Valuing Lead Time in a Dynamic Order-Up-To Inventory System

Jente Van Belle[a,d], Ruben Crevits[b], Tias Guns[c], Wouter Verbeke[a]

[a] *Research Centre for Information Systems Engineering (LIRIS), KU Leuven, Leuven 3000, Belgium, jente.vanbelle@kuleuven.be, wouter.verbeke@kuleuven.be*
[b] *OMP, Wommelgem 2160, Belgium, rcrevits@omp.com*
[c] *Data Lab, Solvay Business School, Vrije Universiteit Brussel, Brussels 1000, Belgium, tias.guns@vub.be*
[d] *Presenting author*

Abstract

The lead time in an inventory system has an impact on inventory (holding and/or shortage) costs as well as on the purchase or production costs. Usually, shorter lead times result in lower inventory costs while also entailing higher purchase or production costs. The former results from a reduced exposure to demand uncertainty, while the latter is due to expediting the production and/or delivery process which typically comes with additional expenses. de Treville et al. (2014) [Production and Operations Management, vol. 23(12), 2103-2117] propose a real-options model to determine the value of lead time in a single-period newsvendor setting by quantifying the effect of reducing (or extending) the lead time on the exposure to forecast uncertainty in terms of a justified increase (or required decrease) in unit purchase or production cost. The difference in unit purchase or production cost quantifies the offset to compensate for the change in exposure to forecast uncertainty, and hence the change in inventory costs. In this paper, we consider a periodic review dynamic order-up-to (OUT) inventory system in a nonstationary demand environment in which forecast distributions are a crucial input to the inventory system. We extend the model proposed by de Treville et al. (2014) to this dynamic OUT setting. Finally, we apply the extended model to a case study and illustrate its use as decision support tool for deciding what lead time to use.

Keywords: Inventory, Decision support, Forecast uncertainty, Lead time, Nonstationary demand.

1. Introduction

When faced with stochastic and nonstationary or time-varying demand, inventory management and order scheduling require accurate short-term probabilistic demand forecasts as inputs. These forecasts, which inherently account for demand uncertainty, are usually produced on a daily, weekly or monthly basis. Fig. 1 depicts the information flow within a typical inventory management system for nonstationary stochastic demand. Given a specific inventory policy, such as the periodic review dynamic order-up-to (OUT) policy (Babai and Dallery, 2009), in each review period, probabilistic demand forecasts together with system parameters form the inputs to a replenishment and inventory optimization model. This optimization model outputs an order quantity which may affect replenishment decisions in subsequent review periods, depending on the inventory control policy in place. The order quantity represents either a production or purchase order, depending on whether the product considered is manufactured in-house or purchased from an external supplier. One important system parameter is the lead time. This is the time between the moment when a decision is made about the quantity to order and the moment when this amount is available to satisfy demand.

In a nonstationary setting, this parameter is naturally linked to the probabilistic demand forecasts needed as input to the system, as the lead time directly determines the (maximum) forecast horizon to be used. Note that while forecasts are typically updated each period, system parameters are kept fixed (in the short-term).

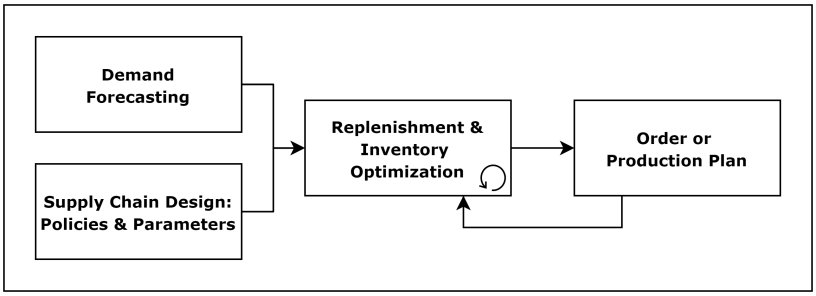


Figure 1: Inventory management system for nonstationary stochastic demand.

Usually, the lead time is assumed to be given, being either deterministic or probabilistic. However, in many practical situations lead time is controllable. Reducing the lead time is considered to come with an extra cost which results from expediting the production and/or delivery process. For instance, if the product is purchased from an outside supplier, the lead time can be reduced by opting for a faster but more expensive transportation mode; and in a manufacturing setting, lead time can be reduced by switching to a more responsive supplier of product components but this will usually require explicit premium payments. Conversely, extending the lead time is expected to come with a decrease in purchase or production cost. In addition to the effect on purchase or production cost, lead time also affects exposure to demand forecast uncertainty. Typically, extending the lead time (and thus the forecast horizon) will result in additional uncertainty. A shorter lead time, on the other hand, generally implies less exposure to forecast uncertainty, resulting in lower safety stock levels and expected shortages for a given target service level. Controlling the lead time thus translates into a trade-off between inventory holding and shortage costs on the one hand, and purchase or production costs on the other.

In this paper, we focus on aligning the controllable lead time with forecasting capabilities taking into account the cost implications of different lead time candidates in a periodic review dynamic OUT inventory system. More specifically, our contribution is that we extend the real-options model proposed by de Treville et al. (2014b), which provides support in making lead time decisions in a single-period newsvendor setting, to this dynamic OUT setting. Given that we are provided with (i) values for the inventory system parameters and a specification of the purchase or production cost for the current lead time and (ii) specifications of forecast uncertainty for different candidate lead times, our extended framework can be used to quantify the required decrease (justified increase) in unit purchase or production cost to compensate for larger (smaller) exposure to forecast uncertainty resulting from an extension (reduction) of the lead time. While both inventory models are suitable when demand is stochastic (and nonstationary), they are different in that excess products and unfilled customer orders are not carried over from one period to the next in the single-period newsvendor model, whereas in a dynamic OUT inventory system (which is a multi-period inventory model per definition) they are carried over to the next period. This difference implies that the inventory models have different revenue streams and cost structures. Moreover, they are also affected differently by a change in lead time as carryover of excess products and unfilled customer

orders implies that we have to account for the uncertainty associated with the *cumulative* demand forecast over the lead time in an OUT system, which is not the case in a newsvendor setting.

The remainder of this paper is structured as follows. In Sec. 2, we discuss related work. The periodic review dynamic OUT inventory system is discussed in Sec. 3, and the framework to value lead time in this setting is covered in Sec. 4. In Sec. 5, a short case study is presented to illustrate the use of the framework as decision support tool. Concluding remarks are formulated in Sec. 6.

2. Literature Review

In this section, we first provide a brief overview of the literature on inventory management in which lead time is treated as a decision variable. We further discuss some works that relate to the identified research gaps in this strand of literature. Second, we briefly discuss related work on generating probabilistic demand forecasts which are necessary inputs to a dynamic inventory system. Finally, as our proposed framework, presented in Sec. 4, is an extension of the model proposed by de Treville et al. (2014b), we explain their model in detail in Sec. 2.3.

2.1. Controllable Lead Time

A classic assumption in inventory management theory is that the lead time is an exogenous parameter, either a constant or a random variable, determined by vendor policies or manufacturing processes. With the rise of the just-in-time (JIT) philosophy, which focuses on efficiency and eliminating the waste of excessive inventory, academics started to research inventory systems with controllable lead times. This is because lead time is controllable in many practical situations and reducing the lead time in an inventory system is an effective way to achieve the JIT goal.

An early study that deals with a controllable lead time in a quantitative manner is van Beek and van Putten (1987). They consider lead time to be a strategic variable that is controllable in the long-term by for example, making arrangements with suppliers to cut supply lead times. Their model is based on trading off the cost of reducing the lead time against logistic savings (i.e., lower levels of inventory and stockouts) achieved by the effect of a shorter lead time on demand uncertainty, with demand modeled as a stationary probability distribution and the cost of reducing the lead time as a rational function of lead time, the momentary lead time and a constant cost parameter. Hill and Khosla (1992) present a more general model for determining the optimal lead time reduction in the sense that they use a more general specification for the cost of reducing the lead time, often referred to in the literature as (lead time) crashing cost.

In a series of other papers, a different functional form is considered for the lead time crashing cost: lead time is decomposed into several components each having a different cost for reducing the lead time with one period (i.e., lead time crashing cost is formalized using a piecewise linear function). Liao and Shyu (1991) were the first to use this setup for optimizing lead time in a continuous review reorder point/order quantity (R,Q) inventory system by minimizing an approximation of the expected total relevant cost (shortages are neglected) with lead time being the unique decision variable. Hariga and Ben-Daya (1999) expand on the above study by considering periodic review inventory control systems. The aforementioned studies all consider a controllable but deterministic lead time. In this work, we follow this approach. However, note that there also exists work in which lead time is considered to be controllable and stochastic (e.g., Chaharsooghi and Heydari (2010)). For stochastic lead times, controllability can relate to average lead time as well as lead time variability.

It is clear that an extensive literature exists regarding the controllability of lead time in inventory systems. The above studies focus on developing more general models with more (and more) decision variables and alternative functional forms for the crashing cost. However, these studies all rely on the following assumptions:

1. *Demands in different time periods are independent and identically distributed (i.i.d.);*
2. *The crashing cost can be explicitly characterized as a function of lead time.*

The i.i.d. assumption allows one to focus on static inventory policies and to specify the distribution of demand over lead time directly as a function of the lead time L, and the demand per time period Y. For instance, if the demand per time period Y is normally distributed $N(\mu, \sigma)$, the distribution of demand during a lead time of length L is $D_L \sim N(L\mu, \sqrt{L}\sigma)$. In combination with an explicit characterization of the crashing cost as a function of lead time, the above allows us to determine the optimal lead time via optimization techniques.

As for the first assumption, in industrial settings, stationary i.i.d. demand is less common than nonstationary demand due to for example, seasonal patterns and trends. If demand is nonstationary, the optimal inventory control policies are also nonstationary or dynamic (Tunc et al., 2011). Regarding the second assumption, Hill and Khosla (1992) already indicated that it is by no means easy to characterize the crashing cost as a function of lead time, since firms typically do not have data available about changes in costs that result from a change in lead time. More recently, de Treville et al. (2014b) proposed a model that provides support in making lead time decisions in the single-period newsvendor problem. Instead of characterizing the crashing cost as a function of lead time, their model outputs the justified increase (or required decrease) in unit purchase or production cost to compensate for the effects of a reduction (or extension) of the lead time. Extensions and applications of the model are presented in de Treville et al. (2014a,b) and Biçer et al. (2018). In this work, we will follow the same approach in that we will not explicitly specify the crashing cost as a function of lead time. However, as discussed in Sec. 1, their model is developed to serve as decision support tool in a single-period newsvendor setting, while we focus on aligning the controllable lead time with forecasting capabilities in a (multi-period) dynamic OUT system.

2.2. Probabilistic Demand Forecasts

A dynamic inventory system requires a probabilistic forecast of demand over the lead time in each period. While there is no single approach to estimating this forecast distribution, a frequently used approach is to first produce point forecasts (mean forecasts) at the highest available frequency and then aggregate these forecasts over the lead time. In a second step, forecast errors that arise from making these point forecasts are used to estimate forecast uncertainty. To this end, it is often assumed that forecast errors are i.i.d. and follow a normal distribution with zero mean (i.e., forecasts are unbiased). Under these assumptions, a similar result as for the lead time demand distribution above is obtained for the forecast uncertainty over the lead time: it is normally distributed with zero mean and standard deviation $\sigma_L = \sqrt{L}\sigma$ with σ the RMSE of the one-step ahead in-sample forecast errors. However, this expression for σ_L was strongly criticized in Chatfield (2000). Barrow and Kourentzes (2016) provide empirical evidence that this estimation procedure results in poor approximations of the out-of-sample forecast uncertainty. This is a consequence of the i.i.d. assumption which ignores (i) potential covariances between the forecast errors of the different time periods forming the cumulative demand over the lead time and (ii) the fact that forecast errors

usually increase for longer forecast horizons. To allow for possible deviations from the i.i.d. assumption, Trapero et al. (2019) recommend to use empirical or data-driven approaches if sufficient data is available. More specifically, they propose to estimate the cumulative forecast uncertainty for each lead time separately and directly based on a collection of out-of-sample lead time demand forecast errors. They present both parametric and nonparametric approaches. Depending on the assumptions adopted, these empirical approaches also allow correcting for bias in lead time demand forecasts and capturing potential deviations from normality. In this paper, we follow this recommendation in that we assume that a separate forecast uncertainty estimate is available for each candidate lead time.

2.3. Valuing Lead Time in the Single-Period Newsvendor Problem

The real-options model proposed by de Treville et al. (2014b) to determine the value of lead time in the single-period newsvendor problem allows one to determine the required decrease (justified increase) in unit purchase or production cost to compensate for higher (lower) demand forecast uncertainty exposure coming from an extension (reduction) of the lead time.

In the single-period newsvendor problem with a lead time of L periods, the company places an (internal) order at time $t-L$ to meet the stochastic demand at delivery time t. By committing the order quantity at time $t-L$, the company inherently exposes itself to demand uncertainty. This gives rise to the use of a probabilistic forecast $Y_{t-L,t}$ of the demand for period t in deciding how much to order at time $t-L$. Ordering too few items will result in lost sales, while ordering too many will result in overage costs. Given a per unit selling price p, a unit purchase or production cost c_L which is assumed to be decreasing in lead time L, and salvage value s, the shortage cost is $C_s = p - c_L$ representing an opportunity cost, while the overage cost is $C_o = c_L - s$. It is well known that the optimal (profit-maximizing) order quantity $K_{t-L,t}$ for a risk-neutral decision maker can be computed as follows:

$$K_{t-L,t} = F^{-1}_{Y_{t-L,t}}(\alpha_L) \tag{1}$$

with $F^{-1}_{Y_{t-L,t}}(.)$ the inverse of the cumulative distribution function (CDF) $F_{Y_{t-L,t}}(x) = P(Y_{t-L,t} \leq x)$ of the forecasted demand $Y_{t-L,t}$ for period t, and α_L the optimal service level. The lead time L thus determines the forecast horizon to be used in forecasting demand for period t. Further, it is assumed that forecast uncertainty is increasing in lead time. The optimal service level α_L is the probability that demand can be fully satisfied and is a function of overage and shortage costs:

$$\alpha_L = \frac{C_s}{C_s + C_o} = \frac{p - c_L}{p - s}. \tag{2}$$

Notice that the optimal service level α_L depends on the lead time L via the unit purchase or production cost c_L. For a derivation of the result for the optimal order quantity $K_{t-L,t}$, one may refer to Ch. 13 of Chopra and Meindl (2013).

At time t, the company observes actual demand of y_t units and makes profit π:

$$\begin{aligned}
\pi(c_L, K_{t-L,t}, y_t) &= p\min(y_t, K_{t-L,t}) - c_L K_{t-L,t} + s\max(K_{t-L,t} - y_t, 0), && (3)\\
&= p(K_{t-L,t} - \max(K_{t-L,t} - y_t, 0)) - c_L K_{t-L,t} + s\max(K_{t-L,t} - y_t, 0), && (4)\\
&= (p - c_L)K_{t-L,t} - (p - s)\max(K_{t-L,t} - y_t, 0). && (5)
\end{aligned}$$

The first and second terms in Eq. (3) represent revenue from regular sales and purchase or production costs, respectively. The third term is the revenue earned from salvaging left-over units.

After rearranging the expression in Eq. (3) to arrive at Eq. (5), de Treville et al. (2014b) rely on the observation that the payoff structure in the second term of this equation resembles the risk exposure experienced by the writer of $(p-s)$ European put options on demand y_t with strike price $K_{t-L,t} \geq 0$ and maturity $L \geq 0$. This analogy allows one to rely on well-known results from the financial option pricing literature, such as the Black-Scholes option pricing formulas, to calculate the expected net present value of profit Π at time $t-L$, that is, the time when the order quantity decision has to be made:

$$\Pi(c_L, K_{t-L,t}, Y_{t-L,t}) = e^{-rL} E_{\mathbb{Q}} \left[(p-c_L) K_{t-L,t} - (p-s) \max(K_{t-L,t} - Y_{t-L,t}, 0) \right], \tag{6}$$

$$= (p-s) e^{-rL} E_{\mathbb{Q}} \left[\left(\frac{p-c_L}{p-s} \right) K_{t-L,t} - \max(K_{t-L,t} - Y_{t-L,t}, 0) \right], \tag{7}$$

with $r \geq 0$ the continuously compounded risk-free rate and $\mathbb{Q}$ a so-called risk-neutral measure. Using a risk-neutral measure instead of the physical one allows one to avoid the incorporation of the decision maker's risk preferences in (i) the discount rate used in the calculation of the expected net present value (so that we can use the risk-free rate r), and in (ii) calculating the optimal order quantity $K_{t-L,t}$. After all, a risk-averse newsvendor is known to optimally order less than the profit-maximizing quantity (Eeckhoudt et al., 1995). The application of risk-neutral valuation relies on the principle of absence of arbitrage. This principle is based on the economic argument that the stochastic payoff of the option, which depends on the probabilistic demand forecast, can be perfectly replicated by constructing a portfolio of tradable financial securities that are correlated with the demand forecast. For more details on the shift from the physical to the risk-neutral measure, one may refer to the original work of de Treville et al. (2014b).

If the company has the possibility to place the order at different moments $t-L$ with $L \in \mathscr{L} = \{0, 1, \ldots, L_{\max}\}$, it will be indifferent towards the adopted lead time L if the expected net present values of the profits Π (at time $t - L_{max}$) for the different lead times are equal:

$$\Pi(c_0, K_t, y_t) = \Pi(c_1, K_{t-1,t}, Y_{t-1,t}) = \cdots = \Pi(c_{L_{\max}}, K_{t-L_{\max},t}, Y_{t-L_{\max},t}), \tag{8}$$

where we have assumed that $r = 0$ to simplify the expression. For $L = 0$, it holds that $K_t = y_t$ as the order quantity is determined only after actual demand is observed. Note that a lead time $L = 0$ can be purely hypothetical as it assumes that (external) production and delivery can happen instantaneously right after demand is observed. Therefore, we obtain for $L = 0$ an expected net present value $\Pi(c_0, K_t, y_t) = (p - c_0) y_t$ and for each $L \in (\mathscr{L} \setminus \{0\})$ the following indifference condition (with respect to $L = 0$):

$$\left(\frac{p-c_0}{p-s} \right) y_t = \left(\frac{p-c_L}{p-s} \right) K^{\mathbb{Q}}_{t-L,t} - E_{\mathbb{Q}} \left[\max(K_{t-L,t} - Y_{t-L,t}, 0) \right], \tag{9}$$

with $K^{\mathbb{Q}}_{t-L,t}$ as in Eq. (1) but computed with respect to the risk-neutral measure $\mathbb{Q}$.

Given the unit purchase or production cost for the current lead time $c_{L_{\text{current}}}$, with L_{current} being a specific element of $\mathscr{L}$, the indifference conditions above allow us to solve for c_L for each $L \in (\mathscr{L} \setminus \{L_{\text{current}}\})$. Consequently, one can compute the required or justified differences in unit purchase or production cost between the different candidate lead times to offset the differences in profit due to changes in exposure to forecast uncertainty.

In the simplest form of the model, de Treville et al. (2014b) assume that the demand forecast evolves in lead time following a geometric Brownian motion with constant volatility. In other

words, they model the demand forecasts $Y_{t-L,t}$ for different lead times L as lognormally distributed random variables. Under this setup, the seminal Black-Scholes option pricing formulas can be used to reformulate the expectation in Eq. (9). The authors also present more general versions of the model by relying on advanced closed-form option pricing formulas that take into account stochastic volatility and heavy-tailedness of the underlying demand forecast distributions.

In Sec. 4, we build on the model outlined above to develop a framework that can be used to make lead time decisions in a (multi-period) dynamic OUT inventory system instead of a single-period newsvendor setting. The dynamics of the former system are described in the next section.

3. Dynamic OUT Inventory System

In this section, we first outline the process of managing inventory via a dynamic OUT system. Next, we discuss a possible approach to obtain the probabilistic lead time demand forecasts needed as input to this system, taking into account the recommendations made by Trapero et al. (2019).

3.1. System Dynamics

The single-item inventory system that we consider in this work is a periodic review dynamic OUT or a state-dependent base stock system in which unmet demands are fully backordered. This model is appropriate when purchase or production costs are linearly increasing in quantity (i.e., fixed ordering or setup costs are negligible). Besides the purchase or production costs, also the holding and backorder costs are considered to evolve linearly in quantity. Further, we consider the setting in which there is no product obsolescence, nor a capacity constraint.

If we manage inventory via this system, the goal is to maintain a target service level α, which can be interpreted in this setting as the percentage of periods in which demand can be satisfied from inventory-on-hand directly (cf. α_L in Sec. 2.3), with minimum inventory and backorders. In order to achieve this goal, at the end of each period t, we first review inventory-on-hand I_t by updating the beginning inventory position I_{t-1}. To this end, (i) items received or produced in that period Q_t are added to inventory and (ii) demand y_t is satisfied (fulfilled or backordered):

$$I_t = I_{t-1} + Q_t - y_t. \tag{10}$$

Second, the quantity to receive or for which production should be finished in period $t+L$ is determined for $L > 1$ as follows:

$$Q_{t+L} = \max(K_{t,L} - I_t - \sum_{i=1}^{L-1} Q_{t+i}, 0), \tag{11}$$

with $K_{t,L}$ the optimal order quantity before taking into account inventory-on-hand I_t and work-in-progress (WIP) quantities $\sum_{i=1}^{L-1} Q_{t+i}$. For $L \leq 1$, the same formula can be used, provided that the WIP term is omitted. Use of the $\max(.,0)$ function corresponds to the assumption that WIP quantities are never canceled.

At the end of each period t, we thus need to calculate the optimal order quantity before inventory adjustments $K_{t,L}$ so as to determine Q_{t+L} via Eq. (11). Obtaining $K_{t,L}$ requires solving a newsvendor problem (see Eq. (1)) with target service level α:

$$K_{t,L} = F^{-1}_{D_{t,L}}(\alpha), \tag{12}$$

with $F^{-1}_{D_{t,L}}(.)$ the inverse CDF of the lead time demand forecast $D_{t,L}$ covering periods $t+1$ to $t+L$. The optimal service level α can be calculated as follows:

$$\alpha = \frac{C_s}{C_s + C_o} = \frac{b}{b+h}, \tag{13}$$

with b and h the unit backorder and holding cost (per period), respectively (Chopra and Meindl, 2013). Note that in practice, it can be very difficult to estimate the unit backorder cost, and therefore often the target service level is specified by expressing the unit backorder cost as a multiple of the unit holding cost. For instance, a target service level of 90% implies that backordering one unit is nine times more expensive than holding one unit of inventory (for one period).

3.2. Lead Time Demand Forecast Distributions

In estimating the optimal quantity before inventory adjustments $K_{t,L}$ via Eq. (12), in each period t, we need as input the CDF of the lead time demand forecast $D_{t,L}$ covering periods $t+1$ to $t+L$. The lead time L thus affects the number of periods that should be taken into account in estimating forecast uncertainty (and the associated forecast horizon(s) to be used).

Recall from Sec. 2.2 that there is no single approach to forecast the distribution of demand over lead time. A frequently used approach is to first produce point forecasts of expected demand over the lead time, $\hat{d}_{t,L}$, covering periods $t+1$ to $t+L$, and use the forecast errors that arise from making these point forecasts in a second step to estimate forecast uncertainty. In this regard, Trapero et al. (2019) recommend the use of a data-driven approach to estimate out-of-sample lead time forecast uncertainty. More specifically, they propose to estimate the (cumulative) forecast uncertainty $\varepsilon_{t,L}$ for each lead time separately and directly based on a collection of out-of-sample lead time demand forecast errors. While this approach allows capturing deviations from the i.i.d. assumption, which underlies uncertainty estimation based on one-step ahead in-sample forecast errors (see Sec. 2.2), it does require the collected lead time demand forecast errors to originate from the same underlying distribution. To this end, one can rely on change point detection methods to check whether the entire collection of lead time demand forecast errors can be used in estimating the forecast uncertainty (see e.g., Aminikhanghahi and Cook (2017) for a survey of change point detection methods). In what follows, we assume that this requirement is always met.

If errors are independent of observed lead time demands, an additive error structure can be used: $D_{t,L} = \hat{d}_{t,L} + \varepsilon_{t,L}$ where the distribution of $\varepsilon_{t,L}$ is based on the N most recent out-of-sample lead time demand forecast errors $e_{s,L} = d_{s,L} - \hat{d}_{s,L}$ with $d_{s,L} = \sum_{i=1}^{L} y_{s+i}$ and $s = t-L-(N-1), \ldots, t-L$. Often a normal distribution is fitted to the N errors. However, a different and nonparametric approach that is shown to work well in Trapero et al. (2019) is to use the sample α-quantile of the N errors to directly estimate $K_{t,L}$. When we expect that the error grows proportionally with the observation value (multiplicative error structure), we can use the lognormal distribution to characterize the lead time demand forecast: $D_{t,L} = \hat{d}_{t,L} e^{\varepsilon_{t,L}}$ with $\varepsilon_{t,L}$ a normal distribution fitted to the N most recent lead time demand forecast errors $e'_{s,L} = \ln(d_{s,L}) - \ln(\hat{d}_{s,L})$.

4. Valuing Lead Time in a Dynamic OUT Inventory System

In this section we develop a framework along the lines of the model presented in Sec. 2.3 for the single-period newsvendor problem to value lead time in the dynamic OUT inventory system as described in Sec. 3. More specifically, for products for which inventory is managed via a dynamic

OUT system, the proposed framework allows determining the required or justified differences in unit purchase or production cost between different candidate lead times to offset differences in inventory (holding plus backorder) costs due to changes in exposure to forecast uncertainty. While the output of our framework is similar to that of the model proposed by de Treville et al. (2014b), recall from Secs. 2.3 and 3.1 that excess products and unfilled customer orders are dealt with differently in the single-period newsvendor problem and the dynamic OUT inventory system. This causes these systems to have different revenue streams and cost structures, and to be affected differently by a change in lead time. Below, we describe our resulting framework which deals with these differences.

4.1. Outline of the Framework

In the dynamic OUT inventory system with a lead time of $L > 0$ periods, in each period $t+L$, the company updates its inventory position I_{t+L} according to Eq. (10) and incurs a total cost ψ_{t+L} consisting of both purchase or production and supply-demand mismatch costs:

$$\psi_{t+L} = c_L Q_{t+L} + h\max(I_{t+L},0) + b|\min(I_{t+L},0)|. \tag{14}$$

The first term in Eq. (14) represents purchase or production costs, while the second and third terms capture holding and backorder costs, respectively. However, as the aim of this framework is to quantify the effect of a change in lead time on these costs, but lead time is not allowed to change very frequently as it requires changing operational processes, we are interested in the long-run average total cost ψ per period:

$$\psi(c_L, I_L) = c_L\bar{y} + h\max(I_L,0) + b|\min(I_L,0)|, \tag{15}$$

with $\bar{y}$ the average actual single-period demand and I_L the average inventory position associated with lead time L. Indeed, the average purchase or production quantity is not affected by changes in lead time as unmet demands are fully backordered and there is no product obsolescence. Demand is thus fulfilled and ordered in full eventually, irrespective of the lead time. On the other hand, changing the lead time does affect the average inventory position since forecast uncertainty over the lead time is taken into account in planning replenishment and inventory (via Eqs. (11) and (12)).

Let us denote with Ψ the expected value of the long-run average total cost ψ per period. By expanding Eq. (10) via repeatedly substituting Eq. (10) into itself, it follows that any inventory position can be decomposed as:

$$I_{t+L} = I_t + \sum_{i=1}^{L} Q_{t+i} - D_{t,L}. \tag{16}$$

An approximation to Eq. (16) can be obtained by relying on Eq. (11). By ignoring the $\max(.,0)$ function, which makes it possible that $K_{t,L}$ is lower than inventory-on-hand plus WIP ($I_t + \sum_{i=1}^{L} Q_{t+i}$), and after some rearrangements, it follows that:

$$I_{t+L} \approx K_{t,L} - D_{t,L}. \tag{17}$$

Using this result, in expectation, we obtain the following expression for Ψ:

$$\Psi(c_L, K_L, D_L) \approx c_L\bar{y} + hE\left[\max(K_L - D_L, 0)\right] + bE\left[\max(D_L - K_L, 0)\right], \tag{18}$$

where K_L is computed as in Eq. (12) but with respect to D_L which represents the distribution of average lead time demand. In Sec. 3.2, different approaches to model $D_{t,L}$ were presented. D_L can be modeled following the same approaches. However, in this case we rely on ε_L, the distribution representing average forecast uncertainty over the lead time, and $\hat{d}_L = L\bar{y} - \mu_L$ (or $\hat{d}_L = L\bar{y}e^{-\mu_L}$) with μ_L the mean of ε_L to capture forecast bias. The distribution representing average forecast uncertainty over the lead time, ε_L, is characterized based on all available most recent lead time demand forecast errors that originate from the same underlying distribution (see Sec. 3.2).

As in Sec. 2.3, it is reasonable to assume that the company will be indifferent towards different candidate lead times $L \in \mathscr{L} = \{0, 1, \ldots, L_{\max}\}$ if the expected net present values of the long-run average total costs per period Ψ are equal. For the single-period newsvendor problem, de Treville et al. (2014b) rely on the concept of risk-neutral valuation to avoid having to take into account risk preferences in discounting the expected values and in calculating the optimal order quantities. However, as the dynamic OUT inventory system inherently involves repetitive decisions on the optimal order quantity, and these decisions typically do not expose the company to costs that are large relative to the company's total worth, it is usually valid to assume that the company acts as a risk-neutral decision maker in this setting (Mian, 2011). This means that we can directly rely on Eq. (12) to determine the optimal order quantity before inventory adjustments, K_L, and that we do not have to take into account risk premiums. If we again assume that $r = 0$ for simplicity, then it holds for $L = 0$ that $D_0 = K_0 = \bar{y}$ so that $\Psi(c_0, K_0, D_0) = c_L\bar{y}$, that is, there are no holding and backorder costs. For each $L \in (\mathscr{L} \setminus \{0\})$ we then obtain the following indifference condition (with respect to $L = 0$):

$$(c_0 - c_L)\bar{y} = hE\left[\max(K_L - D_L, 0)\right] + bE\left[\max(D_L - K_L, 0)\right]. \tag{19}$$

To obtain the required or justified differences in unit purchase or production cost c_L for the different candidate lead times, we can solve the indifference condition(s) above for c_L for each $L \in (\mathscr{L} \setminus \{L_{\text{current}}\})$, starting from the unit purchase or production cost for the current lead time $c_{L_{\text{current}}}$. To this end, we need to evaluate the expectations in Eq. (19). The evaluation of these expectations depends on the approach that is used to model forecast uncertainty, i.e., the distribution of average lead time demand D_L (see Sec. 3.2). For the sake of brevity, however, in what follows we will only focus on an additive error structure in combination with a normal distribution.

4.2. Two Approaches to Evaluate the Expectations in Eq. (19)

Analytical Approach based on Normal Distribution. If D_L is modeled as a normally distributed variable $L\bar{y} - \mu_L + \varepsilon_L$ with $\varepsilon_L \sim N(\mu_L, \sigma_L)$ calibrated based on the entire collected sample of forecast errors $e_{t,L} = d_{t,L} - \hat{d}_{t,L}$, the expectations in Eq. (19) can be evaluated as follows:

$$\begin{aligned}
K_L &= L\bar{y} + \sigma_L\Phi^{-1}(\alpha), && (20)\\
d &= \frac{K_L - L\bar{y}}{\sigma_L}, && (21)\\
E\left[\max(K_L - D_L, 0)\right] &= (K_L - L\bar{y})\Phi(d) + \sigma_L\phi(d), && (22)\\
E\left[\max(D_L - K_L, 0)\right] &= (L\bar{y} - K_L)\left[1 - \Phi(d)\right] + \sigma_L\phi(d), && (23)
\end{aligned}$$

with $\phi(.)$ and $\Phi(.)$ the density function and CDF of a standard normal random variable, respectively. For a derivation of these results, one may refer to Ch. 13 of Chopra and Meindl (2013).

Inventory Simulation Approach. An alternative approach to evaluate the expectations in Eq. (19) is to rely on the results of an inventory simulation. Based on historical sales data for periods $t = 1,\ldots,N_1,\ldots,N_2,\ldots,N_3$, an inventory simulation can be performed for each lead time $L \in (\mathscr{L} \setminus \{0\})$ by sequentially executing the following three steps:

Step 1. Produce lead time demand forecasts $\hat{d}_{t,L}$ from forecast origins $t = N_1,\ldots,N_3 - L$ by gradually expanding the training data used to generate the forecasts. N_1 is the minimum number of observations used to produce these point forecasts;

Step 2. Based on the outputs of the first step, the optimal order quantity before inventory adjustments $K_{t,L}$ can be determined for $t = N_2 + 1 - L,\ldots,N_3 - L$. To determine $K_{t,L}$, we need both the lead time demand point forecast $\hat{d}_{t,L}$ and a collection of out-of-sample lead time demand forecast errors to characterize the uncertainty $\varepsilon_{t,L}$ around this forecast. This collection of out-of-sample errors is updated through time to use the most recent available $N_2 - N_1$ errors for each t. In other words, to estimate forecast uncertainty, a rolling window with a window length of $N_2 - N_1$ is used;

Step 3. By relying on the outputs of the second step, (order quantities Q_{t+L} and) inventory positions I_{t+L} can be simulated for $t = N_2 + 1,\ldots,N_3$ by taking into account actual demand (starting from $Q_{t+L} = 0$ and $I_{t+L} = 0$ for $t \leq N_2$).

The expectations in Eq. (19) can then be evaluated as follows:

$$E\left[\max(K_L - D_L, 0)\right] \approx \frac{1}{N_3 - N_2} \sum_{t=N_2+1}^{N_3} \max(I_{t+L}, 0), \tag{24}$$

$$E\left[\max(D_L - K_L, 0)\right] \approx \frac{1}{N_3 - N_2} \sum_{t=N_2+1}^{N_3} |\min(I_{t+L}, 0)|. \tag{25}$$

Note that $\frac{1}{N_3-N_2}\sum_{t=N_2+1}^{N_3} Q_{t+L} \approx \bar{y}$ should hold for each L, as we consider a system in which all demand is fulfilled and thus ordered, irrespective of the lead time. Finally, also note that this inventory simulation approach requires that sufficient data is available as a first part of the out-of-sample lead time demand forecast errors is only used in estimating forecast uncertainty for $t = N_2 + 1 - L$ in the second step and no inventory positions are simulated for the corresponding time periods. In contrast, the analytical approach presented above uses all available errors to evaluate the expectations in Eq. (19).

5. Numerical Illustration

In this section, we apply the presented framework to a case study in order to illustrate how it can be used to determine what lead time to use, given that we are provided with (i) values for the inventory system parameters and a specification of the purchase or production cost for the current lead time and (ii) historical data to estimate forecast uncertainty for the different lead time candidates considered.

First, we provide more information on the case study and the necessary inputs. Next, we present the application of the proposed framework which allows the company to assess whether it can benefit from changes in lead time.

5.1. Case Study Description

We consider a single high-volume Stock Keeping Unit (SKU) of the company Volys, a Belgian manufacturer of poultry products. For this SKU, replenishment and inventory is managed via the dynamic OUT policy as outlined in Sec. 3.1. Currently, a lead time of 4 days is used for the SKU under consideration. The company wants to assess whether they can benefit from aligning the controllable lead time with its forecasting capabilities, taking into account the cost implications of different candidate lead times $L = 0, \ldots, 5$ days. Hereto, our proposed framework will be used to calculate the required or justified difference in unit production cost c_L for the different candidate lead times so that the current total cost, which comprises inventory holding, backorder and production costs, would remain unchanged. Note that the current total costs are assumed to be minimized in that the target service level used to manage replenishment and inventory is assumed to be optimal (see next subsection).

Inventory System Parameters and Production Cost. In order to optimize replenishment and inventory decisions in a dynamic OUT inventory system, we need to specify the unit holding cost h and the unit backorder cost b. Moreover, for the application of our proposed framework, we also need a specification of the current unit production cost c_4. Recall from Sec. 3.1 that in practice it is usually difficult to estimate b in absolute terms and therefore it is often specified by expressing it as a multiple of h. For the SKU under study, the company targets a service level α of 90%, which implies that backordering one unit is considered nine times more expensive than holding one unit of inventory (for one period). Therefore, if we assume that $h = 1$, it holds that $b = \alpha/(1-\alpha) = 9$. Finally, this relative characterization of costs also requires that c_4 is expressed as a multiple of h. The company estimates that the cost of holding one unit of inventory for one day is currently equal to 25% of the unit production cost, which translates into $c_4/h = 4$.

Demand Forecast Inputs. To forecast lead time demand, the company first makes 1 to 5-days ahead point forecasts, $\hat{y}_{t,t+1}, \ldots, \hat{y}_{t,t+5}$, using exponential smoothing methods. These daily forecasts are then aggregated to obtain the lead time demand point forecast: $\hat{d}_{t,L} = \sum_{i=1}^{L} \hat{y}_{t,t+i}$. To model forecast uncertainty over the lead time, historical out-of-sample lead time demand forecast errors that arise from making these lead time demand point forecasts are used to characterize an additive error structure in combination with a normal distribution. Extreme values in the sample of lead time demand forecast errors are discarded by using a 2% trim (at each end of the sample). 400 lead time demand forecasts and forecast errors are available. The company uses a rolling window containing 250 errors to estimate forecast uncertainty. This setup allows us to simulate inventory positions I_{t+L} for the last 150 periods (as the first 250 periods are needed to initialize the inventory simulations). Recall from Secs. 3.2 and 4.1 that the collection of lead time demand forecast errors used to characterize (average) cumulative forecast uncertainty over the lead time should originate from the same underlying distribution for each of the lead time candidates considered. We assume that this requirement is satisfied in all cases.

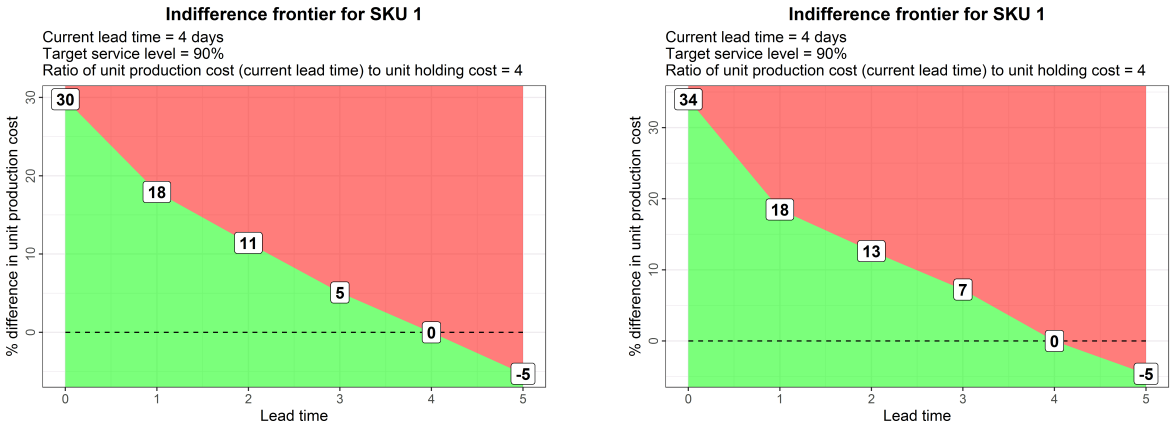
5.2. Valuing Lead Time with the Proposed Framework

The framework proposed in Sec. 4.1 allows us to effectuate the trade-off between production and inventory (holding plus backorder) costs by quantifying for each candidate lead time the (minimum) required decrease or (maximum) justified increase in unit production cost to compensate for a change in exposure to forecast uncertainty resulting from an extension or reduction of the lead

time. By combining the differences in unit production cost (relative to the current unit production cost c_4) for the different candidate lead times, we can obtain a so-called indifference frontier. Each point on this frontier represents a combination of a lead time and a unit production cost difference that results in the same total cost. Such an indifference frontier thus allows answering the following questions for the candidate lead times:

- For a reduction of the lead time: how much more expensive can production (per unit) be so as to obtain equal total cost?
- For an extension of the lead time: how much cheaper should production (per unit) be so as to obtain equal total cost?

Unit production cost indifference frontiers for the example SKU are depicted in Fig. 2. The left panel visualizes the indifference frontier obtained via the analytical approach, whereas in the right panel, we rely on the inventory simulation approach (see Sec. 4.2).



(a) Analytical

(b) Inventory simulation

Figure 2: Unit production cost indifference frontiers for the example SKU.

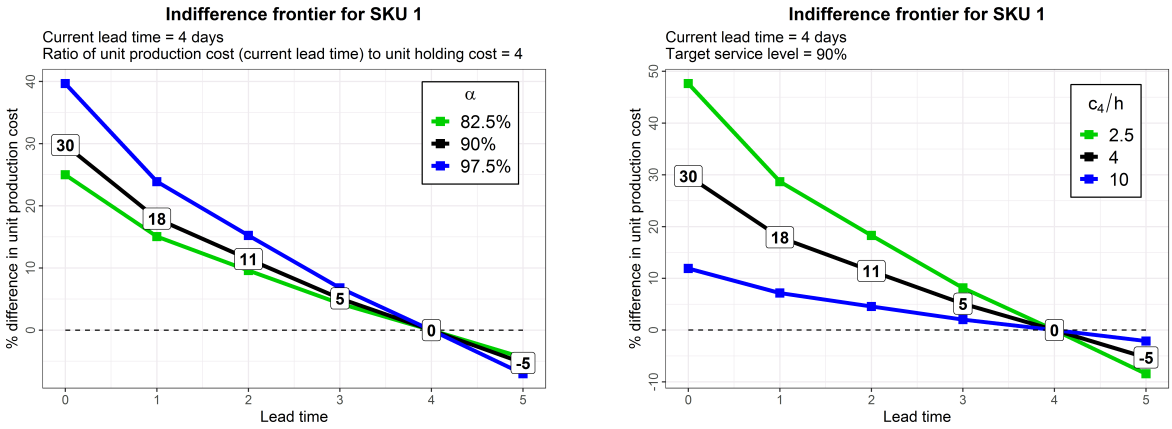
To determine what lead time to use for a specific SKU, an indifference frontier needs to be combined with management estimates of attainable changes in unit production cost, which are based on insights obtained from analyzing changes in operational processes and their associated effects on the production cost via managerial accounting systems. For instance, the indifference frontiers depicted in Fig. 2a and 2b show that an extension of L from 4 to 5 days requires a minimum decrease in unit production cost of 5%. If management finds, for example, that this lead time extension allows allocating less direct labor to the production of this item which results in a unit production cost decrease of more than 5%, this lead time extension would result in total cost savings.

By comparing Figs. 2a and 2b, we observe that there are some small differences between the frontiers obtained via the analytical and inventory simulation approach. A first cause of these observed differences is that deviations from the modeling assumptions only affect the results that rely on the inventory simulations. Moreover, the indifference frontier obtained via the analytical approach is not directly affected by observed actuals. The effect of outlying observations is therefore a second reason for the observed differences. Finally, note that if the actual production quantities

deviate from the optimal ones (as output by the dynamic OUT system) due to operational constraints such as minimum batch sizes or capacity limits that are accounted for in a postprocessing phase, only indifference frontiers based on inventory simulations should be used. While constructing the frontiers based on inventory simulations is computationally more demanding and requires relatively more data, the flexibility that it provides allows it to be used in more situations.

5.3. Effect of Input Parameters on Indifference Frontier

In Fig. 3, for illustrative purposes, we visualize the effects on the indifference frontier of changing (i) the target service level and (ii) the ratio of current unit production cost to unit holding cost, keeping all other inputs unchanged.



(a) Target service level α

(b) c_4/h-ratio

Figure 3: Effect of α and c_4/h-ratio on the indifference frontier obtained via the analytical approach.

From Fig. 3a, it is clear that a higher α causes the indifference frontier to tilt away from the x-axis, while a lower α tilts the frontier towards the x-axis. This follows from the fact that, ceteris paribus, a higher α results in higher inventory (holding plus backorder) costs so that they account for a larger share of the total cost. Hence, trading off these higher inventory costs against production costs gives rise to larger required and justified percentage differences in unit production cost. Note that the relation between α and inventory costs is nonlinear which causes the increase in α from 90% to 97.5% to have a larger effect on the frontier than the decrease in α from 90% to 82.5%.

In Fig. 3b, we visualize the indifference frontiers for c_4/h equal to 2.5, 4 and 10, keeping all other inputs unchanged. These values can be interpreted as follows: holding one unit of inventory for one day is equal to 40%, 25% and 10% of the unit production cost for the current lead time, respectively. From Fig. 3b we observe that a higher value for c_4/h causes the indifference frontier to tilt towards the x-axis, while a lower value tilts the frontier away from the x-axis. This result is intuitive as a higher value for c_4/h implies that the inventory costs represent a smaller share of the total cost, resulting in smaller required and justified percentage differences in unit production cost.

Although both the target service level α and the ratio c_4/h are considered as a given for each SKU, in the context of a large product portfolio, the above insights are useful to help determine for which products the value of lead time should be assessed first.

6. Conclusion

An important parameter in an inventory management system is the lead time. Changing the lead time directly affects the probabilistic demand forecast that is needed as input to the system as the lead time determines the (maximum) forecast horizon to be used. In this way, reducing or extending the lead time causes a change in the exposure to forecast uncertainty, which in turn results in changes in safety stock levels and expected shortages or backorders for a given target service level. However, changing the lead time generally also affects purchase or production costs. Controlling the lead time thus translates into a *trade-off* between *inventory costs* on the one hand, and *purchase or production costs* on the other. In this paper, we extend the model proposed by de Treville et al. (2014b) to quantify the effect of a change in lead time in terms of a justified or required change in unit purchase or production cost from a single-period newsvendor setting to a periodic review dynamic OUT setting.

A limitation of the presented framework is that we only consider the use of a target service level to determine optimal order quantities. However, in practice, sometimes order quantities are set based on a target fill rate instead. If demand is characterized by a parameterized distribution, however, there exists a unique service level that gives rise to the target fill rate. Therefore, the framework as presented (with the normal distribution to model demand) can be readily extended in this direction. Another limitation of our framework is that it only considers demand as a source of uncertainty. In practice, however, inventory systems are sometimes characterized by both demand and supply uncertainty, caused by factors such as quality problems and machine breakdowns. In this regard, an interesting topic for future research is to extend the framework to quantify the effects of a change in lead time on the exposure to both demand and supply uncertainty.

To conclude, we believe that our proposed framework, in combination with a data-driven approach to estimate forecast uncertainty, is a useful tool to support managerial decision making regarding the lead time to use in a periodic review dynamic OUT inventory system.

Funding

This research is funded by VLAIO (HBC.2016.0589) and OM Partners NV.

References

Aminikhanghahi, S., Cook, D.J., 2017. A survey of methods for time series change point detection. Knowledge and Information Systems 51, 339–367.

Babai, M.Z., Dallery, Y., 2009. Dynamic versus static control policies in single stage production-inventory systems. International Journal of Production Research 47, 415–433.

Barrow, D.K., Kourentzes, N., 2016. Distributions of forecasting errors of forecast combinations: Implications for inventory management. International Journal of Production Economics 177, 24–33.

van Beek, P., van Putten, C., 1987. OR contributions to flexibility improvement in production/inventory systems. European Journal of Operational Research 31, 52–60.

Biçer, I., Hagspiel, V., de Treville, S., 2018. Valuing supply-chain responsiveness under demand jumps. Journal of Operations Management 61, 46–67.

Chaharsooghi, S.K., Heydari, J., 2010. LT variance or LT mean reduction in supply chain management: Which one has a higher impact on SC performance? International Journal of Production Economics 124, 475–481.

Chatfield, C., 2000. Time-series forecasting. CRC press.

Chopra, S., Meindl, P., 2013. Supply Chain Management: Strategy, Planning, and Operation. 5 ed., Pearson.

Eeckhoudt, L., Gollier, C., Schlesinger, H., 1995. The risk-averse (and prudent) newsboy. Management science 41, 786–794.

Hariga, M., Ben-Daya, M., 1999. Some stochastic inventory models with deterministic variable lead time. European Journal of Operational Research 113, 42–51.

Hill, A.V., Khosla, I.S., 1992. Models for optimal lead time reduction. Production and Operations Management 1, 185–197.

Liao, C.J., Shyu, C.H., 1991. An analytical determination of lead time with normal demand. International Journal of Operations & Production Management 11, 72–78.

Mian, M.A., 2011. Project economics and decision analysis: Deterministic models. PennWell Books.

Trapero, J.R., Cardós, M., Kourentzes, N., 2019. Empirical safety stock estimation based on kernel and GARCH models. Omega 84, 199–211.

de Treville, S., Biçer, I., Chavez-Demoulin, V., Hagspiel, V., Schürhoff, N., Tasserit, C., Wager, S., 2014a. Valuing lead time. Journal of Operations Management 32, 337–346.

de Treville, S., Schürhoff, N., Trigeorgis, L., Avanzi, B., 2014b. Optimal sourcing and lead-time reduction under evolutionary demand risk. Production and Operations Management 23, 2103–2117.

Tunc, H., Kilic, O.A., Tarim, S.A., Eksioglu, B., 2011. The cost of using stationary inventory policies when demand is non-stationary. Omega 39, 410–415.

Jente Van Belle is Postdoctoral Researcher at KU Leuven.

Ruben Crevits is Data Scientist at OMP and holds a PhD in Business Statistics from KU Leuven.

Tias Guns is Associate Professor in AI at KU Leuven and VUB.

Wouter Verbeke is Associate Professor of Data Science at KU Leuven.