of skewness $\sqrt{\beta_1}$ and the coefficient of kurtosis β_2 that determine the shape parameters γ and δ. Figure 3.2 presents the (β_1, β_2)-space covered by the JTS and illustrates the regions captured by each Johnson family. It also displays the location-scale family of distributions considered by Katircioglu (1996) to provide examples of the standard distributions represented by the JTS. Figure 3.2 indicates that the distributions of the S_U family approximate the Student's t, logistics, and Laplace distributions by matching their first four moments, while the S_B family approximates the exponential, gamma, uniform, and Weibull distributions. It is important to note that the JTS can represent any feasible (finite) first four moments and hence, the distributional shapes represented by the JTS are not limited to the shapes of the standard distributions that appear in Figure 3.2. More specifically, the use of the JTS for demand modeling enables us to solve the inventory problem of interest for any pair of $\sqrt{\beta_1}$ and β_2 that a continuous demand random variable can have. Any mean and variance can also be attained by any one of the Johnson families. Within each family, a distribution is completely specified by the values of the parameters γ, δ, λ, and ξ, and the range of X depends on the family of interest: $X > \xi$ and $\lambda = 1$ for the S_L family; $\xi < X < \xi + \lambda$ for the S_B family; $-\infty < X < \infty$ for the S_U family; and $-\infty < X < \infty$, $\xi = 0$, and $\lambda = 1$ for the S_N family. Figure 3.3 provides examples of the probability density functions (pdfs) represented by each family of the JTS with $\xi = 0$ and $\lambda = 1$. Further examples of the pdfs captured by the JTS can be found in Johnson (1987).

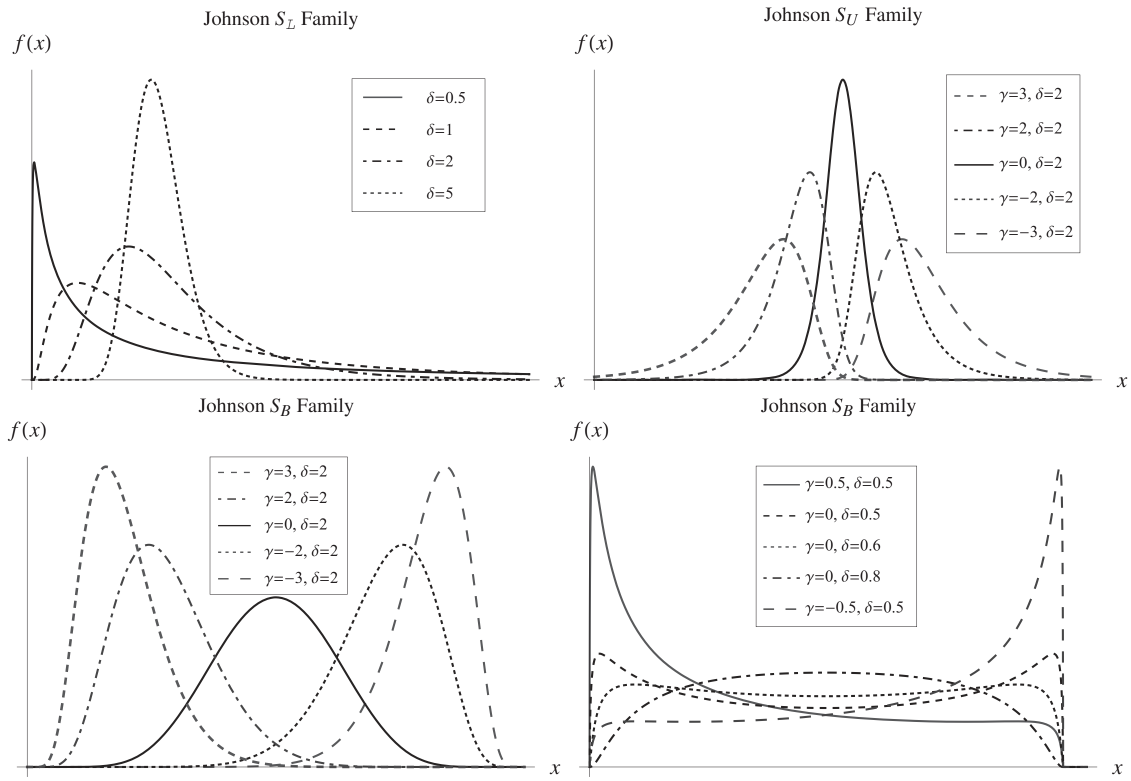
All the relevant revenues and costs associated with the decision to order q units for the forthcoming period are incorporated into a loss function $L(q, x)$ that is piecewise linear in $q - x$. Specifically, we take

$$L(q,x) \quad = \quad \begin{cases} h(q-x) & \text{for } q \geq x, \\ b(x-q) & \text{for } q < x, \end{cases}$$

where h is the unit inventory holding cost (i.e., unit cost of excess stock) and b is the unit cost of shortage. Thus, for each family of the JTS the expected loss function can be written as

$$\bar{L}(q;\gamma,\delta,\xi,\lambda) \quad = \quad \int_{\Omega_L}^{\Omega_U} L(q,x)\mathrm{d}F(x;\gamma,\delta,\xi,\lambda),$$

Figure 3.3: Examples of the pdfs represented by the JTS with $\xi = 0$ and $\lambda = 1$.



where Ω_L and Ω_U are the lower and upper bounds of the support of the Johnson demand distribution. Without loss of generality, we take h as 1 and use $\varphi/(1-\varphi)$ for b, where φ denotes the critical fractile $b/(h+b)$. Thus, we represent the expected loss in terms of the inventory holding cost in the remainder of the chapter.

Since the expected loss function $\bar{L}(q;\gamma,\delta,\xi,\lambda)$ is convex in q, the optimal inventory target q^* is the unique value of q that minimizes $\bar{L}(q;\gamma,\delta,\xi,\lambda)$ in the forthcoming period. We represent the optimal inventory target under complete certainty about the demand distribution function $F(\cdot;\gamma,\delta,\xi,\lambda)$ by $q^* = F^{-1}(\varphi;\gamma,\delta,\xi,\lambda)$. Because the equation $F(q^*;\gamma,\delta,\xi,\lambda) = \varphi$ can be equivalently written as $\Phi\left(\gamma + \delta r\left((q^*-\xi)/\lambda\right)\right) = \varphi$, we have $z_\varphi = \gamma + \delta r((q^*-\xi)/\lambda)$, where z_φ is the φ^{th} quantile of the standard normal random variable Z (i.e., $\mathbb{P}(Z \le z_\varphi) = \varphi$). Therefore, the use of JTS for demand modeling leads to an optimal inventory target of the following form under complete certainty about the demand distribution:

$$q^* = \xi + \lambda r^{-1}\left(-\frac{\gamma}{\delta} + \frac{z_\varphi}{\delta}\right)$$

For ease of presentation in the remainder of the chapter, we assume that the type of the Johnson transformation r is known. We discuss the relaxation of this assumption in Section 3.6. However, the true values of the Johnson demand parameters γ, δ, ξ, and λ remain to be unknown. The decision maker faces the inventory-target estimation problem at the beginning of each period. Then, the demand parameters are estimated from scratch by using all the demand observations that constitute a stationary data set at the time of decision.

3.2.2 Maximum Likelihood Policy (MLP) with Johnson Translation System (JTS)

The inventory target under MLP with JTS is of the same form as the optimal inventory target under complete certainty except that the unknown JTS parameters ξ, λ, γ, and δ are replaced by their MLEs $\hat{\xi}$, $\hat{\lambda}$, $\hat{\gamma}$, and $\hat{\delta}$:

$$\hat{Q} = \hat{\xi} + \hat{\lambda} r^{-1}\left(-\frac{\hat{\gamma}}{\hat{\delta}} + \frac{z_\varphi}{\hat{\delta}}\right)$$

It is important to emphasize the roles played by different Johnson parameters in the distribution of the demand random variable X. If we write $X = \xi + \lambda Y$ with Y denoting the standard Johnson random variable, then we observe that the pdf of X is of the same shape as that of Y. Although the standard deviation of X is λ times that of Y and ξ only affects the expected value of X, parameters ξ and λ have no effect on the values of $\sqrt{\beta_1}$ and β_2. Thus, parameters ξ and λ have no effect on the shape of the density function of X. Since our goal is to investigate the relation between the statistical inventory-target estimation problem and the shape of the demand's density function, we assume that the location parameter ξ and the scale parameter λ are known, and represent $(X - \xi)/\lambda$ by the standard Johnson random variable Y. However, the mean and the variance of Y are still unknown as they are functions of the unknown shape parameters γ and δ.

Our focus on the standard Johnson random variable Y might seem restrictive, but it is less so than it first appears: It suffices to solve the inventory problem for the standard Johnson random variable when there exists expert opinion about the support of the demand distribution. If the demand is known to be larger than ξ' with no upper bound, then a distribution from the S_L family with $\xi = \xi'$ can be used for modeling the demand (i.e., $X = \xi' + Y$). If the demand is known to lie in the interval $(\xi', \xi' + \lambda')$, then a distribution from the S_B family with $\xi = \xi'$ and $\lambda = \lambda'$ can be used for representing the demand (i.e., $X = \xi' + \lambda' Y$). When there exists no such expert opinion, the inventory target is to be estimated by also accounting for the uncertainty around the estimates of the parameters ξ and λ. We describe how to do this in Section 3.6.2.

Since the parameters ξ and λ (> 0) are assumed to be known, we let $\xi = 0$ and $\lambda = 1$ for notational convenience and represent the optimal inventory target under complete certainty about the demand distribution function with $q^* = r^{-1}(-\gamma/\delta + z_\varphi/\delta)$. In the case of having $\xi = \xi'$ and $\lambda = \lambda'$, we set the inventory target to $\xi' + \lambda' q^*$. We are now ready to present the asymptotically unbiased MLEs of parameters $-\gamma/\delta$ and $1/\delta^2$ (Johnson, 1949):

$$-\frac{\hat{\gamma}}{\hat{\delta}} = \frac{1}{n} \sum_{t=1}^{n} r(x_t)$$

$$\frac{1}{\hat{\delta}^2} = \frac{1}{n-1} \sum_{t=1}^{n} \left(r(x_t) - \frac{1}{n} \sum_{\ell=1}^{n} r(x_\ell) \right)^2$$

We adjust the expression of $\hat{\delta}^{-2}$ to assure unbiasedness.

For notational convenience, we denote the random variable $-\hat{\gamma}/\hat{\delta}$ with $\bar{r}$ and use s_r for $1/\hat{\delta}$ in the remainder of the chapter. Since MLEs are invariant under functional transformations, the inventory-target estimate $\hat{Q} = r^{-1}(\bar{r} + z_\varphi s_r)$ is an MLE of $q^* = r^{-1}(-\gamma/\delta + z_\varphi/\delta)$. Furthermore, both $\bar{r}$ and s_r^2 have well-defined sampling distributions:

Proposition 3.2.1. *The distributional properties of random variables $\bar{r}$ and s_r^2 are as follows: (1) $\bar{r}$ and s_r^2 are independent. (2) The sampling distribution of $\bar{r}$, $f_{\bar{r}}(\bar{r})$ is normal with mean $-\gamma/\delta$ and variance δ^{-2}/n. (3) The sampling distribution of s_r^2, $f_{s_r^2}(s_r^2)$ is gamma with shape parameter $(n-1)/2$ and scale parameter $2\delta^{-2}/(n-1)$.*

Proof. (1) Since random variable $r(Y)$ is normally distributed with mean $-\gamma/\delta$ and variance $1/\delta^2$, the independence of random variables $\bar{r}$ and s_r^2 follows from Rohatgi and Saleh (2000). (2) The sampling distribution of $\bar{r}$ follows from the regenerative property of the independent and normally distributed random variables. (3) To derive the sampling distribution of s_r^2, we first multiply the expression for $1/\hat{\delta}^2$ with $\delta^2(n-1)$ and write $\sum_{t=1}^{n}(r(y_t) - \bar{r})^2$ as $\sum_{t=1}^{n}(r(y_t) + \gamma/\delta)^2 - n(\bar{r} + \gamma/\delta)^2$. Hence, the representation of $1/\hat{\delta}^2$ takes the form

$$\delta^2(n-1)s_r^2 + n\delta^2(\bar{r} + \gamma/\delta)^2 = \delta^2 \sum_{t=1}^{n}(r(y_t) + \gamma/\delta)^2,$$

where the right-hand side corresponds to a χ^2 random variable with n degrees of freedom. Because the second term on the left-hand side is the square of a standard normal random variable, it is also a χ^2 random variable, but with one degree of freedom. Consequently, $\delta^2(n-1)s_r^2$ is a χ^2 random variable with $n-1$ degrees of freedom with the following pdf:

$$\frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \left(\delta^2(n-1)s_r^2\right)^{\frac{n-1}{2}-1} \exp\left(-\delta^2(n-1)s_r^2/2\right)$$

Thus, the sampling distribution of s_r^2 is gamma with shape parameter $(n-1)/2$ and scale parameter $2\delta^{-2}/(n-1)$. $\square$

3.2.3 Hayes Inventory Policy (HIP)

Although the inventory-target estimate $\hat{Q} = r^{-1}(\bar{r} + z_\varphi s_r)$ under MLP with JTS is an MLE of the inventory target under complete certainty about the demand distribution, the MLP with JTS ignores the uncertainty around the demand parameter estimates $\bar{r}$ and s_r. Consequently, the MLP with JTS does not consider the uncertainty around the estimated inventory target $\hat{Q}$ and it does not minimize the expected total operating cost (ETOC). We, on the other hand, account for the demand parameter uncertainty and hence, the uncertainty around $\hat{Q}$ by using the Hayes inventory policy (HIP), a new inventory-target estimation policy that identifies improved inventory targets in the presence of limited historical demand data.

We obtain HIP by modifying (i.e., biasing) the inventory-target estimator under MLP with JTS to minimize the ETOC. Specifically, we replace the safety factor z_φ in the inventory-target estimator $\hat{Q} = r^{-1}(\bar{r} + z_\varphi s_r)$ under MLP with JTS by the bias parameter k, and set the inventory-target estimator $\hat{Q}(k)$ under HIP to $r^{-1}(\bar{r} + k s_r)$. We discuss the motivation behind using such a functional form for the inventory-target estimator by focusing on the normally distributed demand. Then, we extend our discussion to other families of the JTS by using the functional relation between the JTS and the normal distribution.

When the demand is normally distributed, the expected loss $\bar{L}(q^*; \gamma, \delta)$ does not depend on the mean demand. Also, underestimating the safety factor z_φ is more costly than overestimating it when the critical fractile is greater than 0.5, and a deliberate upward biasing of the safety factor reduces the expected cost (Silver and Rahmana, 1986). Hence, a way of identifying an ETOC-minimizing inventory-target estimator for the normally distributed demand is to adjust the level of the safety factor by introducing a bias parameter k for the sample standard deviation: $\hat{Q}(k) = \bar{r} + ks_r$. Further discussion on the use of biased estimators for normally distributed demand can be found in Hayes (1969), Weerahandi (1987), and Katircioglu (1996).

Since the random variable $r(Y)$ is normally distributed (Section 3.2.1), we introduce the bias parameter k to the sample standard deviation s_r for each family of the JTS; i.e., $r(\hat{Q}(k)) = \bar{r} + ks_r$. Consequently, we obtain an inventory-target estimator $\hat{Q}(k)$ of the form $r^{-1}(\bar{r}+ks_r)$, where the optimal value of k, k^* solves for the minimum of the expected total operating cost associated with setting the inventory target to $\hat{Q}(k)$. We call the inventory policy that estimates an inventory target in this way Hayes Inventory Policy (HIP) and use it in Section 3.4 for setting inventory targets in the presence of limited historical demand data.

3.2.4 Expected Total Operating Cost

In this section, we use the concept of expected total operating cost introduced by Hayes (1969) to represent the expected one-period cost ETOC$[\hat{Q}(k)]$ associated with setting the inventory target to $\hat{Q}(k) = r^{-1}(\bar{r} + ks_r)$ under HIP:

$$
\begin{aligned}
&\text{ETOC}[\hat{Q}(k)] \qquad (3.1)\\
&= \mathrm{E}_{\bar{r},s_r^2}\left[\bar{L}\left(r^{-1}(\bar{r}+ks_r);\gamma,\delta\right)\right]\\
&= \int_0^{\infty}\int_{-\infty}^{\infty}\int_{\Omega_L}^{r^{-1}(\bar{r}+ks_r)} \left[r^{-1}(\bar{r}+ks_r)-y\right]\ f_Y(y)\ f_{\bar{r}}(\bar{r})\ f_{s_r^2}(s_r^2)\ dy\ d\bar{r}\ ds_r^2\\
&\quad + \frac{\varphi}{1-\varphi}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{r^{-1}(\bar{r}+ks_r)}^{\Omega_U} \left[y-r^{-1}(\bar{r}+ks_r)\right]\ f_Y(y)\ f_{\bar{r}}(\bar{r})\ f_{s_r^2}(s_r^2)\ dy\ d\bar{r}\ ds_r^2
\end{aligned}
$$

In this representation, $\bar{L}(\hat{Q}(k);\gamma,\delta)$ is the expected loss function associated with setting the inventory target of the forthcoming period to $\hat{Q}(k)$; $f_Y(y)$ is the pdf of the standard Johnson random variable Y, i.e.,

$$f_Y(y) = \delta \, \frac{\partial r(y)}{\partial y} \, \phi\left(\gamma + \delta r\left(y\right)\right)$$

with ϕ denoting the pdf of the standard normal random variable; $f_{\bar{r}}(\bar{r})$ is the normal pdf of the estimated demand parameter $\bar{r}$ (Proposition 3.2.1), i.e.,

$$f_{\bar{r}}(\bar{r}) = \frac{1}{\sqrt{2\pi\delta^{-2}/n}} \exp\left(-\left(\bar{r}+\gamma/\delta\right)^2 / \left(2\delta^{-2}/n\right)\right);$$

and $f_{s_r^2}(s_r^2)$ is the gamma pdf of the estimated demand parameter s_r^2 (Proposition 3.2.1), i.e.,

$$f_{s_r^2}(s_r^2) = \frac{\left(2\delta^{-2}/(n-1)\right)^{-(n-1)/2} \left(s_r^2\right)^{(n-3)/2} \exp\left(-(n-1)s_r^2/(2\delta^{-2})\right)}{\Gamma\left((n-1)/2\right)}.$$

The sampling density functions $f_{\bar{r}}(\bar{r})$ and $f_{s_r^2}(s_r^2)$ tell us which values independent random variables $\bar{r}$ and s_r^2 can have and how likely it is for $\bar{r}$ and s_r^2 to assume those values. Thus, the use of these sampling density functions in (3.1) allows us to represent the expected one-period cost by accounting for the uncertainty around the MLEs of the Johnson demand parameters.

Two important insights emerge from the use of the ETOC for capturing the uncertainty around the MLEs of the Johnson demand parameters (Hayes, 1969): (1) The loss function represents the opportunity loss associated with a given choice of the inventory target. Therefore, the value of $\bar{L}(q^*;\gamma,\delta)$ can be interpreted as the expected value of perfect information about the demand distribution. (2) When the demand distribution is not completely specified, ETOC$[\hat{Q}(k)] - \bar{L}(q^*;\gamma,\delta)$ represents the expected value of perfect information about the sampling distribution of the demand parameters for the estimated inventory target $\hat{Q}(k)$. Thus, ETOC$[\hat{Q}(k)] - \bar{L}(q^*;\gamma,\delta)$ can be interpreted as the inaccuracy in the estimation of the inventory target $\hat{Q}(k)$. The next section uses this interpretation for quantifying the inaccuracy in the inventory-target estimation, while Section 3.4 seeks an inventory target that minimizes the ETOC within a class of inventory-target estimators implied by the JTS.

3.3 Inaccuracy Quantification under MLP

We first quantify the inaccuracy in inventory-target estimation in Section 3.3.1. In Section 3.3.1, we provide approximations that show how the length of the historical demand data, the asymmetry of the loss function, and the shape of the demand's density function affect the inaccuracy.

3.3.1 Quantification of the Inaccuracy in Inventory-Target Estimation

For the S_N family of the JTS, the inventory-target estimator $\hat{Q}(k)$ takes the form $\bar{r}+ks_r$. Therefore, the piecewise linear structure of the loss function $L(\hat{Q}(k), y)$ allows $L(\bar{r}+ks_r, y)$ to be written as $s_r L(k,t)$ with t denoting $(y-\bar{r})/s_r$. Consequently, we integrate out the random variables $\bar{r}$ and s_r^2 from the expression in (3.1) and obtain

$$\begin{aligned} \mathrm{ETOC}[\hat{Q}(k)] &= \frac{1}{\sqrt{2\pi}\delta}\sqrt{\frac{n}{n+1}}\int_{-\infty}^{k}(k-t)\left(1+\frac{n}{n^2-1}t^2\right)^{-\frac{n+1}{2}} dt \\ &+ \frac{\varphi}{(1-\varphi)\sqrt{2\pi}\delta}\sqrt{\frac{n}{n+1}}\int_{k}^{\infty}(t-k)\left(1+\frac{n}{n^2-1}t^2\right)^{-\frac{n+1}{2}} dt. \end{aligned} \quad (3.2)$$

$\mathrm{ETOC}[\hat{Q}]$ under MLP with JTS is obtained by evaluating this integral for $k = z_\varphi$. Therefore, the inaccuracy in the inventory-target estimation under MLP with Johnson's S_N family is given by $\Delta_{S_N} = \mathrm{ETOC}[\hat{Q}(z_\varphi)] - \bar{L}(q^*;\gamma,\delta)$, where $\bar{L}(q^*;\gamma,\delta) = \delta^{-1}\phi(z_\varphi)/(1-\varphi)$. Similarly, we use Δ_{S_L}, Δ_{S_U}, and Δ_{S_B} for denoting the inaccuracies in the inventory-target estimation under MLP with Johnson's S_L, S_U, and S_B families, respectively.

For the S_L, S_U, and S_B families of the JTS, however, the structure of the inverse transformation function $r^{-1}(\cdot)$ does not allow us to reduce the loss function $L(\hat{Q}(k), y)$ to a form that is piecewise linear in k. Also, we cannot integrate out the random variables $\bar{r}$ and s_r^2 from the expression in (3.1). Fortunately, the ETOC$[\hat{Q}(k)]$ can be easily evaluated by numerical integration for any family of the JTS. Specifically, we first reduce the three-dimensional integral in (3.1) to a one-dimensional integral for the S_L and S_U families of the JTS. Then, we use Mathematica's built-in numerical integration function to evaluate ETOC$[\hat{Q}(k)]$ by setting the precision goal p and the accuracy goal a to 5 and thus, forcing the integration error to be less than $10^{-a} + |q|\,10^{-p}$ in a result of size q (Wolfram Research, 2008).

3.3.2 Approximation of the Inaccuracy in Inventory-Target Estimation

The quantification of the inaccuracy in the inventory-target estimation requires the evaluation of the three-dimensional integral for the S_B family of the JTS. Therefore, a natural question to ask is whether we can approximate the inaccuracy under MLP with JTS. We answer this question by using the second-order Taylor expansion of $\bar{L}(\hat{Q};\gamma,\delta)$ about the optimal inventory target q^* under complete knowledge about the demand distribution:

$$\bar{L}(\hat{Q};\gamma,\delta) \doteq \bar{L}(q^*;\gamma,\delta) + (\hat{Q}-q^*)\bar{L}'(q^*;\gamma,\delta) + \frac{1}{2}(\hat{Q}-q^*)^2\bar{L}''(q^*;\gamma,\delta)$$

In this representation, $\bar{L}'(q^*;\gamma,\delta)$ and $\bar{L}''(q^*;\gamma,\delta)$ are the first-order and second-order derivatives of the expected loss function $\bar{L}(\cdot)$ evaluated at q^*. As a result of taking the expectation of $\bar{L}(\hat{Q};\gamma,\delta)$ with respect to $\hat{Q}$ and recognizing that $\bar{L}'(q^*;\gamma,\delta)$ is equal to zero, we obtain

$$\text{ETOC}[\hat{Q}] \doteq \bar{L}(q^*;\gamma,\delta) + \frac{1}{2}\left(\mathbb{E}(\hat{Q}^2) - 2q^*\mathbb{E}(\hat{Q}) + q^{*2}\right)\bar{L}''(q^*;\gamma,\delta),$$

where $\bar{L}''(q^*;\gamma,\delta)$ is equal to $f_Y(q^*;\gamma,\delta)/(1-\varphi)$. Then, the inaccuracy $\text{ETOC}[\hat{Q}] - \bar{L}(q^*;\gamma,\delta)$ associated with the inventory-target estimator $\hat{Q}$ can be approximated by

$$\frac{1}{2}\left(\mathbb{E}(\hat{Q}^2) - 2q^*\mathbb{E}(\hat{Q}) + q^{*2}\right)\delta r'(q^*)\frac{\phi(z_\varphi)}{1-\varphi},$$

where $r'(q^*)$ is the first-order derivative of the transformation function $r(\cdot)$ evaluated at q^*. As a result of approximating each of $\mathbb{E}(\hat{Q}^2)$ and $\mathbb{E}(\hat{Q})$ with a second-order Taylor expansion, we obtain the following inaccuracy approximations:

Approximation 3.3.1. *Defining* $g(n) := \Gamma[n/2]/\Gamma[(n-1)/2]$, $u := \delta^{-1}[-\gamma + z_\varphi\sqrt{2/(n-1)}g(n)]$, *and* $v := \delta^{-2}[1/n + (2z_\varphi^2/(n-1))((n-1)/2 - g(n)^2)]$, *we approximate the inaccuracy in the inventory-target estimation as follows:*

(i) For the S_N *family, the approximate inaccuracy* $\Delta^a_{S_N}$ *is given by*

$$\frac{1}{\delta}\left[z_\varphi^2\left(1-\sqrt{\frac{2}{n-1}}g(n)\right)+\frac{1}{2n}\right]\frac{\phi(z_\varphi)}{1-\varphi}.$$

(ii) For the S_L *family with* $q^* = \exp((z_\varphi - \gamma)/\delta)$, *the approximate inaccuracy* $\Delta^a_{S_L}$ *is given by*

$$\frac{\delta}{2q^*}\left[\left(e^u - q^*\right)^2 + e^u v\left(2\ e^u - q^*\right)\right]\frac{\phi(z_\varphi)}{1-\varphi}.$$

(iii) For the S_B *family with* $q^* = (1+\exp((-z_\varphi+\gamma)/\delta))^{-1}$, *the approximate inaccuracy* $\Delta^a_{S_B}$ *is given by*

$$\frac{\delta\ e^u}{2\ q^*(1-q^*)}\left[\frac{\left(1+e^u\right)^2 + \left(2-e^u\right)v}{\left(1+e^u\right)^4} - \frac{\left(2\left(1+e^u\right)^2 + \left(1-e^u\right)v\right)q^*}{\left(1+e^u\right)^3} + \frac{q^{*2}}{e^u}\right]\frac{\phi(z_\varphi)}{1-\varphi}.$$

(iv) For the S_U family with $q^* = \{\exp((z_\varphi - \gamma)/\delta) - \exp((-z_\varphi + \gamma)/\delta)\}/2$, the approximate inaccuracy $\Delta^a_{S_U}$ is given by

$$\frac{\delta}{2\sqrt{q^{*2}-1}}\left[\left(\frac{e^u - e^{-u}}{q^*}\right)^2 + \frac{v}{2}\left(e^{2u} + e^{-2u} - q^*\left(e^u - e^{-u}\right)\right)\right]\frac{\phi(z_\varphi)}{1-\varphi}.$$

The functional form of the approximate inaccuracy $\Delta^a_{S_N}$ suggests that the ETOC reduces to the expected loss with the full knowledge of the demand distribution as n and $\sqrt{2/(n-1)}g(n)$ approach infinity and 1, respectively. That is, the inaccuracy in the inventory-target estimation under MLP diminishes as the length of the demand data increases. Furthermore, $\Delta^a_{S_N}$ increases with the critical fractile φ as long as the underestimation of the inventory target is penalized more heavily than its overestimation (i.e., $\varphi > 0.5$).

Since the expected loss function $\bar{L}(q^*; \gamma, \delta)$ is given by $\delta^{-1}\phi(z_\varphi)/(1-\varphi)$ for the S_N family of the JTS, the percentage approximate inaccuracy $\Delta^a_{S_N}\%$ (i.e., $\Delta^a_{S_N}/\bar{L}(q^*; \gamma, \delta)$ 100%) is $z^2_\varphi[1 - \sqrt{2/(n-1)}g(n)] + 1/2n$, and it depends only on n and φ. For the S_L family of the JTS, on the other hand, the approximate percentage inaccuracy $\Delta^a_{S_L}\%$ also depends on the parameter δ, which controls the shape of the density function by itself; i.e., $\omega = \exp(\delta^{-2})$, $\sqrt{\beta_1} = \sqrt{\omega - 1}(\omega + 2)$, and $\beta_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3$ (Johnson, 1949). Thus, as the shape of the demand's density function deviates from the shape of the normal density function, the factors affecting the inaccuracy in the inventory-target estimation include $\sqrt{\beta_1}$ and β_2.

We conclude this section by noting that we approximate the inaccuracy in the inventory-target estimation for each family of the JTS, while Hayes (1969) does this only for the S_N family. Hayes also quantifies the approximate inaccuracy by obtaining $\mathbb{E}((\hat{Q} - q^*)^2)$ from $\mathrm{Var}(\hat{Q})$, while $\hat{Q} = \bar{r} + z_\varphi s_r$ is not an unbiased estimator of $q^* = -\gamma/\delta + z_\varphi/\delta$. Our approach for inaccuracy approximation differs from that of Hayes in two aspects: (i) Instead of approximating $\mathbb{E}((\hat{Q} - q^*)^2)$ as $\mathrm{Var}(\hat{Q})$, we use $\mathbb{E}(\hat{Q}^2) - 2q^*\mathbb{E}(\hat{Q}) + q^{*2}$ for the quantification of the approximate inaccuracy; (ii) Hayes obtains the inaccuracy $(2 + z_\varphi^2)\phi(z_\varphi)/(4\delta n(1-\varphi))$ for the normal distribution via the use of $\text{Var}(\hat{Q}) = \text{Var}(\bar{r}) + z_\varphi^2 \text{Var}(s_r)$ in which $\text{Var}(\bar{r}) = \delta^{-2}/n$ and $\text{Var}(s_r)$ is approximated by $\delta^{-2}/(2n)$. We, on the other hand, use the exact value of $\text{Var}(s_r)$ to find our closed-form inaccuracy expressions. Consequently, we approximate the inaccuracy in inventory-target estimation more accurately than Hayes (1969) for the normally distributed demand. We investigate the performance of our approximations for the Johnson S_L family in Section 3.5.

3.4 Setting Inventory Targets under HIP

In this section, we seek the optimal bias parameter k^* that leads to the minimum ETOC for the inventory-target estimator $\hat{Q}(k) = r^{-1}(\bar{r} + ks_r)$. For the S_N family of the JTS with $\hat{Q}(k) = \bar{r} + ks_r$, the minimum ETOC is achieved for the bias parameter k^* that minimizes the expression in (3.2). Since the random variable $t\sqrt{n^2/(n^2-1)}$ has the Student's t distribution with n degrees of freedom, k^* is given by $t_{n,\varphi}\sqrt{n^2-1}/n$, where $t_{n,\varphi}$ is the φ^{th} quantile of the Student's t distribution with n degrees of freedom. Therefore, the ETOC-minimizing inventory target is $\bar{r} + t_{n,\varphi}\sqrt{n^2-1}/n s_r$.

For the remaining families of the JTS, however, the loss function cannot be written independent of the MLEs of the demand parameters. This complicates the determination of the optimal value of k that minimizes $\text{ETOC}[\hat{Q}(k)]$. Nevertheless, the optimal bias coefficient k^* is unique under HIP:

Proposition 3.4.1. *The optimal bias coefficient k^* that minimizes* (3.1) *is unique under HIP.*

Proof. We prove the uniqueness of k^* by showing that (i) $\text{ETOC}[\hat{Q}(k)]$ is strictly convex in $\hat{Q}(k)$ and (ii) $\hat{Q}(k)$ is strictly increasing in k: (i) $\text{ETOC}[\hat{Q}(k)]$ can be written as $\int_{\hat{Q}(k)} \bar{L}(\hat{Q}(k);\gamma,\delta) f_{\hat{Q}(k)}(\hat{Q}(k))\, d\hat{Q}(k)$, where $f_{\hat{Q}(k)}$ is the pdf of $\hat{Q}(k)$. ETOC$[\hat{Q}(k)]$ is strictly convex in $Q\hat{(}k)$, because the expected loss function $\bar{L}(\hat{Q}(k);\gamma,\delta)$ is strictly convex in $\hat{Q}(k)$, i.e., $\bar{L}''(\hat{I}(k);\gamma,\delta) = f_Y(\hat{Q}(k))/(1-\varphi) > 0$, and the strict convexity is preserved under non-negative (infinitesimal) linear combinations (Boyd and Vandenberghe, 2004). (ii) Since the inverse transformation function $r^{-1}(\cdot)$ is continuous and the first-order derivative of $r^{-1}(\cdot)$ is positive for all families of the JTS, $\hat{Q}(k)$ is continuous and strictly increasing in k. Therefore, ETOC$[\hat{Q}(k)]$ is increasing in k for $k > k^*$ and decreasing in k for $k < k^*$. Hence, the optimal bias parameter k^* is unique under HIP. $\square$

The existence of a unique value for the optimal bias coefficient allows us to identify the ETOC-minimizing inventory target within a prespecified level of accuracy using a one-dimensional search procedure (Press et al., 2007). Thus, the determination of the optimal bias parameter k^* is a standard, unconstrained minimization problem with a solution that is known to be in a continuous interval $\mathcal{K}$, i.e., $k^* \in \mathcal{K}$ and ETOC$[\hat{Q}(k)] >$ ETOC$[\hat{Q}(k^*)]$ for all $k \in \mathcal{K} \setminus k^*$. Our numerical solution procedure consists of two major steps: (i) The evaluation of ETOC$[\hat{Q}(k)]$ via numerical integration. (ii) The iterative minimization of ETOC$[\hat{Q}(k)]$ to reach an approximation $\hat{k}^*$ of k^* within a predetermined precision. In this step, we let $\mathcal{K} = [0, 2z_\varphi]$ and identify the value of k that minimizes ETOC$[\hat{Q}(k)]$ by using the golden search minimization algorithm with a tolerance parameter of 10^{-3} (Press et al., 2007).

Clearly, k^* depends only on n and φ for the S_N family of the JTS. The independence of k^* from demand's shape parameters continues to hold for the location-scale family of distributions (Katircioglu, 1996). However, in return for an extended flexibility in demand modeling, the optimal bias coefficient k^* becomes dependent on the shape parameters for each of the S_L, S_B, and S_U families of the JTS. Therefore, it is important to enter robust estimates of the Johnson parameters as inputs into the golden search minimization algorithm. We do this by using the MLEs of the shape parameters when there is expert opinion about the bounds of the stochastic demand. If there is not any expert opinion about the bounds of the stochastic demand, then we use the publicly available FITTR1 software of Swain et al. (1988) that fits target distributions from the JTS to independent and identically distributed data using the diagonally weighted least-squares method. It is known that the small-sample properties of the location and scale parameter estimates obtained from this software are superior to those of the MLEs (Biller and Gunes, 2010). Further details of this fitting method can be found in Biller and Nelson (2005).

3.5 Results and Insights

The objective of this section is to investigate the effectiveness of HIP in minimizing the ETOC incurred under MLP with JTS. We experiment with four different distributions from the S_L family of the JTS. Table 3.1 provides their distributional properties (i.e., the shape parameter δ, the coefficient of variation σ/μ, the coefficient of skewness $\sqrt{\beta_1}$, the coefficient of kurtosis β_2, and the 95% quantile of the demand random variable $X_{0.95}$). In particular, distribution S_L I is the most positively skewed with the longest right tail. As δ increases from 1/2 to 5, the shape of the demand's density function approaches the shape of a normal density function (i.e., $\sqrt{\beta_1}$ and β_2 approach 0 and 3, respectively). We adjust the value of parameter γ to achieve a mean demand of 50 units (i.e., $\mu = 50$) for each distribution. Since the shape of the demand's density function depends only on δ for the S_L family of the JTS, the value of parameter γ does not affect the insights about the relation between the shape of the demand's density function and the inventory-target estimation.

We let $n \in \{8, 10, 15, 20, 30, 50\}$ and $\varphi \in \{0.90, 0.95, 0.99\}$ to investigate the effectiveness of HIP as a function of the length of the historical demand data and the asymmetry of the loss function. We provide our results in Tables 3.2, 3.3, and 3.4 for

Table 3.1: Distributional properties of the Johnson S_L distributions with $\xi = 0$, $\lambda = 1$, and mean $\mu = 50$

	δ	σ/μ	$\sqrt{\beta_1}$	β_2	$X_{0.95}$
S_L I	1/2	7.32	414.36	$9.22\ 10^6$	181
S_L II	1	1.31	6.18	113.94	157
S_L III	2	0.53	1.75	8.89	100
S_L IV	5	0.20	0.61	3.67	68

Table 3.2: Inaccuracy (Δ_{S_L}), approximate inaccuracy ($\Delta^a_{S_L}$), ETOC[$\hat{Q}$] under MLP (E), ETOC[$\hat{Q}(k^*)$] under HIP (E*), and CPU time (T) for $\varphi = 0.99$.

	S_L I					S_L II				
n	Δ_{S_L}	$\Delta^a_{S_L}$	E	E*	T	Δ_{S_L}	$\Delta^a_{S_L}$	E	E*	T
8	1367.2	569.4	3177.5	2550.5	0.3	219.0	162.5	630.8	629.1	0.3
10	974.3	503.5	2784.6	2423.3	0.3	170.7	135.0	582.5	581.8	0.3
15	569.6	379.0	2379.9	2238.2	0.3	109.9	94.3	521.7	521.6	0.3
20	403.5	300.7	2213.9	2138.6	0.3	81.0	72.3	492.8	492.8	0.3
30	255.5	211.5	2065.8	2034.2	0.4	53.0	49.2	464.8	464.8	0.5
50	147.6	132.2	1957.9	1947.0	0.6	31.4	30.0	443.2	443.2	0.6
	S_L III					S_L IV				
8	57.1	41.9	176.6	173.2	0.4	14.8	10.0	48.5	45.8	0.3
10	43.6	33.8	163.1	160.7	0.4	11.0	7.9	44.7	42.9	0.3
15	27.2	22.8	146.7	145.5	0.5	6.6	5.2	40.3	39.5	0.4
20	19.6	17.2	139.1	138.4	0.4	4.7	3.9	38.3	37.9	0.4
30	12.6	11.5	132.1	131.8	0.5	2.9	2.6	36.6	36.4	0.5
50	7.3	6.9	126.8	126.7	0.6	1.7	1.6	35.4	35.3	0.6

$\varphi = 0.99$, $\varphi = 0.95$, and $\varphi = 0.90$, respectively. Each of these tables presents the exact inaccuracy in the inventory-target estimation (Δ_{S_L}), the approximate inaccuracy ($\Delta^a_{S_L}$), the ETOC associated with the MLP (ETOC[$\hat{Q}$]) in column E, the ETOC associated with the HIP (ETOC[$\hat{Q}(k^*)$]) in column E*, and the CPU time (in seconds) spent to obtain the ETOC-minimizing inventory target (T). We do not report the value of $\bar{L}(q^*; \gamma, \delta)$ in any of the tables, but it can be easily obtained from the difference ETOC[$\hat{Q}(k)$] $- \Delta_{S_L}$. We code our solution procedures in Wolfram Mathematica 7.0.0 and run the codes in execution mode on an IBM T8300 2.4 Ghz with 2 GB of RAM. Solving the inventory problem of interest requires very little computational effort; i.e., we obtain a solution within a second in each experiment.

Table 3.3: Inaccuracy (Δ_{S_L}), approximate inaccuracy ($\Delta^a_{S_L}$), ETOC[$\hat{Q}$] under MLP (E), ETOC[$\hat{Q}(k^*)$] under HIP (E*), and CPU time (T) for $\varphi = 0.95$.

	S$_L$ I					S$_L$ II				
n	Δ_{S_L}	$\Delta^a_{S_L}$	E	E*	T	Δ_{S_L}	$\Delta^a_{S_L}$	E	E*	T
8	152.0	82.3	740.8	675.2	0.2	49.8	42.7	259.3	256.8	0.2
10	113.0	70.4	701.7	661.1	0.2	39.4	35.1	248.9	247.5	0.3
15	68.8	50.9	657.5	640.1	0.3	26.0	24.1	235.5	234.9	0.3
20	49.5	39.7	638.2	628.5	0.3	19.4	18.3	228.9	228.5	0.3
30	31.7	27.4	620.4	616.2	0.4	12.8	12.4	222.3	222.2	0.4
50	18.4	16.9	607.2	605.6	0.5	7.7	7.5	217.2	217.1	0.5
	S$_L$ III					S$_L$ IV				
8	16.5	14.9	92.6	92.6	0.3	4.8	4.3	29.1	28.9	0.3
10	13.0	12.0	89.1	89.1	0.4	3.7	3.4	28.0	27.9	0.3
15	8.5	8.0	84.6	84.6	0.4	2.4	2.2	26.6	26.6	0.4
20	6.3	6.1	82.4	82.4	0.4	1.8	1.7	26.0	26.0	0.4
30	4.2	4.0	80.3	80.3	0.4	1.1	1.1	25.4	25.4	0.5
50	2.5	2.4	78.6	78.6	0.6	0.7	0.7	24.9	24.9	0.6

Table 3.4: Inaccuracy (Δ_{S_L}), approximate inaccuracy ($\Delta^a_{S_L}$), ETOC[$\hat{Q}$] under MLP (E), ETOC[$\hat{Q}(k^*)$] under HIP (E*), and CPU time (T) for $\varphi = 0.90$.

	S$_L$ I					S$_L$ II				
n	Δ_{S_L}	$\Delta^a_{S_L}$	E	E*	T	Δ_{S_L}	$\Delta^a_{S_L}$	E	E*	T
8	46.3	28.0	378.2	359.3	0.3	22.4	20.1	166.9	165.4	0.3
10	34.9	23.6	366.8	354.7	0.3	17.8	16.4	162.4	161.4	0.3
15	21.6	16.8	353.4	348.1	0.3	11.8	11.2	156.3	155.9	0.3
20	15.6	12.9	347.5	344.5	0.3	8.8	8.5	153.4	153.1	0.3
30	10.0	8.9	341.9	340.6	0.4	5.8	5.7	150.4	150.3	0.5
50	5.8	5.4	337.7	337.2	0.6	3.5	3.5	148.1	148.0	0.6
	S$_L$ III					S$_L$ IV				
8	8.6	8.2	67.2	67.2	0.4	2.7	2.6	22.6	22.6	0.3
10	6.8	6.6	65.5	65.5	0.4	2.1	2.1	22.0	22.0	0.3
15	4.5	4.4	63.1	63.1	0.5	1.4	1.4	21.3	21.3	0.4
20	3.4	3.3	62.0	62.0	0.4	1.0	1.0	20.9	20.9	0.4
30	2.2	2.2	60.9	60.9	0.5	0.7	0.7	20.6	20.5	0.5
50	1.3	1.3	60.0	60.0	0.6	0.4	0.4	20.3	20.3	0.6

Our results suggest a significant amount of inaccuracy in the inventory-target estimation for the highly skewed stochastic demand in the presence of a small amount of historical demand data. Table 3.2 shows that the percent inaccuracy (i.e., $\Delta_{S_L}/(\text{ETOC}[\hat{Q}(k)] - \Delta_{S_L})100\%$) is 75.5% for the S_L I distribution when the historical data set has only 8 observations and the critical fractile is $\varphi = 0.99$. The use of HIP eliminates 45.9% (i.e., $(\text{ETOC}[\hat{Q}] - \text{ETOC}[\hat{Q}(k^*)])/(\text{ETOC}[\hat{Q}] - \bar{L}(q^*;\gamma,\delta))100\%$) of this inaccuracy and achieves a 24.6% (i.e., $(\text{ETOC}[\hat{Q}] - \text{ETOC}[\hat{Q}(k^*)])/\text{ETOC}[\hat{Q}(k^*)]100\%$) reduction in the ETOC. However, the amount of improvement decreases with both n and $1-\varphi$. When $n = 10$, we still improve the ETOC by almost 15%, but by less than 1% and eliminate only 7.4% of the inaccuracy when $n = 50$. Thus, HIP is the most effective in improving the ETOC of an inventory policy with highly skewed stochastic demand when the amount of the historical demand data is small and/or when the shortage of an item costs significantly more than its inventory holding cost. We also observe that the approximate inaccuracy $\Delta^a_{S_L}$ performs very poorly in this case. However, $\Delta^a_{S_L}$ approximates the exact inaccuracy Δ_{S_L} quite well as φ decreases and n increases.

In each table, there is a non-monotonic change in the effectiveness of HIP with respect to the shape parameter δ. For example, when $n = 8$, $\varphi = 0.99$, and the distribution of interest is S_L I in Table 3.2, HIP eliminates 45.9% of the inaccuracy in the inventory-target estimation. These percentages are 0.8%, 6.1%, and 18.2% for the distributions S_L II, S_L III, and S_L IV, respectively. We can explain such a change in the effectiveness of HIP as follows: A small value of δ leads to the optimality of a negatively biased inventory target (i.e., $k^* < z_\varphi$) as has been the case for the exponentially distributed demand in Hayes (1969). A high value of δ, on the other hand, results in a positively biased inventory target (i.e., $k^* > z_\varphi$) as has been observed for the normally distributed demand in Hayes (1969). Since there is a continuum of k^* values for increasing δ, we expect an equality between k^* and z_φ

at a specific value of δ at which the percentage of the inaccuracy eliminated by HIP is zero. Establishing the connection of this observation to $\sqrt{\beta_1}$ and β_2 in the JTS generalizes the findings of Hayes (1969), which have been limited to exponentially and normally distributed demands.

3.6 Implementation Detailes

Section 3.6.1 discusses how to select the appropriate Johnson family. Section 3.6.2 extends the Hayes inventory policy for unknown location and scale parameters.

3.6.1 Selection of the Johnson Family

A question that arises in implementation is which Johnson family to use for modeling the demand. Since the location parameter ξ and the scale parameter λ of the JTS have simple interpretations for the S_L and S_B families, understanding the physical limitations of the underlying stochastic demand and choosing the appropriate Johnson family are often closely related. For the S_L and S_B families, ξ is the lower bound to the values the demand random variable can take. For the S_B family, λ also represents the range, i.e., the difference between the largest and smallest values the demand random variable can assume. Therefore, knowing the lower bound (lower and upper bounds) to the stochastic demand justifies the selection of the S_L (S_B) family of the JTS for demand modeling. Expert opinion about the characteristics of the stochastic demand can also be conveniently incorporated into the demand model and used for the identification of the appropriate Johnson family; e.g., when the pdf of the demand is bimodal (S_B) or positively skewed with a long right tail and a lower bound (S_L). If it is not possible to rely on any knowledge or expertise for selecting a Johnson family, then we can utilize the existing body of work on the determination of the family that best represents the data on hand; see Slifker and Shapiro (1980), Bowman and Shenton (1989), Chou et al. (1998), Chen and Schmeiser (2001), and Niermann (2006).

In this study, we consider a repeated newsvendor setting with a single-product; there are only four different families to consider. Therefore, we implement HIP for each family of the JTS and compare the performances of the resulting inventory targets in the presence of limited historical demand data to decide on which family of the JTS to use for modeling the demand.

3.6.2 Extension of HIP for Unknown ξ and λ

So far, we have assumed that the location parameter ξ and the scale parameter λ are known. We now relax this assumption and discuss the representation of ETOC$[\hat{Q}(k)]$ by also accounting for the uncertainty around the estimates of the parameters ξ and λ. It is important to note that ETOC$[\hat{Q}(k)]$ is still strictly convex in $\hat{Q}(k)$, which is further strictly increasing in k, for all families of the JTS. Therefore, the solution approach in Section 3.4 to find the optimal bias coefficient remains the same.

Johnson's S$_L$ Family

Johnson's S$_L$ distribution with $\lambda = 1$ is well defined only if $\delta > 0$ and $\xi < X_{(1)}$. Therefore, the maximum of the likelihood function at the point $\hat{\xi} = x_{(1)}$ is not achieved. We follow the approach of Biller and Nelson (2005), define the feasible region of the parameter estimation problem as $\xi \leq X_{(1)}$ (i.e., we technically remove $x_{(1)}$ from the support of the distribution), and take the MLE of ξ as $\hat{\xi} = x_{(1)}$. This leads to an inventory-target estimator of the form $\hat{Q}(k) = x_{(1)} + r^{-1}\left(\bar{r} + ks_r\right)$ under HIP. Since the random variable $x_{(1)}$ is independent of $\bar{r}$ and s_r^2, and its sampling density function is given by

$$f_{x_{(1)}}(x_{(1)}) = n\Big(1 - F\left(x_{(1)}; \gamma, \delta, \xi\right)\Big)^{n-1} \frac{\partial F\left(x_{(1)}; \gamma, \delta, \xi\right)}{\partial x_{(1)}},$$

with $x_{(1)} \in (\Omega_L, \infty)$ (Rohatgi and Saleh, 2000), the ETOC associated with setting the inventory target $\hat{Q}(k)$ to $x_{(1)}+r^{-1}\left(\bar{r}+ks_r\right)$ is obtained from $\mathbb{E}_{x_{(1)},\bar{r},s_r^2}(\bar{L}(\hat{Q}(k);\gamma,\delta,\xi))$.

Johnson's S$_B$ Family

We define the feasible region of the parameter estimation problem as $\xi \leq X_{(1)}$ and $\xi + \lambda \geq X_{(n)}$ (i.e., we remove both $x_{(1)}$ and $x_{(n)}$ from the support of the distribution) and take the MLEs of ξ and λ as $x_{(1)}$ and $x_{(n)} - x_{(1)}$. It is well known that the joint density function $f_{x_{(1)},x_{(n)}}(x_{(1)}, x_{(n)})$ of $x_{(1)}$ and $x_{(n)}$ is given by

$$n(n-1)\Big(F\left(x_{(n)};\gamma,\delta,\xi,\lambda\right)-F\left(x_{(1)};\gamma,\delta,\xi,\lambda\right)\Big)^{n-2}\frac{\partial F\left(x_{(1)};\gamma,\delta,\xi,\lambda\right)}{\partial x_{(1)}}\frac{\partial F\left(x_{(n)};\gamma,\delta,\xi,\lambda\right)}{\partial x_{(n)}}$$

with $x_{(1)} \in (\Omega_L, x_{(n)})$ and $x_{(n)} \in (x_{(1)}, \Omega_U)$ (Rohatgi and Saleh, 2000). Furthermore, both $x_{(1)}$ and $x_{(n)}$ are independent of $\bar{r}$ and s_r^2. Thus, the inventory-target estimator under HIP is given by $\hat{Q}(k) = x_{(1)} + (x_{(n)} - x_{(1)})r^{-1}\left(\bar{r} + ks_r\right)$, and ETOC$[\hat{Q}(k)]$ is obtained from $\mathbb{E}_{x_{(1)},x_{(n)},\bar{r},s_r^2}(\bar{L}(\hat{Q}(k);\gamma,\delta,\xi,\lambda))$.

Johnson's S$_U$ Family

The approach we follow for the S$_L$ and S$_B$ families does not extend to the S$_U$ family of the JTS. Although ξ and λ of the S$_U$ family still have no impact on the shape of demand's pdf, the relation of ξ and λ to the position and size of the density function is not simple. Fortunately, the functional form of Johnson's S$_U$ distributions leads to closed-form expressions for the moments of the underlying demand random variable. Consequently, the location and scale parameters of the S$_U$ family are represented by

$$\xi = \mu + \frac{\sqrt{2\ \sigma^2\ \exp(\delta^{-2})}\ \sinh(\gamma/\delta)}{\sqrt{(\exp(\delta^{-2})-1)\,(\exp(\delta^{-2})\ \cosh(2\gamma/\delta)+1)}}$$

and

$$\lambda = \frac{\sqrt{2\ \sigma^2}}{\sqrt{(\exp(\delta^{-2})-1)\,(\exp(\delta^{-2})\ \cosh(2\gamma/\delta)+1)}},$$

where μ and σ^2 are the mean and the variance of the stochastic demand X (Johnson, 1949). Assuming that $x_t, t = 1, 2, \ldots, n$ are independent and identically distributed with finite μ and σ^2, the sample mean $\hat{\mu} = \sum_{t=1}^{n} x_t/n$ and the sample variance $\hat{\sigma}^2 = \sum_{t=1}^{n}(x_t - \hat{\mu})^2/(n-1)$ are the unbiased estimators of μ and σ^2 (i.e., $\mathbb{E}(\hat{\mu}) = \mu$ and $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$) with a joint sampling distribution characterized by Springer (1953) and Bennett (1955). Using the estimator $\bar{r}$ of $-\gamma/\delta$ and the estimator s_r^2 of $1/\delta^2$ provided in Section 3.2 together with $\hat{\mu}$ and $\hat{\sigma}^2$, we estimate ξ and λ as follows:

$$\hat{\xi} = \hat{\mu} + \frac{\sqrt{2\ \hat{\sigma}^2\ \exp(s_r^2)}\ \sinh(-\bar{r})}{\sqrt{(\exp(s_r^2)-1)\,(\exp(s_r^2)\ \cosh(-2\bar{r})+1)}}$$

$$\hat{\lambda} = \frac{\sqrt{2\ \hat{\sigma}^2}}{\sqrt{(\exp(s_r^2)-1)\,(\exp(s_r^2)\ \cosh(-2\bar{r})+1)}}.$$

Thus, the sampling distributions of $\hat{\xi}$ and $\hat{\lambda}$ are obtained from the joint sampling distribution of $\hat{\mu}$, $\hat{\sigma}^2$, $\bar{r}$, and s_r^2. Furthermore, the inventory-target estimator under HIP is given by $\hat{Q}(k) = \hat{\xi} + \hat{\lambda}\ r^{-1}(\bar{r} + ks_r)$ and ETOC$[\hat{Q}(k)]$ is obtained from $\mathbb{E}_{\hat{\xi},\hat{\lambda},\bar{r},s_r^2}(\bar{L}(\hat{Q}(k); \gamma, \delta, \xi, \lambda))$.

3.7 HIP for Service-Level Constraints

So far, we have focused on the minimization of the ETOC function written in terms of the unit inventory holding cost and the unit shortage cost. However, it is often difficult for the inventory manager to choose a value for the unit shortage cost that includes intangible components such as the loss of goodwill (Nahmias, 2005). Motivated by this difficulty, we describe how to use HIP for setting inventory targets subject to service-level constraints. Specifically, we consider a Type 1 service-level criterion in Section 3.7.1 and a Type 2 service-level criterion in Section 3.7.2.

3.7.1 Type 1 Service-Level Constraint

We define the Type 1 service objective as setting the inventory target to a value that fully satisfies the demand of the forthcoming period with an average probability of α. Under complete certainty about the demand cdf $F(\cdot;\gamma,\delta)$, the optimal inventory target q_1^*, which attains the Type 1 service objective, is unique and satisfies $\Phi(\gamma + \delta r(q_1^*)) = \alpha$. Thus, $q_1^* = r^{-1}(-\gamma/\delta + z_\alpha/\delta)$, where $z_\alpha = \Phi^{-1}(\alpha)$ is the safety factor associated with the Type 1 service objective. Estimating $-\gamma/\delta$ by $\bar{r}$ and $1/\delta$ by s_r leads to the inventory-target estimator $\hat{Q}_1 = r^{-1}(\bar{r} + z_\alpha s_r)$ under MLP with JTS. The use of this inventory-target estimator results in an average Type 1 service level of $\mathbb{E}_{\bar{r},s_r^2}(\Phi\left(\gamma + \delta\left(\bar{r} + z_\alpha s_r\right)\right))$, which is not necessarily equal to α. Hence, the inventory-target estimator $\hat{Q}_1$ may not attain the average Type 1 service level α.

It is possible to achieve the target Type 1 service level in the long run by adjusting the bias in the inventory-target estimator $\hat{Q}_1$. Specifically, we replace the safety factor z_α with the bias coefficient k_1, which satisfies $\alpha = \mathbb{E}_{\bar{r},s_r^2}(\mathbb{P}(Z < \gamma + \delta(\bar{r} + k_1 s_r))) = \mathbb{P}(t < k_1)$, where $t = ((Z-\gamma)/\delta - \bar{r})/s_r$. Thus, the optimal bias parameter k_1^* is identified as $t_{n-1,\alpha}\sqrt{(n+1)/n}$, where $t_{n-1,\alpha}$ is the αth quantile of the Student's t random variable with $n-1$ degrees of freedom.

3.7.2 Type 2 Service-Level Constraint

We define the Type 2 service objective as the determination of the inventory target that sets the average proportion of demand satisfied immediately from stock to β. Under complete certainty about the demand distribution $F(\cdot;\gamma,\delta)$, the optimal inventory target q_2^*, which attains the average Type 2 service level of β, is uniquely characterized by $r^{-1}(-\gamma/\delta + z_\beta/\delta)$, where the safety factor z_β satisfies

$$\int_{r^{-1}\left(-\gamma/\delta+z_\beta/\delta\right)}^{\Omega_U} \Big(y - r^{-1}\left(-\gamma/\delta + z_\beta/\delta\right)\Big)\delta\,\frac{\partial r(y)}{\partial y}\,\phi\Big(\gamma + \delta r(y)\Big)\mathrm{d}y = (1-\beta)\,\mu.$$

Estimating $-\gamma/\delta$ by $\bar{r}$ and $1/\delta$ by s_r leads to the inventory-target estimator $\hat{Q}_2 = r^{-1}(\bar{r}+z_\beta s_r)$ under MLP with JTS. However, the use of this inventory-target estimator results in an average Type 2 service level, which is not necessarily equal to β. Hence, the inventory-target estimator $\hat{Q}_2$ may not attain the average Type 2 service level.

To achieve the target Type 2 service level in the long run, HIP suggests the use of an inventory-target estimator of the form $\hat{Q}_2(k_2^*) = r^{-1}(\bar{r} + k_2^* s_r)$, where k_2^* is the optimal bias coefficient satisfying

$$\mathbb{E}_{\bar{r},{s_r}^2}\left(\int_{r^{-1}\left(\bar{r}+k_2^* s_r\right)}^{\Omega_U} \Big(y - r^{-1}\left(\bar{r} + k_2^* s_r\right)\Big)\delta\,\frac{\partial r(y)}{\partial y}\,\phi\Big(\gamma + \delta r\left(y\right)\Big)\mathrm{d}y\right) = (1-\beta)\mu. \tag{3.3}$$

Since $\hat{Q}_2(k_2)$ is strictly increasing in k_2 and the integrand of the representation in (3.3) is decreasing in k_2, the unique optimal bias parameter k_2^* can be easily identified by a one-dimensional search algorithm, which sets the left-hand side of (3.3) to its right-hand side for a given value of β.

3.8 Conclusion

In this chapter, we study the problem of estimating an inventory target from limited historical demand data in a repeated newsvendor setting. We quantify the inaccuracy in the inventory-target estimation under maximum likelihood policy (MLP) as a function of the length of the historical demand data, the critical fractile, and the shape parameters of the demand distribution. We suggest the use of Hayes inventory policy (HIP), instead of MLP, for setting inventory targets in the presence of this inaccuracy. The distinguishing feature of HIP is the joint use of the highly flexible JTS for demand modeling and the ETOC criterion for capturing the uncertainty around the MLEs of the demand parameters. Our solution procedure allows the decision maker to obtain improved inventory targets without making any assumptions about the first four moments of the demand random variable. Our numerical analysis identifies an inaccuracy of 54% in the inventory-target estimation under MLP for a highly skewed stochastic demand when the critical fractile is 0.99 and there are only 10 demand observations available. The management of inventory via HIP leads to a reduction of 15% in the expected total operating cost, while eliminating 37% of the inaccuracy. We also extend the use of HIP for managing inventory subject to Type 1 and Type 2 service-level constraints.

Our study is the first to use JTS for demand modeling in inventory management. The versatility offered by the JTS provides an opportunity to capture a wide variety of demand characteristics. However, this discussion does not extend to slow moving items with intermittent demand; JTS fails to assign positive probability to the event of observing no demand in the forthcoming period. The inventory-target estimation problem in the presence of intermittent demand data is the subject of the next chapter.

Chapter 4

Managing Inventory with Limited History of Intermittent Demand

4.1 Introduction

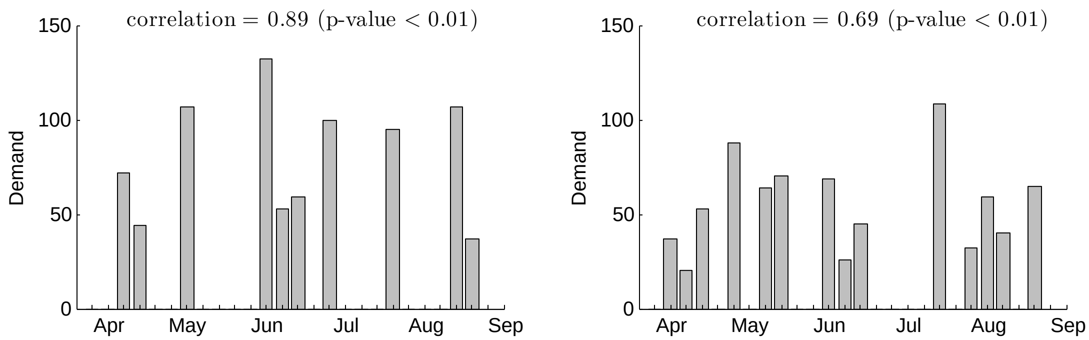
A common challenge faced by many businesses is linking inventory management and historical demand data. The routine practice is to model stochastic demand using probability distributions and then estimate the parameters of these distributions to compute inventory policies – which are known to be optimal only when the true parameter values are used. Therefore, the sequential process of first estimating the parameters and then treating the parameter estimates as if they were the true values casts doubt on the performance of the "optimal" inventory policy. This is clearly not an issue when there is a large amount of demand data and the parameter estimates converge to their true values. In practice, however, the demand history available to support operational decisions is often very short – mainly because the underlying demand generating process does not remain constant indefinitely, and, even if there is a long demand history, it is common to consider only the most recent observations. The discrepancy between the performance of an optimal (but unknown) inventory target and the performance of its estimate from a short demand history is a serious, but often ignored, operational problem.

We address this problem by considering parameter estimation and inventory optimization as a single task, as opposed to the usual paradigm of estimating the parameters first, and then solving an optimization problem based on the estimates. In recent years, this idea has been instrumental in setting inventory targets in the presence of historical data of finite length; see Liyanage and Shanthikumar (2005), Chu et al. (2008), Akcay et al. (2011), and Ramamurthy et al. (2012) for example studies. We also manage inventory by following this principle. However, we do this when the stationary demand process is intermittent.

Intermittent demand is a series of positive realizations that only appear at random periods. It constitutes a challenge in inventory management due to the dual source of variation; i.e., both the demand size and the number of inter-demand periods – the zero-demand periods between two (positive) demands – are uncertain. Intermittent demand may arise for a number of reasons. For example, a smooth demand series at a quarterly or monthly or weekly aggregation can become intermittent at a monthly or weekly or daily bucket. Similarly, the upstream member in a supply chain may observe intermittent demand based on the replenishment policies of downstream members such as order batching and not ordering until the inventory level drops below a certain level.

There is a type of dependence specific to intermittent demand presenting a further challenge: correlation between the size of a positive demand and the number of zero-demand periods preceding the demand. Positive correlation occurs when a long (short) inter-demand period is followed by a high (low) demand size. Negative correlation indicates the opposite relationship between the demand size and the number of inter-demand periods. Willemain et al. (1994) show the existence of correlation in the historical shipment data of items ranging from electrical equipment to health

Figure 4.1: The intermittent histories with demand sizes that are correlated with number of inter-demand periods



products. Eaves (2002) investigates spare parts data from the Royal Air Force and finds statistically significant correlation in 18% of more than ten thousand items. It is concluded that positive correlations are more frequent than negative correlations. More recently, Altay et al. (2012) analyze aircraft service parts data from the U.S. Defense Logistics Agency and identify significant correlation in 35% of the items.

Our experience is as follows. In a data set from a global luxury manufacturer in consumer-product space, the percentage of zero-demand periods varies between 32% and 68% over a six-month period from April to September in 2010. A two-sided t test shows that 292 items (out of 1149) have statistically significant correlation between the demand size and the number of inter-demand periods. We observe that 81% of the 292 items have positive correlation, which can be as high as 0.89. Figure 4.1 illustrates the demand histories for two of these items. We also find that the negative correlation can be as low as -0.83, while the average value of the correlation is 0.49.

We introduce a flexible *copula-based* demand model to capture the correlation between the demand size and the number of inter-demand periods. This flexibility comes with additional parameters to be estimated from historical demand data. Consequently, the discrepancy between the performance of the optimal inventory target and the performance of its estimate becomes a more serious operational problem especially when the demand history is short. In this study, we introduce two finite-sample hypothesis tests to investigate the existence of correlation in a short demand history. If the test suggests sufficient evidence for the existence of correlation, then we incorporate the dependence between the demand size and the number of inter-demand periods into the estimation of the inventory target.

Next, we illustrate the role of statistical estimation errors on the performance of an inventory-target estimate in the simplest possible setting with intermittent demand. We consider the case of deterministic demand size with uncertain demand-arrival periods in the presence of no correlation.

4.1.1 Role of Parameter Uncertainty: An Example with Deterministic Demand

We let the demand size be a fixed quantity Q, but it is unknown when the demand will arrive. The decision maker models the intermittent nature of demand as a Bernoulli process; i.e., either Q units of demand arrives with probability p or zero demand arrives with probability $1 - p$. The comparison of the expected costs at inventory targets between Q and zero shows that the optimal policy – which minimizes the sum of expected holding and shortage costs – has a simple form: Do not keep any inventory if $p < h/(h + b)$, and stock Q units if $p > h/(h + b)$, where h is the unit holding cost and b is the unit shortage cost. If p is equal to $h/(h+b)$, then the decision maker is indifferent between stocking Q and zero units (and any value between Q and zero) as they all result in the same expected cost.

The value of the demand-occurrence probability p is one important piece of information missing to implement the optimal policy in practice, and it has to be estimated from historical data. A way of estimating p is to use the ratio of the number of positive-demand periods to the total number of periods in the demand history; i.e., the sample mean of the Bernoulli process. We denote this estimator of the demand-occurrence probability by $\hat{p}$, which is not only unbiased and consistent, but also achieves the Cramér-Rao lower bound when the sample size tends to infinity. However, the decision maker is not fortunate enough to count on the unbiasedness and large-sample properties of $\hat{p}$ when the historical data is of limited length.

We illustrate this situation with a numerical example: The demand realizations are recorded in the last eleven periods, $h = \$1$, $b = \$9$, $Q = 100$, and the *unknown* value of p is 0.2. Since $p > 0.1$ (i.e., $\$1/(\$1 + \$9)$), it is optimal to order 100 units with an expected cost of \$80. The decision maker implements the "optimal" policy by replacing unknown p with $\hat{p}$. That is, she sets the inventory target to the optimal 100 units only when $\hat{p} > 0.1$. This condition is satisfied only if the nature generates two or more positive realizations in eleven periods – a random event that occurs with 68% probability. For the remaining 32% of the time, the decision maker incorrectly sets the inventory target to zero and bears the *a priori* difference in the expected costs at inventory targets of zero and 100 units. Since the expected cost associated with ordering zero units is \$180, we identify the expected cost of incorrectly estimating p as \$32 (i.e., 0.32 (\$180 – \$80)).

The selection of the threshold as $h/(h + b)$, albeit optimal with known p, ignores the impact of the uncertainty around $\hat{p}$ (i.e., the estimation error) on the expected cost. On the other hand, an alternative threshold which factors the estimation error into inventory-target estimation can reduce the a priori expected cost of incorrect estimation. As an example, suppose that the threshold is 0.01. Then, the decision maker sets the inventory target to the optimal 100 units only when $\hat{p} > 0.01$. This condition is satisfied only if the nature generates one or more positive realizations in eleven periods – a random event with 91.5% probability of occurrence. That is, the decision maker incorrectly sets the inventory target to zero for the remaining 8.5% of the time. Consequently, the selection of 0.01 as an *improved* threshold reduces the probability of incorrectly estimating the optimal inventory target from 32% to 8.5% and the expected cost of incorrect estimation from \$32 to \$8.5 (i.e., 0.085(\$180−\$80)).

4.1.2 Contributions and Outline

Croston's method (Croston, 1972) – a variant of exponential smoothing – is widely used to estimate the mean demand per period from an intermittent demand history with no correlation between demand size and the number of zero-demand periods preceding the demand. Most of the research on intermittent-demand forecasting builds on Croston's method with the following objectives: (i) Improving the statistical properties of the mean demand estimates (Syntetos and Boylan, 2001, 2005); (ii) assessing the variance of the forecast error (Syntetos and Boylan, 2010); and (iii) understanding the interaction between intermittent-demand forecasting and inventory control (Teunter and Sani, 2009; Snyder et al., 2012). Unlike these research streams, we do not forecast the demand but directly estimate the inventory target (as in Hayes 1969, Liyanage and Shanthikumar 2005, and Ramamurthy et al. 2012) in a decision theoretical framework by minimizing the expected cost of incorrect estimation.

More specifically, we summarize our contributions in this study as follows:

1. The implementation of the optimal inventory policy by using the point estimates of the demand-occurrence probability and the demand mean and variance results in an additional cost. We quantify the expected value of this cost – the expected cost of parameter uncertainty – as a function of a threshold variable when the distribution of the positive demand size belongs to a location-scale family for an item with intermittent demand.

2. We illustrate the derivation of the expected cost of parameter uncertainty for exponentially and normally distributed demand sizes.

3. We propose an inventory-target estimation rule – the improved inventory-target estimation – that minimizes the expected cost of parameter uncertainty with respect to the threshold variable. Our approach, thus, combines inventory management and parameter estimation into a single task, and effectively balances under- and overestimation of the optimal inventory target. In certain cases, the use of an optimized threshold eliminates the expected cost of parameter uncertainty completely.

4. Motivated by our analysis of industrial data and the previous literature on empirical analysis of intermittent demand, we present a copula-based demand model to capture the correlation between demand size and the number of zero-demand periods that precede the demand. We characterize the optimal policy when there is no parameter uncertainty.

5. We develop two new hypothesis tests to assess the existence of correlation in limited amount of intermittent demand data. The copula-based representation of intermittent demand allows us to obtain the distribution of the test statistics by sampling from uniform random variables. We find that the test which considers the expected cost of parameter uncertainty tends to reject the independence assumption less frequently than the test which only considers the sampling distribution of the copula-parameter estimator.

The remainder of the chapter is organized as follows. Section 4.2 presents the demand and inventory models and discusses the derivation of the expected cost due to parameter uncertainty. Section 4.3 discusses the improved inventory-target estimation. Section 4.4 presents our copula-based demand model and develops the hypothesis tests to assess the existence of correlation in intermittent demand data. Section 4.5 concludes with a summary of findings and future research directions.

4.2 Statistical Inventory Management with Intermittent Demand

We start our study of managing inventory with limited history of intermittent demand under the assumption of no correlation between the demand size and the number of inter-demand periods in Section 4.2.1. We consider this special case first as it will be the building block of our copula-based demand model in Section 4.4, where we also use it to obtain the test-statistic distribution under the null hypothesis of no correlation in the operational test. Section 4.2.2 presents the inventory model and identifies the optimal policy for known parameters. Section 4.2.3 discusses the practice of implementing the optimal policy by naively treating the parameter estimates as if they were the true values. Section 4.2.4 quantifies the expected cost of parameter uncertainty under this practice, and Section 4.2.5 illustrates this quantification when the demand size is represented with exponential and normal distributions.

4.2.1 Demand Model

We consider a discrete-time model with inventory review periods that are often shorter than the times between successive demand arrivals. Therefore, there are periods in which no demand is received. We model the randomness in demand arrivals by a Bernoulli process; i.e., the number of inter-demand periods, which we denote by $Y \in \{0, 1, 2, \ldots\}$, has the geometric distribution with cumulative distribution function (cdf) $G(y; p) = 1 - (1 - p)^{y+1}$. Thus, the probability of observing a positive demand in any period is equal to p. We let the distribution of (positive) demand size, denoted by X, be a member of the location-scale family of distributions with location parameter τ and scale parameter θ. Therefore, the cdf of X, $F(x; \tau, \theta)$ can be written as $F((x - \tau)/\theta; 0, 1)$, where $F(\cdot; 0, 1)$ is the standardized cdf that does not depend on τ and θ. We let $Z := (X - \tau)/\theta$ be the standardized demand-size random variable with expectation a_1 and variance a_2. For example, $a_1 = a_2 = 1$ and $\tau = 0$ for exponential distribution and $a_1 = 0$ and $a_2 = 1$ for normal distribution. Other members of the location-scale family include gamma and Weibull distributions with fixed shape parameters as well as Cauchy, uniform, logistics, Student's t, and Laplace distributions.

4.2.2 Inventory Model

The decision of how much inventory to keep, if any, is made at the beginning of each period and is contingent on the amount of inventory on hand. We consider linear procurement, holding, and backlogging costs c, h, and b per unit, respectively, and the time lag between procurement and delivery is negligible. Holding and backlogging costs are calculated based on the amount of ending inventory in each period. All the backlogged demand is satisfied before the next period starts, and the decision maker cannot dispose any inventory during the multi-period planning horizon. At the end of the finite planning horizon, the decision maker obtains reimbursement of the procurement cost for each leftover unit and incurs the procurement cost for each backlogged unit. The goal is to find an ordering policy that minimizes the overall expected cost. Without loss of generality, we take the per-unit ordering cost c as zero, and represent the single-period expected cost associated with a nonnegative inventory target q as follows:

$$C(q; p, \tau, \theta) := (1-p)hq + p\left(h\int (q-x)^+ \mathrm{d}F(x;\tau,\theta) + b\int (x-q)^+ \mathrm{d}F(x;\tau,\theta)\right).$$

In the setting described above, the base-stock policy is optimal with complete knowledge of the demand distribution (Porteus, 2002). Furthermore, the optimal inventory target has a simple form:

Proposition 4.2.1. *Let γ_o denote $h/(h+b)$. For $p > \gamma_o$, the optimal inventory target q^* is given by $\tau + \eta(p, \gamma_o)\theta$, where $\eta(p, \gamma_o)$ is equal to $F^{-1}(1 - \gamma_o/p; 0, 1)$. For $p \leq \gamma_o$, q^* is zero and it is optimal not to carry any inventory.*

Proof. The inventory level before ordering, denoted by ν, is the state of the system. The optimality equation is given by

$$V_t(\nu) = \min_{q \geq \max(0,\nu)} \{C(q; p, \tau, \theta) + (1 - p)V_{t+1}(q) + p\,\mathbb{E}(V_{t+1}(q - X))\}$$

for $t = 1, 2, \ldots, N$ and $V_{N+1}(\nu) = 0$ with N the number of periods in the planning horizon. Let $\mathscr{L}_t(q)$ denote the function $C(q; p, \tau, \theta) + (1-p)V_{t+1}(q) + p\mathbb{E}(V_{t+1}(q-X))$.

There are two cases to analyze: (i) If $p > \gamma_o$, then the minimizer of the convex function $C(q; p, \tau, \theta)$ is characterized by the (interior) first-order condition $(1 - p)h + p((h + b)F(q; \tau, \theta) - b) = 0$. Then, the optimal base-stock level q^* has an explicit form given by $\tau + F^{-1}(1 - \gamma_o/p; 0, 1)\theta$. The proof is standard, and we refer the reader to page 70 in Porteus (2002) for details. (ii) If $p \leq \gamma_o$, then the minimizer of the convex function $C(q; p, \tau, \theta)$, and hence, $\mathscr{L}_N(q)$, is zero (i.e., the boundary solution). Therefore, $V_N(\nu)$ is given by $\mathscr{L}_N(0)$ if $\nu \leq 0$ and $\mathscr{L}_N(\nu)$ if $\nu > 0$. The function $\mathscr{L}_{N-1}(q) = C(q; p, \tau, \theta) + (1 - p)V_N(q) + p\mathbb{E}(V_N(q - X))$ is also minimized by the boundary solution since it is a weighted sum of functions, which take their minimums at zero. Consequently, $V_{N-1}(\nu)$ is given by $\mathscr{L}_{N-1}(0)$ if $\nu \leq 0$ and $\mathscr{L}_{N-1}(\nu)$ if $\nu > 0$. The result follows from a recursive argument through the periods $t = N - 1, N - 2, \ldots, 1$. □

The decision maker obtains the optimal inventory target q^* for $p > h/(h + b)$ by minimizing the single-period expected cost knowing that the *myopic* policy is optimal. To implement the policy, however, she obtains the *point estimates* of the unknown parameters p, τ, and θ from the historical demand data, and plugs these estimates into the functional form of q^* which is optimal only when there is no parameter uncertainty. We name this practice as naive inventory-target estimation because it ignores the uncertainty around the parameter estimates which are – inevitably – obtained from a *single* realization of the demand history.

4.2.3 Naive Inventory-Target Estimation

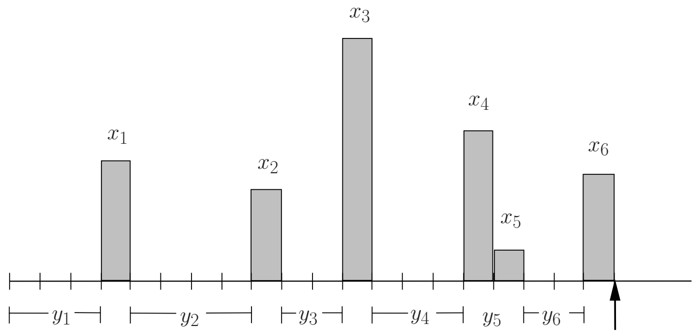
We let $\{(x_i, y_i); i = 1, 2, \ldots, n_d\}$ denote the demand history, where x_i is the size of ith demand, y_i is the number of zero-demand periods preceding x_i, and n_d is the number of periods with demand. Use of a base-stock policy implies ordering just after a period with a demand arrival. Therefore, the number of inter-demand periods is zero at the time of ordering, which is the beginning of period $n + 1$, where n is $n_d + \sum_{i=1}^{n_d} y_i$. We illustrate this intermittent demand history in Figure 4.2.

One practical approach is to estimate the expected (positive) demand size $\mathbb{E}(X)$ by $\bar{x} := \sum_{i=1}^{n_d} x_i/n_d$, the demand-size variance by $s^2 := \sum_{i=1}^{n_d}(x_i - \bar{x})^2/(n_d - 1)$, and the expected number of inter-demand periods $\mathbb{E}(Y)$ by $\sum_{i=1}^{n_d} y_i/n_d$. If one of the parameters τ and θ is known (as is the case for the exponentially distributed demand size), only $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ are estimated. These estimators are unbiased and easy to calculate, and therefore, immediate choices for the decision maker. Consequently, the decision maker implements the policy in Proposition 4.2.1 with the parameter estimates $\hat{\theta} = s/\sqrt{a_2}$, $\hat{\tau} = \bar{x} - a_1\hat{\theta}$, and $\hat{p} = n_d/n$ before the ordering decision is made at the beginning of period $n+1$. We let $\hat{Q}(\hat{p}, \gamma_o)$ denote the resulting inventory-target estimator.

Definition 4.2.1 (Naive Inventory-Target Estimation)**.** *If $\hat{p} > \gamma_o$, then the naive inventory-target estimator $\hat{Q}(\hat{p}, \gamma_o)$ is given by $\hat{\tau} + \eta(\hat{p}, \gamma_o)\hat{\theta}$. If $\hat{p} \leq \gamma_o$, then $\hat{Q}(\hat{p}, \gamma_o)$ is zero.*

As demand observations accumulate over time, there is less uncertainty around the point estimates of the parameters, and the naive inventory-target estimation starts

Figure 4.2: The intermittent demand data $(x_1, y_1) = (40, 3)$, $(x_2, y_2) = (30, 4)$, $(x_3, y_3) = (80, 2)$, $(x_4, y_4) = (50, 3)$, $(x_5, y_5) = (10, 0)$, and $(x_6, y_6) = (35, 2)$ with $n_d = 6$ observed by the end of period $n = 20$, marked by the arrow ↑



working as the optimal policy with no parameter uncertainty. The question is how to manage inventory in the presence of a limited amount of demand data as this knowledge is accumulated.

4.2.4 Quantification of the Expected Cost of Parameter Uncertainty

We assume that the decision maker has access to an intermittent demand history of length n. The objective is to quantify the expected cost of incorrectly estimating the parameters p, τ, and θ in period $n + 1$ under the naive inventory-target estimation. The difference between the expected costs of the naive inventory-target estimator and the optimal inventory target (i.e., $C(\hat{Q}(\hat{p}, \gamma_o); p, \tau, \theta) - C(q^*; p, \tau, \theta)$) is a nonnegative random variable because a different demand history from the *true* data generating process could have led to a different value for $\hat{Q}(\hat{p}, \gamma_o)$. Our objective is to calculate the expected value of this difference, which we interpret as the expected cost of parameter uncertainty.

We start with noting that the demand-occurrence probability estimator $\hat{p}$ takes different values depending on the number of positive realizations in an intermittent demand history:

Lemma 4.2.1. *When the length of the historical demand data is* n*,* $\hat{p}$ *takes the value of* $\omega \in \{0, 1/n, 2/n, \ldots, 1\}$ *with probability*

$$\mathbb{P}(\hat{p} = \omega) = \frac{n!}{(n\omega)!(n - n\omega)!} p^{n\omega}(1 - p)^{n - n\omega}.$$

Proof. The estimator of the demand-occurrence probability, $\hat{p} = n_d/n$, is the sample mean of n independent Bernoulli random variables, each with a success probability of p. Equivalently, $n\hat{p}$ is a binomial random variable with parameters n and p. The probability mass function of $\hat{p}$ is, therefore, obtained from the transformation of the binomial random variable $n\hat{p}$ to $\hat{p}$. □

Naive inventory-target estimation compares the demand-occurrence probability estimate $\hat{p}$ with threshold γ_o, and sets the inventory target as described in Definition 4.2.1. On the other hand, we know from the simple numerical example in Section 4.1.1 that the use of an alternative threshold might improve the operational performance when a decision is based on a limited amount of historical data. Therefore, we quantify the expected cost of parameter uncertainty for an arbitrary threshold $\gamma \in [0, 1]$ in search of an inventory-target estimation rule $\hat{Q}(\hat{p}, \gamma)$ that is better than the naive inventory-target estimation rule $\hat{Q}(\hat{p}, \gamma_o)$. Section 4.3 will discuss the minimization of the expected cost of parameter uncertainty with respect to γ. In the remainder of this section, we focus on quantifying the expected cost of parameter uncertainty as a function of γ.

There are four cases to consider: (1) If $p \leq \gamma_o$ and $\hat{p} \leq \gamma$, then both the optimal inventory target and its estimate are zero, and there is no expected cost due to parameter uncertainty. (2) If $p \leq \gamma_o$ and $\hat{p} > \gamma$, then the optimal inventory target is zero while its estimate is positive, resulting in the additional expected cost of $C(\hat{Q}(\hat{p}, \gamma); p, \tau, \theta) - C(0; p, \tau, \theta)$ due to parameter uncertainty. (3) If $p > \gamma_o$ and $\hat{p} \leq \gamma$, then the optimal inventory target is positive while its estimate is zero;

thus, the additional expected cost is $C(0;p,\tau,\theta)-C(q^*;p,\tau,\theta)$. (4) If $p>\gamma_o$ and $\hat{p}>\gamma$, then both the optimal inventory target and its estimate are positive with $C(\hat{Q}(\hat{p},\gamma);p,\tau,\theta)-C(q^*;p,\tau,\theta)$ as the additional expected cost. Consequently, the expected cost of incorrectly estimating p, averaged over all possible realizations of $\hat{p}$, is given by

$$\sum_{w>\gamma}\Big(C(\hat{Q}(w,\gamma);p,\tau,\theta)-C(0;p,\tau,\theta)\Big)\,\mathbb{P}\left(\hat{p}=w\right)$$

for $p\le\gamma_o$; and,

$$\sum_{w\le\gamma}C(0;p,\tau,\theta)\,\mathbb{P}\left(\hat{p}=w\right)+\sum_{w>\gamma}C(\hat{Q}(w,\gamma);p,\tau,\theta)\,\mathbb{P}\left(\hat{p}=w\right)-C(q^*;p,\tau,\theta)$$

for $p>\gamma_o$. For ease in exposition, we denote this expected cost of incorrectly estimating p with $\Delta_{\hat{\tau},\hat{\theta}}(\gamma;p,\tau,\theta)$. Since the estimators $\hat{\tau}$ and $\hat{\theta}$ obtained from nw demand random variables factor into $\hat{Q}(w,\gamma)$, $\Delta_{\hat{\tau},\hat{\theta}}(\gamma;p,\tau,\theta)$ is a random variable. This leads to the interpretation that $\mathbb{E}(\Delta_{\hat{\tau},\hat{\theta}}(\gamma;p,\tau,\theta))$ is the expected cost due to parameter uncertainty averaged over *all* possible realizations of an intermittent demand history of length n.

Remark 4.2.1. $\mathbb{E}(\Delta_{\hat{\tau},\hat{\theta}}(\gamma_o;p,\tau,\theta))$ *is the expected cost of incorrectly estimating the optimal inventory target under naive inventory-target estimation. Since* $\hat{p}$*,* $\hat{\tau}$*, and* $\hat{\theta}$ *are consistent estimators (Rohatgi and Saleh, 2000),* $\mathbb{E}(\Delta_{\hat{\tau},\hat{\theta}}(\gamma_o;p,\tau,\theta))$ *approaches zero as the length of the demand history increases. In this study, we focus on its magnitude when there is a limited amount of historical demand data for parameter estimation.*

We now introduce two new random variables $\mathcal{U}:=(\hat{\tau}-\tau)/\theta$ and $\mathcal{V}:=\hat{\theta}/\theta$, and reduce the single-period expected cost function to an alternative form:

Lemma 4.2.2. *(i) The distribution of random variables* $\mathcal{U}$ *and* $\mathcal{V}$ *are independent of the parameters* τ *and* θ*. (ii) The expected cost function* $C(q;p,\tau,\theta)$ *can be equivalently* *written as*

$$\theta\, C\left(\frac{q-\tau}{\theta}; p, 0, 1\right) + (1-p)\, h\, \tau.$$

Proof. (i) We let $z_i = (x_i - \tau)/\theta$ and $\bar{z} = (1/n)\sum_{i=1}^{n} z_i$. The distribution functions of z_i and $\bar{z}$ are both independent of τ and θ for any positive integer n. We also make the following observations:

$$\mathbb{P}\left(\frac{\bar{x}-\tau}{\theta} \le \kappa\right) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\tau}{\theta} \le \kappa\right) = \mathbb{P}(\bar{z} \le \kappa)$$

$$\begin{aligned}
\mathbb{P}\left(\frac{\hat{\theta}}{\theta} \le \kappa\right) &= \mathbb{P}\left(\frac{1}{\theta}\sqrt{\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}} \le \kappa\sqrt{a_2}\right) \\
&= \mathbb{P}\left(\sqrt{\sum_{i=1}^{n}\left(\frac{x_i-\tau}{\theta} - \frac{\bar{x}-\beta}{\theta}\right)^2} \le \kappa\sqrt{a_2(n-1)}\right) \\
&= \mathbb{P}\left(\sqrt{\sum_{i=1}^{n}(z_i-\bar{z})^2} \le \kappa\sqrt{a_2(n-1)}\right)
\end{aligned}$$

Thus, the distributions of $(\bar{x}-\tau)/\theta$ and $\mathcal{V} = \hat{\theta}/\theta$ do not depend on τ and θ. Plugging $\hat{\tau} = \bar{x} - a_1\hat{\theta}$ into $\mathcal{U} = (\hat{\tau}-\tau)/\theta$, we see that $\mathcal{U}$ can be written as $(\bar{x}-\tau)/\theta - a_1\mathcal{V}$, and hence, it is also independent of τ and θ. (ii) As a result of transforming X to $\tau + Z\theta$, the expected cost function $C(q; p, \tau, \theta)$ can be rewritten as follows:

$$\begin{aligned}
C(q; p, \tau, \theta) &= (1-p)hq + p\left(h\,\mathbb{E}(q-X)^+ + b\,\mathbb{E}(X-q)^+\right) \\
&= (1-p)h\tau + \theta(1-p)h\left(\frac{q-\tau}{\theta}\right) \\
&\quad + \theta p\left(h\,\mathbb{E}\left(\frac{q-\tau}{\theta} - Z\right)^+ + b\,\mathbb{E}\left(Z - \frac{q-\tau}{\theta}\right)^+\right) \\
&= (1-p)h\tau + \theta C\left(\frac{q-\tau}{\theta}; p, 0, 1\right) \quad \square
\end{aligned}$$

Lemma 4.2.2 leads to the quantification of the expected cost of parameter uncertainty as a linear function of the scale parameter θ and independent of the location parameter τ. This result will play a key role in Section 4.3 as it allows us to minimize the expected cost of parameter uncertainty over the threshold variable γ uniformly for all values of τ and θ.

Proposition 4.2.2. *The expected cost of incorrectly estimating the unknown parameters p, τ, and θ, which is given by $\mathbb{E}\left(\Delta_{\hat{\tau},\hat{\theta}}(\gamma; p, \tau, \theta)\right)$, can be written as $\theta\,\mathbb{E}\left(\Delta_{\mathcal{U},\mathcal{V}}(\gamma; p, 0, 1)\right)$, where $\Delta_{\mathcal{U},\mathcal{V}}(\cdot; p, 0, 1)$ is a function of the random variables $\mathcal{U}$ and $\mathcal{V}$, and the demand-occurrence probability p.*

Proof. We let η_0 denote the value of $F^{-1}(0; 0, 1)$. By Lemma 2(ii), we rewrite the expected cost of incorrectly estimating p for given $\hat{\tau}$ and $\hat{\theta}$ as

$$\theta \sum_{w>\gamma} \left\{ C\left(\frac{\hat{Q}(w,\gamma) - \tau}{\theta}; p, 0, 1\right) - C\left(\frac{(\tau + \eta_0\theta) - \tau}{\theta}; p, 0, 1\right) \right\} \mathbb{P}\left(\hat{p} = w\right)$$

for $p \leq \gamma_o$, while it takes the form

$$\theta \sum_{w\leq\gamma} C\left(\tfrac{(\tau+\eta_0\theta)-\tau}{\theta}; p, 0, 1\right) \mathbb{P}\left(\hat{p} = w\right) + \theta \sum_{w>\gamma} C\left(\tfrac{\hat{Q}(w,\gamma)-\tau}{\theta}; p, 0, 1\right) \mathbb{P}\left(\hat{p} = w\right)$$
$$-\theta C\left(\eta(p, \gamma_o); p, 0, 1\right)$$

for $p > \gamma_o$. Using $\hat{\tau} + \eta(w,\gamma)\hat{\theta}$ as the inventory-target estimator $\hat{Q}(w,\gamma)$, we further rewrite this expression as $\theta\Delta_{\mathcal{U},\mathcal{V}}(\gamma; p, 0, 1)$, where $\Delta_{\mathcal{U},\mathcal{V}}(\gamma; p, 0, 1)$ is given by

$$\sum_{w>\gamma} \left\{ C\left(\mathcal{U} + \eta(w,\gamma)\mathcal{V}; p, 0, 1\right) - C\left(\eta_0; p, 0, 1\right) \right\} \mathbb{P}\left(\hat{p} = w\right)$$

for $p \leq \gamma_o$, and it is given by

$$\sum_{w\leq\gamma} C\left(\eta_0; p, 0, 1\right) \mathbb{P}\left(\hat{p} = w\right) + \sum_{w>\gamma} C\left(\mathcal{U} + \eta(w,\gamma)\mathcal{V}; p, 0, 1\right) \mathbb{P}\left(\hat{p} = w\right) - C\left(\eta(p, \gamma_o); p, 0, 1\right)$$

for $p > \gamma_o$. We know from Lemma 4.2.2(i) that $\mathcal{U} = (\hat{\tau} - \tau)/\theta$ and $\mathcal{V} = \hat{\theta}/\theta$ are independent of the true values of τ and θ. □

4.2.5 Examples

We illustrate the quantification of the expected cost of parameter uncertainty first for exponentially distributed demand size and then for normally distributed demand size. Consistent with our presentation in the previous section, we quantify the expected cost of parameter uncertainty for an arbitrary threshold γ on the unit interval. The use of γ_o in place of γ leads to the expected cost of parameter uncertainty under naive inventory-target estimation.

Exponentially Distributed Demand Size

We let the demand size X be exponentially distributed with mean θ; i.e., $F(x;\theta) = 1 - \exp(-x/\theta)$. Therefore, the optimal inventory target q^* is $\theta \log(p/\gamma_o)$ if $p > \gamma_o$, and zero, otherwise (Proposition 4.2.1). Then, we identify the minimum expected cost with known p and θ as follows:

$$C(q^*;p,\theta) = \begin{cases} \theta p b & \text{for } p \leq \gamma_o, \\ \theta h \left(\log\left(p/\gamma_o\right) - p + 1\right) & \text{for } p > \gamma_o. \end{cases}$$

Without knowing the true values of p and θ, the decision maker obtains $\hat{p} = n_d/n$ and $\bar{x} = \sum_{i=1}^{n_d} x_i/n_d$ from historical data of length n. The naive inventory-target estimator takes the form $\bar{x}\log(\hat{p}/\gamma_o)$ if $\hat{p} > \gamma_o$, and zero, otherwise.

The goal is to characterize the expected cost due to the incorrect estimation of the unknown parameters p and θ for an arbitrary threshold γ. First, we let $\mathcal{V}_w$ denote $\sum_{i=1}^{nw} Z_i/(nw)$ with $Z_i = X_i/\theta$; i.e., $\mathcal{V}_w$ is the average of nw independent and exponentially distributed random variables, each with a mean of one. Lemma 4.2.3

provides a key component to the quantification of the expected cost of parameter uncertainty in Proposition 4.2.3.

Lemma 4.2.3. *Let Z be an exponentially distributed random variable with mean one. The expected value of $(Z - \mathcal{V}_w \log(w/\gamma))^+$, which we denote with $\ell(n, \gamma, w)$, is given by $(\log(w/\gamma)/(nw) + 1)^{-nw}$ for the demand-occurrence probability estimate $w \in \{1/n, 2/n, \dots, 1\}$ greater than γ.*

Proof. Taking the expectation of $(Z - \mathcal{V}_w \log(w/\gamma))^+$ with respect to Z gives

$$\int_{\mathcal{V}_w \log\left(\frac{w}{\gamma}\right)}^{\infty} \left(z - \mathcal{V}_w \log\left(\frac{w}{\gamma}\right)\right) \exp(-z) \mathrm{d}z = \left(\frac{w}{\gamma}\right)^{\mathcal{V}_w}.$$

Since $nw\,\mathcal{V}_w$ is Gamma$(nw, 1)$ distributed, the result follows from the evaluation of the integral

$$\int_0^{\infty} \left(\frac{w}{\gamma}\right)^{-\frac{x}{nw}} \frac{1}{\Gamma(nw)} x^{nw-1} \exp(-x) \mathrm{d}x,$$

where $\Gamma(nw) = (nw - 1)!$ is the gamma function. $\square$

Next, we characterize the expected cost of parameter uncertainty for an arbitrary threshold γ:

Proposition 4.2.3. *For threshold γ and exponentially distributed demand with mean θ, the expected cost due to incorrect estimation of parameters p and θ is given by*

$$\theta \sum_{w > \gamma} \left\{ h \log\left(\frac{w}{\gamma}\right) + p(h + b)\left(\ell(n, \gamma, w) - 1\right) \right\} \mathbb{P}(\hat{p} = w)$$

for $p \leq \gamma_o$; and for $p > \gamma_o$, it is given by

$$\theta p b \sum_{w \leq \gamma} \mathbb{P}(\hat{p} = w) + \theta \sum_{w > \gamma} \left\{ h \log\left(\tfrac{w}{\gamma}\right) - ph + p(h + b)\ell(n, \gamma, w) \right\} \mathbb{P}(\hat{p} = w)$$
$$-\theta h \left(\log\left(\tfrac{p}{\gamma_o}\right) - p + 1\right)$$

Proof. Suppose that $p \leq \gamma_o$ and w, the realized value of $\hat{p}$ is greater than γ. In this case, the optimal inventory target is zero, while the inventory-target estimator $\hat{Q}(w, \gamma)$ takes a positive value. Consequently, the expected cost of parameter uncertainty is given by

$$\mathbb{E}\left(\sum_{w>\gamma}\left(C(\hat{Q}(w,\gamma);p,\theta) - C(0;p,\theta)\right)\mathbb{P}(\hat{p}=w)\right). \tag{4.1}$$

Let $\hat{Q}(w, \gamma) = \bar{x}\log(w/\gamma)$ be the inventory target when the realized value of the demand-occurrence probability is w, and hence, $\bar{x}$ is the sample mean of nw exponentially distributed random variables with mean θ. Note that $C(q; p, \theta) = (1-p)hq + p(h(q-\theta) + (h+b)\mathbb{E}(X-q)^+)$. By plugging in the value of $C(0; p, \theta) = \theta pb$ and noting that $\mathbb{E}(\hat{Q}(w, \gamma)) = \theta \log(w/\gamma)$ and $\mathbb{E}(X - \hat{Q}(w, \gamma))^+ = \theta\,\mathbb{E}(Z - \hat{Q}(w, \gamma)/\theta)^+$, the expression (4.1) can be rewritten as

$$\theta \sum_{w>\gamma}\left\{h\log\left(\frac{w}{\gamma}\right) + p(h+b)\left(\mathbb{E}\left(Z - \frac{\bar{x}\log(w/\gamma)}{\theta}\right)^+ - 1\right)\right\}\mathbb{P}(\hat{p}=w).$$

The replacement of $\mathbb{E}\left(Z - \bar{x}\log(w/\gamma)/\theta\right)^+$ with the closed-form expression derived in Lemma 4.2.3 gives the result for $p \leq \gamma_o$. The proof for $p > \gamma_o$ follows in a similar manner. $\square$

Normally Distributed Demand Size

We now let the demand size X have the normal distribution with mean τ and standard deviation θ. The cdf of the standard normal random variable $Z = (X-\tau)/\theta$ is denoted by $\Phi(\cdot)$. If $p > \gamma_o$, then the optimal inventory target q^* is given by $\tau + \eta(p, \gamma_o)\,\theta$, where $\eta(p, \gamma_o)$ is equal to $\Phi^{-1}(1 - \gamma_o/p)$. Otherwise, q^* is set to zero (Proposition 4.2.1). The minimum expected cost with known p, τ, and θ can be shown to be

$$C(q^*; p, \tau, \theta) = \begin{cases} \tau pb & \text{for } p \leq \gamma_o, \\ \tau(1-p)h + p(h+b)\theta\phi\left(\eta(p,\gamma_o)\right) & \text{for } p > \gamma_o, \end{cases}$$

where $\phi(\cdot)$ is the probability density function (pdf) of the standard normal random variable. The decision maker estimates the unknown parameters p, τ, and θ^2 by $\hat{p} = n_d/n$, $\bar{x} = \sum_{i=1}^{n_d} x_i/n_d$, and $s^2 = \sum_{i=1}^{n_d}(x_i - \bar{x})^2/(n_d - 1)$ from the intermittent demand history of length n. The naive inventory-target estimator is then given by $\bar{x} + \eta(\hat{p}, \gamma_o)\, s$ if $\hat{p} > \gamma_o$, and zero, otherwise.

We now let $\mathcal{U}_w = \sum_{i=1}^{nw} Z_i/(nw)$ and $\mathcal{V}_w^2 = \sum_{i=1}^{nw}(Z_i - \mathcal{U}_w)^2/(nw-1)$ be the sample mean and sample variance of nw independent standard normal random variables: (i) $\sqrt{nw}\,\mathcal{U}_w$ has the standard normal distribution; (ii) $(nw-1)\mathcal{V}_w^2$ has the chi-square distribution with $nw-1$ degrees of freedom; and (iii) $\sqrt{nw}\,\mathcal{U}_w$ and $(nw-1)\mathcal{V}_w^2$ are independent (Rohatgi and Saleh, 2000). We use these properties in Lemma 4.2.4, which is the first step towards characterizing the expected cost of parameter uncertainty in Proposition 4.2.4:

Lemma 4.2.4. *Let Z be a standard normal random variable. The expected value of $(Z - (\mathcal{U}_w + \eta(w,\gamma)\mathcal{V}_w))^+$, which we denote with $\ell(n,\gamma,w)$, is given by*

$$\sqrt{\frac{nw+1}{nw}}\frac{1}{\sqrt{2\pi}}\left(\frac{nw\,\eta(w,n)^2}{(nw)^2-1}+1\right)^{\frac{1-nw}{2}} - \frac{\Gamma\left(\frac{nw}{2}\right)}{\Gamma\left(\frac{nw-1}{2}\right)}\frac{\sqrt{2}\,\eta(w,\gamma)}{\sqrt{nw-1}}\left(1 - T_{nw}\left(\frac{nw\,\eta(w,\gamma)}{\sqrt{(nw)^2-1}}\right)\right)$$

for the demand-occurrence probability estimate $w \in \{2/n, 3/n, \ldots, 1\}$ greater than γ, where $\Gamma(\cdot)$ is the gamma function and $T_{nw}(\cdot)$ is the cdf of the t distribution with nw degrees of freedom.

Proof. Note that $\mathcal{U}_w$ is normally distributed with mean zero and variance $1/(nw)$. Therefore, the random variable $Z - \mathcal{U}_w$ is normally distributed with mean zero and variance $(nw+1)/nw$. By letting ξ denote the standard normal random variable $\sqrt{nw/(nw+1)}\,(Z - \mathcal{U}_w)$, we obtain

$$\mathbb{E}\,(Z - (\mathcal{U}_w + \eta(w,\gamma)\mathcal{V}_w))^+ = \sqrt{\frac{nw+1}{nw}}\;\mathbb{E}\left(\xi - \eta(w,\gamma)\frac{\mathcal{V}_w}{\sqrt{\frac{nw+1}{nw}}}\right)^+. \tag{4.2}$$

Taking the expectation with respect to ξ, the right-hand side of (4.2) is given by

$$\sqrt{\frac{nw+1}{nw}}\,\mathbb{E}\left(\phi\left(\frac{\eta(w,\gamma)\mathcal{V}_w}{\sqrt{\frac{nw+1}{nw}}}\right)-\frac{\eta(w,\gamma)\mathcal{V}_w}{\sqrt{\frac{nw+1}{nw}}}+\frac{\eta(w,\gamma)\mathcal{V}_w}{\sqrt{\frac{nw+1}{nw}}}\Phi\left(\frac{\eta(w,\gamma)\mathcal{V}_w}{\sqrt{\frac{nw+1}{nw}}}\right)\right).$$

There are three terms inside the expectation operator. For ease in exposition, we let ψ be a random variable that has the chi-square distribution with $nw-1$ degrees of freedom and denote its pdf with $\chi(\cdot)$, and rewrite the random variable $\mathcal{V}_w$ in terms of ψ. Then, the first term is given by

$$\begin{aligned}\sqrt{\frac{nw+1}{nw}}\mathbb{E}\left(\phi\left(\frac{\eta(w,\gamma)\mathcal{V}_w}{\sqrt{\frac{nw+1}{nw}}}\right)\right) &= \sqrt{\frac{nw+1}{nw}}\int_0^\infty \phi\left(\eta(w,\gamma)\sqrt{\frac{nw}{(nw)^2-1}}\sqrt{\psi}\right)\chi(\psi)\mathrm{d}\psi.\\ &= \sqrt{\frac{nw+1}{nw}}\frac{1}{\sqrt{2\pi}}\left(\frac{nw\,\eta(w,n)^2}{(nw)^2-1}+1\right)^{\frac{1-nw}{2}}.\end{aligned}$$

The second term $\eta(w,\gamma)\mathbb{E}(\mathcal{V}_w)$ reduces to the following form:

$$\eta(w,\gamma)\mathbb{E}\left(\frac{\sqrt{\psi}}{\sqrt{nw-1}}\right) = \frac{\eta(w,\gamma)}{\sqrt{nw-1}}\frac{\sqrt{2}\,\Gamma\left(\frac{nw}{2}\right)}{\Gamma\left(\frac{nw-1}{2}\right)}.$$

We rewrite the last term as a function of ψ as

$$\begin{aligned}&\eta(w,\gamma)\mathbb{E}\left(\frac{\sqrt{\psi}}{\sqrt{nw-1}}\Phi\left(\frac{\eta(w,\gamma)\sqrt{nw}}{\sqrt{(nw)^2-1}}\sqrt{\psi}\right)\right)\\ &\qquad = \frac{\eta(w,\gamma)}{\sqrt{nw-1}}\int_0^\infty \int_{-\infty}^{\frac{\eta(w,\gamma)\sqrt{nw}}{\sqrt{(nw)^2-1}}\sqrt{\psi}} \sqrt{\psi}\,\phi(z)\chi(\psi)\mathrm{d}z\mathrm{d}\psi. \qquad (4.3)\end{aligned}$$

It is well known that $\mathcal{T} := Z\sqrt{nw-1}/\sqrt{\psi}$ is a Student's t random variable with $nw-1$ degrees of freedom (Rohatgi and Saleh, 2000). The transformation of the random variable Z to $\mathcal{T}\sqrt{\psi}/\sqrt{nw-1}$ allows us to write (4.3) as

$$\frac{\eta(w,\gamma)}{\sqrt{nw-1}} \int_{-\infty}^{\frac{\eta(w,\gamma)\sqrt{nw}}{\sqrt{nw+1}}} \int_{0}^{\infty} \frac{\psi}{\sqrt{nw-1}} \phi\left(t\frac{\sqrt{\psi}}{\sqrt{nw-1}}\right) \chi(\psi)\mathrm{d}\psi\mathrm{d}t.$$

By evaluating the integrals first with respect to ψ and then with respect to t, we reduce it to

$$\frac{\eta(w,\gamma)}{\sqrt{nw-1}} \frac{\sqrt{2}\,\Gamma\left(\frac{nw}{2}\right)}{\Gamma\left(\frac{nw-1}{2}\right)} T_{nw}\left(\frac{nw\,\eta(w,\gamma)}{\sqrt{(nw)^2-1}}\right),$$

where $T_{nw}(\cdot)$ is the cdf of Student's t distribution with nw degrees of freedom. Rearranging the three terms derived above completes the proof. $\square$

We are now ready to characterize the expected cost of parameter uncertainty as a function of γ:

Proposition 4.2.4. *For threshold γ and normally distributed demand with mean τ and standard deviation θ, the expected cost due to the incorrect estimation of parameters p, τ, and θ is given by*

$$\theta \sum_{w>\gamma} \left\{ \frac{\sqrt{2}h\,\eta(w,\gamma)}{\sqrt{nw-1}} \frac{\Gamma\left(\frac{nw}{2}\right)}{\Gamma\left(\frac{nw-1}{2}\right)} + p(h+b)\ell(n,\gamma,w) + \eta_0(bp-(1-p)h) \right\} \mathbb{P}(\hat{p}=w)$$

for $p \leq \gamma_o$; and for $p > \gamma_o$, it is given by

$$\begin{aligned} &\theta \sum_{w\leq\gamma} \{\eta_0(h(1-p)-pb) - h\,\eta(p,\gamma_o)\}\,\mathbb{P}(\hat{p}=w) \\ &-\theta p(h+b)\left\{\phi\left(\eta(p,\gamma_o)\right) - (\gamma_o/p)\eta(p,\gamma_o)\right\} \\ &+\theta \sum_{w>\gamma} \left\{ \frac{\sqrt{2}h\,\eta(w,\gamma)}{\sqrt{nw-1}} \frac{\Gamma\left(\frac{nw}{2}\right)}{\Gamma\left(\frac{nw-1}{2}\right)} + p(h+b)\ell(n,\gamma,w) - h\eta(p,\gamma_o) \right\} \mathbb{P}(\hat{p}=w). \end{aligned}$$

Proof. For $p \leq \gamma_o$, the expected cost of incorrectly estimating the parameters p, τ, and θ is given by

$$\mathbb{E}\left(\sum_{w>\gamma}\left(C(\hat{\tau}+\eta(w,\gamma)\hat{\theta};p,\tau\theta)-C(\tau+\eta_0\theta;p,\tau,\theta)\right)\mathbb{P}(\hat{p}=w)\right). \tag{4.4}$$

Since $C(q;p,\tau,\theta)$ is given by $(1-p)hq+p(h(q-\tau)+(h+b)\mathbb{E}(\tau+Z\theta-q)^+)$, we can rewrite (4.4) as

$$\begin{aligned}
h\theta(1-p)\sum_{w>\gamma}\mathbb{E}\left(\tfrac{\hat{\tau}-\tau}{\theta}+\eta(w,\gamma)\tfrac{\hat{\theta}}{\theta}-\eta_0\right)\mathbb{P}(\hat{p}=w)\\
+p\theta h\sum_{w>\gamma}\mathbb{E}\left(\tfrac{\hat{\tau}-\tau}{\theta}+\eta(w,\gamma)\tfrac{\hat{\theta}}{\theta}-\eta_0\right)\mathbb{P}(\hat{p}=w)\\
+p\theta(h+b)\sum_{w>\gamma}\left(\mathbb{E}\left(Z-\tfrac{\hat{\tau}-\tau}{\theta}-\eta(w,\gamma)\tfrac{\hat{\theta}}{\theta}\right)^+-\mathbb{E}(Z-\eta_0)^+\right)\mathbb{P}(\hat{p}=w).
\end{aligned}$$

Note that $(\hat{\tau}-\tau)/\theta$ is normally distributed with zero mean, and $\sqrt{nw-1}\,\hat{\theta}/\theta$ is the square root of a chi-square random variable with $nw-1$ degrees of freedom. Therefore, $\mathbb{E}((\hat{\tau}-\tau)/\theta)=0$ and $\sqrt{nw-1}\,\mathbb{E}(\hat{\theta}/\theta)=\sqrt{2}\Gamma(nw/2)/\Gamma((nw-1)/2)$. By definition, we also have $\mathbb{E}(Z-\eta_0)^+=-\eta_0$ (i.e., $\Phi(\eta_0)$ is assumed zero). Finally, plugging the closed-form expression obtained for $\mathbb{E}(Z-(\hat{\tau}-\tau)/\theta-\eta(w,\gamma)\hat{\theta}/\theta)^+$ in Lemma 4.2.4 and rearranging the terms give the result. The proof for $p>\gamma_o$ follows in a similar manner. □

4.3 Improved Inventory-Target Estimation

In this section, we switch our focus to the minimization of the expected cost of parameter uncertainty, which has been characterized in Section 4.2. We know from Proposition 4.2.2 that the expected cost of parameter uncertainty $\mathbb{E}\left(\Delta_{\hat{\tau},\hat{\theta}}(\gamma;p,\tau,\theta)\right)$ is equal to $\theta\,\mathbb{E}\left(\Delta_{\mathcal{U},\mathcal{V}}(\gamma;p,0,1)\right)$, and hence, its minimization with respect to threshold γ does not require the knowledge of parameters τ and θ. For ease in exposition, we let $\mathcal{R}(\gamma; p)$ denote $\mathbb{E}\left(\Delta_{\mathcal{U},\mathcal{V}}(\gamma; p, 0, 1)\right)$. Section 4.3.1 discusses the minimization of $\mathcal{R}(\gamma; p)$ over threshold γ along with the implementation details. Section 4.3.2 compares the performance of an inventory-target estimator with optimized threshold to the performance of the naive inventory-target estimator with threshold γ_o.

4.3.1 Minimizing the Expected Cost of Parameter Uncertainty

When the true parameter values are known, Proposition 4.2.1 shows the optimality of a myopic base-stock policy with an explicit form for the optimal inventory target. However, when there is parameter uncertainty, parameter estimators appear in the dynamic programming formulation as additional states. The resulting multi-dimensional dynamic program with continuous state space is subject to the curse of dimensionality. Therefore, we continue to focus on the forthcoming period as in the optimal policy with known parameters. To be specific, the objective is to minimize the expected cost due to parameter uncertainty in period $n+1$, given by $\theta\mathcal{R}(\gamma; p)$, in the presence of a demand history of length n. In this way, we consider the impact of estimation errors directly on the expected cost, which is operationally more relevant than statistical measures such as mean squared error and mean absolute deviation of the parameter estimates.

It is not possible to compute the value of γ that minimizes $\mathcal{R}(\gamma; p)$ directly as $\mathcal{R}(\gamma; p)$ remains a function of the unknown parameter p. Obviously, we would prefer a decision rule (i.e., a value for optimum γ) superior to all others for all values of p. Failing this, the decision maker can be forced to choose between two or more rules, each of which is optimal only over certain values of p. In this study, we introduce the notion of *confidence-interval information* to address this problem. More specifically, we evaluate $\mathcal{R}(\gamma; p)$ over a certain range of p values, which is likely to include the true value of p, and then choose the value of γ that minimizes the average value of $\mathcal{R}(\gamma; p)$.

It is well known that the Wald statistic $(\hat{p} - p)/\sqrt{\hat{p}(1-\hat{p})/n}$ converges to a standard normal random variable with increasing n (Rohatgi and Saleh, 2000). The true value of the demand-occurrence probability p (which is fixed and unknown to the decision maker) is then trapped in the interval

$$\mathcal{P} := \left[\hat{p} - \upsilon_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \upsilon_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

with probability at least $1-\alpha$, where $\upsilon_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$. That is, the decision maker uses the historical data available by the end of period n to understand where the true value of p is likely to lie on the unit interval. The improved inventory-target estimation then corresponds to the following optimization problem:

$$\gamma^* := \underset{\gamma \in [0,1]}{\arg\min} \int_{\mathcal{P} \cap [0,1]} \mathcal{R}(\gamma; p) \mathrm{d}p. \tag{4.5}$$

We note that Akcay et al. (2011) and Ramamurthy et al. (2012) estimate inventory targets by optimizing a frequentist risk function as well, and they suggest using the maximum likelihood estimate of an unknown parameter when the risk function requires the specification of the unknown parameter. It is worth noting that this practice is subsumed in the definition of $\mathcal{P}$ for the special case of α equal to 1. Next, we replace γ_o of the naive inventory-target estimation with γ^* to account for the uncertainty around the parameter estimates:

Definition 4.3.1 (Improved Inventory-Target Estimation)**.** *If $\hat{p} > \gamma^*$, then the improved inventory-target estimator $\hat{Q}(\hat{p}, \gamma^*)$ is given by $\hat{\tau} + \eta(\hat{p}, \gamma^*)\hat{\theta}$. If $\hat{p} \leq \gamma^*$, then $\hat{Q}(\hat{p}, \gamma^*)$ is zero.*

The optimal threshold γ^* can be identified by evaluating the objective function in (4.5) on a fine grid of γ between 0 and 1. While this is clearly not the most efficient approach, it is practical as the optimal threshold can be obtained almost instantly.

4.3.2 Results and Insights

The objective of this section is two-fold: (1) To quantify the expected cost of parameter uncertainty associated with naive inventory-target estimation. (2) To investigate the effectiveness of improved inventory-target estimation in reducing the expected cost of parameter uncertainty. We let the demand size be normally distributed with mean $\tau = 100$ and standard deviation $\theta = 20$. It is important to note that the optimal threshold γ^* of improved inventory-target estimation does not depend on what we specify for τ and θ. We let $n \in \{10, 25, 30\}$ and $\gamma_o \in \{0.25, 0.50, 0.75\}$ to investigate the effectiveness of the improved-inventory target estimation as a function of the length of the demand history and the asymmetry of the unit holding and backlogging costs.

We provide three different results for each scenario in Tables 4.1, 4.2, and 4.3. First, we present the optimal threshold γ^*, a decision rule that minimizes the expected cost of parameter uncertainty before the demand realization in the forthcoming period. Second, we provide the ratio of the expected cost of parameter uncertainty to the minimum expected cost with known parameters; i.e., $\mathcal{E}(\gamma_o; p) := \theta\mathcal{R}(\gamma_o; p)/C(q^*; p, \tau, \theta)100\%$, evaluated at $p \in \{0.2, 0.4, 0.6, 0.8\}$ for the naive inventory-target estimation. Finally, we report $\mathcal{E}(\gamma^*; p)$ for the improved inventory-target estimation. The decision maker uses two pieces of information to compute γ^*: The length of the historical data n and the number of positive demand realizations n_d. We report the values of γ^* when n_d is equal to $p\,100\%$ of the demand history n. It is important to note that the decision maker uses n_d to build the confidence interval $\mathcal{P}$; the uncertainty of n_d (or $\hat{p}$) is still captured by the minimization of

Table 4.1: $\gamma_o = 0.25$ (Unit backlogging cost is greater than the unit holding cost)

		p			
		0.2	0.4	0.6	0.8
$n = 10$	γ^*	1.000	0.183	0.192	0.221
	$\mathcal{E}(\gamma_o; p)$	12.5%	14.0%	7.0%	4.6%
	$\mathcal{E}(\gamma^*; p)$	0.0%	8.1%	5.2%	3.8%
$n = 15$	γ^*	1.000	0.188	0.224	0.229
	$\mathcal{E}(\gamma_o; p)$	10.9%	9.0%	3.5%	2.9%
	$\mathcal{E}(\gamma^*; p)$	0.0%	5.7%	3.0%	1.9%
$n = 30$	γ^*	1.000	0.197	0.231	0.236
	$\mathcal{E}(\gamma_o; p)$	6.9%	4.7%	1.5%	1.2%
	$\mathcal{E}(\gamma^*; p)$	0.0%	2.6%	0.6%	0.5%

the expected cost of parameter uncertainty in (4.5). In each experiment, the optimal threshold γ^* is computed within seconds on a standard desktop system.

In improved inventory-target estimation, the decision maker compares γ^* with the current estimate of the demand-occurrence probability to decide whether to stock inventory. We note that γ^* is equal to 1 in some instances. This means that the decision maker sets inventory target to zero units regardless of the realized demand history since $\hat{p}$ is less than or equal to one for *any* demand history. We observe γ^* to be equal to 1 for $p = 0.2$ and $\gamma_o = 0.25$ in Table 4.1, for $p \in \{0.2, 0.4\}$ and $\gamma_o = 0.50$ in Table 4.2, and for $p \in \{0.2, 0.4, 0.6\}$ and $\gamma_o = 0.75$ in Table 4.3. In these cases, the improved inventory-target estimation mimics the optimal policy with known parameters, which sets the inventory target to zero. Therefore, the expected cost of parameter uncertainty is zero, and we identify the ratio $\mathcal{E}(\gamma^*; p)$ as 0.0%.

We observe the difference between the true value of p and γ_o to be an important driver of the expected cost of parameter uncertainty. For example, when $n = 10$ and $p = 0.2$, we see that $\mathcal{E}(\gamma_o; p)$ is as much as 12.5% for $\gamma_o = 0.25$ (Table 4.1), while it is only 1.1% for $\gamma_o = 0.75$ (Table 4.3). The probability of $\hat{p}$ being less than $\gamma_o = 0.75$ is greater than the probability of $\hat{p}$ being less than $\gamma_o = 0.25$ for $p = 0.2$. Thus, it is more

Table 4.2: $\gamma_o = 0.50$ (Unit backlogging cost is equal to the unit holding cost)

		p			
		0.2	0.4	0.6	0.8
$n = 10$	γ^*	1.000	1.000	0.392	0.460
	$\mathcal{E}(\gamma_o; p)$	1.9%	9.1%	13.1%	8.4%
	$\mathcal{E}(\gamma^*; p)$	0.0%	0.0%	7.0 %	4.9%
$n = 15$	γ^*	1.000	1.000	0.395	0.461
	$\mathcal{E}(\gamma_o; p)$	1.1%	8.9%	8.3%	3.5%
	$\mathcal{E}(\gamma^*; p)$	0.0%	0.0%	5.4%	2.6%
$n = 30$	γ^*	1.000	1.000	0.423	0.474
	$\mathcal{E}(\gamma_o; p)$	0.0%	4.0%	6.5%	1.4%
	$\mathcal{E}(\gamma^*; p)$	0.0%	0.0%	3.1%	1.1%

Table 4.3: $\gamma_o = 0.75$ (Unit holding cost is greater than the unit backlogging cost)

		p			
		0.2	0.4	0.6	0.8
$n = 10$	γ^*	1.000	1.000	1.000	0.693
	$\mathcal{E}(\gamma_o; p)$	1.1 %	3.2%	13.3%	9.5%
	$\mathcal{E}(\gamma^*; p)$	0.0%	0.0%	0.0 %	7.3%
$n = 15$	γ^*	1.000	1.000	1.000	0.699
	$\mathcal{E}(\gamma_o; p)$	0.6%	1.5%	7.0%	8.3%
	$\mathcal{E}(\gamma^*; p)$	0.0%	0.0%	0.0%	5.7%
$n = 30$	γ^*	1.000	1.000	1.000	0.724
	$\mathcal{E}(\gamma_o; p)$	0.0%	0.0%	2.9%	5.5%
	$\mathcal{E}(\gamma^*; p)$	0.0%	0.0%	0.0%	3.9%

likely for the naive inventory-target estimation to align with the optimal policy with known parameters (i.e., stocking zero units) when $p = 0.2$ and $\gamma_o = 0.75$ compared to when $p = 0.2$ and $\gamma_o = 0.25$. This is the main reason why we observe a smaller $\mathcal{E}(\gamma_o; p)$ for $\gamma_o = 0.75$ than for $\gamma_o = 0.25$ in this example. Improved inventory-target estimation achieves 0.0% as the value of $\mathcal{E}(\gamma^*; p)$ for both of these cases.

The intuition on the effectiveness of the improved inventory-target estimation lies behind accounting for the costs of overestimation and underestimation. For example, when $p = 0.6$ and $n = 10$, the optimal threshold γ^* is 0.192 for $\gamma_o = 0.25$ (Table 4.1), while it is 1 for $\gamma_o = 0.75$ (Table 4.3). That is, the improved inventory-target estimation makes it more difficult (in fact, impossible in this example) to set a positive inventory target by choosing $\gamma^* = 1$ greater than $\gamma_o = 0.75$ when the holding cost dominates the backlogging cost (or the overestimation is more costly than underestimation). On the other hand, when the backlogging cost dominates the holding cost (or the underestimation is more costly than overestimation), we observe that $\gamma^* = 0.192$ is less than $\gamma_o = 0.25$, meaning that improved inventory-target estimation dictates ordering more than the naive inventory-target estimate. We note that γ^* can still be smaller than γ_o even though the holding cost dominates the backlogging cost as in the $p = 0.8$ column of Table 4.3, depending on the length of the demand history and the number of positive demand observations. We observe in our experiments with a wider range of cost parameters that the role of γ^* on balancing the costs of underestimation and overestimation becomes more visible when γ_o approaches zero or one.

We know that the estimators $\hat{p}$, $\hat{\tau}$, and $\hat{\theta}$ are all consistent; i.e., they converge to their true values with increasing length of the demand history. Therefore, the expected cost of parameter uncertainty (or the values of $\mathcal{E}(\gamma_o; p)$ and $\mathcal{E}(\gamma^*; p)$) decreases as the length of the demand history increases. For example, when $p = 0.6$ and $\gamma_o = 0.50$ (Table 4.2), we observe that $\mathcal{E}(\gamma_o; p)$ decreases from 13.1% to 6.5% as n increases from 10 to 30, while $\mathcal{E}(\gamma^*; p)$ decreases from 7.0% to 3.1%. To put it another way, while the expected cost of parameter uncertainty disappears with increasing length of the demand history, the decision maker must be aware of the cost of parameter uncertainty if there is a limited amount of historical data.

4.4 Correlated Demand Size and Number of Inter-Demand Periods

Section 4.4.1 presents a copula-based demand model to capture the relation between a demand size and the number of zero-demand periods preceding the demand, and characterizes the optimal policy when there is no parameter uncertainty. Section 4.4.2 introduces a statistical test to assess the existence of correlation in the presence of a *limited* data. Section 4.4.3 introduces an alternative test for the same purpose based on the difference between the expected cost of an inventory-target estimate and the minimum expected cost. We name these tests as *correlation test* and *operational test*, respectively. Contrary to the correlation test – which is solely based on the sampling distribution of the copula-parameter estimator, the operational test accounts for the costs of over- and underestimation of the optimal inventory target in investigating the existence of correlation. Section 4.4.4 describes how the improved inventory-target estimation of Section 4.3 can be extended to intermittent demand with correlation.

4.4.1 Correlated Intermittent Demand Model and Optimal Inventory Policy

When the size of a (positive) demand depends on the number of periods since the last demand, there is a challenge in the characterization of the expected cost function: Modeling the distribution of demand size *conditional* on the number of periods since the last demand, while retaining the distribution of demand size X from the location-scale family along with the geometric distribution for the number of inter-demand periods Y. We address this issue by constructing the joint distribution of X and Y with a *copula*, which allows us to model the univariate distributions of X and Y and the dependence structure between them separately. Sklar's theorem (Nelsen 2006) elucidates the role played by a copula in the relation between a joint distribution and its univariate cdfs:

Theorem 4.4.1 (Sklar's Theorem)**.** *Given the bivariate cdf $H(x, y)$ for the random variables X and Y with univariate cdfs $F(x; \tau, \theta)$ and $G(y; p)$, there exists a copula $\mathscr{C}$ such that $H(x, y) = \mathscr{C}(F(x; \tau, \theta), G(y; p))$. Conversely, if $\mathscr{C}$ is a copula and $F(x; \tau, \theta)$ and $G(y; p)$ are univariate cdfs, then $H(x, y)$ is a bivariate cdf for the random variables X and Y.*

The applications of copulas are relatively few in operations management; Clemen and Reilly (1999), Cooper and Gupta (2006), Corbett and Rajaram (2006), and Aydın et al. (2012) are some of the few examples in decision analysis, revenue management, and supply chain management. To the best of our knowledge, we are the first to use copula in the interface of inventory management and intermittent demand modeling with parameter uncertainty.

More specifically, we construct the bivariate distribution of X and Y using the *normal copula*, which encodes the dependence precisely the same way a bivariate normal distribution does: $H(x, y) = \Phi_2 \{\Phi^{-1}(F(x; \tau, \theta)), \Phi^{-1}(G(y; p)); r\}$ where $\Phi_2\{\cdot, \cdot; r\}$ is the bivariate standard normal cdf with the correlation coefficient r (Nelsen, 2006). We use the normal copula because the copula parameter r can be conveniently estimated by using the tools available to estimate the correlation coefficient of a bivariate standard normal distribution. We next characterize the optimal policy when there is no parameter uncertainty.

Proposition 4.4.1. *Suppose that the demand-size distribution parameters τ and θ, the demand-occurrence probability p, and the copula parameter r are known. Let y denote the number of zero-demand periods since the last demand.*

(i) A base-stock policy is optimal in any time period for all values of y.

(ii) For the demand-occurrence probability $p > \gamma_o$, the optimal inventory target q^* is given by $\tau + F^{-1}(\kappa^*; 0, 1)\,\theta$, where κ^* is the unique value of κ that solves the equation

$$\frac{\Phi_2\left\{\Phi^{-1}(\kappa), \Phi^{-1}(G(y;p)); r\right\} - \Phi_2\left\{\Phi^{-1}(\kappa), \Phi^{-1}(G(y-1;p)); r\right\}}{(1-p)^y p} = 1 - \frac{\gamma_o}{p}. \tag{4.6}$$

(iii) For $p \leq \gamma_o$, the optimal inventory target q^* is zero.

Proof. (i) The state space is defined by the level of inventory before ordering, ν and the number of zero-demand periods since the last demand, y. We now let $C(q|y; p, \tau, \theta, r)$ denote $(1-p)hq + p\,(h\,\mathbb{E}(q - X|y)^+ + b\,\mathbb{E}(X|y - q)^+)$. The optimality equation is written as $V_t(\nu, y) = \min_{q \geq \max(0,\nu)}\{C(q|y; p, \tau, \theta, r) + (1-p)V_{t+1}(q, y+1) + p\,\mathbb{E}(V_{t+1}(q - X|y, 0))\}$ for $t = 1, 2, \ldots, N$ with $V_{N+1}(\nu, y) = 0$. Let $\mathscr{L}_t(q)$ denote $C(q|y; p, \tau, \theta, r) + (1-p)V_{t+1}(q, y+1) + p\mathbb{E}\left(V_{t+1}(q - X, 0)\right)$. We now examine the single-period problem in period N for a particular value of y. Note that the unique minimizer of the convex function $\mathscr{L}_N$ is the base-stock level for period N. Since convexity is preserved under minimization (Theorem A.4, Porteus 2002), V_N is also convex, and the result follows from a recursive argument through the periods $t = N-1, N-2, \ldots, 1$ for any value of y. There are two cases to analyze: If $p > \gamma_0$, then the optimal base-stock level is positive for all y values, while it is zero if $p \leq \gamma_0$. We discuss the minimization of the single-period expected cost function below for each of these cases.

(ii) If $p > \gamma_0$, then the minimizer of the convex function $C(q|y; p, \tau, \theta, r)$ is characterized by the first-order condition $(1-p)h + p((h+b)H_{X|Y}(q|y) - b) = 0$ with $H_{X|Y}(\cdot|y)$ the cdf of the demand size X conditional on the number of zero-demand periods $Y = y$. Using the bivariate distribution of X and Y characterized by Sklar's theorem, $\mathbb{P}(X \leq x, Y = y)$ is obtained from $H(x, y) - H(x, y-1)$ and equal to $\Phi_2\left\{\Phi^{-1}(F(x; \tau, \theta)), \Phi^{-1}(G(y;p)); r\right\} - \Phi_2\left\{\Phi^{-1}(F(x; \tau, \theta)), \Phi^{-1}(G(y-1;p)); r\right\}$ by the definition of the normal copula. We also know $\mathbb{P}(Y = y) = (1-p)^y p$ by the definition of the geometric distribution. Consequently, the distribution of X conditional on $Y = y$; i.e., $\mathbb{P}(X \leq x | Y = y) = \mathbb{P}(X \leq x, Y = y)/\mathbb{P}(Y = y)$, takes the form

$$\frac{\Phi_2\left\{\Phi^{-1}(F(x;\tau,\theta)), \Phi^{-1}(G(y;p)); r\right\} - \Phi_2\left\{\Phi^{-1}(F(x;\tau,\theta)), \Phi^{-1}(G(y-1;p)); r\right\}}{(1-p)^y p}. \tag{4.7}$$

Since this expression is an increasing function of $F(x;\tau,\theta)$, there is a unique value of $F(x,\tau,\theta)$ which makes (4.7) equal to $1 - h/((h+b)p)$ to impose the first-order condition. Let κ^* denote this value of $F(x;\tau,\theta)$. Consequently, the optimal solution q^* of the single-period problem is given by $\tau + F^{-1}(\kappa^*;0,1)\theta$; i.e., $\kappa^* = F(q^*;\tau,\theta)$.

(iii) If $p \leq \gamma_0$, the expected cost $C(q|y;p,\tau,\theta,r)$ is minimized at zero, the boundary solution. □

It is important to note that the solution of (4.6) does not depend on the demand-size parameters due to our copula-based representation of intermittent demand. We will make use of this key property in establishing the distribution of the test statistic introduced in the operational test.

4.4.2 A Sampling-Based Test for Correlation

The copula-based representation of intermittent demand allows the univariate distribution parameters τ, θ, and p, and the copula parameter r to be estimated separately, potentially with different methods. In this section, we continue to use the estimators $\hat{\tau}$, $\hat{\theta}$, and $\hat{p}$ introduced in Section 4.2.3. We then transform the historical data $\{(x_t, y_t); t = 1, 2, \ldots, n_d\}$ into $\{(u_x^i, u_y^i); i = 1, 2, \ldots, n_d\}$ with $u_x^i := F((x_i - \hat{\tau})/\hat{\theta}; 0, 1)$ and $u_y^i := G(y; \hat{p})$, and use the maximum likelihood method to estimate the copula parameter r by fitting the normal copula to the transformed data:

$$\hat{r} := \underset{r \in (-1,1)}{\arg\max} \sum_{i=1}^{n_d} \log \phi_2\{\Phi^{-1}(u_x^i), \Phi^{-1}(u_y^i); r\}, \tag{4.8}$$

where $\phi_2\{\cdot, \cdot; r\}$ is the bivariate standard normal pdf with correlation r. The transformation of historical data into a unit hypercube by the parametric estimates of their univariate cdfs is known as the *inference functions for margins*; we refer the reader to Cherubini et al. (2011) for the details.

Clearly, the estimated value of $\hat{r}$ never takes the exact value of zero. Thus, the immediate task of a decision maker – before using the copula – is to make sure that the historical data carry sufficient information to reject the independence between the demand size and the number of inter-demand periods. There are powerful tools in statistics to test the existence of correlation in bivariate normal data; see Fosdick and Raftery (2012) for a review.

Traditionally, the distribution of a test statistic is based on asymptotic theory. In contrast, we generate test-statistic samples, each of which is obtained from a *limited* amount of demand data, to build an empirical distribution of the test statistic. The so-called bootstrap hypothesis testing has become increasingly attractive due to advances in computing (MacKinnon, 2009). We let the null hypothesis $\mathcal{H}_0$ be $r = 0$ and the two-tailed alternative hypothesis $\mathcal{H}_{\text{A}}$ be $r \neq 0$. The correlation test compares the observed value of the test statistic $\hat{r}$ with the distribution that it would follow if the null hypothesis were true. The null hypothesis is then rejected if $\hat{r}$ is sufficiently extreme relative to this distribution. The decision maker performs the test by only specifying the length of the demand history and the number of positive realizations. We present an outline of the correlation test in Figure 4.3.

Remark 4.4.1. *The copula-based representation of intermittent demand and the assumption of a location-scale family for the distribution of demand size allow the use of standardized demand samples* $\{z_1^b, z_2^b, \ldots, z_{n_d}^b\}$ *instead of the actual demand samples* $\{x_1^b, x_2^b, \ldots, x_{n_d}^b\}$ *– which could be generated only if the true values of* τ *and* θ *were known.*

Figure 4.3: The correlation test with $\mathcal{H}_0 : r = 0$ and $\mathcal{H}_\mathrm{A} : r \neq 0$

Initialization. Specify the number of bootstrap samples (B), the length of historical data (n), the number of positive realizations (n_d), and the significance level ($1-\alpha$).

Set $\hat{p} \leftarrow n_d/n$ and $b \leftarrow 1$.

Compute the test statistic $\hat{r}$ as in (7) from the historical data $\{(x_t, y_t); t = 1, 2, \ldots, n_d\}$.

while $b \leq B$:

Step 1. *Generate the bootstrap sample path*:

Set $i \leftarrow 1$ and $t \leftarrow 1$.

while $t \leq n$

Generate $(\mathcal{Z}^b_{x,i}, \mathcal{Z}^b_{y,i})$ from the bivariate standard normal cdf with correlation $r = 0$.

Obtain $z^b_i = F^{-1}(\Phi(\mathcal{Z}^b_{x,i}); 0, 1)$ and $y^b_i = G^{-1}(\Phi(\mathcal{Z}^b_{y,i}); \hat{p})$.

Set $t \leftarrow t + y^b_i + 1$ and $i \leftarrow i + 1$.

Let n^b_d be the number of nonzero demands in bth bootstrap sample path.

end

Step 2. *Compute the test statistic from the bootstrap sample path*:

Let $\bar{z}$ and s_z be the sample mean and sample standard deviation of $\{z^b_1, z^b_2, \ldots, z^b_{n_d}\}$.

Set $\hat{r}^b \leftarrow \arg\max_{r \in (-1,1)} \sum_{i=1}^{n^b_d} \log \phi_2\{\Phi^{-1}(F((z_i - \bar{z})/s_z; 0, 1)), \Phi^{-1}(G(y^b_i; n^b_d/n)); r\}$.

$b \leftarrow b + 1$

end

Step 3. *Construct the critical region to reject the null hypothesis*:

Sort $\hat{r}^b$, $b = 1, 2, \ldots, B$ in ascending order $(r_{(1)}, r_{(2)}, \ldots, r_{(B)})$.

Reject the null hypothesis $\mathcal{H}_0$ if $\hat{r} < r_{(\lfloor(\alpha/2)B\rfloor)}$ or $\hat{r} > r_{(\lceil(1-\alpha/2)B\rceil)}$.

We let the number of independent bootstrap sample paths B equal to $10,000$. The number of positive realizations n_d is taken as 20%, 40%, 60%, and 80% of n, the length of the demand history. A level $\alpha = 0.05$ test rejects the hypothesis that the correlation is zero in favor of the alternative that it is not zero if the copula-parameter estimate is less than 2.5% quantile or greater than 97.5% quantile. Table 4.4 presents the so-called 95% significance test bounds. For example, the significance test bounds are $\{-0.89, 0.89\}$ for a demand history of length 30 with 6 positive demand realizations. That is, the decision maker has sufficient evidence to reject the independence assumption only if the estimated value of the copula parameter is less than -0.89 or

Table 4.4: 95% significance bounds in the correlation test to reject $r = 0$ against $r \neq 0$

	The percentage of positive realizations in the n-period history			
n	20%	40%	60%	80%
30	{-0.89, 0.89}	{-0.62, 0.62}	{-0.48, 0.48}	{-0.41, 0.41}
50	{-0.70, 0.70}	{-0.46, 0.46}	{-0.36, 0.36}	{-0.31, 0.31}
100	{-0.47, 0.47}	{-0.32, 0.32}	{-0.26, 0.26}	{-0.22, 0.22}
250	{-0.28, 0.28}	{-0.20, 0.20}	{-0.16, 0.16}	{-0.14, 0.14}

greater than 0.89. On the other hand, the significance test bounds are $\{-0.14, 0.14\}$ for a demand history of length 250 with 200 positive demand realizations. In this case, the estimated value of the copula parameter must be less than -0.14 or greater than 0.14 to justify the modeling of correlation in intermittent demand. The significance bounds approach zero (i.e., the critical region becomes larger) as the length of the demand history and the number of positive realizations increase.

In Table 4.5, we focus on the fraction of time that the decision maker rejects the null hypothesis at the level $\alpha = 0.05$ based on 500 replications of the correlation test. In each scenario, we let the percentage of positive realizations be 20% (top) and 40% (bottom) of the demand history to focus on highly intermittent demand histories. Clearly, the fraction of rejections approaches one as the length of the demand history and the true value of the copula parameter (i.e., the strength of correlation) increases. For example, the decision maker rejects the null hypothesis of no correlation 5% of the time when the copula parameter is 0.15 in a demand history of length 30 with 12 positive demand realizations. In this case, the fraction of rejections reaches 94% when the copula parameter is as high as 0.90. It is worth noting that the correlation test only considers the sampling distribution of the copula-parameter estimator to decide whether to model the correlation. We next present an alternative hypothesis test that considers the expected cost associated with the over- and underestimation of the optimal inventory target to investigate the existence of correlation.

Table 4.5: The fraction of time the correlation test rejects $\mathcal{H}_0$ at 95% significance level

		r					
n	n_d	0.15	0.3	0.45	0.6	0.75	0.90
30	6	3%	3%	5%	12%	17%	38%
50	10	5%	7%	13%	32%	62%	85%
100	20	9%	20%	46%	77%	95%	100%
250	50	20%	53%	90%	99%	100%	100%
30	12	5%	9%	22%	37%	75%	94%
50	20	7%	20%	41%	71%	94%	100%
100	40	11%	37%	75%	96%	100%	100%
250	100	31%	80%	99%	100%	100%	100%

4.4.3 Sampling-Based Testing for Correlation: An Operational Alternative

As opposed to searching for the statistical significance of a parameter estimate, a decision maker is more interested in the performance of the decision based on the estimate. In this section, we propose an alternative test considering the sampling distribution of the difference between the expected costs of the optimal inventory target and its estimate.

We start with showing a result analogous to Lemma 4.2.2(ii) to reduce the expected cost function to an alternative form, which will be used to construct the sampling distribution of the test statistic. Let $C(q|y;p,\tau,\theta,r)$ denote the single-period expected cost conditional on the number of periods y since the last demand.

Lemma 4.4.1. *The expected cost $C(q|y;p,\tau,\theta,r)$ can be equivalently written as*

$$\theta C\left(\frac{q-\tau}{\theta}|y;p,0,1,r\right)+h\tau(1-p).$$

Proof. We first derive the pdf of X conditional on $Y = y$, denoted by $f_{X|Y}(x|y)$, as

$$f(x;\tau,\theta)\left\{\Phi\left(\tfrac{\Phi^{-1}(G(y+1;p))-r\Phi^{-1}(F(x;\tau,\theta))}{\sqrt{1-r^2}}\right)-\Phi\left(\tfrac{\Phi^{-1}(G(y;p))-r\Phi^{-1}(F(x;\tau,\theta))}{\sqrt{1-r^2}}\right)\right\},$$

where f is the marginal pdf of X. If we make the transformation $Z_c = (X - \tau)/\theta$ in $f_{X|Y}(\cdot|y)$, then we obtain the pdf of Z_c as

$$\tfrac{1}{\theta}f(z_c;0,1)\left\{\Phi\left(\tfrac{\Phi^{-1}(G(y+1;p))-r\Phi^{-1}(F(z_c;0,1))}{\sqrt{1-r^2}}\right)-\Phi\left(\tfrac{\Phi^{-1}(G(y;p))-r\Phi^{-1}(F(z_c;0,1))}{\sqrt{1-r^2}}\right)\right\}.$$

That is, the marginal-copula representation of intermittent demand allows us to transform the demand size X conditional on $Y = y$ into $\tau + Z_c\theta$ with Z_c the standardized conditional demand. Consequently, the expected cost function $C(q|y; p, \tau, \theta, r)$, which is given by $(1-p)hq + p\left(h\mathbb{E}(q - X|y)^+ + b\mathbb{E}(X|y - q)^+\right)$, can be transformed into

$$(1-p)hq + p\left(h\mathbb{E}(q - (\tau + Z_c\theta))^+ + b\mathbb{E}((\tau + Z_c\theta) - q)^+\right).$$

Rewriting $(1-p)hq$ as $\theta(1-p)h[(q-\tau)/\theta] + h\tau(1-p)$ completes the proof. □

The decision maker implements the optimal policy in Proposition 4.4.1 with an arbitrary threshold γ as follows: If the estimate of the demand-occurrence probability $\hat{p} > \gamma$, then the inventory target is set to $\hat{\tau} + F^{-1}(\hat{\kappa}_\gamma; 0, 1)\hat{\theta}$; and if $\hat{p} \leq \gamma$, then it is set to zero. We let $\hat{\kappa}_\gamma$ the value of κ that solves the equation $\Pi(\kappa; \hat{p}, y, \hat{r}) = 1 - \gamma/\hat{p}$. In this representation, $\Pi(\kappa; \hat{p}, y, \hat{r})$ denotes the left-hand side of (4.6) for notational convenience. The use of γ_o as the value of threshold γ leads to the correlated counterpart of the naive inventory target of Section 4.2.3.

Building on Lemma 4.4.1 and the definitions of $\mathcal{U}$ and $\mathcal{V}$ from Section 4.2.4, the difference between the expected cost of an inventory-target estimate and the minimum expected cost of the optimal inventory target takes the form $\theta\Lambda(\hat{\kappa}_\gamma, \mathcal{U}, \mathcal{V}, y; \gamma_o, p, r)$, where $\Lambda(\hat{\kappa}_\gamma, \mathcal{U}, \mathcal{V}, y; \gamma_o, p, r)$ is given by

$$\begin{cases} C(\mathcal{U} + F^{-1}(\hat{\kappa}_\gamma; 0,1)\mathcal{V} | y; p, 0, 1, r) - C(\eta_0 | y; p, 0, 1, r) & \text{for } \hat{p} > \gamma;\ p \leq \gamma_o, \\ C(\mathcal{U} + F^{-1}(\hat{\kappa}_\gamma; 0,1)\mathcal{V} | y; p, 0, 1, r) - C(F^{-1}(\kappa^*; 0,1) | y; p, 0, 1, r) & \text{for } \hat{p} > \gamma;\ p > \gamma_o, \\ C(\eta_0 | y; p, 0, 1, r) - C(F^{-1}(\kappa^*) | y; p, 0, 1, r) & \text{for } \hat{p} \leq \gamma;\ p > \gamma_o, \\ 0 & \text{for } \hat{p} \leq \gamma;\ p \leq \gamma_o. \end{cases}$$

We use the the nonnegative random variable $\Lambda(\hat{\kappa}_\gamma, \mathcal{U}, \mathcal{V}, y; \gamma_o, p, r)$ as the test statistic. The operational test compares the observed value of the test statistic with the distribution that it would follow under the null hypothesis of no correlation. The null hypothesis is then rejected when the observed value of the test statistic is sufficiently high relative to this distribution. We present the outline of the operational test in Figure 4.4.

Remark 4.4.2. *(i) We present the operational test for an arbitrary threshold γ on the unit interval. In the remainder of the section, we let γ be equal to γ_o for the numerical experiments. (ii) In Step 2, plugging the estimator $s/\sqrt{a_2}$ of $\hat{\theta}$ in $\mathcal{V} = \hat{\theta}/\theta$ leads to $v = s_z/\sqrt{a_2}$ – which is computed from the standardized demand samples $\{z_1^b, z_2^b, \ldots, z_{n_d}^b\}$ – as the sampled value of $\mathcal{V}$. Similarly, plugging the estimator $\bar{x} - a_1\hat{\theta}$ of $\hat{\tau}$ in $\mathcal{U} = (\hat{\tau} - \tau)/\theta$ leads to $u = \bar{z} - a_1 v$ as the sampled value of $\mathcal{U}$.*

The main question we aim to answer is how the consideration of the expected cost of parameter uncertainty affects the decision maker in assessing the need for modeling the correlation. To this end, we use the same experimental design in Table 4.5 and compare the operational test and the correlation test in terms of the fraction of time the decision maker rejects the null hypothesis of no correlation. Table 4.6 presents the fraction of time that the decision maker rejects the independence assumption at the $\alpha = 0.05$ level when the cost ratio $\gamma_0 = h/(h+b)$ is equal to 0.1.

Figure 4.4: The operational test with $\mathcal{H}_0 : r = 0$ and $\mathcal{H}_{\text{A}} : r \neq 0$

Initialization. Specify the number of bootstrap samples (B), the length of historical data (n), the number of positive realizations (n_d), the number of periods since last demand (y), and the significance level $(1-\alpha)$.

Set $\hat{p} \leftarrow n_d/n$ and $b \leftarrow 1$.

Compute the test statistic $\hat{\Lambda} := \Lambda(\hat{\kappa}_\gamma, 0, 1, y; \gamma_o, \hat{p}, 0)$ from historical data $\{(x_t, y_t); t = 1, 2, \ldots, n_d\}$.

while $b \leq B$:

Step 1. Generate the bth bootstrap sample path as in Step 1 of the correlation test.

Step 2. *Compute the test statistic from the bootstrap sample path*:

Set $\hat{p}^b \leftarrow n_d^b/n$.

Let $\bar{z}$ and s_z be the sample mean and sample standard deviation of $\{z_1^b, z_2^b, \ldots, z_{n_d}^b\}$.

Set $\hat{r}^b \leftarrow \arg\max_{r \in (-1,1)} \sum_{i=1}^{n_d^b} \log \phi_2\{\Phi^{-1}(F((z_i - \bar{z})/s_z; 0, 1)), \Phi^{-1}(G(y_i^b; \hat{p}^b)); r\}$.

Set κ_γ^b to the value of κ that solves the equation $\Pi(\kappa; \hat{p}^b, y, \hat{r}^b) = 1 - \gamma/\hat{p}^b$.

Set $u^b \leftarrow \bar{z} - a_1 v^b$ and $v^b \leftarrow s_z/\sqrt{a_2}$.

$\hat{\Lambda}^b \leftarrow \Lambda(\hat{\kappa}_\gamma, u^b, v^b, y; \gamma_o, \hat{p}^b, 0)$.

$b \leftarrow b + 1$

end

Step 3. *Construct the critical region to reject the null hypothesis*:

Sort $\hat{\Lambda}^b$, $b = 1, 2, \ldots, B$ in ascending order $(\Lambda_{(1)}, \Lambda_{(2)}, \ldots, \Lambda_{(B)})$.

Reject the null hypothesis $\mathcal{H}_0$ if $\hat{\Lambda} > \Lambda_{(\lceil (1-\alpha)B \rceil)}$.

We observe that the fraction of time the decision maker rejects the independence assumption is considerably smaller in Table 4.6 compared to Table 4.5 especially when the length of the demand history and the strength of the correlation are not very high. This can be explained by the additional statistical estimation error around the copula parameter when the independence assumption is relaxed. To put it another way, the benefit of accounting for the correlation is dominated by the additional expected cost associated with the incorrect estimation of the copula parameter. This is why the operational test, which considers the expected cost of parameter uncertainty, is in favor of the simpler model with no correlation when the amount of data is limited and the strength of the correlation is low. Not surprisingly, this is the case when the the copula-parameter estimator has the highest variance, and hence, when the estimated value of correlation is subject to highest statistical estimation error.

We also observe that the difference between the inventory holding and backlogging costs – which is ignored by the correlation test – plays an important role when the expected cost of parameter uncertainty is the criterion to decide whether to model the correlation. To illustrate, we present the fraction of time that the decision maker rejects the independence assumption in Table 4.7 when γ_0 is equal to 0.25. Clearly, the fraction of time the decision maker rejects the independence assumption is smaller in Table 4.7 compared to Table 4.6 (bottom). In these tables, the values of n and n_d are both the same, while the difference between the ratio of n_d to n and the value of γ_0 is smaller in Table 4.7. The decision maker is more likely to set a zero inventory target when the ratio of n_d to n is not sufficiently bigger than γ_0, leading to smaller values for the probability of rejecting the independence assumption in Table 4.7. Intuitively, the decision maker is less likely to bother with modeling the correlation since the inventory target is set to zero anyway. Indeed, we observe that the fraction of time the decision maker rejects the independence assumption in the operational test is

Table 4.6: The fraction of time the operational test rejects $\mathcal{H}_0$ at 95% significance level for $\gamma_o = 0.1$

		r					
n	n_d	0.15	0.3	0.45	0.6	0.75	0.90
30	6	0%	0%	0%	0%	0%	2%
50	10	0%	0%	0%	0%	3%	5%
100	20	0%	1%	3%	9%	21%	27%
250	50	3%	7%	26%	33%	46%	69%
30	12	0%	1%	3%	5%	8%	28%
50	20	0%	3%	8%	13%	35%	45%
100	40	0%	8%	28%	46%	49%	58%
250	100	0%	15%	51%	63%	73%	84%

Table 4.7: The fraction of time the operational test rejects $\mathcal{H}_0$ at 95% significance level for $\gamma_o = 0.25$

		r					
n	n_d	0.15	0.3	0.45	0.6	0.75	0.90
30	12	0%	0%	0%	1%	3%	5%
50	20	0%	0%	2%	5%	11%	14%
100	40	0%	2%	5%	13%	41%	53%
250	100	1%	10%	35%	69%	72%	81%

almost zero in all the cases when the ratio of n_d to n is 20% with $\gamma_0 = 0.25$ and when the ratio of n_d to n is 20% and 40% with $\gamma_0 = 0.5$.

A natural question to ask is which statistical test to use for investigating the existence of correlation in an intermittent demand history. The correlation test considers the copula parameter in isolation, and therefore, it is more appropriate if the goal is merely to make an inference about the intermittent demand process. On the other hand, the operational test incorporates the expected cost of incorrectly estimating the copula parameter into the decision of modeling the correlation. Thus, the operational test is more pragmatic to consider when the question is whether the copula-parameter estimate should be used as an input in decision making, as is the case in this study.

4.4.4 Improved Inventory-Target Estimation in the Presence of Correlation

Once the need to model the correlation is warranted, an immediate question is whether an alternative threshold can be used to reduce the expected cost of parameter uncertainty. We briefly address this question in the remainder of the section.

The copula-based representation of intermittent demand makes it very convenient to sample intermittent demand histories, lending itself to the use of Sample Average Approximation (SAA) to find an improved threshold. More specifically, the difference between the expected costs of the optimal inventory target and the estimate of the optimal inventory target – which we characterize as $\theta\Lambda(\hat{\kappa}_\gamma, \mathcal{U}, \mathcal{V}, y; \gamma_o, p, r)$ in Section 4.4.3 – is independently sampled a large number of times and added up to obtain the objective function. The so-called SAA objective function is a deterministic function of the threshold γ. Therefore, any numerical minimization method suitable for a deterministic problem can be applied to approximate the optimal threshold. The method provides consistent estimates of the optimal threshold and the minimum expected cost of parameter uncertainty. We refer the reader to Shapiro et al. (2009) for details.

Without loss of generality, the decision maker can generate the independent sample paths by assuming the scale parameter θ as one. This is because the solution of the SAA objective function does not depend on the value of θ. However, it is necessary to specify the copula parameter r and the demand-occurrence probability p to generate the sample paths. In this case, the sampling distribution of the copula-parameter estimator can be used to obtain the confidence-interval information for the copula parameter r – in addition to the confidence-interval information for the demand-occurrence probability p obtained in Section 4.3. If the statistical tests are not in favor of rejecting the independence assumption, then the decision maker can directly use the improved inventory-target estimation proposed in Section 4.3.

4.5 Conclusion

We consider the inventory-target estimation problem for a single item whose demand history consists of many zero observations because demand does not arrive every review period. We first establish the optimal policy under the assumptions of known demand-occurrence probability and known mean and variance for the demand size: If demand-occurrence probability is less than a threshold, do not keep inventory; otherwise, set a positive inventory target that minimizes the expected holding and shortage costs in the forthcoming period. It is well known that the implementation of this policy – which is only optimal for known parameters – may result in poor operational performance when the parameter estimates are used as if they were the true values.

We first quantify the expected cost of parameter uncertainty as a function of a threshold variable when the demand-size distribution belongs to a location-scale family. We illustrate this quantification for exponentially and normally distributed demand sizes. We then discuss the minimization of the expected cost of parameter uncertainty with respect to the threshold variable. Our approach, thus, combines inventory management and parameter estimation into a single task, and effectively balances under- and overestimation of the optimal inventory target by considering the interplay between holding and backlogging costs. For example, our method suggests using a threshold equal to one, when the demand is highly intermittent and the holding cost dominates the backlogging cost. In this case, the decision maker mimics the optimal policy for known parameters by not keeping any stock regardless of the realized demand history. The optimized threshold variable minimizes the expected cost by accounting for all the possible realizations of the demand history, and hence all the realizations of the demand parameter estimators, in a frequentist framework.

Motivated by our analysis of industrial data and the previous literature on empirical analysis of intermittent demand, we also present a copula-based model to capture the correlation between demand size and the number of zero-demand periods that precede the demand. The main question is when we need to model the correlation despite the statistical estimation error around the copula parameter. The copula-based representation of intermittent demand allows us to develop two new hypothesis tests to assess the existence of correlation in a limited amount of intermittent demand data. We show that a statistical test which accounts for the expected cost of parameter uncertainty tends to reject the independence assumption less frequently than a statistical test which only considers the sampling distribution of the copula-parameter estimator.

In this study, we assume that the intermittent demand process is stationary and the decision maker uses all the demand history for inventory-target estimation. However, when demand is nonstationary, the expected cost of parameter uncertainty can be smaller when only the most recent demand realizations are used in inventory-target estimation. Although we do not address this issue here, our joint estimation-optimization approach is the first step in search of a solution to this important operational problem and it is the subject of future work.

Chapter 5

Data-Driven Newsvendor: A Probabilistic Guarantee for Near Optimality when Demand is Temporally Dependent

5.1 Introduction

An important problem in inventory management is to set inventory targets in the absence of complete information about the true demand generating process. In this paper, we address this problem in a newsvendor setting for stationary and temporally dependent demand. It is well known that the *critical fractile solution* is optimal when the parameters that characterize the distribution of the demand conditional on past demand realizations are known. In practice, however, these parameters are unknown and must be estimated using only a finite (and sometimes, very limited) amount of historical demand data. Consequently, the expected cost associated with an *estimate* of the critical fractile solution can be far from the minimum but unknown expected cost.

The dependence of demand on its past realizations is not uncommon in practice. Erkip et al. (1990) find that the autocorrelation in monthly demand can reach up to 0.70 in a consumer products company. Lee et al. (1997) report dependence between demand realizations over time – especially positive autocorrelation – in high-tech and grocery industries. Lee et al. (2000) show that the sales data of 91% of the items in a supermarket have autocorrelations between 0.26 and 0.89. Similarly, Hosoda et al. (2008) analyze the sales data of soft drink products at a grocery retailer and find that the autocorrelation in weekly demand varies between 0.77 and 0.83. The dependence of random variables on past realizations is also frequent in areas other than inventory management. The daily concentrations of pollutants collected and analyzed in an environmental health project (Peng and Dominici, 2008), the energy generated from a wind farm (Kim and Powell, 2011), and the number of beds demanded in a hospital (Kros and Brown, 2012) are some examples.

A widely used method to model autocorrelation is to construct the temporally dependent process via classical time series. For example, the linear autoregressive (AR) process is often used for demand modeling in inventory management and supply chain analysis (e.g., Lee et al. 2000, Luong 2007, Chen and Lee 2012). It is well known that an AR process can be expressed as a linear combination of independent and normally distributed random shocks. This implies that the marginal demand distribution is normal, which is an assumption vastly made in theory but often violated in practice. In particular, a normal distribution often falls short of an adequate representation of the demand distribution, leading to inaccurate prediction models and poor operational performance. We refer the reader to Akcay et al. (2011) and Ramamurthy et al. (2012) as the examples illustrating the importance of the demand's distributional shape in estimating inventory targets from independent and identically distributed demand data.

For autocorrelated demand, as in this paper, a considerable amount of effort has been devoted to modeling time series with exponential, gamma, geometric, or general discrete marginal distributions (e.g., Block et al. 1990, Tiku et al. 2002, Akkaya and Tiku 2005, Gourieroux and Jasiak 2006, and Jose et al. 2008). Nevertheless, these models often allow only limited control of the dependence structure for a given marginal demand distribution. In addition, there is no one-size-fit-all solution for modeling and estimation of time-series processes. We overcome these challenges by (i) modeling the temporal dependence in a transformed demand process which has a standard uniform marginal distribution and matches the underlying dependence structure of the actual demand process, and (ii) capturing the demand's distributional shape with the empirical demand distribution function. To be specific, we model the temporal dependence in $\{F(X_t); t = 1, 2, \ldots\}$ rather than modeling the temporal dependence in the actual demand process $\{X_t; t = 1, 2, \ldots\}$, where F is the marginal cumulative distribution function (cdf) of the demand process. We do this by using a *copula* which allows a decision maker to avoid any restrictive assumption on the functional form of F. To estimate the critical fractile solution, the decision maker first obtains the empirical demand distribution from the temporally dependent demand data and then uses it in lieu of the true marginal demand distribution while estimating the copula parameters that characterize the temporal dependence. Finally, an estimate of the critical fractile solution is obtained as a function of the empirical demand distribution and the estimated values of the copula parameters.

A natural question to ask is as follows. *How good is an inventory target estimated in this way?* We define the *goodness* of an inventory-target estimate as in Levi et al. (2007) by using the notion of ε-optimality; i.e., an inventory-target estimate is ε-optimal if its expected cost is at most $1 + \varepsilon$ of the minimum but unknown expected cost. Clearly, the ε-optimality of an inventory-target estimator is a random event because the inventory-target estimator is a random variable as a result of being a function of the historical data randomly generated by the true demand process. In this paper, we consider all the possible realizations of a demand history, and hence, all the realizations of an inventory-target estimator in a frequentist framework. Accordingly, we obtain a lower bound on the probability of ε-optimality when the inventory target is a function of the empirical demand distribution and copula-parameter estimators. This lower bound, which we refer as *a probabilistic guarantee for near optimality*, serves as a level of confidence for the decision maker to assure the ε-optimality of an inventory target obtained from a limited amount of demand data.

We summarize our contributions in this paper as follows:

1. To the best of our knowledge, we are the first to investigate the ε-optimality (i.e., near optimality) guarantee of an inventory target estimated from temporally dependent demand data.

2. We introduce a copula-based model that allows the decision maker to assume a demand process with an arbitrary marginal demand distribution and temporal-dependence structure. Autocorrelation is a measure of linear dependence and it may fall short of capturing the real-world dynamics in historical demand data. Our copula-based demand model does not suffer from this limitation. We illustrate it by accounting for the tail dependence – the amount of dependence between the very low or very high valued demand realizations of the two consecutive time periods.

3. For the special case of normal copula, we show that the probabilistic guarantee for ε-optimality decreases with the strength of the autocorrelation. We then provide a lower bound on the number of demand observations necessary to achieve a certain level of probabilistic guarantee when the marginal demand

distribution is known. Furthermore, we provide an upper bound on the number of demand observations – as a function of the copula parameter – such that the decision maker achieves a higher probabilistic guarantee of ε-optimality by simply ignoring the temporal dependence.

4. We propose a sampling-based method to compute a lower bound to the probability of ε-optimality for an arbitrary choice of copula when the marginal demand distribution is unknown to the decision maker. In particular, our method builds on the idea of sampling dependent uniform random variates matching the underlying dependence structure of the demand process – rather than sampling the actual demand which requires the specification of the marginal distribution.

The remainder of the paper is organized as follows. Section 5.2 presents our copula-based demand model and the solution of the newsvendor problem with complete information about the temporally dependent demand process. Section 5.3 provides a two-step estimation method consistent with the marginal-copula representation of demand. Section 5.4 characterizes the ε-optimality for temporally dependent demand, provides structural results for normal copula, and presents our sampling-based method to compute a probabilistic guarantee for ε-optimality. Section 5.5 aims to understand the driving factors behind the probabilistic guarantee, the value of perfect information about autocorrelation, and the value of relaxing the independence assumption. Section 5.6 investigates the impact of tail dependence on inventory-target estimation. Section 5.7 concludes the paper with a summary of findings and future research directions. We provide all the proofs in an Appendix at the end of the chapter.

5.2 The Modeling Framework

Section 5.2.1 introduces our copula-based demand model. Section 5.2.2 presents an overview of the newsvendor model and characterizes the optimal critical-fractile solution when the marginal demand distribution and the temporal dependence structure are known.

5.2.1 Demand Process

We let the demand $\{X_t; t = 1, 2, \ldots\}$ be a stationary first-order Markov process; i.e.,

$$\mathbb{P}\left(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_1 = x_1\right) = \mathbb{P}\left(X_{t+1} = x_{t+1} | X_t = x_t\right).$$

In contrast to using a conditional probability distribution directly, we consider an alternative approach based on *copulas* to model the first-order temporal dependence. We do this by constructing the joint distribution of transformed demand random variables $U_t = F(X_t)$ and $U_{t+1} = F(X_{t+1})$ by a bivariate copula $C(\cdot, \cdot; \boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where the parameter space $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^d$ and d is the number of parameters characterizing the underlying dependence structure.

Definition 5.2.1 (Nelsen 2006)**.** *A bivariate copula is a function* $C : [0, 1] \times [0, 1] \to [0, 1]$*; i.e., a mapping of the unit square into the unit interval with the following properties: (i) For every* (u_t, u_{t+1}) *in* $[0, 1] \times [0, 1]$, $C(u_t, 0; \boldsymbol{\theta}) = C(0, u_{t+1}; \boldsymbol{\theta}) = 0$, *and* $C(u_t, 1; \boldsymbol{\theta}) = u_t$ *and* $C(1, u_{t+1}; \boldsymbol{\theta}) = u_{t+1}$*; (ii) For every* $u'_t, u''_t, u'_{t+1}, u''_{t+1}$ *in* $[0, 1]$ *such that* $u'_t \leq u''_t$ *and* $u'_{t+1} \leq u''_{t+1}$, $C(u''_t, u''_{t+1}; \boldsymbol{\theta}) - C(u''_t, u'_{t+1}; \boldsymbol{\theta}) - C(u'_t, u''_{t+1}; \theta) + C(u'_t, u'_{t+1}; \boldsymbol{\theta}) \geq 0$.

The first condition provides the lower bound on the distribution function and ensures that a bivariate copula is a two-dimensional distribution function with standard uniform marginal distributions. The second condition guarantees the probability of observing a point in $[0,1]\times[0,1]$ to be nonnegative. The use of copulas for representing the joint distribution of a random vector has been studied extensively for the past two decades. We refer the reader to Joe (1997) and Nelsen (2006) for the widely known properties of the copulas. The use of copulas for modeling the temporal dependence of time series data has recently gained attention; see Chen and Fan (2006), Beare (2010), and Patton (2012) as the example studies.

Sklar's theorem (Sklar, 1959) and the stationarity of the demand process allow us to construct the distribution of demand random variables X_t and X_{t+1} with a marginal distribution function F and a copula $C(\cdot,\cdot;\boldsymbol{\theta})$:

Theorem 5.2.1 (Sklar's Theorem.)**.** *Let H be a bivariate distribution function with the continuous marginal cdf F. Then, there exists a bivariate unique copula C such that $H(x_t,x_{t+1}) = C(F(x_t),F(x_{t+1});\boldsymbol{\theta})$ for all $(x_t,x_{t+1}) \in \mathbb{R}^2$. Conversely, if C is a bivariate copula and the marginal cdf F is continuous, then the function $C(F(x_t),F(x_{t+1});\boldsymbol{\theta})$ is a bivariate distribution function with marginal cdf F.*

The advantage of using copula for demand modeling is that we have the freedom to specify the marginal demand distribution and the dependence structure separately. That is, we can choose any arbitrary continuous demand distribution, link the consecutive demand random variables with a copula, and obtain a legitimate bivariate distribution to characterize the first-order time series. We only impose the following technical assumption on the dependence structure:

Assumption 5.2.1. *The copula $C(u_t,u_{t+1};\boldsymbol{\theta})$ is absolutely continuous with respect to Lebesgue measure on $[0,1]\times[0,1]$, and is neither the Fréchet-Hoeffding upper bound (i.e., $C(u_t,u_{t+1};\boldsymbol{\theta}) \neq \min(u_t,u_{t+1})$) nor the Fréchet-Hoeffding lower bound (i.e., $C(u_t,u_{t+1};\boldsymbol{\theta}) \neq \max(u_t+u_{t+1}-1,0)$).*

Assumption 5.2.1 is standard in the context of dependence modeling to rule out the deterministic cases of $X_t = X_{t-1}$ for the upper bound and $X_t = F^{-1}(1-F(X_{t-1}))$ for the lower bound.

There are relatively few applications of copulas in operations management; Clemen and Reilly (1999), Cooper and Gupta (2006), Corbett and Rajaram (2006), and Aydın et al. (2012) are examples in decision analysis, revenue management, and supply chain modeling. The common characteristic of these papers is that they all use copulas to model the joint distribution of random vectors that are independent over time. In this study, we use copulas to model the temporal dependence in a demand process for the first time in operations management. We illustrate our method of demand modeling in Example 5.2.1 by using the *normal copula*, which encodes the dependence precisely the same way a bivariate normal distribution does; we refer the reader to Joe (1997) for other copulas that enable us to capture any form of dependence structure.

Example 5.2.1 (Normal Copula)**.** *Let Φ be the standard normal cdf and $\Phi_2(\cdot,\cdot;\theta)$ be the standard normal bivariate cdf with correlation $\theta \in (-1,1)$. The normal copula is defined as $C(u_t, u_{t+1};\theta) = \Phi_2(\Phi^{-1}(u_t), \Phi^{-1}(u_{t+1});\theta)$. As a result of representing a standard normal random variable Z_t as $\Phi^{-1}(U_t)$ from a standard uniform random variable $U_t = F(X_t)$, it can be easily seen that normal copula constructs a valid joint distribution for demand random variables X_t and X_{t+1}:*

$$\begin{aligned}
\Phi_2\left(\Phi^{-1}(u_t), \Phi^{-1}(u_{t+1});\theta\right) &= \mathbb{P}\left(Z_t \le \Phi^{-1}(u_t), Z_{t+1} \le \Phi^{-1}(u_{t+1})\right) \\
&= \mathbb{P}\left(\Phi^{-1}(U_t) \le \Phi^{-1}(u_t), \Phi^{-1}(U_{t+1}) \le \Phi^{-1}(u_{t+1})\right) \\
&= \mathbb{P}\left(U_t \le u_t, U_{t+1} \le u_{t+1}\right) \\
&= \mathbb{P}\left(F(X_t) \le F(x_t), F(X_{t+1}) \le F(x_{t+1})\right) \\
&= \mathbb{P}\left(X_t \le x_t, X_{t+1} \le x_{t+1}\right)
\end{aligned}$$

The last equation follows because an inequality still holds after applying a monotonically increasing function to both sides of the inequality.

It is worth noting that the dependence structure of a normal copula implies that the demand process $\{X_t; t = 1, 2, \ldots\}$ accepts an Autoregressive-To-Anything (ARTA) process representation of Cario and Nelson (1996). In Example 5.2.1, the ARTA process first takes an AR(1) model $Z_t = \theta Z_{t-1} + Y_t$ with standard normal Z_t and normally distributed independent error term Y_t with mean zero and variance $1 - \theta^2$, and then obtains the demand random variable X_t via the transformation $X_t = F^{-1}(\Phi(Z_t))$. Clearly, the use of normal copula together with normal marginal distribution further reduces the demand process to a classical AR model.

There is a one-to-one mapping between the normal copula parameter θ – which is the autocorrelation coefficient in the AR(1) model described above – and the autocorrelation of the demand process $\{X_t; t = 1, 2, \ldots\}$. We refer reader to Cario and Nelson (1996) for the details of so-called correlation matching problem. In this paper, we focus on estimating the copula parameter directly.

5.2.2 Newsvendor Model

The decision maker aims to set the correct number of units in stock to meet the unknown demand in period $n + 1$. The period starts with zero inventory on hand. Ordering too few incurs a shortage cost of b per unit short, while ordering too many incurs a holding cost of h per unit over. The goal is to minimize the sum of expected shortage and holding costs *conditional* on the most recent demand realization since the demand is known to be a first-order Markov process.

The resulting objective function is convex and minimized by the critical fractile solution (Porteus, 2002). The critical fractile solution requires the knowledge of the distribution of demand X_{n+1} conditional on $X_n = x_n$, which can be derived directly from the copula as follows:

$$
\begin{aligned}
\mathbb{P}(X_{n+1} \leq x_{n+1} | X_n = x_n) &= \mathbb{P}(U_{n+1} \leq u_{n+1} | U_n = u_n) \\
&= \lim_{\delta \to 0} \frac{\mathbb{P}(U_{n+1} \leq u_{n+1}, U_n \in (u_n - \delta, u_n + \delta))}{\mathbb{P}(U_n \in (u_n - \delta, u_n + \delta))} \\
&= \lim_{\delta \to 0} \frac{C(u_n + \delta, u_{n+1}; \boldsymbol{\theta}) - C(u_n - \delta, u_{n+1}; \boldsymbol{\theta})}{2\delta} \\
&= \left. \frac{\partial C(u_n, F(x_{n+1}); \boldsymbol{\theta})}{\partial u_n} \right|_{u_n = F(x_n)}
\end{aligned}
$$

We denote this conditional distribution by $C_{2|1}\left(F(\cdot)|F(x_n); \boldsymbol{\theta}\right)$. The critical fractile solution q^* is then the value of q that solves the first-order condition

$$
C_{2|1}\left(F(q)|F(x_n); \boldsymbol{\theta}\right) - \frac{b}{h+b} = 0, \tag{5.1}
$$

and it can be written as $q^* = F^{-1}(C_{2|1}^{-1}\left(b/(h+b)|F(x_n); \boldsymbol{\theta}\right))$. We next discuss a two-step method to estimate the unknown marginal distribution function F and the d-dimensional copula-parameter vector $\boldsymbol{\theta}$.

5.3 Model Estimation

The marginal-copula representation of the demand process allows the marginal distribution function and copula parameters to be estimated separately, potentially with different methods. In this study, we focus on the case where the decision maker estimates the marginal distribution function with the empirical demand distribution. Subsequently, the copula parameters are estimated via the maximum likelihood method by treating the empirical demand distribution as the true marginal distribution. This method is known as *semiparametric estimation* as well as *canonical maximum likelihood estimation* (Patton, 2012).

The joint density function of the demand random variables X_t and X_{t+1} is given by

$$\frac{\partial^2 C(F(x_t), F(x_{t+1}); \boldsymbol{\theta})}{\partial x_t \partial x_{t+1}} = f(x_t) f(x_{t+1}) c\left(F(x_t), F(x_{t+1}); \boldsymbol{\theta}\right),$$

where $f(\cdot)$ is the marginal probability density function (pdf) and $c(u, v; \boldsymbol{\theta}) := \partial^2 C(u, v; \boldsymbol{\theta})/(\partial u \partial v)$ is the copula density function. Therefore, the density function of X_{t+1} conditional on $X_t = x_t$, which we denote with $f_{2|1}(x_{t+1}|x_t)$, reduces to $f(x_{t+1}) c(F(x_t), F(x_{t+1}); \boldsymbol{\theta})$. The likelihood function of the historical demand data $\{x_t; t = 1, 2, \ldots, n\}$ then takes the form

$$f(x_1) \prod_{t=1}^{n-1} f_{2|1}(x_{t+1}|x_t) = \prod_{t=1}^{n} f\left(x_t\right) \prod_{t=1}^{n-1} c\Big(F(x_t), F(x_{t+1}); \boldsymbol{\theta}\Big). \tag{5.2}$$

It is worth noting that the copula density function of independent random variables takes the value of one and the likelihood function in (5.2) reduces to the likelihood function $\prod_{t=1}^{n} f\left(x_t\right)$ of independent and identically distributed data.

In the first step of semiparametric estimation, the decision maker estimates the marginal distribution function $F(x)$ by using the empirical demand distribution function $F_n(x) := (1/n) \sum_{t=1}^{n} \mathbb{1}(X_t \leq x)$, where $\mathbb{1}(\cdot)$ is the indicator variable taking the value of 1 if $\cdot$ is true and zero otherwise. In the second step, the copula-parameter vector $\boldsymbol{\theta}$ is estimated by maximizing the log-likelihood function after replacing the unknown F by F_n and ignoring the terms that do not depend on $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{t=1}^{n-1} \log c\Big(F_n(x_t), F_n(x_{t+1}); \boldsymbol{\theta}\Big).$$

Finally, the decision maker estimates the critical fractile solution q^* by replacing the marginal distribution function F by F_n and the copula parameters $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ in the functional form of q^*. The inventory-target estimator, which is a function of the demand random variables $\{X_t; t = 1, 2, \ldots, n\}$, then takes the form

$$\hat{Q} = F_n^{-1}\left(C_{2|1}^{-1}\left(\frac{b}{h+b}\middle| F_n(X_n); \hat{\boldsymbol{\theta}}\right)\right), \tag{5.3}$$

where

$$F_n^{-1}(\tau) = \min_{j=1,2,\ldots,n}\left\{X_j : \frac{1}{n}\sum_{t=1}^{n}\mathbb{1}(X_t \leq X_j) \geq \tau\right\}.$$

To the best of our knowledge, Chen and Fan (2006) is the first to show the consistency of the semiparametric copula-parameter estimator $\hat{\boldsymbol{\theta}}$. We next present this result as it plays a crucial role in establishing the consistency of the inventory-target estimator in (5.3).

We first define $\mathscr{F}$ as the space of probability distributions over the support of X_t. For any $F_0 \in \mathscr{F}$, we then let $\|F_0 - F\|_{\mathscr{F}} = \sup_x |(F_0(x) - F(x))/w(F(x))|$ with $w(\cdot)$ a weighting function satisfying the conditions in Lemma 4.1 of Chen and Fan (2006). In addition, we let $\mathscr{F}_\lambda = \{F_0 \in \mathscr{F} : \|F_0 - F\|_{\mathscr{F}} \leq \lambda\}$ for a small $\lambda > 0$. For any $\boldsymbol{\theta_0} \in \Theta$, we use $\|\boldsymbol{\theta_0} - \boldsymbol{\theta}\|$ to denote the usual Euclidean metric.

Proposition 5.3.1 (Chen and Fan 2006)**.** *Let* $\ell(u_1, u_2; \boldsymbol{\theta})$ *denote* $\log c(u_1, u_2; \boldsymbol{\theta})$*, and represent* $\partial\ell(u_1, u_2; \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ *and* $\partial^2\ell(u_1, u_2; \boldsymbol{\theta})/(\partial u_j \partial\boldsymbol{\theta}), j = 1, 2$ *with* $\ell_{\boldsymbol{\theta}}(u_1, u_2; \boldsymbol{\theta})$ *and* $\ell_{\boldsymbol{\theta},j}(u_1, u_2; \boldsymbol{\theta}), j = 1, 2$*, respectively. Suppose that the following conditions hold:*

i. $\mathbb{E}(\ell_{\boldsymbol{\theta}}(U_{t-1}, U_t; \boldsymbol{\theta})) = 0$*, and* $\ell_{\boldsymbol{\theta}}(U_{t-1}, U_t; \boldsymbol{\theta})$ *is Lipschitz continuous at* $\boldsymbol{\theta}$ *for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*.*

ii. $\ell_{\boldsymbol{\theta}}(u_1, u_2; \boldsymbol{\theta})$ *and* $\ell_{\boldsymbol{\theta},j}(u_1, u_2; \boldsymbol{\theta})$*,* $j = 1, 2$ *are well-defined for* $(u_1, u_2, \boldsymbol{\theta}) \in (0, 1) \times (0, 1) \times \boldsymbol{\Theta}$*.*

iii. $\{X_t; t = 1, 2, \ldots\}$ *is* β*-mixing with a decay rate of* $\beta_t = O(t^{-b})$ *for* $b > 0$*.*

iv. $\mathbb{E}(\sup_{\boldsymbol{\theta_0} \in \boldsymbol{\Theta}} \|\ell_{\boldsymbol{\theta}}(U_{t-1}, U_t; \boldsymbol{\theta_0})\| \log(1 + \|\ell_{\boldsymbol{\theta}}(U_{t-1}, U_t : \boldsymbol{\theta_0})\|)) < \infty$*.*

v. $\mathbb{E}(\sup_{\boldsymbol{\theta_0} \in \boldsymbol{\Theta}, F \in \mathscr{F}_\lambda} \|\ell_{\boldsymbol{\theta},j}(F(X_{t-1}), F(X_t); \boldsymbol{\theta_0})\| w(U_{t-2+j})) < \infty$ *for* $j = 1, 2$*.*

Then, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| = o_p(1)$. *That is,* $\mathbb{P}(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| > \epsilon)$ *approaches zero for every* $\epsilon > 0$ *as* n *tends to infinity.*

The first two conditions are standard regularity conditions. The third condition requires the demand process to be β-mixing with a polynomial decay rate. The assumption that $c(u_1, u_2; \boldsymbol{\theta})$ is positive on $(0, 1) \times (0, 1)$ ensures that any process satisfying Assumption 1 is β-mixing with at least a polynomial decay rate; we refer the reader to Chen and Fan (2006) for the definition of β-mixing and further details. For example, the time series $\{X_t; t = 1, 2, \ldots\}$ generated by a normal copula is β-mixing with an exponential decay rate regardless of its marginal distribution. The fourth condition is a moment condition on the score function, while the last condition states that the weighted partial derivatives of the score function must be dominated by a function with a finite first moment.

The asymptotic behavior of the semiparametric copula-parameter estimator and the convergence of the empirical demand distribution to the true marginal distribution imply that, when there is a large number of demand observations, the inventory-target estimator $\hat{Q}$ approaches the critical fractile solution with full knowledge of copula parameters and marginal distribution function. In practice, a decision maker is rarely fortunate enough to observe a large number of demand observations. In contrast, the demand history can be very short, casting doubt on the performance of the inventory-target estimator $\hat{Q}$ in minimizing the sum of expected inventory holding and shortage costs. In this paper, we mainly focus on this expected cost associated with $\hat{Q}$ for *finite* number of demand observations, which is a more relevant measure for the decision maker than the asymptotic properties of the estimators F_n and $\hat{\boldsymbol{\theta}}$.

5.4 ε-Optimality in Inventory-Target Estimation

We let $\hat{q}$ denote the realization of the inventory-target estimator $\hat{Q}$ from the historical demand data $\{x_t; t = 1, 2, \ldots, n\}$ of length n. The expected cost associated with $\hat{q}$ is given by

$$L(\hat{q}|x_n) := h \int_{-\infty}^{\hat{q}} (\hat{q} - x_{n+1}) f_{2|1}(x_{n+1}|x_n) \mathrm{d}x_{n+1} + b \int_{\hat{q}}^{\infty} (x_{n+1} - \hat{q}) f_{2|1}(x_{n+1}|x_n) \mathrm{d}x_{n+1}.$$

The decision maker does not necessarily achieve the minimum expected cost $L(q^*|x_n)$ by using the inventory-target estimate $\hat{q}$ especially when n is small. In this study, we are interested in finding a lower bound to the probability of the difference $L(\hat{q}|x_n) - L(q^*|x_n)$ not exceeding a certain threshold for *all* possible realizations of the demand history. To this end, we let $0 < \varepsilon \leq 1$ and define the ε-optimality in inventory-target estimation as follows:

Definition 5.4.1. *The inventory-target estimate $\hat{q}$ is ε-optimal if its expected cost is at most $1 + \varepsilon$ of the minimum expected cost; i.e., $L(\hat{q}|x_n) \leq (1 + \varepsilon) L(q^*|x_n)$.*

Considering all possible realizations of the demand observation in period n, we let Q^* denote the critical fractile solution which has the functional form $F^{-1}(C_{2|1}^{-1}(b/(h+b)|F(X_n); \boldsymbol{\theta}))$. We aim to provide a lower bound to the probability of the event

$$\left[L(\hat{Q}|X_n) \leq (1 + \varepsilon)\, L(Q^*|X_n) \right]$$

to measure the quality of the inventory-target estimator $\hat{Q}$ in a frequentist framework. If this probability is sufficiently large, then $\hat{Q}$ can be used confidently even though it is estimated from finite number of demand observations. This is because the expected cost of its realization (before the realization of demand in period $n + 1$) cannot be more than $1 + \varepsilon$ of the minimum but unknown expected cost with high probability for *any* historical demand data.

We organize the remainder of this section as follows. Section 5.4.1 generalizes the characterization of ε-optimality for temporally dependent demand. Section 5.4.2 considers the special case of normal copula to provide analytical insights shedding light on the roles of temporal dependence and the length of demand history in inventory-target estimation. Section 5.4.3 switches our focus back to an arbitrary choice of copula and proposes a sampling-based method to compute a lower bound to the probability of ε-optimality.

5.4.1 The ε-Optimality for Temporally Dependent Demand Data

The ε-optimality in inventory-target estimation is introduced by Levi et al. (2007) under the assumption of *independent* and identically distributed demand data. We start with generalizing the concept of ε-optimality for stationary and temporally dependent demand data.

Proposition 5.4.1. *Let $C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}) \geq b/(h+b)-\alpha$ and $C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}) \leq b/(h+b)+\alpha$. The expected cost of the inventory-target estimate $\hat{q}$ is at most $1+\varepsilon$ of the expected cost of the optimal inventory target q^* when α is equal to $(\varepsilon hb/(h+b))/(h+b+\varepsilon\max(b,h))$.*

Proposition 5.4.1 characterizes the value of α so that an inventorytarget estimate $\hat{q}$ is *guaranteed* to be ε-optimal. Therefore, if the value of α is taken as characterized in Proposition 5.3.1, the probability of the event $Y(\hat{Q}, X_n)$ defined as

$$\left[C_{2|1}\left(F(\hat{Q})|F(X_n);\boldsymbol{\theta}\right) \geq \frac{b}{h+b}-\alpha\right] \bigcap \left[C_{2|1}\left(F(\hat{Q})|F(X_n);\boldsymbol{\theta}\right) \leq \frac{b}{h+b}+\alpha\right]$$

gives a *lower bound* to the probability of the ε-optimality of the inventory-target estimator $\hat{Q}$. We denote the probability of the event $Y(\hat{Q}, X_n)$ with δ for ease in exposition in the next section, where we investigate the behavior of this probabilistic guarantee for ε-optimality as a function of the copula parameters and the length of the demand history.

Remark 5.4.1. *Levi et al. (2007) set α to $(\varepsilon/3)\min(b,h)/(h+b)$ when the historical demand data is independent and identically distributed. The lower bound on the probability of ε-optimality is higher when α is set as in Proposition 5.4.1, leading to a tighter bound. This is because $(\varepsilon hb/(h+b))/(h+b+\varepsilon\max(b,h))$ is always greater than $(\varepsilon/3)\min(b,h)/(h+b)$.*

5.4.2 Analytical Insights: Normal Copula

The objective of this section is to provide analytical insights that highlight the relation between the lower bound δ on the probability of ε-optimality, i.e.,

$$\mathbb{P}\left\{L(\hat{Q}|X_n) \leq (1+\varepsilon)\, L(Q^*|X_n)\right\} \geq \delta, \tag{5.4}$$

and the temporal dependence in the demand process as well as the length of the demand history when the temporal dependence is represented by a normal copula.

We start with investigating how the lower bound δ in (5.4) changes as a function of the normal-copula parameter θ, i.e., the autocorrelation coefficient of the transformed AR(1) demand process $\Phi^{-1}(F(X_t)) = \theta\Phi^{-1}(F(X_{t-1}))+Y_t$. Proposition 5.4.2 provides a condition that allows us to characterize the impact of the copula parameter θ on the lower bound δ:

Proposition 5.4.2. *Let $\alpha_{\varepsilon,h,b}$ denote the value of α characterized in Proposition 2. Define $p(\cdot;\theta)$ as the pdf of the random variable $C_{2|1}(F(\hat{Q})|F(X_n);\theta)$. Suppose that*

$$\frac{p\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b};\theta\right)}{p\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b};\theta\right)} > \frac{\Phi^{-1}\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b}\right)\phi\left(\Phi^{-1}\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b}\right)\right)}{\Phi^{-1}\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b}\right)\phi\left(\Phi^{-1}\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b}\right)\right)}, \tag{5.5}$$

where $\Phi^{-1}(\cdot)$ *is the inverse standard normal cdf and* $\phi(\cdot)$ *is the standard normal pdf. Then, the lower bound* δ *on the probability of* ε*-optimality increases in* θ *for* $\theta \in (-1, 0)$*, takes its maximum value when* θ *is zero, and decreases in* θ *for* $\theta \in (0, 1)$*.*

In the remainder of the section, we let the unit holding cost be equal to the unit shortage cost.

Corollary 5.4.1. *The lower bound* δ *for the probability of* ε*-optimality takes its maximum value when there is no temporal dependence with the copula parameter* θ *equal to zero. Furthermore,* δ *decreases as the negative autocorrelation becomes stronger with* θ *approaching from zero to* -1 *or as the positive autocorrelation becomes stronger with* θ *approaching from zero to* 1*.*

Corollary 5.4.1 shows that the probabilistic guarantee for ε-optimality gets smaller as the strength of the autocorrelation increases. That is, the decision maker is less confident about the performance of the inventory-target estimator in (5.3) when there is high dependence between consecutive demand realizations. Intuitively, this can be easily seen by visualizing a demand history with very high autocorrelation. In this case, the demand realizations take values very close to each other, and they do not uniformly span the support of the demand distribution. Consequently, it takes more time for the empirical demand distribution to move away from the impact of initial demand realizations, and hence, to converge to the true marginal demand distribution, leading to smaller probabilistic guarantee for a fixed number of demand realizations in the historical data.

In the remainder of the section, we concentrate on the uncertainty around the copula-parameter estimator through its asymptotic distribution by assuming that the decision maker has access to the true marginal demand distribution. We switch our focus back to finite-sample analysis in Section 5.4.3, where both the marginal distribution function and the copula parameters are unknown to the decision maker without any restrictions on the cost parameters and the functional form of the copula.

We first present an asymptotic approximation to the distribution of the random variable $C_{2|1}(F(\hat{Q})|F(X_n);\theta)$:

Lemma 5.4.1. *For a normal copula with parameter θ, the pdf of the random variable $C_{2|1}(F(\hat{Q})|F(X_n);\theta)$ can be approximated as*

$$p(\cdot;\theta) := \frac{1}{\phi(\Phi^{-1}(\cdot))}\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\, g\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}(\cdot)\right),$$

where $g(\cdot)$ is the pdf of the standard normal-product distribution.

A standard normal-product random variable is the product of two independent standard normal random variables; its cdf is given by $G(\cdot) = \phi^{-1}K_0(|\cdot|)$, where $K_0(\cdot)$ is a modified Bessel function of the second kind (Craig, 1936).

We next provide a lower bound to number of autocorrelated demand observations that is necessary to achieve a certain level of probabilistic guarantee for ε-optimality:

Proposition 5.4.3. *A lower bound to the length of the demand history, n to deliver an approximate probabilistic guarantee of δ for ε-optimality is given by*

$$n \geq \left\lceil \left(\frac{1-\theta^2}{1+\theta^2}\right)\left(\frac{G^{-1}\left(\frac{1+\delta}{2}\right)}{\Phi^{-1}\left(\frac{1}{2}\left(1+\frac{\varepsilon}{2+\varepsilon}\right)\right)}\right)^2\right\rceil,$$

where G^{-1} is the inverse cdf of the standard normal-product distribution.

Remark 5.4.2. *The probabilistic guarantee δ is approximate because we make use of the asymptotic distribution of the normal-copula parameter estimator from Lemma 5.4.1. We build our discussion on finite number of demand observations in Section 5.4.3.*

We immediately observe in Proposition 5.4.3 that the number of demand observations necessary to achieve a certain level of approximate probabilistic guarantee for ε-optimality is a decreasing function of ε and and an increasing function of δ. That is, not surprisingly, the decision maker needs to collect more demand observations to achieve less expected cost with higher confidence. However, a more interesting result arises about the copula parameter: A smaller number of demand observations is required as the strength of the autocorrelation (i.e., the absolute value of the copula parameter θ) increases. This observation tells us the opposite of what is implied in Proposition 5.4.2.

In Proposition 5.4.2, the marginal demand distribution is unknown to the decision maker, and an increase in the strength of the autocorrelation leads to a slower convergence of the empirical demand distribution to the true marginal distribution, leading to a smaller probabilistic guarantee for a fixed number of demand observations. On the other hand, when the marginal demand distribution is known by the decision maker as in Proposition 5.4.3, an increase in the strength of the autocorrelation leads to a smaller variance for the copula-parameter estimator, and hence, an improved performance when the critical fractile solution is obtained from the true marginal demand distribution and hence, a more accurate copula-parameter estimate.

We now investigate whether it is *always* necessary to account for the temporal dependence in the demand process. Intuitively, a decision maker may not find it reasonable to model the autocorrelation and set the inventory target accordingly if there is no visible sign of temporal dependence in the demand process. Even if there is some evidence of temporal dependence, the decision maker may still want to neglect the autocorrelation considering the inevitable statistical estimation error around the copula parameter. Proposition 5.4.4 sheds light on this issue.

Proposition 5.4.4. *Despite the temporal dependence in the demand process* $\{X_t; t = 1, 2, \ldots\}$ *with a nonzero copula parameter* θ *(i.e.,* $|\theta| > 0$*), the decision maker achieves* *a higher probabilistic guarantee for ε-optimality by ignoring the demand autocorrelation when*

$$n \;\leq\; \left\lfloor \frac{1-\theta^2}{1+\theta^2}\,\frac{1}{\eta(\varepsilon)^2}\,K_0^{-1}\left(\pi\Phi\left(\frac{\sqrt{1-\theta^2}\,\eta(\varepsilon)}{|\theta|}\right)\right)^2\right\rfloor,$$

where $\eta(\varepsilon)$ is $\Phi^{-1}\left((1/2)(1+\varepsilon/(2+\varepsilon))\right)$ and $K_0(\cdot)$ is a modified Bessel function of the second kind.

We observe in Proposition 5.4.4 that the decision maker achieves a higher probabilistic guarantee for ε-optimality by ignoring the autocorrelation in the demand process if the number of demand observations is less than a certain threshold. Clearly, this threshold is a decreasing function of $|\theta|$ since the inverse of $K_0(\cdot)$ is an increasing function. We also note that the threshold is an increasing function of ε. That is, the decision maker is better off by simply ignoring the autocorrelation in the demand process when the strength of autocorrelation is small and the accuracy level ε is high in the presence of a limited amount of demand data.

For example, Proposition 5.4.4 reveals that ignoring the autocorrelation in the demand process leads to a higher probabilistic-guarantee on 0.25-optimality when $|\theta| = 0.1$ in the presence of less than 70 demand observations. Nevertheless, as the strength of autocorrelation increases, it quickly becomes necessary to account for the autocorrelation in inventory-target estimation. For instance, ignoring the autocorrelation always leads to a smaller probabilistic-guarantee for 0.25-optimality as soon as the strength of autocorrelation, $|\theta|$ exceeds 0.3. In Section 5.5.2, we investigate when the estimation of the critical fractile solution by ignoring the temporal dependence leads to a higher probabilistic guarantee of ε-optimality if the marginal demand distribution is also unknown.

5.4.3 Identifying the Probabilistic Guarantee for ε-Optimality under General Copula

In this section, we switch our focus back to a general copula to allow the decision maker to capture scale-free temporal dependence properties in the demand process. For example, autocorrelation is only a scalar measure of linear dependence suitable when the demand random variables in two consecutive periods have an elliptical joint distribution; e.g., bivariate normal distribution implied by an AR(1) model with normal random shocks. On the other hand, the joint distribution of the demand random variables in two consecutive periods may not be elliptical. The objective of this section is to provide an arbitrarily close approximation to the lower bound δ defined in (5.4) under the assumption of a general copula that captures such nonlinear asymmetric dependence. We also relax the assumptions of known marginal demand distribution and equal holding and shortage costs, and use the empirical demand distribution to estimate the marginal demand distribution.

Clearly, the probability of the random event $Y(\hat{Q}, X_n)$ – which gives the lower bound δ – is a function of the unknown marginal distribution function F, while our goal is to find a probabilistic guarantee on ε-optimality for any marginal distribution. The marginal-copula representation of the demand process allows us to achieve this goal by waiving the requirement to know the true marginal distribution function of the demand process.

We let $\{U_t; t = 1, 2, \ldots, n\}$ correspond to a series of *dependent* standard uniform random variables represented by the copula function $C(\cdot, \cdot; \boldsymbol{\theta})$, and let $G_n(u) := (1/n) \sum_{t=1}^{n} \mathbb{1}(U_t \leq u)$ be the empirical distribution function built from these uniform random variables. Then, we define a random variable U^* as follows:

$$U^* = \min_{j=1,2,\ldots,n} \left\{ U_j : G_n(U_j) \geq C_{2|1}^{-1} \left(\frac{b}{h+b} \Big| G_n(U_n); \hat{\boldsymbol{\theta}} \right) \right\}. \quad (5.6)$$

We also define the event $\tilde{Y}(U^*, U_n)$ as

$$\left[C_{2|1}\left(U^*|U_n;\boldsymbol{\theta}\right) \geq \frac{b}{h+b} - \alpha\right] \bigcap \left[C_{2|1}\left(U^*|U_n;\boldsymbol{\theta}\right) \leq \frac{b}{h+b} + \alpha\right],$$

which will lead us to the key result to identify the lower bound δ on the probability of ε-optimality for any marginal demand distribution:

Proposition 5.4.5. *The probability of the event $Y(\hat{Q}, X_n)$ is equal to the probability of the event $\tilde{Y}(U^*, U_n)$, where U^* is defined as in (5.6) from a series of dependent standard uniform random variables $\{U_t; t = 1, 2, \ldots, n\}$ generated from the copula $C(\cdot, \cdot; \boldsymbol{\theta})$.*

Proposition 5.4.5 plays a critical role in identifying the lower bound δ because it allows us to focus on the transformed demand random variables $\{U_t; t = 1, 2, \ldots, n\}$ instead of the actual demand process $\{X_t; t = 1, 2, \ldots, n\}$. We next present an algorithm in Figure 5.1 based on this property to approximate the lower bound δ to the ε-optimality probability by sampling dependent uniform random variates – rather than sampling the actual demand which requires the knowledge of the marginal distribution.

Remark 5.4.3. *The decision maker may question the existence of temporal dependence in the demand process and choose to set the inventory target as if the demand realizations were independent and identically distributed. In this case, the algorithm provides the lower bound to the probability of ε-optimality by choosing u^* as $\min_{j=1,2,\ldots,n} \{u_j : G_n(u_j) \geq b/(h+b)\}$ for each sample path.*

The theoretical support for the algorithm above rests upon Hoeffding's inequality for bounded random variables (Hoeffding, 1963):

$$\mathbb{P}\left(|\bar{B} - \mathbb{E}(B)| > \gamma\right) \leq 2\exp(-2M\gamma^2).$$

Figure 5.1: Computation of the lower bound δ to the probability of ε-optimality

Initialization. *Specify an accuracy parameter* $\gamma > 0$ *and a confidence parameter* $\beta \in (0,1)$. *Let* $M = \lceil \log(2/\beta)/(2\gamma^2) \rceil$ *and* $m = 1$.

while $m \leq M$:

Generate dependent standard uniform variates $(u_1, u_2, \ldots, u_n)$ *from the copula* $C(\cdot,\cdot;\boldsymbol{\theta})$.

$\hat{\boldsymbol{\theta}} \leftarrow \operatorname{argmax}\left\{\sum_{t=1}^{n-1} \log c(G_n(u_t), G_n(u_{t+1}); \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$.

$u^* \leftarrow \min_{j=1,2,\ldots,n}\left\{u_j : G_n(u_j) \geq C_{2|1}^{-1}\left(b/(h+b) \mid G_n(u_n); \hat{\boldsymbol{\theta}}\right)\right\}$.

$$B_m \leftarrow \begin{cases} 1 & \textit{if } \left|C_{2|1}(u^*|u_n;\boldsymbol{\theta}) - b/(h+b)\right| \leq (\varepsilon h b/(h+b))/(h+b+\varepsilon\max(h,b)), \\ 0 & \textit{otherwise.} \end{cases}$$

$m \leftarrow m + 1$.

end

Set $\bar{B} := M^{-1}\sum_{m=1}^{M} B_m$.

Return $\bar{B} - \gamma$ *as the value of* δ.

In this representation, $\mathbb{E}(B)$ is the unknown mean of the independently sampled indicator random variables $\{B_m; m = 1, 2, \ldots, M\}$, and it corresponds to the true value of the lower bound δ on the probability of ε-optimality. Hoeffding's inequality immediately leads to the guarantee $\mathbb{P}\left(\mathbb{E}(B) \in [\bar{B} - \gamma, \bar{B} + \gamma]\right) \geq 1 - \beta$. The algorithm provides an arbitrarily close approximation to δ because the values of γ and β can be chosen arbitrarily small to make $\bar{B} - \gamma$ close enough to the true value of the δ with high confidence. To sum up, given the confidence level $1 - \beta$, the expected cost at the inventory-target estimator $\hat{Q}$ is at most $1 + \varepsilon$ of the minimum expected cost with probability at least $\bar{B} - \gamma$.

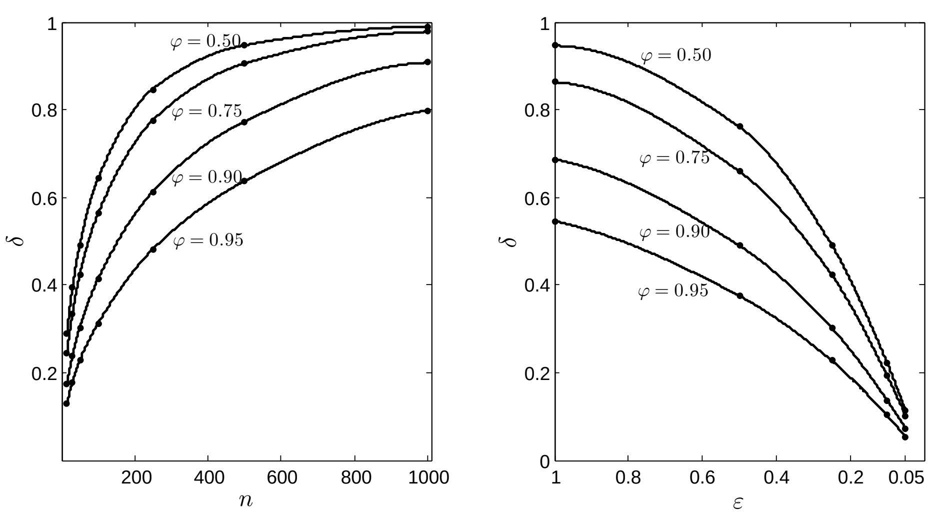
5.5 Results

In this section, we implement the algorithm in Section 5.4.3 to investigate how the lower bound δ on the probability of ε-optimality gets affected by the value of ε, the temporal dependence in the demand process, the length of the demand history, and the cost parameters. We assume that the decision maker does not know the marginal demand distribution and the normal-copula parameter θ is the measure of temporal dependence in the demand process. We relax the latter assumption in Section 5.6. Letting φ denote the critical fractile $b/(h+b)$ for ease in presentation, we set the values of γ and β to 0.001 and 0.05, respectively, in the rest of the chapter.

We first let the demand process be independent and identically distributed with θ equal to zero. The decision maker does not know that the demand realizations are independent and identically distributed, and first builds the empirical demand distribution and then estimates the copula parameter to obtain an estimate of the critical fractile solution as in (5.3).

Figure 5.2 (left) takes the accuracy level ε equal to 0.25 and plots the lower bound δ for increasing length of the demand history. We observe that the lower bound δ increases and approaches one in a faster rate for smaller values of n than for the higher values of n. For example, Figure 5.2 (left) shows that, for $\varphi = 0.50$, the lower bound δ takes the values of 28%, 39%, 49%, 64%, 85%, 95%, and 99% when the number of demand realizations n is equal to 15, 30, 50, 100, 250, 500, and 1000, respectively. That is, the expected cost associated with the inventory-target estimator in (5.3) is at most 1.25 of the expected cost of the optimal inventory target with probability

Figure 5.2: $\theta = 0$ and $\varepsilon = 0.25$ (left); $\theta = 0$ and $n = 50$ (right)



at least 39% when there are 30 demand observations in the historical data. In this case, the probabilistic guarantee of 0.25-optimality reaches 95% when the number of demand realizations increases to 500.

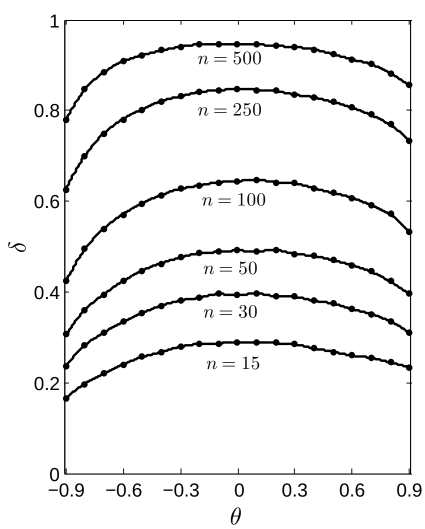
Figure 5.2 (left) also provides an insight into the role of the unit shortage and holding costs in achieving the 0.25-optimality. The lower bound δ on the probability of the 0.25-optimality decreases as the difference between the unit shortage and holding costs increases; i.e., when the asymmetry of the newsvendor's expected cost function is high. This is because the statistical estimation of the data generating process does not account for the shortage and holding costs of the newsvendor model, while the underestimation and overestimation of the optimal critical fractile solution can be penalized far differently. Intuitively, it is very likely for the expected cost of an inventory-target estimate to quickly exceed the minimum expected cost of the critical fractile solution even when the critical fractile solution is just a bit underestimated if the shortage cost is much higher than the holding cost. For example, in the presence of 500 demand observations, we observe that the lower bound δ on the probability of the 0.25-optimality takes the values of 90%, 77%, and 64% when the critical fractile φ is 0.75, 0.90, and 0.95, respectively.

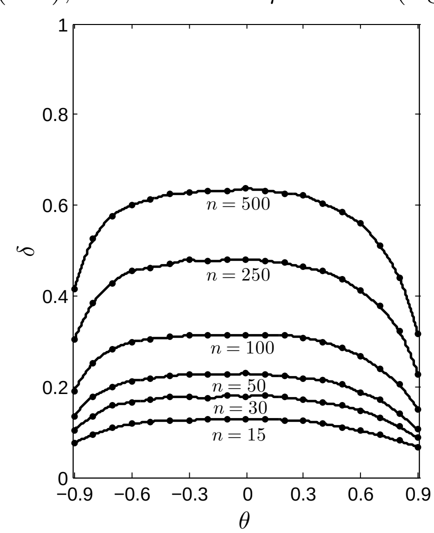
A natural question to ask is how fast the lower bound δ on the probability of the ε-optimality decreases as the value of ε gets smaller. Figure 5.2 (right) provides an answer to this question. We observe that the lower bound δ decreases and approaches zero at a faster rate for smaller values of ε than for the higher values of ε. For example, Figure 5.2 (right) shows that, for $n = 50$ and $\varphi = 0.50$, the lower bound δ takes the values of 95%, 76%, 49%, 22%, and 11% when the accuracy level ε is equal to 1, 0.5, 0.25, 0.1, and 0.05, respectively. In this case, the decision maker concludes that the expected cost of the inventory-target estimator in (5.3) is at most twice the expected cost of the optimal inventory target with probability at least 95%, while the probabilistic guarantee for 0.05-optimality is only 11%. We already know from

Figure 5.2 (left) that the lower bound δ on the probability of ε-optimality decreases as the critical fractile deviates from 0.5. What we further see from Figure 5.2 (right) is the role of ε on this reduction: The absolute decrease in the lower bound δ with the deviation of φ from 0.5 gets smaller as the accuracy level ε approaches zero.

We now let the copula parameter θ take values between -0.9 and 0.9 to represent a wide range of negative and positive autocorrelations in the demand process. Figure 5.3 plots the lower bound δ on the probability of 0.25-optimality as a function of the length of the demand history. For all values of n, we observe that the lower bound δ takes its maximum value when there is no autocorrelation in the demand process; i.e., when θ is equal to zero. Furthermore, we see that the lower bound δ decreases as the absolute value of the copula parameter θ increases. That is, the probabilistic guarantee to assure that the expected cost of an inventory-target estimate is at most 1.25 of the optimal expected cost decreases with the increasing strength of the autocorrelation in the demand process. To put it another way, as the autocorrelation in the demand process increases, the decision maker needs to collect more demand observations to be able to achieve a specified level of probabilistic guarantee. In Proposition 5.4.2, we prove this result for a symmetric expected cost function with $\varphi = 0.5$. We observe

Figure 5.3: $\varepsilon = 0.25$ and $\varphi = 0.5$ (left); $\varepsilon = 0.25$ and $\varphi = 0.95$ (right)





in Figure 5.3 (right) that this is also the case for a highly asymmetric expected cost function with $\varphi = 0.95$.

Table 5.1 and Table 5.2, in which we provide the number of demand observations necessary to achieve an ε-optimality guarantee of 50% for the critical fractile values of 0.5 and 0.95, verifies the need for a larger number of demand observations with increasing strength of autocorrelation. For example, Table 5.1 shows that the expected cost of the inventory-target estimator in (5.3) is at most 1.5 of the optimal expected cost with a probability of at least 50% if the historical demand data includes the past 45 observations of a highly negatively correlated demand process (i.e., $\theta = -0.9$). The required number of demand observations decreases to 14 when there is no temporal dependence (i.e., $\theta = 0$), and then increases to 30 for a highly positively correlated demand process (i.e., $\theta = 0.9$). We explain this behavior in Section 5.4.2 by the slow convergence of the empirical demand distribution to the true marginal distribution in the presence of strong correlation – because it takes more time for the demand history to discard the initial condition (i.e., the demand realizations at the beginning of the process) which controls the later demand realizations, delaying the convergence of

Table 5.1: The number of demand observations necessary to achieve $\delta = 0.5$ with $\varphi = 0.5$

ε	θ						
	-0.9	-0.6	-0.3	0	0.3	0.6	0.9
0.50	45	23	17	14	17	20	30
0.25	145	73	57	52	53	62	87
0.10	749	374	309	290	287	326	401

Table 5.2: The number of demand observations necessary to achieve $\delta = 0.5$ with $\varphi = 0.95$

ε	θ						
	-0.9	-0.6	-0.3	0	0.3	0.6	0.9
0.50	265	108	100	96	102	136	475
0.25	778	322	280	274	299	387	1384
0.10	3812	1581	1390	1370	1442	1892	6738

the empirical demand distribution. Tables 5.1 and 5.2 also shed light on how fast the number of demand observations necessary to achieve a certain level of probabilistic guarantee increases as ε decreases.

5.5.1 The Value of Perfect Information about Autocorrelation

We now investigate the value of the knowledge of the autocorrelation in the demand process. To put it another way, we ask what the increase in the probabilistic guarantee of ε-optimality would be if the copula parameter θ were known by the decision maker. That is, the decision maker estimates the marginal distribution only. Table 5.3 reports the value of the lower bound δ on the probability of ε-optimality obtained in this way in parenthesis next to the value of δ obtained under the assumption of an unknown copula parameter as in the previous section.

The immediate observation is the relatively smaller increase in the lower bound δ with the additional knowledge of the copula parameter in estimating inventory targets. For example, the lower bound δ on the probability of 0.5-optimality is 41% when the copula parameter $\theta = -0.90$ is unknown and estimated as described in Section 5.3. However, when the value of θ is known by the decision maker, the lower bound δ only reaches 45%. Similarly, the value of δ is 51% when the copula parameter $\theta = 0.90$ is unknown by the decision maker, and it reaches 72% when the true value of θ is made available to the decision maker. This is the maximum increase we observe in Table 5.3. The key takeway here is the leading role of the uncertainty around the empirical demand distribution function in driving the probabilistic guarantee for ε-optimality.

Table 5.3: The probabilistic guarantee δ when θ is estimated from a demand history of length $n = 30$ (when θ is known by the decision maker) for $\varphi = 0.5$

θ	ε			
	0.50	0.25	0.10	0.05
-0.9	0.41 (0.45)	0.23 (0.26)	0.10 (0.12)	0.05 (0.06)
-0.6	0.56 (0.62)	0.33 (0.37)	0.15 (0.17)	0.08 (0.09)
-0.3	0.63 (0.70)	0.38 (0.44)	0.17 (0.19)	0.09 (0.10)
0	0.65 (0.73)	0.40 (0.46)	0.18 (0.21)	0.09 (0.11)
0.3	0.63 (0.72)	0.39 (0.45)	0.17 (0.20)	0.09 (0.11)
0.6	0.59 (0.69)	0.37 (0.43)	0.16 (0.19)	0.08 (0.10)
0.9	0.51 (0.72)	0.32 (0.47)	0.14 (0.22)	0.07 (0.11)

5.5.2 The Value of Relaxing the Independence Assumption

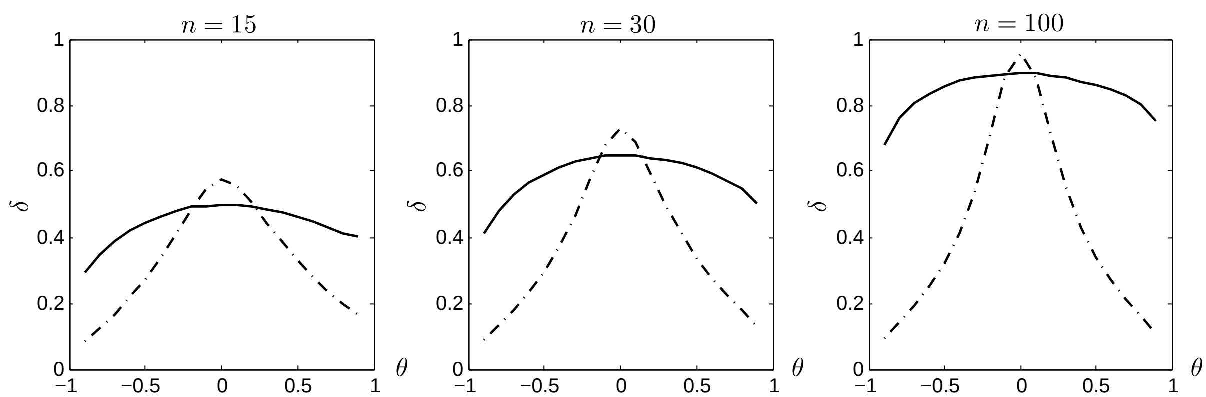
For a decision maker with no significant evidence about the existence of autocorrelation in the demand process, it may not be reasonable to make use of the copula-based demand model introduced in Section 5.2.1. Instead, the decision maker may choose to estimate the critical fractile solution by assuming that the demand process is independent and identically distributed. In this section, considering the inevitable statistical estimation error around the copula parameter, we aim to understand when it is necessary to account for the autocorrelation in the historical demand data and when it is better to ignore it.

Figure 5.4 plots the lower bound δ for the probability of 0.5-optimality when the demand is assumed to be independent (dot-dashed line) and autocorrelated (solid line) by the decision maker. We observe that the insights derived in Proposition 5.4.4 for known marginal distribution continue to hold when the decision maker does not know the true marginal distribution: In each plot, there is an interval of copula-parameter values, where the lower bound δ on the probability of 0.5-optimality is higher if the demand is assumed to be independent.

Specifically, we observe that if the autocorrelation is small enough such that the copula parameter θ takes values from the intervals $(-0.2, 0.2)$, $(-0.15, 0.15)$, and

$(-0.1, 0.1)$, it hurts to account for the autocorrelation in the demand process for a demand history of length 15, 30, and 100, respectively. This can be explained by the fact that the copula-parameter estimator has the highest variance when there is no temporal dependence in the demand process. Intuitively, assuming the copula-parameter estimate as zero is better than estimating its value with high variability for these values of the copula parameter and the length of demand history. On the other hand, if the autocorrelation is strong enough such that the copula parameter falls outside of the intervals $(-0.2, 0.2)$, $(-0.15, 0.15)$, and $(-0.1, 0.1)$ when n is equal to 10, 30, and 100, we observe a significant benefit in Figure 5.4 to achieve a higher probabilistic guarantee by incorporating the autocorrelation in inventory decisions. Clearly, the uncertainties around the copula-parameter estimator and the empirical demand distribution decrease as the length of the demand history increases, and this is why the copula-parameter interval which suggests ignoring the autocorrelation gets smaller and finally becomes negligible with the accumulation of demand realizations.

Figure 5.4: The probabilistic guarantee δ for 0.5-optimality when the decision maker assumes that demand is independent (dot-dashed line) and demand is autocorrelated (solid line) with $\varphi = 0.5$



5.6 Impact of Tail Dependence on the Probabilistic Guarantee for ε-Optimality

Autocorrelation is a measure of *linear* association between a random variable and its previous realizations. The normal copula, for example, allows us to model the first-order stationary demand process $\{X_t; t = 1, 2, \ldots\}$ by representing the random variable $\Phi^{-1}(F(X_t))$ as a linear combination of $\Phi^{-1}(F(X_{t-1}))$ and a random shock; i.e., $\Phi^{-1}(F(X_t)) = \theta\,\Phi^{-1}(F(X_{t-1})) + Y_t$. However, dependence is a more general concept which refers to any type of association between random variables. For example, it can be more likely to see a stronger dependence between low demand realizations in consecutive time periods when the economy is in a temporary recession. In other words, the dependence between low demand values can be higher than the dependence between moderate demand values. Similarly, a high demand can be more likely to be followed by another high demand when the economy is doing well. In these situations, autocorrelation is not sufficient to describe the temporal dependence at the tails of the bivariate distribution that characterizes the uncertainty in the consecutive demand random variables X_t and X_{t+1}.

In this section, we first illustrate the effectiveness of the copula-based time-series modeling in capturing tail dependence and then investigate the impact of tail dependence on the probabilistic guarantee for ε-optimality. The tail dependence is a measure of the joint behavior in the tails of the lower-left quadrant or upper-right quadrant of a bivariate distribution (Nelsen, 2006):

Definition 5.6.1. *The upper-tail dependence τ^U is defined as*

$$\tau^U = \lim_{u \to 1^-} \mathbb{P}\left(F(X_{t+1}) \geq u | F(X_t) \geq u\right),$$

and the lower-tail dependence τ^L *is defined as*

$$\tau^L = \lim_{u \to 0^+} \mathbb{P}\left(F(X_{t+1}) \leq u | F(X_t) \leq u\right).$$

Using the definition of the bivariate copula function $C(\cdot, \cdot; \boldsymbol{\theta})$, an equivalent representation of the tail dependence is given by

$$\tau^U = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u; \boldsymbol{\theta})}{1 - u}$$

for the upper-tail dependence and

$$\tau^L = \lim_{u \to 0^+} \frac{C(u, u; \boldsymbol{\theta})}{u}$$

for the lower-tail dependence. These representations show that the tail dependence is a copula property. For example, it is well known that the normal copula does not allow for tail dependence since the values of τ^U and τ^L are both zero for any value of the normal-copula parameter θ. Many of the other copulas allow tail dependence; we refer the reader to Nelsen (2006) for a review. In the remainder of this section, we illustrate the use of Clayton copula – a widely-used copula for capturing lower-tail dependence – as an alternative to normal copula while we investigate the role of tail dependence in the probabilistic guarantee for ε-optimality.

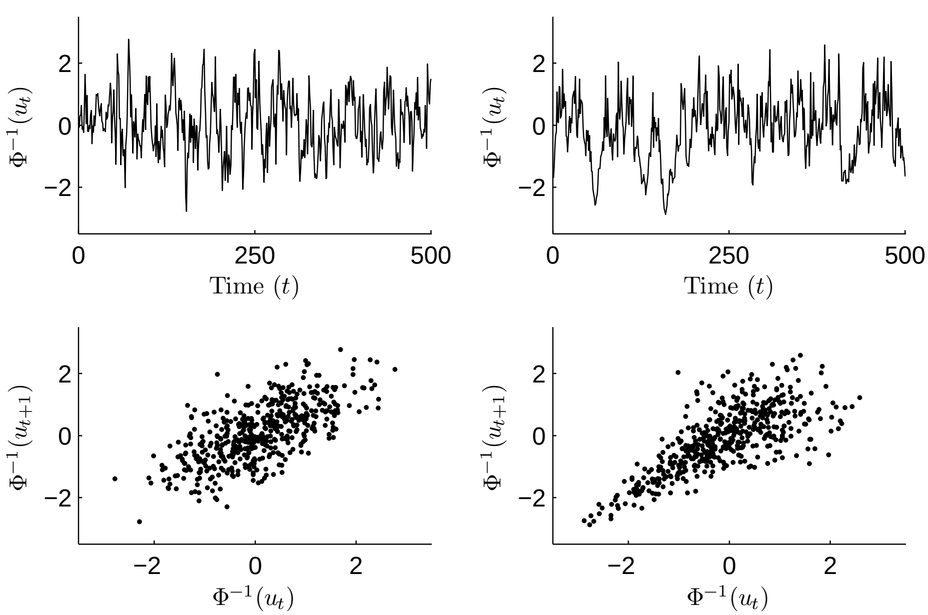
Example 5.6.1 (Clayton Copula)**.** *Let the copula parameter* θ *be positive. The Clayton copula is defined as* $C(u_t, u_{t+1}; \theta) = (u_t^{-\theta} + u_{t+1}^{-\theta} - 1)^{-1/\theta}$*, and it has the lower-tail dependence* $\tau^L = 2^{-1/\theta}$ *and the upper-tail dependence* $\tau^U = 0$.

We next illustrate the potential pitfall of taking autocorrelation as the only measure of temporal dependence. Figure 5.5 shows a 500-period time series generated by a normal copula with parameter 0.7 (plots on left) and by a Clayton copula with parameter 2.015 (plots on right) under the assumption of standard normal marginals.

In this example, both of these time series have exactly the same autocorrelation coefficient, which is equal to 0.7. Clearly, the scatter plot on the right shows a positive lower-tail dependence of $2^{-1/2.015}$ which is impossible to capture by considering the autocorrelation as the only temporal dependence measure. In other words, if the demand history shows the characteristics of tail dependence, normal copula and autocorrelation may not be sufficient to describe the temporal dependence structure in the demand process.

We now investigate the impact of tail dependence on the probabilistic guarantee for ε-optimality. We do this by considering the demand processes illustrated in Figure 5.5, and hence, we focus on the impact of the lower-tail dependence. Since both demand processes have exactly the same autocorrelation but different lower-tail dependencies, we can isolate the impact of tail dependence in our analysis. Figure 5.6 reports the lower bound δ on the probability of ε-optimality for varying levels of ε and data

Figure 5.5: Illustration of a 500-period time series generated by a normal copula (plots on left) and by a Clayton copula (plots on right) with the same autocorrelation coefficient but with different tail dependence



length n. We consider equal holding and shortage costs to limit the impact by the asymmetry of the newsvendor's expected cost function.

The difference between the two curves on each plot of Figure 5.6 highlights the impact of tail dependence on the probabilistic guarantee for ε-optimality. We observe that the presence of tail dependence leads to a smaller probabilistic guarantee despite the identical amounts of linear dependence in the respective demand processes. The impact is more visible as the length of the demand history increases, while it is less apparent for shorter demand histories. Intuitively, we can explain this result by the difficulty of estimating the tail dependence from a limited amount of historical demand data. For example, we observe a negligible difference between the two curves for n less than about 30 – a number too small to be able to observe enough number of very low consecutive demand realizations, and hence, to justify the use of tail dependence in demand modeling.

Table 5.4 provides further support to our observation about the limited impact of tail dependence on the probabilistic guarantee for ε-optimality when n is small. In Table 5.4, we aim to understand whether it is always necessary to capture the

Figure 5.6: The probabilistic guarantee δ for ε-optimality when $\varphi = 0.5$ and the demand process is represented by a normal copula (dot-dashed line) and a Clayton copula (solid line) with the same autocorrelation coefficient but with different tail dependencies as illustrated in Figure 5.5

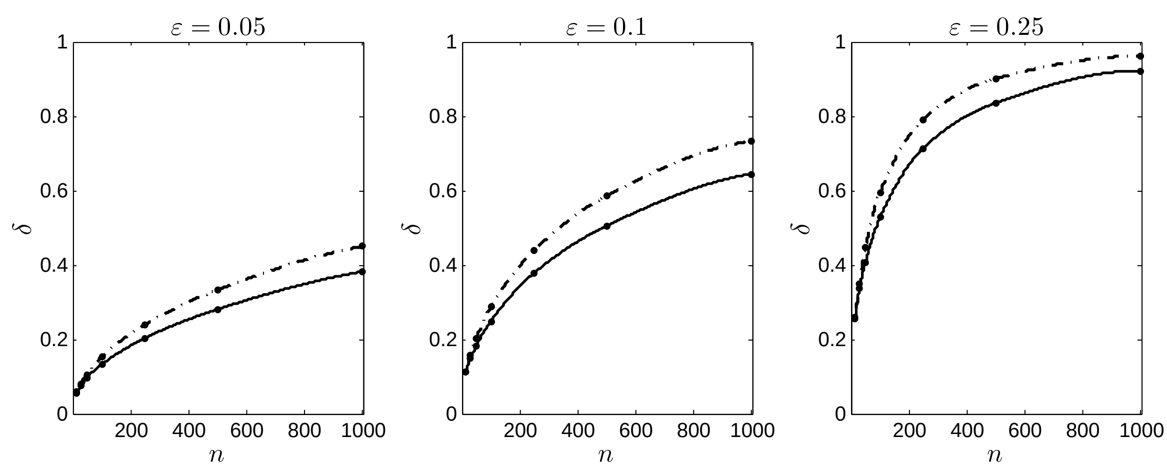


Table 5.4: The probabilistic guarantee δ for 0.25-optimality when $\varphi = 0.50$

	$\theta = 0.5$ ($\tau^L = 0.25$)		$\theta = 2$ ($\tau^L = 0.71$)		$\theta = 5$ ($\tau^L = 0.87$)	
n	δ (Clayton)	δ (Normal)	δ (Clayton)	δ (Normal)	δ (Clayton)	δ (Normal)
15	0.30	0.28	0.26	0.24	0.25	0.22
30	0.39	0.36	0.34	0.30	0.30	0.28
50	0.49	0.44	0.41	0.33	0.34	0.30
100	0.63	0.55	0.53	0.38	0.42	0.33
500	0.93	0.73	0.84	0.42	0.72	0.36

tail dependence in a demand process to achieve a higher probabilistic guarantee for ε-optimality. To this end, we provide the probabilistic guarantee δ when a normal copula – which is unable to capture any tail dependence – is used in inventory-target estimation even there is positive tail dependence in the demand process (i.e., see δ (Normal) column). In the δ (Clayton) column, on the other hand, the decision maker correctly fits a Clayton copula to the tail-dependent historical demand data and obtains the inventory-target estimate accordingly. Table 5.4 indicates a small difference between the values of the probabilistic guarantee in these two cases when there is a limited amount of demand data, suggesting the use of autocorrelation as a sufficient measure of temporal dependence. We observe that the incorporation of tail dependence in inventory decisions becomes more important to achieve a higher probabilistic guarantee for ε-optimality especially when there are strong tail dependence and sufficiently large number of demand observations.

5.7 Conclusion

We consider the problem of estimating the critical fractile solution in a newsvendor setting when the demand depends on its past realizations. As opposed to widely-used linear time-series models with normally distributed random shocks, we introduce a copula-based demand model that allows us to represent the stationary and temporally dependent demand process with any marginal demand distribution and an arbitrary dependence structure. Consequently, the decision maker estimates the marginal demand distribution and the copula parameters that characterize the temporal dependence separately without making any assumptions on the parametric form of the marginal demand distribution. The objective of our study is to identify a probabilistic guarantee for the ε-optimality of an estimate of the critical fractile solution which is obtained from the empirical demand distribution and the estimates of the copula parameters. To the best of our knowledge, we are the first to investigate the ε-optimality guarantee for temporally dependent demand data.

We first consider the special case of a normal copula and prove that the probabilistic guarantee for ε-optimality decreases with the strength of the autocorrelation in the demand process. We then provide a lower bound on the number of demand observations necessary to achieve a certain level of probabilistic guarantee when the marginal distribution is known and the unit holding and shortage costs are the same. In this case, we also provide an upper bound for the number of demand observations such that the decision maker achieves a higher probabilistic guarantee for ε-optimality by simply ignoring the temporal dependence in inventory-target estimation. We then propose a sampling-based method to compute this probabilistic guarantee when the marginal distribution is unknown to the decision maker without any restrictions on the functional form of copula and the cost parameters. Our method builds on the idea of sampling dependent uniform random variates matching the underlying dependence structure of the demand process – rather than the sampling of the actual demand which requires the specification of the marginal demand distribution. We also account for the tail dependence in our copula-based demand model as a measure of association alternative to autocorrelation.

We believe that our copula-based demand model and the sampling-based method of computing a probabilistic guarantee for ε-optimality have application areas not only in inventory management but essentially in any decision problem in which an overageunderage trade-off exists and the random realizations observed in consecutive time periods depend on each other. In this paper, we focus on a single-period problem and the extension of this study in a multi-period decision setting is a potential direction for future research.

Appendix

Proof of Proposition 5.4.1. Suppose $\hat{q}$ is greater than q^*. If the realized demand in period $n+1$ is $x_{n+1} \in (-\infty, \hat{q})$, then the inaccuracy $L(\hat{q}|x_n) - L(q^*|x_n)$ is at most $h(\hat{q} - q^*)$. On the other hand, if $x_{n+1} \in (\hat{q}, \infty)$, then $L(q^*|x_n) - L(\hat{q}|x_n)$ is equal to $b(\hat{q} - q^*)$. It is then possible to bound the inaccuracy as follows:

$$\begin{aligned} L(\hat{q}|x_n) - L(q^*|x_n) &\leq C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta})h(\hat{q} - q^*) \\ &\quad -(1 - C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}))b\,(\hat{q} - q^*) \\ &\leq \left(h\left(\frac{b}{h+b} + \alpha\right) - b\left(\frac{h}{h+b} - \alpha\right)\right)(\hat{q} - q^*) \\ &= \alpha(h+b)\,(\hat{q} - q^*) \end{aligned} \tag{5.7}$$

The first inequality follows from the definition of copula $C_{2|1}$ as the cdf of X_{n+1} conditional on $X_n = x_n$. The second inequality follows from the assumption of $C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}) \leq b/(b+h) + \alpha$. Furthermore, since

$$\begin{aligned} L(q^*|x_n) &\geq \mathbb{E}_{X_{n+1}|x_n}\left(b(X_{n+1} - q^*)^+\right) \\ &\geq \mathbb{E}_{X_{n+1}|x_n}(\mathbb{1}(X_{n+1} > \hat{q})b(\hat{q} - q^*)) \\ &= (1 - C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}))b(\hat{q} - q^*), \end{aligned}$$

the inequality

$$L(q^*|x_n) \geq b\left(\frac{h}{h+b} - \alpha\right)(\hat{q} - q^*) \tag{5.8}$$

is also implied by the assumption of $C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}) \leq b/(b+h) + \alpha$.

Now suppose $\hat{q}$ is less than q^*. If the realized demand in period $n+1$ is $x_{n+1} \in (\hat{q}, \infty)$, then the inaccuracy $L(\hat{q}|x_n) - L(q^*|x_n)$ is at most $b(q^* - \hat{q})$. On the other hand, if $x_{n+1} \in (-\infty, \hat{q})$, then $L(q^*|x_n) - L(\hat{q}|x_n)$ is equal to $h(q^* - \hat{q})$. It is then possible to bound the inaccuracy as follows:

$$\begin{aligned} L(\hat{q}|x_n) - L(q^*|x_n) &\leq (1 - C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}))b(\hat{q} - q^*) \\ &\quad -C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta})h\,(q^* - \hat{q}) \\ &\leq \left(b\left(\frac{h}{h+b} + \alpha\right) - h\left(\frac{b}{h+b} - \alpha\right)\right)(q^* - \hat{q}) \\ &= \alpha(h+b)\,(q^* - \hat{q}) \end{aligned} \tag{5.9}$$

The second inequality follows from the assumption of $C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}) \geq b/(h+b) - \alpha$. Furthermore, since

$$\begin{aligned} L(q^*|x_n) &\geq \mathbb{E}_{X_{n+1}|x_n}\left(h(q^* - X_{n+1})^+\right) \\ &\geq \mathbb{E}_{X_{n+1}|x_n}(\mathbb{1}(X_{n+1} \leq \hat{q})h(q^* - \hat{q})) \\ &= C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta})h(q^* - \hat{q}), \end{aligned}$$

the inequality

$$L(q^*|x_n) \geq h(b/(b+h) - \alpha)(q^* - \hat{q}) \tag{5.10}$$

is also implied by the assumption of $C_{2|1}(F(\hat{q})|F(x_n);\boldsymbol{\theta}) \geq b/(h+b) - \alpha$.

The inequalities (5.7) and (5.9) imply that $L(\hat{q}|x_n) - L(q^*|x_n) \leq \alpha(h+b)|\hat{q} - q^*|$. Similarly, the inequalities (5.8) and (5.10) imply that $L(q^*|x_n) \geq \{hb/(h+b) - \alpha\max(h,b)\}|\hat{q} - q^*|$. Consequently, the value of α that guarantees the ε-optimality of the inventory-target estimate $\hat{q}$ follows from solving for α in the equation $(b+h)\alpha = \varepsilon(hb/(h+b) - \alpha\max(h,b))$. $\square$

Proof of Proposition 5.4.2. Let $\mathcal{L}(\theta)$ denote the random variable $C_{2|1}(F(\hat{Q})|F(X_n);\theta)$. The focus is on the lower bound δ on the probability of ε-optimality, which is defined as the probability of the event

$$\left[\mathcal{L}(\theta) \geq \frac{b}{h+b} - \alpha_{\varepsilon,h,b}\right] \bigcap \left[\mathcal{L}(\theta) \leq \frac{b}{h+b} + \alpha_{\varepsilon,h,b}\right]$$

in Section 5.1. Let δ_a denote $\mathbb{P}(\mathcal{L}(\theta) \leq a)$. Thus, the lower bound δ is given by the difference between $\delta_{b/(h+b)+\alpha_{\varepsilon,h,b}}$ and $\delta_{b/(h+b)-\alpha_{\varepsilon,h,b}}$. The goal is to investigate the sensitivity of the probabilistic guarantee for ε-optimality, δ to the copula parameter θ; i.e.,

$$\frac{\partial \delta}{\partial \theta} = \frac{\partial \delta_{b/(h+b)+\alpha_{\varepsilon,h,b}}}{\partial \theta} - \frac{\partial \delta_{b/(h+b)-\alpha_{\varepsilon,h,b}}}{\partial \theta}. \tag{5.11}$$

We next obtain a characterization for the probability sensitivity $\partial\delta_a/\partial\theta$, which we will later use in (5.11) to obtain $\partial\delta/\partial\theta$. To do this, we first note that the following three conditions are satisfied by

$$\mathcal{L}(\theta) = \Phi\left(\frac{\Phi^{-1}(F(\hat{Q})) - \theta\Phi^{-1}(F(X_n))}{\sqrt{1-\theta^2}}\right)$$

and its first-order derivative

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \left(\frac{-\Phi^{-1}(F(X_n)) + \theta\Phi^{-1}(F(\hat{Q}))}{(1-\theta^2)\sqrt{1-\theta^2}}\right)\phi\left(\frac{\Phi^{-1}(F(\hat{Q})) - \theta\Phi^{-1}(F(X_n))}{\sqrt{1-\theta^2}}\right):$$

1. For any $\theta \in (-1,1)$, $\partial\mathcal{L}(\theta)/\partial\theta$ exists with probability one, and there exists a random variable η such that $\mathbb{E}(\eta) < \infty$ and $|\mathcal{L}(\theta+\Delta\theta) - \mathcal{L}(\theta)| < \eta|\Delta\theta|$ for any $\Delta\theta$ that is close enough to zero.

2. For any $\theta \in (-1,1)$, the cdf $\mathcal{P}(a;\theta)$ of $\mathcal{L}(\theta)$ is continuously differentiable at $a = b/(h+b) - \alpha_{\varepsilon,h,b}$ and $a = b/(h+b) + \alpha_{\varepsilon,h,b}$.

3. Since the conditional expectation function $\mathbb{E}(\partial\mathcal{L}(\theta)/\partial\theta|\mathcal{L}(\theta)=a)$ is given by

$$-\frac{1}{\sqrt{1-\theta^2}}\mathbb{E}\left(\Phi^{-1}(F(X_n))\phi(\Phi^{-1}(a))\right)+\frac{\theta}{1-\theta^2}\Phi^{-1}(a)\phi(\Phi^{-1}(a)),$$

and $\Phi^{-1}(F(X_n))$ is standard normal with the expected value of zero, the conditional expectation $\mathbb{E}(\partial\mathcal{L}(\theta)/\partial\theta|\mathcal{L}(\theta)=a)$ reduces to the following representation:

$$\frac{\theta}{1-\theta^2}\Phi^{-1}(a)\phi(\Phi^{-1}(a)).$$

Therefore, $\mathbb{E}(\partial\mathcal{L}(\theta)/\partial\theta|\mathcal{L}(\theta)=a)$ is continuous at $a=b/(h+b)-\alpha_{\varepsilon,h,b}$ and $a=b/(h+b)+\alpha_{\varepsilon,h,b}$ for any $\theta\in(-1,1)$.

Since $\mathcal{L}(\theta)$ satisfies all the three conditions above, we use the following result by Hong (2009) to identify the probability sensitivity $\partial\delta_a/\partial\theta$ as

$$\frac{\partial\delta_a}{\partial\theta}=-p(a;\theta)\mathbb{E}\left(\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\Big|\mathcal{L}(\theta)=a\right)$$

where $p(a;\theta)$ is the value of $\partial\mathcal{P}(t;\theta)/\partial\theta$ evaluated at $t=a$. Consequently, we obtain

$$\begin{aligned}\frac{\partial\delta}{\partial\theta} &= -\frac{\theta}{1-\theta^2}p\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b};\theta\right)\Phi^{-1}\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b}\right)\phi\left(\Phi^{-1}\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b}\right)\right)\\ &\quad+\frac{\theta}{1-\theta^2}p\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b};\theta\right)\Phi^{-1}\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b}\right)\phi\left(\Phi^{-1}\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b}\right)\right).\end{aligned}$$

Thus, $\partial\delta/\partial\theta=0$ for $\theta=0$. We also note that $\partial\delta/\partial\theta<0$ for $\theta\in(0,1)$ and $\partial\delta/\partial\theta>0$ for $\theta\in(-1,0)$ as long as the inequality in Proposition 5.4.2 is satisfied. □

Proof of Corollary 5.4.1. For $h=b$, $\alpha_{\varepsilon,h,b}$ reduces to $(1/2)(1+\varepsilon/(2+\varepsilon))$, and it holds that

$$\Phi^{-1}\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b}\right)=-\Phi^{-1}\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b}\right)$$

since $\Phi^{-1}(1-\Phi(z)) = -z$ for any z. Furthermore,

$$\phi\left(\Phi^{-1}\left(\frac{b}{h+b}+\alpha_{\varepsilon,h,b}\right)\right) = \phi\left(\Phi^{-1}\left(\frac{b}{h+b}-\alpha_{\varepsilon,h,b}\right)\right)$$

because the standard normal density is symmetric around zero. Consequently, the right-hand side of the inequality in Proposition 5.4.2 simplifies to -1. Since any density function takes nonnegative values, this inequality is always satisfied. $\square$

Proof of Lemma 5.4.1. The inventory-target estimator $\hat{Q}$, which is given by

$$F_n^{-1}\left(\Phi\left(\hat{\theta}\,\Phi^{-1}\left(F_n(X_n)\right)\right)+\Phi^{-1}\left(\frac{b}{h+b}\right)\sqrt{1-\hat{\theta}^2}\right)$$

for an unknown marginal distribution and unknown normal-copula parameter, simplifies to

$$F^{-1}\left(\Phi(\hat{\theta}\Phi^{-1}(F(X_n)))\right)$$

when the true marginal demand distribution is known and the unit holding and shortage costs are equal. In this case, $C_{2|1}(F(\hat{Q})|F(X_n);\theta)$ reduces to

$$\Phi\left(\frac{\left(\hat{\theta}-\theta\right)\Phi^{-1}\left(F(X_n)\right)}{\sqrt{1-\theta^2}}\right),$$

which we denote by $\mathcal{L}(\theta)$. We are interested in finding the distribution of the random variable $\mathcal{L}(\theta)$, which has the cdf defined as

$$\begin{aligned}
\mathcal{P}(t;\theta) &= \mathbb{P}\left(\mathcal{L}(\theta)\leq t\right) \\
&= \mathbb{P}\left(\Phi\left(\frac{\left(\hat{\theta}-\theta\right)\Phi^{-1}\left(F(X_n)\right)}{\sqrt{1-\theta^2}}\right)\leq t\right) \\
&= \mathbb{P}\left(\frac{\left(\hat{\theta}-\theta\right)\Phi^{-1}\left(F(X_n)\right)}{\sqrt{1-\theta^2}}\leq \Phi^{-1}(t)\right). \qquad (5.12)
\end{aligned}$$

We now derive an asymptotic approximation to the distribution of the random variable $\hat{\theta} - \theta$ using the large-sample theory of maximum likelihood estimation for stationary time series. When the marginal demand distribution function F is assumed to be known, the transformed demand series $\{\Phi^{-1}(F(X_t)); t = 1, 2, \ldots\}$ reduces to an AR(1) time series $\{Z_t; t = 1, 2, \ldots\}$ with autocorrelation θ. Since the marginal demand distribution is known, the two-step estimation procedure introduced in Section 3.1 has only the second step, which yields the copula-parameter estimator $\hat{\theta}$ as the solution of $\arg\max\{\ell(\theta) : \theta \in (-1, 1)\}$, where the log-likelihood function $\ell(\theta)$ is given by

$$\ell(\theta) = \sum_{t=1}^{n-1} \left(-\frac{1}{2} \log(1 - \theta^2) - \frac{\theta^2}{2(1 - \theta^2)} (z_t^2 + z_{t+1}^2) + \frac{\theta}{1 - \theta^2} z_t z_{t+1} \right).$$

We first note that $\mathbb{E}(Z_t^2) = 1$ and $\mathbb{E}(Z_t Z_{t+1}) = \theta$, and then use the inverse of Fisher information $\mathbb{E}(-\partial^2 \ell(\theta)/\partial\theta^2)$ to calculate the variance of the copula-parameter estimator $\hat{\theta}$. Consequently, we show that $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal with mean zero and variance $(1 - \theta^2)^2/(1 + \theta^2)$ since the AR(1) time-series model is ergodic and the asymptotic normality results derived for maximum likelihood estimation with independent data carry over (Billingsley, 2008). Let Z denote a standard normal random variable independent of the standard normal random variable $Z_n = \Phi^{-1}(F(X_n))$. The distribution of $\mathcal{L}(\theta)$ in (5.12) can be approximated as

$$\begin{aligned} \mathcal{P}(t; \theta) &= \mathbb{P}\left(\sqrt{\frac{1 - \theta^2}{n(1 + \theta^2)}} Z Z_n \leq \Phi^{-1}(t) \right) \\ &= \mathbb{P}\left(Z Z_n \leq \sqrt{\frac{n(1 + \theta^2)}{1 - \theta^2}} \Phi^{-1}(t) \right) \\ &= G\left(\sqrt{\frac{n(1 + \theta^2)}{1 - \theta^2}} \Phi^{-1}(t) \right), \end{aligned}$$

where G is the cdf of the product of two independent standard normal random variables. The result follows from taking the derivative of $\mathcal{P}(t;\theta)$ with respect to t. $\square$

Proof of Proposition 5.4.3. We first note that the coefficient $\alpha_{\varepsilon,h,b}$ characterized in Proposition 5.4.2 is equal to $(\varepsilon/2)(1/(2+\varepsilon))$ when the unit shortage and holding costs are the same. The goal is to identify the minimum length of the demand history n such that

$$\mathcal{P}\left(\frac{1}{2}\left(1+\frac{\varepsilon}{2+\varepsilon}\right);\theta\right)-\mathcal{P}\left(\frac{1}{2}\left(1-\frac{\varepsilon}{2+\varepsilon}\right);\theta\right)\geq\delta.$$

This inequality can be written as

$$G\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}\left(\frac{1}{2}\left(1+\frac{\varepsilon}{2+\varepsilon}\right)\right)\right) \\ -G\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}\left(\frac{1}{2}\left(1-\frac{\varepsilon}{2+\varepsilon}\right)\right)\right) \;\geq\; \delta$$

by Lemma 5.4.1. Due to the symmetry of the cdf G around zero, this inequality takes the form

$$2\,G\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}\left(\frac{1}{2}\left(1+\frac{\varepsilon}{2+\varepsilon}\right)\right)\right)-1\geq\delta,$$

and the result follows from solving for n in the inequality. $\square$

Proof of Proposition 5.4.4. We know from the definition of the event $Y(\hat{Q},X_n)$ that the lower bound on the probability of ε-optimality is characterized as

$$\mathbb{P}\left(\Phi\left(\frac{(\hat{\theta}-\theta)\Phi^{-1}(F(X_n))}{\sqrt{1-\theta^2}}\right)\leq\frac{1}{2}+\frac{\varepsilon/2}{\varepsilon+2}\right) \\ -\mathbb{P}\left(\Phi\left(\frac{(\hat{\theta}-\theta)\Phi^{-1}(F(X_n))}{\sqrt{1-\theta^2}}\right)\leq\frac{1}{2}-\frac{\varepsilon/2}{\varepsilon+2}\right)$$

when the unit holding and shortage costs are the same and the decision maker knows the marginal demand distribution function F. Using the asymptotic distribution of the copula-parameter estimator $\hat{\theta}$ characterized in Lemma 1, we can equivalently write (5.13) as

$$G\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}\left(\frac{1}{2}\left(1+\frac{\varepsilon}{2+\varepsilon}\right)\right)\right) - G\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}\left(\frac{1}{2}\left(1-\frac{\varepsilon}{2+\varepsilon}\right)\right)\right),$$

which takes the form

$$2G\left(\sqrt{\frac{n(1+\theta^2)}{1-\theta^2}}\Phi^{-1}\left(\frac{1}{2}\left(1+\frac{\varepsilon}{2+\varepsilon}\right)\right)\right) - 1, \tag{5.13}$$

since the standard normal-product distribution is symmetric around zero. The cdf $G(\cdot)$ is given by $\phi^{-1}K_0(|\cdot|)$, and we rewrite (5.13) as

$$\frac{2}{\pi}K_0\left(\sqrt{n}\sqrt{\frac{1+\theta^2}{1-\theta^2}}\eta(\varepsilon)\right) - 1, \tag{5.14}$$

where $\eta(\varepsilon)$ denotes $\Phi^{-1}\left((1/2)(1+\varepsilon/(2+\varepsilon))\right)$. We now suppose that the decision maker ignores the autocorrelation in the demand process by setting $\hat{\theta}$ equal to zero. In this case, (5.13) reduces to

$$\mathbb{P}\left(\Phi\left(-\frac{\theta\,\Phi^{-1}(F(X_n))}{\sqrt{1-\theta^2}}\right) \leq \frac{1}{2} + \frac{\varepsilon/2}{\varepsilon+2}\right) - \mathbb{P}\left(\Phi\left(-\frac{\theta\,\Phi^{-1}(F(X_n))}{\sqrt{1-\theta^2}}\right) \leq \frac{1}{2} - \frac{\varepsilon/2}{\varepsilon+2}\right).$$

Since $\Phi^{-1}(F(X_n))$ is a standard normal random variable, (5.15) can be rewritten as

$$2\,\Phi\left(\frac{\sqrt{1-\theta^2}\,\eta(\varepsilon)}{|\theta|}\right) - 1. \tag{5.15}$$

The result follows from solving the value of n that makes (5.15) greater than or equal to (5.14). $\square$

Proof of Proposition 5.4.5. It follows from the probability integral transformation and the definition of a copula function that the series of standard uniform random variables $\{U_t; t = 1, 2, \ldots, n\}$ is equivalent to $\{F(X_t); t = 1, 2, \ldots, n\}$. We consider the realizations $\{u_t; t = 1, 2, \ldots, n\}$ and $\{x_t; t = 1, 2, \ldots, n\}$. Since only the ordinal relation matters in building an empirical distribution function, the values of $F_n(x_t)$ and $G_n(u_t)$ are the same for $t = 1, 2, \ldots, n$. Therefore, $C_{2|1}^{-1}(b/(h+b)|F_n(x_n); \hat{\boldsymbol{\theta}})$ is equal to $C_{2|1}^{-1}(b/(h+b)|G_n(x_n); \hat{\boldsymbol{\theta}})$. The result follows because $u_t = F(x_t)$, $t = 1, 2, \ldots, n$ and

$$\hat{q} = \min_{j=1,2,\ldots,n} \left\{ x_j : F_n(x_j) \geq C_{2|1}^{-1}(b/(h+b)|F_n(x_n); \hat{\boldsymbol{\theta}}) \right\}$$

imply the equivalence between

$$\min_{j=1,2,\ldots,n} \left\{ u_j : G_n(u_j) \geq C_{2|1}^{-1}(b/(h+b)|G_n(u_n); \hat{\boldsymbol{\theta}}) \right\}$$

and $F(\hat{q})$ for any realizations of $\{U_t; t = 1, 2, \ldots, n\}$ and $\{X_t; t = 1, 2, \ldots, n\}$. $\square$

Chapter 6

Concluding Remarks

A challenge faced by many businesses is linking inventory management and the estimation of demand from historical data. In this dissertation, we consider a decision maker who applies an inventory replenishment formula obtained from a stochastic inventory model, in which the parameters of the true data generating process are estimated from a limited amount of historical demand data. We adopt a frequentist view and assume that the values of these parameters are fixed and only known by nature throughout the dissertation. We develop data-driven methodologies to understand, quantify, and eliminate the expected operational costs that arise from the use of parameter estimates – which are essentially random due to the random nature of the historical demand data – in decision making as if they were the true values. We focus on different types of historical demand data with a common theme of capturing the characteristics of real world demand.

In the first study, we identify a large number of demand histograms with significant levels of asymmetry and tail weight. This naturally sheds doubt on the performance of the inventory-target estimates obtained from historical data under the assumption of normally distributed demand – an assumption vastly made in theory but often violated in practice. We address this issue in a newsvendor setting by modeling the demand random variable with a flexible system of distributions, which captures a wide variety of distributional shapes with asymmetry, peakedness, and tail weight. In the second study, we consider a multi-period inventory management problem, in which inventory review periods are often shorter than the times between successive demand observations. Therefore, the demand process is intermittent, and the demand history contains many zero values. Motivated by industrial data of intermittent demand, we investigate the impact of correlation between demand size and the number of zero-demand periods preceding the demand on the performance of an inventory target estimated from limited demand data. To this end, we introduce a new copula-based demand model to capture the relation between demand size and the number of zero-demand periods preceding the demand. Building on this model, we propose two new finite-sample hypothesis tests to investigate the existence of correlation in an intermittent demand history. Our results show that the test which only considers the sampling distribution of the correlation-coefficient estimator tends to reject the independence assumption more frequently than the other test which considers the expected cost of parameter uncertainty – a measure operationally more relevant to the decision maker.

The overarching theme of these first two studies in the dissertation is to develop inventory-target estimation methods that account for the operational costs of incorrectly estimating the unknown parameters in a demand model. In particular, we combine inventory management and parameter estimation into a single task to balance the costs of under- and overestimation of the optimal inventory target of a product. We do this by minimizing the expected total operating cost in the first study and by minimizing the expected cost of parameter uncertainty in the second study.

We consider a temporally dependent demand process in the third study. In particular, we focus on finding a probabilistic guarantee of the near-optimality of an inventory-target estimator when the decision maker uses the empirical demand distribution without making any restricting assumptions on the parametric form of the demand distribution. We propose a sampling-based method to compute a lower bound to the probability of near optimality by making use of the marginal-copula representation of demand and building on the idea of sampling dependent uniform random variates – rather than the sampling of actual demand which requires the specification of a marginal demand distribution. We analyze the driving factors behind the probabilistic guarantee of near optimality, the value of perfect information about autocorrelation, and the value of relaxing the independence assumption in the presence of limited amount of historical demand data. Our findings also shed light on how the autocorrelation and tail dependence (i.e., the joint behavior at extreme demand realizations in consecutive time periods) in a demand process affect the number of demand observations required to achieve a performance arbitrarily close to the performance of the optimal inventory target. This had been only investigated for independent and identically distributed demand in the inventory management literature.

In this dissertation, the common takeaway in all three studies is that the decision maker should not regard an inventory decision as optimal when there is limited amount of historical demand data available for parameter estimation and the parameter estimates factor into an inventory model. We propose data-driven methods that effectively quantify and minimize the additional expected cost due to the incorrect estimation of unknown parameters. Nevertheless, we assume that the data generating process is stationary throughout the dissertation. We believe that the extension of our methods to work in demand environments that change over time is a potential research direction. Our main premise of using a limited amount of historical demand data is actually a good starting point, as the demand histories available to support operational decisions are often very short in practice – mainly because the underlying demand generating process does not remain constant indefinitely, and, even if there is a long demand history, it is common to consider only the most recent observations. In addition, we consider the problem of managing the inventory of a single product. We believe that the extension of our work to account for multiple product families is another promising research area as the decision maker can get more from the limited amount of demand histories by considering the relations between product families.

Bibliography

Akcay, A., B. Biller, S. Tayur. 2011. Improved inventory targets in the presence of limited historical demand data. *Manufacturing and Service Operations Management* **13**(3) 297–309.

Akkaya, A.D., M.L. Tiku. 2005. Time series AR(1) model for short-tailed distributions. *Statistics* **39**(2) 117–132.

Altay, N., L.A. Litteral, F. Rudisill. 2012. Effects of correlation on intermittent demand forecasting and stock control. *International Journal of Production Economics* **135**(1) 275–283.

Aydın, B., K. Guler, E. Kayış. 2012. A copula approach to inventory pooling problems with newsvendor products. *Handbook of Newsvendor Problems* 81–101.

Azoury, K.S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* **31** 1150–1160.

Bartezzaghi, E., R. Verganti, G. Zotteri. 1999. Measuring the impact of asymmetric demand distributions on inventories. *International Journal of Production Economics* **60-61** 395–404.

Beare, B.K. 2010. Copulas and temporal dependence. *Econometrica* **78**(1) 395–410.

Bennett, B.M. 1955. On the joint distribution of the mean and standard deviation. *Annals of the Institute of Statistical Mathematics* **7** 63–66.

Bertsimas, D., A. Thiele. 2006. A robust optimization approach to inventory theory. *Operations Research* **54**(1) 150–168.

Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Science, Articles in Advance* 1–18.

Biller, B., C. Gunes. 2010. Least-squares and Bayesian inferences for Johnson's S_L and S_B distributions. Working Paper, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA.

Biller, B., B. Nelson. 2005. Fitting time-series input processes for simulation. *Operations Research* **53** 549–559.

Billingsley, P. 2008. *Probability and Measure*. John Wiley & Sons.

Block, H.W., N.A. Langberg, D.S. Stoffer. 1990. Time series models for non-gaussian processes. *Lecture Notes-Monograph Series* 69–83.

Bookbinder, J.H., A.E. Lordahl. 1989. Estimation of inventory reorder level using the bootstrap statistical procedure. *IEE Transactions* **21** 302–312.

Bowman, K.O., L.R. Shenton. 1989. S_B and S_U distributions fitted by percentiles: A general criterion. *Communications in Statistics: Simulation and Computation* **18** 1–13.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, United Kingdom.

Brown, R.G. 1959. Statistical forecasting for inventory control .

Cachon, G.P., C. Terwiesch. 2009. *Matching Supply with Demand: An Introduction to Operations Management*. 2nd edition, McGraw-Hill/Irwin, New York.

Cario, M.C., B.L. Nelson. 1996. Autoregressive to anything: Time-series input processes for simulation. *Operations Research Letters* **19**(2) 51–58.

Chen, H., B.W. Schmeiser. 2001. Stochastic root finding via retrospective approximation. *IIE Transactions* **33** 259–275.

Chen, L., H.L Lee. 2012. Bullwhip effect measurement and its implications. *Operations Research* **60**(4) 771–784.

Chen, X., Y. Fan. 2006. Estimation of copula-based semiparametric time series models. *Journal of Econometrics* **130**(2) 307–335.

Cherubini, U., S. Mulinacci, F. Gobbi, S. Romagnoli. 2011. *Dynamic Copula Methods in Finance*, vol. 625. Wiley.

Chou, Y.M., A.M. Polansky, R.L. Mason. 1998. Transforming non-normal data to normality in statistical process control. *Journal of Quality Technology* **30** 133–141.

Chu, L.Y., J.G. Shanthikumar, Z.M. Shen. 2008. Solving operational statistics via a Bayesian analysis. *Operations Research Letters* **36**(1) 110–116.

Clemen, R.T., T. Reilly. 1999. Correlations and copulas for decision and risk analysis. *Management Science* **45**(2) 208–224.

Cooper, W.L., D. Gupta. 2006. Stochastic comparisons in airline revenue management. *Manufacturing and Service Operations Management* **8**(3) 221–234.

Corbett, C.J., K. Rajaram. 2006. A generalization of the inventory pooling effect to nonnormal dependent demand. *Manufacturing and Service Operations Management* **8**(4) 35–358.

Cornacchia, E., J. Shamir. 2009. *Mastering Lumpy Demand*. ToolsGroup white paper, access date: 04.04.2010, http://www.toolsgroup.com.

Craig, C.C. 1936. On the frequency function of xy. *The Annals of Mathematical Statistics* **7**(1) 1–15.

Croston, J.D. 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* **23**(3) 289–303.

Eaves, A.H.C. 2002. Forecasting for the ordering and stock-holding of consumable spare parts. Ph.D. thesis, Lancaster University.

Erkip, N., W.H. Hausman, S. Nahmias. 1990. Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands. *Management Science* **36** 381–392.

Fortuin, L. 1980. Five popular probability density functions: A comparison in the field of stock-control models. *Journal of Operational Research Society* **31** 937–942.

Fosdick, B.K., A.E. Raftery. 2012. Estimating the correlation in bivariate normal data with known variances and small sample sizes. *The American Statistician* **66**(1) 34–41.

Fukuda, Y. 1960. Estimation problems in inventory control. Research Report No 64, University of California Los Angeles, Los Angeles, CA.

Gallego, G., I. Moon. 1993. The distribution free newsboy problem: review and extensions. *Journal of Operational Research Society* **44**(8) 825–834.

Gallego, G., J.K. Ryan, D. Simchi-Levi. 2001. Minimax analysis for finite-horizon inventory models. *IIE Transactions* **33** 861–874.

Godfrey, G.A., W.B. Powell. 2001. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science* **47**(8) 1101–1112.

Gourieroux, C., J. Jasiak. 2006. Autoregressive gamma processes. *Journal of Forecasting* **25**(2) 129–152.

Gupta, S. 1960. A theory of adjusting parameter estimates in decision models. PhD thesis, Case Institute of Technology, Cleveland, OH, USA.

Hayes, R.H. 1969. Statistical estimation problems in inventory control. *Management Science* **15**(11) 686–701.

Heuts, R., J.V. Lieshout, K. Baken. 1986. An inventory model: What is the influence of the shape of the lead time demand distribution? *Mathematical Methods of Operations Research* **30** B1–B14.

Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**(301) 13–30.

Hong, L.J. 2009. Estimating quantile sensitivities. *Operations Research* **57**(1) 118–130.

Hosoda, T., M.M. Naim, S.M. Disney, A. Potter. 2008. Is there a benefit to sharing market sales information? Linking theory and practice. *Computers & Industrial Engineering* **54**(2) 315–326.

Huh, W.T., R. Levi, P. Rusmevichientong, J.B. Orlin. 2011. Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Operations Research* **59**(4) 929–941.

Huh, W.T., P. Rusmevichientong. 2009. A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* **34**(1) 103–123.

Iglehart, D.L. 1964. The dynamics inventory problem with unknown demand distribution. *Management Science* **10**(3) 429–440.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Johnson, M.E. 1987. *Multivariate Statistical Simulation*. John Wiley & Sons, New York.

Johnson, N.L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* **36**(1-2) 149–176.

Jose, K.K., T. Lishamol, J. Sreekumar. 2008. Autoregressive processes with normal-laplace marginals. *Statistics & Probability Letters* **78**(15) 2456–2462.

Katircioglu, K. 1996. Essays in inventory control. PhD thesis, University of British Columbia, Vancouver, BC, Canada.

Kim, J.H., W.B. Powell. 2011. Optimal energy commitments with storage and intermittent supply. *Operations research* **59**(6) 1347–1360.

Klabjan, D., D. Simchi-Levi, M. Song. 2013. Robust stochastic lot-sizing by means of histograms. *Production and Operations Management.* Published online in advance.

Kottas, F.J., H. Lau. 1980. The use of versatile distribution families in some stochastic inventory calculations. *The Journal of the Operational Research Society* **31** 393–403.

Kros, J.F., E. Brown. 2012. *Health Care Operations and Supply Chain Management: Strategy, Operations, Planning, and Control*. Jossey-Bass.

Kumaran, M., K.K. Achary. 1996. On approximating lead time demand distributions using the generalised lambda type distribution. *The Journal of the Operational Research Society* **47** 395–404.

Lau, H.S., A.H. Lau, C.J. Ho. 1998. Improved moment-estimation formulas using more than three subjective fractiles. *Management Science* **44** 346–351.

Lau, H.S., A. Zaki. 1982. The sensitivity of inventory decisions to the shape of lead time demand distribution. *IIE Transactions* **14** 265–271.

Lee, H.L., V. Padmanabhan, S. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Science* **43** 546–558.

Lee, H.L., K.C. So, C.S. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Science* **46** 626–643.

Levi, R., G. Perakis, J. Uichanco. 2012. The data-driven newsvendor problem: new bounds and insights. Working Paper, Sloan School of Management, MIT, Cambridge, MA, USA.

Levi, R., R. Roundy, D.B. Shmoys. 2007. Provably near-optimal sampling based policies for stochastic inventory control models. *Mathematics of Operations Research* **32**(4) 821–839.

Liyanage, L.H., J.G. Shanthikumar. 2005. A practical inventory policy using operational statistics. *Operations Research Letters* **33**(4) 341–348.

Lordahl, A.E., J.H. Bookbinder. 1994. Order statistics calculation, costs and service in an (s,Q) inventory system. *Naval Research Logistics* **41**(1) 81–97.

Lovejoy, W.S. 1990. Myopic policies for some inventory models with uncertain demand distributions. *Management Science* **36**(6) 724–738.

Luong, H.T. 2007. Measure of bullwhip effect in supply chains with autoregressive demand process. *European Journal of Operational Research* **180** 1086–1097.

MacKinnon, J.G. 2009. Bootstrap hypothesis testing. *Handbook of Computational Econometrics. Chichester, UK: Wiley* 183–213.

Moon, I., G. Gallego. 1994. Distribution free procedures for some inventory models. *The Journal of the Operational Research Society* **45**(6) 651–658.

Naddor, E. 1978. Sensitivity to distributions in inventory systems. *Management Science* **24** 1769–1772.

Nahmias, S. 2005. *Production and Operations Analysis*. 5th edition, McGraw-Hill/Irwin, Burr Ridge, IL.

Nelsen, R.B. 2006. *An Introduction to Copulas*. 2nd edition, Springer, New York.

Niermann, S. 2006. Evolutionary estimation of parameters of Johnson distributions. *Journal of Statistical Computation and Simulation* **76** 185–193.

Patton, A.J. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* **110** 4–18.

Pearson, K. 1895. Skew variations in homogeneous material. *Philosophical Transactions of the Royal Society* **186** 343–414.

Peng, R.D., F. Dominici. 2008. *Statistical methods for environmental epidemiology with R: a case study in air pollution and health*. Springer.

Perakis, G., G. Roels. 2008. Regret in the newsvendor model with partial information. *Operations Research* **56**(1) 188–203.

Porteus, E. W. 2002. *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. 2007. *Numerical Recipes: The Art of Scientific Computing*. 3rd edition, Cambridge University Press, New York.

Ramamurthy, V., J.G. Shanthikumar, Z.M. Shen. 2012. Inventory policy with parametric demand: operational statistics, linear correction, and regression. *Production and Operations Management* **21**(2) 291–308.

Rohatgi, V.K., A.K.Md.E. Saleh. 2000. *An Introduction to Probability and Statistics*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Scarf, H.E. 1958. A min-max solution to an inventory problem. K. J. Arrow, S. Karlin, H. E. Scarf, eds., *Studies in Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA, 201–209.

Scarf, H.E. 1959. Bayes solutions of the statistical inventory problem. *Annals of Mathematical Statistics* **30** 490–508.

Scarf, H.E. 1960. Some remarks on bayes solutions to the inventory problem. *Naval Research Logistics Quarterly* **7**(4) 591–596.

Schmeiser, B.W., S.J. Deutsch. 1977. A versatile four parameter family of probability distributions suitable for simulation. *IIE Transactions* **9** 176–181.

See, C.T., M. Sim. 2010. Robust approximation to multiperiod inventory management. *Operations research* **58**(3) 583–594.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, PA.

Silver, E.A., M.R. Rahmana. 1986. The cost effects of statistical sampling in selecting the reorder point in a common inventory model. *The Journal of the Operational Research Society* **37** 705–713.

Sklar, A. 1959. Fonctions de repartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Universit de Paris* **8** 229–231.

Slifker, J.F., S.S. Shapiro. 1980. The Johnson System: Selection and parameter estimation distributions. *Technometrics* **22** 239–246.

SmartOps Corporation. 2009. *Enterprise Inventory Optimization Empowers Companies to Outperform.* SmartOps white paper, access date: 08.04.2010, http://www.smartops.com.

Snyder, R.D., J.K. Ord, A. Beaumont. 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting* **28**(2) 485–496.

Springer, M.D. 1953. Joint sampling distribution of the mean and standard deviation for probability density functions of doubly infinite range. *The Annals of Mathematical Statistics* **24** 118–122.

Swain, J.J., S. Venkatraman, J.R. Wilson. 1988. Least-squares estimation of distribution functions in Johnson's translation system. *The Journal of Statistical Computation and Simulation* **29** 271–297.

Syntetos, A.A., J.E. Boylan. 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics* **71**(1-3) 457–466.

Syntetos, A.A., J.E. Boylan. 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting* **21**(2) 303–314.

Syntetos, A.A., J.E. Boylan. 2010. On the variance of intermittent demand estimates. *International Journal of Production Economics* **128**(2) 546–555.

Tang, O., R. Grubbstrom. 2006. On using higher-order moments for stochastic inventory systems. *International Journal of Production Economics* **104** 454–461.

Teunter, R., B. Sani. 2009. Calculating order-up-to levels for products with intermittent demand. *International Journal of Production Research* **118**(1) 82–86.

Tiku, M.L., W-K Wong, D.C. Vaughan, G. Bian. 2002. Time series models in non-normal situations: Symmetric innovations. *Journal of Time Series Analysis* **21**(5) 571–596.

Wagner, H.M. 2002. And then there were none. *Operations Research* **50**(1) 217–226.

Weerahandi, S. 1987. Decision-theoretic estimation of optimum inventory levels with natural invariant loss functions. *Statistics & Decisions* **5** 177–188.

Willemain, T.R., C.N. Smart, J.H. Shockor, P.A. DeSautels. 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting* **10**(4) 529 – 538.

Wolfram Research, Inc. 2008. Mathematica, Version 7.0, Champaign, IL, USA.