RC25557 (WAT1509-011) September 3, 2015
Mathematics

IBM Research Report

A Data-Driven, Distribution-Free, Multivariate Approach to the Price-Setting Newsvendor Problem

Pavithra Harsha, Ramesh Natarajan*, Dharmashankar Subramanian
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598 USA

*currently at Amazon, Seattle, WA 98109

A data-driven, distribution-free, multivariate approach to the price-setting newsvendor problem

Pavithra Harsha
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, pharsha@us.ibm.com

Ramesh Natarajan
IBM T. J. Watson Research Center (past affiliation when this research was done)
Amazon, 515 Westlake Avenue N, Seattle, WA 98109 , ramesn@amazon.com

Dharmashankar Subramanian
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, dharmash@us.ibm.com

Many aspects of the classical price-setting newsvendor problem have been studied in the literature and most of the results pertain to the case where the price-demand relationship and demand distribution are explicitly provided. However, in practice, one needs to model and estimate these from historical sales data. Furthermore, many other drivers besides price must be included in the demand response model for statistical accuracy, along with conditional heteroskedasticity effects in the demand distribution. In this paper we develop a practical framework for data-driven, distribution-free, multivariate modeling of the price-setting newsvendor problem, which includes statistical estimation and price optimization methods for estimating the optimal solutions and associated confidence intervals. The specific novelty of the framework is that the relevant statistical estimation methods are carried out in close conjunction with the requirements of the optimization problem, which in this context requires the estimation of three distinct aspects of the demand distribution, namely the mean, quantile and superquantile (also known as conditional value-at-risk, CVaR). We investigate different statistical estimators, which are broadly based on generalized linear regression (GLR), mixed-quantile regression (MQR), and superquantile regression (SQR) respectively. Our results extend the previous literature, notably to incorporate heteroskedasticity in MQR, and to obtain a novel and exact large-scale decomposition method that is computationally efficient for SQR (these extensions are of independent interest, besides the application discussed here). Our detailed computational experiments indicate that quantile-based methods such as MQR and SQR provide better solutions for a wide range of demand distributions, although for certain location-scale demand distributions that are similar to the *Normal* distribution, GLR may be preferable.

Key words: Pricing, newsvendor, statistics: estimation, decomposition algorithm, heteroskedastic least squares, quantile regression, conditional value-at-risk, superquantile regression

---

1. Introduction

The classical price-setting newsvendor problem (and its many variants described for example in Porteus 1990, Khouja 1999) occupies a central and prominent role in pricing and inventory decision theory. In its simplest form, a firm must simultaneously and jointly determine the optimal price and optimal order quantity for a product with a *known* stochastic price-dependent demand, in order to maximize the expected profit during a single inventory period. The newsvendor profit function accounts for underage and overage costs which are respectively associated with scenarios when the demand is higher or lower than the order quantity, as well as the revenue that is obtained from the observed demand.

The topic of coordination of pricing and inventory is of great interest to firms from an operations-management perspective, since it offers up the possibility of obtaining a better solution to the profit maximization problem, particularly when compared to the corresponding uncoordinated case where the pricing and inventory decisions are made independently, or sequentially at best. This coordinated approach has received great impetus with the emergence of internet-based technologies for real-time inventory tracking and dynamic pricing, and is now widely used for revenue management in several application areas (Phillips 2005).

In many of these existing and emerging application areas, however, the relevant demand function is a random variable, whose probability distribution and dependence on price is not explicitly known, and must be modeled and estimated from historical sales data. Furthermore, many other drivers besides price must be included in the demand response model for statistical accuracy, along with conditional heteroskedasticity effects in the demand distribution. The objective of this paper is to address this issue using data-driven, distribution-free, multivariate regression methods to characterize the stochastic demand response, and the primary novelty is that the statistical modeling methods are carried out in close conjunction with the requirements of the optimization problem in the coordinated setting, here, the classical price setting newsvendor.

One motivation for our work is the potential application of the price-setting newsvendor model and related schemes to the emerging electricity smart grid for demand response planning. Here, the electric utility may simultaneously decide on both the scheduled generation and certain demand-shaping price incentives, so as to minimize the impact on the expected operational profits. In this application, other factors besides price, such as weather and time-of-day effects, will influence the demand response and demand variability, and must be taken into consideration in the statistical modeling. For example, consider the energy consumption data in Fig. 1 recorded during the morning peak in a dynamic price experiment at the Olympic Peninsula. Notice the significant impact of temperature on demand and the presence of temperature-dependent heteroskedasticity. The consumption patterns can also display significant daily and weekly dependencies (not shown). One approach here is to estimate a separate demand model from historical data for each combination of the levels of the relevant external factors; however, this is clearly impractical when the number of such combinations is very large. Besides, many potential combinations are rarely observed in the historical data. Therefore, it is preferable to capture these high-dimensional relationships using flexible, multivariate regression methods to obtain robust demand response models (e.g., see, Hastie et al. 2001) for operational planning.

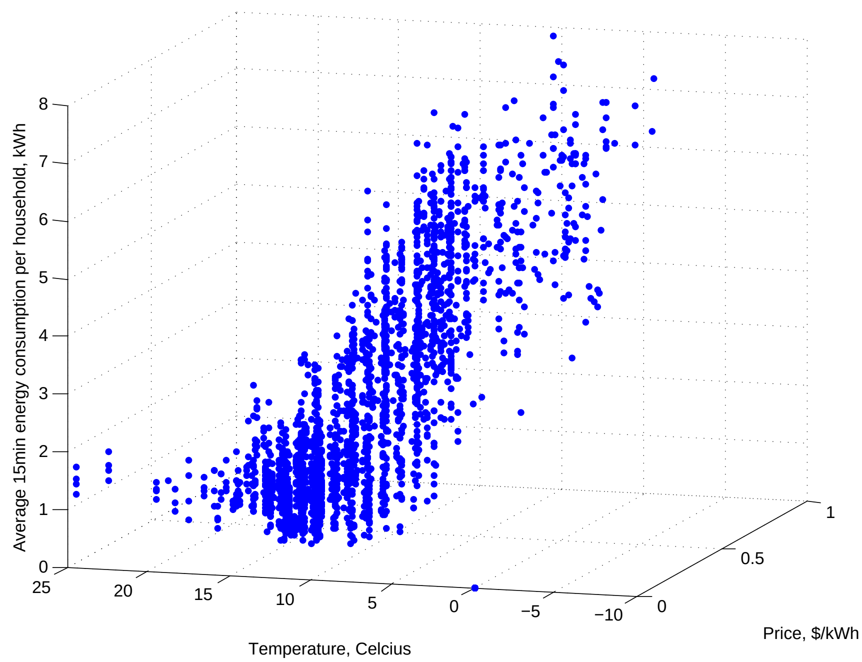

Figure 1 Energy consumption in kWh per household recorded every 15 mins between 7-9am from April 1, 2006 to March 31, 2007 in a dynamic pricing experiment in the Pacific Northwest GridWise Testbed Demonstration Project (Hammerstrom et al. 2007).

The same issues arise for applications of the price-setting newsvendor model in the retail, manufacturing and services sectors. For example, the data in Fig. 2 shows a price-dependent heteroskedasticity in individual stores sales of a product, as well as distributional variations across store locations (perhaps influenced by their individual characteristics and shopper demographics). Moreover, we observe a price-dependent monotonic variance function with the lowest demand variability at the largest price. The presence of non-monotonic variance functions are also quite likely in practice. For example, the lowest demand variability can come in the middle range of prices where one has a good understanding of the market (Raz and Porteus 2006). The methods described in this paper are therefore relevant in all these application areas as well.

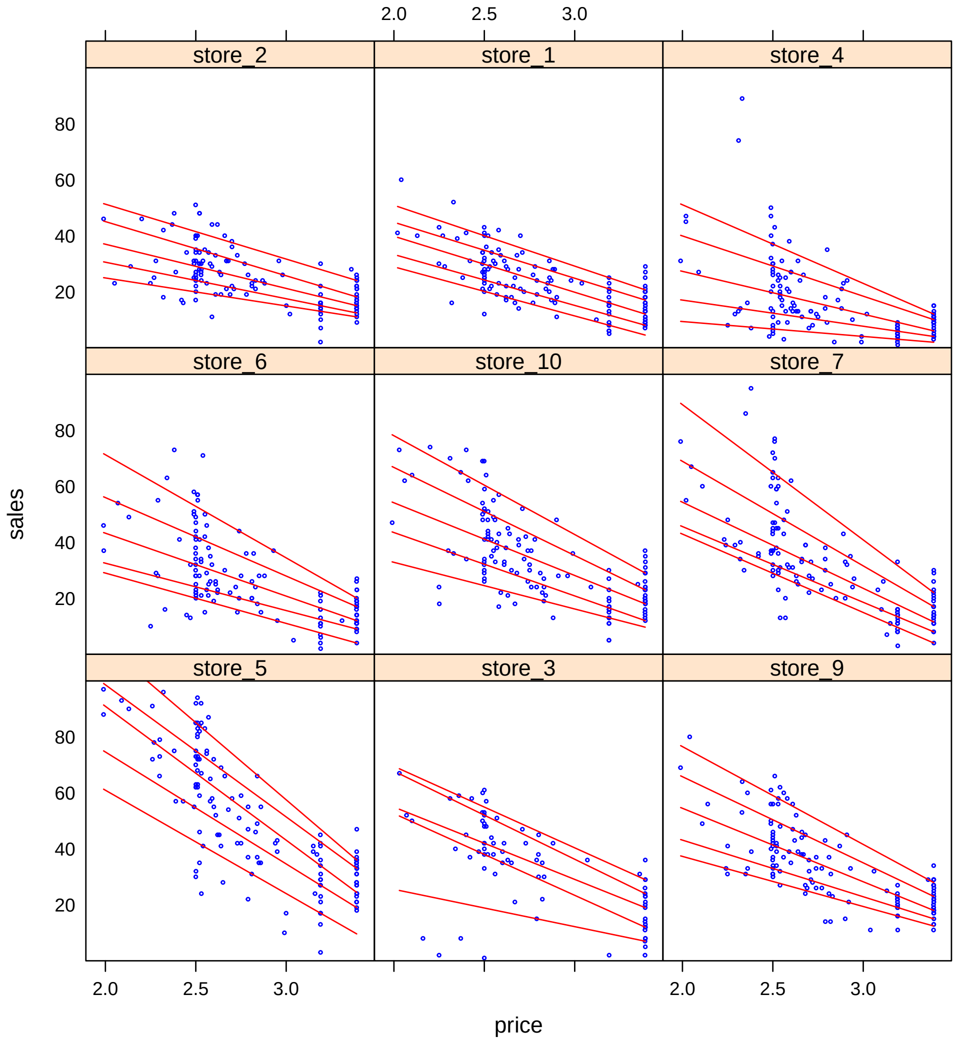

Figure 2 Retail sales data for a non-seasonal product as a function of price across several stores (data anonymized to protect confidentiality). The red lines are the quantiles at levels 0.1,0.25,0.5,0.75 and 0.9 respectively, and individual panels are ordered by increasing mean sales.

1.1. Contributions

This paper focuses on the classical price-setting newsvendor problem and provides results that address the two common variants of this problem: the *lost sales* setting where the excess demand is entirely lost and the *emergency order* setting where the excess demand is also met but at a high cost that is exogenous to the price. Our contributions are as follows and unless otherwise explicitly specified, our contributions refer to either variant of the decision problem.

1. **Practical framework for a data-driven, distribution-free, multivariate modeling approach to the price-setting newsvendor problem:** In this paper we develop a practical framework for data-driven, distribution-free, multivariate modeling of the price-setting newsven-

dor problem, which includes statistical estimation and price optimization methods for estimating the optimal solutions and associated confidence intervals. The specific novelty of the framework is that the relevant statistical estimation methods are carried out in close conjunction with the requirements of the decision problem, which in this context, requires the estimation of three distinct aspects of the demand distribution, namely the mean, quantile and superquantile (also known as conditional value-at-risk, CVaR).

In contrast to the current state of the art for solving the data-driven, price-sensitive newsvendor problem, the proposed framework does not require the complete price-dependent demand distribution prior to the optimization. More specifically, for emergency order setting, the three distinct aspects of the demand distributions, specifically the quantile and the superquantile estimators, are queried exactly at one quantile level. The same is true for the lost sales setting except for the superquantile estimator that is queried at a few different quantile levels as a part of a bounded one dimensional search in the price optimization step. Note that the latter is far from the full distribution.

2. **Consideration of different statistical estimation methods:** We investigate three different distribution-free, multivariate regression methods adapted to the price-setting newsvendor problem; these methods are broadly based on generalized linear regression (GLR), mixed-quantile regression (MQR), and superquantile regression (SQR) respectively. In fact, any combination of these or other new techniques may be used to estimate the three distinct quantities of interest, leading to a profusion of ways for implementing the desired optimization computations.
3. **Extensions to current CVaR estimation methods:** We develop an efficient and exact large-scale decomposition method to solve large instances of the SQR which currently does not scale beyond a few hundred sample points. The proposed algorithm is a novel cutting plane algorithm that is shown to be empirically far more tractable than the original SQR formulation. We also extend the MQR method to to allow conditional homoskedasticity and conditional heteroskedasticity over respectively desired subsets of covariates, a method critical for examples discussed in the introduction. The extensions to CVaR regression methods described in this paper have wider applicability and are of independent interest, besides the price-setting newsvendor application discussed in this paper (e.g., in financial applications).
4. **Computational Experiments and Insights:** We carry out a detailed computational analysis and comparison for a variety of stochastic demand models with different functional and noise characteristics. Our computational experiments highlight that quantile based methods such as MQR and SQR provide better solutions for a wider range of demand models, except for the case when the noise terms have a location-scale form that is similar to the *Normal* distribution (e.g., symmetric, unimodal, and not heavy-tailed) when GLR methods may be preferred.

1.2. Background and Relevant work

We review the background and relevant work in order to motivate the elements of the proposed framework outlined in this paper.

The evolving literature on the coordination of pricing and inventory decisions has been reviewed by Elmaghraby and Keskinocak (2003), Chan et al. (2004), Yano and Gilbert (2005), and more recently by Chen and Simchi-Levi (2010).

A synthesis of the literature on the price-setting newsvendor problem for the lost sales formulation (described in Section 2) is provided in Petruzzi and Dada (1999), which particularly covers the case when the stochastic price-demand relationship is specified in a certain form, e.g., the additive model with a linear demand function (Mills 1959), the multiplicative model with a iso-elastic demand function (Karlin and Carr 1962), and the mixed additive-multiplicative model (Young 1978). In all these cases, the mean demand is specified as a monotonic decreasing function of the price, and the variance is specified as a non-increasing function of price (further, typically, a constant variance is assumed in the additive case, and monotonic decreasing variance is assumed in the multiplicative and mixed additive-multiplicative case). The relevant existence and uniqueness conditions are provided, and the resulting optimal solutions are also compared to the equivalent *risk-free* case (i.e., where the deterministic price-demand relationship has the same functional form as the mean demand in the corresponding stochastic case). Yao et al. (2006) have extended these results to a more general class of price-demand functions for the additive and multiplicative models. Kocabıyıkoğlu and Popescu (2011) provide further generalizations, in particular, including the case when the demand variance may be a non-monotonic function of the price for mixed additive-multiplicative models. Such a non-monotonic variance function is quite likely in practice (e.g., see Lau and Lau (1988), Raz and Porteus (2006), and the discussion below), but was not previously covered by the analysis of the mixed additive-multiplicative models in Petruzzi and Dada (1999).

In Arikan et al. (2007), the model parameters for the additive-multiplicative class of models are estimated from some retail data sets (which were seasonally de-trended to isolate the price-dependent effects for regression modeling). They concluded that, based on statistical fit criteria alone, it was often difficult to distinguish between alternative model classes. Although the estimated models were used to obtain optimal solutions for the price-setting newsvendor problem, the confidence intervals for these solutions were not estimated; consequently, their results were inconclusive in terms of clarifying the impact of model fit and estimation errors on the optimal solutions.

Arikan and Jammernegg (2009) have reviewed a number of approaches for modeling the stochastic price-demand relationship in the literature, and they characterize the following two approaches as being distinctive. The first, due to Lau and Lau (1988), models the first four lower-order moments of the demand distribution as a function of price. The second, due to Raz and Porteus (2006), models the individual quantiles of the demand distribution as piecewise-linear functions of the price. These two "distinctive" approaches are in some sense complementary in terms of specifying price-dependent heteroskedasticity effects in the demand model: in the first, through the moments of the distribution function and in the second, through the levels of the quantile function. However, in both these papers, the model parameters are specified based on subjective assessment of experts, rather than directly estimated from historical sales data. For instance, the model specification is given in terms of the first four moments in Lau and Lau (1988) which are then somewhat subjectively matched to the corresponding moments of a four-parameter beta distribution for performing the newsvendor optimization. Raz and Porteus (2006) recommend obtaining the subjective estimates for the demand variability at a few selected price points, which are then interpolated and extended throughout the price range of interest using piecewise-linear functions.

A fully data-driven approach, by contrast, is not constrained by the need to obtain the subjective estimates in some convenient manner. So, for example, multiple demand drivers can be directly incorporated and estimated in the demand model, whereas the subjective assessment of these multivariate effects would be difficult at best. While the subjective approach is not the focus of this paper, it is nevertheless a useful alternative approach, particularly when there is no historical data (e.g., for new products that have little or no sales history).

For the *standard newsvendor* problem, the only decision variable is the order quantity (and price is not a decision variable). Beutel and Minner (2012) describe an approach where the demand model can comprise of multiple drivers that include the effects of price, price changes and weather. They observe that the inclusion of these additional drivers in the demand modeling substantially improves the accuracy of the demand forecasts, leading to a corresponding reduction in the safety stock requirements. They describe two data-driven modeling approaches for estimating the optimal order quantity for the standard newsvendor problem: the first is based on linear regression for the demand response, with a first-order heteroskedasticity correction for the model coefficient estimates; the second is based on direct estimation of the optimal order quantity using a linear programming formulation that is essentially equivalent to quantile regression (see also Rudin and Vahn (2014), who additionally propose using regularization and kernel-based methods for high-dimensional problems). The two methods in Beutel and Minner (2012) are quite cognate to the first two approaches in our work described below; however, the extensions required for the price-setting newsvendor problem are non-trivial and require the estimation of the CVaR.

Other related data-driven perspectives on the standard newsvendor problem include, for example, the use of bootstrap confidence intervals for newsvendor quantile estimates (Bookbinder and

Lordahl 1989), robust optimization (Scarf et al. 1958, Gallego and Moon 1993, Perakis and Roels 2008), non-parametric approaches in censored data environments (Godfrey and Powell 2001, Huh et al. 2008, Huh and Rusmevichientong 2009, Besbes and Muharremoglu 2013), operational statistics (Liyanage and Shanthikumar 2005, Chu et al. 2008) and sampling-based bounds (Levi et al. 2007, 2011, Rudin and Vahn 2014). None of these papers, however consider price as a decision variable, and barring Rudin and Vahn 2014, they also do not consider multiple drivers in the estimated demand models.

Some of the other variants of the price-setting newsvendor problem in the literature include: (a) the coordination of pricing and inventory for an assortment of products where the demand of any item depends on the prices of all the items (Aydin and Porteus 2008); (b) use of an alternate objective such as a risk-averse profit objective as opposed to a traditional risk-neutral one (Agrawal and Seshadri 2000, Chen et al. 2009); (c) multi-period models with backordered inventory (Federgruen and Heching 1999 and other related work). The focus in these papers is on the existence and uniqueness of the optimal decisions and related structural results (in similar spirit to Petruzzi and Dada 1999).

Our framework requires the estimated CVaR of the multivariate demand distribution, for which we have considered the following two recent approaches to CVaR or superquantile regression. The *mixed quantile regression* method proposed by Chun et al. (2012) estimates the CVaR using a linear programming formulation similar to quantile regression. Their method is restricted to homoskedastic or constant variance distributions and is extended in our paper to include conditional heteroskedastic effects. The *superquantile regression* method proposed by Rockafellar et al. (2014) also estimates CVaR using a linear programming (LP) formulation, which is however derived based on the risk quadrangle (see Rockafellar and Uryasev 2013). Although this method makes no assumptions about homoskedasticity, as indicated by the authors in their paper (as well as by our experience with this method), it does not scale beyond a few hundred sample points because of size explosion in the LP (both in the number of constraints and variables). We propose a novel decomposition method that enable this LP formulation to be scaled to very large sample sizes.

Organization: The organization of the rest of the paper is as follows: In Section 2 we describe the price-setting newsvendor problem and our data-driven distribution-free multivariate regression framework for practical implementations. We describe the three multivariate regression techniques along with our extensions in Section 3. In Section 4, we provide the details of our computational experiments and discuss our observations based on the results. In Section 5, we describe some additional enhancements for the operationalization of the data-driven approach for the price-setting newsvendor problem. Section 6 provides some concluding remarks.

2. Price-Setting Newsvendor Problem

Consider a single-product, profit-maximizing firm, which at the beginning of an inventory period has to set a unit product price $p \in \mathcal{P}$, where $\mathcal{P}$ is a closed and continuous set of feasible prices. Simultaneously, the firm also has to set the order quantity x for stocking the product at the unit procurement cost c. The stochastic price-dependent demand is denoted by $D(p, \mathbf{z})$, where $\mathbf{z}$ denotes the external drivers, as elucidated further below. Any unsold stock units at the end of the inventory period are redeemed at the unit salvage price s. Note that $p > c > s$ is required in order to have a meaningful and non-trivial newsvendor problem.

The external drivers $\mathbf{z}$ in the stochastic demand $D(p, \mathbf{z})$ may include the effects of time-of-day, day-of-week, season, weather, holidays, special events, advertising campaigns, promotional incentives, and so on. Furthermore, when dynamic time-series effects are considered, $\mathbf{z}$ may also include the lagged effects demand, price and other relevant drivers (see, e.g., Hanssens et al. 2003, Leeflang et al. 2000). The values of all external drivers $\mathbf{z}$ are assumed to be known at the beginning of the inventory period. Therefore, any demand drivers that are unknown, unmeasured or uncertain are not included in $\mathbf{z}$, and their effects are considered to be part of the stochastic component of $D(p, \mathbf{z})$.

We consider two common variants of the price-setting newsvendor problem that arise in different applications.

Lost Sales Formulation: For a product retailer, we are primarily concerned with this variant of the price-setting newsvendor problem. Here, if the observed demand $D(p, \mathbf{z})$ exceeds the order quantity x, then the resulting underage is associated with a unit cost $p - c + v$ where v may be the monetary equivalent of the loss of consumer goodwill that is incurred due to the out-of-stock situation. On the other hand, if the observed demand $D(p, \mathbf{z})$ is lower than the order quantity x, then the resulting overage is associated with a unit cost $c - s$. The product retailer aims to maximize the expected profit by jointly optimizing the two decision variables, viz., the unit price p and the order quantity x. As is well known, this optimization problem can be formulated as:

$$\Pi_{ls}(\mathbf{z}): \quad \max_{p \in \mathcal{P}, x} \quad (p-c)E[D(p,\mathbf{z})] - (p-c+v)E\left[D(p,\mathbf{z}) - x\right]^{+} - (c-s)E\left[x - D(p,\mathbf{z})\right]^{+}. \quad (2.1)$$

Emergency Order Formulation: For the electricity provider, by contrast, we are primarily concerned with this variant of the formulation. Here, the order quantity x represents the pre-scheduled electricity generation. If the observed demand $D(p, \mathbf{z})$ exceeds x, then the resulting shortfall is immediately procured from the spot market or from spinning reserve, but at the premium fixed unit procurement cost $m > c$. The resulting underage unit cost $m - c$ is also often referred to as the unanticipated stock-replenishment costs. The unit salvage price $s < c$ is associated with the excess in the pre-scheduled generation; for example, s may represent the contracted per unit sell price with a bulk storage farm. The electricity provider also aims to maximize the expected profitability by jointly optimizing the two decision variables, viz., the unit price p and the order quantity x. Similarly, this optimization problem can be formulated as:

$$\Pi_{eo}(\mathbf{z}): \quad \max_{p\in\mathcal{P},x} \quad (p-c)E[D(p,\mathbf{z})]-(m-c)E\left[D(p,\mathbf{z})-x\right]^{+}-(c-s)E\left[x-D(p,\mathbf{z})\right]^{+}. \tag{2.2}$$

We now describe a reformulation of (2.1) and (2.2) that is suitable for the implementation of the required optimization procedures.

2.1. Optimization Formulation

The properties of the newsvendor objective function have been widely studied (Zipkin 2000). For the lost sales (2.1) and emergency order (2.2) formulations, the objective function is concave in x for given p. Therefore, in both cases, there is a unique optimal solution for x given p, denoted $\mathrm{VaR}_\alpha[D(p,\mathbf{z})]$, and given by

$$\mathrm{VaR}_\alpha[D(p,\mathbf{z})]=\inf\left\{x\geq 0: F_{D(p,\mathbf{z})}(x)\geq\alpha\right\}, \tag{2.3}$$

where $F_{D(p,\mathbf{z})}(.)$ is the cumulative distribution function (c.d.f.) of the random variable $D(p,\mathbf{z})$. The *critical quantile* or the *newsvendor quantile* $\alpha\in[0,1]$ is denoted by α_{ls} for the lost sales formulation, and by α_{eo} for the emergency order formulation respectively, with

$$\alpha_{ls}=\frac{p-c+v}{p-s+v}, \quad \alpha_{eo}=\frac{m-c}{m-s}. \tag{2.4}$$

The quantity $\mathrm{VaR}_\alpha[D(p,\mathbf{z})]$ in Eq. (2.3) is the α-level value-at-risk of $D(p,\mathbf{z})$, or equivalently, the α-level quantile function of $D(p,\mathbf{z})$.

In Eq. (2.4), the value α_{eo} depends only on the known problem parameters m, c and s. The value α_{ls}, however, depends on the decision variable p, in addition to the specified parameters c, s and v. While this distinction between α_{eo} and α_{ls} is important for the respective optimization procedures, as described later below, for notational brevity, and to emphasize the common aspects of the optimization formulation, we suppress the dependence of α_{ls} on p, and also omit the subscripts on α_{ls} and α_{eo} below (except where this distinction is explicitly required).

Then, substituting the conditional optimal value for x from Eq. (2.3) into either (2.1) or (2.2) results in the following reduced objective function which only involves the decision variable p:

$$\Pi_{ls}(\mathbf{z}) \text{ or } \Pi_{eo}(\mathbf{z}): \quad \max_{p\in P} \quad (p-s)E[D(p,\mathbf{z})]-(c-s)\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]. \tag{2.5}$$

Here $\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]$ denotes the α-level conditional value-at-risk of $D(p,\mathbf{z})$, which is defined in Rockafellar and Uryasev (2000) as

$$\mathrm{CVaR}_\alpha[D(p,\mathbf{z})] = \min_x \left[x + \frac{1}{(1-\alpha)} E\left[D(p,\mathbf{z}) - x\right]^+ \right]. \tag{2.6}$$

For a continuous random variable $D(p,\mathbf{z})$, this is identical to the conditional expected value in the upper α tail, given by

$$\begin{aligned} \mathrm{CVaR}_\alpha[D(p,\mathbf{z})] &= E[D(p,\mathbf{z}) | D(p,\mathbf{z}) \geq \mathrm{VaR}_\alpha[D(p,\mathbf{z})]] \\ &\equiv \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_\tau[D(p,\mathbf{z})] d\tau. \end{aligned} \tag{2.7}$$

For discrete or mixed discrete-continuous $D(p,\mathbf{z})$, there is an equivalent definition to Eq. (2.7) which is given later below.

The conditional value-at-risk $\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]$ in Eq. (2.6) arises in diverse disciplines, although the terminology may vary depending on the interpretation of the random variable $D(p,\mathbf{z})$. For example, in the electricity distribution industry, if $D(p,\mathbf{z})$ denotes the stochastic electricity demand, and if α denotes the quantile level of $D(p,\mathbf{z})$ corresponding to the pre-scheduled electricity generation, then $\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]$ is essentially equivalent to the well-known reliability metric *LOLE - Loss of Load in Expectation* (Billinton and Allan 1996, Harsha et al. 2013). In more recent literature, the conditional value-at-risk is also referred to as the *superquantile* (Rockafellar et al. 2014).

2.1.1. Discussion and Perspective The optimization formulation in Eq. (2.5) involves a specific linear combination of the two quantities $E[D(p,\mathbf{z})]$ and $\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]$. Although rarely presented in this form in the operations management literature, Eq. (2.5) is reminiscent of the mean-CVaR objective function used in risk optimization, e.g., Rockafellar and Uryasev (2000). A third quantity $\mathrm{VaR}_\alpha[D(p,\mathbf{z})]$ is required to obtain the optimal order quantity in Eq. (2.3).

In summary, the specification of the three quantities $E[D(p,\mathbf{z})]$, $\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]$ and $\mathrm{VaR}_\alpha[D(p,\mathbf{z})]$ is sufficient to obtain the desired optimal solutions to the price-setting newsvendor problem from Eq. (2.3) and Eq. (2.5). If the stochastic demand function $D(p,\mathbf{z})$ is available in some standard, explicit form, these three quantities can be directly evaluated (e.g., using closed-form expressions available for many standard distributions; see, Andreev et al. 2005, Nadarajah et al. 2014). From the perspective of this paper, however, the optimization formulation in Eq. (2.3) and Eq. (2.5) suggests that it may be fruitful to directly estimate these quantities from the historical sales data, without the intermediate step of explicitly ascertaining $D(p,\mathbf{z})$. A variety of multivariate regression-based techniques can be used for this purpose, which are capable of flexibly modeling the respective functional dependencies on the demand drivers, with minimal assumptions on the form and distribution of $D(p,\mathbf{z})$, as described further below.

2.2. Optimization Algorithms

We outline one possible approach for obtaining the optimal price and optimal order quantity for the optimization formulation described in Section 2.1.

Algorithm 1 Optimization Procedure for Price-Setting Newsvendor Problem (Emergency Order)

Input: Given m, c, s.

1: Estimate $E[D(p,\mathbf{z})]$.
2: Obtain α from Eq. (2.4).
3: Estimate $\text{CVaR}_\alpha[D(p,\mathbf{z})]$, and obtain the corresponding optimal price $p^* \in \mathcal{P}$ from Eq. (2.5).
4: Estimate optimal order quantity $x^* = \text{VaR}_\alpha[D(p^*,\mathbf{z})]$ from Eq. (2.3).

Output: optimal price p^*, and optimal order quantity x^*.

Algorithm 1 is given for the emergency order setting. The corresponding algorithm for the lost sales setting only differs in Steps 2 and 3. In particular, after Step 1, an initial guess of price p_0 is made in Step 2 of the algorithm that results in an initial quantile estimate α_0. Then, Steps 3 and 2 are iteratively executed till some converge criterion is met to obtain an optimal price estimate, p^*. Finally, the corresponding optimal order quantity estimate is obtained from Step 4. A pictorial representation of our framework that enables data-driven, distribution-free, multivariate modeling approaches is provided in Fig. 3, which captures the high-level implementation details of the algorithms for the emergency order and lost sales settings.

We make some remarks about Algorithm 1. First, not only can this algorithm be used when the explicit form of the stochastic demand function $D(p,\mathbf{z})$ is known, but it can also be used when the three quantities $E[D(p,\mathbf{z})]$, $\text{CVaR}_\alpha[D(p,\mathbf{z})]$ and $\text{VaR}_\alpha[D(p,\mathbf{z})]$ are directly estimated from the data. Second, the objective function of problem (2.5) need not be concave in general (the existence and uniqueness results for the price-setting newsvendor problem are given in Kocabıyıkoğlu and Popescu (2011), see Section 1.2 below). Consequently, even though the optimal price p^* always exists (given a continuous objective function (2.5) in a closed set $\mathcal{P}$)), there may be multiple local maxima. However, a standard univariate, derivative-free, non-linear optimization procedure (Brent 1973) can be used to directly obtain the desired optimal solution for (2.5) for both the lost sales and the emergency order settings.

3. Some Relevant Statistical Estimation Methods

As noted in Section 2.1, the optimization formulation in Eq. (2.3) and Eq. (2.5) is completely specified in terms of the three quantities $E[D(p,\mathbf{z})]$, $\text{VaR}_\alpha[D(p,\mathbf{z})]$, and $\text{CVaR}_\alpha[D(p,\mathbf{z})]$, and each

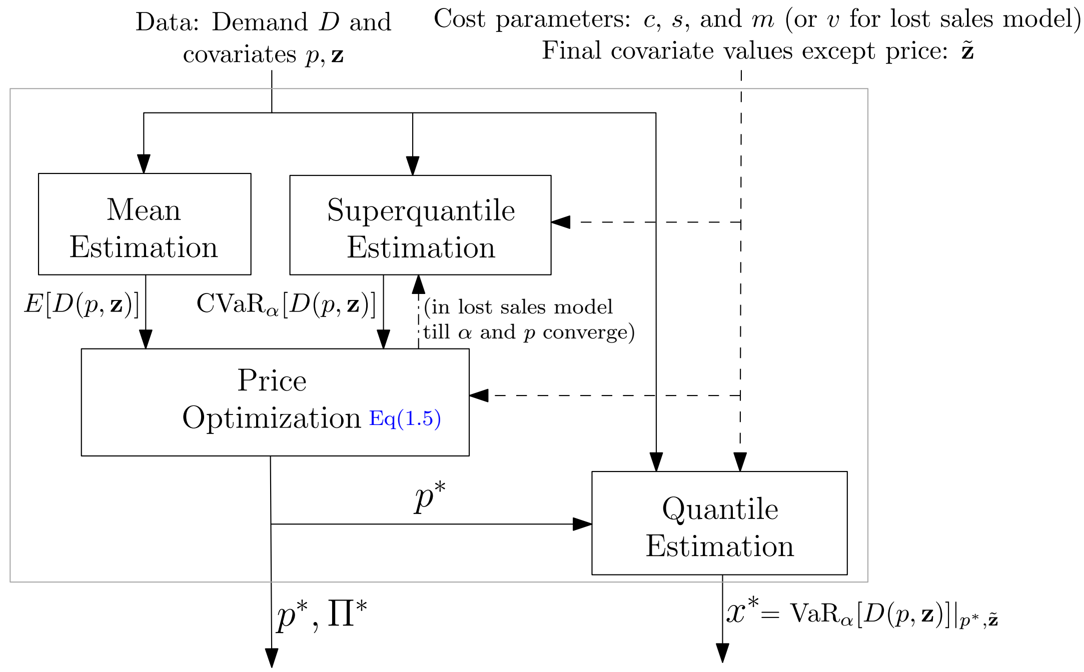

Figure 3 Framework for a practical data-driven approach to the price-setting newsvendor problem

of these can be directly estimated from historical sales data by a variety of different techniques. Therefore, any appropriate combination of these techniques can therefore be used in the optimization formulation, leading to a profusion of ways for implementing the computations.

We consider three specific approaches below, which, broadly speaking, are based on generalized linear regression (GLR), mixed quantile regression (MQR), and superquantile regression (SQR) respectively. The first two approaches are related to the methods previously described in Chun et al. (2012) for estimating conditional risk measures using linear regression and quantile regression, but extended here to include conditional heteroskedasticity effects. The third approach is related to recent work in (Rockafellar et al. 2014) on obtaining direct regression estimates of conditional risk measures. For all three approaches, we provide appropriate extensions, and adaptation to the price-setting newsvendor problem.

The standard notation for regression problems is used in the description below. The response variable is denoted by Y, which is typically the demand $D(p, \mathbf{z})$ itself, and the regression models typically involve the estimation of $E[Y]$, $\mathrm{VaR}_\alpha[Y]$ or $\mathrm{CVaR}_\alpha[Y]$. The set of regression covariates are denoted by $\mathbf{X}$ (or sometimes by $\mathbf{Z}$), and this set typically includes the constant or intercept term (except where explicitly indicated below), along with other terms that involve transformations and interactions of the various external demand drivers $(p, \mathbf{z})$ as appropriate. The regression functions involve linear combinations of these regression covariates with the coefficients being the parameters to be estimated from historical data, based on the appropriate regression formulation. We note that the regression functions are typically linear in the parameters, the ability to include nonlinear and interaction terms in the covariate effects, as well as nonlinear response transformations, provides sufficient generality for modeling a wide range of functional forms required in practical applications (Hastie et al. 2001).

3.1. Generalized Linear Regression (GLR)

The first approach is based on heteroskedastic regression using generalized linear models (GLM). If Y has a known distribution from the exponential family (e.g. see McCullagh and Nelder 1989), then the corresponding GLM regression estimates can be obtained using maximum likelihood. However, the GLM approach can also be based on maximum quasi likelihood (Wedderburn 1974), which only requires a specification of the relation between the first two moments of Y instead of the full distribution. For modeling the response mean $E[Y]$, the regression estimates from maximum quasi likelihood are essentially equivalent to maximum likelihood estimates whenever Y has an exponential family distribution. However, the maximum quasi likelihood estimates retain the desirable properties of consistency, efficiency and asymptotic normality of the maximum likelihood estimates, even if the distribution of Y is not explicitly known. The maximum quasi likelihood formulation can also be extended to incorporate heteroskedasticity modeling (see e.g. Nelder and Pregibon 1987, Davidian and Carroll 1987). This a highly desirable property for demand modeling in the context of the price-setting newsvendor problem.

To describe this approach in full generality, we denote $E[Y] = \mu$, we need a specification of the mean-variance relationship, i.e. $E[(Y-\mu)^2] = \phi V(\mu)$, where $V(\mu)$ is the variance function, and ϕ is the dispersion parameter. For example, a common specification for the variance function is $V(\mu) = \mu^\theta$ for some fixed θ. For the Normal, Gamma and Inverse Gaussian distributions, $\theta = 0, 2$ and 3 respectively. In general any other value of θ can also be specified. Regression models for μ and ϕ can be specified in the form of generalized linear models, and the model parameters can be estimated iteratively using two GLM maximum quasi-likelihood function calls that iteratively estimate the mean and the dispersion. For more information on this heteroskedastic method we refer the reader to papers by Nelder and Pregibon (1987), Davidian and Carroll (1987, 1988) and Nelder and Lee (1992) respectively. We apply this above approach to data-driven newsvendor problem and describe the resulting algorithm next.

Model Specification and Estimation. To fix ideas for our context, we consider the following generating model for the stochastic demand function:

$$Y = \mu + \sqrt{\phi V(\mu)}\epsilon, \tag{3.1}$$

where ϵ is a random variable whose distribution is independent of $\mathbf{X}$ with $E[\epsilon]=0$ and $E[\epsilon^2]=1$, with

$$g(\mu)=\boldsymbol{\beta}^T\mathbf{X}, \quad h(\phi)=\boldsymbol{\gamma}^T\mathbf{Z}, \tag{3.2}$$

where $\boldsymbol{\beta},\boldsymbol{\gamma}$ are the respective regression parameters, and $g:\mathbb{R}\rightarrow \text{Range}(Y)$, and $h:\mathbb{R}\rightarrow\mathbb{R}^+$ are the respective link functions for μ and ϕ, and $\mathbf{X}$ and $\mathbf{Z}$ denote the respective set of covariates in the mean and dispersion models. Note that if $\mathbf{X}$ and $\mathbf{Z}$ have a common subset of covariates, then there may be aliasing in the specification of ϕ and $V(\mu)$, but it is always advantageous to carefully specify $V(\mu)$ so that ϕ can take a simpler form with fewer parameters to capture the additional heteroscedastic variation.

The model in Eq. (3.1) and Eq. (3.2) is equivalent to the additive-multiplicative demand model in the inventory literature, and the covariates $\mathbf{X}$ and $\mathbf{Z}$ represent the demand drivers including price in this model. We outline the steps for estimating μ and ϕ in Algorithm 2.

Algorithm 2 Heteroscedastic regression

Input: Data $\{\mathbf{X}_i,\mathbf{Z}_i,Y_i\}$ for $i=1,...,N$, the variance function $V(\mu)$, and the link functions $g(.)$ and $h(.)$ respectively.

1: Set the initial values for $\hat{\phi}_i$ for $i=1,...,N$.
2: Obtain the mean regression parameters $\hat{\boldsymbol{\beta}}$, using response Y_i, covariates $\mathbf{X}_i$, variance function $V(\mu)$, dispersion $\hat{\phi}_i$, and link function $g(.)$. Set $\hat{\mu}_i=g^{-1}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i)$ and obtain the Pearson residuals $\hat{d}_i=\frac{(Y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}$.
3: Obtain the dispersion regression parameters $\hat{\boldsymbol{\gamma}}$, using response $\hat{d}_i$, covariates $\mathbf{Z_i}$, variance function $V(\phi)=\phi^2$, dispersion 2, and link function $h(.)$. Set $\phi_i=h^{-1}(\hat{\boldsymbol{\gamma}}^T\mathbf{Z}_i)$.
4: Repeat from step 2 till $\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}}$ converge.

Output: Estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ for heteroscedastic regression.

Algorithm 2 has a very modular structure, which can be easily implemented using existing software for fitting GLM models using maximum quasi likelihood. Other modifications to improve iterative convergence, and to obtain unbiased estimates for the dispersion regression parameters in small samples, described in Smyth et al. (2001),

The VaR and CVaR of ϵ can now be obtained from the empirical distribution of the adjusted residuals $\hat{\epsilon}$ where $\hat{\epsilon}_i=\frac{Y_i-\hat{\mu}_i}{\sqrt{\hat{\phi}_i V(\hat{\mu}_i)}}$. We denote the empirical cdf $F_{\hat{\epsilon}}(u)=\frac{1}{N}\sum_{i=1}^N \mathrm{I}_{[\hat{\epsilon}_i]}(u)$, where $\mathrm{I}_{[\hat{\epsilon}_i]}(u)$ is the indicator function which takes the value 1 if $(u-\hat{\epsilon}_i)\geq 0$, and 0 otherwise. Then,

$$\mathrm{VaR}_\alpha[\hat{\epsilon}] = \inf\{u : F_{\hat{\epsilon}}(u)\geq\alpha\}, \tag{3.3}$$

$$\mathrm{CVaR}_\alpha[\hat{\epsilon}] = \lambda_\alpha(\hat{\epsilon})\mathrm{VaR}_\alpha[\hat{\epsilon}] + (1-\lambda_\alpha(\hat{\epsilon}))\,E\left[\hat{\epsilon}|\hat{\epsilon}>\mathrm{VaR}_\alpha[\hat{\epsilon}]\right], \tag{3.4}$$

$$\text{where } \lambda_\alpha(\hat{\epsilon}) = \frac{F_{\hat{\epsilon}}\left(\mathrm{VaR}_\alpha[\hat{\epsilon}]\right)-\alpha}{1-\alpha}.$$

This description of CVaR for discrete distributions is given by Rockafellar and Uryasev (2002).

In summary, given Y, $\mathbf{X}$ and $\mathbf{Z}$, the three quantities of interest for the price-setting newsvendor problem (see Fig. 3 and Section 2.2) are as follows:

$$E[Y] = \hat{\mu}, \tag{3.5}$$

$$\mathrm{VaR}_\alpha[Y] = \hat{\mu} + \sqrt{\hat{\phi} V(\hat{\mu})}\, \mathrm{VaR}_\alpha[\hat{\epsilon}], \tag{3.6}$$

$$\mathrm{CVaR}_\alpha[Y] = \hat{\mu} + \sqrt{\hat{\phi} V(\hat{\mu})}\, \mathrm{CVaR}_\alpha[\hat{\epsilon}], \tag{3.7}$$

where $g(\hat{\mu}) = \hat{\boldsymbol{\beta}}^T \mathbf{X}$ and $h(\hat{\phi}) = \hat{\boldsymbol{\gamma}}^T \mathbf{Z}$. Here, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ are the outputs of Algorithm 2.

3.2. Mixed Quantile Regression (MQR)

The second approach is based on quantile regression which is a method to estimate $\mathrm{VaR}_\alpha[Y]$ given α. We note that $E[D(p,\mathbf{z})] = \mathrm{CVaR}_0[D(p,\mathbf{z})]$ and so a module for estimating $\mathrm{CVaR}_\alpha[D(p,\mathbf{z})]$ given α can obtain the remaning two of the three desired quantities for the price-setting newsvendor problem. One method for the evaluation of $\mathrm{CVaR}_\alpha[Y]$ is motivated from the module for estimating $\mathrm{VaR}_\alpha[Y]$ given α exploiting Eq. (2.7), as described further below.

One motivation for using the quantile regression approach is that it enables a broader class of stochastic demand functions to be modeled. For instance, the specific example below describes a response variable for which different covariates are significant at different quantile levels. This characteristic cannot be modeled using the parameterization that was used for heteroskedastic regression models in Section 3.1.

Example 1. This example describes a stochastic demand function with different demand drivers at different quantile levels. Consider the random variable ϵ, and let

$$Y = \boldsymbol{\beta}^T \mathbf{X} + \boldsymbol{\gamma}_1^T \mathbf{Z_1} \min\{\epsilon, \lambda\} + \boldsymbol{\gamma}_2^T \mathbf{Z_2} \max\{\lambda, \epsilon\}, \tag{3.8}$$

where the constant $\lambda = \mathrm{VaR}_\zeta[\epsilon]$ for some value $\zeta \in (0,1)$. Note that since

$$\mathrm{VaR}_\alpha[\min\{\epsilon,\lambda\}] = \begin{cases} \lambda, & \alpha \geq \zeta, \\ \mathrm{VaR}_\alpha[\epsilon], & \alpha < \zeta. \end{cases}$$

$$\mathrm{VaR}_\alpha[\max\{\epsilon,\lambda\}] = \begin{cases} \mathrm{VaR}_\alpha[\epsilon], & \alpha \geq \zeta, \\ \lambda, & \alpha < \zeta, \end{cases}$$

and similarly since

$$\mathrm{CVaR}_\alpha[\min\{\epsilon,\lambda\}] = \begin{cases} \lambda, & \alpha \geq \zeta, \\ \mathrm{CVaR}_\alpha[\epsilon] + (\lambda - \tau)\frac{1-\zeta}{1-\alpha}, & \alpha < \zeta. \end{cases}$$

$$\mathrm{CVaR}_\alpha[\max\{\epsilon,\lambda\}] = \begin{cases} \mathrm{CVaR}_\alpha[\epsilon], & \alpha \geq \zeta, \\ \frac{\zeta-\alpha}{1-\alpha}\lambda + \frac{1-\zeta}{1-\alpha}\tau, & \alpha < \zeta. \end{cases}$$

where $\tau = \mathrm{CVaR}_{\zeta}[\epsilon]$, we have

$$\mathrm{VaR}_{\alpha}[Y] = \boldsymbol{\beta}^T \mathbf{X} + \mathrm{VaR}_{\alpha}[\min\{\epsilon, \lambda\}] \boldsymbol{\gamma}_1^T \mathbf{Z_1} + \mathrm{VaR}_{\alpha}[\max\{\epsilon, \lambda\}] \boldsymbol{\gamma}_2^T \mathbf{Z_2}.$$

along with a similar expression for $\mathrm{CVaR}_{\alpha}[Y]$.

A situation highlighted in this specific example arises frequently in practice. For example, the uppermost quantiles of the stochastic electricity demand are likely to be quite sensitive to price and weather covariates, whereas the lowermost quantiles are quite insensitive to these covariates. We would like to note that generating model in Eq. (3.8) is an example where the sum of VaR or CVaR of two different random variables (i.e., min and max) is equal to their individual sums. This property does not hold for general random variables.

Model estimation. Given a response Y, covariates X and quantile level α, the estimate for $\mathrm{VaR}_{\alpha}[Y]$ is obtained using quantile regression (Koenker and Bassett 1978) in the form

$$\mathrm{VaR}_{\alpha}[Y] = \boldsymbol{\beta}_v^T \mathbf{X}. \tag{3.9}$$

Quantile regression involves solving the following optimization problem to estimate $\boldsymbol{\beta}_v$:

$$\mathbf{QR}: \quad \hat{\boldsymbol{\beta}}_v = \arg\min_{\boldsymbol{\beta}} \quad \frac{1}{N} \sum_{i=1}^{N} \psi_{\alpha}(Y_i - \boldsymbol{\beta}^T X_i) \tag{3.10}$$

where $\psi_{\theta}(t) = \theta[t]^+ + (1-\theta)[-t]^+$ and $\theta = [0, 1]$. The function $\psi_{\theta}(t)$ is commonly referred to as the quantile loss function and is the quantile weighted sum of the positive and negative deviations between the response variable and its estimate. The goal of QR problem is to minimize this loss function. The QR optimization problem can be rewritten as a linear programming problem and can be solved very efficiently. QR is a standard subroutine available in most commercial software packages.

For a fixed data sample, the estimates of these quantile regression coefficients will depend on α in general, and significance tests can be used to ascertain if these differences are indicative of heteroskedasticity.

One method for the evaluation of $\mathrm{CVaR}_{\alpha}[Y]$ is based on the numerical quadrature of the integral in Eq. (2.7). Hence $\mathrm{CVaR}_{\alpha}[Y]$ will be a linear combination, with appropriate quadrature weights, of $\mathrm{VaR}_{\alpha'}[Y]$ evaluated at certain quadrature nodes α', where $\alpha < \alpha' < 1$ (we typically use quadrature rules that only involve nodes that are in the interior of the interval to avoid the estimation of the extremal quantiles $\alpha' = 0, 1$). From Eq. (3.9), the corresponding estimate for $\mathrm{CVaR}_{\alpha}(Y)$ also has the form (Peracchi and Tanase 2008, Leorato et al. 2012):

$$\mathrm{CVaR}_{\alpha}(Y) = \boldsymbol{\beta}_c^T \mathbf{X}. \tag{3.11}$$

A related method was first described in Rockafellar et al. (2008), and further explored by Chun et al. (2012) where it is aptly referred to as a mixed quantile regression (MQR). However, as presented there, MQR is restricted to the homoskedastic case (in which the the variance does not depend on any of the covariates). We extend this method to incorporate conditional heteroskedasticity. As pointed in the introduction, this is highly desirable property for demand modeling in the context of the price-setting newsvendor problem.

In order to describe this formulation, we let $\sum_{j=1}^{M} w_j \mathrm{VaR}_{\alpha_j}[Y]$ be the discretization of Eq. (2.7), where α_j and w_j denote the quadrature nodes and weights respectively. Then, we have

$$\mathbf{MQR}: \quad \hat{\boldsymbol{\beta}}_c = \arg\min_{\boldsymbol{\beta},\boldsymbol{\tau}_j} \quad \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} w_j \psi_{\alpha_j}\left(Y_i - (\boldsymbol{\tau}_j + \boldsymbol{\beta})^T \mathbf{X}_i\right), \tag{3.12}$$

$$\text{s.t.,} \quad \sum_{j=1}^{M} w_j \boldsymbol{\tau}_j = \mathbf{0}, \tag{3.13}$$

$$\bar{\mathbf{e}}.\boldsymbol{\tau}_j = \mathbf{0} \quad \forall\ j = 1,\ldots,M. \tag{3.14}$$

where $\psi_\theta(t)$ is the loss function for quantile regression and $\bar{\mathbf{e}}$ is a vector such that $\bar{e}_p = 1$ if the p^{th} covariate (not including the intercept) is homoskedastic (i.e., does not impact variance) and 0 otherwise. For example, if a simple mid-point quadrature rule is used for the discretization of Eq. (2.7), then with $\Delta := M^{-1}(1-\alpha)$, we have $w_j = (1-\alpha)^{-1}\Delta$ and $\alpha_j := \alpha + (j-0.5)\Delta,\ j = 1,..,M$. Similarly, if a Gauss-Legendre quadrature rule is used, then $w_j = 0.5(1-\alpha)\delta_j$ and $\alpha_j = 0.5\left[(1-\alpha)\xi_j + (1+\alpha)\right]$ where δ_j and $\xi_j \in (-1,1)$ respectively denote the weights and nodes of the corresponding M-point quadrature rule.

The objective function in problem **MQR** is a weighted sum of individual loss functions for each quantile level α_j, with the corresponding regression function $(\boldsymbol{\beta}+\boldsymbol{\tau}_j)^T\mathbf{X}$. Constraint (3.13) ensures that $\hat{\boldsymbol{\beta}}_c^T\mathbf{X}$ is the desired estimate of $\mathrm{CVaR}_\alpha[Y]$. Constraint (3.14) is more restrictive version of constraint (3.13) and imposes the condition that $\boldsymbol{\tau}_j$ is set to zero for any homoskedastic covariate (besides intercept which is always treated as a heteroskedastic covariate).

Claim 1. *The* **MQR** *formulation by design allows for conditional homoskedasticity and conditional heteroskedasticity over the space of covariates.*

Proof. The space of covariates are divided into homoskedastic and heteroskedastic covariates and this is similar in effect to limiting the set of covariates $\mathbf{Z}$ that are allowed in the dispersion model for heteroskedastic regression in Eq. (3.2). Constraint (3.14) imposes conditional homoskedasticity on the homoskedastic covariates by restricting all quantiles to have an identical coefficients (the elements of the $\boldsymbol{\tau}_j$ are set to zero for any covariates in $\mathbf{X}$ but not in $\mathbf{Z}$) and thus ensuring parallelism in that covariate dimension. On the other hand, constraint (3.13) (which is the only constraint for the heteroskedastic covariates) provides $\boldsymbol{\tau}_j$ a degree of freedom that it can vary across quantiles thus the estimation method can choose to have quantiles that are not parallel to obtain a better fit, similar to Fig. 2. □

Thus to model homoskedasticity in all covariates all $\boldsymbol{\tau}_j$'s are set to zero except for the intercept term. This recovers the MQR formulation in Chun et al. (2012) as a special case. In the most general full heteroskedastic case, where all $\boldsymbol{\tau}_j$'s can be non-zero, the objective function Eq. (3.12) decouples and can be evaluated with independent and individual quantile regressions for each α_j. This is the method discussed in (Peracchi and Tanase 2008, Leorato et al. 2012) and can be implemented with the widely-available quantile regression modules. On the other hand, the extended MQR method described here can limit the subset of covariates over which heteroskedasticity is manifested and requires implementing a specialized linear program.

One difficulty with the individual quantile regression estimates in Eq. (3.10) is that these may not monotonic in α over the range of the covariates $\mathbf{X}$, particularly for mis-specified regression models (see, e.g. Chernozhukov et al. 2010). This long-standing problem called "quantile crossing" is partly mitigated in the homoskedastic case by explicitly enforcing parallel quantile functions via constraint (3.14) in the **MQR** formulation. While this aspect is not pursued here, the resulting CVaR regression estimates may not be monotonic in α over the range of the covariates $\mathbf{X}$, thereby leading to an equivalent "superquantile crossing" problem.

In summary, given Y, $\mathbf{X}$ and the homoskedastic covariates, the three quantities needed to solve the price-setting newsvendor problem (see Fig. 3 and Section 2.2) are as follows:

$$E[Y] = \hat{\hat{\boldsymbol{\beta}}}_c^T \mathbf{X}, \tag{3.15}$$

$$\mathrm{VaR}_\alpha[Y] = \hat{\boldsymbol{\beta}}_v^T \mathbf{X}, \tag{3.16}$$

$$\mathrm{CVaR}_\alpha[Y] = \hat{\boldsymbol{\beta}}_c^T \mathbf{X}, \tag{3.17}$$

where $\hat{\hat{\boldsymbol{\beta}}}_c, \hat{\boldsymbol{\beta}}_v$ and $\hat{\boldsymbol{\beta}}_c$ are outputs of subroutines that solve optimization formulations **MQR**, **QR**, and **MQR** with input quantiles $0, \alpha$ and α respectively.

3.3. Superquantile Regression (SQR)

Recently, Rockafellar et al. (2014) proposed *superquantile regression* for estimating $\mathrm{CVaR}_\alpha[Y]$ conditional on a set of covariates $\mathbf{X}$, using methods that are not based on the discretization of Eq. (2.7) as in Section 3.2. To motivate their approach, we recall that in quantile regression, $\mathrm{VaR}_\alpha[Y]$ is estimated as a linear function of the form $\boldsymbol{\beta}_v^T \mathbf{X}$ by using a suitable error measure (or loss function) for the residual $(Y - \boldsymbol{\beta}_v^T \mathbf{X})$. Similarly, Rockafellar et al. (2014) provide a suitable modification of the error measure used in quantile regression leading to the superquantile regression estimates for $\mathrm{CVaR}_\alpha[Y]$. This modified error measure is based on an auxiliary response variable whose quantiles, by construction, are equivalent to the desired superquantiles $\mathrm{CVaR}_\alpha[Y]$.

An earlier result in Rockafellar et al. (2008) provides the relationship between the three quantities, viz., the error measure, the deviation measure, and the statistic to be estimated in any regression problem. For example, in quantile regression, this error measure is the quantile error (or loss) function $\psi_\alpha(.)$ of the residual, the deviation measure is the difference between CVaR and expectation of the residual, and the statistic is the estimated quantile of the response variable. In Rockafellar et al. (2008) (Theorem 3.2), it was shown that minimizing the error measure for the residual may be decomposed into two steps. The first step is to minimize the corresponding deviation measure of the residual excluding the intercept term to estimate all coefficients but the intercept. The second step is to set the intercept term to the statistic associated with the estimated optimal residual from the first step.

This theory is extended to superquantile regression in Rockafellar et al. (2014), using the modified error measure and a corresponding deviance measure. This leads to a linear programming (LP) formulation for estimating the regression function for $\mathrm{CVaR}_\alpha[Y]$. We refer the reader to Rockafellar et al. (2014) for the full details of their methodology; however, the LP formulation for obtaining the estimates of the superquantile regression coefficients is summarized below.

We note that for any random variable R, the deviation measure for superquantile regression is defined in Rockafellar et al. (2014) as

$$\mathcal{D}_\alpha(R) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{CVaR}_\tau[R]d\tau - E[R], \tag{3.18}$$

while the corresponding statistic is simply $\mathrm{CVaR}_\alpha[R]$. Let us consider regression functions where the constant term is explicit, i.e. $\beta_0 + \boldsymbol{\beta}^T\mathbf{X}$, whilst $\mathbf{X}$ consists of columns for all the covariates except the column of ones unlike the earlier sections. Therefore, the minimization of the error measure in superquantile regression may be equivalently written as

$$\boldsymbol{\beta}_c = \arg\min_{\boldsymbol{\beta}} \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{CVaR}_\tau[Y-\boldsymbol{\beta}^T\mathbf{X}]d\tau - E[Y-\boldsymbol{\beta}^T\mathbf{X}], \tag{3.19}$$

$$\beta_{c,0} = \mathrm{CVaR}_\alpha[Y-\boldsymbol{\beta}_c^T\mathbf{X}]. \tag{3.20}$$

and the estimated superquantile regression is given by $\beta_{c,0}+\boldsymbol{\beta}_c^T\mathbf{X}$. Observe that the random variable R in Eq. (3.18) is set to the residual $Y-\boldsymbol{\beta}^T\mathbf{X}$ without the intercept in Eq. (3.19).

Model Estimation In practice, we have data samples $\{\mathbf{X}_i, Y_i\}$ for $i = 1, \dots, N$ for obtaining the estimates of the regression function for $\mathrm{CVaR}_\alpha[Y]$. Note here that $\mathbf{X}_i$ does not contain the constant 1 covariate for the intercept term. The residual random variable has a discrete support, thereby leading to a cumulative distribution function which has a piecewise constant structure.

This structure, along with the formulation of the CVaR estimation as a minimization problem in Rockafellar and Uryasev (2000), allows Problem (3.19) to be expressed as the following nonlinear mathematical program,

$$\begin{aligned}\boldsymbol{\beta}_c = \underset{\boldsymbol{\beta},U}{\arg\min} \ & \frac{1}{1-\alpha}\sum_{k=N_\alpha}^{N-1}(\kappa_k-\kappa_{k-1})U_k + \frac{1}{(1-\alpha)}\sum_{k=N_\alpha}^{N-1} a_k E[\max\{Y-\boldsymbol{\beta}^T\mathbf{X}-U_k,0\}] \\ & + \frac{1}{N(1-\alpha)}\max_{i=1\dots N}(Y_i-\boldsymbol{\beta}^T\mathbf{X}_i) - \frac{1}{N}\sum_{i=1}^{N}(Y_i-\boldsymbol{\beta}^T\mathbf{X}_i)\end{aligned} \tag{3.21}$$

In the above model, $N_\alpha = \lceil N\alpha \rceil$, and the decision variables include $\boldsymbol{\beta}_c \in \mathbb{R}^n$, $U \in \mathbb{R}^{N-N_\alpha}$ where n is the number of coefficients to be estimated except the intercept term. Further $\kappa_{N_\alpha-1}=\alpha$, and $\kappa_k = \frac{k}{N}$, which capture the various piecewise constant levels within the limits of the integration in Eq. (3.19), and $a_k = \ln(1-\kappa_{k-1}) - \ln(1-\kappa_k)$.

Linearization using additional decision variables, $V \in \mathbb{R}^{N(N-N_\alpha)}$, and $W \in \mathbb{R}$ yields the following linear program as developed in Rockafellar et al. (2014).

$$\begin{aligned}\textbf{SQR}: \ \hat{\boldsymbol{\beta}}_c = \underset{\boldsymbol{\beta},U,V,W}{\arg\min} \ & \frac{1}{1-\alpha}\sum_{k=N_\alpha}^{N-1}(\kappa_k-\kappa_{k-1})U_k + \frac{1}{N(1-\alpha)}\sum_{k=N_\alpha}^{N-1}\sum_{i=1}^{N} a_k V_{ki} && (3.22)\\ & + \frac{1}{N(1-\alpha)}W - \frac{1}{N}\sum_{i=1}^{N}(Y_i-\boldsymbol{\beta}^T\mathbf{X}_i), \\ \text{s.t.,} \quad & V_{ki} \geq Y_i - \boldsymbol{\beta}^T\mathbf{X}_i - U_k, \quad \forall\, k=N_\alpha,\dots,N-1,\ i=1,\dots,N, && (3.23)\\ & V_{ki} \geq 0, \quad \forall\, k=N_\alpha,\dots,N-1,\ i=1,\dots,N, && (3.24)\\ & W \geq Y_i - \boldsymbol{\beta}^T\mathbf{X}_i, \quad \forall\, i=1,\dots,N. && (3.25)\end{aligned}$$

The estimation procedure can then be summarized as an algorithm.

Algorithm 3 Superquantile regression for CVaR estimation

Input: Data $\{\mathbf{X}_i, Y_i\}$ for $i=1,...,N$, and the level α. Note that $\mathbf{X}_i$ does not include the constant 1 covariate corresponding to the intercept term.

1: Set up and solve the linear program **SQR**. The solution produces an estimate for $\hat{\boldsymbol{\beta}}_c$.

2: Obtain constant term, using Eq. (3.20), by computing the empirical CVaR of the residual corresponding to $\hat{\boldsymbol{\beta}}_c$. This is computable as,

$$\hat{\beta}_{c,0} = \frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} R^{(i)} + \Big(\frac{N\alpha - \lfloor N\alpha \rfloor}{N\alpha}\Big) R^{(\lceil N\alpha \rceil)},$$

where $R = (Y - \hat{\boldsymbol{\beta}}_c^T\mathbf{X})$, and $R^{(i)}$ represent the decreasing order statistics of R over the empirical sample $\{\mathbf{X}_i, Y_i\}$, $i=1,\dots,N$, i.e. $R^{(1)} \geq \dots \geq R^{(N)}$.

Output: Estimates $[\hat{\beta}_{c,0}; \hat{\boldsymbol{\beta}}_c]$ for superquantile regression, i.e., $\text{CVaR}_\alpha[Y] = \hat{\beta}_{c,0} + \hat{\boldsymbol{\beta}}_c^T\mathbf{X}$.

The methodology for the price optimization is then similar to that described in Section 3.2, except that the CVaR terms can now estimated using superquantile regression.

3.3.1. An Efficient Algorithm For Superquantile Regression We begin by observing that the above formulation **SQR** involves $\mathcal{O}(N^2)$ number of variables as well as constraints. When N is large, this quadratic complexity makes the above formulation impractical in terms of the computational time needed to solve the linear program. We instead present an alternative linearization of Eq. (3.21), **RSQR** (reformulated SQR), which enables the derivation of an efficient cutting plane algorithm. We begin with a technical observation which is then used to arrive at the alternative linear reformulation.

CLAIM 2. *Let $\mathcal{N} = \{1 \dots N\}$. Then, for any fixed index k, the following equality holds, where $\mathcal{P}(\mathcal{N})$ denotes the power set, i.e. the set of all subsets of $\mathcal{N}$.*

$$\sum_{i=1}^{N} \max\{Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k, 0\} = \max_{J_k \in \mathcal{P}(\mathcal{N})} \sum_{i \in J_k} (Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k) \tag{3.26}$$

Proof. Consider the subset $J_k^* = \{i \in \mathcal{N} \mid Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k > 0\}$. Then, we have

$$\sum_{i=1}^{N} \max\{Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k, 0\} = \sum_{i \in J_k^*} \max\{Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k, 0\} \tag{3.27}$$

since any index $i \notin J_k^*$ contributes zero to the summation on the left hand side. Similarly, it is also evident that,

$$J_k^* \in \arg\max_{J_k \in \mathcal{P}(\mathcal{N})} \sum_{i \in J_k} (Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k) \tag{3.28}$$

This is because any subset, say, $J_k \subset J_k^*$ can be augmented with elements from $J_k^* \setminus J_k$ to strictly increase the objective function, while no superset $J_k \supset J_k^*$ can possibly increase the objective function relative to J_k^* due to its definition. Taken together, Eqs. (3.27–3.28) lead to Eq. (3.26). $\square$

Expressing the expectation in Eq. (3.21) as a finite summation and applying Eq. (3.26) leads to the following formulation.

$$\hat{\boldsymbol{\beta}}_c = \arg\min_{\boldsymbol{\beta}, U, W} \; \frac{1}{1-\alpha} \sum_{k=N_\alpha}^{N-1} (\kappa_k - \kappa_{k-1}) U_k + \frac{1}{N(1-\alpha)} \sum_{k=N_\alpha}^{N-1} a_k \max_{J_k \in \mathcal{P}(\mathcal{N})} \sum_{i \in J_k} (Y_i - \boldsymbol{\beta}^T \mathbf{X}_i - U_k) \tag{3.29}$$

$$+ \frac{1}{N(1-\alpha)} W - \frac{1}{N} \sum_{i=1}^{N} (Y_i - \boldsymbol{\beta}^T \mathbf{X}_i),$$

$$\text{s.t.,} \quad W \geq Y_i - \boldsymbol{\beta}^T \mathbf{X}_i, \quad \forall\, i = 1, \dots, N. \tag{3.30}$$

We may then linearize the above formulation using an exponential number of constraints, with only an $\mathcal{O}(N)$ number of additional variables, T_k, as follows.

$$\mathbf{RSQR}: \; \hat{\boldsymbol{\beta}}_c = \underset{\boldsymbol{\beta},U,W,T}{\arg\min} \; \frac{1}{1-\alpha}\sum_{k=N_\alpha}^{N-1}(\kappa_k - \kappa_{k-1})U_k + \frac{1}{N(1-\alpha)}\sum_{k=N_\alpha}^{N-1} a_k T_k \tag{3.31}$$

$$+ \frac{1}{N(1-\alpha)}W - \frac{1}{N}\sum_{i=1}^{N}(Y_i - \boldsymbol{\beta}^T\mathbf{X}_i),$$

$$\text{s.t.,} \quad T_k \geq \sum_{i\in J_k} Y_i - \boldsymbol{\beta}^T\mathbf{X}_i - U_k, \quad \forall\; J_k \in \mathcal{P}(\mathcal{N}),\; k = N_\alpha,\ldots,N-1, \tag{3.32}$$

$$T_k \geq 0, \qquad \forall\; k = N_\alpha,\ldots,N-1, \tag{3.33}$$

$$W \geq Y_i - \boldsymbol{\beta}^T\mathbf{X}_i, \qquad \forall\; i = 1,\ldots,N. \tag{3.34}$$

The proposed reformulation allows us to derive an efficient decomposition algorithm. We first present the following claim that lets us successfully seed the following algorithm in its very first iteration. Let $\mathbf{RSQR} - \mathbf{RELAX}$ denote a relaxation of $\mathbf{RSQR}$, where we replace $\mathcal{P}(\mathcal{N})$ in constraint Eq. (3.31) with respective subsets $\mathcal{J}_k \subseteq \mathcal{P}(\mathcal{N})$, for each index k.

Claim 3. *Let $\mathcal{J}_k = \{\mathcal{N}\}, \forall\; k$. Then, the corresponding relaxation,* $\mathbf{RSQR} - \mathbf{RELAX}$ *is a bounded linear program.*

Proof. We firstly note that constructing a finite, feasible solution for the corresponding dual LP is sufficient to establish boundedness of the above LP, due to weak duality. The corresponding dual LP is:

$$\max_{p,q} \quad \sum_{i=1}^{N} p_i Y_i + \sum_{k=N_\alpha}^{N-1}\sum_{i=1}^{N} Y_i q_k \tag{3.35}$$

$$\text{s.t.,} \quad q_k \leq \frac{a_k}{N(1-\alpha)}, \qquad \forall\; k = N_\alpha,\ldots,N-1, \tag{3.36}$$

$$q_k = \frac{\kappa_k - \kappa_{k-1}}{N(1-\alpha)}, \qquad \forall\; k = N_\alpha,\ldots,N-1, \tag{3.37}$$

$$\sum_{i=1}^{N} p_i = \frac{1}{N(1-\alpha)}, \tag{3.38}$$

$$\sum_{i=1}^{N} p_i X_{i,l} + \sum_{k=N_\alpha}^{N-1}\sum_{i=1}^{N} X_{i,l} q_k = \frac{1}{N}\sum_{i=1}^{N} X_{i,l}, \quad \forall\; l = 1,\ldots,n, \tag{3.39}$$

$$p_i, q_k \geq 0, \qquad \forall\; i = 1,\ldots,N, \quad k = N_\alpha,\ldots,N-1. \tag{3.40}$$

Consider the candidate solution that evidently satisfies the non-negativity constraints, as well as constraints Eq. (3.37) and Eq. (3.38).

$$\tilde{p}_i = \frac{1}{N^2(1-\alpha)}, \quad \tilde{q}_k = \frac{\kappa_k - \kappa_{k-1}}{N(1-\alpha)} \tag{3.41}$$

Constraint Eq. (3.36) is satisfied because, using a series expansion for the natural logarithm (where each $|\kappa_k| < 1$), we have

$$a_k = \ln(1-\kappa_{k-1}) - \ln(1-\kappa_k) = \sum_{j=1}^{\infty} \frac{\kappa_k^j - \kappa_{k-1}^i}{j} > \kappa_k - \kappa_{k-1} \tag{3.42}$$

The final constraint Eq. (3.39) is also satisfied as verifiable via substitution, where $\forall\ l = 1, \ldots, n$, we have,

$$\begin{aligned}
\text{LHS} &= \sum_{i=1}^{N} X_{i,l} \left(\frac{1}{N^2(1-\alpha)} + \sum_{k=N_\alpha}^{N-1} \frac{\kappa_k - \kappa_{k-1}}{N(1-\alpha)} \right) \\
&= \sum_{i=1}^{N} X_{i,l} \left(\frac{1}{N^2(1-\alpha)} + \frac{1}{N(1-\alpha)} \left(\frac{N-1}{N} - \alpha \right) \right) \\
&= \sum_{i=1}^{N} X_{i,l} \left(\frac{1}{N} \right) = \text{RHS}.
\end{aligned}$$

We also note that feasibility of the above relaxed primal, namely **RSQR** − **RELAX** is self-evident. Taken together, these imply a finite, non-empty optimal solution for **RSQR** − **RELAX**. □

The decomposition algorithm may then be presented as in Algorithm 4.

Theorem 1. *Algorithm 4 converges in finite time and solves problem* **RSQR***, equivalently* **SQR***, upon convergence.*

Proof. Step 2 is guaranteed to result in a finite, non-empty solution in the very first iteration, due to Claim 3 and thereby successfully seeds the delayed constraint-generation procedure. Finite convergence is guaranteed due to the finiteness of the power set, $\mathcal{P}(\mathcal{N})$. Convergence in the LP solution is achieved in step 3, when no new constraint is identifiable for each index k, i.e. J_k^* is already present in $\mathcal{J}_k$. Upon such convergence, it can be seen that the (final) converging linear program **RSQR** − **RELAX** is a relaxation of **RSQR** with respect to the representation of Eq. (3.31), but it also satisfies all the unrepresented constraints from **RSQR**. Thereby, its solution is also optimal for **RSQR**, and equivalently **SQR**. □

In practice, convergence is realized in far fewer iterations than the cardinality of the power set. We present below an empirical investigation of the computational performance of the algorithm.

3.3.2. Computational Results We consider the heteroskedastic generating model **G1** described in Section 4.1 below with a unit Normal error distribution. Fig. 4 shows the average computational time needed to solve formulation **SQR** using Algorithm 3, as well as the proposed reformulation using Algorithm 4, as a function of N, namely, the number of sample points at two quantile levels $\alpha = 0.7, 0.85$. For each value of N and α, the plot shows the mean value and error bars estimated over 200 independently generated data sets, each of size N. All computations

Algorithm 4 Decomposition Algorithm for Superquantile regression used for CVaR estimation

Input: Data $\{\mathbf{X}_i, Y_i\}$ for $i = 1, ..., N$, and the level α. Note that $\mathbf{X}_i$ does not include the constant covariate corresponding to the intercept term.

1: Initialize $\mathcal{J}_k = \{\mathcal{N}\}$, $\forall\ k = N_\alpha, \ldots, N-1$.

2: Solve the relaxed linear program $\mathbf{RSQR-RELAX}$ implied by the current value of $\mathcal{J}_k$. Let $\hat{\boldsymbol{\beta}}$ and $\hat{U}_k$ be the optimal solution values for these variables.

3: For each k, identify the most violated constraint, relative to the full set of constraints in Eq. (3.31). This is computable in $\mathcal{O}(N)$ effort for each k as

$$J_k^* = \{i \in \mathcal{N} \mid Y_i - \hat{\boldsymbol{\beta}}^T \mathbf{X}_i - \hat{U}_k > 0\},$$

and do $\mathcal{J}_k = \mathcal{J}_k \cup J_k^*$. This results in adding a constraint for each k.

4: Repeat steps 2-3, until convergence of the LP solution in step 2. Let the converged solution for variable $\boldsymbol{\beta}$ be denoted as $\hat{\boldsymbol{\beta}}_c$.

5: Obtain constant term, using Eq. (3.20), by computing the empirical CVaR of the residual corresponding to $\hat{\boldsymbol{\beta}}_c$. This is computable as,

$$\hat{\beta}_{c,0} = \frac{1}{N\alpha} \sum_{i=1}^{\lfloor N\alpha \rfloor} R^{(i)} + \Big(\frac{N\alpha - \lfloor N\alpha \rfloor}{N\alpha}\Big) R^{(\lceil N\alpha \rceil)},$$

where $R = (Y - \hat{\boldsymbol{\beta}}_c^T \mathbf{X})$, and $R^{(i)}$ represent the decreasing order statistics of R over the empirical sample $\{\mathbf{X}_i, Y_i\}$, $i = 1, \ldots, N$, i.e. $R^{(1)} \geq \cdots \geq R^{(N)}$.

Output: Estimates $[\hat{\beta}_{c,0}; \hat{\boldsymbol{\beta}}_c]$ for superquantile regression, i.e., $\mathrm{CVaR}_\alpha[Y] = \hat{\beta}_{c,0} + \hat{\boldsymbol{\beta}}_c^T \mathbf{X}$.

were carried out using Matlab/CPLEX on 64-bit Macbook Pro, Intel®Core™ i7 @ 2.5 GHz, 16 GB RAM. While Algorithm 3 fails to acceptably scale beyond a five hundred sample points, the proposed decomposition algorithm performs well even for really large sets.

In our experiments, we observe that it is often more stable to implement $\mathbf{RSQR-RELAX}$ in terms of the scaled covariate $\hat{\mathbf{X}} = (\mathbf{X} - \mu_{\mathbf{X}})/\sigma_{\mathbf{X}}$ without any loss of generality. For some instances of the optimization problem with $N = 5000$, and with greater frequency for larger values of N and α, we observe that CPLEX can sometimes throw an exception of the form: 'optimal solution is available, but with infeasibilities after unscaling' in the unscaled problem, whereas the scaled problem does not encounter these difficulties. Many commercial regression automatically incorporate this form of scaling for better numerical stability, and a similar practice in for Algorithm 4 also seems to be quite beneficial.

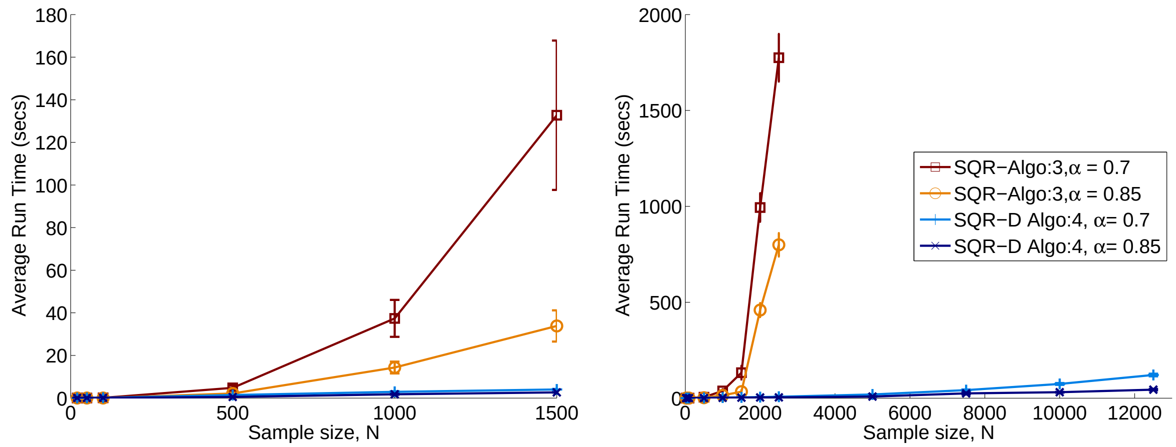

Figure 4 Computational run times for Algorithm 3 and Algorithm 4 as a function of N for $\alpha = 0.7, 0.85$. The left plot is zoomed version of the right plot for the range [0, 1500].

4. Monte-Carlo Simulation Study

In this section, the proposed methodologies for the data-driven, price-setting newsvendor problem described in Section 3, are evaluated through a Monte Carlo simulation study. To fix ideas, we focus on the lost-sales formulation of the price-setting newsvendor problem, although the general conclusions carry over to the emergency order formulation as well.

If the stochastic price-demand function is explicitly known, then the exact optimal solutions to the price-setting newsvendor problem can be directly obtained (e.g., using the methods in Section 2.2). Furthermore, simulated data sets can be generated from these known stochastic price-demand functions, and the estimated optimal solutions for each simulated data set can be obtained using the methods in Section 3. These results can be used to evaluate the statistical properties of the estimated optimal solutions, as well as the coverage and length of their estimated bootstrap confidence intervals, as described further below.

4.1. Simulated Data Sets

Denoting the stochastic price-demand functions by Y, we consider two such explicit functions which are motivated from Eqs. (3.1) and (3.8) respectively (for simplicity of exposition, price is the only demand driver that is included)

$$\textbf{G1.} \qquad Y = \beta_0 + \beta_1 p + \left(\gamma_0 + \gamma_1 p + \gamma_2 p^2\right)\,\epsilon, \tag{4.1}$$

where ϵ is a random variable with mean 0 that is specified further below, and $\beta_0 = 200.0,\ \beta_1 = -35.0,\ \gamma_0 = 36.0,\ \gamma_1 = -12.0,\ \gamma_2 = 2.1.$

$$\textbf{G2.} \qquad Y = \beta_0 + \beta_1 p + \beta_2 p^2 + (\gamma_0 + \gamma_1 p)\ \epsilon^- + \gamma_2 p^2\ \epsilon^+, \tag{4.2}$$

where $\epsilon^- = \min\{\epsilon, 0\}$, $\epsilon^+ = \max\{0, \epsilon\}$ and ϵ is $N(0,1)$ i.e., a Normal distribution with mean 0 and standard deviation 1. Here, $\beta_0 = 215.0$, $\beta_1 = -37.0$, $\beta_2 = -1.5\ \mathrm{CVaR}_\epsilon(0.5) = -1.1968$, $\gamma_0 = 36.0, \gamma_1 = -4.0, \gamma_2 = 3$.

The demand functions in the generating models **G1** and **G2** have means that are decreasing linear functions of price, and variances that are non-monotonic quadratic functions of price (the mean for the generating model **G2** is obtained by evaluating $\mathrm{CVaR}_0[Y]$). As discussed in Section 1.2, price-demand functions with these characteristics are of practical importance.

The parameters for the price-setting newsvendor problem are taken to be $c = 1.0$, $s = 0.5$, and $v = 1.0$. The unit price p, which is the decision variable, is constrained to the interval $(1.5, 4.0)$. These parameter values are inspired by an example in Lau and Lau (1988), although that paper only considered a homoskedastic demand models with normal errors.

For the random variable ϵ in the generating model **G1**, we consider the following distributions:

1. Normal(0,1): Normal distribution with mean 0 and standard deviation 1.
2. Gamma(2,1): Gamma distribution with shape 2 and rate 1 (equivalently with mean 2 and standard deviation $\sqrt{2}$), recentered to have mean 0.
3. Log-normal(0,1): Log-normal distribution with mean 0 and standard deviation 1 on the variable's log-scale, recentered to have mean 0.
4. Student's t(3): Student's t-distribution with 3 degrees of freedom, with mean 0 and and standard deviation $\sqrt{3}$.
5. Mixture(-2,2): Mixture of two normal distributions, $N(2,1)$ and $N(-2,1)$, with equal weight and standard deviations 1 each.

The Gamma(2) and Log-normal(0,1) distributions which are re-centered to have mean 0, are asymmetric distributions. The Student's t(3) distribution is symmetric but is heavy-tailed. The Mixture(2,-2) distribution is also symmetric but is bi-modal unlike the other distributions that are considered in this study.

The exact optimal solutions to the price-setting newsvendor problem for the price-demand functions in generating model **G1** and **G2** are given in Table 1. The corresponding sample estimators for the quantities in Table 1, are respectively denoted by $\hat{p}^*$ for optimal price, $\hat{x}^*$ for optimal order quantity, and $\hat{\Pi}^*$ for optimal profit.

The number of data points in each simulated data set is denoted by N and our results are obtained for values ranging from $N = 50$ to $N = 1500$. The covariate values for p for the individual cases in each simulated data set are obtained by uniform sampling from the allowed range in the

Generating Model	Distribution for ϵ	Optimal Solutions		
		Price, p^*	Order Quantity, x^*	Profit Π^*
G1	Normal	3.32	105.57	178.74
	Gamma	3.28	114.77	167.76
	Lognormal	3.22	113.60	155.85
	Student T	3.28	111.5	169.58
	Mixture	3.34	134.18	184.41
G2	Normal	3.16	119.05	169.04

Table 1 True optima for the lost sales price-setting newsvendor problem with the stochastic demand model.

interval $(1.5, 4.0)$. The number of simulated data sets used in the Monte Carlo evaluation is denoted by N_{mc} and is chosen to be 200.

4.2. Data-driven techniques used for experiments

As discussed in Section 3, any appropriate combination of the techniques for estimating the three quantities $E[D(p,\mathbf{z})]$, $\text{VaR}_\alpha[D(p,\mathbf{z})]$, and $\text{CVaR}_\alpha[D(p,\mathbf{z})]$ completely specify the optimization formulation in Eq. (2.3) and Eq. (2.5) and hence can be used for testing. This leads to a profusion of ways for implementing the computations as shown in Table 2 (just based on the methods described in this paper). In the table, GLR refers to generalized linear regression based on Algorithm 2, Residuals refers to the outputs of GLR combined with Eqs. (3.3–3.4), QR and MQR refer to the formulations **QR** and **MQR** respectively while SQR refers to the decomposition approach presented in Algorithm 4. The subscripts 0 or α represent the input quantile levels associated with the estimation procedures.

$E[D(p,\mathbf{z})]$	$\text{VaR}_\alpha[D(p,\mathbf{z})]$	$\text{CVaR}_\alpha[D(p,\mathbf{z})]$
GLR MQR_0 SQR_0	GLR+Residuals_α QR_α	GLR + Residuals_α MQR_α SQR_α

Table 2 Possible statistical estimation methods to estimate the different quantities of interest

To fix ideas, we implement and compare the following techniques:

- For the generating model, **G1**, we use the GLR method to evaluate $E[D(p,\mathbf{z})]$ and compare three methods that differ in the way they compute $\text{CVaR}_\alpha[D(p,\mathbf{z})]$ and $\text{VaR}_\alpha[D(p,\mathbf{z})]$ respectively. The first method uses the residuals of the GLR for the superquantile and quantile estimations, the second method uses MQR and QR and the last method uses SQR and QR respectively.

The reason we choose the same mean estimator across the three different methods is to do with many reasons: (1) Eq. (4.1) has the same form as Eq. (3.1) and Eq. (3.2) (i.e., set $\mu = \beta_0 + \beta_1 p$, $\sqrt{\phi} = \gamma_0 + \gamma_1 p + \gamma_2 p^2$, $V(\mu) = 1$, with $g(\mu)$ being the identity link function, and $h(\phi)$ being the square-root link function). (2) The mean coefficient retrieval is expected to be pretty good if the weights are a good approximation in the weighted least squares subroutine (step 2 when $V(\mu) = 1$) in GLR (weighted least squares is BLUE i.e., the best linear unbiased estimator). (3) A common mean estimator across methods also enables one to focus on the quality of the superquantile and the quantile estimation methods.

- For the generating model, **G2**, we use two different mean estimation methods: GLR and MQR_0 and compare these against the three different ways described above to estimate the superquantile and quantile (GLR residuals, MQR-QR and SQR-QR).

In our implementations, we define the GLR algorithm to have successful convergence if the number of iterations of the Algorithm 2 is less than 50. For the MQR method we use a simple uniform discretization where $\Delta = 0.01$.

4.3. Performance Metrics - Perfect Hindsight

In this simulation study, our goal is to understand the performance of the data-driven techniques in retrieving the true optimal price and the true optimal order quantity. To avoid information overload we only focus on the effect of the estimated price and order quantity on the true realized objective that we denote by 'Realized $\text{Profit}_{i,m}$' for an finite sample instance i using the method m. We compare this realized objective against the maximum realizable objective, denoted by 'Profit^*' had we offered the true optimal price, p^* and stocked the true optimal order quantity, x^*. This method is often called the perfect hindsight method because it compares the realizable objective attained by the estimation method using finite data against the maximum realizable profit as though in hindsight one can achieve the latter (within some tolerance) with sufficient data.

We use the mean absolute error as a measure of error between the two objectives i.e.,

$$\text{MAE}_{i,m} = \frac{\text{Profit}^* - \text{Realized Profit}_{i,m}}{\text{Profit}^*} \tag{4.3}$$

We estimate this measure of error for every finite sample Monte Carlo data set i using a variety of methods m described in Section 4.2. We present the mean and standard error of the $\text{MAE}_{i,m}$ over all the instances i for each method m.

4.4. Results

We summarize our results about the mean and standard error of $\text{MAE}_{i,m}$ in Table 3 and Table 4. Table 3 focuses on generating model **G1** and presents the results for various distributions of error

N	Normal				Gamma			
	Instances	GLR	SQR	MQR	Instances	GLR	SQR	MQR
50	198	**0.916** (0.101)	1.099 (0.112)	1.010 (0.089)	192	**2.077** (0.208)	2.400 (0.220)	2.435 (0.232)
100	200	**0.410** (0.029)	0.478 (0.036)	0.477 (0.036)	200	**0.752** (0.064)	0.875 (0.074)	0.847 (0.075)
250	200	**0.146** (0.011)	0.171 (0.013)	0.171 (0.013)	200	**0.266** (0.020)	0.312 (0.022)	0.303 (0.021)
500	200	**0.081** (0.006)	0.095 (0.007)	0.095 (0.007)	200	**0.137** (0.011)	0.165 (0.013)	0.161 (0.013)
1000	200	**0.036** (0.003)	0.041 (0.003)	0.041 (0.003)	200	**0.069** (0.005)	0.082 (0.007)	0.082 (0.007)
1500	200	**0.024** (0.002)	0.029 (0.002)	0.029 (0.002)	200	**0.043** (0.003)	0.047 (0.003)	0.046 (0.003)

N	Student t				Lognormal			
	Instances	GLR	SQR	MQR	Instances	GLR	SQR	MQR
50	177	2.117 (0.216)	1.889 (0.177)	**1.816** (0.182)	157	3.160 (0.472)	3.282 (0.356)	**2.989** (0.339)
100	191	1.123 (0.118)	0.938 (0.091)	**0.926** (0.094)	185	1.561 (0.154)	1.805 (0.207)	**1.373** (0.147)
250	198	0.430 (0.033)	**0.334** (0.023)	0.336 (0.024)	186	0.560 (0.046)	0.565 (0.089)	**0.446** (0.035)
500	197	0.282 (0.024)	0.181 (0.014)	**0.177** (0.014)	194	0.342 (0.031)	0.248 (0.023)	**0.245** (0.023)
1000	198	0.120 (0.009)	**0.072** (0.005)	**0.072** (0.005)	197	0.159 (0.013)	0.123 (0.011)	**0.122** (0.011)
1500	197	0.106 (0.014)	**0.052** (0.003)	0.053 (0.004)	198	0.141 (0.013)	**0.087** (0.007)	**0.087** (0.007)

N	Mixture			
	Instances	GLR	SQR	MQR
50	198	5.606 (0.312)	5.064 (0.301)	**4.967** (0.301)
100	200	3.636 (0.252)	3.464 (0.233)	**3.450** (0.236)
250	200	2.295 (0.149)	2.212 (0.128)	**2.210** (0.129)
500	200	1.378 (0.072)	1.273 (0.067)	**1.271** (0.067)
1000	200	1.036 (0.055)	**0.970** (0.048)	0.972 (0.048)
1500	200	0.748 (0.047)	0.687 (0.041)	**0.685** (0.041)

Table 3 Mean and standard error (in brackets) for the MAE of the profit using the generating model **G1**

N	Instances	GLR mean			MQR mean		
		GLR	SQR	MQR	GLR	SQR	MQR
50	189	2.059 (0.238)	2.139 (0.241)	2.264 (0.271)	2.062 (0.24)	2.082 (0.253)	**2.014** (0.235)
100	199	0.681 (0.062)	0.704 (0.056)	0.698 (0.058)	**0.648** (0.054)	0.664 (0.051)	0.678 (0.053)
250	200	0.3 (0.02)	0.235 (0.017)	0.231 (0.017)	0.301 (0.021)	**0.222** (0.016)	0.225 (0.016)
500	200	0.235 (0.015)	0.141 (0.012)	0.138 (0.012)	0.237 (0.015)	**0.136** (0.012)	0.14 (0.012)
1000	200	0.144 (0.007)	0.059 (0.004)	0.058 (0.004)	0.143 (0.007)	**0.056** (0.004)	0.058 (0.004)
1500	200	0.131 (0.005)	0.037 (0.003)	0.037 (0.003)	0.131 (0.005)	**0.036** (0.003)	0.038 (0.003)

Table 4 Mean and standard error (in brackets) for the MAE of the profit using the generating model **G2**

discussed in Section 4.1 and three different estimation procedures discussed in Section 4.2. Table 3 focuses on generating model **G2** and presents the results for six different estimation methods. In each of the tables the first column with the title 'N' refers to the number of data points in each instance of the data set. There were $N_{mc} = 200$ random instances generated for each data set. The second column in each table that has a title 'Instances' denotes the number of instances, amongst the $N_{mc} = 200$, where the GLR algorithm successfully converged. Our mean and standard error results are presented only on the instances where the GLR algorithm successfully converged.

In Table 3 and Table 4, we boldface that method that has the highest mean performance for better (or similar) standard error levels. It can be observed from Table 3 that for generating model **G1** the GLR method tends to outperform the MQR and SQR for the Normal and Gamma distributions especially as N becomes larger. The reverse is true, i.e., GLR tends to underperform compared to MQR and SQR for the other distributions such as Student t, Lognormal and Mixture for larger N. In Table 4 for generating model **G2** again the SQR and MQR methods with either mean estimation method tends to dominate over the GLR based method. This is surprising because generating model **G2** only employs a Normal distribution of error and for generating model **G1** under a Normal distribution GLR method dominated over the MQR and SQR methods. The MQR mean estimation method here performs slightly better than the GLR mean estimation method. For smaller data sets, no method statistically dominates another method although some methods seem to have smaller mean for the same standard errors. We also note that across all the results that it is harder to distinguish the MQR and the SQR methods as their mean performances are very similar with near identical standard error levels. We do see that MQR tends to slightly outperform

SQR for smaller sized data sets in generating model **G1** and SQR tends to slightly outperform in generating model **G2**.

4.5. Discussion

In this section, we summarize our thoughts and lessons learnt from extensive experimentation.
Performance in Estimation: Our experimental results in the context of the price-setting newsvendor show that SQR and MQR result in better solutions for a wide range of generating models over GLR. GLR performs best in Normal and Gamma distributions but if the error distributions are highly asymmetric or heavy tailed or bi-modal (even though symmetric) or possess heteroskedastic effects that cannot fully be explained by a variance predictor (i.e., quantiles that depend differently on the different covariates aside from the effects of the noise), a mean-variance model captured by GLR may not have the best performance. Sometimes, GLR can even fail to converge, more so in some distributions over others.

Our observations seem different from Chun et al. (2012) who compared OLS and MQR methods for superquantile estimation in a homoskedastic setting (GLR is OLS in a homoskedastic setting). The key in a homoskedastic setting to capture the mean accurately while the quantiles and superquantiles capture the effect of the empirical noise distribution from the residuals. Unlike this, in a heteroskedastic model (e.g., models **G1** and **G2**) it is not only important to capture the mean, but also the other quantities accurately. We believe this is where quantile based methods like QR, MQR and SQR are powerful as highlighted in the experiments for generating model **G1**. In summary, for quantile or superquantile estimations, unless one expects the error distributions to have a unimodal, symmetric, non-heavy-tailed or homoskedastic behaviro (i.e., similar to the *Normal* distribution) it is always better to work with quantile-based methods such as QR, MQR or SQR.

For the mean estimation, the experiments for generating model **G2** indicate that quantile-based methods such as MQR and SQR are preferred to GLR whenever the quantiles have different dependence on the covariates (aside from the effects of the noise distribution). However, when this is not the case as in generating model **G1**, methods like GLR outperform MQR and SQR (even in run-time) and we believe this stems from the BLUE property of weighted least squares.

Between MQR and SQR, SQR has some very interesting theoretical properties in terms of risk measure and being part of the risk quadrangle but statistically it is hard to distinguish MQR and SQR.

Based on the above discussion, we gather that it is important the user of these techniques understands the data and uses the insights from the application area together with some of our conclusions to gauge the best technique that suits the data.

Run time comparisons: We focus on the run-times of the individual algorithms of GLR, QR, MQR and SQR as opposed to the overall time for solving a data-driven price-setting newsvendor problem. This is because the price optimization for lost sales requires a non-linear algorithm and this can call the CVaR estimation methods multiple times. We consider the heteroskedastic generating model **G1** with a unit Normal error distribution. Fig. 5 shows the average computational time needed to solve the different estimation methods as a function of N for two different α values (i.e., 0.7 and 0.85). For each value of N and α, the plot shows the mean value of the run times and the corresponding error (i.e., standard deviation) bars estimated over 200 independently generated data sets, each of size N.

GLR as a practical technique for larger data sets is much faster and grows at a negligible rate compared to MQR and SQR at any quantile level, and even more so at quantiles closer to 0. Observe also that the standard deviation of this method is very small. The time taken for MQR in comparison to SQR is in the same order and a bit higher in many cases except for $\alpha = 0.7$ and large N. The run time of MQR can be tuned up or down by decreasing or increasing Δ which is currently set to 0.01. This reduces the number of discretizations over which the quantiles are estimated. The decomposition method of SQR on the other hand uses the finest discretization that can be generated with the residual data set. Observe also that the standard deviations of MQR and SQR to increase for larger N (and smaller α for SQR in particular).

To estimate a single quantile QR is a fast practical routine with similar performance guarantees like GLR but if GLR is already used for the mean evaluation (say as in the context of the price setting newsvendor problem), estimating quantiles or superquantiles from the residual distribution after the main GLR routine is just a few additional algebraic operations.

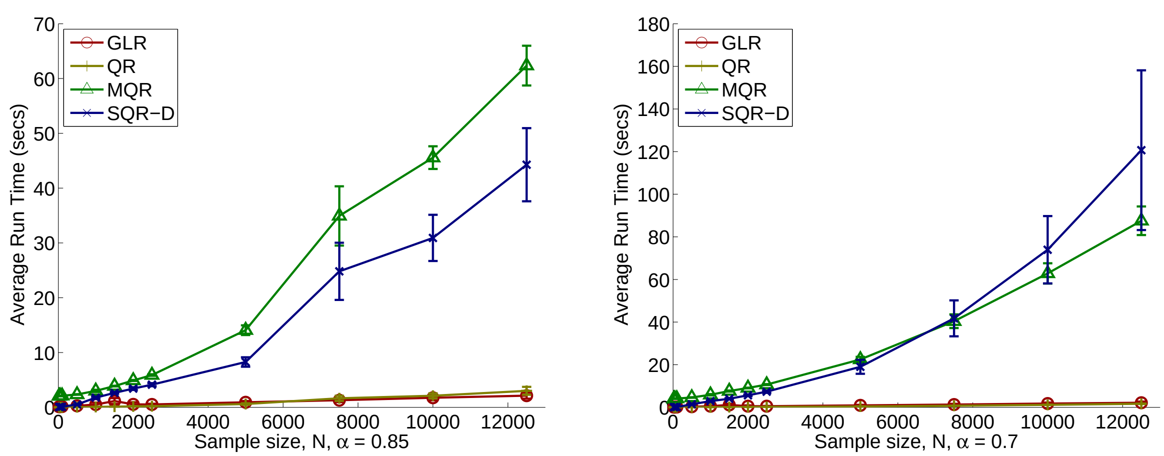

Figure 5 Computational run times for GLR, QR, MQR and SQR-Decomposition as a function of N for $\alpha = 0.85$ and 0.7 respectively. Note the difference in scales in the two plots.

5. Extensions and enhancements for the operational use of the data-driven price setting newsvendor

Statistical Consistency: There are other statistical performance measures that are often used to gauge the performance of an estimation technique in a Monte Carlo setting such as consistency, bias, mean square error and relative efficiency of estimators as the size of the dataset N becomes large (see Keeping (1962) for definitions). These metrics are important to compare the statistical properties different estimators, and fore example, the consistency of CVAR estimation using quantile regression has previously been studied by Leorato et al. (2012) and Chun et al. (2012).

The data-driven price setting newsvendor problem as highlighted in Fig. 3, is seen in this paper to comprise of multiple estimation methods as well as price optimization step. Therefore the consistency (and any other related metrics) depends on the individual estimators, and on their combination in the algorithms. Although this is an interesting direction to pursue both theoretically and empirically, in order to limit the scope of this paper, we just provide one example of the experimental evaluation of the statistical consistency in Fig. 6. The concentration of the density plots for the Monte Carlo estimates of $\hat{\Pi}^*$ for large N (here, for the case when GLR is used for mean, quantile and the superquantile regression) indicates that this approach leads to statistically consistent estimates for the optimal solutions.

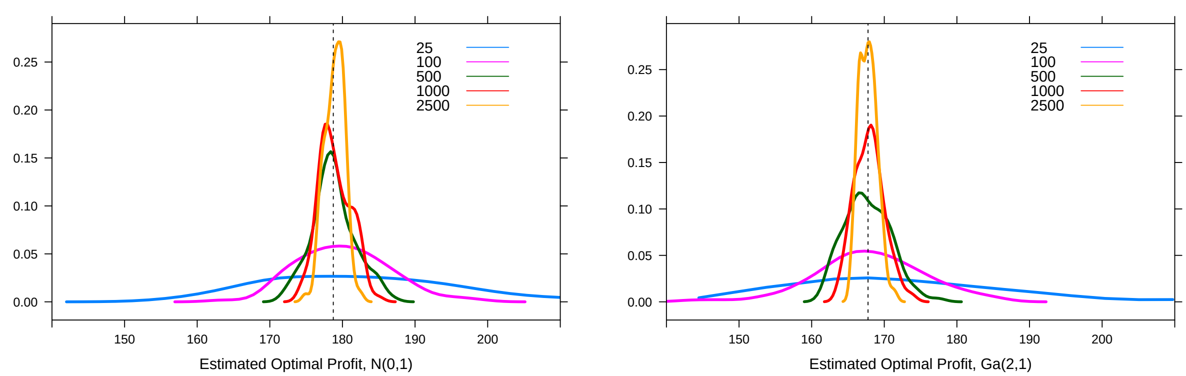

Figure 6 Consistency results for the estimated optimal profit using Heteroscedastic Regression. The distribution of the estimated optimal profit from the Monte Carlo data sets from the generating model **G1** is shown a function of the sample size N.

Bootstrap Confidence Intervals: An important aspect of a practical data-driven methodology for the price-setting newsvendor problem is obtaining confidence intervals for the estimated optimal solutions. We propose to use the non-parametric bootstrap (Efron and Tibshirani 1994) to obtain the desired confidence intervals. The overall approach is similar to that used for heteroskedastic quantile regression (Koenker 2005), and in particular, we use the "paired bootstrap" whereby the bootstrap data sets are generated by sampling entire individual cases with replacement from the original data set.

The coverage and accuracy of the resulting bootstrap confidence intervals can be evaluated through a Monte Carlo simulation study for an explicitly-known stochastic price-demand function. Fig. 7 shows the results of the Monte Carlo evaluation of the 95% bootstrap confidence intervals of the estimated optimal profit $\hat{\Pi}^*$ over $N_{\text{mc}} = 100$ instances using GLR based estimators for mean, quantile and superquantile respectively. For each of the Monte Carlo instance, the 95% bootstrap confidence intervals are presented. The coverage of the bootstrap method is evaluated as the fraction of simulated Monte Carlo data sets for which the exact optimal profit is within the 95% bootstrap confidence intervals for the corresponding estimated optimal profit. Fig. 7 shows that these bootstrap confidence intervals are quite adequate.

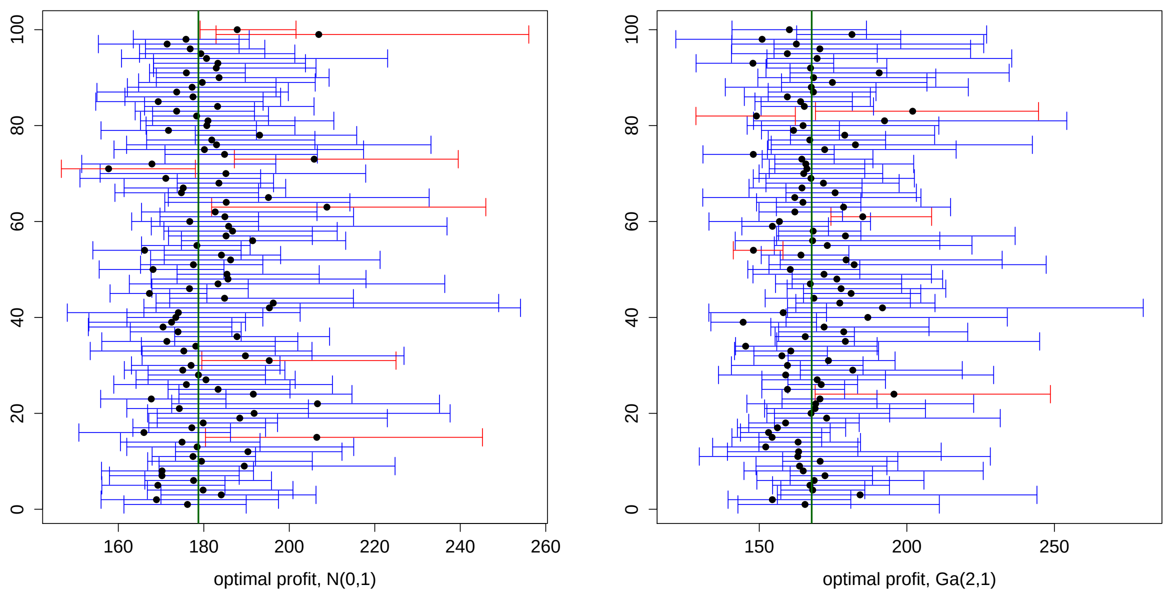

Figure 7 Coverage results for the 95% bootstrap confidence intervals for the estimated optimal profit obtained using heteroscedastic regression for the sample size $N = 50$. The confidence intervals are shown along with the exact optimal profit in each case for $N_{\text{mc}} = 100$ Monte Carlo simulations (the intervals shown marked in red do not contain the exact optimal profit).

Regularization and Feature Selection: Standard techniques for as regularization and feature selection can also be used for each individual estimation method within our framework. These techniques are particularly useful in the big data setting for high-dimensional covariate spaces, since they lead to stable regression estimates with better predictive power. The regression formulation can be generalized to include use regularization terms such as the Lasso L1 norm, Ridge L2 norm, etc., or to even include explicit constraints on the number of features in the estimated regression model. For example, Rudin and Vahn (2014) describe several regularization approaches for quantile regression in the context of the standard newsvendor problem. They also use regularization in the context of multiple items to ensure some features have similar role on predicting demand for multiple items or for the cold start demand estimation problem in new items using other items. Similar regularization terms can be included in each of our statistical estimation methods in these respective settings. Bertsimas and Mazumder (2014) consider the subset selection problem for obtaining the best feature subsets using mixed-integer optimization in the case of linear regression; they provide constraints that can incorporated into the estimation methods considered in this paper.

6. Concluding Remarks

In this paper, we described a practical framework for data-driven, distribution-free, multivariate modeling of the price-setting newsvendor problem. It includes statistical estimation and price optimization methods for estimating the optimal solutions conditional on a broad set of covariates (other than price) and associated confidence intervals. In contrast to the current state of the art for solving the data-driven, price-sensitive newsvendor problem, our framework does not require the complete price-dependent demand distribution prior to the optimization. Relevant statistical estimation methods are carried out only for quantities that are necessary to solve the decision problem, spanning a broad set of covariates including auxiliary measurements (like weather related variables, for e.g.) in addition to primary covariates (like price). These include the following three aspects of the demand distribution, namely the mean, quantile and superquantile (CVaR). We also investigated different statistical estimators which are broadly based on GLR, MQR, and SQR respectively. Our detailed simulations and computational experiments indicate that quantile-based methods such as MQR and SQR provide better solutions for a wide range of demand distributions, although for certain location-scale demand distributions that are similar to the *Normal* distribution, GLR may be preferable.

In this paper, we also presented a novel large-scale decomposition method that is exact and computationally efficient for SQR. We also extended the MQR estimation formulation to allow conditional homoskedasticity and conditional heteroskedasticity over respectively desired subsets of covariates. These CVaR estimation extensions may be of independent interest, e.g., in financial applications.

We expect that many ideas proposed in this paper will useful for developing a data-driven framework for similar or other advanced types of operations models such as multi-item, multi-period and more general risk-aware objectives in the context of the coordination of pricing and inventory.

References

Agrawal, V., S. Seshadri. 2000. Impact of uncertainty and risk aversion on price and order quantity in the newsvendor problem. *Manufacturing & Service Operations Management* **2**(4) 410–423.

Andreev, A., A. Kanto, P. Malo. 2005. On closed-form calculation of CVaR .

Arikan, E., J. Fichtinger, W. Jammernegg. 2007. Single period combined inventory and pricing models. *Management Logistischer Netzwerke* 179–199.

Arikan, E., W. Jammernegg. 2009. The newsvendor problem with a general price dependent demand distribution. *Available at SSRN 1523303* .

Aydin, G., E. L. Porteus. 2008. Joint inventory and pricing decisions for an assortment. *Operations Research* **56**(5) 1247–1255.

Bertsimas, D., R. Mazumder. 2014. Exact subset selection in regression via a modern optimization lens. Based on the 2014-2015 Philip McCord Morse Lecture.

Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Science* **59**(6) 1407–1424.

Beutel, A., S. Minner. 2012. Safety stock planning under causal demand forecasting. *International Journal of Production Economics* **140**(2) 637–645.

Billinton, R., R. N. Allan. 1996. *Reliability evaluation of power systems*. Plenum Press.

Bookbinder, J. H., A. E. Lordahl. 1989. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE transactions* **21**(4) 302–312.

Brent, R. P. 1973. *Algorithms for minimization without derivatives*. Courier Dover Publications.

Chan, L. M. A., Z. J. M. Shen, D. Simchi-Levi, J. L. Swann. 2004. Coordination of pricing and inventory decisions: A survey and classification. *International Series in Operations Research and Management Science* 335–392.

Chen, X., D. Simchi-Levi. 2010. Pricing and inventory management. *The Handbook of Pricing Management, Oxford University Press, Oxford, UK* .

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

Chernozhukov, V., I. Fernández-Val, A. Galichon. 2010. Quantile and probability curves without crossing. *Econometrica* **78**(3) 1093–1125.

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

Chun, S. Y., A. Shapiro, S. Uryasev. 2012. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research* **60**(4) 739–756.

Davidian, M., R. J Carroll. 1987. Variance function estimation. *Journal of the American Statistical Association* **82**(400) 1079–1091.

Davidian, M., R. J. Carroll. 1988. A note on extended quasi-likelihood. *Journal of the Royal Statistical Society. Series B (Methodological)* 74–82.

Efron, B., R. J. Tibshirani. 1994. *An introduction to the bootstrap*, vol. 57. CRC press.

Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science* **49**(10) 1287–1309.

Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Operations Research* **47**(3) 454–475.

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

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

Hammerstrom, D. J., R. Ambrosio, J. Brous, T. A. Carlon, D. P. Chassin, J. G. DeSteese, R. T. Guttromson, G. R. Horst, O. M. Järvegren, R. Kajfasz, et al. 2007. Pacific northwest gridwise testbed demonstration projects. *Part I. Olympic Peninsula Project* .

Hanssens, D. M., L. J. Parsons, R. L. Schultz. 2003. *Market response models: Econometric and time series analysis*. Kluwer Academic.

Harsha, P., M. Sharma, R. Natarajan, S. Ghosh. 2013. An inventory theory framework for the analysis of probabilistic demand response schemes. *IEEE Transactions on Smart Grid* **4**(4) 2274–2284.

Hastie, T., R. Tibshirani, J. Friedman. 2001. The elements of statistical learning theory.

Huh, W. T., R. Levi, P. Rusmevichientong, J. B. Orlin. 2008. Adaptive data-driven inventory control policies based on kaplan-meier estimator. *Book of Abstracts*. 42.

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

Karlin, S., C. R. Carr. 1962. Prices and optimal inventory policy. *Studies in Applied Probability and Management Science* 159–172.

Keeping, E. S. 1962. *Introduction to statistical inference*, vol. 26. Courier Dover Publications.

Khouja, M. 1999. The single-period newsvendor problem: literature review and suggestions for future research. *Omega* **27**(5) 537–553.

Kocabıyıkoğlu, A., I. Popescu. 2011. An elasticity approach to the newsvendor with price-sensitive demand. *Operations Research* **59**(2) 301–312.

Koenker, R. 2005. *Quantile regression*. 38, Cambridge university press.

Koenker, R., G. Bassett. 1978. Regression quantiles. *Econometrica* **46**(1) 33–50.

Lau, A. H. L., H. S. Lau. 1988. The newsboy problem with price-dependent demand distribution. *IIE transactions* **20**(2) 168–175.

Leeflang, P. S. H., D. R. Wittink, M. Wedel, P. A. Naert. 2000. *Building models for marketing decisions*. Kluwer Academic.

Leorato, S., F. Peracchi, A. V. Tanase. 2012. Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics & Data Analysis* **56**(4) 768–784.

Levi, R., G. Perakis, J. Uichanco. 2011. The data-driven newsvendor problem: new bounds and insights. Submitted and in revision.

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

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

McCullagh, P., J. A. Nelder. 1989. Generalized linear models (monographs on statistics and applied probability 37). *Chapman Hall, London* .

Mills, E. S. 1959. Uncertainty and price theory. *The Quarterly Journal of Economics* **73**(1) 116–130.

Nadarajah, S., B. Zhang, S. Chan. 2014. Estimation methods for expected shortfall. *Quantitative Finance* **2**(14) 271–291.

Nelder, J. A., Y. Lee. 1992. Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. *Journal of the Royal Statistical Society. Series B (Methodological)* 273–284.

Nelder, J. A., D. Pregibon. 1987. An extended quasi-likelihood function. *Biometrika* **74**(2) 221–232.

Peracchi, F., A. V. Tanase. 2008. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry* **24**(5) 471–493.

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

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

Phillips, R. 2005. *Pricing and revenue optimization*. Stanford Business Books.

Porteus, E. L. 1990. Stochastic inventory theory. *Handbooks in operations research and management science* **2** 605–652.

Raz, G., E. L. Porteus. 2006. A fractiles perspective to the joint price/quantity newsvendor model. *Management science* **52**(11) 1764–1777.

Rockafellar, R. T., J. O. Royset, S. I. Miranda. 2014. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research* **234**(1) 140–154.

Rockafellar, R. T., S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* **2** 21–41.

Rockafellar, R. T., S. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26**(7) 1443–1471.

Rockafellar, R. T., S. Uryasev. 2013. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* **18**(1) 33–53.

Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2008. Risk tuning with generalized linear regression. *Mathematics of Operations Research* **33** 712–729.

Rudin, C., G-Y. Vahn. 2014. The big data newsvendor: Practical insights from machine learning analysis. Working Paper, MIT Sloan School of Management.

Scarf, H., K. J. Arrow, S. Karlin. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* **10** 201–209.

Smyth, G. K., A. F. Huele, A. P. Verbyla. 2001. Exact and approximate reml for heteroskedastic regression. *Statistical modelling* **1**(3) 161–175.

Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the gaussewton method. *Biometrika* **61**(3) 439–447.

Yano, C. A., S. M. Gilbert. 2005. Coordinated pricing and production/procurement decisions: A review. *Managing Business Interfaces* 65–103.

Yao, L., Y. Chen, H. Yan. 2006. The newsvendor problem with pricing: Extensions. *International Journal of Management Science and Engineering Management* **1**(1) 3–16.

Young, L. 1978. Price, inventory and the structure of uncertain demand. *New Zealand Oper. Res* **6**(1) 157–177.

Zipkin, P. H. 2000. *Foundations of inventory management*, vol. 2. McGraw-Hill New York.