university of groningen / faculty of economics and business

2019003-OPERA

The Data-driven Newsvendor Problem: Achieving On-target Service Levels

May 2019

Niels van der Laan
Ruud H. Teunter
Ward Romeijnders
Onur A. Kilic

university of groningen

faculty of economics and business

SOM is the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:
- Economics, Econometrics and Finance
- Global Economics & Management
- Innovation & Organization
- Marketing
- Operations Management & Operations Research
- Organizational Behaviour

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 7068/3815

www.rug.nl/feb/research

university of groningen / faculty of economics and business

The Data-driven Newsvendor Problem: Achieving On-target Service Levels

Niels van der Laan
University of Groningen, Faculty of Economics and Business, Department of Operations
n.van.der.laan@rug.nl

Ruud H. Teunter
University of Groningen, Faculty of Economics and Business, Department of Operations

Ward Romeijnders
University of Groningen, Faculty of Economics and Business, Department of Operations

Onur A. Kilic
University of Groningen, Faculty of Economics and Business, Department of Operations

The data-driven newsvendor problem: achieving on-target service levels

Niels van der Laan Ruud H. Teunter Ward Romeijnders
Onur A. Kilic

May 24, 2019

Abstract

The classical approach to the newsvendor problem is to first estimate the demand distribution (or assume it to be given) and then determine the optimal order quantity. Data-driven approaches that base the inventory level directly on a vector of historical demand and feature observations, offer an alternative. We show that existing data-driven approaches suffer from overfitting, resulting in below-target service levels. We propose new data-driven approaches that do achieve on-target service levels, even if the number of observations is limited. We accomplish this through distributionally robust optimization, i.e. we optimize with respect to a set of distributions that could have generated the historical data. We demonstrate numerically for a range of demand specifications and sample sizes, that our methods achieve on-target service levels without making overly conservative inventory decisions.

1 Introduction

One of the key issues in inventory control is demand forecasting [Silver et al., 1998, Thomopoulos, 2015]. The traditional approach to inventory control is to first forecast demand based on historic demand observations, and then determine the safety stock and corresponding stock position based on the forecast accuracy [Axsäter, 2015]. To improve the accuracy, one can exploit the dependency of demand on related external variables, called features, such as price (changes), weather, or season [Beutel and Minner, 2012]. For example, during a sale, demand is expected to increase, and in summer, demand for beachwear items will be larger than in winter. Moreover, recently a number of so-called data-driven approaches have been suggested that integrate forecasting into decision making, that is, determine the stock position directly based on all available historical data. Most authors consider the single period newsvendor setting as it offers a natural starting point. For example, Levi et al. [2015], Wang et al. [2016], and Oroojlooyjadid et al. [2018] propose data-driven approaches for the case where only historical demand observations are available and Beutel and Minner [2012], Ban and Rudin [2018], and Bertsimas and Kallus [2018] explicitly take into account features.

Unfortunately, data-driven approaches for solving the newsvendor problem are prone to overfitting. As a result, although the in-sample quality of the generated solutions may be high, the out-of-sample performance can be poor, especially if the number of observations is small. Typically, merely asymptotic optimality of the generated solutions can be proven. For example, Levi et al. [2015], Ban and Rudin [2018], and Bertsimas and Kallus [2018] prove that as the number of observations goes to infinity, the solutions generated by their approaches converge to the optimal solution with respect to the true underlying demand distribution. However, in practice limited historical data is available, and the performance of their solutions may be poor.

A second shortcoming of the current literature on the data-driven newsvendor problem is that the focus has mostly been on expected cost minimization instead of service level constraints, in spite of the fact that service level constraints are widely used in practice and typically contract-enforced [Liang and Atkins, 2013]. To our knowledge, Beutel and Minner [2012] are the only ones to consider service level constraints in the presence of features using what we will call the *hindsight approach*. They observe that while the in-sample service level is on-target, the out-of-sample service level is below target as a result from overfitting.

A key observation in this paper is that service level constraints are in fact chance constraints, a special type of constraint known from the stochastic programming literature (see Prekopa [1970] and Birge and Louveaux [1997]). Based on this observation, we exploit results from chance-constrained programming theory by applying them to our setting. This leads to the following contributions.

- (Section 3) We demonstrate that the hindsight approach, which was also considered by Beutel and Minner [2012], leads to achieved service levels far below the targets. This is done by deriving worst-case lower bounds on the probability that the service level restriction is met. These bounds are near zero for realistic sample sizes ($N < 130$). Furthermore, we show that even if the sample size is relatively large for practical situations ($N = 200$), the achieved service level is still considerably below the target.
- (Section 4) We propose several new data-driven approaches to the data-driven newsvendor problem with service level constraints. The advantage of our approaches over existing ones is that we achieve on-target service levels, even if the sample size is small. We accomplish this through *distributionally robust optimization*, that is, we optimize with respect to a set of probability distributions that could have generated the historical data. As the sample size increases, this set of distributions becomes smaller, and thus we maintain asymptotic optimality with respect to the true underlying demand distribution. Moreover, the various methods we propose make different distributional assumptions providing flexibility of use to the decision maker.
- (Section 5) We conduct numerical experiments in order to assess the performance of our approach for realistic demand models and to compare our approach to existing ones. We find that our approaches are more reliable. Indeed, we achieve on-target service levels, even for small sample sizes, without being overly conservative.

The remainder of this paper is organized as follows. In Section 2, we review the literature on data-driven approaches to the newsvendor problem and on chance-constrained programming theory. In Section 3, we reinterpret an existing data-driven approach by Beutel and Minner [2012] developed for service level constraints. Next, we propose new approaches to the data-driven newsvendor problem with service level constraints in Section 4. We numerically test the performance of these approaches in Section 5, and conclude in Section 6.

2 Problem Description and Literature Review

2.1 Problem Description

The newsvendor problem is one of the simplest and most well-known models in safety stock optimization. Over the years, many variants and facets of the original problem have been studied, see e.g. Raz and Porteus [2006], Cachon and Kök [2007], Olivares et al. [2008], Baron et al. [2015], Käki et al. [2015], and Kirshner and Ovchinnikov [2018]; and Qin et al. [2011] for a survey. The problem is to set an inventory level I such that stochastic demand D is met with a prescribed probability $1-\alpha$, where $\alpha \in (0,1)$, while minimizing the expected surplus inventory $\mathbb{E}(I-D)^+$, where $(s)^+ := \max\{0, s\}$, $s \in \mathbb{R}$. Alternatively, the objective is to minimize total expected costs, consisting of underage and overage costs, which amount to b and h per unit, respectively. Thus, the newsvendor problem in case of a service level constraint is given by

$$\min_{I \geq 0} \left\{ \mathbb{E}(I-D)^+ : \quad \mathbb{P}[I \geq D] \geq 1-\alpha \right\} \tag{1}$$

and in case of a cost-minimization objective by

$$\min_{I \geq 0} \mathbb{E}[b(D-I)^+ + h(I-D)^+]. \tag{2}$$

If the cumulative distribution function (cdf) F of demand D is known, then the optimal solutions in (1) and (2) are $I^*_{SL} := F^{-1}(1-\alpha)$ and $I^*_{EC} := F^{-1}\left(\frac{b}{b+h}\right)$, respectively, where $F^{-1}(x) := \min\{t : F(t) \geq x\}$ is the inverse cdf of D. For given demand distributions, the two problems are equivalent, as one can choose b and h such that $b/(b+h) = 1-\alpha$ for any $\alpha \in (0,1)$, or vice versa. In practice, however, the distribution of demand is unknown and (1) is typically harder to solve (approximately) than (2) as feasibility comes into play. Indeed, I^*_{SL} cannot be computed if F is unknown, and thus the goal is to choose an inventory level $I \geq I^*_{SL}$, such that $I - I^*_{SL}$ is as small as possible. In contrast, in case of a cost minimization objective, an inventory level I should be chosen as close as possible to I^*_{EC}, but $I \geq I^*_{EC}$ is not required.

2.2 Data-driven Approaches to the Newsvendor Problem

In recent years, many firms have started to collect more data and explore data science techniques for improved decision making. In line with this development, data-driven approaches have also been suggested for the newsvendor problem. Most of these approaches assume that there are not only historical observations on demand, but also on related external variables, called features, such as price, customer data, Twitter feeds, and weather forecasts. More formally, the decision maker has historical data consisting of N demand observations $D_1, \ldots, D_N$, and of corresponding feature observations $x_1, \ldots, x_N \in \mathbb{R}^d$. We assume that x_i contains a constant, and we will write $x_i = (1, \tilde{x}_i)$. The decision maker faces the problem of determining the optimal inventory level I upon observing the feature vector x_{N+1} in the review period $N+1$. More formally, the decision maker determines the decision rule $I = q(X_{N+1})$ that solves (1) or (2).

The optimal decision rule can be learned through data with machine learning techniques. Data-driven approaches for the case where x contains just a constant (i.e. $d = 1$) are proposed by Levi et al. [2015] (sample average approximation), Wang et al. [2016] (likelihood robust optimization), and Oroojlooyjadid et al. [2018] (deep learning). Other methods use feature data as well as demand data. This approach is taken by Beutel and Minner [2012] (empirical risk minimization (ERM)), Ban and Rudin [2018] (ERM, ERM with regularization, and kernel optimization), and Bertsimas and Kallus [2018] (k nearest neighbours, local regression, classification and regression trees, and random forests). Beutel and Minner [2012] are the only ones to consider the more difficult and arguably more practical service constraint variant (1) of the data-driven newsvendor problem and their approach is discussed further in Section 3.

A risk of machine learning approaches to the newsvendor problem is that they are prone to overfitting. This means that the generated decision rule performs well in-sample, but poorly out-of-sample. While some authors prove asymptotic optimality of their approaches [Levi et al., 2015, Ban and Rudin, 2018, Bertsimas and Kallus, 2018], finite sample performance may be poor, especially if the number of features is large. This can have different implications for the decision rule obtained in this way, depending on whether the objective is to minimize expected costs or to achieve a prescribed service level. In case of an expected cost minimization objective, the decision rule is suboptimal, whereas in case of a service level constraint, the decision rule can be infeasible with respect to the service level constraint.

Indeed, the only existing study on the data-driven newsvendor problem under a service level constraint, showed that overfitting indeed leads to infeasible solutions. To develop alternative approaches, we will interpret the service level constraint as a chance constraint (in Section 3), allowing us to apply results from stochastic programming that will be reviewed in the next subsection.

2.3 Chance-constrained Programming

Consider the following restriction on a vector of decision variables $r \in \mathbb{R}^d$

$$\mathbb{P}[G(r,\xi) \leq 0] \geq 1-\alpha, \tag{3}$$

and the corresponding optimization problem

$$\min_{r\in X}\{c^\top r: \quad \mathbb{P}[G(r,\xi) \leq 0] \geq 1-\alpha\}, \tag{4}$$

where ξ is a random vector with support $\Xi \subseteq \mathbb{R}^d$ and probability measure $\mathbb{P}$, the vector $c \in \mathbb{R}^d$ contains cost coefficients, and $G: \mathbb{R}^d \times \Xi \to \mathbb{R}$ is a known function. The restriction in (3) imposes that a random goal constraint holds with prescribed probability $1-\alpha$ and is known as a chance constraint. To see how chance constraints relate to the data-driven newsvendor problem with service level constraints, suppose that the decision rule q, which links an inventory level to an observation of the feature vector x, is linear in x, i.e. $q(x) = r^\top x$, where r is a vector of decision variables. Then, a service level constraint is represented by

$$\mathbb{P}[D - r^\top x \leq 0] \geq 1-\alpha,$$

which is a special case of (3) by defining $\xi = (x, D)$ and $G(r,\xi) = D - r^\top x$.

While chance constraints have many applications, they are generally intractable [Nemirovski and Shapiro, 2006]. The reason is that even for fixed r, the probability $\mathbb{P}[G(r,\xi) \leq 0]$ requires computing a multi-dimensional integral. Furthermore, the feasible region defined by (3) is

$$\Big\{r \in X : \mathbb{P}[G(r,\xi) \leq 0] \geq 1-\alpha\Big\},$$

which is in general non-convex, even if X is convex. This implies that optimization under (3) is very challenging. Moreover, in practical situations the distribution of ξ is unknown.

In order to deal with these difficulties, approximations of chance constraints have been developed in the literature. The sample average approximation [Luedtke and Ahmed, 2008] and the scenario approximation [Calafiore and Campi, 2006] are of special interest to our analysis and we discuss them now. Sample average approximation (SAA) replaces the distribution of the random vector ξ by the empirical distribution of a sample $\tilde{\xi} := \{\xi^{(1)}, \ldots, \xi^{(N)}\}$ drawn from the distribution of ξ. For example, in the data-driven newsvendor problem, the historical demand and feature observations constitute such a sample. By replacing the distribution of ξ with the empirical distribution of the sample $\tilde{\xi}$, we obtain the approximating constraint

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(G(r,\xi^{(i)}) \leq 0) \geq 1-\alpha, \tag{5}$$

where $\mathbb{1}$ denotes the indicator function. The advantage of this approximation is that it can be solved efficiently [Luedtke et al., 2010, Küçükyavuz, 2012], even though the feasible region defined by (5) is non-convex. Moreover, the following result due to Pagnoncelli et al. [2009] states that under some regularity conditions the solution obtained by SAA is optimal in the true problem (4) as the sample size $N \to \infty$.

Theorem 1. *Suppose that $G(r,\cdot)$ is continuous for every $r \in \mathbb{R}^d$. Furthermore, assume that there exists an optimal solution $\bar{r}$ of the true chance constrained programming problem (4) such that for every $\varepsilon > 0$ there exists an r with $||r - \bar{r}|| < \varepsilon$ and $\mathbb{P}[G(r,\xi) \leq 0] > 1-\alpha$. Then, with probability 1, the optimal solution and value of the SAA problem*

$$\min_{r\in X}\left\{c^\top r: \quad \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(G(r,\xi^{(i)}) \leq 0) \geq 1-\alpha\right\}$$

coincide with the optimal solution and value of the true problem in (4) as the sample size $N \to \infty$.

Proof. See Pagnoncelli et al. [2009]. □

Theorem 1 formalizes the intuition that the approximation of the distribution of the random vector ξ improves as the sample size increases. However, for finite N, there is no guarantee on the solution quality of the solution obtained by SAA. Moreover, if a decision vector r is feasible in the approximating chance constraint (5), then r need not even be feasible with respect to the true chance constraint (3) [Pagnoncelli et al., 2009].

Alternatively, the more conservative *scenario approximation* (ScA) imposes that

$$G(\boldsymbol{r}, \boldsymbol{\xi}^{(i)}) \leq 0, \quad i = 1, \ldots, N.$$

The scenario approximation has the advantage that it is convex in r if $G(\cdot, \xi)$ is convex for every $\xi \in \Xi$, which allows for straightforward optimization. In the setting of the data-driven newsvendor problem, ScA comes down to selecting a decision rule such that demand would have been met in all previous periods. In contrast, SAA only requires that demand would have been met in a fraction $1 - \alpha$ of the previous periods.

In fact, both SAA and ScA run the risk of selecting an infeasible decision rule, but this risk is smaller for ScA. In the literature, bounds have been derived on the minimal sample size required to achieve a prescribed risk level. Indeed, Calafiore and Campi [2006] derive such a bound for ScA, and Luedtke and Ahmed [2008] do so for the SAA case. In Theorem 2, we present the bound for ScA. We omit the bound for SAA, because it severely overstates the required sample size for a given risk level.

Theorem 2. *Suppose that $G(\cdot, \xi)$ is convex for every $\xi \in \Xi$. If*

$$N \geq 2/\alpha \log(1/\delta) + 2d + \frac{2d}{\alpha} \log(2/\alpha),$$

then the optimal solution of the scenario approximation problem

$$\min_{r \in X} \{c^\top r : \quad G(r, \xi^{(i)}) \leq 0, \quad i = 1, \ldots, N\},$$

is feasible with respect to the true chance constraint (3) *with probability at least* $1 - \delta$.

Proof. See Calafiore and Campi [2006]. □

Restated, Theorem 2 says that for a given sample size N, the optimal solution of the ScA problem is feasible with respect to (3) with probability at least

$$1 - \left(\frac{2}{\alpha}\right)^d \exp\left\{\alpha \left(d - \frac{N}{2}\right)\right\}. \tag{6}$$

Note that this bound is increasing in N and converges to 1 as $N \to \infty$. Furthermore, the bound is decreasing in d, which implies that in order to obtain a reliable solution through SAA, we need more observations as the dimension of the data increases.

3 The Hindsight Approach

In order to formulate the hindsight approach, we impose that the decision rule q is linear in x, that is, $q(x) = r^\top x$ for some $r \in \mathbb{R}^d$, so that the problem reduces to finding the optimal value of r. This linearity assumption is not very restrictive, because non-linear relationships can be modelled by including non-linear transformations of x as additional features. The hindsight approach is to find r by solving the following mixed-integer linear programming problem

$$\min_{r,y,\gamma} \sum_{i=1}^{N} y_i \tag{7}$$

s.t.

$$y_i \geq r^\top x_i - D_i, \qquad i = 1, \ldots, N, \tag{8}$$

$$r^\top x_i + \gamma_i M \geq D_i, \qquad i = 1, \ldots, N, \tag{9}$$

$$\sum_{i=1}^{N} \gamma_i \leq \alpha N, \tag{10}$$

$$y_i \geq 0, \quad \gamma_i \in \{0, 1\}, \quad i = 1, \ldots, N, \tag{11}$$

where M is a large positive constant. This approach essentially replaces the unknown joint distribution of (x, D) by the empirical distribution of the historical data. Indeed, the value of r that solves (7)-(11) is such that a service level of at least $1 - \alpha$ would have been achieved in the past, while minimizing the total inventory level. To see this, note that (9) imposes that demand is met in period i, unless $\gamma_i = 1$. By (10), the number of periods in which demand is not met is at most αN, implying that demand is met in at least $(1 - \alpha)N$ out of N periods. Furthermore, (8) combined with $y_i \geq 0$ implies that y_i represents surplus inventory in period i, and thus the objective in (7) is to minimize the total surplus inventory during the past.

We now argue that the hindsight approach is equivalent to an SAA of a chance constraint. Recall from Section 2 that the service level constraint is in fact a chance constraint of the form

$$\mathbb{P}[D - r^\top x \leq 0] \geq 1 - \alpha. \tag{12}$$

If we assume that (x, D) follows a stationary distribution, that is, the joint distribution of (x, D) does not change over time, then the historical feature and demand observations $(x_i, D_i), i = 1, \ldots, N$, constitute a sample from this distribution. The corresponding SAA of (12) is

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(D_i - r^\top x_i \leq 0) \geq 1 - \alpha. \tag{13}$$

To see that the hindsight approach comes down to (13), note that in (7)-(11), $\gamma_i = \mathbb{1}(D_i - r^\top x_i > 0)$, and thus (10) is equivalent to (13).

Using Theorem 1, we infer that the hindsight approach yields the optimal decision rule as $N \to \infty$, provided (x, D) follows a stationary distribution. However, for finite N, there is no such guarantee. In fact, the decision rule obtained through the hindsight approach need not be feasible. Even stronger, the decision rule obtained by solving the more conservative ScA may be infeasible. Nevertheless, Theorem 2 enables us to derive a lower bound on the probability that ScA yields a feasible decision rule, the *reliability* of ScA. Note that the reliability of the hindsight approach is lower than that of ScA, because ScA yields more conservative decision rules than the hindsight approach. Figure 1 shows the bound on the reliability of ScA for the case where we observe just one feature in addition to the constant term. Moreover, we show estimates of the true probability that the hindsight approach and ScA yield infeasible decision rules. In addition, Figure 2 shows the estimated expected service level of the hindsight approach as well as ScA. These estimates are obtained using simulation under the following demand specification, taken from Beutel and Minner [2012]:

$$D_i = a + b x_i + u_i,$$

where a and b are drawn from uniform distributions on $[1000, 2000]$ and $[-1000, -500]$ and where u_i follows a normal distribution with mean zero and variance such that the coefficient of variation at mean price equals 0.3. The feature x_i represents the price level and is drawn from a normal

Figure 1: Lower bound on the reliability of ScA, along with estimated reliabilities of ScA and the hindsight approach under normality, for one feature and $1 - \alpha = 90\%$.

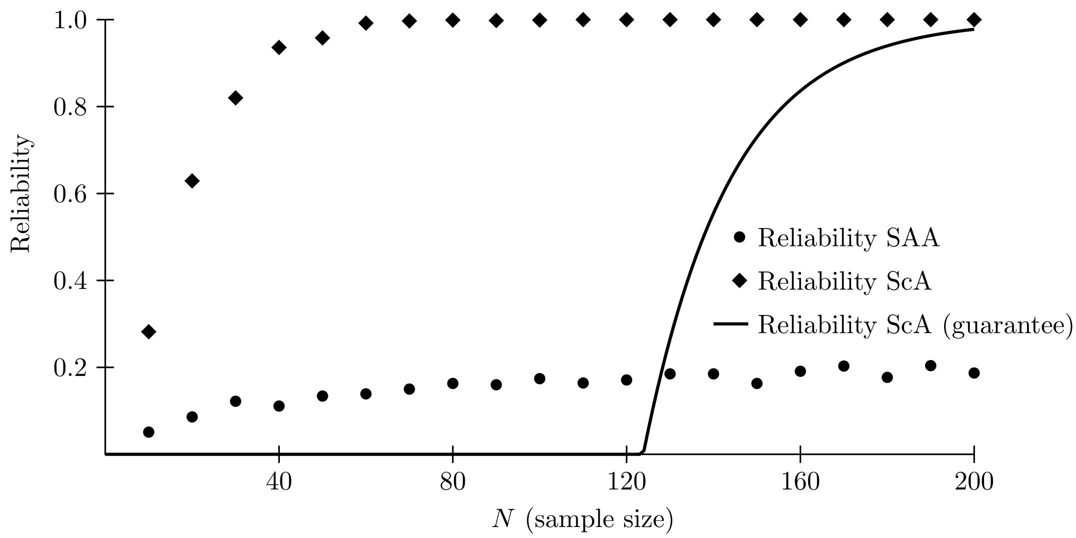
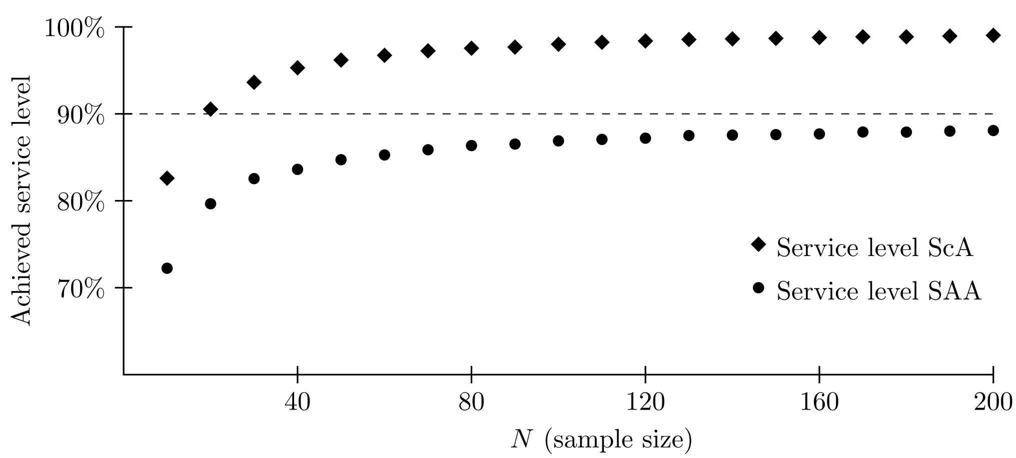


Figure 2: Achieved service level of the hindsight approach and ScA, if we observe demand and one feature, and where $1 - \alpha = 90\%$.



distribution with mean 0.5 and standard deviation 0.25. Both demand and price are truncated at zero to prevent negative values.

Figure 1 shows that we need at least 130 samples in order to obtain a theoretical feasibility guarantee for the ScA decision rule. It turns out that this bound is conservative under the aforementioned demand specification. Indeed, under this specification, the ScA decision rule is feasible with high probability if we have at least a sample size of at least 50. However, the reliability of the hindsight approach is at most 20%, even if $N = 200$. Moreover, Figure 2 shows that the corresponding expected service level is consistently below the 90% target. On the other hand, the ScA decision rule generally achieves too high service levels if $N \geq 30$, resulting in high surplus inventories; and while it performs on target for $N = 20$, this is not supported by any theoretical guarantee.

In Figures 1 and 2, we consider the case where the dimension of the data equals two, that is, we observe one feature next to demand. Recall from Section 2.3 that as the dimension of the data increases, we need even more samples to ensure the feasibility of the ScA decision rule. In order to investigate this further, consider the minimal sample size $N^*(d,\alpha)$ such that the bound in (6) is positive for given d and α. In other words, $N^*(d,\alpha)$ is the minimal sample size such that there is a theoretical guarantee that ScA yields a feasible decision rule. We emphasize that ScA is more conservative than the hindsight approach, and thus we need even more samples to obtain a feasibility guarantee for the hindsight approach.

It follows by setting $\delta = 1$ in Theorem 2 that

$$N^*(d,\alpha) = \left\lceil 2d + \frac{2d}{\alpha}\log(2/\alpha) \right\rceil,$$

where we use the round-up operators to ensure integrality of the sample size. Note that, ignoring the round-up operators, $N^*(d,\alpha)$ increases linearly in d and approaches infinity as $\alpha \to 0$. For example, if we observe 10 features and wish to achieve a service level of 95%, then we need at least a sample size of $N^*(11, 0.05) = 1646$ to have any sort of feasibility guarantee for our decision rule. For a service level of 99%, this increases to $N^*(11, 0.01) = 11679$. These numbers are unrealistic in any practical setting, and thus we propose new data-driven approaches to the newsvendor problem.

4 New Data-driven Approaches for the Newsvendor Problem

Our approaches are based on distributionally robust optimization, which means that we optimize with respect to a set of distributions that could have generated the observed feature and demand data. That is, we consider the the *ambiguous chance constraint*

$$\min\{f(\boldsymbol{x}) : \boldsymbol{x} \in X, \quad \mathbb{P}[G(\boldsymbol{x},\boldsymbol{\xi}) \leq 0] \geq 1-\alpha \quad \forall \mathbb{P} \in \mathcal{P}\}, \tag{14}$$

where the *ambiguity set* $\mathcal{P}$ is a set of probability measures. We consider the ambiguous chance constraint representing a service level constraint

$$\mathbb{P}[D - r^\top x \leq 0] \geq 1-\alpha \quad \forall \mathbb{P} \in \mathcal{P}. \tag{15}$$

Recently, the literature on distributionally robust optimization has grown significantly [Gabrel et al., 2014], and various alternatives for the specification of $\mathcal{P}$ have been proposed [see e.g. Erdoğan and Iyengar, 2006, Zymler et al., 2013, Jiang and Guan, 2016, Hanasusanto et al., 2017, Chen et al., 2018]. We opt to use state-of-the-art approaches with ambiguity sets based on first- and second order moment information, the Wasserstein distance between two probability measures, and Kullback Leibler (KL) divergence.

4.1 Moment Information

Our first approach assumes that we only know the mean μ and covariance matrix Σ of the joint feature and demand distribution; no further distributional assumptions are required. Indeed, we consider the ambiguity set defined as the set of all probability measures that conform with these moments. Note that μ and Σ can be estimated by their sample counterparts using the historical feature and demand observations. If we do so, then the empirical measure $\hat{\mathbb{P}}_N$ lies in the ambiguity set, implying that this approach is more conservative than the hindsight approach. A result due to Zymler et al. [2013] enables us to derive a safe approximation of the ambiguous service level constraint (15). Namely, (15) holds if the optimal value of the following semi-definite program (SDP)

$$\min_{\lambda, M} \lambda + \frac{1}{\alpha} \mathrm{tr}(\Omega M) \tag{16}$$

s.t.

$$M \in \mathbb{S}^{d+1}, \quad \lambda \in \mathbb{R} \tag{17}$$

$$M \succeq O_{d+1}, \quad M \succeq \begin{pmatrix} O_m & 0_m & -\frac{1}{2}\tilde{r} \\ 0_m{}^\top & 0 & \frac{1}{2} \\ -\frac{1}{2}\tilde{r}^\top & \frac{1}{2} & -r_0 - \lambda \end{pmatrix} \tag{18}$$

is non-positive, where the second-order moment matrix Ω is defined as

$$\Omega = \begin{pmatrix} \Sigma + \mu\mu^\top & \mu \\ \mu^\top & 1 \end{pmatrix},$$

the set $\mathbb{S}^{d+1}$ contains all symmetric matrices of order $d+1$, and $r = (r_0, \tilde{r})$ (recall that $x_i = (1, \tilde{x}_i)$ and thus $r^\top x_i = r_0 + \tilde{r}^\top \tilde{x}_i$). We thus propose solving the following SDP in order to obtain a decision rule for the newsvendor problem

$$\min_{r,y,\lambda,M} \sum_{i=1}^{N} y_i$$

s.t.

(17)-(18)

$$\lambda + \frac{1}{\alpha} \mathrm{tr}(\Omega M) \leq 0 \tag{19}$$

$$y_i \geq r_0 + \tilde{r}^\top \tilde{x}_i - D_i, \qquad y_i \geq 0, \qquad i = 1, \dots, N.$$

Here, y_i has the interpretation of surplus in period i, similar as in the hindsight approach. Constraint (19) enforces that the optimal value of the SDP in (16)-(18) is non-positive, which ensures that (15) is satisfied.

4.2 Wasserstein Distance

Next, we consider ambiguity sets defined in terms of the Wasserstein distance between probability measures, which is a frequently used tool in distributionally robust optimization [see e.g. Yang, 2017, Esfahani and Kuhn, 2018, Hanasusanto and Kuhn, 2018, Zhao and Guan, 2018]. More precisely, we optimize with respect to all probability measures that are within a fixed distance θ from a central probability measure, which we take to be the empirical distribution $\hat{\mathbb{P}}_N$ of the feature and demand observations. Note that if $\theta = 0$, then the ambiguity set is a singleton consisting of just $\hat{\mathbb{P}}_N$ and our approach reduces to the hindsight approach. Thus, our approach is in general more conservative than the hindsight approach, but we maintain asymptotic optimality by choosing θ suitably. In particular, if the sample size increases, then $\hat{\mathbb{P}}_N$ is a better approximation of the true distribution and we can choose a smaller value of θ.

More formally, we use the following ambiguity set specification, taken from Chen et al. [2018],

$$\mathcal{P}(\theta) = \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^d) : W_p(\mathbb{P}, \hat{\mathbb{P}}_N) \leq \theta\},$$

where $\mathcal{P}(\mathbb{R}^d)$ denotes the set of all probability distributions on $\mathbb{R}^d$ and $W_p(\mathbb{P}, \hat{\mathbb{P}}_N)$ is the Wasserstein distance of order p between $\mathbb{P}$ and $\hat{\mathbb{P}}_N$. The Wasserstein distance of order p is defined as

$$W_p(\mathbb{P}, \hat{\mathbb{P}}_N) = \inf_{\mathbb{Q} \in \mathcal{P}(\mathbb{P}, \hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{Q}} ||\xi_1 - \xi_2||_p,$$

where $\mathcal{P}(\mathbb{P}, \hat{\mathbb{P}}_N)$ is the set of all distributions on $\mathbb{R}^d \times \mathbb{R}^d$ such that it has marginal distributions $\mathbb{P}$ and $\hat{\mathbb{P}}_N$, and $|| \cdot ||_p$ is the L_p norm on $\mathbb{R}^d$. The Wasserstein distance has the interpretation of the minimum transportation cost necessary to move probability mass from $\hat{\mathbb{P}}_N$ to obtain $\mathbb{P}$, and thus $\hat{\mathbb{P}}_N$ and $\mathbb{P}$ are similar if $W_p(\mathbb{P}, \hat{\mathbb{P}}_N)$ is small.

A good choice of the value of θ should be large enough to guarantee that the true joint distribution $\mathbb{Q}$ of features and demand lies in $\mathcal{P}(\theta)$, but should not be too large, which would lead to making overly conservative decisions. As mentioned before, the value of θ should therefore depend on the sample size N, as we expect that $\hat{\mathbb{P}}_N$ becomes a better approximation of $\mathbb{Q}$ as N increases. Fournier and Guillin [2015] formalize this intuition by showing that, under some moment conditions, $\hat{\mathbb{P}}_N \in \mathcal{P}(\theta)$ with probability $1 - \delta$ if $\theta = g\left(\log(1/\delta)/N\right)$, where g is some increasing sublinear function. If we use the Wasserstein distance of order 1, then θ should scale roughly as $(\log(1/\delta)/N)^{1/d}$, where d is the length of the feature vector, i.e. the dimension of the data. Moreover, if we ensure that $\theta \to 0$ as $N \to \infty$, then Theorem 1 implies that the resulting solution is optimal as $N \to \infty$.

Chen et al. [2018] derive a mixed-integer conic program reformulation of a single ambiguous chance constraint with a Wasserstein ambiguity set. Based on their formulation, we propose our second approach for the data-driven newsvendor problem, which is to solve the following mixed-integer program:

$$\begin{aligned}
&\min_{r,y,q,s,t} \sum_{i=1}^{N} y_i \\
&\text{s.t.} \\
&\frac{1}{N}\sum_{j=1}^{N} s_j + \theta||(1, \tilde{r})||_\infty \le \alpha t && (20)\\
&r^\top x_i - D_i + M q_i \ge t - s_i, \quad i = 1, \dots, N \\
&M(1 - q_i) \ge t - s_i, \quad i = 1, \dots, N \\
&y_i \ge r^\top x_i - D_i, \quad i = 1, \dots \\
&y_i \ge 0,\ s_i \ge 0,\ q_i \in \{0, 1\}, \quad i = 1, \dots, N,
\end{aligned}$$

where M is a sufficiently large constant, and the non-linear constraint (20) can be enforced by adding $2d - 1$ linear inequalities. This formulation is based on the Wasserstein distance of order 1, and thus we use $\theta = (1/N)^{1/d}$. We remark that the results by Chen et al. [2018] enable us to use Wasserstein distances of orders different than 1, but numerical evidence indicates that the performance is comparable.

4.3 Kullback Leibler Divergence

Finally, we consider ambiguity sets based on KL divergence, which is also widely applied in distributionally robust optimization [see e.g. Calafiore, 2007, Bayraksan and Love, 2015, Wang et al., 2016]. The KL divergence between two measures $\mathbb{P}$ and $\mathbb{P}_0$ (the reference distribution) with corresponding probability density functions f and f_0 is defined as

$$D_{KL}(\mathbb{P}, \mathbb{P}_0) = \int_{\mathbb{R}^d} \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) f_0(\xi) d\xi,$$

where $\phi(t) = t\log(t) - t + 1$. This definition assumes that density functions exist for both probability measures, but similar definitions exists for discrete distributions. Similar as for the ambiguity set based on the Wasserstein distance, this approach requires the specification of a reference distribution $\mathbb{P}_0$, for which we consider both the empirical measure and a fitted normal distribution. The corresponding ambiguity set is defined as

$$\mathcal{P}(\theta) := \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^d) : D_{KL}(\mathbb{P}, \mathbb{P}_0) \leq \theta\}$$

It turns out that, as shown by Jiang and Guan [2016], optimization with respect to an ambiguous chance constraint using KL divergence is equivalent to optimization with respect to the reference distribution $\mathbb{P}_0$, with an adjusted risk level $1 - \alpha' \geq 1 - \alpha$. In other words, by choosing a more conservative risk level, we hedge against distributional uncertainty. More precisely, Jiang and Guan [2016] show that α' should be chosen as

$$\alpha' = 1 - \inf_{s \in (0,1)} \left\{ \frac{e^{-\theta} s^{1-\alpha} - 1}{s - 1} \right\}. \tag{21}$$

Denote the feature vector by x and write $x = (1, \tilde{x})$. We consider two reference distributions of $(\tilde{x}, D)$, namely the empirical distribution $\hat{\mathbb{P}}_N$ and a fitted normal distribution $\mathbb{P}_{\mathcal{N}}$, whose mean $\mu = (\mu_{\tilde{x}}, \mu_D)$ and covariance matrix Σ are estimated by their sample counterparts. Furthermore, we take $\theta = (1/N^2)^{1/d}$ to reflect the fact that as the sample size increases, the reference distribution approximates the true distribution more closely, but at a slower rate if the dimension d of the data is large. In order to solve the ambiguous service level constraint (15) with KL divergence with $\mathbb{P}_0 = \hat{\mathbb{P}}_N$, we simply use the hindsight approach with the adjusted risk level $1 - \alpha'$. To solve (15) with $\mathbb{P}_0 = \mathbb{P}_{\mathcal{N}}$, we use a result by Nemirovski [2012] to rewrite the service level constraint $\mathbb{P}_{\mathcal{N}}[D - r^\top x \leq 0] \geq 1 - \alpha'$ with $r = (r_0, \tilde{r})$ as

$$\mu_D - r_0 - \tilde{r}^\top \mu_{\tilde{x}} + \Phi(1 - \alpha') \sqrt{(\tilde{r}^\top, -1)\Sigma \begin{pmatrix} \tilde{r} \\ -1 \end{pmatrix}} \leq 0, \tag{22}$$

where Φ is the cumulative distribution function of the standard normal distribution. Note that (22) is convex in r if $1 - \alpha' \geq 1/2$.

5 Numerical Experiments

5.1 Setup

We use Monte Carlo sampling to compare several data-driven approaches to the newsvendor problem. In particular, we consider the hindsight approach; the scenario approximation; and distributionally robust optimization with ambiguity sets constructed using moment information, the Wasserstein distance, and KL divergence. The ambiguity set based on the Wasserstein distance is centered around the empirical distribution $\hat{\mathbb{P}}_N$, whereas for KL divergence we use $\hat{\mathbb{P}}_N$ and the fitted normal distribution $\mathbb{P}_{\mathcal{N}}$ as reference distributions. For comparison, we additionally solve (15), where $\mathcal{P}$ is a singleton consisting of $\mathbb{P}_{\mathcal{N}}$. For easy reference, we provide abbreviations of these approaches in Table 1.

Table 1: Data-driven approaches: abbreviations.

Approach	Abbreviation
Classical approaches	
Hindsight approach	HA
Scenario approximation	ScA
Optimize with respect to $\mathbb{P}_{\mathcal{N}}$	NOR
Distributionally robust optimization	
Moment information	MI
Wasserstein distance	WD
KL divergence relative to $\hat{\mathbb{P}}_N$	KLE
KL divergence relative to $\mathbb{P}_{\mathcal{N}}$	KLN

In our numerical experiments, we sample feature and demand observations N times under a particular demand specification, where $N = 10, 20, \dots, 100$. Next, we apply the data-driven approaches on this sample to obtain a decision rule which bases the inventory decision directly on the feature observations. Then, for each such decision rule, we estimate the corresponding service level and average surplus inventory using out-of-sample estimation with a sample of size 10^6, sampled from the same demand specification. By repeating this *experiment* 1000 times, we obtain accurate estimates of the service levels and average surplus inventories achieved by the various approaches.

We consider six demand specifications, which are adapted from Beutel and Minner [2012]. First, we generate price x_i (feature) and demand D_i observations obeying $D_i = a + bx_i + u_i$, where x_i is drawn from a normal distribution with mean 0.5 and standard deviation 0.25, and the disturbance u_i is drawn from zero-mean normal distribution, whose standard deviation is chosen such that at mean price the coefficient of variation of demand is cv, where $cv \in \{0.3, 0.5\}$. In each experiment, the values of a and b are drawn from uniform distributions on $[1000, 2000]$ and $[-1000, -500]$, respectively. Note that price and demand are truncated to prevent negative values of x_i and D_i. Second, we let u_i follow a gamma distribution, whose parameters are chosen such that $E[D_i] = a + bx_i$, and again the coefficient of variation of demand at mean price equals cv. Finally, we consider a nonlinear demand specification: $D_i = a + b\exp(p_i) + u_i$, where u_i follows a normal distribution. For this specification, a is sampled from a uniform distribution on $[3000, 4000]$, to prevent that almost all demand observations are truncated to zero.

5.2 Results

In Table 2, we report the estimated mean service levels and average surplus inventories for each approach and demand specification. We report the mean service level, because typical firms stock many items and care mostly about the average service level. In general, we prefer approaches that achieve the lowest surplus inventories while achieving on or above target service levels.

Overall, KLN and WD show the most promising performance. They achieve the lowest average surplus inventories while meeting the service level constraints, even for small sample sizes. We support this claim by analysing the results of each approach in more detail.

5.2.1 Classical Approaches.

In line with the analysis in Section 3, HA suffers significantly from overfitting. In particular, HA achieves service levels that are up to 12% below target for all demand specifications. Performance is worse if the sample size is small, but even if $N = 100$, it performs 3% below target. As a result, HA consistently achieves lower costs than the other methods. However, since service level constraints are often contract-enforced, HA is not useful in practice.

Table 2: Estimated service levels (se < 0.005) and average surplus inventories. Target service level $1 - \alpha = 95\%$. Average surplus inventories in italics if the service level is on or above target. Minimal feasible average surplus inventory in boldface.

	HA		ScA		NOR		WD		MI		KLN		KLE	
N	SL	Surplus	SL	Surplus	SL	Surplus	SL	Surplus	SL	Surplus	SL	Surplus	SL	Surplus
Normal demand specification, $cv = 0.3$														
10	0.83	457.6	0.83	457.6	0.89	518.3	0.91	615.7	1.00	*1334.4*	0.97	***852.8***	0.83	457.6
20	0.85	427.5	0.91	559.1	0.93	535.9	0.95	***729.4***	1.00	*1394.5*	0.98	*770.1*	0.91	559.1
30	0.90	498.1	0.94	621.0	0.93	548.2	0.94	645.8	1.00	*1424.4*	0.98	***738.1***	0.94	621.0
40	0.90	483.7	0.95	*669.9*	0.94	558.1	0.95	*698.6*	1.00	*1450.9*	0.98	*723.2*	0.95	***669.9***
50	0.92	505.1	0.96	*678.2*	0.94	543.3	0.95	***648.0***	1.00	*1417.2*	0.98	*686.3*	0.96	*678.2*
60	0.91	508.4	0.97	*731.6*	0.94	568.1	0.95	*696.9*	1.00	*1476.2*	0.98	*702.7*	0.95	***617.8***
70	0.92	524.1	0.97	*739.8*	0.94	558.8	0.95	*662.9*	1.00	*1455.7*	0.97	*681.1*	0.95	***628.0***
80	0.92	518.4	0.98	*768.3*	0.95	***569.2***	0.95	*685.6*	1.00	*1481.7*	0.97	*685.0*	0.96	*658.1*
90	0.93	521.3	0.98	*755.4*	0.95	***548.7***	0.95	*654.4*	1.00	*1433.9*	0.97	*654.2*	0.95	*596.8*
100	0.92	517.2	0.98	*787.1*	0.95	***559.7***	0.95	*670.8*	1.00	*1461.8*	0.97	*661.3*	0.96	*620.4*
Normal demand specification, $cv = 0.5$														
10	0.83	756.3	0.83	756.3	0.89	845.8	0.91	937.3	1.00	*2172.8*	0.97	***1389.2***	0.83	756.3
20	0.85	706.2	0.91	925.4	0.92	873.9	0.96	***1115.9***	1.00	*2270.7*	0.98	*1254.5*	0.91	925.4
30	0.90	823.7	0.94	1028.7	0.93	895.1	0.94	979.3	1.00	*2323.8*	0.98	***1204.4***	0.94	1028.7
40	0.90	799.8	0.95	*1110.2*	0.94	910.7	0.95	***1065.5***	1.00	*2365.2*	0.98	*1179.4*	0.95	*1110.2*
50	0.92	835.6	0.96	*1124.1*	0.94	886.3	0.95	***978.4***	1.00	*2309.7*	0.97	*1118.9*	0.96	*1124.1*
60	0.91	841.0	0.97	*1212.8*	0.94	927.3	0.95	*1059.9*	1.00	*2407.2*	0.97	*1146.3*	0.95	***1023.2***
70	0.92	867.1	0.97	*1226.5*	0.94	910.7	0.95	***999.2***	1.00	*2369.0*	0.97	*1109.2*	0.95	*1040.1*
80	0.92	857.7	0.98	*1274.1*	0.94	928.3	0.95	***1043.5***	1.00	*2413.8*	0.97	*1116.6*	0.96	*1090.3*
90	0.93	862.7	0.98	*1252.8*	0.94	894.6	0.95	*989.4*	1.00	*2335.0*	0.97	*1065.9*	0.95	***988.4***
100	0.92	855.7	0.98	*1305.5*	0.94	912.4	0.95	***1019.4***	1.00	*2380.3*	0.97	*1077.5*	0.96	*1027.7*
Gamma demand specification, $cv = 0.3$														
10	0.83	506.0	0.83	506.0	0.89	542.6	0.91	675.3	0.99	*1366.3*	0.97	***879.9***	0.83	506.0
20	0.86	463.4	0.91	631.4	0.92	555.4	0.96	*815.5*	1.00	*1421.0*	0.97	***790.7***	0.91	631.4
30	0.90	551.3	0.94	723.0	0.92	564.3	0.94	703.1	1.00	*1448.5*	0.97	***755.0***	0.94	723.0
40	0.89	503.3	0.95	*748.5*	0.92	550.7	0.95	*744.4*	1.00	*1419.9*	0.96	***710.2***	0.95	*748.5*
50	0.92	556.4	0.96	*802.8*	0.93	562.7	0.95	***702.4***	1.00	*1448.5*	0.96	*706.7*	0.96	*802.8*
60	0.91	540.3	0.97	*843.5*	0.93	566.8	0.95	*733.8*	1.00	*1462.3*	0.96	*698.6*	0.95	***686.5***
70	0.92	570.1	0.97	*874.6*	0.93	566.0	0.95	*704.9*	1.00	*1462.4*	0.96	***687.3***	0.95	*714.3*
80	0.92	553.5	0.98	*908.3*	0.93	568.4	0.95	*726.6*	1.00	*1467.3*	0.96	***681.6***	0.96	*747.9*
90	0.93	576.9	0.98	*927.5*	0.93	567.8	0.95	*704.1*	1.00	*1465.4*	0.96	***674.0***	0.95	*684.6*
100	0.92	546.7	0.98	*922.3*	0.93	552.0	0.95	*702.8*	1.00	*1428.9*	0.96	***650.0***	0.96	*685.1*
Gamma demand specification, $cv = 0.5$														
10	0.83	863.5	0.83	863.5	0.88	885.6	0.91	1089.9	0.99	*2209.0*	0.95	***1426.5***	0.83	863.5
20	0.86	770.4	0.91	1099.4	0.91	910.9	0.96	*1335.7*	1.00	*2308.0*	0.96	***1290.1***	0.91	1099.4
30	0.90	931.3	0.94	1245.6	0.91	921.4	0.94	1107.0	1.00	*2354.5*	0.96	***1229.8***	0.94	1245.6
40	0.89	859.1	0.95	*1339.1*	0.92	912.6	0.95	*1208.5*	1.00	*2335.4*	0.96	***1173.3***	0.95	*1339.1*
50	0.91	964.4	0.96	*1447.5*	0.92	933.2	0.94	1128.3	1.00	*2392.8*	0.96	***1170.0***	0.96	*1447.5*
60	0.91	908.1	0.97	*1529.9*	0.92	930.3	0.95	*1181.4*	1.00	*2387.0*	0.95	***1144.2***	0.95	*1194.3*
70	0.92	979.7	0.97	*1604.8*	0.93	942.6	0.95	***1128.2***	1.00	*2416.2*	0.95	*1141.8*	0.95	*1266.3*
80	0.92	958.1	0.98	*1665.1*	0.93	959.0	0.95	*1196.1*	1.00	*2458.6*	0.95	***1147.7***	0.96	*1342.9*
90	0.93	978.1	0.98	*1681.2*	0.93	938.6	0.95	*1128.9*	1.00	*2413.6*	0.95	***1112.7***	0.95	*1192.1*
100	0.92	971.2	0.98	*1756.1*	0.93	958.6	0.95	*1178.5*	1.00	*2456.2*	0.95	***1125.8***	0.96	*1260.5*
Exponential demand specification, $cv = 0.3$														
10	0.83	932.0	0.83	932.0	0.90	1049.7	0.91	1177.7	1.00	*2697.8*	0.98	***1725.6***	0.83	932.0
20	0.85	871.4	0.91	1142.4	0.93	1090.5	0.95	***1410.1***	1.00	*2828.9*	0.98	*1566.1*	0.91	1142.4
30	0.90	1009.6	0.94	1259.3	0.93	1107.9	0.94	1235.6	1.00	*2876.3*	0.98	***1491.4***	0.94	1259.3
40	0.90	971.0	0.95	*1347.6*	0.94	1118.8	0.95	***1333.0***	1.00	*2906.3*	0.98	*1449.6*	0.95	*1347.6*
50	0.92	1032.4	0.96	*1392.0*	0.94	1109.5	0.94	1243.8	1.00	*2892.7*	0.98	*1401.5*	0.96	***1392.0***
60	0.91	1018.8	0.97	*1467.4*	0.94	1136.5	0.95	*1325.7*	1.00	*2951.8*	0.98	*1405.7*	0.95	***1239.5***
70	0.92	1051.9	0.97	*1485.7*	0.94	1119.4	0.95	***1252.9***	1.00	*2915.1*	0.97	*1364.4*	0.96	*1261.6*
80	0.92	1035.8	0.98	*1535.3*	0.95	***1136.7***	0.95	*1302.7*	1.00	*2956.9*	0.97	*1367.9*	0.96	*1315.7*
90	0.93	1063.2	0.98	*1542.8*	0.95	***1117.0***	0.95	*1255.7*	1.00	*2917.2*	0.97	*1331.6*	0.95	*1218.0*
100	0.92	1041.3	0.98	*1589.6*	0.95	***1127.4***	0.95	*1282.5*	1.00	*2943.7*	0.97	*1332.0*	0.96	*1252.9*
Exponential demand specification, $cv = 0.5$														
10	0.83	1531.4	0.83	1531.4	0.89	1705.2	0.91	1844.4	1.00	*4376.5*	0.97	***2799.3***	0.83	1531.4
20	0.85	1433.0	0.91	1882.8	0.92	1772.3	0.95	***2216.1***	1.00	*4599.5*	0.98	*2543.0*	0.91	1882.8
30	0.90	1663.6	0.94	2078.1	0.93	1803.7	0.94	1928.4	1.00	*4680.1*	0.98	***2426.5***	0.94	2078.1
40	0.90	1600.3	0.95	*2225.5*	0.94	1820.1	0.95	***2089.4***	1.00	*4724.2*	0.98	*2356.7*	0.95	*2225.5*
50	0.92	1702.6	0.96	*2299.2*	0.94	1805.3	0.94	1943.6	1.00	*4703.2*	0.97	***2278.8***	0.96	*2299.2*
60	0.91	1680.7	0.97	*2426.3*	0.94	1850.3	0.95	*2069.2*	1.00	*4801.6*	0.97	*2287.2*	0.95	***2046.2***
70	0.92	1735.4	0.97	*2456.6*	0.94	1819.0	0.95	***1945.0***	1.00	*4730.7*	0.97	*2215.4*	0.95	*2083.5*
80	0.92	1710.1	0.98	*2540.3*	0.94	1848.7	0.95	***2037.1***	1.00	*4804.4*	0.97	*2223.4*	0.96	*2174.5*
90	0.93	1754.6	0.98	*2551.2*	0.94	1816.3	0.95	***1959.9***	1.00	*4739.6*	0.97	*2164.0*	0.95	*2011.7*
100	0.92	1719.0	0.98	*2629.1*	0.94	1833.1	0.95	***2001.5***	1.00	*4782.0*	0.97	*2164.7*	0.96	*2069.8*

ScA attempts to repair the overfitting property of HA in a rather crude way by completely ignoring the real target service level and choosing an artificial target service level of 100%. Interestingly, for a realistic target service level of 95%, this still does not give the desired service performance for small sample sizes ($N \leq 30$). On the other hand, for larger sample sizes, ScA results in overly conservative decisions.

Next, we observe that NOR almost always performs below target, even if the normality assumption holds up. This is because the population moments of the normal distribution are estimated with error. This error does vanish as N increases, but this requires a sufficiently large sample size e.g. $N \geq 80$ if demand has a coefficient of variation of 0.3. Furthermore, if the normality assumption is violated, e.g. in the case of a gamma demand specification, then NOR always performs poorly, indicating that it is necessary to hedge against inaccurate parameter estimates as well as distributional misspecifications, as is done in KLN.

5.2.2 Distributionally Robust Optimization.

Note from Table 2 that MI consistently achieves service levels of 99% and 100%, which far exceed the target. In other words, MI is far too conservative, resulting in high average surplus inventories. The reason is that the ambiguity set based on moment information is too large, and does not shrink as N increases. In other words, the fact that we are able to estimate the true feature and demand distribution more accurately if N is larger is not reflected by MI. Moreover, the approximation of the ambiguous chance constraint in (16)-(18) is conservative, resulting in even higher service levels.

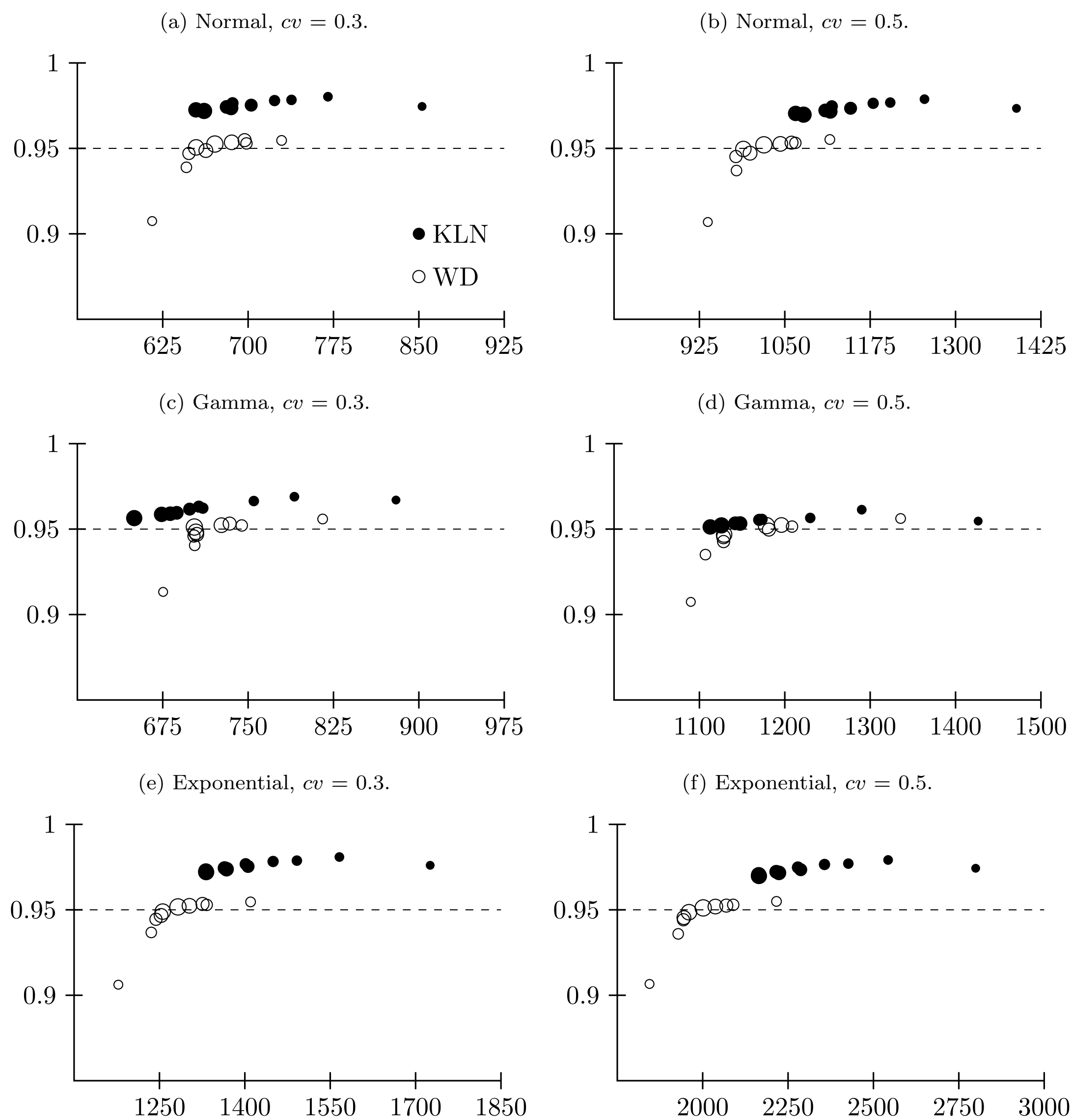
KLE counters the overfitting property of HA by considering all distributions such that the KL divergence relative to the empirical distribution is below a fixed threshold. Recall from Section 4 that this comes down to choosing an artificial target service level which is between $(1-\alpha)100\%$ and 100%. While KLE performs better relative to HA for large sample sizes, it still underperforms by up to 12% for small sample sizes. In fact, if $N = 10$, then both KLE and ScA generate exactly the same decision rule as HA. The reason is that HA achieves an *in-sample* ready rate of at least $(1-\alpha)100\%$, but then, due to the discrete nature of ready rate constraints, HA ensures that demand is met in all 10 cases, resulting in an in-sample service level of 100%. Since KLE and ScA aim for a higher in-sample service level, they arrive at exactly the same decision rule as HA.

Moreover, the KLE ambiguity set only contains discrete distributions which assign probability mass to the observations in the historical dataset. As a result, KLE performs poor if the number of observations is small. The WD and KLN approaches overcome this limitation in different ways, namely by using the Wasserstein distance instead of KL divergence and by using $\mathbb{P}_{\mathcal{N}}$ instead of $\hat{\mathbb{P}}_N$ as the reference distribution, respectively. In particular, the WD ambiguity set also contains distributions that assign probability mass to unobserved feature-demand combinations, and the KLN ambiguity set contains continuous distributions (including $\mathbb{P}_{\mathcal{N}}$).

Indeed, WD and KLN are overall the best performing methods. Out of all approaches, WD achieves the lowest average surplus inventories while meeting the service level restriction on 19 out of 60 instances. Moreover, WD achieves service levels within 1% of the target for all instances with $N \geq 20$. Only if $N = 10$, then WD performs 4% below target. For these instances, KLN is a more reliable alternative than WD: KLN consistently performs on or above target for the range of demand specifications and sample sizes that we consider. However, when more than 10 observations are available, KLN frequently achieves service levels that are up to 3% above target, resulting in average surplus inventories that are higher than strictly necessary. While this suggests that KLN is overly conservative, we observe that in case of a gamma demand specification with $cv = 0.5$, KLN achieves service levels that are much closer to the target. In other words, the performance of KLN cannot be improved further by choosing a narrower ambiguity set without sacrificing performance for other realistic demand specifications.

A head-to-head comparison of WD and KLN reveals that KLN outperforms WD on 31 out of 60 instances, i.e. KLN achieves lower average surplus inventory levels while meeting the service level constraint. Thus, it is not immediately clear which approach should be preferred. To further analyse the differences between the two approaches, consider Figure 3, which shows the achieved service level vs. the average surplus inventory level for both WD and KLN. Figure 3(c) clearly shows that KLN is superior to WD in case of a gamma demand specification with $cv = 0.3$. On the other hand, based on Figures 3(e) and 3(f), WD should be preferred for larger sample sizes. In general we observe that, for a given demand specification, KLN achieves relatively consistent service levels for all sample sizes, while there is considerable spread in the average surplus inventory levels. In contrast, WD achieves average surplus inventory levels that are less dispersed, at the cost of achieving below target service levels for small sample sizes. We conclude that KLN should be preferred for very small sample sizes ($N \leq 20$), while WD is better suited for medium to large sample sizes ($N > 20$).

Figure 3: Estimated mean service level (y-axis) vs. average surplus inventory levels (x-axis). Point size proportional to N (sample size).



6 Conclusion

We consider data-driven approaches to the newsvendor problem, in which surplus inventories have to be minimized while meeting a service level constraint. We assume that there are historical observations not only on demand, but also on *features*: related variables that may improve demand forecasts, such as Twitter feeds, weather data, and price (changes). First, we demonstrate that existing approaches to this problem suffer from overfitting: they achieve below target service levels, limiting their use. These approaches only become reliable if the data contains many observations of every feature. Moreover, this problem worsens if more features are included in the data.

We propose new approaches based on distributionally robust optimization, that is, we optimize with respect to an *ambiguity set*, containing probability distributions that could have generated the observed data. Our approaches are more reliable than existing ones, especially if we have a limited number of observations. Indeed, a numerical study shows that for a range of demand specifications, approaches based on distributionally robust optimization achieve on-target service levels. Two approaches stand out in this respect, namely WD and KLN. These approaches use ambiguity sets based on the Wasserstein distance relative to the empirical distribution and Kullback Leibler divergence relative to a fitted normal distribution, respectively. Based on our results, we overall prefer WD over KLN, because WD generally achieves service levels that are closer to the target. In contrast, KLN frequently overshoots the target, resulting in higher costs than strictly necessary. Only for very small sample sizes, KLN should be preferred because WD performs below target. Interestingly, for specific demand specifications KLN does lead to more efficient solutions than WD. Thus, an avenue for future research is how KLN can be amended to a broader range of demand specifications.

While we consider only the single-item case, our analysis can be generalized to multiple items. Indeed, item specific service levels can be captured by multiple single chance constraints, whereas joint chance constraints can be used for joint service level requirements. Furthermore, our analysis applies to ready rate restrictions, which impose that demand is met in $(1-\alpha)100\%$ of the cases. A frequently used alternative is fill rate restrictions, i.e. at least a prespecified fraction $(1-\beta)$ of demand is met. In this setting, distributionally robust optimization can be applied as well to prevent overfitting. This provides another direction for future research.

References

S. Axsäter. *Inventory control*, volume 225 of *International Series in Operations Research Management Science*. Springer, 2015.

G.-Y. Ban and C. Rudin. The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108, 2018.

O. Baron, M. Hu, S. Najafi-Asadolahi, and Q. Qian. Newsvendor selling to loss-averse consumers with stochastic reference points. *Manufacturing & Service Operations Management*, 17(4):456–469, 2015.

G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. *Tutorials in Operations Research*, pages 1–19, 2015.

D. Bertsimas and N. Kallus. From predictive to prescriptive analytics. *arXiv preprint arXiv:1402.5481v4*, 2018.

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

J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

G. P. Cachon and A. G. Kök. Implementation of the newsvendor model with clearance pricing: How to (and how not to) estimate a salvage value. *Manufacturing & Service Operations Management*, 9(3):276–290, 2007.

G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.

G. C. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

Z. Chen, D. Kuhn, and W. Wiesemann. Data-driven chance constrained programs over wasserstein balls. *arXiv preprint arXiv:1809.00210*, 2018.

E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1-2):37–61, 2006.

P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3):471–483, 2014.

G. A. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 2018.

G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.

R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158(1-2):291–327, 2016.

A. Käki, J. Liesiö, A. Salo, and S. Talluri. Newsvendor decisions under supply uncertainty. *International Journal of Production Research*, 53(5):1544–1560, 2015.

S. N. Kirshner and A. Ovchinnikov. Heterogeneity of reference effects in the competitive newsvendor problem. *Manufacturing & Service Operations Management*, 2018.

S. Küçükyavuz. On mixing sets arising in chance-constrained programming. *Mathematical programming*, 132(1-2):31–56, 2012.

R. Levi, G. Perakis, and J. Uichanco. The data-driven newsvendor problem: new bounds and insights. *Operations Research*, 63(6):1294–1306, 2015.

L. Liang and D. Atkins. Designing service level agreements for inventory management. *Production and Operations Management*, 22(5):1103–1117, 2013.

J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming*, 122(2):247–272, 2010.

A. Nemirovski. On safe tractable approximations of chance constraints. *European Journal of Operational Research*, 219(3):707–718, 2012.

A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

M. Olivares, C. Terwiesch, and L. Cassorla. Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science*, 54(1):41–55, 2008.

A. Oroojlooyjadid, L. Snyder, and M. Takáč. Applying deep learning to the newsvendor problem. *arXiv preprint arXiv:1607.02177v4*, 2018.

B. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: theory and applications. *Journal of optimization theory and applications*, 142(2):399–416, 2009.

A. Prekopa. On probabilistic constrained programming. In *Proceedings of the Princeton symposium on mathematical programming*, volume 113, page 138. Princeton, NJ, 1970.

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

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

E. A. Silver, D. F. Pyke, and R. Peterson. *Inventory management and production planning and scheduling*, volume 3. Wiley New York, 1998.

N. T. Thomopoulos. *Demand forecasting for inventory control*. Springer, 2015.

Z. Wang, P. W. Glynn, and Y. Ye. Likelihood robust optimization for data-driven problems. *Computational Management Science*, 13(2):241–261, 2016.

I. Yang. A convex optimization approach to distributionally robust markov decision processes with wasserstein distance. *IEEE control systems letters*, 1(1):164–169, 2017.

C. Zhao and Y. Guan. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

List of research reports

14001-OPERA: Germs, R. and N.D. van Foreest, Optimal control of production-inventory systems with constant and compound poisson demand

14002-EEF: Bao, T. and J. Duffy, Adaptive vs. eductive learning: Theory and evidence

14003-OPERA: Syntetos, A.A. and R.H. Teunter, On the calculation of safety stocks

14004-EEF: Bouwmeester, M.C., J. Oosterhaven and J.M. Rueda-Cantuche, Measuring the EU value added embodied in EU foreign exports by consolidating 27 national supply and use tables for 2000-2007

14005-OPERA: Prak, D.R.J., R.H. Teunter and J. Riezebos, Periodic review and continuous ordering

14006-EEF: Reijnders, L.S.M., The college gender gap reversal: Insights from a life-cycle perspective

14007-EEF: Reijnders, L.S.M., Child care subsidies with endogenous education and fertility

14008-EEF: Otter, P.W., J.P.A.M. Jacobs and A.H.J. den Reijer, A criterion for the number of factors in a data-rich environment

14009-EEF: Mierau, J.O. and E. Suari Andreu, Fiscal rules and government size in the European Union

14010-EEF: Dijkstra, P.T., M.A. Haan and M. Mulder, Industry structure and collusion with uniform yardstick competition: theory and experiments

14011-EEF: Huizingh, E. and M. Mulder, Effectiveness of regulatory interventions on firm behavior: a randomized field experiment with e-commerce firms

14012-GEM: Bressand, A., Proving the old spell wrong: New African hydrocarbon producers and the ‘resource curse’

14013-EEF: Dijkstra P.T., Price leadership and unequal market sharing: Collusion in experimental markets

14014-EEF: Angelini, V., M. Bertoni, and L. Corazzini, Unpacking the determinants of life satisfaction: A survey experiment

14015-EEF: Heijdra, B.J., J.O. Mierau, and T. Trimborn, Stimulating annuity markets

14016-GEM: Bezemer, D., M. Grydaki, and L. Zhang, Is financial development bad for growth?

14017-EEF: De Cao, E. and C. Lutz, Sensitive survey questions: measuring attitudes regarding female circumcision through a list experiment

14018-EEF: De Cao, E., The height production function from birth to maturity

14019-EEF: Allers, M.A. and J.B. Geertsema, The effects of local government amalgamation on public spending and service levels. Evidence from 15 years of municipal boundary reform

14020-EEF: Kuper, G.H. and J.H. Veurink, Central bank independence and political pressure in the Greenspan era

14021-GEM: Samarina, A. and D. Bezemer, Do Capital Flows Change Domestic Credit Allocation?

14022-EEF: Soetevent, A.R. and L. Zhou, Loss Modification Incentives for Insurers Under ExpectedUtility and Loss Aversion
14023-EEF: Allers, M.A. and W. Vermeulen, Fiscal Equalization, Capitalization and the Flypaper Effect.

14024-GEM: Hoorn, A.A.J. van, Trust, Workplace Organization, and Comparative Economic Development.

14025-GEM: Bezemer, D., and L. Zhang, From Boom to Bust in de Credit Cycle: The Role of Mortgage Credit.

14026-GEM: Zhang, L., and D. Bezemer, How the Credit Cycle Affects Growth: The Role of Bank Balance Sheets.

14027-EEF: Bružikas, T., and A.R. Soetevent, Detailed Data and Changes in Market Structure: The Move to Unmanned Gasoline Service Stations.

14028-EEF: Bouwmeester, M.C., and B. Scholtens, Cross-border Spillovers from European Gas Infrastructure Investments.

14029-EEF: Lestano, and G.H. Kuper, Correlation Dynamics in East Asian Financial Markets.

14030-GEM: Bezemer, D.J., and M. Grydaki, Nonfinancial Sectors Debt and the U.S. Great Moderation.

14031-EEF: Hermes, N., and R. Lensink, Financial Liberalization and Capital Flight: Evidence from the African Continent.

14032-OPERA: Blok, C. de, A. Seepma, I. Roukema, D.P. van Donk, B. Keulen, and R. Otte, Digitalisering in Strafrechtketens: Ervaringen in Denemarken, Engeland, Oostenrijk en Estland vanuit een Supply Chain Perspectief.

14033-OPERA: Olde Keizer, M.C.A., and R.H. Teunter, Opportunistic condition-based maintenance and aperiodic inspections for a two-unit series system.

14034-EEF: Kuper, G.H., G. Sierksma, and F.C.R. Spieksma, Using Tennis Rankings to Predict Performance in Upcoming Tournaments

15001-EEF: Bao, T., X. Tian, X. Yu, Dictator Game with Indivisibility of Money

15002-GEM: Chen, Q., E. Dietzenbacher, and B. Los, The Effects of Ageing and Urbanization on China's Future Population and Labor Force

15003-EEF: Allers, M., B. van Ommeren, and B. Geertsema, Does intermunicipal cooperation create inefficiency? A comparison of interest rates paid by intermunicipal organizations, amalgamated municipalities and not recently amalgamated municipalities

15004-EEF: Dijkstra, P.T., M.A. Haan, and M. Mulder, Design of Yardstick Competition and Consumer Prices: Experimental Evidence

15005-EEF: Dijkstra, P.T., Price Leadership and Unequal Market Sharing: Collusion in Experimental Markets

15006-EEF: Anufriev, M., T. Bao, A. Sutin, and J. Tuinstra, Fee Structure, Return Chasing and Mutual Fund Choice: An Experiment

15007-EEF: Lamers, M., Depositor Discipline and Bank Failures in Local Markets During the Financial Crisis

15008-EEF: Oosterhaven, J., On de Doubtful Usability of the Inoperability IO Model

15009-GEM: Zhang, L. and D. Bezemer, A Global House of Debt Effect? Mortgages and Post-Crisis Recessions in Fifty Economies

15010-I&O: Hooghiemstra, R., N. Hermes, L. Oxelheim, and T. Randøy, The Impact of Board Internationalization on Earnings Management

15011-EEF: Haan, M.A., and W.H. Siekman, Winning Back the Unfaithful while Exploiting the Loyal: Retention Offers and Heterogeneous Switching Costs

15012-EEF: Haan, M.A., J.L. Moraga-González, and V. Petrikaite, Price and Match-Value Advertising with Directed Consumer Search

15013-EEF: Wiese, R., and S. Eriksen, Do Healthcare Financing Privatisations Curb Total Healthcare Expenditures? Evidence from OECD Countries

15014-EEF: Siekman, W.H., Directed Consumer Search

15015-GEM: Hoorn, A.A.J. van, Organizational Culture in the Financial Sector: Evidence from a Cross-Industry Analysis of Employee Personal Values and Career Success

15016-EEF: Te Bao, and C. Hommes, When Speculators Meet Constructors: Positive and Negative Feedback in Experimental Housing Markets

15017-EEF: Te Bao, and Xiaohua Yu, Memory and Discounting: Theory and Evidence

15018-EEF: Suari-Andreu, E., The Effect of House Price Changes on Household Saving Behaviour: A Theoretical and Empirical Study of the Dutch Case

15019-EEF: Bijlsma, M., J. Boone, and G. Zwart, Community Rating in Health Insurance: Trade-off between Coverage and Selection

15020-EEF: Mulder, M., and B. Scholtens, A Plant-level Analysis of the Spill-over Effects of the German *Energiewende*

15021-GEM: Samarina, A., L. Zhang, and D. Bezemer, Mortgages and Credit Cycle Divergence in Eurozone Economies

16001-GEM: Hoorn, A. van, How Are Migrant Employees Manages? An Integrated Analysis

16002-EEF: Soetevent, A.R., Te Bao, A.L. Schippers, A Commercial Gift for Charity

16003-GEM: Bouwmeerster, M.C., and J. Oosterhaven, Economic Impacts of Natural Gas Flow Disruptions

16004-MARK: Holtrop, N., J.E. Wieringa, M.J. Gijsenberg, and P. Stern, Competitive Reactions to Personal Selling: The Difference between Strategic and Tactical Actions

16005-EEF: Plantinga, A. and B. Scholtens, The Financial Impact of Divestment from Fossil Fuels

16006-GEM: Hoorn, A. van, Trust and Signals in Workplace Organization: Evidence from Job Autonomy Differentials between Immigrant Groups

16007-EEF: Willems, B. and G. Zwart, Regulatory Holidays and Optimal Network Expansion

16008-GEF: Hoorn, A. van, Reliability and Validity of the Happiness Approach to Measuring Preferences

16009-EEF: Hinloopen, J., and A.R. Soetevent, (Non-)Insurance Markets, Loss Size Manipulation and Competition: Experimental Evidence

16010-EEF: Bekker, P.A., A Generalized Dynamic Arbitrage Free Yield Model

16011-EEF: Mierau, J.A., and M. Mink, A Descriptive Model of Banking and Aggregate Demand

16012-EEF: Mulder, M. and B. Willems, Competition in Retail Electricity Markets: An Assessment of Ten Year Dutch Experience

16013-GEM: Rozite, K., D.J. Bezemer, and J.P.A.M. Jacobs, Towards a Financial Cycle for the US, 1873-2014

16014-EEF: Neuteleers, S., M. Mulder, and F. Hindriks, Assessing Fairness of Dynamic Grid Tariffs

16015-EEF: Soetevent, A.R., and T. Bružikas, Risk and Loss Aversion, Price Uncertainty and the Implications for Consumer Search

16016-HRM&OB: Meer, P.H. van der, and R. Wielers, Happiness, Unemployment and Self-esteem

16017-EEF: Mulder, M., and M. Pangan, Influence of Environmental Policy and Market Forces on Coal-fired Power Plants: Evidence on the Dutch Market over 2006-2014

16018-EEF: Zeng,Y., and M. Mulder, Exploring Interaction Effects of Climate Policies: A Model Analysis of the Power Market

16019-EEF: Ma, Yiqun, Demand Response Potential of Electricity End-users Facing Real Time Pricing

16020-GEM: Bezemer, D., and A. Samarina, Debt Shift, Financial Development and Income Inequality in Europe

16021-EEF: Elkhuizen, L, N. Hermes, and J. Jacobs, Financial Development, Financial Liberalization and Social Capital

16022-GEM: Gerritse, M., Does Trade Cause Institutional Change? Evidence from Countries South of the Suez Canal

16023-EEF: Rook, M., and M. Mulder, Implicit Premiums in Renewable-Energy Support Schemes

17001-EEF: Trinks, A., B. Scholtens, M. Mulder, and L. Dam, Divesting Fossil Fuels: The Implications for Investment Portfolios

17002-EEF: Angelini, V., and J.O. Mierau, Late-life Health Effects of Teenage Motherhood

17003-EEF: Jong-A-Pin, R., M. Laméris, and H. Garretsen, Political Preferences of (Un)happy Voters: Evidence Based on New Ideological Measures

17004-EEF: Jiang, X., N. Hermes, and A. Meesters, Financial Liberalization, the Institutional Environment and Bank Efficiency

17005-EEF: Kwaak, C. van der, Financial Fragility and Unconventional Central Bank Lending Operations

17006-EEF: Postelnicu, L. and N. Hermes, The Economic Value of Social Capital

17007-EEF: Ommeren, B.J.F. van, M.A. Allers, and M.H. Vellekoop, Choosing the Optimal Moment to Arrange a Loan

17008-EEF: Bekker, P.A., and K.E. Bouwman, A Unified Approach to Dynamic Mean-Variance Analysis in Discrete and Continuous Time

17009-EEF: Bekker, P.A., Interpretable Parsimonious Arbitrage-free Modeling of the Yield Curve

17010-GEM: Schasfoort, J., A. Godin, D. Bezemer, A. Caiani, and S. Kinsella, Monetary Policy Transmission in a Macroeconomic Agent-Based Model

17011-I&O: Bogt, H. ter, Accountability, Transparency and Control of Outsourced Public Sector Activities

17012-GEM: Bezemer, D., A. Samarina, and L. Zhang, The Shift in Bank Credit Allocation: New Data and New Findings

17013-EEF: Boer, W.I.J. de, R.H. Koning, and J.O. Mierau, Ex-ante and Ex-post Willingness-to-pay for Hosting a Major Cycling Event

17014-OPERA: Laan, N. van der, W. Romeijnders, and M.H. van der Vlerk, Higher-order Total Variation Bounds for Expectations of Periodic Functions and Simple Integer Recourse Approximations

17015-GEM: Oosterhaven, J., Key Sector Analysis: A Note on the Other Side of the Coin

17016-EEF: Romensen, G.J., A.R. Soetevent: Tailored Feedback and Worker Green Behavior: Field Evidence from Bus Drivers

17017-EEF: Trinks, A., G. Ibikunle, M. Mulder, and B. Scholtens, Greenhouse Gas Emissions Intensity and the Cost of Capital

17018-GEM: Qian, X. and A. Steiner, The Reinforcement Effect of International Reserves for Financial Stability

17019-GEM/EEF: Klasing, M.J. and P. Milionis, The International Epidemiological Transition and the Education Gender Gap

2018001-EEF: Keller, J.T., G.H. Kuper, and M. Mulder, Mergers of Gas Markets Areas and Competition amongst Transmission System Operators: Evidence on Booking Behaviour in the German Markets

2018002-EEF: Soetevent, A.R. and S. Adikyan, The Impact of Short-Term Goals on Long-Term Objectives: Evidence from Running Data

2018003-MARK: Gijsenberg, M.J. and P.C. Verhoef, Moving Forward: The Role of Marketing in Fostering Public Transport Usage

2018004-MARK: Gijsenberg, M.J. and V.R. Nijs, Advertising Timing: In-Phase or Out-of-Phase with Competitors?

2018005-EEF: Hulshof, D., C. Jepma, and M. Mulder, Performance of Markets for European Renewable Energy Certificates

2018006-EEF: Fosgaard, T.R., and A.R. Soetevent, Promises Undone: How Committed Pledges Impact Donations to Charity

2018007-EEF: Durán, N. and J.P. Elhorst, A Spatio-temporal-similarity and Common Factor Approach of Individual Housing Prices: The Impact of Many Small Earthquakes in the North of Netherlands

2018008-EEF: Hermes, N., and M. Hudon, Determinants of the Performance of Microfinance Institutions: A Systematic Review

2018009-EEF: Katz, M., and C. van der Kwaak, The Macroeconomic Effectiveness of Bank Bail-ins

2018010-OPERA: Prak, D., R.H. Teunter, M.Z. Babai, A.A. Syntetos, and J.E. Boylan, Forecasting and Inventory Control with Compound Poisson Demand Using Periodic Demand Data

2018011-EEF: Brock, B. de, Converting a Non-trivial Use Case into an SSD: An Exercise

2018012-EEF: Harvey, L.A., J.O. Mierau, and J. Rockey, Inequality in an Equal Society

2018013-OPERA: Romeijnders, W., and N. van der Laan, Inexact cutting planes for two-stage mixed-integer stochastic programs

2018014-EEF: Green, C.P., and S. Homroy, Bringing Connections Onboard: The Value of Political Influence

2018015-OPERA: Laan, N. van der, and W. Romeijnders, Generalized aplha-approximations for two-stage mixed-integer recourse models

2018016-GEM: Rozite, K., Financial and Real Integration between Mexico and the United States

2019001-EEF: Lugalla, I.M., J. Jacobs, and W. Westerman, Drivers of Women Entrepreneurs in Tourism in Tanzania: Capital, Goal Setting and Business Growth

2019002-EEF: Brock, E.O. de, On Incremental and Agile Development of (Information) Systems

2019003-OPERA: Laan, N. van der, R.H. Teunter, W. Romeijnders, and O.A. Kilic, The Data-driven Newsvendor Problem: Achieving On-target Service Levels.

www.rug.nl/feb