

Chapter 2

Literature Review

In the following we will review literature on unobservable lost sales estimation, assortment planning and behavioral operations management.

With our focus on retail, we exclude literature on assortment planning in production from our analysis. In contrast to retail where substitution is consumer-driven, demand substitution in production systems is usually controlled by the supplier side. If demand cannot be filled, downward substitution takes place and excess demand is satisfied from a superior product (see for example Hsu and Bassok 1999). In the assortment planning literature, three main types of substitution can be distinguished based on their causes. If substitution occurs as a response to a temporary out-of-stock situation, it is called stockout-based substitution. If customers substitute one product for another to realize savings from price differences, it is referred to as price-based substitution. Assortment-based substitution takes place when a product is not carried by the store at all and thus, a customer chooses another variant from the available assortment set instead of her favorite product. For a comprehensive review on the assortment planning problem and related aspects, the interested reader is referred to Kök et al. (2008) and Pentico (2008).

2.1 Unobservable Lost Sales

The existing literature distinguishes between parametric and non-parametric approaches. The former is based on theoretical demand distributions such as the normal in Nahmias (1994), who suggests to approximate mean and standard deviation from sales data given an order-up-to level inventory policy. The censored part of the right tail of the normal distribution is calculated by taking into account that the distribution function is symmetric about its mean. A major advantage of the normal distribution function is its wide applicability to large datasets due to the

Law of Large Numbers. In contrast, a Poisson process is generally more suitable for discrete and small datasets as in Conrad (1976) or the compound Poisson in Springael and van Nieuwenhuysse (2005).

But, as Agrawal and Smith (1996) emphasize, another important requirement is that the distribution chosen is also capable of capturing the effects of demand variation as present in retailing. They derive the parameters of the negative binomial distribution by matching the sample mean and the frequency of observing zero demand to the observed frequency of demand.

Lau and Lau (1996) propose a nonparametric model that does not require any prior distributional assumptions based on the product limit method (Kaplan and Meier 1958) and daily sales patterns obtained from previous observations.

Berk et al. (2007) use Bayesian updates for obtaining the parameter values of the negative binomial, Gamma, Poisson and normal distribution for the censored newsvendor problem. They rely on an approximation of the posterior distribution by matching the first two moments given that one parameter is known (e.g. mean or variance). Lu et al. (2006) consider Bayesian updates in the context of durable goods for a general distribution function and apply their findings to the normal distribution with known variance. Lu et al. (2008) analytically investigate the benefits from overstocking to learn about the true demand in a Bayesian setting. Tan and Karabati (2004) suggest an updating mechanism to achieve a desired service level by iteratively adjusting the inventory level. Jain et al. (2013) also use Bayesian updates, but additionally take the timing of sales transactions before a stockout occurs into account.

2.2 Assortment Planning

Pentico (1974) is among the first to address the assortment planning problem with stochastic demand. A single-period newsvendor solution is obtained under the assumption that customer arrivals occur before any demand is filled. A “no-crossover” assumption prohibits stockout-based substitution.

van Ryzin and Mahajan (1999) gain theoretical insights on the trade-off between inventory costs and product variety benefits. There is only assortment-based but no stockout-based substitution. Their analysis is restricted to cases where all variants offered have the same retail price-cost ratio. By adding variety to an assortment, total demand increases but comes at the expense of potential cannibalization effects if demand of the other items in the assortment decreases. Fast and costly replenishment strategies are desirable for fashion items for which demand is characterized by purchase behavior that depends on previous demand and therefore allows forecasting future sales once the season has started. In contrast, demand is assumed to be independent for casual items and scale economies should be leveraged by using large-scale store formats.

Another extension of van Ryzin and Mahajan (1999) is proposed by Maddah and Bish (2007) who additionally take pricing into account in a newsvendor setting with assortment-based substitution. Demand is represented by a multinomial logit (*MNL*) model with a mixed multiplicative and additive form. For a high customer arrival rate or large mean demand, the optimal prices in an assortment have equal profit margins.

Topaloglu (2013) also builds on the model of van Ryzin and Mahajan (1999). He extends the model by varying the assortments offered over a selling period. Topaloglu (2013) sets up a nonlinear model to determine which products to offer in an assortment and for how long each product should be offered. He uses the *MNL* model in order to solve the nonlinear model.

Miller et al. (2010) address the assortment planning problem for infrequently purchased goods. They assess the robustness of the optimal assortment for changing customer preferences. Therefore, they develop a *MNL* model with heterogeneous utilities as well as two other choice models. In addition to the retailer's objective function with choice probabilities, they establish upper and lower bounds on the expected profit by assuming that the customer purchases the most respectively least profitable product. They apply adaptive conjoint analysis to online purchase data in order to obtain individual product utilities and form consideration sets for each customer. By comparing results on customer preferences to a retailer's market share, they show that their estimates are reasonable. They find that increased heterogeneity leads to higher profit uncertainty but at the same time allows the shift of customers to more profitable products, thus resulting in higher expected profits.

Sauré and Zeevi (2013) study a dynamic assortment planning problem where the retailer learns about customer preferences by varying the set of products offered. There is only limited shelf space and the retailer must select which products to offer. They address the trade-off between learning out about customer preferences versus offering the best set of products (when no learning takes place anymore). They suggest policies to quickly find the best set of products and how to identify products as suboptimal that should not be carried.

2.3 Assortment Planning with Stockout-Based Substitution

Smith and Agrawal (2000) develop a base-stock inventory model with stockout-based substitution that determines the optimal assortment to be carried as well as inventory levels subject to a service level constraint. They show how further constraints such as shelf space can be incorporated into their approach. A logit choice model is used to determine substitution probabilities. Demand is also dependent on the inventory policy. In a comparison of single and multiple substitution attempts, they find that the more items are stocked the smaller is the effect of allowing for more substitution attempts since the probability of finding a suitable item approaches one.

Mahajan and van Ryzin (2001b) extend the model of Smith and Agrawal (2000) by introducing dynamic consumer substitution where the number of substitution attempts is not restricted and substitution rates depend on the availability of substitutes in a given assortment. They model demand and substitution as a general choice process. The profit-maximization problem is solved with a stochastic sample path gradient algorithm which is compared to heuristic policies. The setting with dynamic substitution is compared to static substitution where demand is independent of the current on-hand inventory levels. An important finding of their analysis is that the profit function is not quasi-concave in inventory levels. Furthermore, larger amounts of popular items and fewer amounts of unpopular items should be stocked in an inventory system with substitution compared to a traditional newsvendor.

Kök and Fisher (2007) develop a practice-motivated approach to determine the optimal assortment from sales data. Given their focus on products with long shelf life and high service level, the demand function is obtained from log-linear regression, ignoring unobservable lost sales. Parameters for assortment-based substitution are estimated from stores with varying assortments calibrated on full-assortment stores. The approach is then extended to possible out-of-stock situations. Stockout-based substitution rates are derived from individual store sales data using the expectation-maximization (*EM*) algorithm. Input data required include time of purchase, customer arrivals at different levels and number of product units sold. Other factors influencing purchase behavior such as price, weather and promotional activities are also incorporated. Finally, an iterative heuristic combined with a local search algorithm is applied to solve the assortment optimization problem. They find that stores should aim at higher inventory levels of goods with high demand variance thus hedging against potential lost sales. The amount of inventory to be carried of products with large case sizes depends on the available shelf space.

Hopp and Xu (2008) formulate an attraction model with a factor for each product that depends on quality and price. Multiple substitution attempts are modeled by a static approximation as a simplification of the dynamic substitution approach of Mahajan and van Ryzin (2001b). Different settings of price, service and assortment competition are studied. In a duopoly with price, service and assortment competition, product variety diminishes compared to a monopoly in order to avoid price competition whereas the total number of products and thus inventory level increases.

Honhon et al. (2010) consider the assortment planning problem with stockout-based substitution. Demand is classified into different customer types whereas each type has a certain ranking of purchase preferences. Prices remain fixed in this model. The optimal assortment is determined for a fixed proportion of each customer type and a heuristic is provided for the more general case with random proportions. They find that the optimal set of assortment possesses a certain structure in terms of newsvendor fractiles and underage cost.

Yücel et al. (2009) combine assortment planning with the supplier selection problem in the presence of quality issues and dynamic substitution behavior. For each of these aspects, a cost function is included in the overall objective function that is furthermore subject to shelf space constraints and constraints on the quantity

of each product that can be supplied. In their analysis, they show that ignoring one of the factors substitution, supplier selection or shelf space limitations results in a significant loss of profit.

2.4 Stockout-Based Substitution in a Fixed Assortment

Parlar and Goyal (1984) address the newsvendor problem with two products and stockout-based substitution with exogenous rates. The assortment is assumed to be fixed and demand is represented by some type of density function. They show that the expected profit as a function of order quantities of the two products is strictly concave. The concavity can only be shown if the retail price of one product lies within a certain range of the substitute's price. Upper and lower bounds of this range can be calculated from the substitution probabilities of the two products.

Mahajan and van Ryzin (2001a) consider inventory competition in an oligopoly. Each firm offers one product and customer demand depends on the availability of the substitute product, i.e. stockout-based substitution takes place. Demand is modeled as a utility-maximizing choice process from which a *MNL* model is chosen for the numerical example. The authors prove the existence and uniqueness of a Nash equilibrium. They provide support for a competitive overstocking effect: In a symmetric setting where all firms face the same costs, prices and choice probability, the firms would be better off if they all decreased their inventory levels in the Nash equilibrium. However, none of the firms would decide to do so because each of them would have an incentive to choose a higher inventory level than its competitors (prisoners' dilemma). Consequently, they all end up with higher inventory levels and lower profits. The profits of each firm decrease with an increasing number of firms in the market.

Netessine and Rudi (2003) compare two settings where stockout-based substitution takes place. In the centralized setting, one decision maker determines the order quantities of all products that may be mutual substitutes. In the decentralized setting, a decision maker only determines the order quantities of one product. Order quantities of potential substitutes are determined by other decision makers. Given deterministic substitution rates, excess demand is reallocated to another product in a given assortment. In the decentralized setting, expected profits are optimized for each product separately. Compared to the standard newsvendor solution, the optimal order quantity is increased by the substituted demand from other products. Therefore, the decentral decision makers have to take the anticipated inventory levels of the other products into account which results in a Nash equilibrium. In the centralized setting, inventory levels are chosen to maximize the expected profit for all products. Demand switching into both directions—to and from a certain product—results in adjusting the newsvendor inventory up- and downward. While inventory levels in the decentralized setting are usually higher than the centralized solution, the authors show also a counterexample where decentralized inventory levels may also be lower for some products.

2.5 Joint Pricing and Inventory Planning with Substitution

Aydin and Porteus (2008) consider the joint inventory and pricing problem for multiple products based on Petruzzi and Dada (1999). Investigating the effect of prices in a given assortment, they allow for price-based substitution. Given several assumptions on the relationship between product price and attractiveness, they establish different demand models, e.g., comprising of the logit and multiplicative competitive interaction functions. The profit function is not jointly quasi-concave in prices and inventory levels, but based on the first-order conditions, they find a price-inventory level vector that maximizes their profit function. They find that a unit cost increase of one product results in lower optimal prices for the other products. Additionally, if a product's attractiveness is quality-dependent, higher quality of one product leads to higher optimal prices for all products.

The problem considered by Zhao and Atkins (2008) is an extension of Petruzzi and Dada (1999) and closely linked to Hopp and Xu (2008) but evaluating simultaneous price and inventory competition. Products are sold by competing vendors, and customers are allowed one substitution attempt to the product offered by another seller in case that their favorite variant is out-of-stock. This type of substitution behavior is represented by price-independent spill rates. Demand is a general function of price competition with a stochastic component. They demonstrate that the profit function is jointly quasiconcave in inventory and price and establish a Nash equilibrium under certain assumptions. For a linear demand function, they show that increasing spill rate raises prices, inventory and safety stock and results in a positive effect on profits. This positive effect can be enhanced by price competition if the spill rate is high. The direction of the effect of increased price competition on prices and safety stocks depends on the level of the spill rate.

Karakul and Chan (2008) address the joint procurement and pricing problem with substitution. Given the fixed price of an existing product, the (higher) price for a newly-introduced product is determined taking into consideration potential cannibalization effects from the existing product. They show that considering substitutability results in higher prices and safety stock of the new product but lower safety stock of the existing one and overall increased profitability.

2.6 Behavioral Operations Management

Human decision makers often deviate from the normative solution. In the past decade, several studies analyzed this behavior and potential causes in the newsvendor setting. Starting with Schweitzer and Cachon (2000), newsvendor decisions were analyzed in laboratory experiments. Schweitzer and Cachon (2000) provide evidence that human decision makers consistently order too few units in high profit settings and too many in low profit settings. This bias towards mean demand was named the pull-to-center bias. Bolton and Katok (2008) show that the pull-to-center

bias also persists after many rounds and (almost) no learning takes place (Bolton and Katok 2008; Ho et al. 2010). The pull-to-center bias is no matter of experience since it is not only observed in experiments with students but also with managers who already have experience with similar ordering tasks (Bolton et al. 2012).

Several potential causes for this pull-to-center bias have been identified: One potential explanation by Schweitzer and Cachon (2000) is demand chasing where the decision maker anchors on his order quantity and adjusts towards observed demand.

A second potential explanation is ex-post inventory error minimization which describes the minimization of anticipated regret from not matching demand (Schweitzer and Cachon 2000). Kremer et al. (2014) find additional evidence for decision maker's preference to avoid ex-post inventory errors. If participants in an experiment are offered additional information on demand which would help to reduce the mismatch between demand and supply, i.e. the ex-post inventory error, they are willing to pay a price that exceeds the benefit from eliminating risk based on several risk utility functions.

A third potential explanation is anchoring on mean demand and insufficient adjustment (Schweitzer and Cachon 2000; Schiffels et al. 2014). It describes the tendency of human decision makers to facilitate the decision making process by choosing mean demand as reference point that is then adjusted.

Another potential explanation was developed by Su (2008). Su (2008) assumes that the newsvendor decision maker chooses the order quantity from a set of alternative solutions where more attractive alternatives (i.e. order quantities yielding higher profits) are chosen with higher probability. This decision process is modeled as logit choice model where the decision makers make random errors.

Kremer et al. (2010) show that ex-post inventory minimization, anchoring and adjustment, and demand chasing are valid in the presence of framing. They present a newsvendor situation in two different ways: one group of participants is aware of facing a standard newsvendor problem whereas the other group plays a lottery experiment. The first group shows a significantly stronger tendency towards mean anchoring and demand chasing than the second group. This result contrasts Su (2008)'s findings since the explanation of random errors should hold for both cases and not depend on the frame.

Building on the ex-post inventory error minimization framework and prospect theory (Kahneman and Tversky 1979), Ho et al. (2010) include reference-dependent preferences in a multilocation newsvendor model. They model the reference-dependent preferences as psychological costs which the decision makers associate with leftovers and stockouts. Decision makers associate higher psychological costs with leftover inventory than with stockouts.

Ren and Croson (2013) identify overconfidence as an explanation for the pull-to-center bias. If the newsvendor is too confident in his estimates, he underestimates the variance which results in an order quantity closer to mean demand than optimal. The authors suggest corrective measures based on coordinating contracts.

Becker-Peth et al. (2013) show that knowledge about behavioral biases helps to improve contract design. They find that order quantities under a buyback contract

do not always lie between mean demand and the optimal order quantity. Further, they study how decisions under a buyback contract are affected by the contract parameters finding that the chosen order quantities not only depend on the critical ratio. The decision makers value revenues from sales differently than from returning items to the supplier. Becker-Peth et al. (2013) use these findings and the data collected in their experiments to fit parameters of adjusted buyback contracts. They test these adjusted buyback contracts in another set of experiments and show that they capture the decision making process better than without behavioral parameters.

Other studies investigate, for example, the effect of censoring (Feiler et al. 2013; Rudi and Drake 2014) or forecasting (Kremer et al. 2011). For more general reviews on behavioral operations not only in the newsvendor context refer to Katok (2011) and Bendoly et al. (2010).

Retail Analytics

Integrated Forecasting and Inventory Management for
Perishable Products in Retailing

Sachs, A.-L.

2015, XVII, 111 p. 14 illus., Softcover

ISBN: 978-3-319-13304-1