Math. Program., Ser. A (2015) 152:1–32
DOI 10.1007/s10107-014-0776-y

FULL LENGTH PAPER

Distributionally robust multi-item newsvendor problems with multimodal demand distributions

Grani A. Hanasusanto · Daniel Kuhn · Stein W. Wallace · Steve Zymler

Received: 18 December 2012 / Accepted: 29 March 2014 / Published online: 24 April 2014

Abstract We present a risk-averse multi-dimensional newsvendor model for a class of products whose demands are strongly correlated and subject to fashion trends that are not fully understood at the time when orders are placed. The demand distribution is known to be multimodal in the sense that there are spatially separated clusters of probability mass but otherwise lacks a complete description. We assume that the newsvendor hedges against distributional ambiguity by minimizing the worst-case risk of the order portfolio over all distributions that are compatible with the given modality information. We demonstrate that the resulting distributionally robust optimization problem is NP-hard but admits an efficient numerical solution in quadratic decision rules. This approximation is conservative and computationally tractable. Moreover, it achieves a high level of accuracy in numerical tests. We further demonstrate that disregarding ambiguity or multimodality can lead to unstable solutions that perform poorly in stress test experiments.

Mathematics Subject Classification 90C15 · 90C22

G. A. Hanasusanto (✉) · S. Zymler
Department of Computing, Imperial College London, London SW7 2AZ, UK
e-mail: g.hanasusanto11@imperial.ac.uk

D. Kuhn
Management of Technology and Entrepreneurship Institute,
École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

S. W. Wallace
Department of Business and Management Science,
Norwegian School of Economics, 5045 Bergen, Norway

1 Introduction

The multi-product newsvendor problem describes a fundamental dilemma faced by the seller of different perishable goods with fixed prices and uncertain demands [21]. The underlying challenge is to determine the optimal inventory levels of the products at the beginning of a business cycle, that is, before reliable estimates of the demands are available. This amounts to balancing the opportunity costs of lost sales due to stockouts versus the write-off of unsold inventory. If the seller is risk-neutral and therefore minimizes expected costs, the multi-product newsvendor model naturally decomposes into several single-item newsvendor problems. In this case the inventory levels of the different products can be chosen independently of one another. A risk-averse seller, in contrast, needs to coordinate the inventory levels in a joint effort in order to exploit diversification opportunities, see e.g. Choi et al. [8]. Using a risk-averse decision criterion can be vital for the newsvendor's long-term financial survival. Indeed, while a risk-neutral strategy eventually dominates any other strategy with probability one if the newsvendor problem is solved over and over again and demands are ergodic, the corresponding period-wise costs tend to display considerable variability. Thus, newsvendors with finite financial resources can be forced into bankruptcy relatively easily after just a few consecutive weak business cycles.

In this paper we investigate newsvendor problems whose demand distributions are known to be multimodal but otherwise lack a complete description. Throughout the paper we use the term 'multimodality' in a slightly informal way to refer to the tendency of a distribution to display several spatially disjoint regions of increased probability. The terms 'multimodal', 'bimodal', 'bimodality' etc. are to be interpreted analogously. Multimodality of demands is observed, for instance, in the following generic situations.

- *New products:* It is often difficult to predict if a newly launched product will become a top seller or a shelf warmer. In this case it is natural to assign the product a bimodal demand distribution.
- *Large customers:* Newsvendors serving many small customers with independent demands face an aggregate demand that is approximately Gaussian. In the presence of a large customer that accounts for a large share of the sales, however, demand may be multimodal due to irregular bulk orders.
- *New market entrants:* The emergence of a new major competitor can have a significant impact on demand. If the timing of market entry is uncertain and/or if there is a non-negligible possibility that many customers will switch to the new competitor, then the demand tends to exhibit a bimodality.
- *Fashion trends:* Apparel-type products are heavily influenced by fashion trends. These products are typically differentiated by style (e.g. design, color, clothing material), and at the time of inventory selection it may be unclear which style becomes popular. The corresponding demands are thus governed by multimodal distributions, where each mode reflects a particular popularity state.

Multimodality of demands is difficult to reconcile with the majority of newsvendor models considered in the literature, which typically assume unimodal, and often multi-normal, demand distributions. This bias towards unimodality and normality is partly caused by convenience as normal distributions lend themselves to mathematical manipulations. Of course, normality assumptions frequently enjoy strong theoretical justification. So the costs paid for the normality assumption, mostly in terms of inaccuracy, are more than offset by the generality of the results, which are often analytical and in a closed form.

In this paper we investigate newsvendor models in which unimodality or normality are not justifiable, and hence the price paid for these assumptions is potentially high. Such models have received little attention so far. Vaagen and Wallace [44] discuss production planning of high-technology sportswear where fashion plays a major role in the market. As these products are generally manufactured at specialized facilities in the Far East, before being sent to the USA and Europe to meet the season's demand, the producers find themselves in a newsvendor-like setting. Production must be determined before demand, including the leading fashion trends, become fully known. As each product class (such as mountain jackets for men) come in many models and colors—reflecting the possible fashion trends and customer needs—we are in fact facing a multi-dimensional newsvendor problem with correlated demands. Overviews of assortment planning models are found in Kök et al. [24] and Mahajan and Van Ryzin [26].

Many other products suffer from the same problem: They are primarily functional products, and quality plays a major role in the marketplace. But a miss on the fashion trend may lead to very low demand. Examples include snowboards, skis, surfboards, and certain types of watches and mobile phones. Sometimes the principle of postponement can be used, as was demonstrated by Benetton's T-shirts [40], but for the products we envisage here this is rarely possible for technical reasons. For example, a high-quality mountain jacket cannot be colored after arrival in Europe.

Vaagen and Wallace [44] demonstrate that not only is unimodality unreasonable for these products, but the assumption leads to solutions with unfavourable risk-reward tradeoffs. They therefore point out that numerical results, despite being less general than results from analytical models, are necessary in order to understand the fashion aspect of demand properly. Not many articles seem to take this numerical route. A notable exception is Kök and Fisher [23], but they are not concerned with multimodality.

Another critical assumption adopted in the vast majority of newsvendor models is that the demand distribution is precisely known. Kök et al. [24] explicitly confirm that this is the standard assumption. However, this assumption is rarely justifiable. In reality it may at best be possible to estimate a few low-order moments or to identify the range and some basic modality and symmetry properties of the demand distribution. This means that there is an *ambiguity set* of many different distributions that are consistent with the available information, and the decision maker has no possibility to single out the 'true' distribution among all conceivable distributions. Solving the newsvendor problem in view of an arbitrarily chosen distribution from the given ambiguity set can lead to decisions that perform poorly under the true demand distribution. This shortcoming can be rectified by adopting a *distributionally robust* approach which seeks a decision that achieves the best performance in view of the worst-case distribution within the ambiguity set. We emphasize that this decision criterion enjoys strong justification from the theory of choice in economics, see e.g. Ellsberg [15] and Gilboa and Schmeidler [17].

Scarf [35] proposed the first distributionally robust newsvendor model assuming that only the first two moments of the (univariate) demand distribution are known and derived an analytical expression for the optimal order quantity. This model was later extended in various directions by Gallego and Moon [16], but still only first- and second-order moments were assumed to be known. In order to mitigate the conservativeness of this worst-case approach, Perakis and Roels [31] derive order quantities that minimize the newsvendor's maximum opportunity cost from choosing a particular demand distribution within the ambiguity set. Natarajan et al. [27] introduce asymmetry into the robust newsvendor problem by using mean, variance and semivariance information to design ambiguity sets. This work was motivated by the observation that the optimal ordering quantity can be substantially overestimated if the demand distribution is skewed, see Bartezzaghi et al. [3]. All approaches discussed so far focus on closed-form solutions for single-product newsvendor problems. Wang et al. [46] propose an efficient computational procedure to solve a distributionally robust newsvendor problem with an ambiguity set containing all distributions that achieve a certain level of likelihood given a sequence of historical demand observations. Generic distributionally robust stochastic programming problems have been investigated, for example, by Dupačová [13,14], Shapiro and Ahmed [37], Delage and Ye [10], and Bertsimas et al. [6].

In this paper we study typical newsvendor situations where the demand distribution can be modeled as a mixture of m modes with known sizes (probability masses), locations (conditional mean values) and shapes (conditional covariance matrices). No other distributional information is assumed to be available. As it is impossible to construct a unique demand distribution from the given data, we embrace the distributionally robust approach and introduce a *multimodal ambiguity set* comprising all distributions matching the given conditional moments. We later generalize this ambiguity set to account for uncertainty in the probabilities and the conditional moments of the modes.

We highlight the following main contributions of this paper.

- We formulate a distributionally robust multi-product newsvendor model with a worst-case mean-risk objective. The worst case is taken over all demand distributions within the multimodal ambiguity set. This model captures generic inventory management situations that arise, for instance, when new products are introduced, demand is dominated by large customers, new competitors enter the market or products are subject to fashion trends.
- We prove that the distributionally robust multi-item newsvendor problem is NP-hard and disclose a connection to Manhattan norm maximization problems over ellipsoids.
- We prove that our newsvendor model has an exact reformulation as a semidefinite program (SDP) where the number of linear matrix inequalities scales exponentially with the number of products. By using quadratic decision rules to approximate the recourse decisions, we further show that our model admits a conservative SDP approximation that only involves a polynomial number of linear matrix inequalities. The resulting approximate problem is computationally tractable and can be solved efficiently with interior point algorithms.

- We show that our quadratic decision rule approximation achieves an acceptable level of accuracy on a wide spectrum of randomly generated test instances and distinctly outperforms the exact model in terms of the problem sizes it can handle. We further demonstrate that disregarding either the multimodality or the ambiguity of the demand distribution can lead to solutions that perform poorly if the data-generating distribution differs only slightly from the newsvendor's estimate.

The rest of the paper develops as follows. In Sect. 2 we introduce our distributionally robust multi-item newsvendor problem and characterize the underlying ambiguity set. We demonstrate that this model has an exact reformulation as an SDP of exponential size. By using quadratic decision rules, we further show that our model admits a conservative SDP approximation of polynomial size. Section 3 discusses extensions of the basic model and refinements of the decision rule approximation. The computational complexity of the model is analyzed in Sect. 4. Section 5 reports on numerical results that analyze the scalability of the proposed model and highlight the benefit of truthfully accounting for the multimodality and ambiguity of the demand distributions.

Notation We use lower-case bold face letters to denote vectors and upper-case bold face letters to denote matrices. We define $\boldsymbol{e} \in \mathbb{R}^n$ as the vector with all elements equal to 1, and we let $\boldsymbol{e}_i \in \mathbb{R}^n$ be the i-th standard basis vector. The space of symmetric matrices of dimension n is denoted by $\mathbb{S}^n$. For any two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$, we let $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Tr}(\mathbf{XY})$ be the trace scalar product, while the relation $\mathbf{X} \succcurlyeq \mathbf{Y}$ ($\mathbf{X} \succ \mathbf{Y}$) implies that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite (positive definite). We define diag($\boldsymbol{v}$)$\in \mathbb{S}^n$ as a diagonal matrix with the vector $\boldsymbol{v} \in \mathbb{R}^n$ on its main diagonal. Random variables are always represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. Furthermore, for $x \in \mathbb{R}$ we define $x^+ = \max(x, 0)$.

2 Risk-averse multi-product newsvendor problem

A newsvendor trades in different products with a non-negligible order lead time. The products are indexed by $i = 1, \ldots, n$. At the beginning of the business cycle, the newsvendor orders x_i units of product i at the wholesale price c_i. Only after the goods have been delivered, the demand ξ_i for product i is revealed, and an amount $y_i(\xi_i)$ is sold at the retail price v_i. Of course, the quantity sold must neither exceed the demand nor the initial stock, that is, $0 \le y_i(\xi_i) \le \min(x_i, \xi_i)$. Any unsold stock $[x_i - y_i(\xi_i)]^+$ is cleared at the salvage price g_i, and any unsatisfied demand $[\xi_i - y_i(\xi_i)]^+$ incurs a stock-out cost of b_i per unit. Throughout this paper we will assume that $c_i < v_i$ and $g_i < v_i$. This implies that the optimal sales decisions are of the form $y_i^*(\xi_i) = \min(x_i, \xi_i)$.

In the remainder we let $\boldsymbol{x} = (x_i)_{i=1}^n$ be the vector of order quantities, $\boldsymbol{\xi} = (\xi_i)_{i=1}^n$ the uncertain demand vector and $\boldsymbol{y}^*(\boldsymbol{\xi}) = \min(\boldsymbol{x}, \boldsymbol{\xi})$ the vector of optimal sales quantities, where 'min' stands for component-wise minimization. Moreover, $\boldsymbol{c} = (c_i)_{i=1}^n$ and $\boldsymbol{v} = (v_i)_{i=1}^n$ denote the vectors of wholesale and retail prices, while $\boldsymbol{g} = (g_i)_{i=1}^n$ and $\boldsymbol{b} = (b_i)_{i=1}^n$ denote the vectors of salvage prices and stock-out costs, respectively. Using this notation, the total cost incurred by an order portfolio $\boldsymbol{x}$ under demand scenario $\boldsymbol{\xi}$ can be represented as a piecewise linear loss function of the form

$$L(\boldsymbol{x}, \boldsymbol{\xi}) = \boldsymbol{c}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}^*(\boldsymbol{\xi}) - \boldsymbol{g}^\top (\boldsymbol{x} - \boldsymbol{y}^*(\boldsymbol{\xi})) + \boldsymbol{b}^\top (\boldsymbol{\xi} - \boldsymbol{y}^*(\boldsymbol{\xi}))\,.$$

2.1 Stochastic model

Assume first that the demand can be modeled by a random vector $\tilde{\boldsymbol{\xi}}$ with a *known* probability distribution $\mathbb{P}$, and let $\mathbb{E}_{\mathbb{P}}(\cdot)$ denote the expectation operator under $\mathbb{P}$. Moreover, assume that the newsvendor is risk-averse and uses the Conditional Value-at-Risk (CVaR) at level $\epsilon \in (0, 1)$ to quantify the riskiness of an order portfolio. A risk-sensitive decision criterion is necessary whenever the newsvendor's financial resources are not ample enough to survive a streak of several above-average losses. Our motivation for using CVaR as the risk measure is threefold. First, the CVaR at level ϵ has an intuitive interpretation as the average of the $100 \times \epsilon\%$ worst outcomes of the loss distribution and is widely used in industry. Second, CVaR is known to be a coherent risk-measure in the sense of Artzner et al. [1] and is consistent with first- second-order stochastic dominance [29]. Finally, the use of CVaR will lead to computationally tractable conic optimization models that can be solved efficiently with readily available solvers. Recall that the CVaR at level ϵ with respect to the distribution $\mathbb{P}$ is formally defined as

$$\mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})) = \min_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left([L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) - \beta]^+ \right) \right\} ,$$

see [34]. In this framework, a newsvendor concerned about the expected costs as well as the risk of the order portfolio will solve the following mean-risk optimization model

$$\underset{\boldsymbol{x} \in \mathbb{R}^n_+}{\text{minimize}} \quad \lambda \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})) + (1 - \lambda) \mathbb{E}_{\mathbb{P}}(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})) , \tag{1}$$

where $\lambda \in [0, 1]$ can be viewed as a parameter reflecting the newsvendor's risk-aversion. Substituting the known explicit expression for $\boldsymbol{y}^*(\boldsymbol{\xi})$ into the definition of $L(\boldsymbol{x}, \boldsymbol{\xi})$ yields

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\xi}) &= \boldsymbol{c}^\top \boldsymbol{x} - \boldsymbol{v}^\top \min(\boldsymbol{x}, \boldsymbol{\xi}) - \boldsymbol{g}^\top (\boldsymbol{x} - \min(\boldsymbol{x}, \boldsymbol{\xi})) + \boldsymbol{b}^\top (\boldsymbol{\xi} - \min(\boldsymbol{x}, \boldsymbol{\xi})) \\ &= \boldsymbol{d}^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{\xi} + \boldsymbol{h}^\top \max(\boldsymbol{x} - \boldsymbol{\xi}, \boldsymbol{0}) , \end{aligned} \tag{2}$$

where $\boldsymbol{d} = \boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b}$ and $\boldsymbol{h} = \boldsymbol{v} + \boldsymbol{b} - \boldsymbol{g}$. Since $\boldsymbol{v} > \boldsymbol{g}$ and $\boldsymbol{b} \geq \boldsymbol{0}$ by assumption, we conclude that $\boldsymbol{h} > \boldsymbol{0}$. This in turn implies that the loss function is jointly convex in $\boldsymbol{x}$ and $\boldsymbol{\xi}$. Problem (1) is therefore equivalent to the following convex single-stage stochastic program.

$$\underset{\beta \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n_+}{\text{minimize}} \quad \lambda \beta + \mathbb{E}_{\mathbb{P}} \left(\frac{\lambda}{\epsilon} \left[L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) - \beta \right]^+ + (1 - \lambda) L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right) \tag{3}$$

Remark 2.1 By the translation invariance of coherent risk measures, the newsvendor is indifferent between a certain loss of size $\lambda \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})) + (1 - \lambda) \mathbb{E}_{\mathbb{P}}(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}))$ and the random loss $L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})$. Thus, the objective of problem (1) has a physical interpretation as the certainty equivalent of $L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})$.

2.2 Distributionally robust model

In practice a newsvendor has to estimate the demand distribution $\mathbb{P}$ using limited structural information and historical data. Therefore, the assumption of full and accurate information about $\mathbb{P}$ is unrealistic. Replacing $\mathbb{P}$ in problem (1) with a noisy estimate $\hat{\mathbb{P}}$ (and thereby pretending that $\hat{\mathbb{P}}$ is the true demand distribution) typically results in order portfolios with a poor performance under the true distribution $\mathbb{P}$.

Instead of pretending to have full knowledge of $\mathbb{P}$, we will henceforth acknowledge our lack of distributional information: we will assume that $\mathbb{P}$ is ambiguous and only known to be an element of a given set $\mathcal{P}$, which can be viewed as a confidence region in the space of demand distributions. Under this kind of distributional ambiguity the newsvendor will solve the following variant of problem (1).

$$\underset{\boldsymbol{x}\in\mathbb{R}^n_+}{\text{minimize}} \quad \lambda \sup_{\mathbb{P}\in\mathcal{P}} \ \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x},\tilde{\boldsymbol{\xi}})) + (1-\lambda) \sup_{\mathbb{P}\in\mathcal{P}} \ \mathbb{E}_{\mathbb{P}}(L(\boldsymbol{x},\tilde{\boldsymbol{\xi}})) \tag{4}$$

Problem (4) constitutes a distributionally robust counterpart of (1). Its objective is to minimize a convex combination of the worst-case CVaR and the worst-case expectation of the loss function, where the worst case is taken with respect to all distributions in the set $\mathcal{P}$. This objective constitutes again a coherent risk measure [48], and its numerical value can be interpreted as the certainty equivalent of $L(\boldsymbol{x},\tilde{\boldsymbol{\xi}})$.

Remark 2.2 A less conservative variant of the robust problem (4) is obtained by requiring the distribution used in the CVaR and the expectation of the losses to be equal. This results in the following optimization model that accomodates a worst-case mean-risk objective.

$$\underset{\boldsymbol{x}\in\mathbb{R}^n_+}{\text{minimize}} \quad \sup_{\mathbb{P}\in\mathcal{P}} \left[\lambda \ \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x},\tilde{\boldsymbol{\xi}})) + (1-\lambda) \ \mathbb{E}_{\mathbb{P}}(L(\boldsymbol{x},\tilde{\boldsymbol{\xi}}))\right] \tag{5}$$

Problem (5) is in line with our intuition that there is a unique probabilistic model characterizing the random demand $\tilde{\boldsymbol{\xi}}$. Nevertheless, we will work with problem (4) in the remainder of the paper for the following reasons.

1. Numerical experiments show that the optimal order portfolios for the models (4) and (5) are qualitatively equal.
2. The optimal order portfolios resulting from model (4) are Pareto efficient in the sense that they offer an optimal trade-off between worst-case risk and worst-case expected loss. The corresponding Pareto frontier can be computed efficiently by using the techniques developed in the remainder of this paper. In contrast, the optimal portfolios corresponding to (5) are not Pareto efficient in the above sense. Moreover, they even fail to be Pareto efficient if the mean and risk are evaluated under any crisp distribution $\mathbb{P} \in \mathcal{P}$ or with respect to some adaptable worst-case distribution that is extremal in (5) and thus changes with λ. Note that (4) is equivalent to a model that minimizes worst-case risk subject to a (suitably chosen) upper bound on the worst-case expected losses. In contrast, problem (5) does not admit a similar reformulation.

From now on we assume that only certain properties of the distribution $\mathbb{P}$ are known. In particular, we assume that $\mathbb{P} = \sum_{j=1}^{m} p_j \mathbb{P}_j$ is known to be a mixture of m distinct distributions $\mathbb{P}_1, \ldots, \mathbb{P}_m$, with known mixture probabilities $p_1, \ldots, p_m \in [0, 1]$, $\sum_{j=1}^{m} p_j = 1$. We further assume that each $\mathbb{P}_j$ is only known to be supported on a non-degenerate ellipsoid Ξ_j and to have a known mean value $\boldsymbol{\mu}_j$ and a known covariance matrix $\boldsymbol{\Sigma}_j \succ 0$, $j = 1, \ldots, m$. Each ellipsoidal support set Ξ_j is representable as

$$\Xi_j = \left\{ \boldsymbol{\xi} \in \mathbb{R}^n : (\boldsymbol{\xi} - \boldsymbol{\nu}_j)\boldsymbol{\Lambda}_j^{-1}(\boldsymbol{\xi} - \boldsymbol{\nu}_j) \leq \delta_j^2 \right\}, \tag{6}$$

where $\boldsymbol{\nu}_j \in \mathbb{R}^n$, $\boldsymbol{\Lambda}_j \in \mathbb{S}^n$, $\boldsymbol{\Lambda}_j \succ \mathbf{0}$ and $\delta_j \in \mathbb{R}_+$ determine the center, shape and size of Ξ_j, respectively. In other words, $\mathbb{P}$ is a mixture of m ambiguous distributions, which are unknown beyond their supports and first- and second-order moments. Thus, $\mathbb{P}$ is only known to belong to the ambiguity set

$$\mathcal{P} = \sum_{j=1}^{m} p_j \, \mathcal{P}(\Xi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \tag{7}$$

where

$$\mathcal{P}(\Xi, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left\{ \mu \in \mathcal{M}_+ : \int_{\Xi} \mu(\mathrm{d}\boldsymbol{\xi}) = 1, \int_{\Xi} \boldsymbol{\xi} \mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu}, \int_{\Xi} \boldsymbol{\xi}\boldsymbol{\xi}^\intercal \mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \right\}$$

denotes the set of all distributions supported on Ξ that share the same mean value $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}^n$, $\boldsymbol{\Sigma} \succ \mathbf{0}$, while $\mathcal{M}_+$ denotes the cone of non-negative Borel measures on $\mathbb{R}^n$. The set addition in (7) is understood in the sense of Minkowski.

The m mixture components or modes of the ambiguous demand distribution can be viewed as different popularity states of the products. If a particular product is assigned a high/low expected value in state j, that is, if its component in the mean vector $\boldsymbol{\mu}_j$ exceeds/undershoots some prescribed threshold, then this product is considered to be popular/unpopular in the jth state. The support set Ξ_j and the covariance matrix $\boldsymbol{\Sigma}_j$ provide information about the conditional demand uncertainty in state j. The choice of ellipsoidal support sets has distinct computational advantages as it facilitates a conic reformulation of the distributionally robust newsvendor problem (4) in Sect. 2.3. Moreover, the Ξ_j are conveniently found by clustering algorithms that cover the demand data by minimum-volume ellipsoids [42].

The newsvendor model introduced above has several important applications as outlined in the introduction. A prototypical example where multimodality and ambiguity are pertinent features of the demand distribution arises in the textile apparel industry. Consider a supplier selling different variants of a well-defined fashion apparel category that are distinguished by style and color. The joint demand distribution has two states representing separated clusters of probability mass. In the first state some variants (e.g., textiles in bright colors) are popular while others (e.g., textiles in pastel colors) are unpopular, and in state 2 the situation is reversed. A detailed case study centered around this scenario is described in [44], where the demand distribution is modeled as a mixture of two lognormal modes. As typical business cycles in the fashion apparel industry are of the order of quarters or even years, however, it is unlikely that the supplier has enough data to infer a unique demand distribution. Instead, one can at best construct an ambiguity set of distributions that are compatible with the given data.

In the remainder of this section we derive semidefinite programming (SDP) based reformulations of the worst-case CVaR in (4) under the ambiguity set (7). As the worst-case expected loss coincides with the worst-case CVaR for $\epsilon = 1$, these results will enable us to reformulate (4) as an SDP.

2.3 Worst-case conditional value-at-risk

By using the definition of CVaR and a stochastic min-max theorem by Shapiro and Kleywegt [39], we can reexpress the worst-case CVaR in the objective function of problem (4) as

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})) = \inf_\beta \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}([L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) - \beta]^+)\,. \tag{8}$$

Note that the inner maximization problem on the right hand side of the above equation is equivalent to the following semi-infinite linear program

$$\begin{array}{lll} \sup\limits_{\mu_1,\ldots,\mu_m\in\mathcal{M}_+} & \displaystyle\sum_{j=1}^m p_j \int_{\Xi_j} \max(0, L(\boldsymbol{x}, \boldsymbol{\xi}) - \beta)\; \mu_j(\mathrm{d}\boldsymbol{\xi}) & \\ \text{s.t.} & \left.\begin{array}{l} \displaystyle\int_{\Xi_j} \mu_j(\mathrm{d}\boldsymbol{\xi}) = 1 \\ \displaystyle\int_{\Xi_j} \boldsymbol{\xi}\mu_j(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu}_j \\ \displaystyle\int_{\Xi_j} \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}\mu_j(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^\mathsf{T} \end{array}\right\} & \forall j = 1, \ldots, m\,, \end{array} \tag{9}$$

where the optimization variables are the non-negative measures $\mu_1, \ldots, \mu_m$. The constraints restrict $\mu_1, \ldots, \mu_m$ to be probability measures with prescribed first- and second-order moments. We remark that the worst-case expectation problem (9) is separable with respect to the m states, but the worst-case CVaR problem (8) is not as β must be equal in all states. We now apply the variable transformation $p_j\mu_j \to \mu_j$, $j = 1, \ldots, m$, and then formulate the dual of problem (9), see e.g.[36].

$$\begin{array}{ll} \inf & \displaystyle\sum_{j=1}^m p_j(y_j + \boldsymbol{y}_j^\mathsf{T}\boldsymbol{\mu}_j + \langle \mathbf{Y}_j, \boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^\mathsf{T}\rangle) \\ \text{s.t.} & y_j \in \mathbb{R}, \quad \boldsymbol{y}_j \in \mathbb{R}^n, \quad \mathbf{Y}_j \in \mathbb{S}^n \quad \forall j = 1, \ldots, m \\ & y_j + \boldsymbol{y}_j^\mathsf{T}\boldsymbol{\xi} + \langle \mathbf{Y}_j, \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}\rangle \geq \max(0, L(\boldsymbol{x}, \boldsymbol{\xi}) - \beta) \quad \forall \boldsymbol{\xi} \in \Xi_j, \;\; j = 1, \ldots, m \end{array} \tag{10}$$

Because $\boldsymbol{\Sigma}_j \succ \mathbf{0}$ for all $j = 1, \ldots, m$, it can be shown that strong duality holds [22]. Therefore, the optimal values of (9) and (10) coincide. By redefining the decision variables in (10) as

$$\mathbf{M}_j = \begin{bmatrix} \mathbf{Y}_j & \frac{1}{2}\boldsymbol{y}_j \\ \frac{1}{2}\boldsymbol{y}_j^\top & y_j \end{bmatrix} \quad \forall j = 1, \ldots, m\,,$$

and by introducing $\boldsymbol{\Omega}_j$ to denote the second-order moment matrices of the conditional distributions,

$$\boldsymbol{\Omega}_j = \begin{bmatrix} \boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top & \boldsymbol{\mu}_j \\ \boldsymbol{\mu}_j^\top & 1 \end{bmatrix} \quad \forall j = 1, \ldots, m\,,$$

problem (10) can be simplified to

$$\begin{array}{ll} \inf\limits_{\mathbf{M}_1,\ldots,\mathbf{M}_m \in \mathbb{S}^{n+1}} & \sum\limits_{j=1}^{m} p_j \langle \boldsymbol{\Omega}_j, \mathbf{M}_j \rangle \\ \text{s.t.} & \left[\boldsymbol{\xi}^\top\ 1\right] \mathbf{M}_j \left[\boldsymbol{\xi}^\top\ 1\right]^\top \geq \max(0, L(\boldsymbol{x},\boldsymbol{\xi}) - \beta) \quad \forall \boldsymbol{\xi} \in \Xi_j,\ j = 1, \ldots, m\,. \end{array} \tag{11}$$

Note that the semi-infinite constraint in (11) can be decomposed as follows.

$$\left[\boldsymbol{\xi}^\top\ 1\right] \mathbf{M}_j \left[\boldsymbol{\xi}^\top\ 1\right]^\top \geq 0 \quad \forall \boldsymbol{\xi} \in \Xi_j,\ j = 1, \ldots, m \tag{12a}$$

$$\left[\boldsymbol{\xi}^\top\ 1\right] \mathbf{M}_j \left[\boldsymbol{\xi}^\top\ 1\right]^\top \geq L(\boldsymbol{x},\boldsymbol{\xi}) - \beta \quad \forall \boldsymbol{\xi} \in \Xi_j,\ j = 1, \ldots, m \tag{12b}$$

In the remainder we will invoke the $\mathcal{S}$-lemma [32] to reexpress these semi-infinite constraints in terms of linear matrix inequalities (LMI). To this end, we introduce the constant matrices

$$\mathbf{W}_j = \begin{bmatrix} \boldsymbol{\Lambda}_j^{-1} & -\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\nu}_j \\ -(\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\nu}_j)^\top & \boldsymbol{\nu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\nu}_j - \delta_j^2 \end{bmatrix} \quad \forall j = 1, \ldots, m\,, \tag{13}$$

where $\boldsymbol{\nu}_j$, $\boldsymbol{\Lambda}_j$, and δ_j are the parameters characterizing the ellipsoid Ξ_j. The $\mathcal{S}$-lemma then implies that (12a) is equivalent to m LMIs of the form

$$\forall j = 1, \ldots, m\ \exists \alpha_j \in \mathbb{R}_+ : \mathbf{M}_j + \alpha_j \mathbf{W}_j \succcurlyeq \mathbf{0}. \tag{14}$$

Constraint (12b) has a less benign structure due to the max terms in the definition of the loss function. Such constraints have been studied systematically in the context of classical robust optimization by Gorissen and den Hertog [19]. In the remainder we propose two methods to reexpress constraint (12b) in terms of finitely many LMIs. The first approach consists in a complete expansion of the maximization operations in the definition of $L(\boldsymbol{x},\boldsymbol{\xi})$ and provides a lossless reformulation of (12b). Unfortunately, this approach results in $\mathcal{O}(m \times 2^n)$ LMI constraints and is therefore computationally burdensome. The second approach exploits quadratic decision rules to derive a conservative but tractable approximation for (12b).

2.3.1 Approach 1: Complete expansion

By substituting the expression (2) for the loss function into (12b) we obtain

$$\left[\boldsymbol{\xi}^\top\ 1\right]\mathbf{M}_j\left[\boldsymbol{\xi}^\top\ 1\right]^\top \geq \boldsymbol{d}^\top\boldsymbol{x}+\boldsymbol{b}^\top\boldsymbol{\xi}+\boldsymbol{h}^\top\max(\boldsymbol{x}-\boldsymbol{\xi},\mathbf{0})-\beta \quad \forall\boldsymbol{\xi}\in\Xi_j,\ \ j=1,\ldots,m. \tag{15}$$

Thus, for any fixed $\boldsymbol{x}$ and β the quadratic function $\left[\boldsymbol{\xi}^\top\ 1\right]\mathbf{M}_j\left[\boldsymbol{\xi}^\top\ 1\right]^\top$ is required to exceed the piecewise-linear convex function on the right-hand side of (15), which can also be viewed as the upper envelope of 2^n affine functions (recall that $\boldsymbol{h} > \mathbf{0}$). Indeed, the term

$$\boldsymbol{h}^\top\max(\boldsymbol{x}-\boldsymbol{\xi},\mathbf{0}) = \sum_{i=1}^{n} h_i\max(x_i-\xi_i,0) \tag{16}$$

represents a superposition of n max functions with independent arguments and has therefore 2^n pieces of linearity. Hence, (15) is equivalent to the requirement that $\left[\boldsymbol{\xi}^\top\ 1\right]\mathbf{M}_j\left[\boldsymbol{\xi}^\top\ 1\right]^\top$ exceeds 2^n affine functions individually. In order to formalize the above arguments, we let $\mathbf{B}\in\mathbb{R}^{n\times 2^n}$ denote the matrix whose columns represent all binary numbers with n digits in ascending order. For example, for $n=2$ we set

$$\mathbf{B}=\begin{bmatrix}0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1\end{bmatrix}.$$

Furthermore, we introduce functions $\mathcal{I}_k:\mathbb{R}^n\to\mathbb{R}^n$ for $k=1,\ldots,2^n$, which are defined through

$$[\mathcal{I}_k(\boldsymbol{h})]_i=\begin{cases} h_i & \text{if} B_{ik}=1,\\ 0 & \text{otherwise,}\end{cases} \qquad \forall i=1,\ldots,n.$$

Using this notation, we can reexpress the semi-infinite constraint (15) as

$$\left[\boldsymbol{\xi}^\top\ 1\right]\mathbf{M}_j\left[\boldsymbol{\xi}^\top\ 1\right]^\top \geq \boldsymbol{d}^\top\boldsymbol{x}+\boldsymbol{b}^\top\boldsymbol{\xi}+\mathcal{I}_k(\boldsymbol{h})^\top(\boldsymbol{x}-\boldsymbol{\xi})-\beta \quad \forall\boldsymbol{\xi}\in\Xi_j,\ \ j=1,\ldots,m,$$
$$k=1,\ldots,2^n,$$

which by the $\mathcal{S}$-lemma is equivalent to

$$\forall j=1,\ldots,m,\ k=1,\ldots,2^n\ \exists\gamma_{jk}\in\mathbb{R}_+:\mathbf{M}_j+\gamma_{jk}\mathbf{W}_j$$
$$\succcurlyeq\begin{bmatrix}\mathbf{0} & \frac{1}{2}(\boldsymbol{b}-\mathcal{I}_k(\boldsymbol{h}))\\ \frac{1}{2}(\boldsymbol{b}-\mathcal{I}_k(\boldsymbol{h}))^\top & (\boldsymbol{d}+\mathcal{I}_k(\boldsymbol{h}))^\top\boldsymbol{x}-\beta\end{bmatrix}.$$

Substituting the above LMI reformulations of (12) into (11) yields the following equivalent reformulation of the semi-infinite linear program (9).

$$\begin{array}{ll} \inf & \sum_{j=1}^{m} p_j \langle \mathbf{\Omega}_j, \mathbf{M}_j \rangle \\ \text{s.t.} & \mathbf{M}_j \in \mathbb{S}^{n+1}, \quad \alpha_j,\ \gamma_{jk} \in \mathbb{R}_+ \quad \forall j = 1, \ldots, m,\ k = 1, \ldots, 2^n \\ & \mathbf{M}_j + \alpha_j \mathbf{W}_j \succcurlyeq \mathbf{0} \quad \forall j = 1, \ldots, m \\ & \mathbf{M}_j + \gamma_{jk} \mathbf{W}_j \succcurlyeq \begin{bmatrix} \mathbf{0} & \frac{1}{2}(\boldsymbol{b} - \mathcal{I}_k(\boldsymbol{h})) \\ \frac{1}{2}(\boldsymbol{b} - \mathcal{I}_k(\boldsymbol{h}))^\intercal & (\boldsymbol{d} + \mathcal{I}_k(\boldsymbol{h}))^\intercal \boldsymbol{x} - \beta \end{bmatrix} \quad \begin{array}{l} \forall j = 1, \ldots, m \\ \forall k = 1, \ldots, 2^n \end{array} \end{array} \tag{17}$$

Furthermore, substituting (17) into (8) yields the desired SDP representation of the worst-case CVaR.

$$\begin{array}{ll} \sup_{\mathbb{P} \in \mathcal{P}} & \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})) \\ = \inf & \beta + \frac{1}{\epsilon} \sum_{j=1}^{m} p_j \langle \mathbf{M}_j, \mathbf{\Omega}_j \rangle \\ \text{s.t.} & \beta \in \mathbb{R},\ \mathbf{M}_j \in \mathbb{S}^{n+1}, \quad \alpha_j,\ \gamma_{jk} \in \mathbb{R}_+ \ \forall j=1, \ldots, m,\ k=1, \ldots, 2^n \\ & \mathbf{M}_j + \alpha_j \mathbf{W}_j \succcurlyeq \mathbf{0} \quad \forall j = 1, \ldots, m \\ & \mathbf{M}_j + \gamma_{jk} \mathbf{W}_j \succcurlyeq \begin{bmatrix} \mathbf{0} & \frac{1}{2}(\boldsymbol{b} - \mathcal{I}_k(\boldsymbol{h})) \\ \frac{1}{2}(\boldsymbol{b} - \mathcal{I}_k(\boldsymbol{h}))^\intercal & (\boldsymbol{d} + \mathcal{I}_k(\boldsymbol{h}))^\intercal \boldsymbol{x} - \beta \end{bmatrix} \quad \begin{array}{l} \forall j = 1, \ldots, m \\ \forall k = 1, \ldots, 2^n \end{array} \end{array} \tag{18}$$

Although (18) constitutes a finite SDP, it involves $\mathcal{O}(m \times 2^n)$ LMI constraints and therefore fails to scale with the number of products. As a result, the problem is solvable only for small to medium values of n. Note that for $\epsilon = 1$ the SDP (18) provides an exact reformulation of the worst-case expectation in (4). In "Appendix" we describe a method to construct extremal distributions for the worst-case CVaR problem (8). These extremal distributions will be useful to perform targeted stress tests for different order portfolios in Sect. 5.

2.3.2 Approach 2: Quadratic decision rule approximation

A conservative approximation for the semi-infinite constraint (15) that avoids the exponential number of constraints characteristic for the full expansion model is obtained by introducing new decision variables $\boldsymbol{q}_j, \boldsymbol{\ell}_j, \boldsymbol{z}_j \in \mathbb{R}^n$, $j = 1, \ldots, m$, and imposing the constraints

$$\left.\begin{array}{l} \left[\boldsymbol{\xi}^\intercal\ 1\right] \mathbf{M}_j \left[\boldsymbol{\xi}^\intercal\ 1\right]^\intercal \geq \boldsymbol{d}^\intercal \boldsymbol{x} + \boldsymbol{b}^\intercal \boldsymbol{\xi} + \boldsymbol{h}^\intercal (\boldsymbol{\xi} \circ \boldsymbol{q}_j \circ \boldsymbol{\xi} + \boldsymbol{\ell}_j \circ \boldsymbol{\xi} + \boldsymbol{z}_j) - \beta \\ \boldsymbol{\xi} \circ \boldsymbol{q}_j \circ \boldsymbol{\xi} + \boldsymbol{\ell}_j \circ \boldsymbol{\xi} + \boldsymbol{z}_j \geq \max(\boldsymbol{x} - \boldsymbol{\xi}, \mathbf{0}) \end{array}\right\} \\ \forall \boldsymbol{\xi} \in \Xi_j,\ j = 1, \ldots, m, \tag{19}$$

where '$\circ$' denotes the Hadamard product. The approximation thus arises from sandwiching a separable quadratic function between a non-separable quadratic function and a piecewise linear separable max function. This is equivalent to replacing the individual max functions in (15) with quadratic overestimators. We refer to these overestimators as *quadratic decision rules*. We remark that quadratic decision rules are generally known to provide tractable approximations for dynamic optimization problems [5, p. 382], but they can also render some problems infeasible. However, this may not happen in our setting. As Ξ_j is compact, any continuous function on Ξ_j is bounded above by a constant and—a fortiori—a (separable) quadratic function. It is thus clear that for any fixed $\boldsymbol{x}$ one can find $\boldsymbol{q}_j$, $\boldsymbol{\ell}_j$, $\boldsymbol{z}_j$ and $\mathbf{M}_j$ satisfying (19). As the max function has only linear asymptotic growth, one can show that (19) remains feasible even if Ξ_j is inflated to the entire space.

We now demonstrate that (19) constitutes a tractable constraint system. Indeed, note that the first constraint in (19) can be reexpressed via the $\mathcal{S}$-lemma as

$$\begin{aligned}&\left[\boldsymbol{\xi}^\intercal\ 1\right]\mathbf{M}_j\left[\boldsymbol{\xi}^\intercal\ 1\right]^\intercal \geq \boldsymbol{d}^\intercal\boldsymbol{x}+\boldsymbol{b}^\intercal\boldsymbol{\xi}+\boldsymbol{h}^\intercal(\boldsymbol{\xi}\circ\boldsymbol{q}_j\circ\boldsymbol{\xi}+\boldsymbol{\ell}_j\circ\boldsymbol{\xi}+\boldsymbol{z}_j)-\beta\ \forall\boldsymbol{\xi}\in\Xi_j,\ j=1,\dots,m\\ &\iff \forall j=1,\dots,m\ \exists\gamma_j\in\mathbb{R}_+\ :\ \mathbf{M}_j+\gamma_j\mathbf{W}_j\succcurlyeq\begin{bmatrix}\operatorname{diag}(\boldsymbol{h}\circ\boldsymbol{q}_j) & \frac{1}{2}(\boldsymbol{b}+\boldsymbol{h}\circ\boldsymbol{\ell}_j)\\ \frac{1}{2}(\boldsymbol{b}+\boldsymbol{h}\circ\boldsymbol{\ell}_j)^\intercal & \boldsymbol{d}^\intercal\boldsymbol{x}+\boldsymbol{h}^\intercal\boldsymbol{z}_j-\beta\end{bmatrix}.\end{aligned}\tag{20a}$$

The second constraint can be reformulated as

$$\begin{aligned}&\boldsymbol{\xi}\circ\boldsymbol{q}_j\circ\boldsymbol{\xi}+\boldsymbol{\ell}_j\circ\boldsymbol{\xi}+\boldsymbol{z}_j\geq\max(\boldsymbol{x}-\boldsymbol{\xi},\boldsymbol{0})\quad\forall\boldsymbol{\xi}\in\Xi_j,\ j=1,\dots,m\\ &\iff q_{ji}\xi_i^2+\ell_{ji}\xi_i+z_{ji}\geq x_i-\xi_i,\quad q_{ji}\xi_i^2+\ell_{ji}\xi_i+z_{ji}\geq 0\\ &\quad\ \forall\xi_i\in\Xi_{ji},\ i=1,\dots,n,\ j=1,\dots,m\,,\end{aligned}$$

where $\Xi_{ji}=\{\xi_i\in\mathbb{R}:\frac{(\xi_i-\nu_{ji})^2}{[\Lambda_j]_{ii}}\leq\delta_j^2\}$ denotes the marginal projection of Ξ_j onto the i-th coordinate. By introducing the constant matrices

$$\mathbf{W}_{ji}=\begin{bmatrix}1 & -\nu_{ji}\\ -\nu_{ji} & \nu_{ji}^2-\delta_j^2[\Lambda_j]_{ii}\end{bmatrix}\quad\forall i=1,\dots,n\,,$$

and by invoking the $\mathcal{S}$-lemma, the second constraint is equivalent to

$$\begin{array}{l}\forall i=1,\dots,n\\ \forall j=1,\dots,m\end{array}\exists\phi_{ji},\psi_{ji}\in\mathbb{R}_+\ :\ \begin{array}{l}\begin{bmatrix}q_{ji} & \frac{1}{2}\ell_{ji}\\ \frac{1}{2}\ell_{ji} & z_{ji}\end{bmatrix}+\phi_{ji}\mathbf{W}_{ji}\succcurlyeq\mathbf{0}\\ \begin{bmatrix}q_{ji} & \frac{1}{2}(\ell_{ji}+1)\\ \frac{1}{2}(\ell_{ji}+1) & z_{ji}-x_i\end{bmatrix}+\psi_{ji}\mathbf{W}_{ji}\succcurlyeq\mathbf{0}\,.\end{array}\tag{20b}$$

Thus, the semi-infinite constraints (19) are equivalent to the tractable LMIs (20a) and (20b). Substituting these LMIs into (11) yields an upper bound approximation for the semi-infinite linear program (9), which in turn gives rise to the following conservative SDP-based approximation for the worst-case CVaR

$$\begin{aligned}&\sup_{\mathbb{P}\in\mathcal{P}}\quad \mathbb{P}\text{-CVaR}_\epsilon(L(\boldsymbol{x},\tilde{\boldsymbol{\xi}}))\\ &\leq\inf\quad \beta+\frac{1}{\epsilon}\sum_{j=1}^m p_j\langle\mathbf{M}_j,\boldsymbol{\Omega}_j\rangle\end{aligned}$$

$$
\begin{aligned}
\text{s.t.}\quad & \beta \in \mathbb{R}, \quad \mathbf{M}_j \in \mathbb{S}^{n+1}, \quad \boldsymbol{q}_j,\ \boldsymbol{\ell}_j,\ \boldsymbol{z}_j \in \mathbb{R}^n \forall\, j = 1, \ldots, m \\
& \alpha_j,\ \gamma_j,\ \phi_{ji},\ \psi_{ji} \in \mathbb{R}_+ \quad \forall i = 1, \ldots, n,\ j = 1, \ldots, m \\
& \mathbf{M}_j + \alpha_j \mathbf{W}_j \succcurlyeq \mathbf{0} \quad \forall j = 1, \ldots, m \\
& \mathbf{M}_j + \gamma_j \mathbf{W}_j \succcurlyeq \begin{bmatrix} \operatorname{diag}(\boldsymbol{h} \circ \boldsymbol{q}_j) & \frac{1}{2}(\boldsymbol{b} + \boldsymbol{h} \circ \boldsymbol{\ell}_j) \\ \frac{1}{2}(\boldsymbol{b} + \boldsymbol{h} \circ \boldsymbol{\ell}_j)^\intercal & \boldsymbol{d}^\intercal \boldsymbol{x} + \boldsymbol{h}^\intercal \boldsymbol{z}_j - \beta \end{bmatrix} \forall j = 1, \ldots, m \\
& \left.\begin{aligned} & \begin{bmatrix} q_{ji} & \frac{1}{2}\ell_{ji} \\ \frac{1}{2}\ell_{ji} & z_{ji} \end{bmatrix} + \phi_{ji} \mathbf{W}_{ji} \succcurlyeq \mathbf{0} \\ & \begin{bmatrix} q_{ji} & \frac{1}{2}(\ell_{ji}+1) \\ \frac{1}{2}(\ell_{ji}+1) & z_{ji} - x_i \end{bmatrix} + \psi_{ji} \mathbf{W}_{ji} \succcurlyeq \mathbf{0} \end{aligned}\right\} \begin{aligned} & \forall i = 1, \ldots, n \\ & \forall j = 1, \ldots, m \end{aligned}
\end{aligned} \tag{21}
$$

Unlike the exact SDP reformulation (18), problem (21) involves only $\mathcal{O}(m \times n)$ LMI constraints, which ensures its computational tractability. Note that for $\epsilon = 1$ the SDP (21) provides a tractable approximation for the worst-case expectation in (4).

The main results of this section can be summarized as follows. By using the full expansion approach to re-express the worst-case CVaR and the worst-case expectation in (4) as the optimal values of two conic programs, we obtain an exact SDP reformulation of the distributionally robust newsvendor problem. However, the resulting SDP scales exponentially with the number of products. In contrast, using the quadratic decision rule approach to approximate the worst-case CVaR and worst-case expectation results in a computationally tractable conservative approximation.

Remark 2.3 One could also envisage richer ambiguity sets that contain information about higher-order moments, directional deviations or location and dispersion measures from robust statistics. In these cases it would still be possible to derive conic reformulations of the distributionally robust newsvendor problem by using methods from [18,33,47,49] and the references therein.

3 Extensions

We now discuss extensions that enhance the realism of the basic newsvendor model and develop refinements of the decision rule approximation. Section 3.1 presents a model that accounts for uncertainty in the state probabilities and conditional moments. Section 3.2 develops tractable partial expansion bounds that allow us to better control the trade-off between approximation quality and computational effort.

3.1 Second layer of robustness

In practice, the statewise second-order moment matrices $\boldsymbol{\Omega}_j$, $j = 1, \ldots, m$, and the state probability vector $\boldsymbol{p} \in \mathbb{R}^m_+$ must be estimated from noisy data and are therefore corrupted by estimation errors. We now derive an extended newsvendor model that is immunized against this type of parameter uncertainty.

We assume that each $\boldsymbol{\Omega}_j$ is only known to be contained in a box uncertainty set (or multi-dimensional interval) of the form $[\underline{\boldsymbol{\Omega}}_j, \overline{\boldsymbol{\Omega}}_j]$. To ensure that this set contains a valid second-order moment matrix, we require that $[\overline{\mathbf{\Omega}}_j]_{n+1,n+1} = [\underline{\mathbf{\Omega}}_j]_{n+1,n+1} = 1$ and that there exists $\mathbf{\Omega}_j \succ \mathbf{0}$ with $\underline{\mathbf{\Omega}}_j \le \mathbf{\Omega}_j \le \overline{\mathbf{\Omega}}_j$. Separate uncertainty sets of this type for the mean vector and the covariance matrix were considered in [28]. Our model is slightly different in that we define a single uncertainty set for the second-order moment matrix. This approach has two advantages. First, the resulting SDP reformulation of the newsvendor problem accommodates fewer explicit constraints and is therefore leaner. Second, the combined uncertainty set is less conservative as the positive semidefiniteness of $\mathbf{\Omega}_j$ also impacts the first-order moments. This will typically push the worst-case mean vector away from the corner points of its marginal uncertainty set.

We further assume that $\boldsymbol{p} \in \mathbb{R}^m_+$ is only known to be close to a nominal (e.g., estimated) state probability vector $\hat{\boldsymbol{p}} \in \mathbb{R}^m_+$ with respect to the χ^2-distance. Thus, $\boldsymbol{p}$ is contained in the uncertainty set

$$\Delta = \left\{ \boldsymbol{p} \in \mathbb{R}^m_+ : \boldsymbol{e}^\mathsf{T}\boldsymbol{p} = 1,\ \sum_{j=1}^m (p_j - \hat{p}_j)^2/p_j \le \delta_{\boldsymbol{p}} \right\}, \tag{22}$$

where $\delta_{\boldsymbol{p}} \ge 0$ controls the level of robustness. The χ^2-distance belongs to the class ϕ-divergences [30], and the construction of uncertainty sets for ambiguous probability vectors via ϕ-divergences was propagated in [4]. Our motivation for using the χ^2-distance to construct Δ is threefold. First, Δ is determined by a single size parameter $\delta_{\boldsymbol{p}}$ which can easily be calibrated, e.g., via cross-validation. Secondly, the χ^2-distance guarantees that any weight vector $\boldsymbol{p} \in \Delta$ assigns nonzero probability to all states that have nonzero probability under the nominal weight vector $\hat{\boldsymbol{p}}$. Finally, the structure of Δ implied by the χ^2-distance has distinct computational benefits that will become evident in Proposition 3.1.

Remark 3.1 From a purely computational point of view, the uncertainty set Δ could also be designed differently. As will become clear from the discussion below, we only need to require that the maximization problem $\max_{\boldsymbol{p}\in\Delta} \boldsymbol{\varphi}^\mathsf{T}\boldsymbol{p}$ admits a strong dual and that the dual has a tractable conic representation for any fixed $\boldsymbol{\varphi} \in \mathbb{R}^m$. For example, we could allow for (nonempty) polyhedral uncertainty sets of the form $\Delta = \{\boldsymbol{p} \in \mathbb{R}^m_+ : \boldsymbol{e}^\mathsf{T}\boldsymbol{p} = 1,\ \mathbf{A}\boldsymbol{p} \le \boldsymbol{b}\}$, where the matrix $\mathbf{A}$ and the vector $\boldsymbol{b}$ have appropriate dimensions.

An ambiguity set for the demand distribution that accounts both for the uncertainty in the state probability vector as well as the statewise second-order moment matrices can then be defined as

$$\mathcal{P} = \left\{ \sum_{j=1}^m p_j \mu_j : \begin{array}{l} \boldsymbol{p} \in \Delta, \quad \mu_j \in \mathcal{M}_+ \quad \forall j = 1, \dots, m \\ \underline{\mathbf{\Omega}}_j \le \int_{\Xi_j} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mu_j(\mathrm{d}\boldsymbol{\xi}) \le \overline{\mathbf{\Omega}}_j \quad \forall j = 1, \dots, m \end{array} \right\}. \tag{23}$$

If we use the ambiguity set (23) instead of (7) in Sect. 2.3, the semi-infinite linear program (9) becomes

$$\begin{array}{lll} \sup\limits_{\boldsymbol{p}\in\Delta} & \sup\limits_{\mu_1,\ldots,\mu_m\in\mathcal{M}_+} & \sum\limits_{j=1}^m p_j \displaystyle\int_{\Xi_j} \max(0, L(\boldsymbol{x},\boldsymbol{\xi})-\beta)\ \mu_j(\mathrm{d}\boldsymbol{\xi}) \\ & \text{s.t.} & \underline{\boldsymbol{\Omega}}_j \le \displaystyle\int_{\Xi_j} \left[\boldsymbol{\xi}^\intercal\ 1\right]^\intercal \left[\boldsymbol{\xi}^\intercal\ 1\right] \mu_j(\mathrm{d}\boldsymbol{\xi}) \le \overline{\boldsymbol{\Omega}}_j \quad \forall j=1,\ldots,m. \end{array} \tag{24}$$

Dualizing the inner maximization problem yields

$$\begin{array}{lll} \sup\limits_{\boldsymbol{p}\in\Delta} & \inf & \sum\limits_{j=1}^m p_j \langle \overline{\boldsymbol{\Omega}}_j, \overline{\mathbf{M}}_j\rangle - \langle \underline{\boldsymbol{\Omega}}_j, \underline{\mathbf{M}}_j\rangle \\ & \text{s.t.} & \overline{\mathbf{M}}_j, \underline{\mathbf{M}}_j \in \mathbb{S}^{n+1}, \quad \overline{\mathbf{M}}_j,\ \underline{\mathbf{M}}_j \succeq \mathbf{0} \quad \forall j=1,\ldots,m \\ & & \left[\boldsymbol{\xi}^\intercal\ 1\right]\left(\overline{\mathbf{M}}_j - \underline{\mathbf{M}}_j\right)\left[\boldsymbol{\xi}^\intercal\ 1\right]^\intercal \ge \max(0, L(\boldsymbol{x},\boldsymbol{\xi})-\beta) \\ & & \qquad \forall \boldsymbol{\xi}\in\Xi_j,\ j=1,\ldots,m\,, \end{array} \tag{25}$$

and strong duality holds for (24) and (25) because each box $[\underline{\boldsymbol{\Omega}}_j, \overline{\boldsymbol{\Omega}}_j]$ contains a strictly positive semidefinite matrix [22]. To further simplify (25), we need the following technical proposition.

Proposition 3.1 *Define Δ as in (22). For any $\boldsymbol{\varphi}\in\mathbb{R}^m$, both the optimal value and a maximizer of the worst-case expectation problem* $\max_{\boldsymbol{p}\in\Delta} \boldsymbol{\varphi}^\intercal\boldsymbol{p}$ *can be obtained by solving the second-order cone program*

$$\begin{array}{ll} \max & \boldsymbol{\varphi}^\intercal \boldsymbol{p} \\ \text{s.t.} & \boldsymbol{p},\boldsymbol{r}\in\mathbb{R}^m_+, \quad \boldsymbol{e}^\intercal\boldsymbol{p}=1, \quad \boldsymbol{e}^\intercal\boldsymbol{r}\le\delta_{\boldsymbol{p}} \\ & \sqrt{(p_j-\hat{p}_j)^2+\frac{1}{4}p_j^2+r_j^2} \le \frac{1}{2}p_j + r_j \quad \forall j=1,\ldots,m\,. \end{array} \tag{26}$$

The optimal value can also be computed by solving the following second-order cone program dual to (26).

$$\begin{array}{ll} \min & \delta_{\boldsymbol{p}}\zeta - \eta - 2\hat{\boldsymbol{p}}^\intercal\boldsymbol{r} + 2\zeta\,\hat{\boldsymbol{p}}^\intercal\boldsymbol{e} \\ \text{s.t.} & \zeta\in\mathbb{R}_+, \quad \eta\in\mathbb{R}, \quad \boldsymbol{r},\boldsymbol{s}\in\mathbb{R}^m \\ & \varphi_j \le s_j, \quad s_j+\eta\le\zeta, \quad \sqrt{4r_j^2+(s_j+\mu)^2} \le 2\zeta - s_j - \eta \quad \forall j=1,\ldots,m \end{array} \tag{27}$$

Proof The claim follows immediately from [4, Theorem 4.1]. □

If we identify φ_j with the optimal value of the jth inner minimization problem in (25), then (25) can be re-expressed as $\max_{\boldsymbol{p}\in\Delta} \boldsymbol{\varphi}^\intercal\boldsymbol{p}$. By Proposition 3.1, (25) is thus equivalent to

$$\begin{array}{ll} \inf & \delta_{\boldsymbol{p}}\zeta - \eta - 2\hat{\boldsymbol{p}}^\intercal\boldsymbol{r} + 2\zeta\,\hat{\boldsymbol{p}}^\intercal\boldsymbol{e} \\ \text{s.t.} & \zeta\in\mathbb{R}_+,\ \eta\in\mathbb{R}, \quad \boldsymbol{r},\boldsymbol{s}\in\mathbb{R}^m, \quad \overline{\mathbf{M}}_j,\ \underline{\mathbf{M}}_j\in\mathbb{S}^{n+1},\ \overline{\mathbf{M}}_j,\ \underline{\mathbf{M}}_j\succeq\mathbf{0} \quad \forall j=1,\ldots,m \\ & s_j+\eta\le\zeta, \quad \sqrt{4r_j^2+(s_j+\eta)^2} \le 2\zeta - s_j - \eta \quad \forall j=1,\ldots,m \\ & \langle \overline{\boldsymbol{\Omega}}_j, \overline{\mathbf{M}}_j\rangle - \langle \underline{\boldsymbol{\Omega}}_j, \underline{\mathbf{M}}_j\rangle \le s_j \quad \forall j=1,\ldots,m \\ & \left[\boldsymbol{\xi}^\intercal\ 1\right]\left(\overline{\mathbf{M}}_j - \underline{\mathbf{M}}_j\right)\left[\boldsymbol{\xi}^\intercal\ 1\right]^\intercal \ge \max(0, L(\boldsymbol{x},\boldsymbol{\xi})-\beta) \quad \forall\boldsymbol{\xi}\in\Xi_j \quad \forall j=1,\ldots,m\,. \end{array} \tag{28}$$

By applying the full expansion and quadratic decision rule approaches of Sects. 2.3.1 and 2.3.2 to the semi-infinite constraints in (28), we obtain an exact SDP reformulation and a tractable conservative SDP approximation for (24), respectively. Substituting these SDPs into (8) then yields an exact SDP reformulation and a tractable conservative SDP approximation for the worst-case CVaR. For the sake of conciseness, we omit these routine calculations and merely state the SDP reformulation of the worst-case CVaR obtained from the full expansion approach.

$$
\begin{array}{ll}
\inf & \beta + \frac{1}{\epsilon}(\delta_{\boldsymbol{p}}\zeta - \eta - 2\hat{\boldsymbol{p}}^\intercal \boldsymbol{r} + 2\zeta\hat{\boldsymbol{p}}^\intercal \boldsymbol{e}) \\
\text{s.t.} & \zeta \in \mathbb{R}_+,\ \beta,\eta\in\mathbb{R},\ \boldsymbol{r},\boldsymbol{s}\in\mathbb{R}^m,\ \overline{\mathbf{M}}_j,\ \underline{\mathbf{M}}_j\in\mathbb{S}^{n+1},\ \overline{\mathbf{M}}_j,\ \underline{\mathbf{M}}_j\geq\mathbf{0}\quad\forall j=1,\ldots,m \\
& \alpha_j,\ \gamma_{jk}\in\mathbb{R}_+\quad\forall j=1,\ldots,m,\ k=1,\ldots,2^n \\
& s_j+\eta\leq\zeta,\quad \sqrt{4r_j^2+(s_j+\eta)^2}\leq 2\zeta-s_j-\eta\quad\forall j=1,\ldots,m \\
& \langle\overline{\mathbf{\Omega}}_j,\overline{\mathbf{M}}_j\rangle-\langle\underline{\mathbf{\Omega}}_j,\underline{\mathbf{M}}_j\rangle\leq s_j\quad\forall j=1,\ldots,m \\
& \overline{\mathbf{M}}_j-\underline{\mathbf{M}}_j+\alpha_j\mathbf{W}_j\succcurlyeq\mathbf{0}\quad\forall j=1,\ldots,m \\
& \overline{\mathbf{M}}_j-\underline{\mathbf{M}}_j+\gamma_{jk}\mathbf{W}_j\succcurlyeq\begin{bmatrix}\mathbf{0} & \frac{1}{2}(\boldsymbol{b}-\mathcal{I}_k(\boldsymbol{h})) \\ \frac{1}{2}(\boldsymbol{b}-\mathcal{I}_k(\boldsymbol{h}))^\intercal & (\boldsymbol{d}+\mathcal{I}_k(\boldsymbol{h}))^\intercal\boldsymbol{x}-\beta\end{bmatrix}\ \begin{array}{l}\forall j=1,\ldots,m\\ \forall k=1,\ldots,2^n\ .\end{array}
\end{array}
\tag{29}
$$

Remark 3.2 As in Sect. 6, we can construct an extremal distribution for the worst-case CVaR problem over the generalized ambiguity set (23). To this end, we set $\varphi_j = \langle\overline{\mathbf{\Omega}}_j,\overline{\mathbf{M}}_j^*\rangle - \langle\underline{\mathbf{\Omega}}_j,\underline{\mathbf{M}}_j^*\rangle$, $j = 1,\ldots,m$, where $\overline{\mathbf{M}}_j^*$ and $\underline{\mathbf{M}}_j^*$ are taken from an optimal solution of the SDP (29). Then, we calculate worst-case state probabilities $\boldsymbol{p}^* \in \operatorname{argmax}_{\boldsymbol{p}\in\Delta}\boldsymbol{\varphi}^\intercal\boldsymbol{p}$ by solving the second-order cone program (26). Finally, we can derive an extremal distribution for the inner maximization problem in (24) with $\boldsymbol{p} = \boldsymbol{p}^*$ by using the procedure outlined in Sect. 6 (with obvious minor modifications). By construction, this discrete distribution is extremal for the worst-case CVaR problem.

Remark 3.3 In the presence of data scarcity it may be difficult to estimate the statewise covariances for all pairs of products, and the newsvendor may only have access to first-order moment information. This situation is conveniently captured by an ambiguity set of the type (23), where the upper and lower bounds on the first order moments coincide, while the upper and lower bounds for all second-order moments are set to $+\infty$ and $-\infty$, respectively. Assuming for simplicity that the state probabilities are precisely known, one can show that the SDP reformulation of the worst-case CVaR then reduces to

$$
\begin{array}{ll}
\inf & \beta + \frac{1}{\epsilon}\sum_{j=1}^m p_j(y_j + \boldsymbol{\mu}_j^\intercal\boldsymbol{y}_j) \\
\text{s.t.} & \beta\in\mathbb{R},\quad \boldsymbol{y}_j\in\mathbb{R}^n,\quad y_j\in\mathbb{R},\quad \alpha_j,\ \gamma_{jk}\in\mathbb{R}_+\quad\forall j=1,\ldots,m,\ k=1,\ldots,2^n \\
& \begin{bmatrix}\mathbf{0} & \frac{1}{2}\boldsymbol{y}_j \\ \frac{1}{2}\boldsymbol{y}_j & y_j\end{bmatrix}+\alpha_j\mathbf{W}_j\succcurlyeq\mathbf{0}\quad\forall j=1,\ldots,m \\
& \begin{bmatrix}\mathbf{0} & \frac{1}{2}\boldsymbol{y}_j \\ \frac{1}{2}\boldsymbol{y}_j & y_j\end{bmatrix}+\gamma_{jk}\mathbf{W}_j\succcurlyeq\begin{bmatrix}\mathbf{0} & \frac{1}{2}(\boldsymbol{b}-\mathcal{I}_k(\boldsymbol{h})) \\ \frac{1}{2}(\boldsymbol{b}-\mathcal{I}_k(\boldsymbol{h}))^\intercal & (\boldsymbol{d}+\mathcal{I}_k(\boldsymbol{h}))^\intercal\boldsymbol{x}-\beta\end{bmatrix}\ \begin{array}{l}\forall j=1,\ldots,m\\ \forall k=1,\ldots,2^n\ .\end{array}
\end{array}
$$

In the extreme case when no distributional information beyond the support is available, the worst-case CVaR problem reduces to a classical robust optimization problem of the type studied in [2,5].

3.2 Partial expansion approach to improve the approximation quality

Recall that the quadratic decision rule approximation replaces each of the n max terms in (16) by a quadratic overestimator. A less conservative approximation is obtained by selecting $n_e < n$ products, expanding the corresponding n_e max terms and overestimating the remaining $n - n_e$ max terms by quadratic decision rules. There are $\binom{n}{n_e}$ possibilities to choose n_e out of n products. If n_e is kept fixed, this number scales polynomially with n. Each choice provides a different upper bound on the worst-case CVaR that can be computed by solving a tractable SDP. We refer to the best (lowest) of these upper bounds as the partial expansion bound. This bound can be computed in polynomial time, and its computation can be substantially accelerated by solving the $\binom{n}{n_e}$ individual SDPs in parallel.

One can show that in the absence of support constraints, at least $n_e = 2$ terms must be expanded for the partial expansion bound to improve on the quadratic decision rule bound of Sect. 2.3.2.

Remark 3.4 In order to improve the approximation quality one could also use higher order polynomials to overestimate the max functions in (16). In fact, it has been shown in [7] that one can optimize efficiently over certain classes of polynomial decision rules that admit a sum-of-squares decomposition.

4 Complexity analysis

In this section we formally prove that the distributionally robust multi-item newsvendor problem (4) is NP-hard and that the use of approximation techniques (e.g. based on quadratic decision rules) is therefore justified. Before we embark on the proof of the main result, we first show that optimizing the Manhattan norm (1-norm) over an ellipsoid is computationally intractable.

Lemma 4.1 *Maximizing the Manhattan norm over an ellipsoid of the form* (6) *is* NP-*hard.*

Remark 4.1 More rigorously, the lemma can be paraphrased as follows: Computing the maximum Manhattan norm over an ellipsoid of the form (6) to within accuracy ε cannot be achieved in time polynomial in the description of the ellipsoid and $\log \frac{1}{\varepsilon}$ unless P = NP.

Proof of Lemma 4.1 The claim follows from a hardness result for general matrix norm problems due to Steinberg [41]. Given $n \in \mathbb{N}$ and a positive definite matrix $\mathbf{A} \in \mathbb{S}^n$, it is NP-hard to compute

$$\|\mathbf{A}\|_{2,1} = \max_{\|\boldsymbol{\xi}\|_2 \leq 1} \|\mathbf{A}\boldsymbol{\xi}\|_1 ,$$

see [41, p. 22]. For any $n \in \mathbb{N}$ and $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{A} \succ 0$, we can construct an instance of the ellipsoid Ξ_1 by setting the parameters in (6) to $\boldsymbol{\nu}_1 = \mathbf{0}$, $\boldsymbol{\Lambda}_1 = \mathbf{AA}$ and $\delta_1 = 1$. By construction, we then have

$$\max_{\|\boldsymbol{\xi}\|_2 \leq 1} \|\mathbf{A}\boldsymbol{\xi}\|_1 = \max_{\|\mathbf{A}^{-1}\boldsymbol{\xi}\|_2 \leq 1} \|\boldsymbol{\xi}\|_1 = \max_{\boldsymbol{\xi}^\top(\mathbf{AA})^{-1}\boldsymbol{\xi} \leq 1} \|\boldsymbol{\xi}\|_1 = \max_{\boldsymbol{\xi} \in \Xi_1} \|\boldsymbol{\xi}\|_1 . \tag{30}$$

Thus, maximizing the Manhattan norm over Ξ_1 is equivalent to computing $\|\mathbf{A}\|_{2,1}$. We conclude that it is impossible to efficiently compute the maximum Manhattan norm over an ellipsoid unless P = NP. □

We are now ready to state the first main result of this section.

Theorem 4.1 *The worst-case CVaR problem* (8) *and the distributionally robust multi-item newsvendor problem* (4) *are* NP*-hard even for an ambiguity set of the type* (7) *with* $m = 1$.

Proof Our proof follows the reasoning of Delage [9, Proposition 3.5.4] and Bertsimas et al. [6, § 3.1]. By the results at the beginning of Sect. 2.3, we know that for $m = 1$ the worst-case CVaR problem (8) is equivalent to

$$\begin{aligned} \inf_{\beta \in \mathbb{R},\, \mathbf{M} \in \mathbb{S}^{n+1}} \quad & \beta + \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}_1, \mathbf{M} \rangle \\ \text{s.t.} \quad & \left[\boldsymbol{\xi}^\top \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \ 1\right]^\top \geq \max(0, \boldsymbol{d}^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{\xi} + \boldsymbol{h}^\top \max(\boldsymbol{x} - \boldsymbol{\xi}, \mathbf{0}) - \beta) \ \forall \boldsymbol{\xi} \in \Xi_1 . \end{aligned} \tag{31}$$

The separation version of this problem can be stated as follows.

Separation problem: Given $\mathbf{M}$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Lambda}$, $\boldsymbol{x}$, $\boldsymbol{d}$, $\boldsymbol{b}$, $\boldsymbol{h}$, $\boldsymbol{\nu}$, β, ϵ and δ_1, check whether the semi-infinite constraint in (31) is satisfied. If it is violated, then find $\beta \in \mathbb{R}$, $\mathbf{M} \in \mathbb{S}^{n+1}$ and $\boldsymbol{\xi} \in \Xi_1$ such that

$$\left[\boldsymbol{\xi}^\top \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \ 1\right]^\top < \max(0, \boldsymbol{d}^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{\xi} + \boldsymbol{h}^\top \max(\boldsymbol{x} - \boldsymbol{\xi}, \mathbf{0}) - \beta) .$$

Given an instance of the Manhattan norm maximization problem over an ellipsoid Ξ_1 of the form (6), we can construct an instance of the separation problem with parameters $\mathbf{M} = \mathbf{0}$, $\boldsymbol{x} = \mathbf{0}$, $\boldsymbol{d} = \mathbf{0}$, $\boldsymbol{b} = \boldsymbol{e}$, $\boldsymbol{h} = 2\boldsymbol{e}$ and with a support set that coincides with the ellipsoid supplied to the norm maximization problem. The resulting instance of the separation problem asks whether $\max_{\boldsymbol{\xi} \in \Xi_1} \|\boldsymbol{\xi}\|_1 \leq \beta$. If this inequality could be verified efficiently for arbitrary values of β, we could devise an efficient bisection algorithm to solve the norm maximization problem over Ξ_1. Thus, by Lemma 4.1 the separation problem is NP-hard. NP-hardness of the worst-case CVaR problem (8) follows now immediately from the equivalence of separation and optimization, see Grötschel et al. [20]. NP-hardness of the worst-case expectation problem and the distributionally robust multi-item newsvendor problem (4) can be shown in a similar way. Details are omitted for brevity of exposition. □

Next, we demonstrate that the distributionally robust newsvendor problem (4) remains NP-hard even in the absence of support constraints, that is, for $m = 1$ and $\Xi_1 = \mathbb{R}^n$. This case could be seen as less realistic because it allows for unbounded demands. However, our complexity results are easy to derive in view of the preparatory work above and may also be of interest outside the realm of newsvendor problems. Our proof will be based on a reduction of the NP-complete partition problem.

Partition Problem
Instance. Given is a vector $\boldsymbol{w} \in \mathbb{N}^n$ of positive integers.
Question. Is there a vector $\boldsymbol{\xi} \in \{-1, 1\}^n$ such that $\boldsymbol{w}^\top \boldsymbol{\xi} = 0$?

Given an instance of the partition problem with input $\boldsymbol{w} \in \mathbb{N}^n$, we set

$$\tau = \frac{3}{8\|\boldsymbol{w}\|_2^2}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{B} & \frac{\boldsymbol{w}}{2\|\boldsymbol{w}\|_2} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \frac{\boldsymbol{w}}{2\|\boldsymbol{w}\|_2} \end{bmatrix}^\top . \tag{32}$$

Here, the columns of the matrix $\mathbf{B} \in \mathbb{R}^{n\times(n-1)}$ consist of $(n-1)$ arbitrary *orthonormal* vectors orthogonal to $\boldsymbol{w}$. This matrix can be constructed efficiently, for example, via the Gram-Schmidt procedure.

We first prove a fundamental result about quadratic majorants of the Manhattan norm.

Theorem 4.2 *Checking whether a quadratic function $q(\boldsymbol{\xi})$ majorizes the Manhattan norm is* NP*-hard, i.e., there is no efficient algorithm to check whether $q(\boldsymbol{\xi}) \geq \|\boldsymbol{\xi}\|_1$ for all $\boldsymbol{\xi} \in \mathbb{R}^n$ unless* P = NP.

The proof of Theorem 4.2 will rely on the following technical lemma.

Lemma 4.2 *There exists a solution to the partition problem with input $\boldsymbol{w} \in \mathbb{N}^n$ if and only if there exists $\boldsymbol{\xi} \in \mathbb{R}^n$ such that $\boldsymbol{\xi}^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{\xi} + n - \tau < 2\|\boldsymbol{\xi}\|_1$, where τ and $\boldsymbol{\Lambda}$ are defined as in* (32).

Proof To establish the 'only if' part, suppose $\hat{\boldsymbol{\xi}} \in \{-1, 1\}^n$ with $\boldsymbol{w}^\top \hat{\boldsymbol{\xi}} = 0$ is a solution to the partition problem. Thus, we find

$$\hat{\boldsymbol{\xi}}^\top \boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{\xi}} + n - \tau = n + n - \tau < 2n = 2\|\hat{\boldsymbol{\xi}}\|_1 ,$$

where we use the fact that $\hat{\boldsymbol{\xi}}^\top \boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{\xi}} = n$ for all integer vectors $\hat{\boldsymbol{\xi}} \in \{-1, 1\}^n$ with $\boldsymbol{w}^\top \hat{\boldsymbol{\xi}} = 0$.

To establish the 'if' part, suppose there exists $\boldsymbol{\xi} \in \mathbb{R}^n$ with $\boldsymbol{\xi}^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{\xi} + n - \tau < 2\|\boldsymbol{\xi}\|_1$, implying that

$$\begin{aligned} 0 &> \min_{\boldsymbol{\xi}\in\mathbb{R}^n} \boldsymbol{\xi}^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{\xi} + n - \tau - 2 \max_{\boldsymbol{r}\in\{-1,1\}^n} \boldsymbol{r}^\top \boldsymbol{\xi} \\ &= \min_{\boldsymbol{\xi}\in\mathbb{R}^n,\ \boldsymbol{r}\in\{-1,1\}^n} \boldsymbol{\xi}^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{\xi} + n - \tau - 2\boldsymbol{r}^\top \boldsymbol{\xi} \\ &= \min_{\boldsymbol{r}\in\{-1,1\}^n} n - \tau - \boldsymbol{r}^\top \boldsymbol{\Lambda} \boldsymbol{r} . \end{aligned}$$

This is equivalent to asserting the existence of $\boldsymbol{r} \in \{-1, 1\}^n$ with $n - \tau < \boldsymbol{r}^\top \boldsymbol{\Lambda} \boldsymbol{r}$. By the construction of $\boldsymbol{\Lambda}$ in (32), we conclude that this $\boldsymbol{r}$ satisfies

$$\begin{aligned} n - \tau &< (\mathbf{B}_1^\top \boldsymbol{r})^2 + \cdots + (\mathbf{B}_{n-1}^\top \boldsymbol{r})^2 + \frac{1}{4\|\boldsymbol{w}\|_2^2}(\boldsymbol{w}^\top \boldsymbol{r})^2 \\ &= \|\boldsymbol{r}\|_2^2 - \frac{1}{\|\boldsymbol{w}\|_2^2}(\boldsymbol{w}^\top \boldsymbol{r})^2 + \frac{1}{4\|\boldsymbol{w}\|_2^2}(\boldsymbol{w}^\top \boldsymbol{r})^2 \\ &= n - \frac{3}{4\|\boldsymbol{w}\|_2^2}(\boldsymbol{w}^\top \boldsymbol{r})^2, \end{aligned}$$

where the first equality follows from Parserval's identity. As $\tau = \frac{3}{8\|\boldsymbol{w}\|_2^2}$, we have that $(\boldsymbol{w}^\top \boldsymbol{r})^2 < \frac{1}{2}$. Since both $\boldsymbol{w}$ and $\boldsymbol{r}$ are integer vectors, we may conclude that $\boldsymbol{w}^\top \boldsymbol{r} = 0$. Thus the claim follows. □

Proof of Theorem 4.2 Select an arbitrary $\boldsymbol{w} \in \mathbb{N}^n$ and construct $\boldsymbol{\Lambda}$ as in (32). By Lemma 4.2, $\boldsymbol{w}$ solves the partition problem iff the quadratic function $q(\boldsymbol{\xi}) = \frac{1}{2}(\boldsymbol{\xi}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\xi} + n - \tau)$ is *not* a majorant for the Manhattan norm. Thus, there can be no efficient algorithm to check whether a quadratic function majorizes the Manhattan norm unless P = NP. □

The second main result of this section is a corollary of Theorem 4.2.

Theorem 4.3 *The worst-case CVaR problem* (8) *and the distributionally robust multi-item newsvendor problem* (4) *are* NP-*hard even for an ambiguity set of the type* (7) *with $m = 1$ and $\Xi_1 = \mathbb{R}^n$.*

Proof The proof parallels that of Theorem 4.1. Details are omitted for brevity. □

5 Numerical results

All semidefinite programming problems arising in this section are solved with MOSEK by using the Yalmip [25] interface, while all stochastic programming problems are solved with CPLEX via the ILOG C++ interface. All experiments are run on a 3.4 GHz machine with 8 GB RAM.

5.1 Optimality gap and scalability

The first experiment is designed to assess the accuracy and the tractability of the quadratic decision rule and the partial expansion approximations as compared to the exact full expansion model. For each fixed number of products n, we generate 100 random instances of the distributionally robust newsvendor problem (4) with the ambiguity set (7). We solve these instances both exactly using the complete expansion approach, as well as approximately using the quadratic decision rules and the partial expansion approach with $n_e = 2$. All instances share the following global parameters: $m = 2$, $\lambda = 0.5$, $\epsilon = 5\,\%$, $p = 0.5$ and $\Xi_1 = \Xi_2 = \mathbb{R}^n$. We also fix the retail prices to 10 for each product, and we set the salvage values and stockout costs to 10 and 25 % of the corresponding selling prices, respectively.

For each fixed number of products n, the 100 random instances are constructed as follows. We sample the vector $\boldsymbol{c}$ uniformly from $[3, 8]^n$, thereby ensuring that the retail prices exceed the wholesale prices, which is a prerequisite for a profitable business operation. Similarly, the mean demand vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are sampled from independent uniform distributions on $[5, 100]^n$, while the vectors of standard deviations $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ of the demand in the two states of the world are sampled from independent uniform distributions on $[0.1\boldsymbol{\mu}_1, \boldsymbol{\mu}_1]$ and $[0.1\boldsymbol{\mu}_2, \boldsymbol{\mu}_2]$, respectively. Here, we tie the standard deviations in the two states to their respective mean values as we expect the demand variability to scale with the mean demand.

The correlations among different demands are assumed to be independent of the state. In order to generate the corresponding correlation matrix $\mathbf{C}$, we first sample a random matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ with independent standard normally distributed elements and set $\mathbf{C} = \text{diag}(\boldsymbol{u})\mathbf{U}\text{diag}(\boldsymbol{u})$, where $\mathbf{U} = \mathbf{S}^\intercal\mathbf{S}$ and $\boldsymbol{u}$ is the vector whose ith element is defined as $u_i = 1/\sqrt{U_{ii}}$. Therefore, the covariance matrices corresponding to the two modes of the demand distribution are given by $\boldsymbol{\Sigma}_1 = \text{diag}(\boldsymbol{\sigma}_1)\mathbf{C}\text{diag}(\boldsymbol{\sigma}_1)$ and $\boldsymbol{\Sigma}_2 = \text{diag}(\boldsymbol{\sigma}_2)\mathbf{C}\text{diag}(\boldsymbol{\sigma}_2)$, respectively.

We solve each instance indexed by $k = 1, \ldots, 100$ using the complete expansion, the quadratic decision rule and the partial expansion approaches, and we denote the corresponding optimal objective values by z_k^*, $\hat{z}_k$ and $\bar{z}_k$, respectively. The relative optimality gaps of the quadratic decision rule and partial expansion approximations are defined as $100\,\% \times \left|(\hat{z}_k - z_k^*)/z_k^*\right|$ and $100\,\% \times \left|(\bar{z}_k - z_k^*)/z_k^*\right|$. Figure 1 (left) displays the median as well as maximum and minimum optimality gaps over the 100 instances as we vary the number of products. We observe that the median gaps of the quadratic decision rule (partial expansion) approximation are uniformly bounded above by 4 % (2 %), while the maximum gaps never exceed 7 % (5 %). This means that both approximations achieve a level of accuracy that is acceptable for practical purposes as stochastic programs necessarily constitute imprecise models of reality that can never truthfully capture all physical, commercial, legal, etc. aspects of a real decision-making situation.

Figure 1 (right) shows a semi-log plot of the solver times for the 100 instances as we vary the number of products. As expected, the runtime of the complete expansion formulation scales exponentially with n and exceeds 6,000 s already for 15 products.

The runtime for the partial expansion approach scales more gracefully with n. It exceeds 6,000 s if there are more than 38 products. Finally, the quadratic decision rule

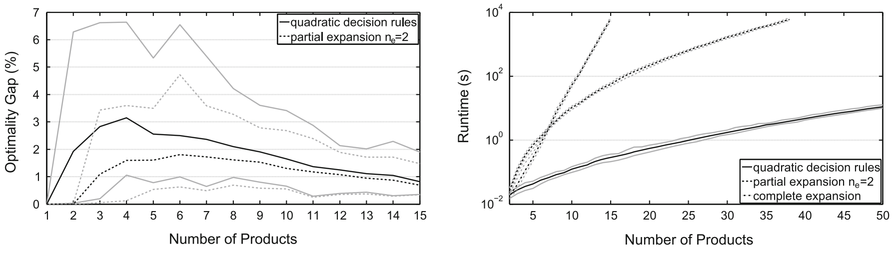

Fig. 1 Optimality gaps (*left*) and scalability properties (*right*) of the quadratic decision rule and partial expansion approximations as well as the complete expansion approach. *Gray lines* represent extreme values not considered to be outliers

approximation allows us to solve instances with up to 50 products in less than 12 s. We conclude that the quadratic decision rule approximation is conservative, achieves an acceptable level of accuracy and distinctly outperforms the exact full expansion model in terms of the problem sizes it can handle. Therefore, we will exclusively use the quadratic decision rule approximation to solve all distributionally robust newsvendor problems in the remainder of this section.

5.2 Stress test analysis

In this section we discuss the experimental results of a stylized newsvendor problem involving $n = 3$ products subject to a bimodal demand distribution. The main insights gained from this study extend to more general models with several products and multimodal demand distributions. We set the retail price and the wholesale price of each product to 10 and 5, respectively. We also set the salvage value and the stockout cost to 1 and 2.5, respectively. Finally, we set the parameters encoding the newsvendor's risk preferences to $\epsilon = 5\,\%$ and $\lambda = 0.8$.

We denote by $\bar{\mathbb{P}}$ the newsvendor's best estimate of the data-generating distribution, which may be derived from historical demand data or from expert knowledge. In our experiments we assume that $\bar{\mathbb{P}}$ is an equiprobable mixture of two truncated normal distributions $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \rho_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \rho_2)$. Here, we assume that $\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \rho_j)$ is obtained by truncating the normal distribution $\mathcal{N}(\boldsymbol{\mu}_j, \frac{F_n(\rho_j^2)}{F_{n+2}(\rho_j^2)}\boldsymbol{\Sigma}_j)$ outside of the ellipsoid $\{\boldsymbol{\xi} : (\boldsymbol{\xi} - \boldsymbol{\mu}_j)^\intercal \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\xi} - \boldsymbol{\mu}_j) \leq \frac{F_n(\rho_j^2)}{F_{n+2}(\rho_j^2)}\rho_j^2\}$, where F_k, $(k \in \mathbb{N})$, denotes the cumulative distribution function of a χ^2-distribution with k degrees of freedom. By construction, $\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \rho_j)$ has mean value $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$ for each $j = 1, 2$ [43]. We select the statewise means and variances as shown in Table 1, where $\mu_+ = 30$, $\mu_- = 15$ and $\sigma = 5$. We further assume that the statewise demands are pairwise correlated with correlation coefficients of 50 %. Finally, we set $\rho_1 = \rho_2 = \sqrt{F_n^{-1}(99\,\%)}$, which means that $\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \rho_j)$ is obtained by truncating only 1 % of the mass of the normal distribution $\mathcal{N}(\boldsymbol{\mu}_j, \frac{F_n(\rho_j^2)}{F_{n+2}(\rho_j^2)}\boldsymbol{\Sigma}_j)$, $j = 1, 2$. Observe that product 1 is unpopular (has low demand) and product 3 is popular (has high demand) in state 1, while product 3 is unpopular and product 1 is popular in state 2. In contrast, the popularity of product 2 is independent of the state.

Table 1 Marginal moments of the product demands

		Product		
State	Parameter	1	2	3
1	Mean	μ_-	$\frac{1}{2}(\mu_- + \mu_+)$	μ_+
	Variance	σ^2	σ^2	σ^2
2	Mean	μ_+	$\frac{1}{2}(\mu_- + \mu_+)$	μ_-
	Variance	σ^2	σ^2	σ^2

An ambiguity-neutral newsvendor solves the stochastic program (1) using $\bar{\mathbb{P}}$, while an ambiguity-averse newsvendor hedges against the possibility that the true data-generating distribution may deviate from $\bar{\mathbb{P}}$ and thus solves the distributionally robust problem (4) using an ambiguity set $\bar{\mathcal{P}}$ that contains $\bar{\mathbb{P}}$. Our experiments are based on the ambiguity set of Sect. 3.1 that caters for two layers of robustness. We set $m = 2$, $\hat{p}_1 = \hat{p}_2 = 0.5$, $\delta_{\boldsymbol{p}} = 0$ and

$$\overline{\boldsymbol{\Omega}}_j = \begin{bmatrix} (1+\tau)^2(\boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^\intercal) & (1+\tau)\boldsymbol{\mu}_j \\ (1+\tau)\boldsymbol{\mu}_j^\intercal & 1 \end{bmatrix},$$

$$\underline{\boldsymbol{\Omega}}_j = \begin{bmatrix} (1-\tau)^2(\boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^\intercal) & (1-\tau)\boldsymbol{\mu}_j \\ (1-\tau)\boldsymbol{\mu}_j^\intercal & 1 \end{bmatrix} \forall j = 1, 2\,,$$

where $\tau \in [0, 1]$ quantifies the level of moment uncertainty. The parameters characterizing the ellipsoidal support Ξ_j are set to $\boldsymbol{\nu}_j = \boldsymbol{\mu}_j$, $\boldsymbol{\Lambda}_j = \boldsymbol{\Sigma}_j$ and $\delta_j = \rho_j\sqrt{F_n(\rho_j^2)/F_{n+2}(\rho_j^2)}$ for $j = 1, 2$.

We can also envision a naive newsvendor who mistakenly believes that the demand distribution is unimodal or decides to approximate the true bimodal demand distribution by a unimodal one. We denote the naive newsvendor's best estimate of the data-generating distribution by $\hat{\mathbb{P}}$. In our experiments we will assume that $\hat{\mathbb{P}}$ follows a truncated normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho)$ with $\boldsymbol{\mu} = \sum_{j=1}^{2} p_j\boldsymbol{\mu}_j$, $\boldsymbol{\Sigma} = \sum_{j=1}^{2} p_j(\boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j\boldsymbol{\mu}_j^\intercal) - \boldsymbol{\mu}\boldsymbol{\mu}^\intercal$ and $\rho = \sqrt{F_n^{-1}(99\,\%)}$. By construction, $\bar{\mathbb{P}}$ and $\hat{\mathbb{P}}$ thus share the same (unconditional) first- and second-order moments.

An ambiguity-neutral naive newsvendor solves the stochastic program (1) using $\hat{\mathbb{P}}$, while an ambiguity-averse naive newsvendor solves the distributionally robust problem (4) using an ambiguity set $\hat{\mathcal{P}}$ that contains $\hat{\mathbb{P}}$. In our experiments we use an ambiguity set of the type considered in Sect. 3.1 with $m = 1$, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$. All other parameters are chosen as in the case of $\bar{\mathcal{P}}$. Throughout this section we assume that all moments used in the construction of ambiguity sets are free of estimation errors, that is, they constitute true parameters.

We now compare the robustness of the order portfolios obtained from the stochastic program (1) and the distributionally robust optimization problem (4) by stress testing them under a contaminated demand distribution [11,12]. Thus, we assess how different order portfolios perform under a data-generating distribution of the form $\mathbb{P}(\upsilon) = (1-\upsilon)\bar{\mathbb{P}} + \upsilon\mathbb{P}^*$, where $\mathbb{P}^*$ represents a contaminant of $\bar{\mathbb{P}}$ and $\upsilon \in [0, 1]$ encodes the contamination level. A meaningful choice for $\mathbb{P}^*$ is the extremal distribution of the worst-case CVaR problem (8) for an order portfolio that is optimal in (1), see also [6, § 4.1.1].

The step-by-step procedure for the stress test analysis is summarized as follows. We solve the stochastic program (1) using the sample average approximation [38] with $N = 50{,}000$ samples drawn from $\bar{\mathbb{P}}$ ($\hat{\mathbb{P}}$) to obtain the optimal stochastic order portfolio $\bar{\boldsymbol{x}}_s$ ($\hat{\boldsymbol{x}}_s$). We also solve the distributionally robust optimization problem (4) under the ambiguity set $\bar{\mathcal{P}}$ ($\hat{\mathcal{P}}$) to obtain the optimal robust order portfolio $\bar{\boldsymbol{x}}_r$ ($\hat{\boldsymbol{x}}_r$). Next, we construct $\mathbb{P}^*$ as an extremal distribution of the worst-case CVaR problem

(8) with $\mathcal{P} = \bar{\mathcal{P}}$ and $\boldsymbol{x} = \bar{\boldsymbol{x}}_s$. Finally, we evaluate the quality of the order portfolios $\bar{\boldsymbol{x}}_s$, $\hat{\boldsymbol{x}}_s$, $\bar{\boldsymbol{x}}_r$ and $\hat{\boldsymbol{x}}_r$ under the contaminated distribution $\mathbb{P}(\upsilon) = (1-\upsilon)\bar{\mathbb{P}} + \upsilon\mathbb{P}^*$ for different contamination levels $\upsilon \in [0, 1]$.

The left panels of Fig. 2 display the mean-CVaR values of the order portfolios $\bar{\boldsymbol{x}}_s$ (stochastic bimodal), $\hat{\boldsymbol{x}}_s$ (stochastic unimodal), $\bar{\boldsymbol{x}}_r$ (robust bimodal) and $\hat{\boldsymbol{x}}_r$ (robust unimodal) under $\mathbb{P}(\upsilon)$ as we vary the contamination level $\upsilon \in [0, 1]$, while the right panels visualize the advantage (in terms of relative mean-CVaR values) of the distributionally robust portfolios $\bar{\boldsymbol{x}}_r$ (bimodal) and $\hat{\boldsymbol{x}}_r$ (unimodal) over the respective stochastic portfolios $\bar{\boldsymbol{x}}_s$ and $\hat{\boldsymbol{x}}_s$. The results are shown for different levels of moment uncertainty, that is, for $\tau = 0$ (top), $\tau = 0.1$ (middle) and $\tau = 0.2$ (bottom).

For small contamination levels the data-generating distribution $\mathbb{P}(\upsilon)$ is well approximated by the best estimate $\bar{\mathbb{P}}$. Thus, the stochastic order portfolios have a small advantage over the robust ones. As the contamination levels increase, both the ambiguity-neutral and the ambiguity-averse newsvendors face higher risk-adjusted losses. However, the losses associated with the stochastic portfolios increase more rapidly, with the consequence that the robust portfolios clearly dominate the stochastic ones for a wide spectrum of contamination levels. In the absence of moment uncertainty ($\tau = 0$), the robust portfolios start to win at a contamination level of $\upsilon = 20\,\%$. For a medium (high) level of moment uncertainty $\tau = 0.1$

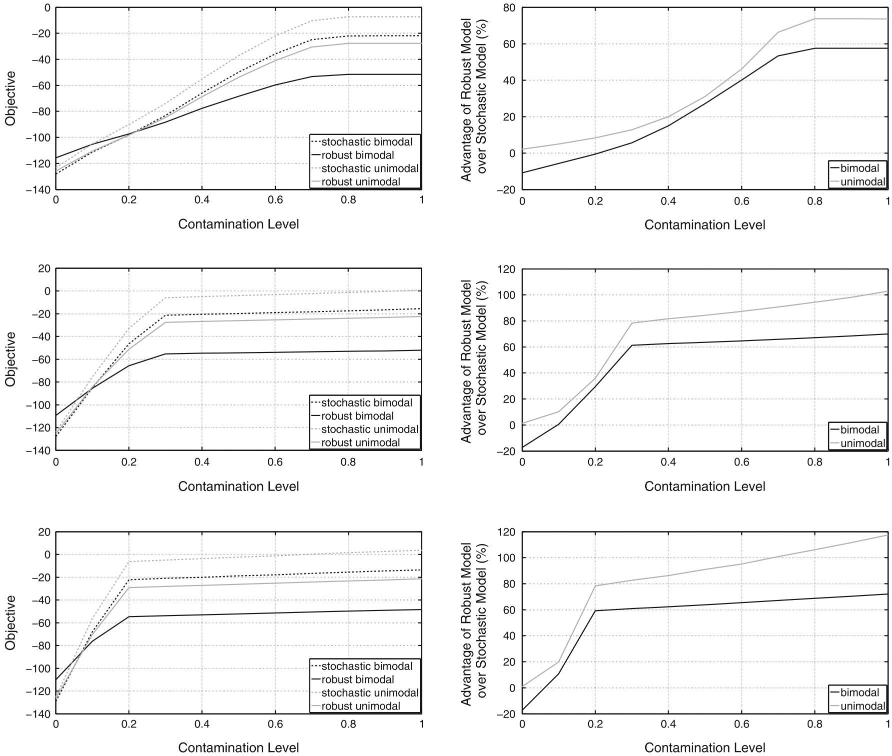

Fig. 2 Mean-CVaR objectives of stochastic and robust models (*left*) and percentage advantage of robust model over stochastic model (*right*) for different levels of moment uncertainty: $\tau = 0$ (*top*), $\tau = 0.1$ (*middle*), and $\tau = 0.2$ (*bottom*)

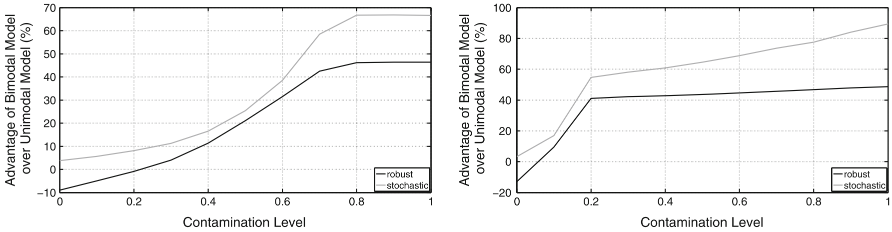

Fig. 3 Percentage advantage of bimodal model over unimodal model for $\tau = 0$ (*left*) and $\tau = 0.2$ (*right*)

($\tau = 0.2$), the robust portfolios already win for a contamination level as low as 10 % (5 %).

The advantage of the robust portfolios is more dramatic if the newsvendor is unaware of (or ignores) the bimodal structure of the demand distribution. In fact, Fig. 2 indicates that the ambiguity-averse naive newsvendor is better off than the ambiguity-neutral naive newsvendor for *all* contamination levels and irrespective of the degree of moment uncertainty. Thus, adopting a distributionally robust approach can compensate to a certain extent for the lack of modality information. Figure 3 provides further insights into the value of modality information. It shows the relative advantage of the bimodal portfolios $\bar{\boldsymbol{x}}_r$ (robust) and $\bar{\boldsymbol{x}}_s$ (stochastic) over the respective unimodal portfolios $\hat{\boldsymbol{x}}_r$ and $\hat{\boldsymbol{x}}_s$ for different levels of moment uncertainty. The models that correctly account for the bimodality outperform their unimodal counterparts for most contamination levels. We remark that modality information becomes more valuable in the presence of moment uncertainty, which we are actually likely to encounter in reality.

We ran the experiments of this section also for different values of the nominal state probabilities $\hat{\boldsymbol{p}}$ and uncertainty levels $\delta_{\boldsymbol{p}}$. The results of these additional experiments are qualitatively in agreement with those reported in this section. Further details are omitted for brevity of exposition.

Acknowledgments We are indebted to Karthik Natarajan for valuable discussions on the topic of this paper. We also thank two anonymous reviewers whose comments led to substantial improvements of this paper. This research was supported by EPSRC under Grant EP/I014640/1.

6 Appendix: Extremal distributions

In order to facilitate targeted stress tests for different order portfolios in Sect. 5, we describe here a method to construct extremal distributions for the worst-case CVaR problem (8). The general recipe is to first solve the SDP (18) and to substitute the resulting optimal β^* into (17). An extremal distribution that achieves the supremum in (8) can then be inferred from the dual solution of the SDP (17) by adapting a procedure for generic worst-case expectation problems with piecewise linear objective functions and without support information; see [6, 10]. We extend this procedure so that it can handle support information. Our construction of the extremal distribution will rely on the following corollary of [45, Section 2.1].

Proposition 6.1 *Let* $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^n : [\boldsymbol{\xi}^\top\ 1]\mathbf{W}[\boldsymbol{\xi}^\top\ 1]^\top \le 0\}$ *be the ellipsoid specified by the matrix*

$$\mathbf{W} = \begin{bmatrix} \mathbf{\Lambda}^{-1} & -\mathbf{\Lambda}^{-1}\boldsymbol{\nu} \\ -(\mathbf{\Lambda}^{-1}\boldsymbol{\nu})^{\top} & \boldsymbol{\nu}^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{\nu} - \delta^2 \end{bmatrix}$$

with $\mathbf{\Lambda} \in \mathbb{S}^n$, $\mathbf{\Lambda} \succ \mathbf{0}$, $\boldsymbol{\nu} \in \mathbb{R}^n$ *and* $\delta \in \mathbb{R}$. *If*

$$\begin{bmatrix} \mathbf{Z} & z \\ z^{\top} & 1 \end{bmatrix} \succcurlyeq \mathbf{0} \quad \text{and} \quad \left\langle \mathbf{W}, \begin{bmatrix} \mathbf{Z} & z \\ z^{\top} & 1 \end{bmatrix} \right\rangle \leq 0\,,$$

then one can construct a discrete distribution supported on Ξ *with mean* z *and covariance matrix* $\mathbf{Z} - zz^{\top}$.

Proof This proposition strenghtens a result proved in [45, Section 2.1] in the special case when Ξ is an ellipsoid. The covariance matrix of the discrete distribution constructed in [45] is only upper bounded (in a positive semidefinite sense) by but not necessarily equal to $\mathbf{Z} - zz^{\top}$.

By using an eigenvalue decomposition, we can factorize $\mathbf{Z} - zz^{\top}$ as $\sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{w}_i^{\top}$. Thus, we have

$$\left\langle \mathbf{W}, \begin{bmatrix} \mathbf{Z} & z \\ z^{\top} & 1 \end{bmatrix} \right\rangle \leq 0 \Longleftrightarrow z^{\top}\mathbf{\Lambda}^{-1}z + \sum_{i=1}^{n} \boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{w}_i - 2\boldsymbol{\nu}\mathbf{\Lambda}^{-1}z + \boldsymbol{\nu}^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{\nu} - \delta^2 \leq 0\,. \tag{33}$$

Assume first that $\boldsymbol{w}_i = \mathbf{0}$ for all $i = 1, \ldots, n$. Then, (33) implies that $z \in \Xi$. The Dirac distribution that concentrates unit mass at z is thus supported on Ξ, and it has mean z and covariance matrix $\mathbf{Z} - zz^{\top} = \mathbf{0}$. Assume now that $\boldsymbol{w}_i \neq \mathbf{0}$ for all $i \in \mathcal{N}$, where $\mathcal{N}$ is a nonempty subset of $\{1, \ldots, n\}$. As $\mathbf{\Lambda}^{-1} \succ \mathbf{0}$, we have $\boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{w}_i > 0$ if $i \in \mathcal{N}$; $= 0$ otherwise. Next, we define the constants

$$\begin{aligned} s &= -(z^{\top}\mathbf{\Lambda}^{-1}z + \sum_{i\in\mathcal{N}} \boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{w}_i - 2\boldsymbol{\nu}\mathbf{\Lambda}^{-1}z + \boldsymbol{\nu}^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{\nu} - \delta^2) \geq 0\,, \\ t &= -\sum_{i\in\mathcal{N}} \boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{w}_i < 0\,. \end{aligned}$$

For each $i \in \mathcal{N}$ we construct the following univariate quadratic function in ω.

$$\begin{aligned} f_i(\omega) &= \left\langle \mathbf{W}, \begin{bmatrix} (z+\omega\boldsymbol{w}_i)(z+\omega\boldsymbol{w}_i)^{\top} & z+\omega\boldsymbol{w}_i \\ (z+\omega\boldsymbol{w}_i)^{\top} & 1 \end{bmatrix} \right\rangle + s \\ &= (z+\omega\boldsymbol{w}_i)^{\top}\mathbf{\Lambda}^{-1}(z+\omega\boldsymbol{w}_i) - 2\boldsymbol{\nu}^{\top}\mathbf{\Lambda}^{-1}(z+\omega\boldsymbol{w}_i) + \boldsymbol{\nu}^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{\nu} - \delta^2 + s \\ &= \boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{w}_i\omega^2 + 2\boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}(z-\boldsymbol{\nu})\omega + z^{\top}\mathbf{\Lambda}^{-1}z - 2\boldsymbol{\nu}\mathbf{\Lambda}^{-1}z + \boldsymbol{\nu}^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{\nu} - \delta^2 + s \\ &= \boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}\boldsymbol{w}_i\omega^2 + 2\boldsymbol{w}_i^{\top}\mathbf{\Lambda}^{-1}(z-\boldsymbol{\nu})\omega + t \end{aligned}$$

Since $\boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i > 0$ and $t < 0$, the quadratic function $f_i(\omega)$ has two real roots $\omega_i > 0$ and $\omega_{i+n} < 0$. By construction, as $f_i(\omega_i) = f_i(\omega_{i+n}) = 0$ and $s \geq 0$, the points $z + \omega_i \boldsymbol{w}_i$ and $z + \omega_{i+n} \boldsymbol{w}_i$ are contained in Ξ. This allows us to define a discrete probability distribution $\mathbb{P}$ supported on Ξ with $2|\mathcal{N}|$ atoms at $z + \omega_i \boldsymbol{w}_i$ and $z + \omega_{i+n} \boldsymbol{w}_i$, $i \in \mathcal{N}$, and with

$$\left.\begin{aligned} \mathbb{P}(\{z + \omega_i \boldsymbol{w}_i\}) &= \kappa_i = \frac{\boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i}{(1 - \omega_i/\omega_{i+n}) \sum_{i' \in \mathcal{N}} \boldsymbol{w}_{i'}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_{i'}} \\ \mathbb{P}(\{z + \omega_{i+n} \boldsymbol{w}_i\}) &= \kappa_{i+n} = -\kappa_i \tfrac{\omega_i}{\omega_{i+n}} = \frac{\boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i}{(1 - \omega_{i+n}/\omega_i) \sum_{i' \in \mathcal{N}} \boldsymbol{w}_{i'}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_{i'}} \end{aligned}\right\} \forall i \in \mathcal{N}.$$

Since $\boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i > 0$ and $\omega_i/\omega_{i+n} < 0$, we have $\kappa_i > 0$ for all $i \in \mathcal{N}$. Next, observe that

$$\sum_{i \in \mathcal{N}} \kappa_i + \kappa_{i+n} = \sum_{i \in \mathcal{N}} \frac{[(1 - \omega_{i+n}/\omega_i) + (1 - \omega_i/\omega_{i+n})] \boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i}{(1 - \omega_i/\omega_{i+n})(1 - \omega_{i+n}/\omega_i) \sum_{i' \in \mathcal{N}} \boldsymbol{w}_{i'}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_{i'}} = 1\,.$$

Thus $\mathbb{P}$ constitutes a valid discrete probability distribution. Furthermore, the first- and second-order moments of $\mathbb{P}$ can be expressed as

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] &= \sum_{i \in \mathcal{N}} \kappa_i (z + \omega_i \boldsymbol{w}_i) + \kappa_{i+n}(z + \omega_{i+n} \boldsymbol{w}_i) = z + \sum_{i \in \mathcal{N}} (\kappa_i \omega_i + \kappa_{i+n} \omega_{i+n}) \boldsymbol{w}_i\,, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}^\top] &= z z^\top + \sum_{i \in \mathcal{N}} (\kappa_i \omega_i + \kappa_{i+n} \omega_{i+n})(z \boldsymbol{w}_i^\top + \boldsymbol{w}_i z^\top) + \sum_{i \in \mathcal{N}} (\kappa_i \omega_i^2 + \kappa_{i+n} \omega_{i+n}^2) \boldsymbol{w}_i \boldsymbol{w}_i^\top. \end{aligned}$$

For all $i \in \mathcal{N}$ we have

$$\kappa_i \omega_i + \kappa_{i+n} \omega_{i+n} = \kappa_i \omega_i - \kappa_i (\omega_i/\omega_{i+n}) \omega_{i+n} = 0$$

and

$$\begin{aligned} \kappa_i \omega_i^2 + \kappa_{i+n} \omega_{i+n}^2 &= \frac{\boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i}{\sum_{i' \in \mathcal{N}} \boldsymbol{w}_{i'}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_{i'}} \left(\frac{\omega_i^2 \omega_{i+n}}{\omega_{i+n} - \omega_i} + \frac{\omega_i \omega_{i+n}^2}{\omega_i - \omega_{i+n}} \right) \\ &= \frac{t}{\sum_{i' \in \mathcal{N}} \boldsymbol{w}_{i'}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_{i'}} \left(\frac{\omega_i}{\omega_{i+n} - \omega_i} - \frac{\omega_{i+n}}{\omega_{i+n} - \omega_i} \right) \\ &= \frac{-t}{\sum_{i' \in \mathcal{N}} \boldsymbol{w}_{i'}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_{i'}} = 1\,, \end{aligned}$$

where the second equality follows from the relation $\omega_i \omega_{i+n} = t/(\boldsymbol{w}_i^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{w}_i)$. Thus, $\mathbb{P}$ has all the desired properties, that is, $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = z$, $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}^\top] = \mathbf{Z}$ and $\mathbb{P}(\tilde{\boldsymbol{\xi}} \in \Xi) = 1$. $\square$

To avoid clutter in the subsequent derivation, we introduce the notational shorthands

$$\boldsymbol{w}_0 = \mathbf{0}, \quad w_0 = 0, \quad \boldsymbol{w}_k = \boldsymbol{b} - \mathcal{I}_k(\boldsymbol{h}), \quad w_k = (\boldsymbol{d} + \mathcal{I}_k(\boldsymbol{h}))^\top \boldsymbol{x} - \beta^* \quad \forall k = 1, \ldots, 2^n .$$

Using these definitions, we can replace $\max(0, L(\boldsymbol{x}, \boldsymbol{\xi}) - \beta^*)$ with $\max_{k=0,\ldots,2^n} \boldsymbol{w}_k^\top \boldsymbol{\xi} + w_k$, which facilitates a streamlined reformulation of the SDP (17).

$$\begin{array}{ll} \inf & \displaystyle\sum_{j=1}^{m} p_j \langle \boldsymbol{\Omega}_j, \mathbf{M}_j \rangle \\ \text{s.t.} & \mathbf{M}_j \in \mathbb{S}^{n+1}, \quad \gamma_{jk} \in \mathbb{R}_+ \quad \forall j = 1, \ldots, m, \ k = 0, \ldots, 2^n \\ & \mathbf{M}_j + \gamma_{jk} \mathbf{W}_j \succcurlyeq \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{w}_k \\ \frac{1}{2}\boldsymbol{w}_k^\top & w_k \end{bmatrix} \quad \begin{array}{l} \forall j = 1, \ldots, m \\ \forall k = 0, \ldots, 2^n \end{array} \end{array}$$

The dual of the above problem is given by

$$\begin{array}{ll} \sup & \displaystyle\sum_{j=1}^{m} \sum_{k=0}^{2^n} \boldsymbol{w}_k^\top \boldsymbol{z}_{jk} + w_k z_{jk} \\ \text{s.t.} & z_{jk} \in \mathbb{R}, \quad \boldsymbol{z}_{jk} \in \mathbb{R}^n, \quad \mathbf{Z}_{jk} \in \mathbb{S}^n \quad \forall j = 1, \ldots, m, \ k = 0, \ldots, 2^n \\ & \displaystyle\sum_{k=0}^{2^n} \begin{bmatrix} \mathbf{Z}_{jk} & \boldsymbol{z}_{jk} \\ \boldsymbol{z}_{jk}^\top & z_{jk} \end{bmatrix} = p_j \boldsymbol{\Omega}_j \quad \forall j = 1, \ldots, m \\ & \left\langle \mathbf{W}_j, \begin{bmatrix} \mathbf{Z}_{jk} & \boldsymbol{z}_{jk} \\ \boldsymbol{z}_{jk}^\top & z_{jk} \end{bmatrix} \right\rangle \le 0, \quad \begin{bmatrix} \mathbf{Z}_{jk} & \boldsymbol{z}_{jk} \\ \boldsymbol{z}_{jk}^\top & z_{jk} \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall j = 1, \ldots, m, \ k = 0, \ldots, 2^n . \end{array} \tag{34}$$

Strong duality holds as the primal minimization problem is strictly feasible. This implies that the optimal value of (34) is equal to that of the worst-case expectation problem (9). We now denote by $\mathbf{Z}_{jk}^*$, $\boldsymbol{z}_{jk}^*$, and z_{jk}^* for $j = 1, \ldots, m$ and $k = 0, \ldots, 2^n$ an optimal solution of the SDP (34). Assume first that all z_{jk}^* are strictly positive. For any fixed j and k we thus construct the positive semidefinite matrix

$$\begin{bmatrix} \mathbf{Z}_{jk}^*/z_{jk}^* & \boldsymbol{z}_{jk}^*/z_{jk}^* \\ (\boldsymbol{z}_{jk}^*/z_{jk}^*)^\top & 1 \end{bmatrix} \succcurlyeq \mathbf{0},$$

which satisfies, together with $\mathbf{W}_j$, the conditions of Proposition 6.1. Hence, we can construct a discrete distribution $\mathbb{P}_{jk}^*$ supported in Ξ_j with mean $\boldsymbol{z}_{jk}^*/z_{jk}^*$ and covariance matrix $\mathbf{Z}_{jk}^*/z_{jk}^* - \boldsymbol{z}_{jk}^*/z_{jk}^*(\boldsymbol{z}_{jk}^*/z_{jk}^*)^\top$. In the remainder we will argue that $\mathbb{P}^* = \sum_{j=1}^{m} \sum_{k=0}^{2^n} z_{jk}^* \mathbb{P}_{jk}^*$ is the sought extremal distribution.

From the equality constraints in (34) it is clear that $\mathbb{P}^*$ is feasible in (9). Thus,

$$\begin{aligned} \sum_{j=1}^{m} \sum_{k=0}^{2^n} \boldsymbol{w}_k^\top \boldsymbol{z}_{jk}^* + w_k z_{jk}^* &= \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\max_{k=0,\ldots,2^n} \boldsymbol{w}_k^\top \tilde{\boldsymbol{\xi}} + w_k \right] \\ &\ge \mathbb{E}_{\mathbb{P}^*} \left[\max_{k=0,\ldots,2^n} \boldsymbol{w}_k^\top \tilde{\boldsymbol{\xi}} + w_k \right] \end{aligned}$$

$$= \sum_{j=1}^{m} \sum_{\ell=0}^{2^n} z_{j\ell}^* \mathbb{E}_{\mathbb{P}_{j\ell}^*} \left[\max_{k=0,\ldots,2^n} \boldsymbol{w}_k^\mathsf{T} \tilde{\boldsymbol{\xi}} + w_k \right]$$

$$\geq \sum_{j=1}^{m} \sum_{k=0}^{2^n} z_{jk}^* \mathbb{E}_{\mathbb{P}_{jk}^*} [\boldsymbol{w}_k^\mathsf{T} \tilde{\boldsymbol{\xi}} + w_k] = \sum_{j=1}^{m} \sum_{k=0}^{2^n} \boldsymbol{w}_k^\mathsf{T} \boldsymbol{z}_{jk}^* + w_k z_{jk}^*,$$

where the first equality follows from strong duality. As all inequalities in the above expression must be binding, $\mathbb{P}^*$ is indeed extremal in (9).

Assume now that $z_{jk}^* = 0$ for some j and k. The positive semidefiniteness constraint in (34) then implies that $\boldsymbol{z}_{jk}^* = \mathbf{0}$, while the trace constraint implies $\langle \boldsymbol{\Lambda}_j^{-1}, \mathbf{Z}_{jk} \rangle \leq 0$, see also the definition of $\mathbf{W}_j$ in (13). Our assumption $\boldsymbol{\Lambda}_j^{-1} \succ \mathbf{0}$ thus implies that $\mathbf{Z}_{jk}^* = \mathbf{0}$. This argument suggests that all pairs of indices j and k for which $z_{jk}^* = 0$ can be ignored altogether. In general, an extremal distribution for the worst-case CVaR problem (8) can thus be constructed as

$$\mathbb{P}^* = \sum_{j=1}^{m} \sum_{\substack{k=0 \\ z_{jk}^* \neq 0}}^{2^n} z_{jk}^* \mathbb{P}_{jk}^* \,.$$

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Financ. **9**(3), 203–228 (1999)
2. Bandi, C., Bertsimas, D.: Tractable stochastic analysis in high dimensions via robust optimization. Math. Program. **134**(1), 23–70 (2012)
3. Bartezzaghi, E., Verganti, R., Zotteri, G.: Measuring the impact of asymmetric demand distributions on inventories. Int. J. Prod. Econ. **60–61**, 395–404 (1999)
4. Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., Rennen, G.: Robust solutions of optimization problems affected by uncertain probabilities. Manag. Sci. **59**(2), 341–357 (2012)
5. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
6. Bertsimas, D., Doan, X.V., Natarajan, K., Teo, C.-P.: Models for minimax stochastic linear optimization problems with risk aversion. Math. Oper. Res. **35**(3), 580–602 (2010)
7. Bertsimas, D., Iancu, D., Parrilo, P.: A hierarchy of near-optimal policies for multistage adaptive optimization. IEEE Trans. Automat. Contr. **56**(12), 2809–2824 (2011)
8. Choi, S., Ruszczyński, A., Zhao, Y.: A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. Oper. Res. **59**(2), 346–364 (2011)
9. Delage, E.: Distributionally Robust Optimization in Context of Data-driven Problems. PhD thesis, Stanford University, Palo Alto, CA (2009)
10. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**, 595–612 (2010)
11. Dupačová, J.: Stress testing via contamination. In: Marti, K., et al. (eds.) Coping with Uncertainty: Modeling and Policy Issues, Lecture Notes in Economics and Mathematical Systems, vol. 58, pp. 29–46. Springer, Berlin (2006)
12. Dupačová, J., Kopa, M.: Robustness in stochastic programs with risk constraints. Ann. Oper. Res. **200**(1), 55–74 (2012)
13. Dupačová, J.: The minimax approach to stochastic programming and an illustrative approach. Stochastics **20**(1), 73–88 (1987)

14. Dupačová, J. (as Žáčková): On minimax solutions of stochastic linear programming problems. Časopis pro pěstování matematiky **91**(4), 423–430 (1966)
15. Ellsberg, D.: Risk, ambiguity, and the Savage axioms. Q. J. Econ. **75**(4), 643–669 (1961)
16. Gallego, G., Moon, I.: The distribution free newsboy problem: review and extensions. J. Oper. Res. Soc. **44**(8), 825–834 (1993)
17. Gilboa, I., Schmeidler, D.: Maxmin expected utility with non-unique prior. J. Math. Econ. **18**(2), 141–153 (1989)
18. Goh, J., Sim, M.: Distributionally robust optimization and its tractable approximations. Oper. Res. **58**(4), 902–917 (2010)
19. Gorissen, B.L., den Hertog, D.: Robust counterparts of inequalities containing sums of maxima of linear functions. Eur. J. Oper. Res. **227**(1), 30–43 (2012)
20. Grötschel, M., Lovász, L., Schrijver, A.: The ellipsoid method and its consequences in combinatorial optimization. Combinatorica **1**, 169–197 (1981)
21. Hadley, G., Whitin, T.M.: Analysis of Inventory Systems. Prentice-Hall International Series in Management. Prentice-Hall, Englewood Cliffs, NJ (1963)
22. Isii, K.: The extrema of probability determined by generalized moments (i) bounded random variables. Ann. Inst. Stat. Math. **12**(2), 119–134 (1960)
23. Kök, A.G., Fisher, M.: Demand estimation and assortment optimization under substitution: methodology and application. Oper. Res. **55**(6), 1001–1021 (2007)
24. G. Kök, A., Fisher, M., Vaidyanathan, R.: Assortment planning: review of literature and industry practice. In: Agrawal, B., Smith, S.A. (eds.) Retail Supply Chain Management, pp. 99–154. Springer, Berlin (2009)
25. Löfberg, J.: YALMIP: a toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD Conference (2004)
26. Mahajan, S., van Ryzin, G.: Retail inventories and consumer choice. In: Tayur, S., Ganesham, R., Magasine, M. (eds.) Quantitative Methods in Supply Chain Management, pp. 201–209. Kluwer Publishers, Amsterdam (1998)
27. Natarajan, K., Sim, M., Uichanco, J.: Asymmetry and ambiguity in newsvendor models. Technical report, National University of Singapore (2008)
28. Natarajan, K., Sim, M., Uichanco, J.: Tractable robust expected utility and risk models for portfolio optimization. Math. Financ. **20**(4), 695–731 (2010)
29. Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. **13**(1), 60–78 (2002)
30. Pardo, L.: Statistical Inference Based on Divergence Measures. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, Taylor & Francis Group, Boca Raton (2006)
31. Perakis, G., Roels, G.: Regret in the newsvendor model with partial information. Oper. Res. **56**(1), 188–203 (2008)
32. Pólik, I., Terlaky, T.: A survey of the S-lemma. SIAM Rev. **49**(3), 371–418 (2007)
33. Popescu, I.: A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. Math. Oper. Res. **30**(3), 632–657 (2005)
34. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2002)
35. Scarf, H.E.: A min-max solution to an inventory problem. In: Arrow, K.J., Karlin, S., Scarf, H.E. (eds.) Studies in Mathematical Theory of Inventory and Production, pp. 201–209. Stanford University Press, Stanford (1958)
36. Shapiro, A.: On duality theory of conic linear problems. In: Goberna, M.A., Lopez, M.A. (eds.) Semi-Infinite Programming: Recent Advances, pp. 135–165. Kluwer Academic Publishers, Dordrecht (2001)
37. Shapiro, A., Ahmed, S.: On a class of minimax stochastic programs. SIAM J. Optim. **14**(1), 1237–1249 (2004)
38. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009)
39. Shapiro, A., Kleywegt, A.J.: Minimax analysis of stochastic problems. Optim. Methods Softw. **17**(3), 523–542 (2002)
40. Signorelly, S., Heskett, J.L.: Benetton. In: Harvard Business School Case # 685020 (1985)
41. Steinberg, D.: Computation of Matrix Norms with Applications to Robust Optimization. Master's thesis, Technion-Israel Institute of Technology (2005)

42. Sun, P., Freund, R.M.: Computation of minimum-volume covering ellipsoids. Oper. Res. **52**(5), 690–706 (2004)
43. Tallis, G.M.: Elliptical and radial truncation in normal populations. Ann. Math. Stat. **34**(3), 940–944 (1963)
44. Vaagen, H., Wallace, S.W.: Product variety arising from hedging in the fashion supply chains. Int. J. Prod. Econ. **114**(2), 431–455 (2008)
45. Vandenberghe, L., Boyd, S., Comanor, K.: Generalized Chebyshev bounds via semidefinite programming. SIAM Rev. **49**(1), 580–595 (2007)
46. Wang, Z., Glynn, P.W., Ye, Y.: Likelihood robust optimization for data-driven newsvendor problems. Technical report, Stanford University (2009)
47. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Working paper (2013)
48. Zhu, S., Fukushima, M.: Worst-case conditional value-at-risk with application to robust portfolio management. Oper. Res. **57**(6), 1155–1168 (2009)
49. Zuluaga, L., Peña, J.: A conic programming approach to generalized Tchebycheff inequalities. Math. Oper. Res. **30**(2), 369–388 (2005)