# Robust newsvendor problem with autoregressive demand

May 19, 2014

Emilio Carrizosa[a], Alba V. Olivares-Nadal [1][,a] and Pepa Ramírez-Cobo[b]

[a]*Instituto de Matemáticas de la Universidad de Sevilla (Spain),*

[b]*Department of Statistics and Operational Research, Universidad de Cádiz (Spain),*

**Abstract**

This paper explores the classic 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 forecasting method, which copes with non-stationary time series, is formulated. A closed-form expression for the optimal solution is found for the problem 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 two versions of the classic approach, in which either the demand distribution is unknown, and assumed to have no autocorrelation, 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.

*Keywords*: Distribution-free newsboy problem; autoregressive process; uncertainty set; minimax; robust optimization; forecasting.

---

[1]Corresponding author: Alba V. Olivares-Nadal, Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla, Av. Reina Mercedes, s/n, 41012 Sevilla, Spain. *Phone number:* +34 637 872 829 *Email address:* `aolivares@us.es` (Alba V. Olivares-Nadal)

# 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 (Weatherford and Pfeifer (1994)), health insurances (Rosenfield (1986); Eeckhoudt et al. (1991)), scheduling (Baker and Scudder (1990)), retailers and managers order quantity decision in sports and fashion industries (Gallego and Moon (1993)), etc.

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 $c$ and sells it at price $v$, then the expected revenue is maximized by buying exactly $Q^* = F^{-1}\left(1 - \dfrac{c}{v}\right)$ 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 (1999); Petruzzi and Dada (1999); Qin et al. (2011).

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...), for example, for small sample sizes. To overcome these and other related problems, some distribution-free approaches have been considered in the literature, Scarf (1958) 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 (1993), which provided an extension to Scarf's solution, and Yue et al. (2006), 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. (2002); Dana and Petruzzi (2001); Godfrey and Powell (2001). However, as pointed out in See and Sim (2010); Bandi and Bertsimas (2012), 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 them to an assumed distribution function may generate drastic errors in the inventory policy. Moreover, demand is in fact usually correlated along time, so assuming demands for each period are independent and identically distributed is in practice unrealistic (Lee et al., 2000; Graves,

1999; Kahn, 1987).

For the reasons mentioned above, some authors have studied inventory models with time-correlated demand, including $AR$ models (Aviv, 2002; Reyman, 1989; Johnson and Thompson, 1975), compound Poisson processes (Shang and Song, 2003), martingale models of forecast evolution (Dong and Lee, 2003; Lu et al., 2006; Wang et al., 2012), factor models (See and Sim, 2010) or estimation via Kalman filter (Aviv, 2003). Most of these papers either assume perfect knowledge of the distribution function (Levi et al., 2008; Aviv, 2003, 2002; Shang and Song, 2003; Wang et al., 2012; Reyman, 1989) or are focused in calculating and optimizing bounds of the objective function. In some cases, as See and Sim (2010); Lu et al. (2006); Dong and Lee (2003), those bounds are distribution-free, but no optimal solutions are obtained. 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. However, the problem to be solved in the remaining cases is extremely tractable due to its structural properties: it is a low-dimensional convex problem for which accurate solutions are easy to obtain. Moreover, we do not only cope with temporal demand but also take into account robustness in terms of uncertainty and risk aversion, which provides great novelty to this paper.

A different matter for discussion in the SPP literature is the choice of the objective function. In most paper the expected revenue is maximized. However, depending on the decision maker's preferences, it may be reasonable (if not necessary) to use other criteria, such as optimizing the probability of achieving a target profit, see for example Kabak and Schiff (1978) or Lau (1980), the Return of Investment (Thakkar et al. (1983)), the Cost-Volume-Profit, the CVaR (Chen et al. (2009)) or other risk-averse policies (Eeckhoudt et al. (1995); Choi et al. (2011)).

Robustness issues have also been addressed. 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, as well as Bertsimas and Thiele (2006) adopt such a criterion, although both approaches enforce independence of the demand along time. On the contrary, our approach would address the worst-case analysis while coping with time-correlated demands. The above-mentioned paper of Yue et al. (2006), as well as Perakis and Roels (2008); Zhu et al. (2013); Jiang et al. (2011) consider the minimax regret decision criterion instead.

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. Specifically, our main contri-

3

butions are:

1. We consider the demand as a time series with non-negligible autocorrelation coefficients. For simplicity, the basic yet efficient 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 uncertainty sets (Bandi and Bertsimas, 2012) to estimate the demand forecast. As the goal is to minimize the losses in the worst-case realization of the parameters, the lower bound obtained for the forecasted demand will characterize the optimal solution. A closed-form expression for the optimal solution is obtained in the case $p = 1$, while for $p \geq 2$ the problem turns into a tractable nonlinear convex optimization program, solved numerically.

3. We show that our approach outperforms two 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.

In this paper we perform a worst-case analysis, seeking the policy maximizing the worst-case revenue. This means buying as many units as given by the worst case scenario, which takes place when the demand is as scarce as possible when no shortage penalties are imposed. Therefore, a robust forecasting method will be developed in order to obtain a prediction interval for the demand, whose lower bound is chosen as the solution for the robust newsvendor problem. Since demand is assumed to follow an autoregressive process, we devote Section 2 to discuss the $AR(p)$ forecasting method and its robust counterpart, formulated using the concept of uncertainty sets. The robust model is written as a nonlinear convex optimization problem, and a closed-form solution is given for $p = 1$. In Section 3 we accommodate the robust forecasting method to the proposed newsvendor problem, obtaining the optimal solution for the particular case of demand following an $AR(1)$ process. In Section 3.2 we present numerical examples, where the robust autoregressive model is tested against two different but classic prediction methods, which are adapted for the newsvendor problem in Section 3.1. Last section is devoted to concluding remarks and extensions.

# 2 Robust forecasting method for $AR(p)$ processes

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 $c, \theta_1, ..., \theta_p$ are coefficients and $\{a_t, \ t > 0\}$ is the sequence of i.i.d model's error terms with expected value $\mu_a$ and variance $\sigma_a^2$, for all $t > 0$. Two realizations of autoregressive processes with different values of $p$ are illustrated in Figure 1. For a more detailed description of the autoregressive process and interpretation of the coefficients in the model, see for example Box et al. (2008).
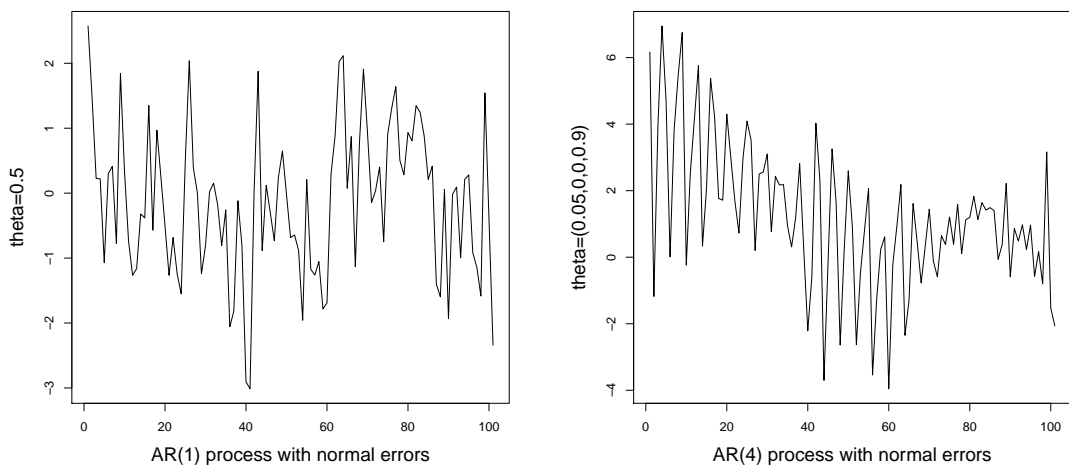


Figure 1: Examples of different $AR(p)$ processes generated with normal errors.

If the parameters in (1) are given (either they are known or estimated from sample data 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. This phenomenon will be illustrated in detail in Section 3. The influence of the error distribution over the demand can be seen in Figure 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. Hence, if real situations are wanted to be encompassed into autoregressive processes, the assumption of normality must be left aside.

The next section addresses the robust counterpart of (2), for which no probability distribution is imposed on the errors.

## 2.1 Statement of the $ARUS(p)$ optimization problem

In this section we describe how to obtain a one-step robust prediction interval $\left[\underline{X}_{T+1}, \overline{X}_{T+1}\right]$ for the $AR(p)$, under the assumption that a realization of the process up to time $T$ is available. To do this we use the concept of uncertainty sets (see Bandi and Bertsimas (2012)). For this reason, we call our model AutoRegressive process based on Uncertainty Sates, in short $ARUS(p)$.The value $\underline{X}_{T+1}$ is obtained by solving the $ARUS(p)$ problem, defined as

$$\min_{c,\theta_1,...,\theta_p,a_1,...,a_{T+1}} X_{T+1} \tag{3}$$

$$\text{s.t}$$

$$\left| \frac{1}{T-p} \sum_{t=p+1}^{T} a_t \right| \leq \Gamma_1, \tag{4}$$

$$\frac{1}{T-p} \sum_{t=p+1}^{T} a_t^2 \leq \Gamma_2, \tag{5}$$

$$X_t = c + \sum_{k=1}^{p} \theta_k X_{t-k} + a_t \quad t = p+1, ..., T+1, \tag{6}$$

$$|a_{T+1}| \leq \Delta \tag{7}$$

Similarly, the maximization of $X_{T+1}$ yields $\overline{X}_{T+1}$. The rationale behind the constraints in the $ARUS(p)$ problem is the following. Constraints (4) and (5) force both the mean and variance of the observed errors to be bounded,

so as to bound the randomness of the variables. In (6) the consecutive values of the process are expressed in terms of the previous $p$ values, according to the definition of the $AR(p)$ process as in (1). Note that, as no restrictions over the $\theta_k$ $k = 1, ..., p$ are made, nonstationary processes can be addressed by the $ARUS(p)$. Finally, (7) implies that the absolute value of the random value $a_{T+1}$ (which represents the prediction error) is bounded above by some constant $\Delta$. The values of $\Gamma_1, \Gamma_2$ and $\Delta$ 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 (2012). Since the errors terms are independent, then Central Limit Theorem (CLT) yields an uncertainty defined as

$$\mathcal{U}\left(\Gamma_1^*\right) = \left\{ (a_{p+1}, \ldots, a_T) \quad \text{s.t} \quad \left| \frac{1}{T-p} \sum_{t=p+1}^{T} a_t \right| \leq \frac{\Gamma_1^* \sigma_a \sqrt{T-p}}{T-p} + \mu_a \right\},$$

which in combination with (6) yields

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

Therefore, constraint (4) may be rewritten as

$$\frac{1}{T-p} \left| \sum_{t=p+1}^{T} X_t - c - \sum_{k=1}^{p} \theta_k X_{t-k} \right| \leq \frac{\Gamma_1^* \sigma_a \sqrt{T-p}}{T-p} + \mu_a. \tag{8}$$

The value $\Gamma_1^*$ in (8) is a small constant that influences the accuracy of the fit. Since $\mu_a$ and $\sigma_a$ are unknown in practice, they need to be estimated. For $\mu_a$, note that there is no loss of generality in assuming $\mu_a = 0$: the model defined by (1) is equivalent to

$$X_t = (c + \mu_a) + \sum_{k=1}^{p} \theta_k X_{t-k} + \varepsilon_t, .$$

$\varepsilon_t$ having zero mean. As regard the value of $\sigma_a$ in (8), we suggest to use

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

where $(1 + \nu/100)$ indicates a perturbation (depending on the value of $\nu$) of $\sigma_0$, which denotes the optimal value to the problem of minimizing the variance of the errors up to time $T$

$$\min_{a_1,\ldots,a_T} \frac{1}{T-p} \sum_{t=p+1}^{T} a_t^2 = \min_{c,\theta_1,\ldots,\theta_p} \frac{1}{T-p} \sum_{t=p+1}^{T} \left( X_t - c - \sum_{k=1}^{p} \theta_k X_{t-k} \right)^2. \quad (9)$$

Consider now the constraint (5). We can proceed similarly as before and substitute $\Gamma_2$ by a certain perturbation of the minimum value attained by the errors' variance:

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

where $\gamma_2$ is the optimal value of

$$\min_{c,\theta_1,\ldots,\theta_p,a_1,\ldots,a_{T+1}} \frac{1}{T-p} \sum_{t=1}^{T} a_t^2 \qquad (10)$$

s.t

$$\left| \frac{1}{T-p} \sum_{t=p+1}^{T} a_t \right| \leq \Gamma_1 \qquad (11)$$

$$X_t = c + \sum_{k=1}^{p} \theta_k X_{t-k} + a_t, \quad t = 2, \ldots, T. \qquad (12)$$

Finally, consider (7). The choice of $\Delta$ 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 $\Delta$ with the values $a_1, a_2, \ldots, a_T$. Since problem (10)-(12) returns the optimal $c^\star, \theta_1^\star, \ldots, \theta_p^\star$ 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 $\Delta$ is the empirical $q-$th quantile of the sample $(a_1, \ldots, a_T)$, solution to problem (10)-(12), for some value of $q$, large enough.

In conclusion, once the perturbation parameters $\nu$ and $\beta$ are set, the problems (9) and (10)-(12) are solved, and the values of $\sigma_0$, $\gamma_2$ and $\Delta$ are found, then the $ARUS(p)$ optimization problem is rewritten as

$$-\Delta + \min_{c,\theta_1,\ldots,\theta_p} \left( c + \sum_{k=1}^{p} \theta_k X_{T+1-k} \right) \tag{13}$$

s.t

$$\left| \frac{1}{T-p} \left( \sum_{t=p+1}^{T} X_t - c - \sum_{k=1}^{p} \theta_k X_{t-k} \right) \right| \leq \frac{\Gamma_1^*(1+\nu/100)\sigma_0\sqrt{T-p}}{T-p} \tag{14}$$

$$\frac{1}{T-p} \left( \sum_{t=p+1}^{T} X_t - c - \sum_{k=1}^{p} \theta_k X_{t-k} \right)^2 \leq (1+\beta/100)\gamma_2 \tag{15}$$

and similarly, the maximization problem is defined. Details on how to solve (16)-(15) are given in Section 2.2.

Now we introduce a mild extension of the model above. In some real-world situations, the nature of the time series requires to predict the 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 Carrizosa et al. (2013)). Therefore, in such cases it seems natural to consider a modified version of the $ARUS(p)$ problem, in which the constraint

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

is added. The new problem will be called $ARUS(p)_{[a,b]}$ from now on.

Since the objective function of the original $ARUS(p)$ minimization problem is convex, then it increases in all directions from the global minimum. This implies that once the solution $\underline{X}_{T+1}$ is found, then the solution to the $ARUS(p)_{[a,b]}$ is necessarily $\underline{X}_{T+1}$, if $\underline{X}_{T+1} \in [a,b]$, or on the contrary it equals $a$ (respectively $b$), if the objective function is increasing (respectively, decreasing) in $[a, b]$. Finally, a similar reasoning applies to the case of the $ARUS(p)$ maximization problem, yielding the next proposition:

**Proposition 1.** *The solution to the minimization $ARUS(p)_{[a,b]}$ is either the solution of the original minimization problem $ARUS(p)$, $\underline{X}_{T+1}$, either $a$ or $b$. An equivalent result is obtained for the $ARUS(p)_{[a,b]}$ maximization problem.*

## 2.2   Solution to the $ARUS(p)$ problem

The $ARUS(p)$ problem defined by (3)-(7) is written in terms of a linear objective function to be optimized in a convex region.

The following result provides a re-formulation of (3)-(7) in such a way that additional properties concerning its solutions are derived. In particular, a closed-form solution for the $ARUS(1)$ is obtained from the novel formulation.

**Proposition 2.** *The minimization $ARUS(p)$ problem defined by (3)-(7) is equivalent to*

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

$$s.t \tag{16}$$

$$\Gamma_2 - H(\boldsymbol{\theta}) \geq 0, \tag{17}$$

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

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

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

*Proof.* Proof See Appendix A. ☐

**Proposition 3.** *Equations (16)-(17) define a convex optimization problem.*

*Proof.* Proof See Appendix B. ☐

The $ARUS(p)$ problem as in (16)-(17) is a smooth convex optimization problem to be solved numerically. 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 (16)-(17) is reached at one of these values for $\theta_1^\star$:*

$$\theta_1^{\star(1)} = \frac{-C_{1,0} \pm \sqrt{C_{1,0}^2 + V_1\left(\Gamma 2 - V_0 - \Gamma_1^2\right)}}{-V_1}, \tag{18}$$

$$\theta_1^{\star(2)} = \frac{-C_{1,0} \pm \sqrt{C_{1,0}^2 + V_1\left(\Gamma 2 - V_0\right)}}{-V_1}, \tag{19}$$

$$\theta_1^{\star(3)} = \frac{-C_{1,0}(V_1 + a^2) \pm \sqrt{C_{1,0}^2(V_1 + a^2)^2 + a^2 V_1(V_1 + a^2)\left(\Gamma 2 - V_0 - C_{1,0}\right)}}{-V_1}, \tag{20}$$

*where*

$$V_k = var\left(X^k\right), \quad C_{k,h} = cov\left(X^k, X^h\right) \tag{21}$$

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

$$X^k = \left(X_{p+1-k} \ldots, X_{T-k}\right), \tag{22}$$

*and $a = \dfrac{-1}{T-1} \displaystyle\sum_{t=2}^{T} X_{t-1} + X_T$.*

*Proof.* Proof See Appendix C. □

# 3 Solution to the robust newsvendor model with autoregressive demand

Once developed the robust forecasting method for autoregressive processes, a prediction interval for the demand is obtained. The worst case scenario for the newsvendor with no shortage costs takes place when the demand reaches its lower value. Hence the optimal robust solution for the newsvendor problem with autoregressive demand is the lower extreme of the prediction interval obtained with the robust forecasting method. Therefore, as developed in Section 2.2, we obtain a closed expression for the optimal solution of our robust newsvendor problem when $p = 1$. For greater values of $p$ the optimal solutions are calculated numerically, although obtaining highly accurate results due to problem convexity properties.

The robust forecasting method accomplishes the task of determining a prediction interval for the demand, providing the optimal solution for the worst case approach. However, negative solutions may be obtained if the plain $ARUS(p)$ is applied, thus a non-negativity constraint must be added to properly fits the demand prediction. Hence, the $ARUS(p)_{[0,+\infty)}$ would properly address the newsvendor problem with autoregressive demand as it will provide the optimal solution in the worst case scenario while assuring feasibility for the demand. Summarizing, the optimal solution for the robust newsvendor problem with autoregressive demand is the value obtained by minimizing the $ARUS(p)_{[0,+\infty)}$ problem defined in Section 2.

In order to illustrate the performance of the presented robust $AR(p)$ forecasting method in the context of the newsvendor problem with autoregressive demand, the proposed approach will be compared with two classic approaches, which will be called *static* and $AR(p)$. The so-called *static* method is based on the classical newsvendor approach and the second one is derived

from basic autoregressive forecasting method, briefly introduced at the beginning of Section 2. Now in Section 3.1 we are going to accommodate those approaches to the newsvendor problem so as to be able to apply them later to some numerical experiments. In the first approach, the demand distribution will be unknown and estimated from the sample observations (assumed to be independent), as in the second approach normality over the error termns will be assumed.

## 3.1   Benchmark methods versus the robust approach

In this section we describe in detail both the benchmark approaches (*static* and $AR(p)$) as well as the way to apply the $ARUS(p)$ model to the newsvendor problem.

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

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

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 cdf $F$ for a large enough sample, is considered. Note that under this approach the temporal dependence in the data is ignored.

In the case of the classical $AR(p)$ approach, we saw in Section 2 that the forecast was assumed to follow a normal distribution. Therefore, the optimal solution would be

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

where $\Phi_{(\hat{X}_{T+1}, \sigma_a)}^{-1}$ is the inverse cdf of a normal distribution with mean $\hat{X}_{T+1}$ and standard deviation $\sigma_a$. Since this approach may lead to negative solutions and the demand always takes non-negative values, the quantity $Q^\star = \max\left\{0, Q_{AR(p)}^*\right\}$ will be considered instead.

Because of the same reason, as an alternative to the $ARUS(p)$ problem (Equations 3-7) we will focus on the $ARUS(p)_{[0,+\infty)}$ problem defined in Section 2, whose optimal solution is

$$Q_{T+1}^* = \max\left\{0, \hat{Q}_{T+1}\right\}$$

$\hat{Q}_{T+1}$ being the solution of the minimization $ARUS(p)$ problem. Consider the $ARUS(p)$ constraint (7), which bounds by $\Delta$ the absolute value of the $(T+1)$-th error term. In the context of the newsvendor problem it seems reasonable to relate the value of $\Delta$ to the ratio $\frac{c}{v}$, which is a measure of the decision maker's risk aversion. From Schweitzer and Cachon (2000), in the case of high profit products (that is, $\frac{c}{v} \leq 0.5$), risk aversion may be reduced by allowing the decision maker to buy more items. This implies that $\hat{Q}_{T+1}$ is allowed to be higher, or, in other words, the length of the prediction interval is permitted to be smaller, which may be obtained by reducing the value of $\Delta$. Thus, in the context of the newsvendor problem, it makes sense to choose $\Delta$ as the empirical $\frac{c}{v}$-th quantile of the sample $(a_1, \ldots, a_T)$, although other choices of $\Delta$, depending on the decision's maker criterion, may be sensible too. After some testing, we have chosen the $\left(\frac{c}{v}\right)^2$-th quantile of $(a_1, \ldots, a_T)$ under three types of profit products, namely, low- $\left(\frac{c}{v} = 0.75\right)$, neutral- $\left(\frac{c}{v} = 0.5\right)$, and high-profit $\left(\frac{c}{v} = 0.25\right)$ products.

## 3.2 Numerical illustrations

Now we are going to test the performance of our approach against those proposed 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.

### 3.2.1 Synthetic data generation and experiments design

In this work the performance of the robust $AR(p)$ forecasting method 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 of 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 Huh et al. (2011). They 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.

A different aspect to be considered when simulating the data is the strength of the temporal dependence. In our experiments, two values of $\boldsymbol{\theta}$ were set. Note that for $p = 1$ the coefficient $\theta_1$ represent the lag$-1$ auto-correlation coefficient thus, in order to test the methods on highly and low correlated time series, $\boldsymbol{\theta} = \theta_1 = 0.9$ and $\boldsymbol{\theta} = \theta_1 = 0.5$ were fixed. Figure 2 illustrate the demand time series generated with the different values of the correlation and errors distributions. As mentioned before, other values of $p$ have been tested with different values of $\boldsymbol{\theta}$ but the conclusions obtained were analogous to those for $p = 1$.
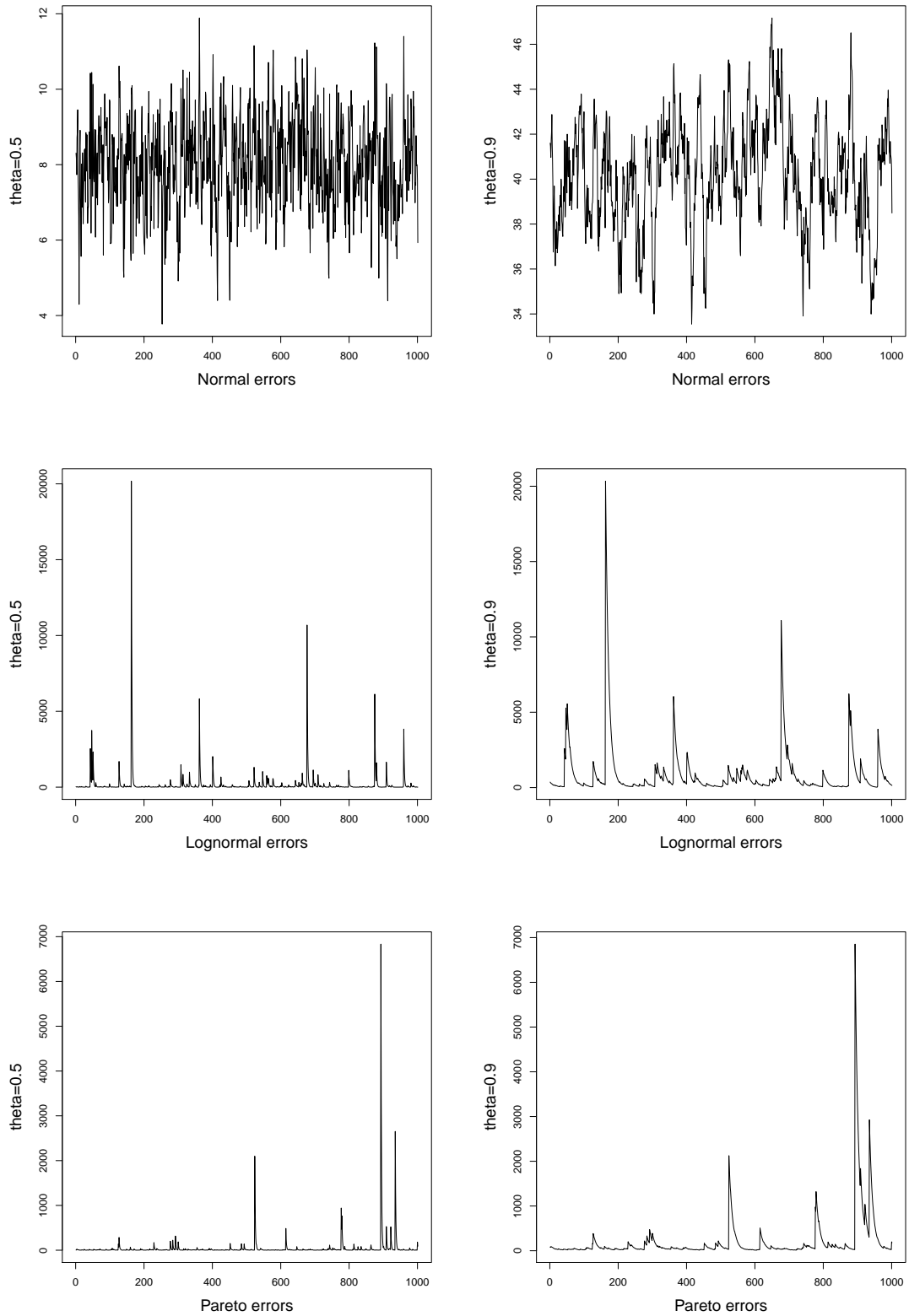
Figure 2: Examples of high and low correlated autoregressive demands generated with errors following $N(4, 1)$, $LN(0, 3)$ and $Par(1, 1)$.

Finally, as commented in the previous section, the risk seeker strategy where $\Delta$ is the $\left(\frac{c}{v}\right)^2$-th quantile of $(a_1, \ldots, a_T)$, has been adopted under three types of profit products: $\frac{c}{v} \in \{0.75, \ 0.5, \ 0.25\}$, representing low-, neutral- and high-profit products, respectively. The perturbation parameters $\beta$ and $\nu$ of the $ARUS(p)$ approach were both set to 5, allowing therefore a 5% perturbation over the minimum variance and standard deviation, respectively. The parameter $\Gamma_1^*$ was set as the 0.95 quantile of the standard normal distribution following the reasoning of Bandi and Bertsimas (2012).

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 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 of products to buy for instant $T + 1$ $Q^\star$ having available the demand historical records for $t = 1, ..., T$

2. The demand in instant $T + 1$ is realized, and the revenue is calculated by using the following expression:

$$R(Q) = \min \{Q, D_{T+1}\} - \left(\frac{c}{v}\right) Q.$$

### 3.2.2   Results

The results obtained are illustrated by Tables 1-2, and Figures 3-7. Tables 1-2 show both the average revenue and the frequency of losses for the different approaches (namely, static, classic $AR(1)$ prediction method, and robust $ARUS(1)$ method) under three different statistical distributions for the error terms ($Par(1, 1)$, $LN(0, 3)$ and $N(4, 1)$). In Table 1 a higher level of dependence than in Table 2 ($\boldsymbol{\theta} = 0.9$ versus $\boldsymbol{\theta} = 0.5$) is considered.

16

| $\boldsymbol{\theta} = 0.9$ | | $Par(1,1)$ | | $LN(0,3)$ | | $N(4,1)$ | |
|---|---|---|---|---|---|---|---|
| $c/v$ | Method | Avg. rev. | % loss | Avg. rev. | %loss | Avg. rev. | %loss |
| 0.25 | Static | 25.70 | 7.50 | 157.64 | 27.50 | 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 |
| | $AR(1)$ | 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 |
| | $AR(1)$ | 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 |

Table 1: Average revenue and frequency of losses obtained for high auto-correlated series ($\boldsymbol{\theta} = 0.9$) under the three considered approaches (Static, classic $AR(1)$ forecasting method, and $ARUS(1)$), under Pareto, Lognormal and normally distributed error terms.

From Table 1 several conclusions can be obtained. First, we point out that under normally distributed errors, the three 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. In both the Pareto and Lognormal cases, the robust approach outperforms the other two, being the $LN(0,3)$ under neutral-profit product ($c/v = 0.5$) an exception. In all cases, the static method, which does not take into account the correlation of the data, presents the poorest performance, especially when the average revenue is considered. The frequency of losses is always zero under the 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.

| $\boldsymbol{\theta} = 0.5$ | | $Par(1,1)$ | | $LN(0,3)$ | | $N(4,1)$ | |
|---|---|---|---|---|---|---|---|
| $c/v$ | Method | Avg. rev. | % loss | Avg. rev. | %loss | Avg. rev. | %loss |
| | Static | 2.72 | 28.40 | 9.29 | 47.70 | 5.56 | 0.00 |
| 0.25 | $AR(1)$ | -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 |
| | Static | 1.16 | 26.00 | 2.80 | 35.70 | 3.48 | 0.00 |
| 0.50 | $AR(1)$ | 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 |
| | Static | 0.35 | 17.70 | 0.61 | 21.90 | 1.60 | 19.00 |
| 0.75 | $AR(1)$ | 2.29 | 0.00 | 9.31 | 0.00 | 1.66 | 9.00 |
| | $ARUS(1)$ | 1.96 | 0.00 | 9.39 | 0.00 | 1.65 | 5.00 |

Table 2: Average revenue and frequency of losses obtained for low autocorrelated series ($\boldsymbol{\theta} = 0.5$) under the three considered approaches (Static, classic $AR(1)$ forecasting method, and robust method), under Pareto, Lognormal and normally distributed error terms.

Consider now Table 2, in which low correlated series are analyzed. Again in this case the three methods behave similarly under the $N(4,1)$ distribution. It is interesting to note however, how for the low-profit products case, all approaches attain a fraction of runs with losses bigger than zero, being the static approach the most extreme one. It is of interest to highlight the poor performance of the $AR(1)$ forecasting method in this case, 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). In this case, in which data are not highly correlated, the static approach does not behave as poorly in the previous example.

In conclusion, it could be said that the robust autoregressive approach is more stable than the other two approaches: it usually performs better or equivalently to the other methods in terms of average revenue and always outperforms them when minimizing the frequency of losses.

As an alternative illustration of the different prediction methods' performance in the least favorable scenarios, we provide Figures 3-5, which depict the predicted empirical cdf of the revenue in the interval of probabilities $[0, 0.2]$. In Figure 3 the error terms follow a $N(4,1)$ distribution, while in Figures 4 and 5, the errors are assumed to be $LN(0,3)$ and $Par(1,1)$ dis-

tributed, respectively. Time series with moderate and high autocorrelation ($\theta = 0.5$, $\theta = 0.9$ respectively) are considered, and the results are given in the left and right column. In addition, the top row of the figures presents the results obtained under a high-profit product, while the central and bottom panels are devoted the the neutral- and low-profit cases, respectively.
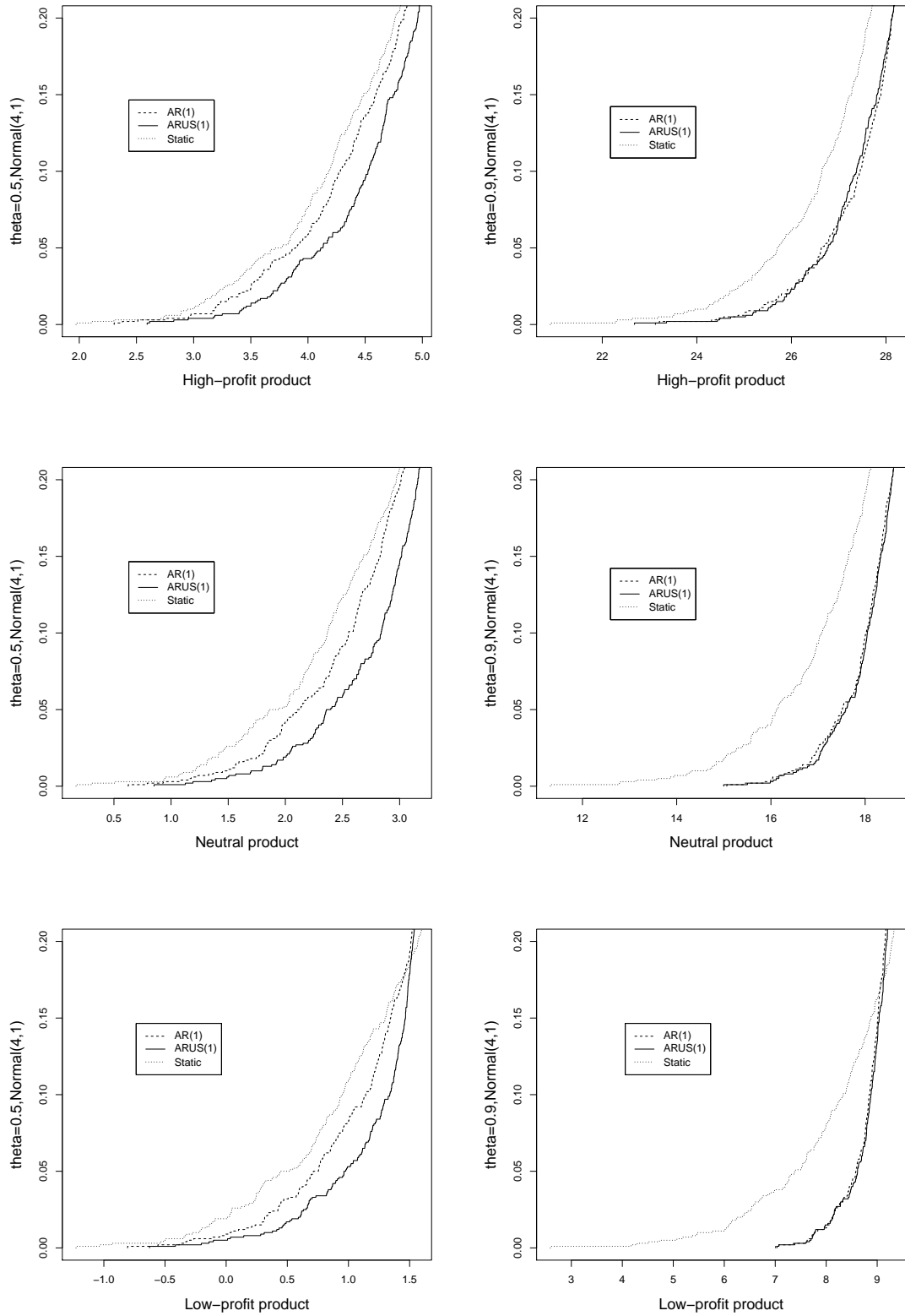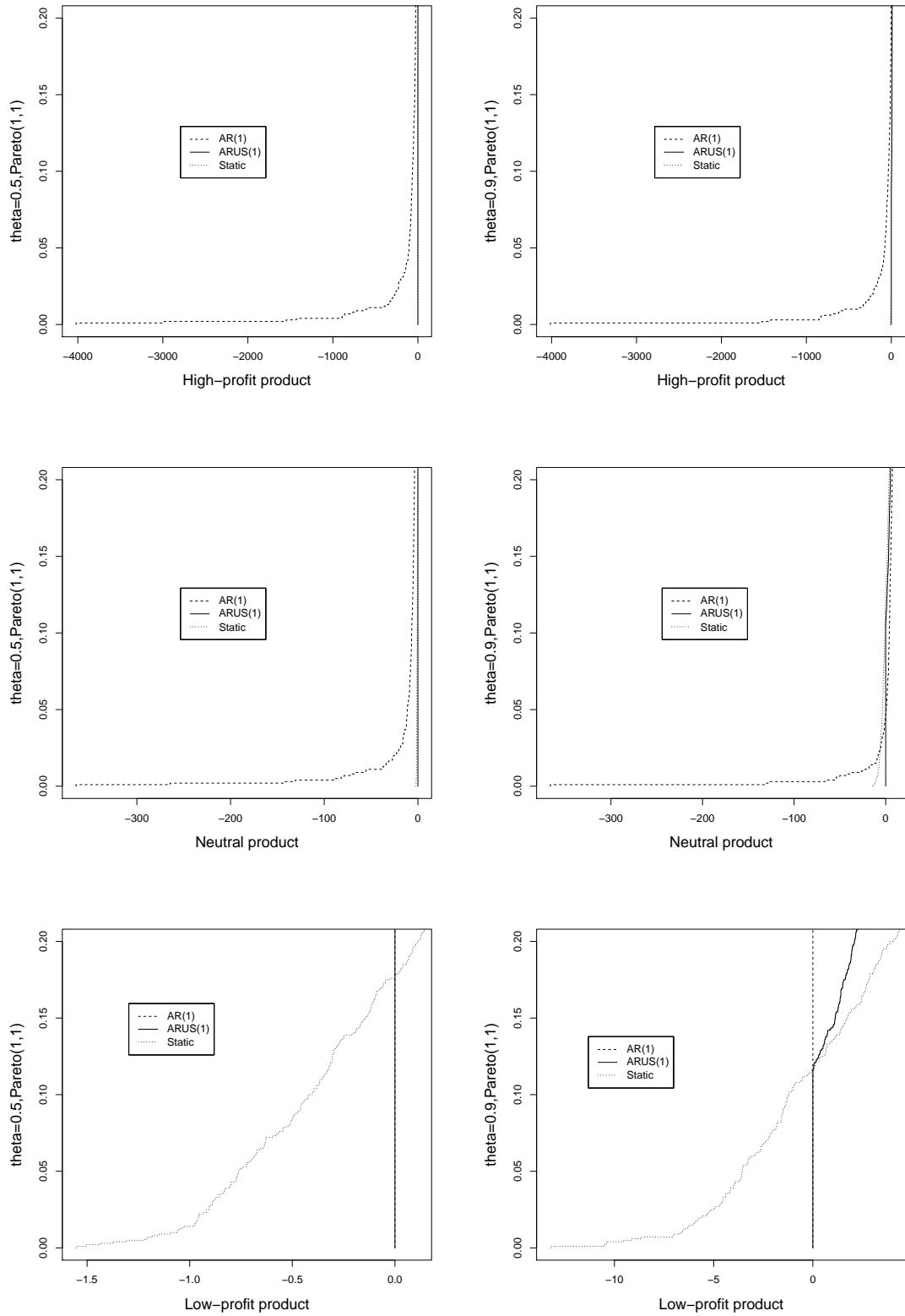
20

Figure 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 distributed error terms.
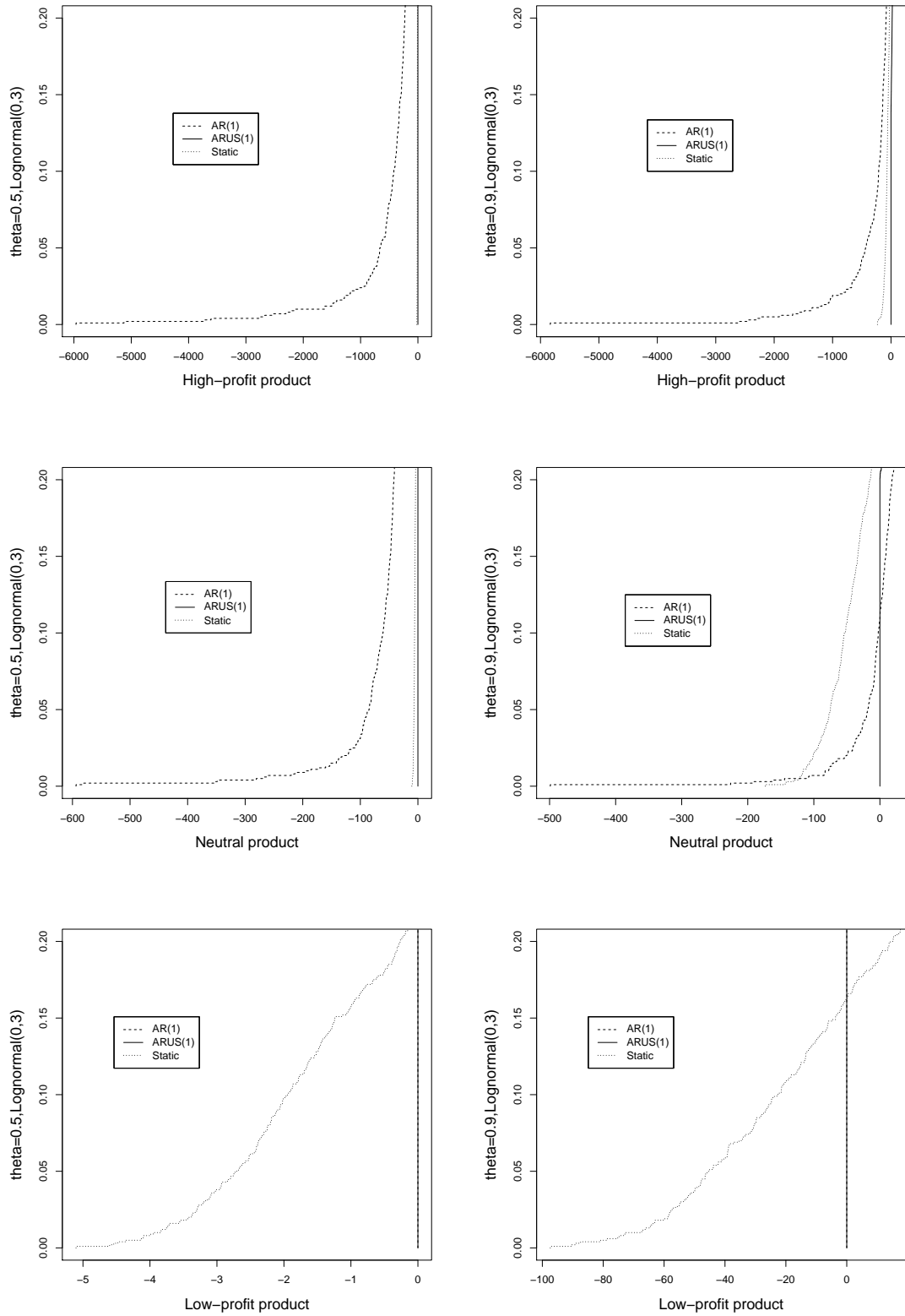
Figure 4: Empirical cdf of the revenue under the static (dotted line), $AR(1)$ forecasting method (dashed line) and the $ARUS(1)$ approach (solid line), under lognormally distributed error terms.

Figure 5: Empirical cdf of the revenue under the static (dotted line), $AR(1)$ forecasting method (dashed line) and the $ARUS(1)$ approach (solid line), under error terms distributed as a $Par(1,1)$ distribution.

Several conclusions can be obtained from Figures 3-5. 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 approach presents the poorest performance. Figures 4 and 5 illustrate the same phenomena reported on Tables 1 and 2 for demand with heavy-tailed errors: the higher the profit of a product is, the worse the classic $AR(1)$ forecasting method performs. Only in the case of low-profit products this approach performs similarly to its robust counterpart, but it can observed how the performance of the $AR(1)$ approach approximates to that of the static one when the profit increases. In order to clarify the comparison between the static and $ARUS(1)$ methods see Figures 6 and 7, which provide the same information as Figures 4 and 5 for the high- and neutral-profit products without considering the robust approach.
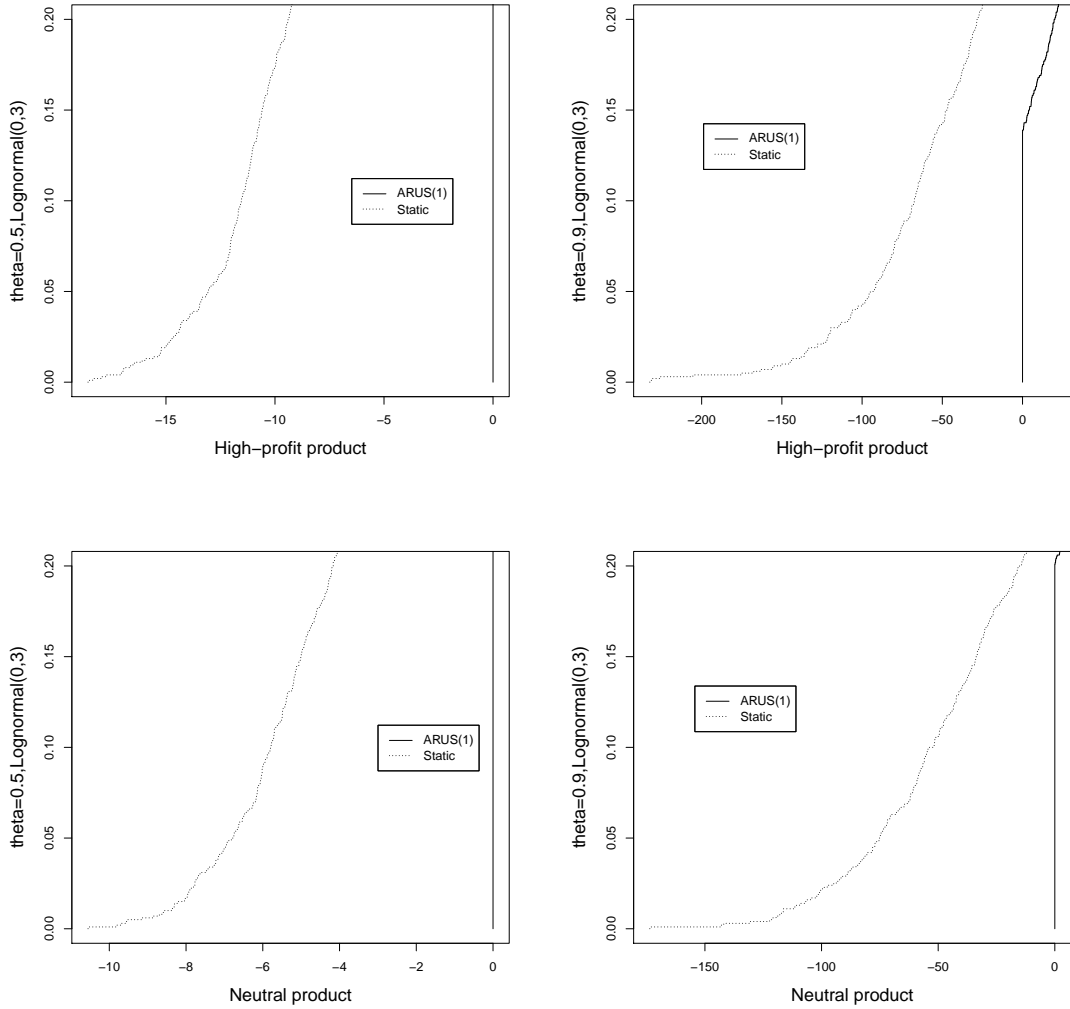
Figure 6: Empirical cdf of the revenue under the static (dotted line) and the $ARUS(1)$ approach (solid line), under lognormally distributed error terms. The values $\boldsymbol{\theta} = 0.5$ and $\boldsymbol{\theta} = 0.9$ are considered in the left and right columns. High- and neutral-profit product strategies are represented in the top and bottom panels, respectively.
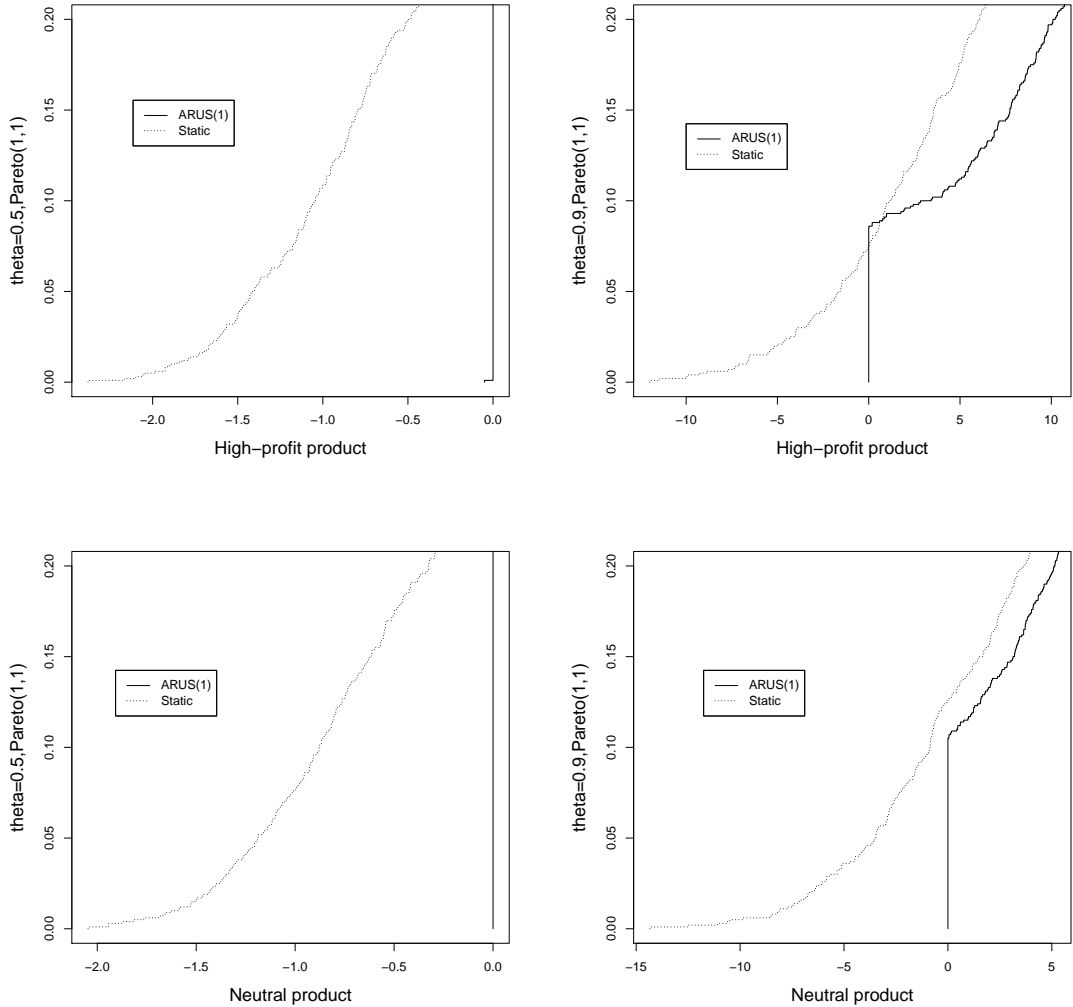
24

Figure 7: Empirical cdf of the revenue under the static (dotted line) and the $ARUS(1)$ approach (solid line), under error terms distributed as a $Par(1,1)$ distribution. The values $\boldsymbol{\theta} = 0.5$ and $\boldsymbol{\theta} = 0.9$ are considered in the left and right columns. High- and neutral-profit product strategies are represented in the top and bottom panels, respectively.

# 4 Concluding remarks and extensions

In this paper we have considered a novel approach to the classic newsvendor problem. First, we incorporate temporal 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 in this work 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 $p = 1$. The performance of the proposed approach is compared to two traditional competing methods. The results show that the robust method outperforms the other approaches in terms of average revenue and obtains better results in term of robustness. In very few occasions runs with losses are obtained.

Because of its tractability and simplicity, this paper considers the single-period newsvendor problem as a test to illustrate the importance of accurate predictions. We are aware of the interest of solving the multistage counterpart, which can be used in a wider range of applications. The complications that such a problem involves define a challenging task that we hope to address in the future. Extensions to more general inventory models, such as the multi-item problem, deserve further attention and careful analysis. In the newsvendor context, shortage penalties and salvage values per unit can be considered as possible extensions to include in the $ARUS(p)$ approach proposed here, as in the case of the multi-product SPP correlated demands amongst items may be taken into account. Future prospects concerning this work also include to formulate robust versions of more sophisticated time series models.

# Acknowledgement

# Appendix A. Proof of Proposition 2

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

**Lemma 1.** *If*

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

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

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

$$\Gamma_2 - \frac{1}{T-p} \sum_{t=p+1}^{T} \varphi_t^2 < \frac{-1}{(T-p)^2} \left( \sum_{t=p+1}^{T} \varphi_t \right)^2 \tag{23}$$

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

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

but from (23)

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

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. $\qquad\qquad\square$

**Proof of Proposition 2:** From (4),

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

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

and (5) implies

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

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

Then given a fixed $\boldsymbol{\theta}$, (4)-(6) is written as:

$$-\Delta + \min \quad \left( c + \sum_{k=1}^{p} \theta_k X_{T+1-k} \right)$$

$$\text{s.t} \quad \begin{cases} c \geq \max \{L_1(\boldsymbol{\theta}), L_2(\boldsymbol{\theta})\} \\ c \leq \min \{U_1(\boldsymbol{\theta}), U_2(\boldsymbol{\theta})\} \\ \sqrt{\Gamma_2 - H(\boldsymbol{\theta})} \geq 0 \end{cases} ,$$

or equivalently as

$$-\Delta + \min \quad \left( \max \{L_1(\boldsymbol{\theta}), L_2(\boldsymbol{\theta})\} + \sum_{k=1}^{p} \theta_k X_{T+1-k} \right)$$

$$\text{s.t} \quad \left\{ \sqrt{\Gamma_2 - H(\theta)} \geq 0 \right.$$

which is equivalent to (16)-(17). $\square$

# Appendix B. Proof of Proposition 3

We provide two lemmas needed for the proof of Proposition 3.

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

*Proof.* 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.

$\square$

**Lemma 3.** *The feasible region of the problem defined by (16)-(17) is convex.*

*Proof.* Proof The feasible region of the problem defined by (16)-(17) is the set $F = \{\boldsymbol{\theta} \text{ s.t. } \Gamma_2 - H(\boldsymbol{\theta}) \geq 0\}$. After some algebra, we have

$$\Gamma_2 - H(\boldsymbol{\theta}) = \Gamma_2 - \frac{1}{T-p} \sum_{t=p+1}^{T} \left( \sum_{k=1}^{p} \theta_k X_{t-k} \right)^2 + \frac{2}{T-p} \sum_{t=p+1}^{T} \sum_{k=1}^{p} \theta_k X_{t-k} X_t -$$

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

From the definitions in $(21)$ it can be seen that

$$\Gamma_2 - H(\boldsymbol{\theta}) = \Gamma_2 - \sum_{k=1}^{p} \theta_k^2 V_k - 2 \sum_{k=1}^{p} \theta_k \sum_{r=k+1}^{p} \theta_r C_{k,r} + 2 \sum_{k=1}^{p} \theta_k C_{k,0} - V_0 \quad (24)$$

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

**Proof of Proposition 3:** Since the feasible region is convex then, 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. $\square$

# 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 $\Gamma_1$ is supposed to be positive. Since the objective function is linear on $\theta_1$, 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 (24) 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 (18).

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 $\theta_1$ 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 minimun is reached at $\Gamma_2 - H(\theta_1) = 0$,

(2.2) The minimun is reached at $\sqrt{\Gamma_2 - H(\theta_1)} = \Gamma_1$,

(2.3) The optimal $\theta_1$ is the one such that $\dfrac{\partial F_2(\theta_1)}{\partial \theta_1} = 0$.

Case (2.2) has been already solved. Consider the case (2.1). Then, after substituting $p = 1$ in (24), expression (19) is obtained. Finally, the problem to be solved for the case (2.3) turns into

$$
\begin{array}{c}
-\Delta + \min F_2(\theta_1) \\
\text{s.t} \begin{cases} \Gamma_2 - H(\theta_1) > 0 \\ \sqrt{\Gamma_2 - H(\theta_1)} \leq \Gamma_1 \end{cases}
\end{array}
\tag{25}
$$

An optimal solution to (25) is obtained in the case (2.3) if there exists $\theta_1$ such that $\dfrac{dF_2(\theta_1)}{d\theta_1} = 0$ , which in addition satisfies the constraints of problem (25). Such value $\theta_1$ 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:

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

from which the next quadratic function is obtained:

$$
\theta_1^2(V_1^2 + a^2 V_1) - 2\theta_1 C_{1,0}(V_1 + a^2) - a^2(\Gamma_2 - V_0 - C_{1,0}),
\tag{26}
$$

which is equal to zero if and only if $\theta_1 = \theta_1^{\star(3)}$ as in the expression (20). $\square$

# References

Aviv, Y. (2002). Gaining benefits from joint forecasting and replenishment processes: The case of auto-correlated demand. *Manufacturing & Service Operations Management*, 4(1):55–55.

Aviv, Y. (2003). A time-series framework for supply-chain inventory managemet. *Operations Research*, 51:210–227.

Baker, K. and Scudder, G. (1990). Sequencing with earliness and tardiness penalties. *Operations Research*, 38:22–36.

Bandi, C. and Bertsimas, D. (2012). Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming*, 134:23–70.

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

Box, E., Jenkins, G., and Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics.

Carrizosa, E., Olivares-Nadal, A. V., and Ramírez-Cobo, P. (2013). Time series interpolation via global optimization of moments fitting. *European Journal of Operational Research*, 230(1):97 – 112.

Chen, Y. F., Xu, M., and Zhang, Z. G. (2009). Technical notea risk-averse newsvendor model under the cvar criterion. *Operations Research*, 57(4):1040–1044.

Choi, S., Ruszczyński, A., and Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research*, 59(2):346–364.

Dana, J. D. and Petruzzi, N. C. (2001). Note: The newsvendor model with endogenous demand. *Management Science*, 47(11):1488–1497.

Ding, X., Puterman, M. L., and Bisi, A. (2002). The censored newsvendor and the optimal acquisition of information. *Operations Research*, 50(3):517–527.

Dong, L. and Lee, H. L. (2003). Optimal policies and approximations for a serial multiechelon inventory system with time-correlated demand. *Operations Research*, 51(6):969–980.

Eeckhoudt, L., Gollier, C., and Schlesinger, H. (1991). Increases in risk and deductible insurance. *Journal of Economic Theory*, 55:435–440.

Eeckhoudt, L., Gollier, C., and Schlesinger, H. (1995). The risk-averse (and prudent) newsboy. *Management Science*, 41(5):786–794.

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

32

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

Graves, S. C. (1999). A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management*, 1(1):50–61.

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

Jiang, H., Netessine, S., and Savin, S. (2011). Technical noterobust newsvendor competition under asymmetric information. *Operations research*, 59(1):254–261.

Johnson, G. and Thompson, H. (1975). Optimality of myopic inventory policies for certain dependent demand processes. *Management Science*, 21:1303–1307.

Kabak, I. and Schiff, A. (1978). Inventory models and management objectives. *Sloan Manage Rev*, 10:53–59.

Kahn, J. A. (1987). Inventories and the volatility of production. *The American Economic Review*, pgs. 667–679.

Khouja, M. (1999). The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega*, 27:537–553.

Lau, H. (1980). The newsboy problem under alternative optimization objectives. *Journal of Operational Research Society*, 31:525–535.

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

Levi, R., Roundy, R. O., Shmoys, D. B., et al. (2008). Approximation algorithms for capacitated stochastic inventory control models. *Operations Research*, 56(5):1184–1199.

Lu, X., Song, J., and Regan, A. (2006). Inventory planning with forecast upadtes: Approximate solutions and cost error bounds. *Operations Research*, 54:1079–1097.

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

Petruzzi, N. C. and Dada, M. (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research*, 47(2):183–194.

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., and Seref, M. M. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research*, 213(2):361–374.

Reyman, G. (1989). State reduction in a dependent demand inventory model given by a time series. *European Journal of Operational Research*, 41:174–180.

Rosenfield, D. (1986). Optimal management of tax-sheltered employee reimbursement programs. *Interfaces*, 16:68–72.

Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, pgs. 201–209.

Schweitzer, M. and Cachon, G. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, 46(3):404–420.

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

Shang, K. H. and Song, J.-S. (2003). Newsvendor bounds and heuristic for optimal policies in serial supply chains. *Management Science*, 49(5):618–638.

Thakkar, R., Finley, D., and Liao, W. (1983). A stochastic demand cvp model with return on investment criterion. *Contemp Account Res*, 1:77–86.

Wang, T., Atasu, A., and Kurtulus, M. (2012). A multiordering newsvendor model with dynamic forecast evolution. *Manufacturing & Service Operations Management*, 14:472–484.

Weatherford, L. and Pfeifer, P. (1994). The economic value of using advance booking of orders. *Omega*, 22:105–111.

Yue, J., Chen, B., and Wang, M. (2006). Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136.

Zhu, Z., Zhang, J., and Ye, Y. (2013). Newsvendor optimization with limited distribution information. *Optimization Methods and Software*, 28:640–667.