Modeling Industrial Lot Sizing Problems: A Review
Raf Jans, Zeger Degraeve

To cite this version:

HAL Id: hal-00512940
https://hal.archives-ouvertes.fr/hal-00512940
Submitted on 1 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Modeling Industrial Lot Sizing Problems: A Review

<table>
<thead>
<tr>
<th>Journal:</th>
<th>International Journal of Production Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>TPRS-2005-IJPR-0281.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>State-of-the-Art Review</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>27-Apr-2006</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Jans, Raf; RSM Erasmus University, School of Management Degraeve, Zeger; London Business School, Decision Sciences</td>
</tr>
<tr>
<td>Keywords:</td>
<td>MODELLING, LOT SIZING</td>
</tr>
<tr>
<td>Keywords (user):</td>
<td></td>
</tr>
</tbody>
</table>
MODELING INDUSTRIAL LOT SIZING PROBLEMS: A REVIEW

Raf Jans
RSM Erasmus University
PO Box 1738, 3000 DR Rotterdam, The Netherlands
rjans@rsm.nl

Zeger De graeve
London Business School
Regent’s Park, London NW1 4SA, U.K.
zdegraeve@london.edu

Accepted July 5, 2006

Abstract

In this paper we give an overview of recent developments in the field of modeling deterministic single-level dynamic lot sizing problems. The focus of this paper is on the modeling of various industrial extensions and not on the solution approaches. The timeliness of such a review stems from the growing industry need to solve more realistic and comprehensive production planning problems. First, several different basic lot sizing problems are defined. Many extensions of these problems have been proposed and the research basically expands in two opposite directions. The first line of research focuses on modeling the operational aspects in more detail. The discussion is organized around five aspects: the set ups, the characteristics of the production process, the inventory, demand side and rolling horizon. The second direction is towards more tactical and strategic models in which the lot sizing problem is a core substructure, such as integrated production-distribution planning or supplier selection. Recent advances in both directions are discussed. Finally, we give some concluding remarks and point out interesting areas for future research.
1. Introduction

In this review, we will discuss models that have been developed for optimizing production planning and inventory management. Lot sizing models determine the optimal timing and level of production. They can be classified according to their time scale, the demand distribution and the time horizon. The famous Economic Order Quantity model (EOQ) assumes a continuous time scale, constant demand rate and infinite time horizon. The extension to multiple items and constant production rates is known as the Economic Lot Scheduling Problem (ELSP) (Elmaghraby 1978, Zipkin 1991). The subject of this review is the dynamic lot sizing problem with a discrete time scale, deterministic dynamic demand and finite time horizon. We will see that lot sizing models will incorporate more and more scheduling aspects. These scheduling models essentially determine the start and finish times of jobs (scheduling), the order in which jobs are processed (sequencing) and the assignment of jobs to machines (loading). Lawler et al. (1993) give an extensive overview of models and algorithms for these problems.


This review has a threefold contribution. Since the excellent reviews of Kuik et al. (1994) and Drexl and Kimms (1997) the research on dynamic lot sizing has further grown substantially. First of all, this paper fills a gap by providing a comprehensive overview of the latest literature in this field. Second, this paper aims to provide a general review and an extensive list of references for researchers in the field. Although this literature review is very extensive, we realize that it is impossible to be exhaustive. We realize that a model and its solution approach are inherently linked: more complex models demand also more complex solution approaches to solve them. However, in this paper we focus on the modeling aspect as much as possible in order to create some structure in the ever growing literature. This focus also distinguishes
this paper from other lot sizing reviews. A recent review of solution approaches can be found in Jans and Degraeve (2006). We show that the lot sizing problem is a core substructure in many applications by reviewing both more operational and tactical or strategic problems. Third, a comprehensive review further allows us to indicate new areas for further research. The power of production planning theory comes from the ability to solve more and more complex industrial problems. Whereas the early models where usually more compact, capturing the main trade-off, the extensions focus more and more on incorporating relevant industrial concerns. Therefore, this review is also very timely.

2. Lot Sizing Models

2.1. The single item uncapacitated lot sizing problem

The simplest form of the *dynamic lot sizing problem* is the single item uncapacitated problem:

\[
\text{Min } \sum_{t=1}^{m} (vc_t x_t + sc_t y_t + hc_t s_t)
\]

s.t. \( s_{t-1} + x_t = d_t + s_t \) \( \forall t \in T \) \hspace{1cm} (1)

\( x_t \leq sd_{mk} y_t \) \( \forall t \in T \) \hspace{1cm} (2)

\( x_t, s_t \geq 0; y_t \in \{0,1\} \) \( \forall t \in T \) \hspace{1cm} (3)

We have three key variables in each period \( t \): the production level \( (x_t) \), the set up decision \( (y_t) \) and the inventory variable \( (s_t) \). With each of these key variables is a cost associated: \( vc_t \), \( sc_t \) and \( hc_t \) are respectively the variable production cost, set up cost and holding cost in period \( t \). \( T \) is the set of all periods in the planning horizon and \( m \) is the last period. Demand for each period, \( d_t \), is known and \( sd_{mk} \) is the cumulative demand for period \( t \) until \( k \). The objective is to minimize the total cost of production, set up and inventory (1). We find here the same basic trade-off between set ups and inventory which is also present in the EOQ formula. Demand can be met from production in the current period or inventory left over from the previous period (2). Any excess is carried over as inventory to the next period. In each period we need a set up if we want to produce anything (3). As there is no ending inventory in an
optimal solution, production is limited by the remaining cumulative demand. Finally, the production and inventory variables must be positive and the set up variables are binary (4). This problem was first discussed in the seminal paper by Wagner and Whitin (1958). Zangwill (1969) showed that this problem is actually a fixed charge network problem. For a 5 period problem, the network can be depicted as shown in Figure 1. The arcs \((0, t)\) correspond to the production variables \(x_t\) and have an associated unit flow cost of \(vc_t\). If the production is strictly positive, i.e. \(x_t > 0\), then there is also a fixed cost of \(sc_t\) on the arc. The arcs \((t, t+1)\) correspond to the inventory variables \(s_t\) and have a unit flow cost of \(hc_t\). In network terms we say that node 0 is the supply or source node, nodes 1 to 5 are the demand nodes and the demand balance equations (2) correspond to the conservation of flow constraints.

**Fig. 1.** Network for the single item uncapacitated lot sizing problem

![Network Diagram](network-diagram.png)

2.2. Capacitated Multi-Item Lot Sizing Problem (CLSP)

Of course, companies do not have an unlimited capacity and usually they make more than one product. Any realistic model has to take this into account. How these two elements are modeled, depends on the mode of production and the choice of the time period. In the *large bucket* model, several items can be produced on the same machine in the same time period. In the *small bucket* model, a machine can only produce one type of product in one period.

The capacitated multi-item lot sizing problem (CLSP) is the typical example of a *large bucket* model. There are \(n\) different items that can be produced and \(P\) is the set
of all these items. In each period, only a limited production capacity $cap_i$ is available. Producing one unit of product $i$ consumes $vt_i$ units of capacity. The formulation is as follows:

$$\text{Min } \sum_{i \in P} \sum_{t \in T} \left( sc_{it} y_{it} + vc_{it} x_{it} + hc_{it} s_{it} \right)$$  \hspace{1cm} (5)$$

s.t. \hspace{1cm} s_{i,t-1} + x_{it} = d_{it} + s_{it} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (6)$$

$$x_{it} \leq My_{it} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (7)$$

$$\sum_{i \in P} vt_i x_{it} \leq cap_i \hspace{1cm} \forall t \in T \hspace{1cm} (8)$$

$$x_{it}, s_{it} \geq 0, y_{it} \in \{0,1\} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (9)$$

We observe that in this formulation, product specific variables and parameters now have an extra index $i$ to identify the item. For each item we have the demand balance equations (6) and set up constraints (7). The main difference with the uncapacitated model is the addition of the capacity constraint (8). In the set up constraint (7), the ‘big M’ is usually set equal to $\min\{cap_i, vt_i, sd_{it}\}$, as such the production is now limited by both the capacity and remaining demand (8).

The Continuous Set Up Lot Sizing Problem (CSLP) is a small bucket model:

$$\text{Min } \sum_{i \in P} \sum_{t \in T} \left( g_{it} z_{it} + sc_{it} y_{it} + vc_{it} x_{it} + hc_{it} s_{it} \right)$$  \hspace{1cm} (10)$$

s.t. \hspace{1cm} s_{i,t-1} + x_{it} = d_{it} + s_{it} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (11)$$

$$\sum_{i \in P} y_{it} \leq 1 \hspace{1cm} \forall t \in T \hspace{1cm} (12)$$

$$vt_i x_{it} \leq cap_i y_{it} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (13)$$

$$z_{it} \geq y_{it} - y_{i,t-1} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (14)$$

$$x_{it}, s_{it} \geq 0, y_{it}, z_{it} \in \{0,1\} \hspace{1cm} \forall i \in P, \forall t \in T \hspace{1cm} (15)$$

The new variable $z_{it}$ is the start up variable and there is an associated start up cost of $g_{it}$. A start up occurs when the machine is set up for an item for which it was not set up in the previous period. The objective function (10) minimizes the total cost of start ups, set ups, variable production and inventory. We still have the regular demand constraints (11). Further, we have the single mode constraint (12), imposing that at most one type of product can be made in each time period. For each item, production
can be up to capacity if there is a set up (13). The start up variables are modeled in constraint (14). There will only be a start up if the machine is set up for an item for which it was not set up in the previous period. A set up can be carried over to the next period if production of the same product is continued. Finally, the set up and start up variables are binary (15). Karmarkar and Schrage (1985) consider this problem without set up costs and called it the product cycling problem. Karmarkar et al. (1987) study the single item version of the CSLP, both for the uncapacitated and capacitated case. This problem is also referred to as lot sizing with start up costs (Wolsey 1989, Sandbothe 1991).

The Discrete Lot Sizing and Scheduling Problem (DLSP) is a small bucket lot sizing model with a discrete production policy: if there is any production in a period, it must be at full capacity. The generic model (Fleischmann 1990) has a similar structure as the CSLP (10)-(15), except that the capacity and set up constraint (13) becomes an equality:

\[ vt x_{it} = cap_i y_{it} \quad \forall i \in P, \forall t \in T \]  

(14)

Note that the production variable can be substituted out through this constraint. Jordan and Drexl (1998) showed the equivalence between DLSP for a single machine and the batch sequencing problem.

3. Further Extensions of Lot Sizing Models

Production planning problems are often classified according to the hierarchical framework of strategic, tactical and operational decision making (e.g. Bitran and Tirupati 1993). Depending on the decision horizon and level of aggregation, lot sizing models are usually classified as either tactical or operational models. A yearly master production schedule at the plant level is used for tactical planning. Production sequencing and loading are operational decisions and determining the lot sizes for products in the next month falls somewhere in between. We observe that the basic lot sizing models from the previous section are extended in two different directions. On one hand, lot sizing formulations include more operational and scheduling issues in order to model more accurately the production process, costs and demand side. We
organize these extensions around four topics: set ups, production, inventory and demand, but clearly some extensions relate to more than one of them. Here we also discuss the use of these models in a rolling horizon way. On the other hand, these models are incorporated into more tactical and strategic problems for which the operational lot sizing decisions are a core substructure.

3.1 Operational models

3.1.1. Extension on the set ups

Sometimes, there are not only set ups for individual items, called minor set ups, but there is a joint or major set up as well, which is incurred when at least one product is produced. These joint costs are used to model general economies of scale in manufacturing or procurement. This problem is extensively studied (Veinott 1969, Atkins and Iyogun 1988) and is sometimes referred to as the coordinated replenishment problem (Kao 1979, Chung et al. 1996, 2000, Robinson and Gao 1996, Robinson and Lawrence 2004) or the joint replenishment problem (Joneja 1990, Federgruen and Tzur 1994). In the case where the orders can be shipped via multiple modes, there is a different set up structure associated with each mode (Jaruphongsa et al. 2005). Production planning models without set ups have also been considered. Bowman (1956) shows that for the problem with convex cost functions, this problem can be solved as a transportation problem. Lotfi and Chen (1991) and Hindi (1995) discuss the capacitated case. On the other hand, sometimes the only objective is to minimize the costs of set ups or start ups. This is the case for the Changeover Scheduling Problem (CSP) (Glassey 1968, Hu et al. 1987, Blocher and Chand 1996a,b, Blocher et al. 1999). The problem assumes a discrete production policy and is as such related to the DLSP. A changeover is performed when production is switched to another product. In the DLSP terminology, this was called a start up. No inventory holding, production or set up costs are considered. Miller and Wolsey (2003) consider the discrete lot sizing problem with set ups but without start ups. Note that there is sometimes confusion in the literature between the terms set up and start up and we use the definition according to Vanderbeck (1998). In the production smoothing problem (Zangwill 1966b, Korgaonker 1977), a penalty proportional to the changes in production level is charged instead of a set up cost.
Several authors (e.g. Manne 1958, Kleindorfer and Newson 1975, Newson 1975, Trigeiro et al. 1989, Diaby et al. 1992, Du Merle et al. 1997, Armantano et al. 1999, Gopalakrishnan et al. 2001, Degraeve and Jans 2003, Hindi et al. 2003, Jans and Degraeve 2004) consider set up times for the CLSP. The set up times represent the capacity lost due to cleaning, preheating, machine adjustments, calibration, inspection, test runs, change in tooling, etc., when the production for a new item starts. The set up time must be accounted for in the capacity constraint. Salomon et al. (1991) and Cattrysse et al. (1993) consider start up times for the DLSP. They assume that start up times must equal an integral multiple of the time bucket, but it is also possible to model start up times which are a fraction of the time bucket (De Matta and Guignard 1994, Jans and Degraeve 2004). Vanderbeck (1998) formulates a CSLP with fractional start up times.

An often considered critique on the CLSP states that this model does not allow a set up to be carried over from one period to the next, even if the last product in one period and the first product in the next are the same. This has led to new models which allow for such a set up carry over, at the expense of introducing additional binary variables (Gopalakrishnan et al. 1995, Sox and Goa 1999, Gopalakrishnan 2000, Porkka et al. 2003, Gupta and Magnusson 2005). This problem is also referred to as the capacitated lot sizing problem with linked lot sizes (Suerie and Stadtler 2003). Computational results show that this model leads to considerable cost savings through the set up carry over (Gopalakrishnan et al. 2001). The Proportional Lot Sizing and Scheduling Problem (PLSP) relaxes the restriction of allowing production for only one product in each time period as imposed by the DLSP and CSLP. In the PLSP at most two different items can be produced in each time period. There is still at most one set up in each period, but the set up from the previous period can be carried over to the next period. Hence, if two items are produced in period \( t \), then the first item must be the same as the last item in the previous period. Drex1 and Haase (1995, 1996) discuss this model and extensions such as set up times and multiple machines. A further refinement allows the set up times to be split between two periods (Suerie 2006). Kimms (1996, 1999) presents the multi-level version of the PLSP.
The PLSP and the model with set up carry over are examples of lot sizing problems that incorporate more and more sequencing aspects. A further step for the CLSP is to determine a sequence for all the products within a time period, and not just for the first and last one. This is necessary if set up costs or times are sequence dependent (Dilts and Ramsing 1989, Haase 1996, Fleischmann and Meyr 1997, Kang et al. 1999, Laguna 1999, Clark and Clark 2000, Haase and Kimms 2000, Meyr 2000, Gupta and Magnusson 2005). In the General Lot Sizing and Scheduling Problem (Fleischmann and Meyr 1997), the macro-periods are divided into a fixed number of micro-periods with variable length, which allows the sequencing of products. Fleischmann (1994) considers the DLSP with sequence dependent start up costs. His heuristic procedure is based on the transformation of the problem into a Traveling Salesman Problem with Time Windows. Salomon et al. (1997) describe an algorithm for the DLSP with both sequence dependent start up costs and start up times. De Matta and Guignard (1994) also model sequence dependent cost and set up times in a DLSP in the process industry. In his review on change-over modeling Wolsey (1997) studies sequence dependent start ups for the CSLP. Belvaux and Wolsy (2000, 2001) present a comprehensive lot sizing model, including sequence dependent costs or times and switch off variables. Potts and Van Wassenhove (1992) discuss the integration of scheduling and lot sizing from a scheduling perspective.

Small set up costs and times are essential for implementing a successful Just-In-Time approach. Set up cost and time reduction programmes require an initial capital investment and result in a more flexible production. Zangwill (1987) points out that some intuitive implications of a set up reduction in an EOQ environment do not necessarily hold in the context of dynamic lot sizing. Mekler (1993), Diaby (1995) and Denizel et al. (1997) offer models to evaluate the tradeoff between the cost and benefits of a set up time reduction within a dynamic lot sizing framework. The set up times and costs are variables and depend on previous investment decisions. Another way to achieve lower set up costs is through learning. According to the theory of the learning curve, production costs decrease as cumulative output increases over time. Chand and Sethi (1990) present a lot sizing model with learning in set ups. Set up costs depend on the total number of set ups up to now and there is a declining set up cost for successive set ups. Benefits from smaller lot sizes are captured in terms of reduced set up costs. Tzur (1996) provide a more general model where the costs of a
set up still depend on the number of previous set ups, but can both decrease or increase, as long as the total set up costs are nondecreasing. Examples of increasing set up costs are the cases where the set up cost increases with the usage of the machine or when some maintenance is necessary after a specific number of set ups. Learning can also decrease the set up time (Pratsini 2000).

Almost all of the dynamic lot sizing models assume that production is done on reliable machines. Kuhn (1997) analyses the effects of set up recovery with machine breakdowns and corrective maintenance for the single item uncapacitated lot sizing problem. In a first case, the assumption is made that the set up is totally lost after a breakdown. In a second case, the costs of resuming production of the same item after a breakdown is lower compared to the original set up cost.

3.1.2. Extensions on the Production

In some manufacturing environments, production is done in batches (Lipmann 1969, Lee 1989, Pochet and Wolsey 1993, Constantino 1998, Van Vyve 2003). In the mathematical formulation, the set up variable $y_{it}$ becomes general integer instead of binary and indicates the number of batches produced. Every time production exceeds a multiple of the batch size, a new set up cost is incurred. This is for example the case in an environment where production is limited by a tank size. Each time one has to fill the tank again a set up cost is incurred, even if the same item is produced. This can also be interpreted as a stepwise cargo cost function (Lee 2004) where the capacity of each cargo is limited. Ben-Khedher and Yano (1994) assume that containers, which may be only partially filled, are assigned to trucks and there is a fixed charge for each truck used. Elmaghraby and Bawle (1972), Dorsey et al. (1974), Van Vyve (2003) and Li et al. (2004) impose that production is done in exact multiples of the batch size. Hence these models assume a discrete production policy but do not consider start up interactions over time. Manufactured units may not be available instantaneously, but arrive only in inventory after the whole batch has been completed (Brüggemann and Jahnke 1994, 2000). Stowers and Palekar (1997) and Bhatia and Palekar (2001) consider a variant of the joint replenishment lot sizing problem where products belonging to the same family can only be made in a fixed proportion to each other. A product can be part of several different families. This type of production occurs in the
manufacturing of metal or plastic plates and die-cast parts and in some chemical production problems. It is referred to as lot sizing with *strong set up interaction*. In an oil refinery, sets of products are produced simultaneously in same process, but the rate depends on the mode of operation (Persson et al. 2004). Love (1973) extends the lot sizing problem by introducing *lower and upper bounds on the production*. Production below some level is not allowed because of technical constraints or in order to make full use of the resource (Anderson and Cheah 1993, Constantino 1998, Mercé and Fontan 2003, Lee 2004).

In many production processes, *tools* such as dies or molds are required and they are often shared among several products. Tools, machines and products are interrelated as there are compatibility requirements between them (Brown et al. 1981). The problem is further complicated as there is only a limited availability of both the machines and tools. Jans and Degraeve (2004) model such a production planning problem for a tire manufacturer where the number of molds is a limiting factor. Akturk and Onen (2002) also integrate a lot sizing and tool management problem.

We observed that the boundaries between lot sizing and scheduling are fading with the introduction of sequence dependent set up costs and times. Lasserre (1992) and Dauzère-Péres and Lasserre (1994) provide a further example of this by integrating a classical multi-period lot sizing problem with a *job shop scheduling* problem. The lot sizing decision determines the due dates and processing times of the jobs. The capacity constraints are modelled at machine level by the regular job-shop precedence relations and disjunctive constraints. It is also an integration of discrete and continuous time planning models.

When *multiple parallel machines* are available, the lot sizing problem does not only include the timing and level of production, but also the allocation of production to machines. As such the loading decision has to be considered as well. Özdamar and Birbil (1998), Özdamar and Barbarosoğlu (1999), Kang et al. (1999), Clark and Clark (2000) and Belvaux and Wolsey (2000) extend the Capacitated Lot Sizing Problem with *multiple machines* with different production efficiencies. The DLSP and CSLP have also been extended with multiple identical machines (Lasdon and Terjung 1971).
and multiple machines with different efficiencies (Salomon et al. 1991, De Matta and Guignard 1994\textsuperscript{a,b}, 1995, Jans and Degraeve 2004\textsuperscript{a}).

*Production costs* can also change. Quantity discounts are sometimes considered in a lot sizing model, mostly in the case of purchasing decisions. Many quantity discount schemes result in a piecewise linear cost function (Shaw and Wagelmans 1998). Usually three types of quantity discounts are considered. The all-units discount (Prentis and Khumawala 1989, Chung et al. 1996) gives a reduction in the purchase price on all the units of a product if you buy more than a specific amount. Degraeve and Roodhooft (2000) model a multi-item purchasing environment where the discount is given on the total amount bought. Chan et al. (2002) propose a modified all-unit discount structure: if the total cost is higher than the total cost at the start of the next quantity interval, you only pay the lower cost. In the case of the incremental quantity discount (Diaby and Martel 1993, Chung et al. 1996, 2000), the reduction is only valid for the amounts in a specific interval. A third alternative is the truckload discount scheme (Li et al. 2004), where a less-than-truckload rate is charged until the total cost equal the truckload rate. If the total quantity is more than a truckload, this same scheme is applied for the excess quantity. Chu et al. (2005) consider a general economies-of-scale function for the ordering costs. When the production costs are actually distribution costs, van Norden and van de Velde (2005) argue that there is a dual cost structure. Any amount up to a reserved capacity is charged at a specific cost, and any amount above is charged at a higher cost.

*Cyclical schedules*, where the time between subsequent set ups is constant, are often used in practice. Bahl and Ritzman (1984) use cyclical schedules in a lower bounding heuristic. Campbell and Mabert (1991) impose such cyclical schedules for the CLSP with set up times. In their study, costs increases only by 5\% on average compared to the best non-cyclical schedules.

In the distribution, inventory control and production planning literature, there is a growing interest in *reverse logistics* (Fleischmann et al. 1997). Items return from the customers to the manufacturer and can be reused, either directly or after remanufacturing. Remanufacturing includes testing, repair, disassembly and reassembly operations. Another option is to recycle the scrap material or reuse some
parts as components. Taking these aspects into consideration requires an adaptation in
the production planning models (Fleischmann et al. 1997, Guide 2000). Few dynamic
lot sizing models have been proposed to accommodate such changes. Richter and
Sombrutzki (2000) extend the ULS with remanufacturing. Demand can be met either
from newly manufactured products or from return products which have been
remanufactured. These two product categories have different set up costs (van den
Heuvel 2004) or a joint set up cost (Teunter et al. 2005). In practice, the two
categories also have different unit (re)manufacturing costs and there is the possibility
of disposal of some of the returned products (Richter and Weber 2001). Yang et al.
(2005) analyze the problem with concave costs. Beltrán and Krass (2002) consider the
case where the returned goods are in good enough condition to be resold immediately
without remanufacturing. In their model, demand can be negative due to the returns
and they allow for the disposal of excess inventory. Kelle and Silver (1989) model a
similar problem, but take into account the uncertainty in the arrival of the returned
goods. They impose a service level constraint and next transform the problem into an
ULS with negative demands. In a mathematical programming model for order
quantity determination in a purchasing context, Degraeve and Roodhooft (1999)
icorporate additional revenues due to repurchases of old products by the supplier.

3.1.3. Extensions on the inventory

The inventory can also be bounded by upper and lower limits (Love 1973, Swoveland
by imposing a lower bound on the inventory in each period. Fixed charges on the
stocks (Van Vyve and Ortega 2004) are useful for an environment with complex
stocking operations or for situations where there are combinatorial constraints on
stocks, such as in the chemical industry where only one type of product can be stored
in a tank.

Martel and Gascon (1998) make a subtle change to the classical uncapacitated lot
sizing problem in a purchase context. The unit purchase cost can vary over time and
the holding cost is price-dependent, whereas in the standard model, the unit inventory
holding cost can vary, but is known in advance and does not depend on the purchase price. The inventory cost is calculated as the purchase price multiplied by a constant inventory holding charge.

Lot sizing problems have also been extended with the issue of *perishable inventory*. Veinott (1969) permits the proportional growth or deterioration of inventory. Hsu (2000) and Chu et al. (2005) consider the uncapacitated single item lot sizing problem with an age dependent inventory cost as well as an age dependent deterioration rate where a part of the inventory is lost by carrying it to the next period. Jain and Silver (1994) look at the problem with random life time perishability. According to some stochastic process, the total inventory becomes either worthless or remains usable for at least the next period.

### 3.1.4. Extension on the demand

By allowing *backlogging* (e.g. Zangwill 1966, Pochet and Wolsey 1988, Federgruen and Tzur 1993) demand can be met by production in a later period at a specific cost. Backlogging corresponds in fact to a negative inventory level. The objective function includes the backlog cost. We can use inventory from the previous period, allow backlog or produce now to satisfy demand, build up inventory or satisfy backlog from a previous period. The case with backlogging is also a single source fixed charge network problem (Zangwill 1969). Backlogging results in a flow from demand point $t$ to $t-1$, in the opposite direction of the inventory flow. Swoveland (1975) imposes that orders are not backlogged for more than a prescribed number of periods. The extension with backlogging is also considered for the DLSP (Jans and Degraeve 2004) and for the coordinated replenishment problem (Robinson and Gao 1996). Hsu and Lowe (2001) study the problem with age dependent backlog costs. Another way of satisfying demand if not enough products are available in time is product *substitution* (Hsu et al. 2005). Demand for e.g. a low quality product can also be met by offering the high quality product at the price of the low quality product. It is also possible that a product first need to be transformed, which leads to an extra conversion cost.
Lot sizing models with \textit{stockouts} (Sandbothe and Thompson 1990, 1993, Aksen et al. 2003) have been proposed as an alternative to situations where backlogging is allowed. When demand cannot be met in time, \textit{lost sales} are incurred instead of backlogging. A variable $l_t$ representing the lost sales is added into the demand equation and the cost of a stockout is properly accounted for in the objective function.

Considering \textit{sales} (Brown et al. 1981, Kang et al. 1999, Loparic, Pochet and Wolsey 2001) instead of fixed demands leads to a profit maximization approach instead of the traditional cost minimization. The demand equation is extended with a variable $v_t$ for the sales and an upper bound of $d_t$ is imposed on these potential sales. The unit selling price is given. Hung and Chien (2000) model a profit maximization approach with different demand classes that have different profitabilities. The two models of maximizing sales and minimizing costs with lost sales are equivalent as demand can be rewritten as the sum of the sales and lost sales.

Lee et al. (2001) discuss the single item uncapacitated dynamic lot sizing problem with a \textit{demand time window}. For each demand an earliest and latest delivery date is specified and demand can be satisfied in this period without penalty. They prove that there exists an optimal solution in which demand is not split: the complete demand for a specific order is covered by production from the same period. Hwang and Jaruphongsa (2006) analyse the case with a speculative cost structure. An extension to a two-echelon supply chain is provided in Jaruphongsa et al. (2004). Another extension is the incorporation of time windows for the suppliers who are shipping the goods via a crossdock (Lim et al. 2005). Also Wolsey (2006) makes a distinction between production time windows and delivery time windows.

3.1.5. \textit{Time Horizon}

Schedules are usually implemented in a \textit{rolling horizon} fashion. Only the first period of a plan is implemented and the demand forecast is updated by looking one period further. A new plan is calculated using this updated input. Experiments (Baker 1977, Blackburn and Millen 1980) have indicated that the Wagner-Whitin algorithm is no longer optimal in a rolling horizon framework and simple lot sizing heuristics may
outperform optimal algorithms for small planning horizons. New research (Simpson 2001), however, indicates that the Wagner-Whitin rule still outperforms all the other heuristics in a wide variety of cases. For an extensive review on rolling horizons and related literature we refer the reader to Chand et al. (2002). The rolling horizon approach can result in nervousness of the planning in the sense that schedules must be frequently changed. These changes might result in extra costs (Kazan et al. 2000). Recently, two methods have been proposed to mitigate the end-of-horizon effect basically by looking beyond the current planning horizon. Stadtler (2000) takes only a fraction of the costs for the last set up period into account. The last set up may be advantageous beyond the planning horizon, so only a proportion of the set up cost has to be borne within the planning horizon. Fisher et al. (2001) impose an appropriate level of ending inventory by assigning a positive value to this ending inventory in the objective function. This positive value is calculated by estimating the future set up costs that are avoided as a result of the ending inventory. Experiments in both papers indicate that these adaptations are quite effective in general and they usually outperform simple heuristics or the unadapted WW algorithm. Van den Heuvel and Wagelmans (2005) point out that the superior performance of Fisher et al. (2001) is mainly due to the availability of accurate information about future demand.

3.2. Tactical and strategic models

Hierarchical production planning (Hax and Meal 1975, Bitran, Haas and Hax 1981, Graves 1982, Bitran and Tirupati 1993) is a sequential procedure for solving production planning at different levels of aggregation. First, decisions are taken at the highest level and they set the limitations for the decisions at a lower level. Items are aggregated into families and families into types. A type is a set of items that have a similar demand pattern and have one aggregate production rate. Within a family, items share the same set up. In the aggregate model, at the type level, the main decision is the determination of the level for regular and overtime capacity and the total production for a type. This decision sets bounds on the model for the family production planning. Here the objective is to minimize the total set up cost for all families within a type. Within the limits of the family batch sizes, production quantities are determined for each item. An application of this type of planning is
described in Liberatore and Miller (1985) and Oliff and Burch (1985). There is a lot of interaction between the different levels and the sequential optimization does usually not result in a global optimum. The reason that detailed integrated models are yet scarce is of course their complexity, which make them difficult to solve. Yet there exist some models which integrate lot sizing and decisions at higher hierarchical levels.

Aggregate planning extends the lot sizing models further at a more tactical level by including labor resource decisions such as hiring and firing (e.g. Dzielinski, Baker and Manne 1963, Thomas and McClain 1993, Nam and Logendran 1992, Aghezzaf 2000). The decision on set up cost and time reduction, as discussed in the previous section, can also be viewed as the integration of lot sizing in a more tactical decision. Another example is the integration of lot sizing and capacity expansion decisions. Rao (1976) considers a lot sizing model where additional capacity can be bought in each period. The extra capacity is also available in all the subsequent periods, in contrast with temporary capacity expansion from overtime. Rajagopalan and Swaminathan (2001) optimize the capacity acquisition, production and inventory decisions over time in an environment with increasing demand. There is the following trade-off: Capacity investments can be postponed by building up inventory earlier. On the other hand, buying additional capacity can lead to smaller inventories by reducing lot sizes. Bradley and Arntzen (1999) discuss a case study where simultaneously considering capacity and inventory decisions leads to superior financial results. In the model of Atamturk and Hochbaum (2001) variable demand can be met by production, inventory, subcontracting or capacity acquisition. A second way of modeling capacity expansion problems is to define the demand as the demand for the incremental capacity (Luss 1982, Chand et al. 2000, Hsu 2002) and there is no modeling of the underlying production lot sizing anymore.

Interest in modeling the interaction between the production stage and upstream (suppliers) and downstream (distribution) activities is growing. Whereas previously these relationships were considered at an aggregate level, it is useful to consider inbound logistics, production and distribution simultaneously at an operational level. The general message of these models is that strategic or tactical decisions, such as
investment decisions or supplier selection, should take operational concerns, in this case lot sizing, into account.

The lot sizing structure shows up as a subproblem in the supplier selection problem considered by Degraeve and Roodhooft (1999, 2000) and Degraeve, Labro and Roodhooft (2000). They model the total cost of ownership at four different levels, i.e. supplier, order, batch and unit level. These costs include the regular order cost, purchase price and inventory holding cost, but can also take into account quality differences, discounts, reception costs for batches and the cost for managing the relationship with the supplier. Basnet and Leung (2005) study a special case with a supplier-dependent order cost.

Global models for supply chain optimization simultaneously consider production, transportation and demand planning. The coordination of production and distribution planning results in cost savings compared to separate optimization of these activities (Chandra and Fisher 1994). Herer and Tzur (2001) consider a multi-location single-item problem where transshipments of stock between locations at the same distribution level are allowed. Transshipments can be beneficial in a deterministic environment due to the presence of fixed order costs or differences in holding costs. Another extension is the inclusion of capacitated vehicles to model the transportation between two stages (Anily and Tzur 2005, 2006). A case study in the fertilizer industry (Haq et al. 1991) describes a multi-level structure with plants, warehouses and retailers and incorporates set up costs and times, production and distribution lead times and the costs for production, set up, inventory and transportation. Martin et al. (1993), Diaby and Martel (1993), Kaminsky and Simchi-Levi (2003), Lee et al. (2003), Jolayemi and Olorunniwo (2004) and Sambasivan and Yahya (2005) discuss models for integrated production, distribution and inventory planning. Bhutta et al. (2003) also take capacity investment decision into account, next to the production and distribution problems. Federgruen and Tzur (1999) study a two-echelon distribution network with one warehouse and many retailers and model it as a dynamic lot sizing problem. Timpe and Kallrath (2000) describe an actual application in the chemical process industry combining batch and campaign production with change-overs in a multi-site setting, raw materials inventory management, transportation between production sites and sales point, inventories at the sales points, different prices for different customers and external purchase possibilities. Their objective function is maximizing the total contribution. Van Hoesel et al. (2002) consider a serial supply
chain with one manufacturer and multiple warehouses and they globally optimize the production, inventory and transportation decisions. A general overview of interesting issues for global supply chain optimization at the supplier, plant and distribution stage is given in Erengüç et al. (1999).

4. Conclusions and New Research Directions

While some extensions are motivated from the literature or general practical observations (e.g. Jaruphongsa et al. 2005, Van Vyve and Ortega 2004), many of the extensions discussed in Section 3 are inspired by a specific real life problem. Typically, the authors relate their model to a case observed in practice: “This paper was motivated by a production planning problem encountered while working with a multinational manufacturing firm” (Miller and Wolsey 2003); “Motivated by a problem faced by a large manufacturer of a consumer product, we explore the interaction between production planning and capacity acquisition decisions.” (Rajagopalan and Swaminathan 2001); “A real-world problem in a company manufacturing steel rolled products provided motivation to this research” (Sambasivan and Yahya 2005). This provides sound industrial validation for the research on extended lot-sizing models. Some of these papers try to fit the real life case into a more generally applicable model and test it on randomly generated data sets. This is usually done to analyse their algorithm and check the impact of some specific parameters (e.g. Gupta and Mangusson 2005, Haase and Kims 2000, Porkka et al. 2003, Robinson and Lawrence 2004, Sambasivan and Yahya 2005.). Other papers analyse the real life problem, provide a new formulation and perform computational experiments on real life data (e.g. Belvaux and Wolsey 2000, 2001, Degraeve and Roodhooft 1999, De Matta and Guignard 1994, Grunow et al. 2002, Jans and Degraeve 2004², Kang et al. 1999, Meyer 2000, Miller and Wolsey 2003, Rajagopalan and Swaminathan 2001). For some case studies, the practical implementation and actual company impact is also discussed in the paper (Bradley and Arntzen 1999, Haq et al. 1991, Lee and Chen 2002, Liberatore and Miller 1985, Martin et al. 1993, Oliff and Burch 1985, Timpe and Kallrath 2000, Van Wassenhove and De Bodt 1983). The applications of deterministic lot-sizing can be found in many different industries: injection moulding (Brown et al. 1981, Van Wassenhove and De Bodt 1983), glass industry (Martin et al. 1993), consumer products (Rajagopalan and

The numerous extensions of the basic lot sizing problem show that it can be used to model a variety of industrial problems. Boundaries between lot sizing and scheduling are fading and further integration of lot sizing, sequencing and loading constitute a challenging research track. Lot sizing on parallel machines is just one example for new research opportunities. Also the increased attention to model specific characteristics of the production process and to accurately represent costs will be valuable in solving real life planning problems. Further, the integration of lot sizing into more global models opens an interesting area for further research. In the case where products have to be manufactured and shipped to different distribution centers, retailers or end customers, it makes sense to consider production and distribution simultaneously at an operational level. In such a situation we should consider fixed and variable costs for both production and transportation and coordinate lot sizing, vehicle loading and routing decisions. New models could also take into account the coordination between multiple plants or further downstream activities such as packing. Another research direction is coordination of lot sizing with decisions from other functional areas such as demand planning and pricing decisions (van den Heuvel and Wagelmans 2006, Geunes et al. 2006, Deng and Yano 2006) in marketing, as is done for other lot sizing models (Goyal and Gunasekaran 1995, Kim and Lee 1998, Abad 1996, 2003).

Pochet (2001) indicates that modeling production planning problems in the process industry constitute a promising area for new research, whereas most of the lot sizing literature is focused on discrete manufacturing. Some distinguishing characteristics,
such as the use of flexible recipes, the existence of by-products, the integration of lot sizing and scheduling, storage constraints and a focus on profit maximization, affect the planning and scheduling. A further general discussion can be found in Crama et al. (2001) and Kallrath (2002, 2005), whereas some specific problems are presented in Smith-Daniels and Ritzman (1988), Selen and Heuts (1990), Heuts et al. (1992), Grunow et al. (2002, 2003), Rajaram and Karmarkar (2004), Persson et al. (2004).

One of the major limitations of the lot sizing models that we discuss in this review is the assumption of deterministic demand and processing times. In many manufacturing environments, there is some degree of uncertainty. As a consequence, the performance evaluation criteria differ in a deterministic and stochastic environment. The deterministic lot sizing models formulate the problem as a trade-off between inventory costs and set up costs. However, they fail to capture the problem of queuing and congestion found in stochastic manufacturing environments (Banker et al. 1988, Rummel 2000). Karmarkar (1987) examines the relationship between lot sizes and lead times in such a stochastic environment. He argues that the lead time is a good proxy for many costs, as larger lead times adversely affect the work-in-process inventory, the safety stock and responsiveness to the customer. Karmarkar shows that there exists a convex relationship between lot sizes and lead times. Small batches lead to more set ups, a higher utilization rate and consequently to longer lead times. This is the saturation effect. On the other hand, larger lot sizes lead to increased lead times due to the batching effect. Deterministic and stochastic lot sizing models also suggest qualitatively different optimal solutions. From basic queuing theory it is known that when the capacity utilization increases to 100%, the waiting times become infinite. This suggests that some slack capacity should be included in the planning. In deterministic capacitated lot sizing models, however, the capacity constraint will be binding in many periods and so there is no slack capacity available. In a stochastic environment, the schedule stability also becomes an important objective (Bourland and Yano 1994). Stochastic inventory models (e.g. Eppen and Martin 1988, Zipkin 2000) and lot sizing models based on queuing theory (e.g. Karmarkar 1987, Suri et al. 1993, Lambrecht et al. 1998), are more appropriate to capture the complexity of a stochastic environment. The combination of such models are used to analyse integrated production-inventory systems where inventory replenishment rules trigger production and can cause queues in the manufacturing environment (e.g. Van Nyen et
al. 2005). Hence, one should be careful in the choice of model and verify that the underlying assumptions and trade-offs are a good approximation of the reality. For a further discussion of deterministic versus stochastic planning models, we refer to Karmarkar (1993). Zhang and Graves (1997) incorporate random machine failures in their analysis of cyclic schedules for products with a stable demand rate. Some research has been done to incorporate uncertainty into the dynamic lot sizing problem as well such as stochastic demand (Bookbinder and Tan 1988, Sox and Muckstadt 1996, 1997, Sox et al. 1999, Tarim and Kingsman 2004, Guan et al. 2006), stochastic lead times (Nevison and Burstein 1984), uncertainty in demand timing (Burstein et al. 1984, Gutiérrez et al. 2004), or a combination of demand and supply uncertainty (Anderson 1989).

Finally, the interaction between modeling and algorithms will play an important role in future research. The inclusion of industrial concerns lead to larger and more complex models and consequently more complex algorithms are needed to solve them. Solution approaches for integrated models will be based on previous research on the separate models. Existing knowledge about the structure and properties of a specific subproblem can be exploited in solving integrated models. Lot sizing problems are challenging because many extensions are very hard to solve. Jans and Degraeve (2005) review several techniques to tighten the formulations (Dantzig-Wolfe decomposition, Lagrange relaxation, cutting planes and variable redefinition) and to obtain good quality solutions using (meta-) heuristics. The development of algorithms based on the combination of some of these techniques has already led to promising results. Vanderbeck (1998) combines branch-and-price and cutting planes for solving the CLSP. Dantzig-Wolfe decomposition can be combined with Lagrange relaxation to speed up the column generation process, either by using Lagrange relaxation to solve the master (Cattrysse et al. 1993, Jans and Degraeve 2004a) or by using Lagrange relaxation to generate new columns (Degraeve and Jans 2003). To obtain stronger bounds, Lagrange relaxation is applied to a variable redefinition reformulation (Jans and Degraeve 2004b). Many more opportunities for combining algorithms are still largely unexplored.
References


Hsu, V.N., Dynamic Economic Lot Size Model with Perishable Inventory. *Management Science*, 2000, 46 (8), 1159-1169.


Özdamar, L., Barbarosoğlu, G., Hybrid heuristics for the multi-stage capacitated lot sizing and loading problem. *Journal of the Operational Research Society*, 1999,


