Computers & Operations Research 68 (2016) 123–133

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Robust newsvendor problem with autoregressive demand

Emilio Carrizosa [a], Alba V. Olivares-Nadal [a,*], Pepa Ramírez-Cobo [b]

[a] Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Instituto de Matemáticas de la Universidad de Sevilla, Av. Reina Mercedes, s/n, 41012 Sevilla, Spain
[b] Department of Statistics and Operational Research, Universidad de Cádiz, Spain

ARTICLE INFO

Available online 1 December 2015

Keywords:
Distribution-free newsboy problem
Autoregressive process
Uncertainty set
Minimax
Robust optimization
Forecasting

ABSTRACT

This paper explores the single-item newsvendor problem under a novel setting which combines temporal dependence and tractable robust optimization. First, the demand is modeled as a time series which follows an autoregressive process $AR(p)$, $p \geq 1$. Second, a robust approach to maximize the worst-case revenue is proposed: a robust distribution-free autoregressive method for the newsvendor problem, which copes with non-stationary time series, is formulated. A closed-form expression for the optimal solution is found for $p=1$; for the remaining values of p, the problem is expressed as a nonlinear convex optimization program, to be solved numerically. The optimal solution under the robust method is compared with those obtained under three versions of the classic approach, in which either the demand distribution is unknown, and autocorrelation is neglected, or it is assumed to follow an $AR(p)$ process with normal error terms. Numerical experiments show that our proposal usually outperforms the previous benchmarks, not only with regard to robustness, but also in terms of the average revenue. Extensions to multiperiod and multiproduct models are also discussed.

1. Introduction

The single-period problem (SPP), also known as the newsvendor problem, is a simple yet rich inventory model which has been widely studied in the Operations Research field due to its versatility and applicability to many business decision problems, in fields such as managing booking and capacity in airlines companies [57], health insurances [49,23], scheduling [4], retailers and managers order quantity decision in sports and fashion industries [26].

The basic version of the problem consists in making a one-step decision on the quantity Q to be bought of one single perishable product under the assumption that the demand is a random variable with known distribution F. If the decision maker buys each unit at cost u and sells it at price v, then the expected revenue is maximized by buying exactly $Q^* = F^{-1}\left(1-\frac{u}{v}\right)$ units.

Numerous variants of the classical SPP have been proposed in the literature; some of them will be discussed below, but for a fuller account of the subject we refer the reader to Khouja [38], Petruzzi and Dada [46], and Qin et al. [47].

The traditional assumption that the demand probability distribution is known may be unrealistic in many cases. In addition, if the demand is inferred from sample data, then the resulting estimate may lack of desirable statistical properties (consistency, asymptotic normality, etc.), for example, for small sample sizes. To overcome these and other related problems, some distribution-free approaches have been considered in the literature, Scarf [50] being the first to give a closed-form solution to the newsvendor problem when only the demand mean and the variance are assumed to be known. Two more remarkable distribution-free works are Gallego and Moon [26], which provided an extension to Scarf's solution, and Yue et al. [58], in which the demand density function is assumed to belong to a specific family of density functions. Other articles which cope with demand uncertainty are Ding et al. [21], Dana and Petruzzi [20], and Godfrey and Powell [28]. However, as pointed out in See and Sim [52] and Bandi and Bertsimas [5], not only the assumption of known distribution of the demand may be too strong, but also to estimate the mean and variance from the sample data and accommodate such estimates to an assumed distribution function may generate drastic errors in the inventory policy. Moreover, demand is in fact usually correlated along time, so assuming that demands for each period are independent and identically distributed is in practice unrealistic [41,29,37]. Some authors have studied inventory models with time-correlated demand, including AR models [2,48,35], compound Poisson processes [53], martingale models of forecast evolution [22,44,56], factor models [52] or estimation via Kalman filter [3]. Most of these papers either assume perfect knowledge of the distribution function [42,2,3,53,56,48] or are focused in calculating bounds of the objective function, which are distribution-

* Corresponding author. Tel.: +34 637 872 829.
E-mail address: aolivares@us.es (A.V. Olivares-Nadal).

http://dx.doi.org/10.1016/j.cor.2015.11.002
0305-0548/

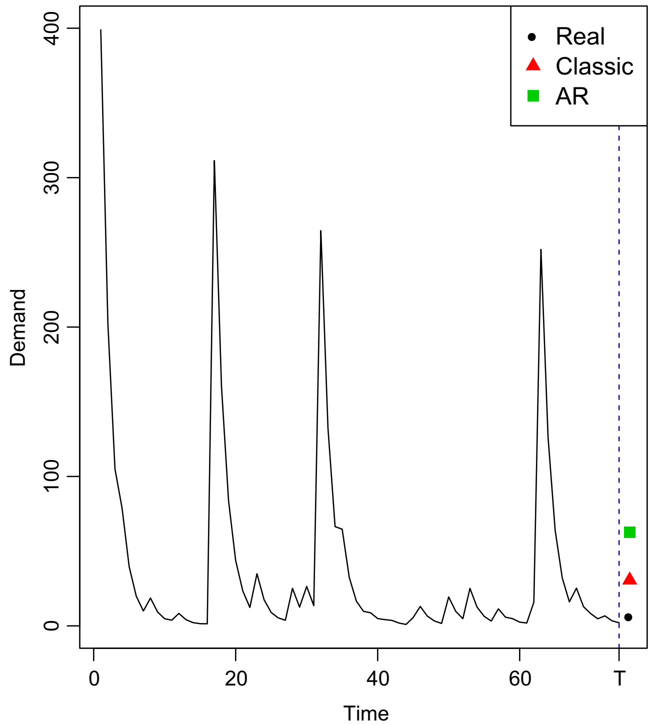

Fig. 1. Demand time series generated from an autoregressive process with heavy-tailed errors, and real and predicted revenues under two classic approaches (classic newsvendor and *AR*).

free in See and Sim [52], Lu et al. [44], and Dong and Lee [22]. In contrast, in the work developed here no distributional assumptions are made and the optimal solution is obtained with a closed expression for a particular case, while the problem to be solved in the remaining cases is extremely tractable due to its structural properties: it is a low-dimensional convex problem. Moreover, we do not only cope with temporal demand but also take into account robustness in terms of uncertainty and risk aversion, which provides novelty to this paper.

Different inventory policies yield rather different revenues. For instance, consider Fig. 1, which depicts a time series for the demand, assumed to follow an autoregressive process with log-normal errors. It can be observed that, in this case, conventional methods such as the classic newsvendor and autoregressive approaches, which will be described in detail in Section 3.1, can lead to losses. In particular, for the data of Fig. 1, the classic model yields a negative revenue of -1.21, while when assuming the basic *AR* model revenues decrease to -9.21. In contrast, robust approaches are usually too conservative and avoid ordering any quantity of product [12,43]. The approach proposed in this paper successfully copes with heavy-tailed demands and usually outperforms both classic and autoregressive approaches in terms of average revenue and also probability of losses, while avoiding over conservativeness of previous robust approaches.

In recent years, risk-averse models have received increasing attention in the inventory literature (see [19] and the references therein). There, instead of maximizing the expected revenues, other utilities are optimized. Depending on the decision maker's preferences, it may be reasonable, for example, optimizing the probability of achieving a target profit (see for example [36] or [39]), the Return of Investment [54], the Cost-Volume-Profit, the CVaR [17] or other risk-averse policies [24,19]. The above-mentioned paper of Yue et al. [58], as well as Perakis and Roels [45], Zhu et al. [60], and Jiang et al. [34], considers the minimax regret decision criterion instead. The robust approach in the newsvendor problem deals with uncertainty in the demand while minimizing the impact over the optimal solution of the worst-case scenario. For example, the landmark Scarf's rule adopts such a criterion, although it enforces independence of the demand along time. Bertsimas and Thiele [12] also propose a robust inventory approach, where uncertainty intervals for the demand are supposed to be already given. On the contrary, our approach would address the worst-case analysis while coping with time-correlated demands and including information of the historical observations of the demand into the model.

In this paper we address the newsvendor problem from a new perspective, integrating a distribution-free design with temporal dependence in the demand, into a robust optimization approach. Throughout the paper we perform a worst-case analysis, seeking the policy maximizing the worst-case revenue. Specifically, our main contributions are:

1. We consider the demand as a time series with non-negligible autocorrelation coefficients. For simplicity, the basic yet versatile autoregressive process of some order p, $AR(p)$, is used as time series model. We follow a distribution-free approach in the sense that no distributional assumption is imposed over the error terms of the autoregressive model.
2. We implement a robust optimization method based on the uncertainty sets of Bandi and Bertsimas [5], where the goal is to minimize the losses in the worst-case realization of the parameters. For the particular case when the uncertainty set is modeled with the l_2-norm, a closed-form expression for the optimal solution is obtained in the case $p=1$. For $p \geq 2$ the problem turns into a tractable nonlinear convex optimization program, solved numerically.
3. We show that our approach outperforms three different classic approaches. In the first one, the demand distribution is assumed to be unknown and is estimated from the sample observations, which are assumed to be independent; in the second one, the demand distribution is assumed to follow an $AR(p)$ process with normal error terms; the third method is the robust distribution-free solution of Scarf [50].
4. We briefly discuss the robust multi-product newsvendor problem with demands correlated over time and between products, and the multi-period case.

The paper is organized as follows. In next section we briefly introduce autoregressive processes and model various robust newsvendor problems with autoregressive demands. Specifically, we formulate the single-item single-period case in terms of an optimization problem in Section 2.1. We discuss the choice of parameters in Section 2.1.1, while in Section 2.1.2 we show that, for a particular case, the problem is a smooth convex optimization problem, and we obtain a closed-form solution for $p=1$. A brief extension to the multi-period case is outlined in Section 2.1.3, where the robust modelling of the autoregressive processes is integrated into the inventory model of Bertsimas and Thiele [12]. An extension to the multiple item newsvendor problem is carried out in Section 2.2, where the demands of the products are assumed to follow a Vector Autoregressive Process. In Section 3 we design and present some numerical examples, where the robust autoregressive model is tested against three different but classic methods, outlined in Section 3.1. Data generation and presentation of results are addressed in Sections 3.2 and 3.3 respectively. Last section is devoted to concluding remarks and extensions.

2. The model

We start this section with a short discussion on autoregressive processes, which will model the demand of our SPP. Because of their simplicity and versatility, autoregressive processes have been widely used to model time series in different contexts where the temporal dependence is significant. A time series $\{X_t,\ t>0\}$ follows an autoregressive process of order $p \geq 1$ (noted $AR(p)$) if it can be expressed in the form

$$X_t = c + \sum_{k=1}^{p} \theta_k X_{t-k} + a_t, \tag{1}$$

where $c, \theta_1, \ldots, \theta_p$ are coefficients and $\{a_t,\ t>0\}$ is the sequence of i.i.d model's error terms with expected value μ_a and variance $\sigma_a{}^2$, for all $t>0$. For a more detailed description of the autoregressive process and interpretation of the coefficients in the model, see for example Box et al. [14].

If the parameters in (1) are given (either they are known or estimated from sample data, available up to time T), then (1) can be used to forecast the process. In particular, if the errors are assumed to follow a normal distribution, then the $(1-\alpha)\%$ prediction interval for a forecasted value at time $T+1$ is given by

$$\hat{X}_{T+1} \pm z_{1-\alpha/2}\sigma_a \tag{2}$$

where $\hat{X}_{T+1}$ is the estimated forecast, and $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution. When the errors are not normally distributed, the use of (2) may lead to inaccurate results, as will be illustrated in detail in Section 3. The influence of the error distribution over the demand can be seen in Fig. 2, from which it is clear that normally distributed errors may be too restrictive as they are not able to capture extreme behavior of the demand, which is common in practice [13].

Now we describe how to obtain a robust solution for the single period newsvendor problem with $AR(p)$ demand, under the assumption that a realization of the process up to time T is available and no probability distribution is imposed on the errors. To do this we use the uncertainty sets of Bandi and Bertsimas [5].

2.1. The single-product case

In this section we focus on the single-item newsvendor with autoregressive demand, obtaining a closed-form solution for a particular case when the demand follows and $AR(1)$ process. The solution we obtain is robust in two senses. First, we optimize revenue under the worst-case scenario of the demand. Second, the forecast of the demand is distribution-free, although it takes into account autocorrelation. We undertake this task via a robust optimization problem with uncertainty sets [7,8]: the data is not deterministic but required to belong to an uncertainty set, and all the constraints of the optimization problem must hold for every valid value of the uncertainty set.

Ben-Tal and Nemirovski [9] show that nominal solutions of optimization problems may become drastically unfeasible under small perturbations in the data. Although they recognize the need of robustifying the nominal problem, they also admit that this is done under a cost over the optimal value. Therefore, risk aversion against uncertainty in the data has often been addressed by alleviating the impact of the worst case scenario over the solution [10,11,6]. As mentioned in Hanasusanto et al. [32], the worst case approach is strongly justified from the theory of choice in economics [25,27]. Furthermore, this risk averse approach has been frequently used in the newsvendor context [50,26,55,31,32]. Although this type of robust solutions might be too conservative, Bertsimas and Thiele [12] show that the choice of the uncertainty sets can play an important role to help avoiding this over conservativeness. In this section we formulate our robust newsvendor problem with autoregressive demand by means of a minimax optimization problem. We give some guidelines on how to choose the parameters of our model so as to take into account risk preferences and avoid over conservativeness.

Assume that we possess the historical data of the demand up to time T, $\{X_t\}_{t=1}^{T}$, but it is unknown for next period. Moreover assume that such demand roughly follows an autoregressive process as in (1). Let Q denote the quantity of product to order, assumed to belong to an interval $\mathbb{Q} = [\underline{Q}, \overline{Q}]$. If the demand takes the value X_{T+1}, then a revenue $R(Q, X_{T+1})$ is obtained. Typically, when no shortage costs nor salvage values are taken into account, R is calculated using the following expression:

$$R(Q, X_{T+1}) = \min\{Q, X_{T+1}\} - \left(\frac{u}{v}\right)Q. \tag{3}$$

where u and v are the unit cost and selling price respectively. The goal is to find Q maximizing the revenue $R(Q, X_{T+1})$ under the worst-case scenario for the demand X_{T+1}, assumed to vary in a given uncertainty set.

We formulate the robust single-item newsvendor problem as follows:

$$\max_{Q \in \mathbb{Q}} \min_{X_{T+1}} \quad R(Q, X_{T+1}) \tag{4}$$

$$\text{s.t.} \quad \left| \frac{1}{T-p} \sum_{t=p+1}^{T} a_t \right| \leq \Gamma_1, \tag{5}$$

$$\|(a_{p+1}, \ldots, a_T)\|_q \leq \Gamma_2, \tag{6}$$

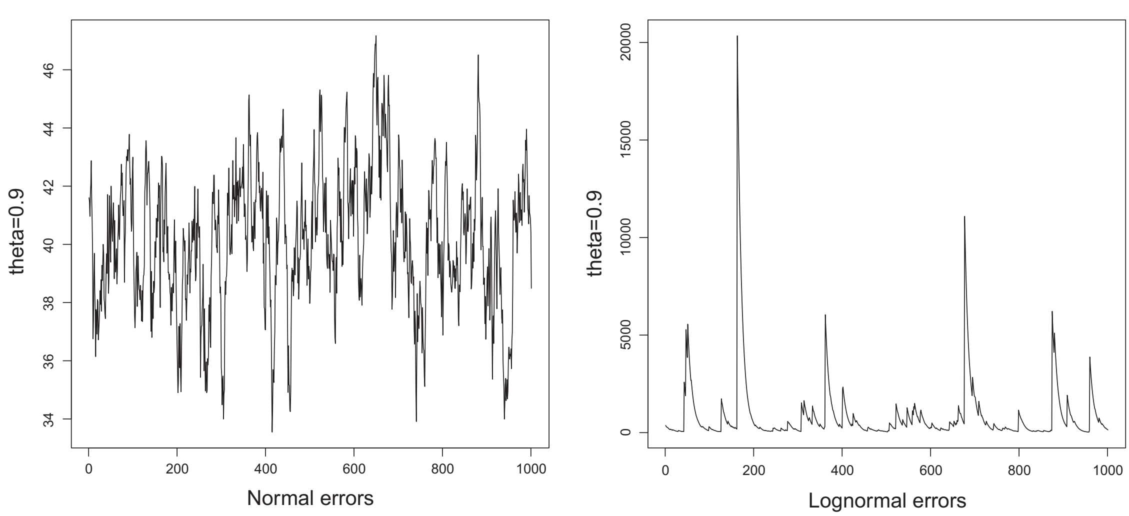

Fig. 2. Examples of highly correlated autoregressive demands generated with errors following $N(4,1)$ and $LN(0,3)$.

$$X_t = c + \sum_{k=1}^{p} \theta_k X_{t-k} + a_t \quad t = p+1, \ldots, T+1, \tag{7}$$

$$|a_{T+1}| \leq \Delta \tag{8}$$

$$a \leq X_{T+1} \leq b \tag{9}$$

The rationale behind the constraints is the following. The vector $(a_{p+1}, \ldots, a_T)$ is the vector of residuals in the $AR(p)$ model in constraint (7). Constraint (6) forces the error vector to be small, by bounding its l_q norm. Typical choices are $q=1$, and thus we make small the absolute value of the residuals, and for $q=2$ we bound their variance. On the other hand constraint (5) requires the mean of the observed errors to be bounded. Note that, as no constraints over the θ_k $k=1,\ldots,p$ are made, nonstationary processes can be addressed by problem (4–9). Finally, (8) implies that the absolute value of the random value a_{T+1} (which represents the prediction error) is bounded above by some constant Δ. In some real-world situations, the nature of the time series requires to obtain a value of X_{T+1} within a specific interval. This is for example the case of rainfall data and exchange rates, which take non-negative values, or unemployment rates and diseases prevalence, which must lie in [0, 1] (see [16]). Constraint (9) expresses this requirement.

Denote the set of constraints (5–9) by $\mathbb{X}$, and fix $Q_0 \in \mathbb{Q}$. As the function $R(Q_0, \cdot)$ decreases in X_{T+1}, then $\min_{X_{T+1} \in \mathbb{X}} R(Q_0, X_{T+1}) = R(Q_0, \underline{X}_{T+1})$, where $\underline{X}_{T+1}$ is either a, b, or the solution to problem *ARUS*(p), defined as:

$$\min_{\substack{c,\theta_1,\ldots,\theta_p,\\ a_1,\ldots,a_{T+1}}} \quad X_{T+1} \tag{ARUS(p)}$$

s.t. $\quad (5), (6), (7), (8)$

We will refer to this problem as an Autoregressive process based on Uncertainty Sets, in short *ARUS*(p). Then, the solution Q^* to problem (4–9) is also the solution to

$$\max_{Q \in \mathbb{Q}} \left(\min\left\{ Q, \underline{X}_{T+1} \right\} - \left(\frac{u}{v}\right) Q \right).$$

As $R(Q, \underline{X}_{T+1}) = \min\left\{Q, \underline{X}_{T+1}\right\} - \left(\frac{u}{v}\right)Q$ is piecewise linear concave function in Q, with a global maximum at $\underline{X}_{T+1}$, then the solution to problem (4–9) is finally given by:

$$Q^* = \begin{cases} \underline{Q} & \text{if } \underline{Q} \geq \underline{X}_{T+1} \\ \underline{X}_{T+1} & \text{if } \underline{Q} \leq \underline{X}_{T+1} \leq \overline{Q} \\ \overline{Q} & \text{if } \underline{X}_{T+1} \geq \overline{Q} \end{cases} \tag{10}$$

2.1.1. Choice of parameters

The values of Γ_1, Γ_2 and Δ in the formulation of the *ARUS*(p) problem are chosen according to the practitioner's criterion. Next, we describe a procedure to select such parameters using the concept of uncertainty sets along the lines of Bandi and Bertsimas [5], which define the following uncertainty set for i.i.d. random variables $Y_1, \ldots, Y_r$ using the Central Limit Theorem (CLT):

$$\mathcal{U}(\Gamma_1^*) = \left\{ (Y_1, \ldots, Y_r) \quad \text{s.t.} \left| \frac{1}{r} \sum_{i=1}^{r} Y_i \right| \leq \frac{\Gamma_1^* \sigma_Y \sqrt{r}}{r} + \mu_Y \right\}, \tag{11}$$

where μ_Y and σ_Y stand respectively for the mean and the standard deviation of the random variables $Y_1, \ldots, Y_r$, and the value Γ_1^* is a small constant that influences the accuracy of the fit. Since the error terms $a_{p+1}, \ldots, a_T$ are assumed i.i.d., we can define an uncertainty set of the form (11) for them. As there is no loss of generality in assuming $\mu_a = 0$, in combination with (7) we would have the following uncertainty set:

$$\mathcal{U}(\Gamma_1^*) = \left\{ (c, \theta_1, \ldots, \theta_p) \quad \text{s.t.} \frac{1}{T-p} \left| \sum_{t=p+1}^{T} X_t - c - \sum_{k=1}^{p} \theta_k X_{t-k} \right| \leq \frac{\Gamma_1^* \sigma_a \sqrt{T-p}}{T-p} \right\}.$$

Since σ_a is unknown in practice, we suggest to set it to:

$$\sigma_a \approx (1 + \nu/100)\sigma_0,$$

where $(1+\nu/100)$ indicates a perturbation (depending on the value of ν) of σ_0, which denotes the optimal value to the problem of minimizing the variance of the errors $a_{p+1}, .., a_T$.

Consider now the constraint (6). For $q=2$ this constraint bounds the variance of the errors. In this case we can proceed similarly as before and substitute Γ_2 by a certain perturbation of the minimum value attained by the errors' variance:

$$\Gamma_2 \approx (1 + \beta/100)\gamma_2,$$

where γ_2 is the obtained by minimizing the variance of the errors subject to constraints (5) and (7).

Finally, consider (8). The choice of Δ is crucial, since it bounds the value of the prediction error a_{T+1}. Since the sequence $\{a_t, t>0\}$ is i.i.d., it seems reasonable to relate Δ with the values $a_1, a_2, \ldots, a_T$. Since we have already obtained the optimal $c^\star, \theta_1^\star, \ldots, \theta_p^\star$ feasible for (5) and (7) and for which the errors' variance is minimum, then it is straightforward to obtain the values $a_1^*, \ldots, a_T^*$ (by substituting $c^\star, \theta_1^\star, \ldots, \theta_p^\star$ into (1)). Therefore, one possible choice of Δ is the empirical k-th quantile of the sample $(a_1^*, \ldots, a_T^*)$, for some value of k, large enough. In the context of the newsvendor problem it seems reasonable to relate the value of Δ to the *profitability* of the product $\frac{u}{v}$, which should influence the decision maker's risk aversion. From Schweitzer and Cachon [51], in the case of high profit products (that is, $\frac{u}{v} \leq 0.5$), risk aversion may be reduced by allowing the decision maker to buy more items. This implies that $\underline{X}_{T+1}$ is allowed to be higher, which may be obtained by reducing the value of Δ. Thus, in the context of the newsvendor problem, it makes sense to choose Δ as the empirical $U\left(\frac{u}{v}\right)$–th quantile of the sample $(a_1, \ldots, a_T)$, where U is a utility function that depends on the decision maker's risk aversion.

2.1.2. Properties and closed-form solution for $q=2$

In this section we analyze the properties of the robust newsvendor problem with autoregressive demand (4–9) for $q=2$ and we obtain a closed-form solution when the demand follows an *AR* (1).

Consider problem

$$-\Delta + \min_{\boldsymbol{\theta}} \left(\frac{-\sum_{t=p+1}^{T} \varphi_t(\boldsymbol{\theta})}{T-p} - \min\left\{ \Gamma_1, \sqrt{\Gamma_2 - H(\boldsymbol{\theta})} \right\} + \sum_{k=1}^{p} \theta_k X_{T+1-k} \right) \tag{12}$$

$$\text{s.t.} \quad \Gamma_2 - H(\boldsymbol{\theta}) \geq 0, \tag{13}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, and $\varphi_t(\boldsymbol{\theta})$, $H(\boldsymbol{\theta})$ are defined as

$$\varphi_t(\boldsymbol{\theta}) = \sum_{k=1}^{p} \theta_k X_{t-k} - X_t,$$

$$H(\boldsymbol{\theta}) = \frac{1}{T-p} \sum_{t=p+1}^{T} \varphi_t(\boldsymbol{\theta})^2 - \frac{1}{(T-p)^2} \left(\sum_{t=p+1}^{T} \varphi_t(\boldsymbol{\theta}) \right)^2,$$

and fix the parameters Γ_1, Γ_2 and Δ. Then the next results hold.

Proposition 1. *The solution to problem (4–9) is either* $\underline{Q}$, $\overline{Q}$, *a, b, or the solution to (12–13).*

Proof. See Appendix A.

Proposition 2. *Eqs.* (12) and (13) *define a smooth convex optimization problem.*

Proof. See Appendix B.

Moreover, the global optimum in the case $p=1$ can be obtained in closed form as the next result shows.

Theorem 1. *For $p=1$, the optimal solution to the minimization problem defined by (12) and (13) is reached at one of these values for* $\theta_1^\star$:

$$\theta_1^{\star(1)} = \frac{-C_{1,0} \pm \sqrt{C_{1,0}^2 + V_1\left(\Gamma_2 - V_0 - \Gamma_1^2\right)}}{-V_1}, \tag{14}$$

$$\theta_1^{\star(2)} = \frac{-C_{1,0} \pm \sqrt{C_{1,0}^2 + V_1(\Gamma_2 - V_0)}}{-V_1}, \tag{15}$$

$$\theta_1^{\star(3)} = \frac{-C_{1,0}(V_1+a^2) \pm \sqrt{C_{1,0}^2(V_1+a^2)^2 + a^2 V_1 (V_1+a^2)(\Gamma_2 - V_0 - C_{1,0})}}{-V_1}, \tag{16}$$

where

$$V_k = var(X^k), \quad C_{k,h} = cov(X^k, X^h) \tag{17}$$

respectively denote the variance of X^k and covariance matrix between X^k and X^i where

$$X^k = (X_{p+1-k} \ldots, X_{T-k}), \tag{18}$$

and $a = \frac{-1}{T-1} S_1 + X_T$.

Proof. See Appendix C.

2.1.3. The multi-period case

It is reasonable to study the multi-period newsvendor problem whenever the product perishes at a rate that extends to more than one period. Bertsimas and Thiele [12] propose a robust multi-period inventory approach in which prediction intervals for the demand are supposed to be given. Although this approach deals with holding and shortage costs, it does not provide any information about the demand or guidelines on how to obtain its forecasts. Here we propose to use our modeling of the robust autoregressive demand to obtain prediction intervals that can be embedded into the multi-period approach of Bertsimas and Thiele [12]. In this way we are able to (i) take into account the temporal autocorrelation of the demand, (ii) provide robust forecasts for the demands, which fits in with the robust approach Bertsimas and Thiele [12] propose, and (iii) intimately bound the demand estimation with the inventory problem, as the forecasts can be done by taking into account the risk aversion of the user, which depends on the profitability of the product $\frac{u}{v}$.

Once provided the extremes values $\underline{X}_{T+1}$ and $\overline{X}_{T+1}$, a multi-stage prediction approach, as e.g. in Cheng et al. [18], is possible: we minimize and maximize the *ARUS*(p) at each time step and use the obtained values for the next period. Note that if the demand follows and *AR*(p) process then X_{T+l} depends on the p previous values, thus the prediction interval obtained for the unknown data in $X_{T+l-1}, \ldots, X_{T+l-p}$ may affect the forecast on instant $T+l$. The robust multi-stage approach we propose to estimate the demand consists of simply adding the constraints:

$$\underline{X}_{T+m} \le X_{T+m} \le \overline{X}_{T+m} \quad m = l-p, \ldots, l-1$$

to the *ARUS*(p) optimization problem when X_{T+m} is unknown. The values of $\underline{X}_{T+m}$ and $\overline{X}_{T+m}$ have been previously obtained by minimizing (respectively maximizing) the *ARUS*(p) problem for X_{T+m}. Note that the value of parameter Δ can be modified on each step to deal with variations in the selling price and the cost.

2.2. The multi-product case

In this section we analyze the extension to the multiple item newsvendor problem when the products are not replaceable. Many papers, such as Zhang [59], Lau and Lau [40], Abdel-Malek and Montanari [1], Choi et al. [19], have treated this problem considering demands possibly correlated amongst products. Without exception, they assume that the demands' marginal density functions are given for each product and disregard temporal correlation. In contrast, we propose a distribution-free approach and we assume that the demands of the products are not only correlated over time but also correlated between items. Specifically, we extend the model in (1) by assuming that the demands follow a Vector Autoregressive Process of or order p, *VAR* (p) thereafter. Such a process can be written as follows:

$$\mathbf{X}_t = \mathbf{c} + \sum_{k=1}^{p} B_k \mathbf{X}_{t-k} + \mathbf{a}_t \tag{19}$$

where $\mathbf{X}_t = \left(X_t^1, \ldots, X_t^N\right)'$ contains the demand of the N products in time instant t, $\mathbf{c} \in \mathbb{R}^N$ is a vector of constants, and $B_k \in \mathbb{R}^{N\times N}$ gathers the coefficients relating the values of the time series with the demand k periods behind. Finally $\mathbf{a}_t \in \mathbb{R}^N$ represents the contemporaneous error terms for all demands series at time instant t.

For the multi-product case we consider an uncertainty set gathering the constraints (5) and (6) for each product, while including the demands' correlation between time and items expressed by (19). In order to bound the prediction errors in such a way that a contemporaneous dependence between the shocks of the different products can be addressed, we add the constraint $\left\|\left(\frac{a_{T+1}^1}{\Delta^1}, \ldots, \frac{a_{T+1}^N}{\Delta^N}\right)\right\|_s \le 1$ to our multiproduct problem. Note that for $s=\infty$, this is equivalent to requiring (8) to hold for each product $j = 1, \ldots, N$.

In conclusion, we model the robust multi-item newsvendor with correlated demand as:

$$\max_{\mathbf{Q}\in\mathbb{Q}} \min_{\substack{X_{T+1}, \mathbf{c}, \\ B_k, \mathbf{a}_t}} \mathbf{R}(Q^1, \ldots, Q^r, X_{T+1}^1, \ldots, X_{T+1}^r) \tag{20}$$

$$\text{s.t.} \quad \left|\frac{1}{T-p}\sum_{t=p+1}^{T} a_t^j\right| \le \Gamma_1^j, \quad j=1,..,N \tag{21}$$

$$\|(a_{p+1}^j, \ldots, a_T^j)\|_q \le \Gamma_2^j, \quad j=1,..,N \tag{22}$$

$$\mathbf{X}_t = \mathbf{c} + \sum_{k=1}^{p} B_k \mathbf{X}_{t-k} + \mathbf{a}_t \quad t = p+1, \ldots, T+1 \tag{23}$$

$$\left\|\left(\frac{a_{T+1}^1}{\Delta^1}, \ldots, \frac{a_{T+1}^N}{\Delta^N}\right)\right\|_s \le 1 \tag{24}$$

Here $\mathbf{R}(Q^1, \ldots, Q^r, X_{T+1}^1, \ldots, X_{T+1}^r) = \sum_{j=1}^N \mathbf{R}_j(Q^j, X_{T+1}^j)$ and $\mathbb{Q}$ is a set of constraints on the quantity of products to acquire. Note that problem (20–24) takes into account the correlation of the demand in two aspects (time and products) while providing a distribution free approach. Define the function

$$\phi(Q^1, \ldots, Q^r) = \min_{\text{s.t. } (21)-(24)} \sum_{j=1}^{N} \mathbf{R}_j(Q^j, X_{T+1}^j)$$

which gives, for quantities $Q^1, \ldots, Q^r$, the worst-case revenue. Observe that, for each $\mathbf{R}_j$ as in (3), the function ϕ is concave and thus (20–24) is a concave maximization problem. It will be linearly constrained if, as customary in the literature, capacity, weight or budget constraints are written as linear functions [30,1].

3. Numerical illustrations

Now we are going to test the performance of our single-product approach against those described in Section 3.1. To make the results as complete as possible we have checked the obtained average revenue and small quantiles for a large number of simulated data sets with different properties, such as the correlation, distribution of errors or seasonality. To further explore the behavior of our approach, different values of p have been considered to generate possibly periodic $AR(p)$ processes, but only results for $p=1$ are included here as the same performance was observed for the other cases tested.

The proposed approach will be compared with three classic approaches, which will be called *static*, *Scarf* and $AR(p)$. In the first approach, the demand distribution will be unknown and estimated from the sample observations (assumed to be independent), the second assumes that the distribution function for the demand is uncertain, as in the third approach normality in the error terms will be assumed.

3.1. Benchmark methods

In this section we describe in detail the three benchmark approaches: *static*, *Scarf* and $AR(p)$.

If the demand distribution F were known and the expected revenue were to be maximized, then the optimal quantity $Q_s^\star$ would be given by a specific quantile of the distribution function, which depends on the cost (u) and selling prices (v) as follows:

$$Q_s^\star = F^{-1}\left(1-\frac{u}{v}\right)$$

Note that under this approach the temporal dependence in the data is ignored, so we call it *static* thereafter. In practice the distribution of the demand is unknown and therefore in the static approach, an estimation of the distribution function must be employed. Usually, the empirical distribution function $\hat{F}$, which converges to the true cumulative density function (cdf) F for a large enough sample, is considered.

The so-called Scarf's rule is the solution to the robust optimization problem

$$\min_{Q \geq 0} \max_{f \in H(\mu,\sigma)} G_f(Q)$$

where $H(\mu,\sigma)$ denotes the set of distributions with mean μ and variance σ, and $G_f(Q)$ is the expected total cost. The solution to this problem is given by

$$Q_S^\star = \begin{cases} 0 & \text{if } \frac{u}{v}\left(1+\frac{\sigma^2}{\mu^2}\right) > 1 \\ \mu+\frac{\sigma}{2}\left(\frac{1-2\frac{u}{v}}{\sqrt{\frac{u}{v}\left(1-\frac{u}{v}\right)}}\right) & \text{if } \frac{u}{v}\left(1+\frac{\sigma^2}{\mu^2}\right) < 1. \end{cases} \quad (25)$$

This solution is robust because it assumes uncertainty over the distribution function of the demand. However, as our numerical results show, this solution may suffer from an excessive conservativeness.

In the case of the classic $AR(p)$ approach the forecast was assumed to follow a normal distribution. Therefore, under normality, the optimal solution for the newsvendor problem would be

$$Q_{AR(p)} = \Phi^{-1}_{(\hat{X}_{T+1},\sigma_a)}\left(1-\frac{u}{v}\right)$$

where $\Phi^{-1}_{(\hat{X}_{T+1},\sigma_a)}$ is the inverse cdf of a normal distribution with mean $\hat{X}_{T+1}$ and standard deviation σ_a. Since $Q_{AR(p)}$ may be negative and the demand always takes non-negative values, the quantity $Q^\star_{AR(p)} = \max\{0, Q_{AR(p)}\}$ will be considered instead.

3.2. Synthetic data generation and experiments design

The performance of the different inventory policies is illustrated by different simulational experiments, for which samples of the $AR(p)$ process (representing the demand series) are artificially generated. In this section we describe how the synthetic $AR(p)$ data have been generated, and we specify the choice of parameters for our model.

We have generated the demand series $\{D_t,\ t>0\}$ following an $AR(p)$ process as in (1). Three different distributions for the error terms were tested in our experiments, all of them chosen so as to generate non-negative demand series. First, $a_t \sim N(4,1)$; also, we found of interest to check the behavior of the methods when heavy-tails are incorporated in the generator model, as done in [33], who choose Pareto and Lognormal distributions to check the performance of their inventory approach under samples of time-independent demand generated with heavy-tailed distributions. Therefore, we also set $a_t \sim LN(0,3)$ and $a_t \sim Par(1,1)$, where LN and Par denote the standard Lognormal and Pareto distributions respectively.

A different aspect to be considered when simulating the data is the strength of the temporal dependence. In our experiments, two values of θ were set. Note that for $p=1$ the coefficient θ_1 represent the lag-1 autocorrelation coefficient thus, in order to test the methods on highly and moderately correlated time series, $\boldsymbol{\theta}=\theta_1=0.9$ and $\boldsymbol{\theta}=\theta_1=0.5$ were fixed. Fig. 2 illustrates demand time series generated with the highest value of the correlation and both normal and lognormal error distributions. As mentioned before, other values of p have been tested with different values of θ but the conclusions obtained were analogous to those for $p=1$.

Finally, after some testing, the risk seeker strategy, where Δ is the $\left(\frac{u}{v}\right)^2$-th quantile of $(a_1,\dots,a_T)$, has been adopted under three types of profit products: $\frac{u}{v} \in \{0.75,\ 0.5,\ 0.25\}$, representing low-, neutral- and high-profit products respectively. In this way the proposed robust approach avoids being too conservative. The perturbation parameters β and ν of the $ARUS(p)$ were both set to 5, allowing therefore a 5% perturbation over the minimum variance and standard deviation respectively. The parameter Γ_1^* was set as the 0.95 quantile of the standard normal distribution following the reasoning of Bandi and Bertsimas [5].

A total of 1000 series of length $T+1=1000$ were generated for each $\boldsymbol{\theta}$ and each error's probability distribution. The first T=999 values have been used as train set in order to estimate the parameters of the inventory policies proposed in Section 3.1 and the $ARUS(p)$, and the remaining value $t=1000$ has been used as validation set. Therefore, the next process has been followed in order to calculate the revenue of the different approaches:

1. Determine the optimal quantity Q^* of products to buy for instant $T+1$ having available the demand historical records for $t=1,\dots,T$.
2. The demand in instant $T+1$ is realized, and then the revenue is calculated by using the expression (3).

3.3. Results

The results obtained are illustrated in Table 1 and Fig. 3. Table 1 shows the average revenue and the frequency of losses for the different approaches (namely, static, Scarf, classic $AR(1)$ prediction method, and robust $ARUS(1)$ method) under three different

Table 1
Average revenue and frequency of losses obtained for highly and moderately autocorrelated series under the four considered approaches (Static, Scarf, classic $AR(1)$ forecasting method, and $ARUS(1)$), under Pareto, Lognormal and normally distributed error terms.

θ	c/v	Method	$Par(1,1)$		$LN(0,3))$		$N(4,1)$	
			Avg. rev.	% loss	Avg. rev.	%loss	Avg. rev.	%loss
0.9	0.25	Static	25.70	7.50	157.64	27.50	29.16	0.00
		Scarf	11.76	12.90	46.26	15.30	29.16	0.00
		$AR(1)$	145.75	22.50	583.51	36.10	29.58	0.00
		$ARUS(1)$	160.22	0.00	674.83	0.00	29.45	0.00
	0.50	Static	13.01	12.60	65.09	23.50	19.02	0.00
		Scarf	1.52	3.20	0.09	0.10	19.02	0.00
		$AR(1)$	118.60	4.50	490.27	10.90	19.52	0.00
		$ARUS(1)$	105.22	0.00	437.73	0.00	19.49	0.00
	0.75	Static	4.94	11.60	20.19	16.30	9.25	0.00
		Scarf	0.00	0.00	0.00	0.00	9.21	0.00
		$AR(1)$	50.51	0.00	187.70	0.00	9.64	0.00
		$ARUS(1)$	52.25	0.00	217.49	0.00	9.64	0.00
0.5	0.25	Static	2.72	28.40	9.29	47.70	5.56	0.00
		Scarf	0.00	0.20	0.00	0.00	5.56	0.00
		$AR(1)$	−27.63	82.00	−145.12	87.80	5.61	0.00
		$ARUS(1)$	7.23	0.10	33.08	0.00	5.46	0.00
	0.50	Static	1.16	26.00	2.80	35.70	3.48	0.00
		Scarf	0.00	0.00	0.00	0.00	3.48	0.00
		$AR(1)$	3.02	63.60	4.63	76.90	3.55	0.00
		$ARUS(1)$	4.16	0.00	19.19	0.00	3.51	0.00
	0.75	Static	0.35	17.70	0.61	21.90	1.60	1.90
		Scarf	0.00	0.00	0.00	0.00	1.60	2.60
		$AR(1)$	2.29	0.00	9.31	0.00	1.66	0.90
		$ARUS(1)$	1.96	0.00	9.39	0.00	1.65	0.50

statistical distributions for the error terms ($Par(1,1)$, $LN(0,3)$ and $N(4,1)$). In Table 1 two levels of dependence ($\boldsymbol{\theta}=0.9$ versus $\boldsymbol{\theta}=0.5$) are considered.

From Table 1 several conclusions can be obtained. We first analyze highly autocorrelated series. First, we point out that under normally distributed errors, the four competing approaches perform similarly in terms of both the average revenue and frequency of losses. However, significant differences are found when the errors follow a distribution with heavier tails. Second, in both the Pareto and Lognormal cases, the robust approach outperforms the other three in terms of both the average revenue and percentage of losses, with the $LN(0,3)$ being under neutral-profit product ($u/v=0.5$) an exception, as the $AR(1)$ attains a better average revenue. Third, the methods that do not take into account the correlation of the data present the poorest performance, Scarf being the worst approach, usually yielding a higher percentage of losses that the $ARUS(1)$ and a lower average revenue. Fourth, the frequency of losses is always zero under the proposed robust approach, while it may be moderately high for both the static approach and the classic $AR(1)$ prediction method when high-profit products are considered.

Consider now moderately correlated series. Although the overall picture is analogous to that of the highly correlated case, some differences arise. First, it is interesting to note how for the low-profit products case with normally distributed errors all approaches attain a fraction of runs with losses bigger than zero, Scarf being the most extreme one. Second, it is worth highlighting the poor performance of the $AR(1)$ forecasting method under heavy-tailed distributions for the errors. Note that for high-profit products negative average revenues are obtained, the frequency of losses being extraordinarily high (this last phenomenon is also observed under neutral-profit products). On the contrary, Scarf's rule is too conservative, always attaining a zero average revenue for heavy-tailed demands. In this case, in which data are not highly correlated, the static approach does not behave as poorly as in the previous example.

In conclusion, it could be said that the robust autoregressive approach is more stable than the static and $AR(1)$ approaches, and overcomes extreme conservativeness of the Scarf method: it usually performs better or equivalently to all methods in terms of average revenue and always outperforms or is equivalent when minimizing the frequency of losses. Last, it is interesting to note that the higher the profit of a product is, the worse the classic $AR(1)$ forecasting method performs.

As an alternative illustration of the different prediction methods' performance in the least favorable scenarios, we provide Fig. 3, which depicts the predicted empirical cdf of the revenue in the interval of probabilities $[0,0.2]$. For the sake of abbreviation, only the neutral products case, i.e., for $\frac{u}{v}=0.5$, has been depicted; for the remaining plots the reader is referred to the electronic companion of this paper [15]. In the top of Fig. 3 the error terms follow a $N(4,1)$ distribution, while in the central and bottom panels the errors are assumed to be $LN(0,3)$ and $Par(1,1)$ distributed respectively. Time series with moderate and high autocorrelation ($\boldsymbol{\theta}=0.5$ and $\boldsymbol{\theta}=0.9$ respectively) are considered, and the results are given in the left and right columns.

Several conclusions can be obtained from Fig. 3. First, it can be deduced that for demand with normally distributed error the classic $AR(1)$ and robust $ARUS(1)$ approaches perform equivalently for highly correlated demand (right column), while for demand with lower correlation the $ARUS(1)$ approach outperforms the classic $AR(1)$. In both cases, the static and Scarf approaches present a poorer performance. Fig. 3 illustrates the same phenomena reported on Table 1 for demand with heavy-tailed errors: the classic $AR(1)$ forecasting method performs poorly, obtaining highly negative revenues. Although the Scarf approach outperforms both the static and classic autoregressive approaches, it is worse than the $ARUS(1)$.

4. Concluding remarks and extensions

In this paper we have considered a novel approach to the classic newsvendor problem. First, we incorporate temporal

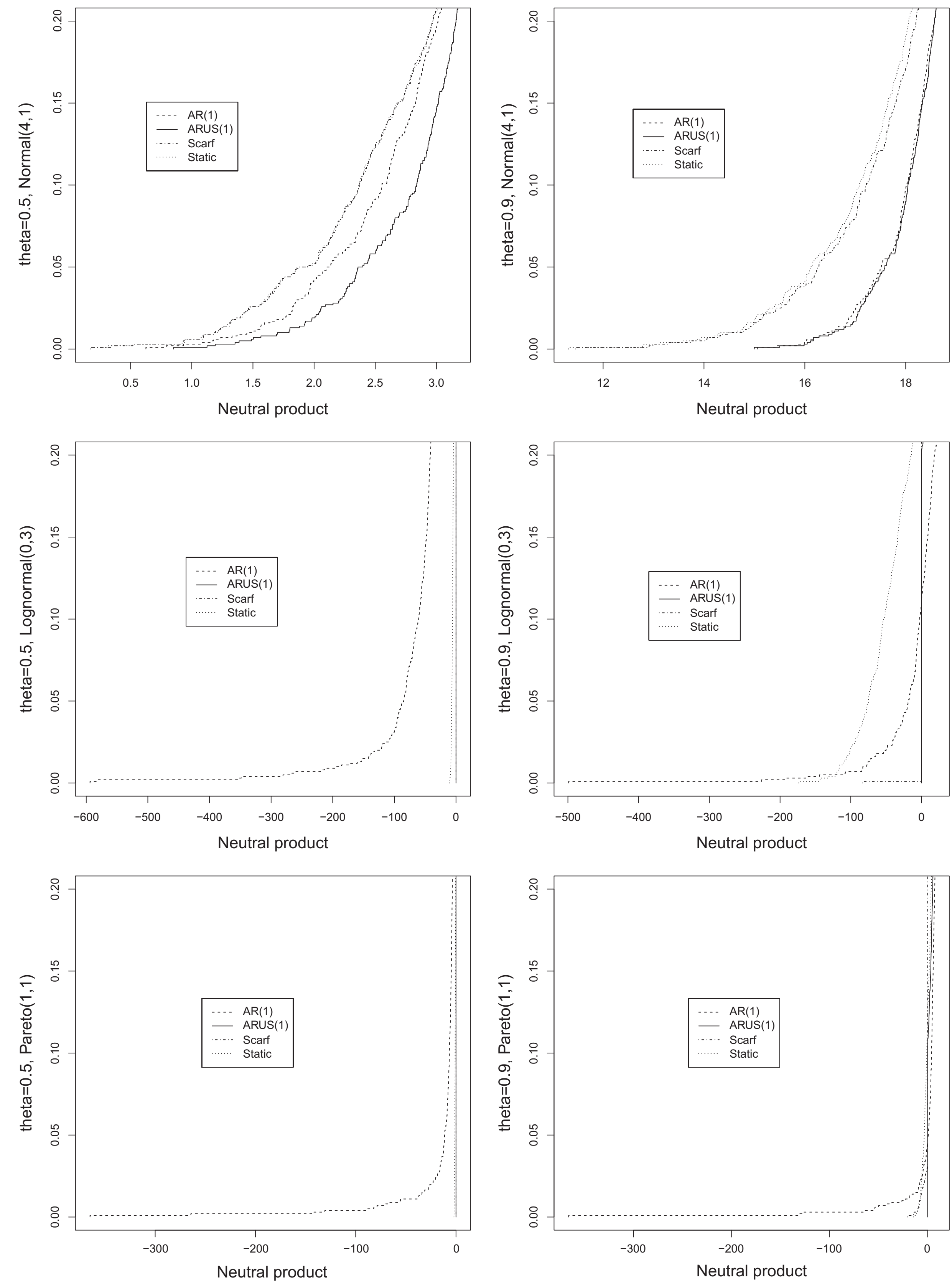

Fig. 3. Empirical cdf of the revenue under the static (dotted line), *AR*(1) forecasting method (dashed line) and the *ARUS*(1) approach (solid line), under normally (top), lognormally (central) and pareto (bottom) distributed error terms in the case of neutral products.

dependence by assuming that the demand follows an autoregressive process, and the forthcoming demand is to be forecasted from historical data. Second, the common assumption of normally distributed error terms in *AR* models is not made, as a distribution free counterpart is proposed. Moreover, a robust approach is used, and a closed form of the optimal solution is derived for the case $p=1$ and $q=2$. The performance of the proposed approach is compared to three traditional competing methods. The results show that the robust method outperforms the other approaches in terms of average revenue and obtains better results in terms of robustness. In very few occasions runs with losses are obtained.

We have briefly explored a robust extension to the multi-period counterpart. We aim to integrate our robust autoregressive approach into the robust inventory model of Bertsimas and Thiele [12], who assume that uncertainty intervals for the demand are already given. We propose to provide such intervals, whose lengths depend on the profitability of the product (or equivalently on the risk aversion of the user), by applying the robust modeling for autoregressive demands *ARUS*(p).

We have shown how to extend the model to a robust multi-item newsvendor problem with demands autocorrelated over time by products. Specifically, we have considered demands following a *VAR*(p) process. Extensions to more general inventory models, such as the replaceable items case, deserve further attention and careful analysis. Future prospects concerning this work would also include to formulate robust versions of more sophisticated time series models. Moreover, in the newsvendor context, shortage penalties and salvage values per unit can be considered as possible extensions to include in the robust approaches proposed here.

Acknowledgments

The research of the authors is supported by Grants MTM2012-36163, Spain, P11-FQM-7603 and FQM-329, Andalucía, all financed in part with EU ERD Funds.

Appendix A. Proof of Proposition 1

First, we provide a lemma that will be necessary for the proof.

Lemma 1. *If*

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\left(c+\sum_{k=1}^{p}\theta_k X_{t-k}-X_t\right)^2 \le \Gamma_2, \quad \text{for all } c,\theta_1,\dots,\theta_p,$$

then $\Gamma_2 - H(\boldsymbol{\theta}) \ge 0$.

Proof. Note first that if $\Gamma_2 - H(\boldsymbol{\theta}) < 0$ then

$$\Gamma_2 - \frac{1}{T-p}\sum_{t=p+1}^{T}\varphi_t^2 < \frac{-1}{(T-p)^2}\left(\sum_{t=p+1}^{T}\varphi_t\right)^2 \tag{26}$$

and

$$\frac{1}{T-p}\sum_{t=p+1}^{T}(c+\varphi_t)^2 = c^2+\frac{2c}{T-p}\sum_{t=p+1}^{T}\varphi_t+\frac{1}{T-p}\sum_{t=p+1}^{T}\varphi_t^2$$

Assume that there exists $(c,\boldsymbol{\theta})$ such that $\Gamma_2 - H(\boldsymbol{\theta}) < 0$ and $\frac{1}{T-p}\sum(c+\varphi_t)^2 \le \Gamma_2$. Then,

$$c^2+\frac{2c}{T-p}\sum_{t=p+1}^{T}\varphi_t \le \Gamma_2 - \frac{1}{T-p}\sum_{t=p+1}^{T}\varphi_t^2$$

but from (26)

$$c^2+\frac{2c}{T-p}\sum_{t=p+1}^{T}\varphi_t < -\frac{1}{(T-p)^2}\left(\sum_{t=p+1}^{T}\varphi_t\right)^2,$$

and therefore

$$\left(c+\frac{1}{T-p}\sum_{t=p+1}^{T}\varphi_t\right)^2 = c^2+\frac{2c}{T-p}\sum_{t=p+1}^{T}\varphi_t+\frac{1}{(T-p)^2}\left(\sum_{t=p+1}^{T}\varphi_t\right)^2 < 0$$

which is a contradiction.□

Proof of Proposition 1. From (5),

$$c \le U_1(\boldsymbol{\theta}) = \Gamma_1 - \frac{1}{T-p}\sum_{t=p+1}^{T}\varphi_t(\boldsymbol{\theta}),$$

$$c \ge L_1(\boldsymbol{\theta}) = -\Gamma_1 - \frac{1}{T-p}\sum_{t=p+1}^{T}\varphi_t(\theta).$$

and (6) implies

$$c \le U_2(\boldsymbol{\theta}) = \frac{-1}{T-p}\sum_{t=p+1}^{T}\varphi_t(\boldsymbol{\theta})+\sqrt{\Gamma_2-H(\boldsymbol{\theta})},$$

$$c \ge L_2(\boldsymbol{\theta}) = \frac{-1}{T-p}\sum_{t=p+1}^{T}\varphi_t(\boldsymbol{\theta})-\sqrt{\Gamma_2-H(\boldsymbol{\theta})},$$

Then given a fixed $\boldsymbol{\theta}$, (5)–(7) is written as:

$$-\Delta+\min\left(c+\sum_{k=1}^{p}\theta_k X_{T+1-k}\right)$$
$$\text{s.t.}\quad \begin{cases} c \ge \max\{L_1(\boldsymbol{\theta}),L_2(\boldsymbol{\theta})\} \\ c \le \min\{U_1(\boldsymbol{\theta}),U_2(\boldsymbol{\theta})\} \\ \sqrt{\Gamma_2-H(\boldsymbol{\theta})} \ge 0 \end{cases},$$

or equivalently as

$$-\Delta+\min\left(\max\{L_1(\boldsymbol{\theta}),L_2(\boldsymbol{\theta})\}+\sum_{k=1}^{p}\theta_k X_{T+1-k}\right)$$
$$\text{s.t.}\quad \left\{\sqrt{\Gamma_2-H(\theta)} \ge 0\right.$$

which is equivalent to (12) and (13). □

Appendix B. Proof of Proposition 2

We provide two lemmas needed for the proof of Proposition 2.

Lemma 2. *The function* $H(\boldsymbol{\theta})$ *is convex.*

Proof. Since $H(\boldsymbol{\theta})$ is an estimator of the variance of $\varphi(\boldsymbol{\theta})$, it can be rewritten as

$$H(\boldsymbol{\theta}) = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(\varphi_t(\boldsymbol{\theta})-\frac{1}{(T-p)}\sum_{t=p+1}^{T}\varphi_t(\boldsymbol{\theta})\right)^2$$

which is a convex function on $\boldsymbol{\theta}$ since each term is the square of an affine function.

Lemma 3. *The feasible region of the problem defined by (12) and (13) is convex.*

Proof. The feasible region of the problem defined by (12) and (13) is the set $F=\{\boldsymbol{\theta} \text{ s.t. } \Gamma_2-H(\boldsymbol{\theta}) \ge 0\}$. After some algebra, we have

$$\Gamma_2-H(\boldsymbol{\theta}) = \Gamma_2-\frac{1}{T-p}\sum_{t=p+1}^{T}\left(\sum_{k=1}^{p}\theta_k X_{t-k}\right)^2$$
$$+\frac{2}{T-p}\sum_{t=p+1}^{T}\sum_{k=1}^{p}\theta_k X_{t-k}X_t-\frac{1}{T-p}\sum_{t='+1}^{T}X_t^2$$

$$+\frac{1}{(T-p)^2}\left(\sum_{t=p+1}^{T}\sum_{k=1}^{p}\theta_k X_{t-k}\right)^2$$

From the definitions in (19) it can be seen that

$$\Gamma_2-H(\boldsymbol{\theta})=\Gamma_2-\sum_{k=1}^{p}\theta_k^2 V_k-2\sum_{k=1}^{p}\theta_k\sum_{r=k+1}^{p}\theta_r C_{k,r}+2\sum_{k=1}^{p}\theta_k C_{k,0}-V_0 \quad (27)$$

Since by Lemma 2, H is convex, the region defined by $\Gamma_2-H(\boldsymbol{\theta})\geq 0$ is convex.

Proof of Proposition 2. Since the feasible region is convex, it suffices to prove that the objective function is convex too. From Lemma 2, $H(\boldsymbol{\theta})$ is convex and therefore

$$-\min\left\{\Gamma_1,\sqrt{\Gamma_2-H(\boldsymbol{\theta})}\right\}$$

is also convex. On the other hand, $\sum_{t=p+1}^{T}\varphi_t(\boldsymbol{\theta})$ is linear on $\boldsymbol{\theta}$, thus

$$-\Delta+\min_{\boldsymbol{\theta}}\left(\frac{-\sum_{t=p+1}^{T}\varphi_t(\boldsymbol{\theta})}{T-p}-\min\left\{\Gamma_1,\sqrt{\Gamma_2-H(\boldsymbol{\theta})}\right\}+\sum_{k=1}^{p}\theta_k X_{T+1-k}\right)$$

is convex.□

Appendix C. Proof of Theorem 1

Two main cases may be distinguished:

1. Either $\min\left\{\Gamma_1,\sqrt{\Gamma_2-H(\theta_1)}\right\}=\Gamma_1$,
2. or $\min\left\{\Gamma_1,\sqrt{\Gamma_2-H(\theta_1)}\right\}=\sqrt{\Gamma_2-H(\theta_1)}$

Consider the first case. Then, the problem to be solved is written as

$$-\Delta-\Gamma_1+\min\left(\frac{-\sum_{t=p+1}^{T}\varphi_t(\theta_1)}{T-p}+\sum_{k=1}^{p}\theta_k X_{T+1-k}\right)$$
$$\text{s.t.}\quad\begin{cases}\Gamma_2-H(\theta_1)\geq 0\\ \sqrt{\Gamma_2-H(\theta_1)}\geq\Gamma_1\end{cases}$$

Note that the first constraint is redundant, as Γ_1 is supposed to be positive. Since the objective function is linear on θ_1, the optimum is reached at the frontier of the feasible region

$$F=\left\{\theta_1 \text{ s.t. } \Gamma_1=\sqrt{\Gamma_2-H(\theta_1)}\right\}.$$

Consider (27) with $p=1$. Then

$$\Gamma_2-\Gamma_1^2-H(\theta_1)=-\theta_1^2\, var(X^1)+2\theta_1\, cov(X^1,X^0)+(\Gamma_2-var(X^0)-\Gamma_1^2)$$

which is equal to zero if $\theta_1=\theta_1^{\star(1)}$ as in (14).

Assume now that $\min\left\{\Gamma_1,\sqrt{\Gamma_2-H(\theta_1)}\right\}=\sqrt{\Gamma_2-H(\theta_1)}$. Then the objective function to be minimized can be written as

$$A_2(\theta_1)=-\Delta+\frac{-1}{T-1}\sum_{t=2}^{T}(\theta_1 X_{t-1}-X_t)-\sqrt{\Gamma_2-H(\theta_1)}+\theta_1 X_T,$$

and let $F_2(\theta_1)=A_2(\theta_1)+\Delta$. This function is convex and therefore, the optimal solution for the unconstrained problem is reached when the derivatives are null, where

$$\frac{dF_2(\theta_1)}{d\theta_1}=a+\frac{\frac{1}{T-1}\sum_{t=2}^{T}(\theta_1 X_{t-1}-X_t)b_t}{\sqrt{\Gamma_2-H(\theta_1)}},$$

and where $a=\frac{-1}{T-1}\sum_{t=2}^{T}X_{t-1}$ and $b_t=X_{t-1}-\frac{1}{T-1}\sum_{t_0=2}^{T}X_{t_0-1}$. Note that $\frac{dF_2(\theta_1)}{d\theta_1}$ exists if and only if $\Gamma_2-H(\theta_1)>0$. Thus, the optimal θ_1 is either the one which the partial derivative of F_2 in the feasible region of the considered problem is zero or it is found at the frontier of such feasible region. Three cases may therefore be considered

(2.1) The minimum is reached at $\Gamma_2-H(\theta_1)=0$.
(2.2) The minimum is reached at $\sqrt{\Gamma_2-H(\theta_1)}=\Gamma_1$.
(2.3) The optimal θ_1 is the one such that $\frac{\partial F_2(\theta_1)}{\partial\theta_1}=0$.

Case (2.2) has been already solved. Consider the case (2.1). Then, after substituting $p=1$ in (27), expression (15) is obtained. Finally, the problem to be solved for the case (2.3) turns into

$$-\Delta+\min F_2(\theta_1)$$
$$\text{s.t.}\quad\begin{cases}\Gamma_2-H(\theta_1)>0\\ \sqrt{\Gamma_2-H(\theta_1)}\leq\Gamma_1\end{cases}\quad (28)$$

An optimal solution to (28) is obtained in the case (2.3) if there exists θ_1 such that $\frac{dF_2(\theta_1)}{d\theta_1}=0$, which in addition satisfies the constraints of problem (28). Such value θ_1 is obtained by

$$\left(-a\sqrt{\Gamma_2-H(\theta_1)}\right)^2=\left(\frac{1}{T-1}\sum_{t=2}^{T}\varphi_t(\theta_1)b_t\right)^2.$$

After some algebra, the right term is expressed as:

$$\frac{1}{(T-1)^2}\left(\sum_{t=2}^{T}\varphi_t(\theta_1)b_t\right)^2=\left(cov\left(\theta_1X^1,X^1\right)-cov\left(X^0,X^1\right)\right)^2,$$

from which the next quadratic function is obtained:

$$\theta_1^2(V_1^2+a^2V_1)-2\theta_1C_{1,0}(V_1+a^2)-a^2(\Gamma_2-V_0-C_{1,0}),\quad (29)$$

which is equal to zero if and only if $\theta_1=\theta_1^{\star(3)}$ as in the expression (16).□

References

[1] Abdel-Malek LL, Montanari R. On the multi-product newsboy problem with two constraints. Comput Oper Res 2005;32(8):2095–116.
[2] Aviv Y. Gaining benefits from joint forecasting and replenishment processes: the case of auto-correlated demand. Manuf Serv Oper Manag 2002;4(1):55.
[3] Aviv Y. A time-series framework for supply-chain inventory management. Oper Res 2003;51:210–27.
[4] Baker K, Scudder G. Sequencing with earliness and tardiness penalties. Oper Res 1990;38:22–36.
[5] Bandi C, Bertsimas D. Tractable stochastic analysis in high dimensions via robust optimization. Math Program 2012;134:23–70.
[6] Ben-Tal A, El Ghaoui L, Nemirovski A. Robust optimization. Princeton, New Jersey: Princeton University Press; 2009.
[7] Ben-Tal A, Nemirovski A. Robust convex optimization. Math Oper Res 1998;23 (4):769–805.
[8] Ben-Tal A, Nemirovski A. Robust solutions of uncertain linear programs. Oper Res Lett 1999;25(1):1–13.
[9] Ben-Tal A, Nemirovski A. Robust solutions of linear programming problems contaminated with uncertain data. Math Program 2000;88(3):411–24.
[10] Bertsimas D, Brown DB, Caramanis C. Theory and applications of robust optimization. SIAM Rev 2011;53(3):464–501.
[11] Bertsimas D, Copenhaver MS. Characterization of the equivalence of robustification and regularization in linear, median, and matrix regression; 2014. arXiv preprint arXiv:1411.6160.
[12] Bertsimas D, Thiele A. A robust optimization approach to inventory theory. Oper Res 2006;54:150–68.
[13] Bimpikis K, Markakis MG. Inventory pooling under heavy-tailed demand. ⟨http://www.econ.upf.edu/~mmarkakis/inventory_pooling.pdf⟩; 2015.
[14] Box E, Jenkins G, Reinsel G. Time series analysis: forecasting and control. Wiley series in probability and statistics; 2008.

[15] Carrizosa E, Olivares-Nadal A, Ramırez-Cobo P. Robust newsvendor problem with autoregressive demand; 2014. Preprint.
[16] Carrizosa E, Olivares-Nadal AV, Ramírez-Cobo P. Time series interpolation via global optimization of moments fitting. Eur J Oper Res 2013;230(1):97–112.
[17] Chen YF, Xu M, Zhang ZG. Technical note-a risk-averse newsvendor model under the cvar criterion. Oper Res 2009;57(4):1040–4.
[18] Cheng H, Tan P-N, Gao J, Scripps J. Multistep-ahead time series prediction. In: Advances in knowledge discovery and data mining. Heidelberg: Springer; 2006. p. 765–74.
[19] Choi S, Ruszczyński A, Zhao Y. A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. Oper Res 2011;59(2):346–64.
[20] Dana JD, Petruzzi NC. Note: the newsvendor model with endogenous demand. Manag Sci 2001;47(11):1488–97.
[21] Ding X, Puterman ML, Bisi A. The censored newsvendor and the optimal acquisition of information. Oper Res 2002;50(3):517–27.
[22] Dong L, Lee HL. Optimal policies and approximations for a serial multiechelon inventory system with time-correlated demand. Oper Res 2003;51(6):969–80.
[23] Eeckhoudt L, Gollier C, Schlesinger H. Increases in risk and deductible insurance. J Econ Theory 1991;55:435–40.
[24] Eeckhoudt L, Gollier C, Schlesinger H. The risk-averse (and prudent) newsboy. Manag Sci 1995;41(5):786–94.
[25] Ellsberg D. Risk, ambiguity, and the savage axioms. Q J Econ 1961:643–69.
[26] Gallego G, Moon I. The distribution free newsboy problem: review and extensions. J Oper Res Soc 1993;44:825–34.
[27] Gilboa I, Schmeidler D. Maxmin expected utility with non-unique prior. J Math Econ 1989;18(2):141–53.
[28] Godfrey GA, Powell WB. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. Manag Sci 2001;47(8):1101–12.
[29] Graves SC. A single-item inventory model for a nonstationary demand process. Manuf Serv Oper Manag 1999;1(1):50–61.
[30] Hadley G, Whitin TM. Analysis of inventory systems. Englewood Cliffs, New Jersey: Prentice Hall; 1963.
[31] Han Q, Du D, Zuluaga LF. Technical note-a risk-and ambiguity-averse extension of the max-min newsvendor order formula. Oper Res 2014;62(3):535–42.
[32] Hanasusanto GA, Kuhn D, Wallace SW, Zymler S. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. Math Program 2014:1–32.
[33] Huh WT, Levi R, Rusmevichientong P, Orlin JB. Adaptive data-driven inventory control with censored demand based on Kaplan–Meier estimator. Oper Res 2011;59(4):929–41.
[34] Jiang H, Netessine S, Savin S. Technical note-robust newsvendor competition under asymmetric information. Oper Res 2011;59(1):254–61.
[35] Johnson G, Thompson H. Optimality of myopic inventory policies for certain dependent demand processes. Manag Sci 1975;21:1303–7.
[36] Kabak I, Schiff A. Inventory models and management objectives. Sloan Manag Rev 1978;10:53–9.
[37] Kahn JA. Inventories and the volatility of production. Am Econ Rev 1987:667–79.
[38] Khouja M. The single-period (news-vendor) problem: literature review and suggestions for future research. Omega 1999;27:537–53.
[39] Lau H. The newsboy problem under alternative optimization objectives. J Oper Res Soc 1980;31:525–35.
[40] Lau H-S, Lau AH-L. The multi-product multi-constraint newsboy problem: applications, formulation and solution. J Oper Manag 1995;13(2):153–62.
[41] Lee HL, So KC, Tang CS. The value of information sharing in a two-level supply chain. Manag Sci 2000;46(5):626–43.
[42] Levi R, Roundy RO, Shmoys DB, et al. Approximation algorithms for capacitated stochastic inventory control models. Oper Res 2008;56(5):1184–99.
[43] Lin J, Ng TS. Robust multi-market newsvendor models with interval demand data. Eur J Oper Res 2011;212(2):361–73.
[44] Lu X, Song J, Regan A. Inventory planning with forecast updates: approximate solutions and cost error bounds. Oper Res 2006;54:1079–97.
[45] Perakis G, Roels G. Regret in the newsvendor model with partial information. Oper Res 2008;56:188–203.
[46] Petruzzi NC, Dada M. Pricing and the newsvendor problem: a review with extensions. Oper Res 1999;47(2):183–94.
[47] Qin Y, Wang R, Vakharia AJ, Chen Y, Seref MM. The newsvendor problem: review and directions for future research. Eur J Oper Res 2011;213(2):361–74.
[48] Reyman G. State reduction in a dependent demand inventory model given by a time series. Eur J Oper Res 1989;41:174–80.
[49] Rosenfield D. Optimal management of tax-sheltered employee reimbursement programs. Interfaces 1986;16:68–72.
[50] Scarf H. A min–max solution of an inventory problem. In: Studies in the mathematical theory of inventory and production; 1958. p. 201–9.
[51] Schweitzer M, Cachon G. Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. Manag Sci 2000;46(3):404–20.
[52] See C-T, Sim M. Robust approximation to multiperiod inventory management. Oper Res 2010;58(3):583–94.
[53] Shang KH, Song J-S. Newsvendor bounds and heuristic for optimal policies in serial supply chains. Manag Sci 2003;49(5):618–38.
[54] Thakkar R, Finley D, Liao W. A stochastic demand cvp model with return on investment criterion. Contemp Account Res 1983;1:77–86.
[55] Vairaktarakis GL. Robust multi-item newsboy models with a budget constraint. Int J Prod Econ 2000;66(3):213–26.
[56] Wang T, Atasu A, Kurtulus M. A multiordering newsvendor model with dynamic forecast evolution. Manuf Serv Oper Manag 2012;14:472–84.
[57] Weatherford L, Pfeifer P. The economic value of using advance booking of orders. Omega 1994;22:105–11.
[58] Yue J, Chen B, Wang M. Expected value of distribution information for the newsvendor problem. Oper Res 2006;54(6):1128–36.
[59] Zhang B. Multi-tier binary solution method for multi-product newsvendor problem with multiple constraints. Eur J Oper Res 2012;218(2):426–34.
[60] Zhu Z, Zhang J, Ye Y. Newsvendor optimization with limited distribution information. Optim Methods Softw 2013;28:640–67.