A SIMULATION-BASED APPROACH TO CAPTURING AUTOCORRELATED DEMAND PARAMETER UNCERTAINTY IN INVENTORY MANAGEMENT

Alp Akcay
Bahar Biller
Sridhar Tayur

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA, 15213, USA

ABSTRACT

We consider a repeated newsvendor setting where the parameters of the demand distribution are unknown, and we study the problem of setting inventory targets using only a limited amount of historical demand data. We assume that the demand process is autocorrelated and represented by an Autoregressive-To-Anything time series. We represent the marginal demand distribution with the highly flexible Johnson translation system that captures a wide variety of distributional shapes. Using a simulation-based sampling algorithm, we quantify the expected cost due to parameter uncertainty as a function of the length of the historical demand data, the critical fractile, the parameters of the marginal demand distribution, and the autocorrelation of the demand process. We determine the improved inventory-target estimate accounting for this parameter uncertainty via sample-path optimization.

1 INTRODUCTION

The common practice in production and inventory management is to estimate the unknown parameters of the demand distribution using a finite (and sometimes, very limited) amount of real-world data, and then to replace the unknown parameters in the inventory model with these estimates. This practice, however, ignores the uncertainty around the estimated demand parameters (i.e., parameter uncertainty), and accounts only for stochastic uncertainty (i.e., the uncertainty due to stochastic demand before the stocking decision). Consequently, the inventory manager often obtains inaccurate estimates that do not necessarily minimize the expected cost, which arises from the mismatch between demand and inventory. Hayes (1969), Liyanage and Shanthikumar (2005), and Akcay, Biller, and Tayur (2011a) discuss the shortcomings of ignoring demand parameter uncertainty (i.e., plugging the estimates of the demand parameters into the critical fractile solution formula) in a newsvendor setting when the demand process is independent over time. For the first time we study this problem considering an autocorrelated demand process.

While parameter uncertainty is a considerable concern in effective management of inventories, it is indeed a general problem to be addressed in manufacturing and service settings when a finite amount of data is used to estimate the unknown parameters of a model. For example, Chick (2001) shows that accounting for parameter uncertainty in the simulation of an M/M/1 queueing system improves the estimates of the mean queue length and the expected percent availability of the server. Similarly, Zouaoui and Wilson (2004) demonstrate that the parameter uncertainty amounts up to 80% of the total uncertainty around the mean waiting time estimate of an M/G/1 queuing system. Our paper contributes to this literature by demonstrating the importance of parameter uncertainty in an inventory control setting and by minimizing the *total* cost function including not only the expected cost due to stochastic demand uncertainty but also the expected cost arising from the parameter uncertainty. However, this approach can be challenging as it is often not possible to express the total cost function in closed form unless the demand process is independent over

time. We overcome this challenge by approximating the total cost function via simulation when the demand process is autocorrelated, and achieve the following objectives: (i) We quantify the impact of parameter uncertainty on inventory-target estimation, and (ii) we determine the inventory-target estimate by hedging against not only the stochastic uncertainty but also the uncertainty around the parameters estimated from historical data. Furthermore, we achieve these goals under a fairly flexible demand (input) model that captures real-world demand characteristics. Proofs of the results presented in this paper are available in Akcay, Biller, and Tayur (2011b).

There exists a well-established stream of research to represent input-model parameter uncertainty, especially in discrete-event stochastic simulation outputs. We refer the reader to Barton, Nelson, and Xie (2010) and the references therein for a recent review. A widely used approach is based on Bayesian resampling methods; see Chick (2001), Zouaoui and Wilson (2003), Ng and Chick (2006), and Biller and Corlu (2011) for example studies. In this paper, we make no assumptions regarding the availability of prior information about unknown demand parameters nor assume a demand distribution that learns from historical demand data in a Bayesian fashion. Instead, we capture the uncertainty around the parameter estimates by their sampling distributions. That is, we take the view that the inventory manager is a frequentist and estimates the unknown parameters based only on a flexible input model that has been adopted for the observable demand data. Our solution method is applicable not only in inventory management but also in general manufacturing settings where a simulationist makes decisions using models with inputs obtained from observed data.

The flexibility in the demand model is important to generate new insights about the role of parameter uncertainty in inventory-target estimation, which has been only studied for independent and identically demand processes. The first-order autoregressive (AR(1)) process is often used to represent the autocorrelated demand in inventory management; e.g., Lee, So, and Tang (2000) and Luong (2007). However, the linearity of the AR(1) model in normally distributed random shocks implies a normal marginal demand distribution (Mallows 1967). Since the demand in practice exhibits varying levels of variability, asymmetry, and tail weight (Akcay, Biller, and Tayur 2011a), it is important to model the temporal dependency while being flexible enough to capture a wide variety of distributional shapes. Indeed, there exists a considerable body of literature on modeling time series with marginal distributions from specific distribution families (Lewis, McKenzie, and Hugus 1989). Nevertheless, these models often allow only limited control of the dependence structure, and a different model is required for each type of marginal distribution. To overcome these limitations, we represent the autocorrelated demand with the Autoregressive-To-Anything (ARTA) process introduced by Cario and Nelson (1996) for modeling and generating a stationary time series with an arbitrary marginal distribution. Furthermore, we allow the inventory manager to avoid any assumptions about the first four moments (i.e., mean, variance, coefficient of skewness, and coefficient of kurtosis) 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 the JTS for demand modeling provides the flexibility of capturing (unimodal and bimodal) distributional shapes with different levels of asymmetry, peakedness, and tail weights. We refer the reader to Kuhl et al. (2010) for details on the JTS within the framework of simulation input modeling.

The paper is organized as follows. In Section 2 we present the demand and inventory models. We describe the quantification of the expected cost due to ARTA parameter uncertainty in Section 3 and the determination of the inventory target considering ARTA parameter uncertainty in Section 4. We conclude with a summary of the paper in Section 5.

2 MODEL DEVELOPMENT

In Section 2.1, we present the ARTA time series for modeling the autocorrelation in the demand process. We describe the JTS to represent the marginal demand distribution in Section 2.2, and discuss the newsvendor inventory model in Section 2.3.

2.1 ARTA Time Series

We use the first-order ARTA time series (ARTA(1)) to allow our autocorrelated demand process to assume an arbitrary marginal distribution. Although it is straightforward to extend the ARTA demand model to capture high orders of autocorrelation, we only focus on the first-order demand autocorrelation due to the difficulty in obtaining accurate estimates of high-order autocorrelations from limited historical data (Wei 1990). Specifically, ARTA(1) models the stationary demand process $\{X_t; t = 1, 2, \ldots\}$ with the marginal distribution parameter vector Ψ, the marginal cumulative distribution function (cdf) $F(\cdot; \Psi)$, and the first-order autocorrelation ρ as the transformation of a base process $\{Z_t; t = 1, 2, \ldots\}$, which is a stationary AR(1) process of the form $Z_t = rZ_{t-1} + Y_t$ with a standard normal marginal distribution. In this representation, $r \in (-1, 1)$ is the lag-one autocorrelation of the base process; i.e., $r = \mathbb{E}[Z_t Z_{t+1}]$ with $\mathbb{E}$ the expectation operator. It further determines the autocorrelations at higher lags; i.e., $\mathbb{E}[Z_t Z_{t+l}] = r^l$ for $l \geq 1$. The lag-one autocorrelation ρ of the demand process $\{X_t; t = 1, 2, \ldots\}$ (i.e., $\rho = (\mathbb{E}[X_t X_{t+1}] - \mathbb{E}[X_t]^2)/\text{Var}[X_t]$) can be easily obtained from the lag-one autocorrelation r of the base process $\{Z_t; t = 1, 2, \ldots\}$ via

$$\rho \quad = \quad \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F^{-1}\left(\Phi(z_t); \Psi\right) F^{-1}\left(\Phi(z_{t+1}); \Psi\right) \vartheta_r (z_t, z_{t+1}) \, dz_t dz_{t+1} - \mathbb{E}[X_t]^2}{\text{Var}[X_t]}, \quad (1)$$

where Φ is the standard normal cdf and ϑ_r denotes the standard bivariate normal probability density function (pdf) with correlation r. The quantities Y_t, $t = 2, 3, \ldots$ are, on the other hand, independent and identically distributed normal random variables with mean zero and variance $1 - r^2$. Thus, the transformation $F^{-1}\left(\Phi(Z_t); \Psi\right)$ ensures that X_t has the marginal distribution function $F(\cdot; \Psi)$. This modeling approach works for *any* marginal distribution, although $F^{-1}(\cdot; \Psi)$ may have to be evaluated numerically when there is no exact closed-form expression.

Next, we characterize the distribution of the demand in two consecutive time periods for the ARTA(1) process. This bivariate demand distribution, denoted by H, is used to obtain the conditional demand distribution for identifying the inventory target that minimizes the expected cost in the newsvendor model:

Proposition 1. *The dependence structure of the ARTA(1) demand process $\{X_t; t = 1, 2, \ldots\}$ is uniquely captured in the two-dimensional normal copula*

$$H(x_t, x_{t+1}; \Psi, r) = \Phi_2\left(\Phi^{-1}\left(F(x_t; \Psi)\right), \Phi^{-1}\left(F(x_{t+1}; \Psi)\right); r\right),$$

where Φ^{-1} is the functional inverse of Φ and $\Phi_2(\cdot, \cdot; r)$ is the two-dimensional standard normal cdf with the correlation r satisfying (1) *for ρ.*

This result is an implication of the Sklar's theorem (Sklar 1959), which allows us to link the (common) marginal distributions of X_t and X_{t+1} in order to construct a valid bivariate distribution for the ARTA(1) process. Consequently, for a differentiable marginal cdf $F(\cdot; \Psi)$ and the normal copula function $\Phi_2(\cdot, \cdot; r)$, the joint pdf of X_t and X_{t+1} is given by

$$h(x_t, x_{t+1}; \Psi, r) \quad = \quad f(x_t; \Psi) f(x_{t+1}; \Psi) c_r (F(x_t; \Psi), F(x_{t+1}; \Psi); r),$$

where $f(\cdot; \Psi)$ is the pdf associated with cdf $F(\cdot; \Psi)$ and c_r is the copula density function defined by $\partial^2 \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v))/(\partial u \partial v)$. In other words, the joint pdf $h(\cdot; \Psi, r)$ can be written as the product of the marginal pdfs of X_t and X_{t+1} and their copula density function that encodes all the information about the temporal dependence structure of the demand process. This characterization allows us to write the transition density function $h(x_{t+1}|X_t = x_t; \Psi, r)$ as the product of the marginal pdf $f(x_{t+1}; \Psi)$ and the normal copula density function $c_r (F(x_t; \Psi), F(x_{t+1}; \Psi); r)$ given by

$$\frac{\vartheta_r\left(\Phi^{-1}\left(F(x_t; \Psi)\right), \Phi^{-1}\left(F(x_{t+1}; \Psi)\right)\right)}{\phi\left(\Phi^{-1}(F(x_t; \Psi))\right) \phi\left(\Phi^{-1}(F(x_{t+1}; \Psi))\right)}$$

with ϕ the standard normal pdf. Thus, we obtain the cdf of X_{t+1} conditional on $X_t = x_t$ as

$$
\begin{aligned}
H(x_{t+1}|X_t = x_t; \Psi, r) &= \int_{-\infty}^{x_{t+1}} f(x_{t+1}; \Psi) c_r(F(x_t; \Psi), F(v; \Psi); r) \, dv, \\
&= \Phi\left(\frac{\Phi^{-1}(F(x_{t+1}; \Psi)) - r\Phi^{-1}(F(x_t; \Psi))}{\sqrt{1 - r^2}} \right).
\end{aligned}
\tag{2}
$$

Section 2.3 uses this distribution function to identify the critical fractile solution that minimizes the newsvendor's expected cost when the ARTA demand parameters Ψ and r are known.

2.2 Johnson Translation System

The JTS for random demand X_t is defined by a cdf of the form

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

with $(\gamma, \delta, \xi, \lambda)' \equiv \Psi$, where γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, and $\eta(\cdot)$ is one of the following transformations:

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

There is a unique family (choice of η) for each feasible combination of the coefficient of skewness and the coefficient of kurtosis that determine the shape parameters γ and δ. Furthermore, the JTS provides a close approximation for many of the standard distributions. For example, the Johnson S_U family captures the Student's t distribution, logistic distribution, and Laplace distribution, while the Johnson S_B family captures the exponential distribution, gamma distribution, uniform distribution, and Weibull distribution. In general, the JTS can represent any set of (finite) first four moments, and hence the distributional shapes represented by the JTS are not limited to the shapes of standard distributions. The use of the JTS for demand modeling, therefore, enables us to solve the inventory problem of interest for any pair of coefficient of skewness and coefficient of kurtosis that a continuous demand 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 γ, δ, λ, and ξ, and the range of X_t depends on the family: $X_t > \xi$ and $\lambda = 1$ for the S_L family; $\xi < X_t < \xi + \lambda$ for the S_B family; $-\infty < X_t < \infty$ for the S_U family; and $-\infty < X_t < \infty$, $\xi = 0$, and $\lambda = 1$ for the S_N family. Examples of the pdfs captured by the JTS can be found in Johnson (1987).

2.3 Inventory Model

In this section, we discuss the newsvendor model to determine the stocking quantity of a single item before the realization of the demand. We assume that the length of the historical demand data is n, and there is no inventory on hand at the beginning of period $n + 1$. The order of I items arrives instantaneously, and all remaining units, if any, are disposed and the shortages are written off at the end of the period $n + 1$. All the relevant revenues and costs associated with the decision to order I units for period $n + 1$ are incorporated into the loss function $L(I, X_{n+1})$ that is piecewise linear in $I - X_{n+1}$. Specifically, we take

$$
L(I, X_{n+1}) = \begin{cases}
\lambda_h(I - X_{n+1}) & \text{for } I \geq X_{n+1}, \\
\lambda_s(X_{n+1} - I) & \text{for } I < X_{n+1},
\end{cases}
$$

where λ_h is the unit inventory holding cost and λ_s is the unit cost of shortage. Without loss of generality, we take λ_h as 1, denote the critical fractile $\lambda_s/(\lambda_s + \lambda_h)$ by φ, and use $\varphi/(1-\varphi)$ for λ_s. We write the expected loss function, conditional on the most-recent demand realization, as follows:

$$\bar{L}(I|X_n = x_n; \Psi, r) = \int_{-\infty}^{\infty} L(I, x_{n+1}) \mathrm{d}H(x_{n+1}|X_n = x_n; \Psi, r).$$

Since the expected loss function $\bar{L}$ is convex in I, the optimal inventory target I^* (i.e., the value of I that minimizes $\bar{L}(I|X_n = x_n; \Psi, r)$ with known Ψ and r) is given by $H^{-1}(\varphi|X_n = x_n; \Psi, r)$. The use of the JTS for the marginal demand distribution and the normal copula for the joint distribution of the ARTA(1) demand process enables the characterization of this unique critical fractile solution by

$$\Phi\left(\frac{\gamma + \delta\eta\left(\frac{I^* - \xi}{\lambda}\right) - r\left(\gamma + \delta\eta\left(\frac{x_n - \xi}{\lambda}\right)\right)}{\sqrt{1 - r^2}}\right) = \varphi.$$

Denoting the φth quantile of the standard normal distribution function by τ_φ, we represent the optimal inventory target, also called critical fractile solution, by

$$I^* = \xi + \lambda\eta^{-1}\left(\frac{r\left(\gamma + \delta\eta\left(\frac{x_n - \xi}{\lambda}\right)\right) + \tau_\varphi\sqrt{1 - r^2} - \gamma}{\delta}\right).$$

3 INVENTORY-TARGET ESTIMATION

In Section 3.1, we discuss the estimation of the ARTA demand parameters and the critical fractile solution. We introduce the notion of inaccuracy in inventory-target estimation and present an algorithm for its approximation in Section 3.2.

3.1 Two-Stage Maximum Likelihood Estimation

The likelihood function of the demand data $x_1, x_2, \ldots, x_n$ is given by $f(x_1; \Psi) \prod_{t=1}^{n-1} h(x_{t+1}|X_t = x_t; \Psi, r)$ by iteratively conditioning on the demand value of the preceding period for periods $2, 3, \ldots, n$. Since $h(x_{t+1}|X_t = x_t; \Psi, r)$ is given by the product of the marginal pdf $f(x_{t+1}; \Psi)$ and the normal copula density function $c_r(F(x_t; \Psi), F(x_{t+1}; \Psi); r)$ (Section 2.1), the likelihood function can be alternatively written as

$$\prod_{t=1}^{n} f(x_t; \Psi) \prod_{t=1}^{n-1} c_r(F(x_t; \Psi), F(x_{t+1}; \Psi); r).$$

Taking the logarithm of this likelihood function leads to the log-likelihood function

$$\ell(\Psi, r) = \sum_{t=1}^{n} \log f(x_t; \Psi) + \sum_{t=1}^{n-1} \log c_r(F(x_t; \Psi), F(x_{t+1}; \Psi); r) \tag{3}$$

in terms of the unknown ARTA demand parameters $\Psi = (\gamma, \delta, \xi, \lambda)'$ and r. What is important to recognize is that the first term on the right-hand side in (3) is the log-likelihood function if the demand process were independent; thus, it does not depend on the autocorrelation r. That is, the copula representation of the ARTA demand process (Proposition 1) allows us to separate the estimation of the marginal distribution parameters from the estimation of the demand autocorrelation and perform the following two-stage estimation: (1) Estimate the marginal distribution parameter vector Ψ by $\tilde{\Psi} = \text{argmax}_\Psi \{\ell_1(\Psi) = \sum_{t=1}^{n} \log f(x_t; \Psi)\}$. Since the Johnson pdf is given by

$$f(x_t; \gamma, \delta, \xi, \lambda) = \frac{\delta}{\lambda(2\pi)^{1/2}} \eta'\left(\frac{x_t - \xi}{\lambda}\right) \exp\left\{-\frac{1}{2}\left(\gamma + \delta\eta\left(\frac{x_t - \xi}{\lambda}\right)\right)^2\right\}$$

with η' the first-order derivative of the transformation function, we obtain $(\tilde{\gamma}, \tilde{\delta}, \tilde{\xi}, \tilde{\lambda})' \equiv \tilde{\Psi}$ from the maximization of

$$\ell_1(\gamma, \delta, \xi, \lambda) = n \log \delta - n \log \lambda + \sum_{t=1}^{n} \log \eta' \left(\frac{x_t - \xi}{\lambda} \right) - \frac{1}{2} \sum_{t=1}^{n} \left(\gamma + \delta \eta \left(\frac{x_t - \xi}{\lambda} \right) \right)^2 - \frac{n}{2} \log(2\pi).$$

(2) Estimate the autocorrelation r via $\tilde{r} = \text{argmax}_r \{ \ell_2(r) = \sum_{t=1}^{n-1} \log c_r \left(F(x_t; \tilde{\Psi}), F(x_{t+1}; \tilde{\Psi}); r \right) \}$ with $\tilde{\Psi}$ obtained in the first stage. Using the functional form of the copula density c_r and the cdf of the JTS, we reduce the log-likelihood function $\ell_2(r)$ to

$$\ell_2(r) = -\frac{n-1}{2} \log(1 - r^2) - \frac{1}{2(1-r^2)} \sum_{t=1}^{n-1} \left\{ \tilde{z}_t^2 + \tilde{z}_{t+1}^2 - 2r\tilde{z}_t\tilde{z}_{t+1} \right\}$$

with $\tilde{z}_t = \tilde{\gamma} + \tilde{\delta}\eta((x_t - \tilde{\xi})/\tilde{\lambda})$. This estimation method is known as the method of inference functions for margins (Joe 1997), and it is computationally easier than the full maximum likelihood method as it exploits the attractive feature of the copulas, that the dependence structure is characterized independently of the marginal distributions. Furthermore, Joe (1997) establishes the strong consistency and asymptotic normality of these two-stage estimators.

3.2 Quantifying the Impact of Parameter Uncertainty on Inventory-Target Estimation

The optimal inventory target is often estimated by replacing Ψ and r with their estimates $\tilde{\Psi}$ and $\tilde{r}$ in the theoretically correct formula for the critical fractile solution under complete information (i.e., maximum likelihood policy). Specifically, the inventory-target estimate is obtained as

$$\tilde{I} = \tilde{\xi} + \tilde{\lambda}\eta^{-1} \left(\frac{\tilde{r} \left(\tilde{\gamma} + \tilde{\delta}\eta \left(\frac{x_n - \tilde{\xi}}{\tilde{\lambda}} \right) \right) + \tau_\varphi \sqrt{1 - \tilde{r}^2} - \tilde{\gamma}}{\tilde{\delta}} \right). \tag{4}$$

It is important to note that the inventory-target estimate $\tilde{I}$ is a function of the historical data $x_1, x_2, \ldots, x_n$, and it is one particular way of estimating the critical fractile solution. By using $I(x_1, x_2, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ for the general form of an inventory-target estimate, we represent the total expected cost (i.e., the sum of expected costs due the stochastic demand uncertainty and the parameter uncertainty) associated with $I(x_1, x_2, \ldots, x_n)$, which has $\tilde{I}$ as a special case, by

$$\int_{\mathscr{X}} \bar{L}(I(x_1, x_2, \ldots, x_n) | X_n = x_n; \Psi, r) \prod_{t=1}^{n} f(x_t; \Psi) \prod_{t=1}^{n-1} c_r \left(F(x_t; \Psi), F(x_{t+1}; \Psi); r \right) dx_1 dx_2 \ldots, dx_n, \tag{5}$$

where $\mathscr{X}$ is the domain of the demand vector in $\mathbb{R}^n$. Hayes (1969) introduced a total expected cost function of a similar form under the name of Expected Total Operating Cost (ETOC) but for an independent and identically distributed demand process. We factor the autocorrelation in the demand process into the ETOC function by the product of the copula densities. Although this makes it impossible to analytically evaluate the ETOC function, we are able to evaluate the n-dimensional integral in (5) via Monte-Carlo integration (Robert and Casella 1999). The convenience of generating autocorrelated samples from ARTA distribution also lends itself to a simulation-based approximation.

Hayes (1969) defined the difference between the ETOC associated with the inventory-target estimate $I(x_1, x_2, \ldots, x_n)$ and the cost of the optimal inventory target under complete certainty as the *inaccuracy* in inventory-target estimation. We let Ω be the cost attributable to the incorrect estimation of the inventory target; i.e., $\Omega \equiv \bar{L}(I(x_1, x_2, \ldots, x_n) | X_n = x_n; \Psi, r) - \bar{L}(I^* | X_n = x_n; \Psi, r)$. The inaccuracy is then given by $\mathbb{E}[\Omega]$, representing the impact of parameter uncertainty on inventory-target estimation. We interpret this

expectation as the limit of a sequence of random variables that are obtained by randomly sampling from the joint density function of the demand random variables $X_1, X_2, \ldots, X_n$. This results in the following algorithm approximating $\mathbb{E}[\Omega]$:

Algorithm 1. *Approximation of the inaccuracy in inventory-target estimation*

    Initialization. *Let $b = 1$ and $B = \infty$, and initialize the confidence level α and the tolerance ε.*

    while $b \leq B$

    Step 1. *Construct the bth ARTA sample path $\{x_t^b; t = 1, 2, \ldots, n\}$:*

        Generate a standard uniform random number u^b and set $z_0^b = \Phi^{-1}(u^b)$. Let $t \leftarrow 1$.

        while $t \leq n$

        Generate y_t^b from the normal distribution with mean 0 and variance $1 - r^2$.

        Set $z_t^b \leftarrow r z_{t-1}^b + y_t^b$ and $x_t^b \leftarrow \xi + \lambda \eta^{-1}\left((z_t^b - \gamma)/\delta\right)$.

        Set $t \leftarrow t + 1$.

        end

    Step 2. *Obtain the two-stage maximum likelihood estimates $\tilde{\gamma}_b$, $\tilde{\delta}_b$, $\tilde{\xi}_b$, $\tilde{\lambda}_b$, and $\tilde{r}_b$ from the sample path $\{x_t^b; t = 1, 2, \ldots, n\}$, and identify the maximum likelihood inventory-target estimate:*

$$\tilde{I}_b = \tilde{\xi}_b + \tilde{\lambda}_b \eta^{-1}\left(\left\{\tilde{r}_b\left(\tilde{\gamma}_b + \tilde{\delta}_b \eta\left((x_n^b - \tilde{\xi}_b)/\tilde{\lambda}_b\right)\right) + \tau_\varphi \sqrt{1 - \tilde{r}_b^2} - \tilde{\gamma}_b\right\}/\tilde{\delta}_b\right)$$

    Step 3. *Obtain the critical fractile solution under known ARTA parameters:*

$$I_b^* = \xi + \lambda \eta^{-1}\left(\left\{r\left(\gamma + \delta \eta\left((x_n^b - \xi)/\lambda\right)\right) + \tau_\varphi \sqrt{1 - r^2} - \gamma\right\}/\delta\right)$$

    Step 4. *Set $\omega_b \leftarrow \bar{L}(\tilde{I}_b | X_n = x_n; \gamma, \delta, \xi, \lambda, r) - \bar{L}(I_b^* | X_n = x_n; \gamma, \delta, \xi, \lambda, r)$;*

        $\bar{\Omega}_b \leftarrow \sum_{i=1}^b \omega_i / b$;

        $v_b \leftarrow \sum_{i=1}^b (\omega_i - \bar{\Omega}_i)^2/(b-1)$ *(set $v_B \leftarrow \infty$ for $b = 1$).*

    Step 5. *Update B and b by $B \leftarrow \tau_{1-\alpha/2}^2 v_b/(\bar{\Omega}_b \varepsilon)^2$ and $b \leftarrow b + 1$, where $\tau_{1-\alpha/2} \equiv \Phi^{-1}(1 - \alpha/2)$.*

    end

    Return $\bar{\Omega}_B$ *as the approximation to $\mathbb{E}[\Omega]$.*

The random variable $(\mathbb{E}[\Omega] - \bar{\Omega}_B)/\sqrt{v_B/B}$ is approximately standard normal for large values of B, the number of independent sample paths used for approximating the inaccuracy in the inventory-target estimation. This leads to the identification of the confidence bounds:

$$\lim_{B \to \infty} \mathbb{P}\left(\left|\frac{\mathbb{E}[\Omega] - \bar{\Omega}_B}{\sqrt{v_B/B}}\right| \leq \tau_{1-\alpha/2}\right) = 1 - \alpha.$$

Thus, a $100(1 - \alpha)\%$ confidence interval for $\mathbb{E}[\Omega]$ is obtained as

$$\left[\bar{\Omega}_B - \tau_{1-\frac{\alpha}{2}}\sqrt{\frac{v_B}{B}}, \bar{\Omega}_B + \tau_{1-\frac{\alpha}{2}}\sqrt{\frac{v_B}{B}}\right];$$

i.e., choosing $B \geq \tau_{1-\frac{\alpha}{2}}^2 v_B/(\varepsilon \bar{\Omega}_B)^2$ guarantees the half-length of the confidence interval on the inaccuracy in the inventory-target estimation to be less than or equal to $\varepsilon 100\%$ of $\bar{\Omega}_B$ with probability approximately $(1 - \alpha)100\%$. In addition, the convergence speed for approximating the inaccuracy in the inventory-target estimation is $O_\mathbb{P}(B^{-1/2})$, which is independent of the length of the historical demand data.

4 CAPTURING PARAMETER UNCERTAINTY IN INVENTORY-TARGET ESTIMATION

In Section 4.1, we introduce a bias parameter into the maximum-likelihood inventory-target estimate (4) and discuss the minimization of the ETOC function, including both stochastic and parameter uncertainties, over this bias parameter. We provide a numerical analysis that investigates the inaccuracy reduction achieved by our policy in Section 4.2.

4.1 Minimizing the Expected Total Operating Cost via Sample-Path Optimization

The objective of this section is to reduce the inaccuracy quantified in Section 3.2 by accounting for the demand parameter uncertainty in inventory-target estimation. We address this problem by generalizing the class of estimators implied by the maximum likelihood policy. Specifically, we replace τ_φ in the functional formula (4) with a bias parameter k, set the inventory-target estimator to

$$\tilde{I}(k) \;=\; \tilde{\xi} + \tilde{\lambda}\,\eta^{-1}\left(\frac{\tilde{r}\left(\tilde{\gamma} + \tilde{\delta}\eta\left(\frac{x_n - \tilde{\xi}}{\tilde{\lambda}}\right)\right) + k\sqrt{1 - \tilde{r}^2} - \tilde{\gamma}}{\tilde{\delta}}\right),$$

and combine parameter estimation with inventory-target optimization by solving the problem

$$\min_{k \in \mathbb{R}}\left\{\text{ETOC}[\tilde{I}(k)] - \bar{L}\left(I^*|X_n = x_n; \Psi, r\right)\right\}.$$

Since $\bar{L}\left(I^*|X_n = x_n; \Psi, r\right)$ does not depend on k, the problem of interest reduces to the determination of the optimal bias parameter $k^* = \operatorname{argmin}\{\text{ETOC}[\tilde{I}(k)] : k \in \mathbb{R}\}$ so as to obtain an improved inventory-target estimator $\tilde{I}(k^*)$. Since $\tilde{I}(k)$ is strictly increasing in k, it is possible to conveniently determine k^* even though the function $\text{ETOC}[\tilde{I}(k)]$ is not necessarily convex in k:

Proposition 2. $\text{ETOC}[\tilde{I}(k)]$ *is strictly increasing in k for $k > k^*$ and strictly decreasing in k for $k < k^*$. Consequently, the optimal bias parameter k^* is unique and satisfies*

$$\mathbb{E}\left[\frac{\partial \tilde{I}(k)}{\partial k}\left(\Phi\left(\frac{\gamma + \delta\eta\left(\frac{\tilde{I}(k) - \xi}{\lambda}\right) - r\left(\gamma + \delta\eta\left(\frac{X_n - \xi}{\lambda}\right)\right)}{\sqrt{1 - r^2}}\right) - \varphi\right)\right] = 0. \tag{6}$$

The challenge in the identification of k^* is that $\text{ETOC}[\tilde{I}(k^*)]$ cannot be directly computed as the inventory manager does not know the true demand parameters $\gamma, \delta, \xi, \lambda$ and r that appear in (6). We overcome this challenge by replacing the unknown demand parameters with their maximum likelihood estimates in the ETOC minimization problem. It is worth pointing out that we do not simply use the parameter estimates to obtain the inventory-target estimate as in the maximum likelihood policy; instead, we use these parameter estimates in the minimization of the ETOC function. Our numerical analysis shows that the additional inaccuracy introduced by using the estimates of the parameters but not their true values is not statistically significant.

In the remainder of the section, we describe how to solve the root-finding problem in (6). For notational convenience, we let $g(k)$ denote $\mathbb{E}[G(k; X_1, X_2, \ldots, X_n)]$ with

$$G(k; X_1, X_2, \ldots, X_n) = \frac{\partial \tilde{I}(k)}{\partial k}\left(\Phi\left(\frac{\gamma + \delta\eta\left(\frac{\tilde{I}(k) - \xi}{\lambda}\right) - r\left(\gamma + \delta\eta\left(\frac{X_n - \xi}{\lambda}\right)\right)}{\sqrt{1 - r^2}}\right) - \varphi\right).$$

Since the expectation in (6) does not have a closed-form representation, we cannot find the value of k that solves the equation $g(k) = 0$ by using standard root-finding procedures. Therefore, we treat our ETOC minimization problem as a stochastic root-finding problem (SRFP) (see Pasupathy and Kim (2011) for a recent review) and determine a value for the bias parameter k by using a consistent estimator of $g(k)$. We solve the SRFP on hand by the sample-path optimization technique, also known as Sample Average Approximation (SAA) (Shapiro 2004). That is, we generate an appropriately chosen number of sample paths (i.e., ARTA demand vectors of dimension n) and use these sample paths to turn the SRFP into a deterministic root-finding problem that is solved to a predetermined precision. More specifically, we solve the sample-path problem $G_N(k) = 0$, where $G_N(k) = \sum_{b=1}^{N} G(k; X_1^b, X_2^b, \ldots, X_n^b)/N$ is a consistent estimator

of $g(k)$ and $\{X_1^b, X_2^b, \ldots, X_n^b\}$, $b = 1, 2, \ldots, N$, are the N independent and identically distributed ARTA demand vectors. If we let k_N denote the solution of this sample-path problem, then k_N converges to k^* almost surely as N approaches infinity.

In this paper, however, we use a refinement of the generic SAA technique to gain overall efficiency. Instead of solving a single problem constructed with a large number of sample paths, we first generate a sequence of problems by progressively increasing the number of sample paths, and then solve the resulting deterministic root-finding problems with progressively decreasing error tolerances. This method, which is known as Retrospective Approximation (RA), was originally designed by Chen and Schmeiser (2001) in the context of one-dimensional SRFPs. The key to achieve computational efficiency via the RA technique is that early retrospective iterations, based upon a small number of sample paths, provide information for later iterations. Since the ARTA parameter estimation is the most computationally demanding part of our inventory-target estimation procedure, keeping the number of sample paths small in early iterations takes us quickly to close proximity of k^*. Because we choose the initial solution of the deterministic root-finding algorithm within close proximity to k^*, not much computational effort is expended for solving the sample-path problems in later iterations. Furthermore, as the number of sample paths increases, the standard error of the retrospective estimate of k^* decreases quickly, and the solutions obtained in the latter retrospective iterations converge quickly to k^*.

Algorithm 2. *Approximation of k^* to obtain improved inventory-target estimates*

    Step 1. *Initialization:*
        (1.1) Set the initial sample-path size to $N_1 = 100$ with the sample-path increase rule as $N_i = \lfloor 1.1N_{i-1} \rfloor$ for the retrospective iterations $i = 2, 3, \ldots$ and the error tolerance decrease as $\varepsilon_i = 0.1/\sqrt{N_i}$ for $i = 1, 2, \ldots;$
        (1.2) Define the initial solution $\bar{k}_0$ as τ_φ and set $i \leftarrow 1;$
    Step 2. *Generate N_i independent n-dimensional ARTA demand vectors and construct the deterministic root finding problem $G_{N_i}(k) = 0;$*
    Step 3. *Use the Newton-Raphson method to solve $G_{N_i}(k) = 0$ with the starting solution $\bar{k}_i$ and the error tolerance ε_i to obtain the ith root estimate k_{N_i}; i.e., return k_{N_i} if $|k_{N_i} - \tilde{k}_i^*| < \varepsilon_i$ with $\tilde{k}_i^*$ the solution of $G_{N_i}(k) = 0;$*
    Step 4. *Calculate the root estimate $\bar{k}_i$ as the weighted sum of the solutions $\{k_{N_j}\}_{j=1}^i$; i.e., $\bar{k}_i = (\sum_{j=1}^i N_j)^{-1} \sum_{j=1}^i N_j k_{N_j};$*
    Step 5. *If $i \geq 10$ and $\bar{k}_i - \bar{k}_{i-1} < 0.001$, then return $k^* = \bar{k}_i$; if not, set $i \leftarrow i + 1$ and go to Step 2.*

The selection of the sample-path increase rule as well as the error-tolerance decrease rule in Step (1.1) has a critical effect on the solution quality as a function of the computational time. We address this issue by choosing the sample-path increase and error tolerance decrease rules from the classes suggested by Pasupathy (2010) to achieve overall efficiency. We terminate the algorithm by requiring at least 10 retrospective iterations because the termination criterion in Step 5 can lead to premature stopping when the number of retrospective iterations is small. The retrospective solutions $\{\bar{k}_i; i = 1, 2, \ldots\}$ obtained in Algorithm 2 converge to k^* with probability one; see Theorem 2 in Pasupathy (2010). The numerical analysis in Section 4.2 shows that the finite-sample convergence of the retrospective solutions to k^* is also very fast; i.e., the number of iterations to achieve the stopping criterion does not exceed 39 in our experiments.

4.2 Results

This section presents the results of our numerical analysis with the following objectives: (i) To quantify the inaccuracy (i.e., the expected cost due to parameter uncertainty) associated with the maximum-likelihood inventory-target estimate $\tilde{I}$; (ii) To investigate the reduction in inaccuracy by using the improved inventory-target estimate $\tilde{I}(k^*)$ instead of $\tilde{I}$. We assume that the marginal demand distribution is from the Johnson S_N and S_L families with mean 100 and varying levels of coefficient of variation. To be specific, Table 1

Table 1: Demand is Johnson S_N with mean 100 and coefficient of variation 0.1.

					r				
	-0.9	-0.8	-0.7	-0.6	0.6	0.7	0.8	0.9
Minimum Expected Cost	11.6	16.0	19.0	21.3	21.3	19.0	16.0	11.6
Inaccuracy with $\tilde{I}$	6.2	9.2	11.3	12.8	16.2	14.2	12.4	9.4
Inaccuracy with $\tilde{I}(k^*)$	3.5	4.6	5.6	6.2	7.3	6.7	5.6	4.1
k^*	3.075	3.071	3.079	3.083	3.188	3.232	3.257	3.340

Table 2: Demand is Johnson S_L with mean 100 and coefficient of variation 0.5.

						r				
	-0.9	-0.8	-0.7	-0.6	-0.5	0.5	0.6	0.7	0.8	0.9
Min. Exp. Cost	69.7	105.6	134.5	156.1	175.2	176.5	156.6	133.6	105.0	70.2
Inacc. with $\tilde{I}$	39.8	55.5	72.2	86.4	102.3	110.6	102.4	92.4	72.8	51.7
Inacc. with $\tilde{I}(k^*)$	25.4	37.1	51.9	64.2	77.4	81.9	75.1	63.6	47.2	31.0
k^*	2.934	2.897	2.851	2.802	2.792	2.824	2.871	2.937	2.978	3.095

Table 3: Demand is Johnson S_L with mean 100 and coefficient of variation 5.

					r		
	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3
Minimum Expected Cost	512.2	944.8	1408.7	1786.7	2101.8	2401.0	2620.5
Inaccuracy with $\tilde{I}$	235.3	447.8	740.9	956.1	1103.7	1316.8	1642.6
Inaccuracy with $\tilde{I}(k^*)$	226.0	447.9	732.7	924.6	1034.7	1185.7	1401.7
k^*	2.480	2.341	2.261	2.198	2.155	2.121	2.077
	0.9	0.8	0.7	0.6	0.5	0.4	0.3
Minimum Expected Cost	526.6	940.1	1404.0	1822.5	2086.3	2430.7	2616.0
Inaccuracy with $\tilde{I}$	264.7	514.8	814.2	1194.0	1244.8	1426.5	1688.6
Inaccuracy with $\tilde{I}(k^*)$	254.6	513.2	775.3	1086.1	1115.3	1266.2	1397.6
k^*	2.513	2.298	2.211	2.138	2.120	2.059	2.067

presents the minimum expected cost under complete knowledge of demand parameters, the inaccuracy values associated with $\tilde{I}$ and $\tilde{I}(k^*)$, and the value of the bias parameter k^* returned by Algorithm 2 when the demand is from the Johnson S_N family with a coefficient of variation that is equal to 0.1. Table 2 and Table 3 present these results for the Johnson S_L demand with coefficient of variations equal to 0.5 and 5. Since the inaccuracy in inventory-target estimation is most problematic when the asymmetry of the newsvendor's loss function is high and the length of the historical data is limited, we let $\varphi = 0.99$ and $n = 10$. We report the inaccuracy results by setting the relative error of the inaccuracy to be $\leq 1\%$ with at least 95% probability; i.e., $\alpha = 0.95$ and $\varepsilon = 0.01$ in Algorithm 1. It is important to note that while we provide Algorithm 1 to approximate the inaccuracy associated with $\tilde{I}$, it is straightforward to approximate the inaccuracy of the inventory-target estimate $\tilde{I}(k^*)$ by replacing τ_φ with k^* in Step 2.

We observe that the inaccuracy in inventory-target estimation can be largely compared to the minimum expected cost with complete knowledge of demand parameters; and our improved inventory-target estimates leads to considerable reduction in this inaccuracy. For example, an inventory manager who builds the inventory target by simply using the estimates of the demand parameters as if they were the true values may end up with 81% (i.e., 9.4/11.6) greater expected cost than the optimal inventory target when the autocorrelation is 0.9 (Table 1). In this particular case, we further see that the use of improved inventory-target estimate eliminates 56% (i.e., $(9.4-4.1)/9.4$) of the inaccuracy. The use of bias parameter k^* instead of τ_φ provides similar improvements for all other demand distributions and autocorrelation values in our

experiments. It is important to notice in Table 3 that the bias parameter returned by Algorithm 2 is not always greater than τ_φ depending on the marginal demand distribution and the value of autocorrelation. That is, the improved inventory-target estimate $\tilde{I}(k^*)$ can be not only greater but also less than $\tilde{I}$ while hedging against demand parameter uncertainty.

5 CONCLUSION

This paper studies the problem of estimating inventory targets in a newsvendor setting when the demand process is autocorrelated and only a limited amount of historical data is available. Specifically, we represent the autocorrelated demand process with the highly flexible ARTA model that allows us to capture the autocorrelation in the demand process and a wide variety of distributional shapes for the marginal demand. We approximate the expected cost due to demand parameter uncertainty using a simulation-based sampling algorithm. Finally, we use sample-path optimization to obtain the value of the bias parameter to reduce the impact of parameter uncertainty in inventory-target estimation. Numerical analysis demonstrates the effectiveness of the resulting procedure in eliminating the inaccuracy in inventory-target estimation.

REFERENCES

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

Akcay, A., B. Biller, and S. Tayur. 2011b. "Inventory-target estimation in the presence of finite autocorrelated demand data". Tepper Working Paper 2011–E3. Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA.

Barton, R. R., B. L. Nelson, and W. Xie. 2010, December. "A framework for input uncertainty analysis". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 1189–1198. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Biller, B., and C. G. Corlu. 2011. "Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs". *Operations Research* 59 (3): 661–673.

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

Chen, H., and B. W. Schmeiser. 2001. "Stochastic root finding via retrospective approximation". *IIE Transactions* 33 (3): 259–275.

Chick, S. E. 2001. "Input distribution selection for simulation experiments: Accounting for input uncertainty". *Operations Research* 49 (5): 744–758.

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

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

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

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

Kuhl, M. E., J. S. Ivy, E. K. Lada, N. M. Steiger, M. A. F. Wagner, and J. R. Wilson. 2010. "Univariate input models for stochastic simulation". *Journal of Simulation* 4 (2): 81–97.

Lee, H. L., K. C. So, and C. S. Tang. 2000. "The value of information sharing in a two-level supply chain". *Management Science* 46 (5): 626–643.

Lewis, P. A. W., E. McKenzie, and D. K. Hugus. 1989. "Gamma processes". *Stochastic Models* 5 (1): 1–30.

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

Luong, H. T. 2007. "Measure of bullwhip effect in supply chains with autoregressive demand process". *European Journal of Operational Research* 180 (1): 1086–1097.

Mallows, C. L. 1967. "Linear processes are nearly Gaussian". *Journal of Applied Probability* 4 (2): 313–329.

Ng, S. H., and S. E. Chick. 2006. "Reducing parameter uncertainty for stochastic systems". *ACM Transactions on Modeling and Computer Simulation* 16 (1): 26–51.

Pasupathy, R. 2010. "On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization". *Operations Research* 58 (4): 889–901.

Pasupathy, R., and S. Kim. 2011. "The stochastic root-finding problem: Overview, solutions, and open questions". *ACM Transactions on Modeling and Computer Simulation* 21 (3): 1–23.

Robert, C. P., and G. Casella. 1999. *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

Shapiro, A. 2004. "Monte Carlo sampling methods". In *Handbooks in Operations Research and Management Science, Stochastic Programming*, edited by A. Ruszczynski and A. Shapiro, 353–426. Amsterdam: North-Holland.

Sklar, A. 1959. "Fonctions de répartition à n dimensions et leurs marges". *Publications de l'Institut de Statistique de L'Université de Paris* 8 (1): 229–231.

Wei, W. W. S. 1990. *Time Series Analysis: Univariate and Multivariate Methods*. New York: Addison Wesley.

Zouaoui, F., and J. R. Wilson. 2003. "Accounting for parameter uncertainty in simulation input modeling". *IIE Transactions* 35 (9): 781–792.

Zouaoui, F., and J. R. Wilson. 2004. "Accounting for input-model and input-parameter uncertainties in simulation". *IIE Transactions* 36 (11): 1135–1151.

AUTHOR BIOGRAPHIES

ALP AKCAY is a Ph.D. candidate of Operations Management and Manufacturing at Carnegie Mellon University. His primary research interests include data-driven decision making under uncertainty with particular applications in inventory management and finance.

BAHAR BILLER is an Associate Professor of Operations Management and Manufacturing at Carnegie Mellon University. Her primary research interest lies in the area of computer simulation experiments for stochastic systems and more specifically, in the simulation methodology for dependent input processes with applications to financial markets and supply chains.

SRIDHAR TAYUR is the Ford Distinguished Research Chair and Professor of Operations Management at Carnegie Mellon University's Tepper School of Business. He has published many scholarly publications in journals such as *Operations Research*, *Management Science*, and *MSOM Journal*, is co-editor of the widely referenced book, *Quantitative Models for Supply Chain Management*, and has served on the editorial boards of *Operations Research*, *Journal of Optimization and Engineering*, *NRLQ*, *MSOM Journal* and *Management Science*. Dr. Tayur also served as President of Manufacturing and Services Operations Management (MSOM) Society. He is the founder of the software company SmartOps Corporation (2000-) as well as the founder of OrganJet Corporation (2011-).