Practice-inspired contributions to inventory theory

Prak, Derk Rutger Jordi

Chapter 7

Conclusion & discussion

In this thesis we have investigated several discrepancies between theoretical inventory control models and the practical purpose that they serve. The main topic - addressed from different angles in chapters 2, 3, and 4 - is the lack of an interface with demand forecasting in the current inventory control literature. In Chapter 5 we studied the decoupling of stock review moments and ordering possibilities. Finally, in Chapter 6, we extended the Repair Kit Problem - often-times applied in maintenance and after-sales industries - to include replenishment lead times and fixed ordering costs. The different chapters thus relate to various problems around bridging the gap between inventory control theory and practice.

Chapters 2 and 3 showed that ignoring forecasting uncertainty in inventory decisions leads to substantially too low inventories and therefore poor customer service, especially when parameters are estimated based on a small number of observations or when the lead time is large. The latter is related to a particularly important observation, overlooked in leading textbooks, namely that future forecast errors are correlated. Ignoring this correlation leads to an underestimated lead time demand variance. For mean-stationary, normally distributed demand and an (r, Q) ordering policy, the corrected order decision that acknowledges the uncertainty in the mean and variance is still a simple, closed-form expression, utilizing the Student's t-distribution (Chapter 2).

The approach taken in Chapter 3 led to a general framework that can accommodate any demand model, parameter estimator, and inventory model. It is devised from Bayesian principles, but whereas the standard in the Bayesian literature is to let the demand model dictate the estimator, this framework allows for using any parameter estimator, as long as its error distribution can be derived exactly or approximately. If the demand model incorporates a trend or auto-correlation, or in general if the demand model includes more unknown parameters, then the effect of ignoring parameter uncertainty intensifies, and cost savings achieved by properly addressing this uncertainty can be up to 80%. A finding of Chapter 3 that substantially eases the method's practical applicability is that the asymptotic normal approximation of the estimator's error distribution performs almost as well as its exact derivation for sample sizes of at least 10 historical demands. Furthermore, large cost savings are still possible even if the demand model is misspecified, especially when using this approximate error distribution.

A further interesting observation is that by addressing parameter uncertainty in the inventory decision not only the expected cost is reduced, but also its variance. The typical result is an inventory mark-up that leads to slightly increased costs when the parameter is already overestimated, but also to a very large cost savings when the parameter is underestimated. This simultaneously reduces cost volatility and average cost.

The lack of a connection between the literature strands on demand forecasting and inventory control also shows in the mismatch between forecasting techniques and unknown demand parameters. To perform inventory control in practice, one needs to fit a demand distribution, i.e. estimate its unknown parameters. However, demand forecasting techniques are solely concerned with yielding an as good as possible estimate of future demand per period. Companies store their historical demands periodically, and forecasting and inventory control software packages use these period demands as input to again create period demand forecasts. Contrarily, order decisions are made continuously, based on continuous review inventory models. It has been pointed out in Chapter 4 that period demand forecasts cannot be used to estimate demand parameters at the individual customer level. The compound Poisson demand distribution requires an estimate for the arrival rate of individual customers and their individual demand sizes. Size-Interval forecasting methods such as the one proposed by Croston (1972) yields separate estimates for the number of periods between two periods with positive demand and the total period demand size. However, mis-using such methods to estimate compound Poisson demand parameters - as current integrated demand forecasting and inventory control software packages do - leads to way too high inventories, and thus to unnecessarily high costs and overshot service levels.

In order to consistently estimate compound Poisson demand parameters from periodic data, the inventory control literature suggests the standard method-of-moments estimator, without motivation. We showed that this estimator has a severe bias in small samples, leading to bad performance especially when the demand pattern is of an intermittent nature. Based on Croston's idea of separately estimating the customer arrival rate and the average demand size we presented an alternative method-of-moments estimator for the parameters of a compound Poisson distribution from periodic data, which significantly outperformed the standard method-of-moments estimator and performed comparably to maximum likelihood, while being closed-form and easy to implement in current software packages.

Next to the interface with demand forecasting, we have studied two other aspects of inventory control. The first aspect is the timing of order placements and stock reviews (see Chapter 5). We have shown that the typical assumption in the literature, which is that orders are placed and delivered at the same moments as when stock reviews take place, is suboptimal. If orders are allowed to arrive also in-between stock reviews, then they are typically not placed at or around the stock review. Contrarily, safety stocks are built up at a diminishing rate throughout the period where the exact stock level is not known, but ordering is halted shortly before the stock review. At this review, there is typically some excess safety stock, which is then first depleted until new safety stock is built up again. If we compare this ordering scheme with the standard that only orders at the inventory review point, then we see large cost savings because of the much smoother inventory build-up. Safety stocks, and thus costs, decrease in the level of randomness in the demand, i.e. the level of uncertainty. This means that further cost savings can be achieved by including (partially) known information on demands that have occurred, for example by making a split-up between the known stream of incoming demands and the unknown loss due to theft, misplacement, etc.

The second aspect, studied in Chapter 6, is the inclusion of fixed order cost and a positive lead time in the Repair Kit Problem, concerned with the question how many parts of various types an engineer has to carry on a tour. This tour may consist of a random number of jobs, each requiring random numbers of various parts. This problem has thus far only been studied assuming free overnight restocking, but we have shown that once order costs and lead times are introduced, highly different order policies arise. One should weigh the job fill rate increase and the expected holding cost increase. We have studied a real case and found two different kinds of order policy behavior. In many cases there are decreasing marginal returns from stock level increases, but for some engineers an S-shaped curve is obtained, showing that a 'critical mass' of items are required on many jobs, and that other more specific items are only useful on certain jobs. Most importantly, the typical heuristic mechanism applied by previous authors that adds parts to the kit based on the ratio of its expected demand and price, does not work well in this more realistic setting.

Discussion & future research

The main conclusion of this thesis, shining through in all chapters, is that more attention should be devoted to the practical applicability of inventory control research. Models that assume full knowledge of a certain demand distribution cannot be used without modification in the real world, where such knowledge is not available. Models that impose that orders are placed solely at the same points in time where the stock level is reviewed lead to unnecessarily high peak inventories and related costs. Finally, models that ignore replenishment lead times and order costs, or fail to make a complete cost-benefit trade-off when selecting which parts to stock, lead to suboptimal compositions and therefore too high costs for a given service level.

The key to making better decisions in practice lies not only with the development of inventory control models, but also with their users - practitioners and software developers. They should carefully design their demand monitoring and storage systems to facilitate proper parameter estimation and decision making. This starts with the way they store their demand data, per transaction or periodically. In the latter case, also the period length should be decided on. On the other hand, users should reflect on their actual business situation and choose the inventory control model that best suits their needs. Finally - and this is where software developers play a key role - the bridge should be made from data to inputs for that specific inventory model, including demand forecasting, parameter estimation, and estimation uncertainty.

The literature should also be more focused on the combined process to move from data to decision in practice. Currently, the research fields of demand forecasting and inventory control operate almost in complete isolation. On the forecasting side, future demand per period is predicted with the sole purpose of producing that forecast, without bearing in mind that demand forecasts are primarily an aid to make better (inventory) decisions. Contrarily, on the inventory control side, optimal replenishment strategies are derived assuming a certain demand distribution of which all parameters are known. On the forecasting side even data manipulation such as temporal aggregation is applied to lower forecast errors of aggregated demands, whereas for demand parameter estimation it would be optimal to store and use demand data at the individual customer level.

This also leads to a peculiar situation in integrated demand forecasting and inventory control packages. These take demand per period as input and produce a period demand forecast, but subsequently derive from this the policy parameters of a continuous review inventory system. After all, at the moment that a single customer demand causes the inventory level to drop below a certain reorder level, dedicated inventory control software will immediately signal a replenishment order. Here we observe exactly the same discrepancy that we also see in the literature: periodic inputs and forecasts, but continuous inventory monitoring and decision making.

Even if parameter estimators are appropriately selected for the demand process at hand, then still the problem of dealing with parameter uncertainty remains and is very relevant, as in practice typically only very few demand observations are available (since the last structural change of the demand process). If an optimal replenishment policy that was derived assuming known demand parameters, is no longer optimal if estimates are used instead. Chapters 2 and 3 studied this problem and proposed a framework to generally overcome this problem. However, it can be computationally challenging to apply the method of Chapter 3 when multiple parameters are estimated. On the other hand, the approximate option does work well if a reasonable number of historical demands is available.

Since the beginning of this century an alternative view on inventory control research has slowly developed in parallel to the classical view: data-driven inventory control. The philosophy of this research strand is to create a direct connection between the data and the replenishment strategy, rather than following a sequential forecasting and inventory control process. This fundamentally different approach is promising because of its much more flexible way of modelling the demand and inventory situation. The first suggestion to derive optimal inventory parameters directly from observed demand (without distributional assumptions) came from Iyer and Schrage (1992), but only after the applications to the newsvendor model by Godfrey and Powell (2001) and Levi et al. (2007, 2015) data-driven inventory control started to gain some more traction. A new idea which can be regarded as a step towards 'Big Data Inventory Control' was proposed by Beutel and Minner (2012) and Sachs and Minner (2014), who take the inventory level as a dependent variable in a linear regression on various different variables that may explain demand. The popularity increase of data-driven inventory control could be catalysed by recent advances in machine learning for data science, increasing computer power, and the explosive growth of data storage and sharing within the supply chain.

Data-driven, integrated demand forecasting and inventory control frameworks, based on or complemented by artificial intelligence, provide a promising outlook to enhance the practical focus of inventory control research. An important conceptual contribution to integrate machine learning with managerial decision problems was made by Bertsimas and Kallus (2018), and first applications are given by Ban and Rudin (2018), Oroojlooyjadid et al. (2018), Taigel and Meller (2018) and Meller et al. (2018). Other authors have successfully applied deep learning to find good solutions to analytically unsolvable inventory models (Bravo and Shaposhnik, 2018; Gijsbrechts et al., 2018). The development of these literature streams will inevitably enhance the way in which inventories are managed. This lays out many new research opportunities, both on developing this new strand and on its interface with classical demand forecasting and inventory control literature.