Abstract: The aim of stochastic inventory control is to determine the timing of issuing replenishment order and the corresponding order quantity subject to uncertainty of demand and/or other system parameters. In this literature survey we focus on the single-item single-stocking location stochastic inventory control problem. More specifically, we survey exact and heuristic models under stationary and non-stationary demand according to uncertainty strategies proposed by Bookbinder and Tan (1988).

Keywords: inventory control, stochastic lot sizing, static uncertainty, dynamic uncertainty, static-dynamic uncertainty.

1. INTRODUCTION

Inventory control focuses on the trade-off that arises when a decision maker aims at meeting customer demand whilst simultaneously maximising profitability of operations.

At the heart of inventory control we find three key questions (Silver, 1981):
(1) How often should the inventory status be observed?
(2) When should an order be placed?
(3) What should be the order quantity?

The earliest published academic study on inventory control, or “lot sizing,” was carried out by Harris (1913). This research came from author’s working experience to determine the most economical quantity of replenishment orders. He considered six factors (unit cost, set-up cost, interest and depreciation on stock, movement and manufacturing interval) to investigate how the total cost varied with the change of order size, and derived the square-root formula for reordering quantity, which is now known as Economic Order Quantity (EOQ) for the deterministic lot sizing problem.

Over decades, a vast amount of literature investigated lot-sizing models to progressively embed more realistic assumptions. This led to several variants according to demand characteristics and parameter settings. Table 1 presents the general classification of lot sizing problem with review papers. A survey of existing review works is presented in (Bushuev et al., 2015).

As shown in Table 1, lot sizing problems are categorised as deterministic and stochastic according to demand type. For deterministic problems, customer demand is constant or dynamic over the time horizon. The EOQ model tackles the case in which demand is constant. Extensions of this model were discussed in (Drake and Marley, 2014). Wagner and Whitin (1958) first proposed the exact formulation and solution methods for the dynamic lot sizing problem; their initial work was improved by Zabel (1964) and Bahl and Taj (1991). Researchers also investigated heuristics algorithms for dynamic lot sizing problems such as Silver and Meal (1973; 1978), Khan et al. (2014) and Beck et al. (2015).

This paper attempts to survey the papers for single-item lot sizing problem under stochastic (stationary and non-stationary) demand in terms of uncertainty strategies. We do not merely list or summarise papers in this area or its extension; our contributions to the literature on lot sizing problem are the following.

- We develop a categorisation of stochastic lot sizing problem under stationary and non-stationary demand respectively to show how the literature evolved over the last several decades.
- We demonstrate the development of literature from the perspective of replenishment policies based on Bookbinder and Tan’s (1988) uncertainty strategies, and presents exact formulation and heuristic algorithms for both types of demand.

In the rest of paper, section 2 will briefly introduce stochastic lot sizing problems, uncertainty strategies and replenishment policies. Sections 3 and 4 will review papers for stationary and non-stationary demand, respectively, in terms of replenishment policies and solution methods.

Table 1. Lot sizing problems and review papers

<table>
<thead>
<tr>
<th>Classification</th>
<th>Review Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Review</td>
<td>(Silver, 1981, 2008), (Glock et al., 2014)</td>
</tr>
<tr>
<td>Deterministic Demand</td>
<td>(Drake and Marley, 2014)</td>
</tr>
<tr>
<td>Stochastic Demand</td>
<td>(Brahimi et al., 2006, 2017)</td>
</tr>
<tr>
<td>Lead Time</td>
<td>(Ritchie, 1986)</td>
</tr>
<tr>
<td>Capacitated Stock</td>
<td>(Dural-Selcuk et al., 2016)</td>
</tr>
<tr>
<td>Perishable Item</td>
<td>(Das, 1975)</td>
</tr>
<tr>
<td>Lateral Transshipment</td>
<td>(Karimi et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>(Nahmias, 1992), (Janssen et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>(Paterson et al., 2011)</td>
</tr>
</tbody>
</table>
2. STOCHASTIC LOT SIZING PROBLEM AND UNCERTAINTY STRATEGIES

The simplest case for stochastic lot sizing is the News-vendor problem addressed by (Edgeworth, 1888), which is concerned with controlling a single item over a single time period.

In the early Sixties, Wagner and Whitin’s work (1958) on dynamic lot sizing was extended into the stochastic lot sizing problem considering single-item single-stocking location inventory control problem with multiple time periods. To deal with the uncertainty of demands, three uncertainty control strategies were discussed and adopted by Bookbinder and Tan (1988): the “static,” the “static-dynamic” and the “dynamic uncertainty.”

Under the static uncertainty strategy, the timing (R) and the quantity (Q) of an order are predefined before the system operates. This policy is generally known as (R, Q) policy.

The static-dynamic uncertainty applies a fixed reordering timing, at which the replenishment quantity is determined. The inventory level required to maintain (the order-up-to-level: S) is also set in advance; however, the actual replenishment quantity is decided only at the time the order is issued. The policy is usually denoted as the (R, S) policy. An alternative to the (R, S) policy is the (s, Q) policy; in an (s, Q) policy, an order of size Q is issued when inventory falls below or at the reorder threshold s.

The replenishment policy under dynamic uncertainty strategy is the (s, S) policy, where s denotes the reorder point and S the order-up-to-level. If current stock decreases to s, the system will place a replenishment order to restore the current stock to the level S. (s, S) policy allows full flexibility. Scarf (1959) showed that if the holding and shortage cost are linear, the optimal policy in each period is of (s, S) type.

3. STOCHASTIC LOT SIZING PROBLEM WITH STATIONARY DEMAND

When the distribution of stochastic demand remains the same form over the time horizon, finite or infinite, the lot sizing problem is stationary. Research on stationary lot sizing problems initially focused on structural properties of optimal control policies. However, the model solving in majority of formulations is of high complexity; thus a wide range of heuristics approaches were derived for various types of replenishment policies.

3.1 (s, Q) Policy

The research on (s, Q) systems — often referred to as (r, Q) systems — started with Galliher et al. (1959), which compared two systems with arbitrary form stationary probability distribution and with Poisson distribution. They found that the value of r should be increased along with the increase of Q in variance of replenishment time to maintain the optimality of the solution.

Hadley and Whitin (1962) presented an exact solution to the problem where penalty cost was applied for backordering. Hadley and Whitin also derived an heuristic algorithm that ignored the possibility during lead time that demand exceeded the quantity ordered and stockout.

Atkins and Iyogun (1988) proposed a decomposition method to derive a tight lower bound for stochastic joint replenishment method under (r, Q) system.

Browne and Zipkin (1991) discussed the (r, Q) system when demand was discrete and (multivariate) diffusion based on Hadley and Whitin’s discussion.

Federgruen and Zheng (1992) presented an exact algorithm for the discrete cases. In this method, the optimal parameters were found by incrementally enlarging the interval of base-stock policies over which the average was taken until the total cost stopped decreasing. Rosling (1999) provided a revised version for Federgruen and Zheng’s method to improve the computational complexity. In Rosling’s method, Q was initially set to be its lower bound, and the corresponding r was obtained by Rosling’s discrete square root algorithm. The number of iterations required in Rosling’s method was still large and the upper bound was not tight, which meant that it was still pseudo-polynomial.

After Hadley and Whitin (1962), several works discussed exact models and heuristic solution methods for stationary problem under various assumptions.

Johansen and Thorstenson (1996) considered (r, Q) system with Poisson demand and lost sales. They formulated an exact model and designed a policy-iteration algorithm for discounted cases. Gallego (1998) derived a distribution-free solution and provided upper bounds on the optimal long run average cost and on the optimal batch size. Lau and Lau (2002) proposed a method using spreadsheet’s direct optimisation to solve a (r, Q) system with backordering. Shen et al. (2009) proposed a recursive procedure for determining the exact policy costs for (r, Q) policy with Poisson demand and constant lead time. Mhada et al. (2013) addressed the model when lead time was of exponential distribution and the demand was an exponential unreliable manufacturing plant, aiming for a constant mean production rate. Drezner and Scott (2015) derived the approximate formulas for the optimal solution for the particular case of an exponential demand distribution and simple formulas for the general Poisson demand distribution. Bright and Rossetti (2013) provided a comparison among algorithms for the unconstrained (r, Q) inventory system by evaluating computational performance and solution accuracy of the algorithms for a series of randomly generated instances.
3.2 \((s, S)\) Policy

Given the complexity in computing parameters for \((s, S)\) policy, there are many works developing efficient algorithms for stationary \((s, S)\) policy. This research started with Iglehart (1963), which gave bounds for the sequences of \(\{s_n\}\) and \(\{S_n\}\). Iglehart also investigated the limiting behaviors of \(\{s_n\}\) and \(\{S_n\}\) over the infinite time horizon under backordering and lead-time settings. It was proved that the sequences \(\{s_n\}\) and \(\{S_n\}\) contained convergent subsequences, and every limit point of the sequence \(\{S_n\}\) was a minimum for the cost function. Then, if the cost function had a unique optimum, \(\{S_n\}\) converged.

After Iglehart (1963), Veinott Jr and Wagner (1965) derived a computational approach from renewal theory and stationary analysis, and generalized it for the unit interval range of value for discount factor \(\alpha\). The algorithm started with the condition that \(\alpha < 1\); furthermore, a resolution was found to guarantee the optimum of \((s, S)\) policy was determined by computation.

Richards (1975) examined the condition under which the inventory position was uniformly distributed, that if and only if demands were of unit size. Richards also proved that this condition was independent of the lead time distribution and of demand distribution. Sahin (1982) also gave mathematical results on \((s, S)\) inventory models based on renewal junction. He proved that the total cost function was pseudo-convex if the underlying renewal function is concave. His conclusion guaranteed that every local minimum was a global minimum of total cost, and permitted the efficient computation of the optimal policy parameters through a one-dimensional search routine.

To solve the problem efficiently, Archibald and Silver (1978) developed a series of formulae to calculate the cost for \((s, S)\) policies recursively, when the inventory system was continuously reviewed with discrete compound Poisson demand. The algorithm started with finding the optimal \(s\) for a given \(n\), where \(n = S - s - 1\); and then \(S\) would be determined by finding an optimal \(n\), where \(n\) is a local optimum. Archibald and Silver found that the value \(n\) increased until a local minimum pair of \((s, n)\) was found. Tighter bounds were developed in Veinott Jr and Wagner’s (1965).

Inspired by Archibald and Silver’s research, Federgruen and Zipkin (1984) proposed an algorithm to compute \((s, S)\) policy starting with any arbitrary parameter pair. The algorithm was based on an adaptation of the general policy-iteration method for solving Markov decision problem, where the special structure of \((s, S)\) policies was exploited in several ways. However, this algorithm was easily trapped in local optima due to the quasi-convex nature of the cost function. This algorithm was improved by Zheng and Federgruen (1991) based on properties of cost function of \((s, S)\) system and the tight lower and upper bounds for two parameters, which were iteratively and easily updated and converged monotonically. Zheng and Federgruen also exploited a characterisation of the cost function to allow fast updates by only altering the value of \(s\). Being different from Zheng and Federgruen’s (1991) method, Feng and Xiao (2000) introduced a dummy cost factor and an auxiliary function to search for the optimal cost value. Feng and Xiao’s algorithm revised the dummy cost based on the sign of auxiliary function and identified the non-prospective set of \(S\) to reduce search efforts. The numerical test in (Feng and Xiao, 2000) showed that this algorithm saved on average more than 30% evaluation effort compared with Zheng and Federgruen’s algorithm.

4. STOCHASTIC LOT SIZING PROBLEM WITH NON-STATIONARY DEMAND

As Graves (1999) claimed, one main reason for the continuing development of inventory theory is to embed more realistic assumptions that are in line with production manufacturing demand into theoretical inventory models. Under the majority of practical industry circumstances, demand is not only stochastic but also non-stationary.

4.1 \((R, Q)\) Policy

Sox (1997) studied the dynamic lot sizing problem with known cumulative function of demand in each time period. In his settings, costs were non-stationary and backordering was allowed. A mixed integer non-linear program was applied to formulate this problem. The model described the immediate cost incurred at the end of each time period by the loss function in inventory theory, which was convex but non-linear. By substituting cumulative order quantity into the immediate cost function, inventory variables were eliminated so that the objective function can be separable in respect of cumulative order quantity variables. Given this separability, a solution algorithm was derived based on Wagner-Whitin’s algorithm (1958) for deterministic problems by adding additional feasibility constraints. The algorithm transformed the objective cost function into a series of multi-period newsvendor problems with various constraints by decomposition, and conducted rolling-horizon implementation for obtaining the optimal solution.

This policy was also investigated by Vargas (2009) to determine the optimal solution over the entire finite planning horizon for dynamic lot sizing problem with stochastic and non-stationary demand, where a demand density was known. The model applied assumptions in Wagner-Whitin (1958) and introduced penalty cost for backordering. By using stochastic dynamic programming, his model was shown to be equivalent to solving a shortest path problem.
in a specified acyclic network. Vargas also provided an optimisation algorithm with rolling horizon with two stages: (1) to determine optimal replenishment quantities for any sequence of replenishment point, and (2) to sort out the optimal sequence of replenishment from above. The results obtained by this approach presented a better performance than the previous known methods.

4.2 \((R,S)\) Policy

Tarim and Kingsman (2004) formulated the problem as a mixed integer programming (MIP). It modeled the total expected cost by minimising the summation of holding and ordering cost under the constraint that the probability that the closing inventory in each time period was a certain non-negative value. This formulation allowed the simultaneous determination of reorder point and size. Tarim and Kingsman (2006) provided another MIP formulation where the objective function was obtained by the mean of piecewise linearisation. The accuracy of the approximation can be adjusted by introducing new breakpoints.

Tarim et al. (2011) provided an efficient computational approach to solve the MIP model in (Tarim and Kingsman, 2004). The algorithm converted the relaxation of the original MIP model to a shortest path problem implemented by branch-and-bound procedures. This algorithm also considered the case of infeasibility, where the solution would generate a tight lower bound for the optimal cost, and it can be modified to obtain a feasible solution to generate an upper bound.

Özen et al. (2012) considered both penalty cost and service level, proved that the optimal policy was the base stock policy for both penalty and service-level constrained models, and for the capacity limitations and minimum order quantity requirements.

More recently, Rossi et al. (2015) extended Tarim and Kingsman’s MIP model (2006) into a mixed integer linear programming (MILP) formulation for non-stationary stochastic demand. The model first applied loss function and its complementary function to describe total cost. A piecewise linearisation approach was utilized to convert cost function from non-linear form to linear. The research also considered several service level measures (\(\alpha\) service level on each period, \(\beta^{\text{sc}}\) service level independently for each replenishment cycle and the classic \(\beta\) service level) by adding various constraints. Moreover, the model considered penalty cost for backordering, and it can be adapted into lost sales scheme by introducing a parameter that presented the selling price per product in order to take into account the associated opportunity cost, which related to the demand that was not immediately satisfied under control policy as a prerequisite. Therefore, this research enabled the modelling for several variants with a fully linear formulation.

4.3 \((s,S)\) Policy

Computing \((s,S)\) policy parameters under non-stationary demand is a challenging task. The classic Silver and Meal heuristic algorithm (Silver and Meal, 1973) for deterministic demand was extended by Silver (1978) and Askin (1981) to study the lot sizing problem under nonstationary stochastic demand.

Silver’s algorithm was a stochastic version of (Silver and Meal, 1973). It used a deterministic model to calculate the number of periods that each order must cover; when this replenishment plan was known, the associated safety stocks were then myopically determined.

Askin explicitly included the cost effects of probabilistic demand in the choice of the number of periods for which to order (Askin, 1981). In contrast to Silver (1978), Askin used a least cost per unit time approach to determine the number of periods the immediate replenishment must cover.


Dural-Selcuk et al. implemented computational experiments of Askin (1981) and Bollapragada and Morton (1999) on a new common test bed. For both approaches they observed relatively large optimality gaps of 3.9% and 4.9%, respectively (Dural-Selcuk et al., 2016).

To overcome these shortcomings, Xiang et al. (2018) introduced a mixed integer non-linear programming (MINLP) formulation for \((s,S)\) system applying piecewise linearisation approximation method proposed by Rossi et al. (2015). Xiang et al. also derived an heuristic algorithm with binary search. Both solution methods outperformed the previous heuristics in computational efficiency for short and long time horizon tests. The comparison between two proposed algorithms found that binary search required a significantly less time than the MINLP.

Finally, Kilic and Tarim (2011) provided the grounds for measuring system nervousness in \((R,S)\) and \((s,S)\) system when demand was non-stationary. Tunc et al. (2013) conducted a numerical study to compare these policies for cost-effectiveness. Tunc et al. (2011) applied the stationary policy as the approximation to the optimal non-stationary system and found that stationary policies may be an efficient approximation to the optimal non-stationary system when demand information was of high uncertainty with a low penalty cost.

5. CONCLUSION

This paper surveyed works focusing on the single-item single-stocking location stochastic lot sizing problem under stationary and non-stationary demand. Conceptual maps were provided to classify works in the literature that appeared since the late Fifties. By separating works on grounds of demand (non)stationarity, we presented the development of the literature from the perspective of replenishment policies based on Bookbinder and Tan’s (1988) uncertainty strategies. For both types of demand, we investigated existing exact as well as heuristic formulations.


