Journal of Operations Management 32 (2014) 337–346

Contents lists available at ScienceDirect

Journal of Operations Management

journal homepage: www.elsevier.com/locate/jom

Valuing lead time

Suzanne de Treville[a,*], Isik Bicer[a], Valérie Chavez-Demoulin[a], Verena Hagspiel[a], Norman Schürhoff[a,b], Christophe Tasserit[a], Stefan Wager[c]

[a] University of Lausanne, Faculty of Business and Economics, 1015 Lausanne, Switzerland
[b] Swiss Finance Institute, 40 Bd du Pont-d'Arve, 1211 Geneva, Switzerland
[c] Stanford University, Department of Statistics, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Received 14 June 2014
Accepted 16 June 2014
Available online 9 July 2014

Keywords:
Option theory
Manufacturing lead time
Supply-chain mismatch cost
Functional products

ABSTRACT

When do short lead times warrant a cost premium? Decision makers generally agree that short lead times enhance competitiveness, but have struggled to quantify their benefits. Blackburn (2012) argued that the marginal value of time is low when demand is predictable and salvage values are high. de Treville et al. (2014) used real-options theory to quantify the relationship between mismatch cost and demand volatility, demonstrating that the marginal value of time increases with demand volatility, and with the volatility of demand volatility. We use the de Treville et al. model to explore the marginal value of time in three industrial supply chains facing relatively low demand volatility, extending the model to incorporate factors such as tender-loss risk, demand clustering in an order-up-to model, and use of a target fill rate that exceeded the newsvendor profit-maximizing order quantity. Each of these factors substantially increases the marginal value of time. In all of the companies under study, managers had underestimated the mismatch costs arising from lead time, so had underinvested in cutting lead times.

1. Introduction

The widespread belief that time-based manufacturing offsets cost advantages from low-cost producers with longer lead times was dashed by the massive wave of offshoring from developed countries that began in the 1980s. Blackburn (2012, p. 397), an initial proponent of time-based competition (Blackburn, 1991), observed that over the past decades supply chains have gotten longer instead of shorter, and the flow of goods through the chains has become slower rather than faster. He asked, "Supply chains pose the following conundrum for time-based competition: if time is so valuable, then why are supply chains so long?" Provocatively, Blackburn demonstrated that the marginal value of time is low for a wide class of product demand structures, and challenged researchers to identify the elements that make short lead times valuable. The demand structure modeled by Blackburn (2012) entailed the following assumptions:

1. The forecast is not expected to evolve over time: the forecast distribution for demand over a given time period in the near future is the same as that at a more distant point in time.
2. The cost of overstocking is limited to the inventory-holding cost, with no obsolescence or perishability.
3. Demand, although potentially highly variable, is predictable.

In the early years of time-based competition it was taken for granted that lead-time reduction always provided value. This belief was challenged by Fisher (1997), who categorized products as functional or innovative depending on the predictability of demand, and proposed that supply chains for functional products should emphasize efficiency over flexibility. The demand characteristics assumed by Blackburn (2012) correspond to functional products, and the low marginal value of time predicted by Blackburn's model is consistent with Fisher's recommendation that paying a cost premium for flexibility is not warranted for such products.

The Fisher framework, in contrast, proposes that innovative products stand to benefit from flexible supply chains: When demand is unpredictable, we can expect a higher marginal value of time. Fisher observed, however, that it is often difficult to determine whether a product is innovative or functional, as many products that appear at first glance to be functional incur high market mediation (mismatch) costs and thus qualify as innovative. He listed

* Corresponding author. Tel.: +41 21 692 3448.
E-mail address: suzanne.detreville@unil.ch (S. de Treville).

http://dx.doi.org/10.1016/j.jom.2014.06.002
0272-6963/

product characteristics that can be helpful in designating a product as functional or innovative (e.g., contribution margin, margin of forecast error, product variety). Tools that allow a manager to quantify demand predictability have not been available, however, so such designation has remained largely qualitative.

de Treville et al. (2014) proposed a model that uses quantitative finance tools to optimize sourcing decisions in the face of evolutionary demand risk. The authors demonstrate that when the forecast evolves over time and demand volatility is high or stochastic, the marginal value of time is high and investment in lead-time reduction is warranted.[1] The de Treville et al. (2014) model provides a useful foundation for quantifying demand unpredictability. We apply this model to products in three industrial settings – which products are difficult to classify as functional or innovative at first glance – to explore how the marginal value of time changes with demand characteristics.

The marginal value of time is captured in the de Treville et al. (2014) model as an indifference frontier between make-to-order and long-lead-time production. This *cost-differential frontier* compares production at a given non-zero lead time to make-to-order production, showing the cost reduction required to compensate for the resulting demand-volatility exposure. Savings that match the cost-differential-frontier value should render decision makers indifferent between the two alternatives in the absence of other forms of supply risk. If a long-lead-time producer offers the product at a cost that is cheaper than the make-to-order cost by a percentage that exceeds the required cost differential, then the cost differential covers the supply–demand mismatch cost for that lead time. If, however, the offered cost differential lies below the cost-differential frontier, the supply–demand mismatch cost is greater than the cost reduction offered by the long-lead-time supplier. The cost-differential frontier is based on the assumption that the order quantity at a given lead time is that which maximizes profit, corresponding to that obtained under the standard newsvendor model.

In each of the three supply chains that we analyzed, managers underestimated the cost of long lead times. They did not consider forecast evolution in supply-chain planning, and underestimated volatility by considering sales rather than demand data. Tender-loss risk and demand clustering that occurred from promotional campaigns and order batching were addressed through forecasting, rather than considered as sources of demand unpredictability. In two of the companies, the order quantity was based on a target fill rate rather than the newsvendor profit-maximizing service level. We extended the de Treville et al. (2014) model to quantify the impact of these increases in demand unpredictability on the marginal value of time, showing that the increase was high enough to warrant lead-time reduction.

Our results do not consider additional supply risk that arises from long lead times due, for example, to lead-time variability, or to quality or logistics problems. We also do not consider supply-chain coordination or pipeline-inventory costs that arise from extending the supply chain. These results should thus be considered as a lower bound on the marginal value of time.

After a review of the literature in Section 2 and of forecast evolution in Section 3, Section 4 describes an application of the model to the Nissan Europe supply chain. In Section 5 we add the possibility that demand would suddenly fall to zero to demand volatility, and apply the model to the GSK Vaccines supply chain. In Section 6 we consider the impact of clustered demand volatility on the marginal value of time, applying the model to the supply chain of a Nestlé Switzerland product. In Section 7 we summarize and conclude.

[1] Lead time refers to the elapsed time between committing production and delivery: a firm that holds inventory to offer a short *delivery* lead time will be subject to supply–demand mismatch costs from holding that inventory.

2. Literature review

That short lead times can be a source of competitiveness is well established in the literature. Suri (1998) provides an in-depth review of how to reduce manufacturing lead times (see also Hopp and Spearman, 1996), and how to use short lead times to gain competitive advantage. Fisher et al. (1997), however, observed that companies struggle to reduce their lead times, and to quantify the impact of lead-time reduction on profit.

For short-life-cycle products, short lead times represent one element of the Quick Response approach that aims to reduce demand-risk exposure (e.g., Iyer and Bergen, 1997). Quick Response calls for partial lead-time reduction in combination with information management to bring supply closer to demand (Abernathy et al., 1999). Fisher and Raman (1996) describe a representative implementation of Quick Response at an apparel manufacturer that combined estimation of the coefficient of variation of demand from the distribution of forecast estimates, observation of early sales, and access to short-lead-time "reactive" capacity. Quick Response achieves much of its reduction in supply–demand mismatch costs with quite limited lead-time reduction: much sourcing is done from long-lead-time suppliers, and reactive suppliers have lead times that, while reduced by about half of that of the "speculative" capacity, remain too long to permit production to order (Iyer and Bergen, 1997). Many of the benefits of lead-time reduction can be achieved through techniques such as postponement (Lee and Billington, 1995) without actually shortening the supply chain. Thus, even for fashion (innovative) products, the result from Blackburn (2012) that lead-time reduction cannot be assumed to always provide value is upheld. Allon and Van Mieghem (2010) considered dual sourcing between a low-cost, long-lead-time supplier and a higher-cost, reactive supplier. They showed that allocating a small portion of demand to the reactive supplier sufficed to minimize total cost, although the allocation to the reactive supplier increased when positive demand autocorrelation increased the variance of demand.

Lead-time reductions allow the order decision to be made based on an updated demand forecast. Thus, the forecast evolution process affects the marginal value of time. Milner and Kouvelis (2005) consider a product whose demand evolves over time following the martingale model of forecast evolution (Graves et al., 1986; Heath and Jackson, 1994), showing the reduction in mismatch cost obtained from an ability to place a second order closer to the delivery date. Gallego and Özer (2001) present a model that quantifies the value of receiving demand information further in advance of the delivery date, showing that the performance of the system improves as order information is received earlier. Thus, the value of lead-time reduction decreases when firms have other alternatives to obtain demand information. Wang and Tomlin (2009) captured the impact of forecast updating on lead-time policy, assuming a multiplicative Markovian forecast-update process (Hausman, 1969). These authors consider lead-time stochasticity as a type of supply risk, showing that as lead-time reliability decreases, firms facing demand volatility either order earlier (increasing the full lead-time period) or pay a premium to increase lead-time reliability.

As mentioned in Section 1, real-options theoretic models can be used to calculate the impact of demand volatility on the cost-differential frontier. de Treville et al. (2014) showed that when the demand forecast evolves at a constant rate, and the order quantity used is that which maximizes profit (corresponding to the newsvendor critical fractile), the cost-differential frontier increases in demand volatility at a decreasing rate, with incremental lead time reduction of little value. This functional form explains some of the offshoring that has occurred over the past couple of decades. A local producer that can produce based on accurate demand information because of short lead times may be well positioned to

compete against an offshore supplier with long lead times. This may not be the case, however, for a local producer with lead times that are long enough to cause demand-volatility exposure. Suppose, for example, that a local supplier has a relative lead time of 0.5. Under constant volatility, the cost differential required to compensate for the offshore supplier's lead time may well not be much greater than that for the local supplier. Being located close to the market facilitates but does not guarantee short lead times.

The assumption of constant demand volatility does not hold when the forecast evolves with clustered information arrival, such as occurs when promotional campaigns shift demand forward or backward in time. de Treville et al. (2014) used the cost-differential frontier to illustrate two key managerial insights under stochastic demand volatility. First, the frontier is increasing in the volatility of volatility. Second, stochastic volatility increases the marginal value of incremental lead time. Whereas under constant volatility increasing relative lead time from 0.7 to 1 might require a low cost differential, under stochastic volatility a similar increase in lead time might require a much higher cost differential.

3. Modeling forecast evolution

The forecast-evolution process has a major impact on how lead time affects supply–demand mismatch costs. Suppose that a retailer can place an order at any time between $t = 0$ (the long-lead-time option) and $t = T$ (make to order), and that at each time $t \in [0, T]$ the retailer has access to a forecast F_t of the eventual demand D. Following the martingale model of forecast evolution (Hausman, 1969; Graves et al., 1986; Heath and Jackson, 1994), we assume that the forecast process is rational in the following sense:

1. The forecast will eventually converge to the correct answer, such that $F_T = D$. This is equivalent to assuming that a make-to-order producer produces the right amount of goods.
2. The forecast updates are unbiased, such that for any subsequent times $0 \le t \le t' \le T$, the expected value of the forecast update is zero: $\mathbb{E}[F_{t'} - F_t | F_t] = 0$.

These assumptions do not specify the forecast process F_t; rather they only place weak constraints on what constitutes an acceptable forecast model. In particular, this framework can support a wide variety of forecast processes with, for example, jumps or stochastic volatility.

In the following sections, we examine three industrial cases, and show how different models for F_t are appropriate for each of them. Choosing a good model for F_t is crucial, as it will substantially impact the shape of the cost-differential frontier. Following de Treville et al. (2014), we first consider a baseline case where we assume that the forecast process F_t follows a geometric Brownian motion with a constant instantaneous volatility σ. We parametrize our analysis in terms of the relative lead time $RLT = t/T$.

To give a better idea of what a constant-volatility forecast process looks like, we show some sample paths of the geometric Brownian motion in Fig. 1. The relative lead time between forecast and delivery increases along the horizontal axis from left to right, with the forecast being drawn from a lognormal marginal density with a marginal variance that increases linearly in lead time. The volatility parameter thus increases with the square root of time, yielding the "square-root-of-time" rule well known in both finance and inventory theory. For a given volatility parameter, the demand density becomes wider as relative lead time increases. Densities for relative lead times of 0.05 (solid line), 0.1 (dashed line), and 1 (line with dashes and dots) are shown in Fig. 2. The coefficient of variation ρ of the demand density for the full lead time determines

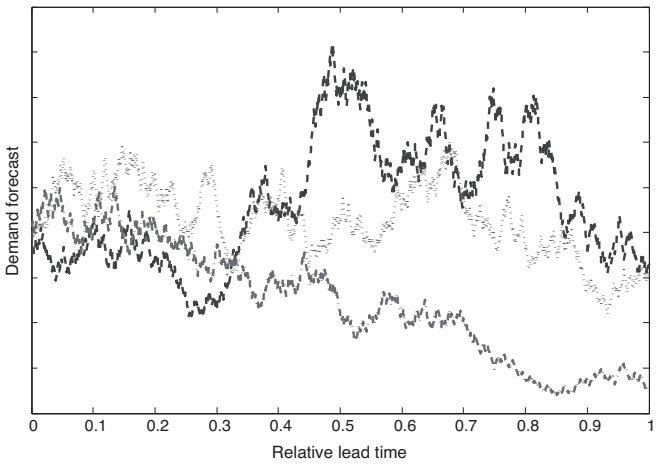

Fig. 1. Three possible forecast-evolution paths under geometric Brownian motion for a given demand process. The relative lead time refers to the time remaining when the forecast is made. A relative lead time of 0 refers to a forecast made on the delivery date, so all three paths depart from an actual-to-forecast ratio of 1 at the left of the graph.

the constant instantaneous volatility parameter σ via the following formula $\sigma = \sqrt{\log(\rho^2 + 1)}$.

The Black–Scholes model – considered as the workhorse of financial engineering – assumes that the behavior of securities prices is described by a constant-volatility process. Even though true volatility is often stochastic and securities prices may experience jumps, the Black–Scholes model is generally considered to give a reasonable first idea of option value (e.g., Hull, 1996). The same holds with demand forecast-evolution processes. The cost-differential frontier rises with the volatility of demand volatility (de Treville et al., 2014), and with jumps (Bicer et al., 2013), thus the constant-volatility assumption provides a lower bound on the marginal value of time. The more complicated models considered in Sections 5 and 6 are refinements of this constant-volatility baseline: From the baseline we quantify the increase in the marginal value of time from tender-loss risk and the stochastic volatility that arises from demand-information clustering.

Inventory-theory models typically specify a demand density – normal, uniform, or empirical, for example – rather than a demand forecast-evolution process. When we make the forecast-evolution process our focus, the appropriate marginal density for a given lead time emerges. In the absence of information to the contrary, it is reasonable to begin by assuming constant volatility, hence

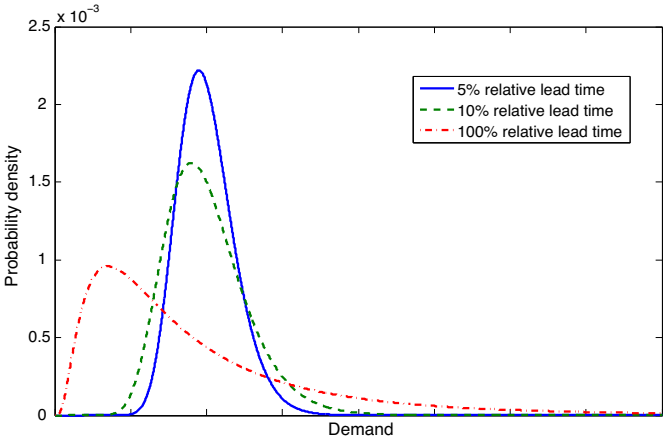

Fig. 2. As the relative lead time increases, the associated marginal density becomes wider, increasing the expected mismatch cost for a given target quantile.

lognormal demand with a known marginal density. Let us note in passing that the lognormal density eliminates the problem of negative demand incurred with the normal distribution for higher coefficients of variation. If the forecast evolves based on a stochastic instantaneous volatility, the marginal density for a given lead time remains lognormal.

When the ratio between actual and median demand follows a lognormal density, managers are often able to estimate distribution parameters using their intuition, specifying (1) a demand peak that they would expect to encounter as a multiple of median demand, (2) how often such a demand peak would be expected to occur, and (3) whether the forecast lies at the median of the actual-to-forecast ratio. Suppose that managers' intuition is that demand can be expected to double median weekly demand 2 weeks per 50-week year. An actual-to-median ratio of 2 is thus estimated to lie at the 96th percentile, or $\Phi^{-1}(1-2/50)=1.75$ multiplicative standard deviations above the median (i.e., the geometric mean.[2] Then $2=e^{z\sigma}=e^{\Phi^{-1}(1-2/50)\sigma}$, and $\sigma=\log(2)/\Phi^{-1}(1-2/50)\approx 0.4$. When empirical demand data does not fit a lognormal density, it is likely that the forecast evolution includes jumps. The empirical density can be decomposed into a lognormal component and adjustments to its skewness and kurtosis. Bicer et al. (2013) used an Edgeworth-series expansion to calculate the change in mismatch cost that results from these new values of skewness and kurtosis, noting as well that the changes to skewness and kurtosis can be used to get an idea of the magnitude and direction of jumps. To summarize, the forecast-evolution process is described by an instantaneous volatility that is either constant or stochastic, as well as by the intensity of any jumps that might occur. Jumps result in a marginal density that differs from the lognormal. Our experience has been that managers are better able to share their intuition about the forecast-evolution process than about the parameters of an empirical marginal density for a given lead time (corresponding to the results by Schweitzer and Cachon, 2000), and our approach estimates the value of lead time directly from this process. If, however, the available demand data takes the form of an empirical marginal density for a given lead time, one can use that empirical density to get an understanding of the forecast-evolution process by first determining whether it corresponds to a lognormal, and if not, what kinds of jumps are suggested by the differences in skewness and kurtosis relative to the lognormal.

4. Constant volatility: Nissan Europe

Nissan Europe has historically frozen production schedules 8 weeks in advance in order to permit production rationalization, estimated to reduce the assembly cost between 1% and 2%. This baseline Black–Scholes form of the cost-differential frontier, discussed above, allows us to estimate the cost of demand-volatility exposure as lead time increases from 1 week (allowing Nissan Europe to assemble cars to order) to 8 weeks.

Although lean production emphasizes producing what the customer wants, a common practice in the Toyota Production System – carried over to lean production – is to aggregate orders over 6–8 weeks and then schedule production and material orders as evenly as possible over that period (Womack et al., 1990). Liker (2004) quotes Fujio Cho, who was then president of Toyota Motor Corporation, arguing that leveling the production schedule plays a fundamental role in effective lean production. Liker argues that the cost of the customer waiting a few weeks to receive an order is more than made up by the benefit of eliminating variability from the production schedule. This level-scheduling approach is currently followed by Nissan Europe plants.

The cost-differential frontier allows Nissan Europe to explore this trade-off in finer detail. A simplified version of the trade-off is as follows: a car's profit depends on whether it is sold immediately after production. Holding costs combine with "incentives" (concessions made to sell cars when supply is higher than demand) to eliminate the profit if not sold immediately. The production schedule is fixed 8 weeks in advance, with the resulting cost savings estimated by Nissan Europe to be between 1% and 2%. In weeks where demand exceeds the production quantity, the sales opportunity is lost because the customer chooses an alternative brand in the same price class. The trade-off that we explored was how the mismatch cost from waiting to finalize the production schedule until demand was observed compares to the efficiency gain from freezing the production schedule.

We normalized the 8-week lead time to 1. Because of the schedule freeze, production is currently committed at time $t=0$ for delivery at $t=1$. Not freezing the production schedule implies a production commitment at $t=1$ for immediate delivery. It is also possible to weigh the advantages and disadvantages of reducing but not eliminating the lead time: A production commitment made at $t=0.5$ drops the lead time to $1-0.5=0.5=4$ weeks.

Following discussions with Nissan managers, we normalized the price for the higher-margin vehicle to 100 and the cost if produced after observing actual demand (thus not freezing the production schedule) to 50. If the car is not sold immediately, its value is reduced by a combination of the inventory holding cost and the expected incentive value required to move cars that do not immediately sell. This was estimated by management to amount to around 20% of production cost, leaving a residual value normalized to 40. The resulting profit-maximizing critical fractile is 0.83 for the short lead time. Our initial assumption is that the order quantity is that which maximizes profit for all lead times, with the critical fractile increasing to reflect the lower cost incurred at longer lead times. A cost differential of 8%, for example, increases the critical fractile from 0.83 to 0.9.

Time-series data for demand was not available. Historical sales data gave a truncated picture of demand, as it eliminated stockouts and did not indicate what part of sales occurred after incentives. We considered demand peaks that were 1.2, 1.6, and 2 times median demand 1 week out of 100, so at the 99th percentile. This implied constant instantaneous volatility parameters of $\sigma=0.07$, 0.2, and 0.3, respectively. The marginal density of the actual-to-median ratio for a lead time $RLT=1-t$ is logNormal(0, $\sigma^2\times RLT$). Results from computing the cost-differential frontier with these choices for σ are shown in Fig. 3. Manual calculation of the cost differential for a given lead time and volatility is described in Appendix A. At 7% volatility, the 1–2% savings from freezing the production schedule approximately compensate for demand-volatility exposure, with supply–demand mismatch costs that are about the same as the benefits of rationalization. As demand volatility increases to 20% (30%), the required cost reduction increases to over 5% (7%), indicating that the mismatch costs are likely to exceed the benefits of freezing the production schedule by a considerable margin. Prior to this analysis freezing the production schedule was not questioned. Managers' intuition was that demand volatility was more likely to be in the 20–30% range than at the 7% maximum in order for the advantages of the schedule freeze to outweigh the resulting mismatch cost. Thus these cost-differential frontier results – even though based on managerial intuition about the volatility level rather than historical demand data – put the topic on the table for discussion.

Extension: reduction in residual value. The Nissan Europe analysis presented above is based on a relatively high residual value. If market conditions reduce this residual value, the marginal value of time increases. In Fig. 4, we begin with the curve from Fig. 3 where

[2] For a full description of use of the multiplicative standard deviation, see Limpert et al. (2001).

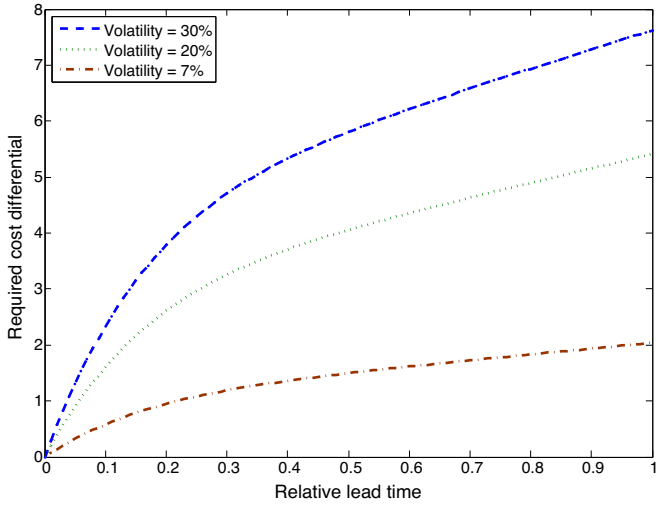

Fig. 3. The cost-differential frontier is increasing in volatility: At a constant instantaneous volatility of 7%, the cost of demand-volatility exposure for the full lead time is approximately the same as the 1–2% benefit from freezing the production schedule. As the volatility increases to 30%, the cost of demand-volatility exposure exceeds the expected benefit.

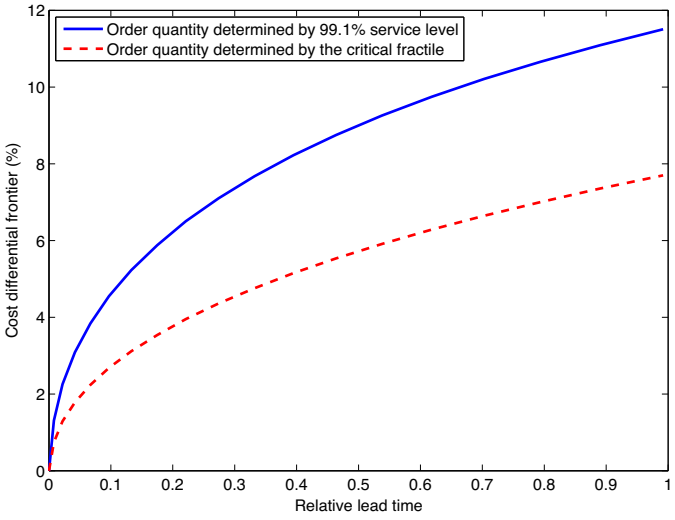

Fig. 5. The cost-differential frontier increases when the service level exceeds that which maximizes profit: The lower curve shows a 30% constant volatility and profit-maximizing service level for price = 100, make-to-order cost = 50, and salvage value = 40. The top curve shows the required cost differential when a 99.4% fill rate is required, which corresponds to a 99.1% service level.

the demand volatility is low enough that the gains from freezing the production schedule compensate for the increase in demand-volatility exposure (constant volatility = 7%) at a residual value of 40 (80% of production cost), then reduce the residual value to 0 and 20 (0–40% of production cost). As the residual value approaches 0, the required cost differential increases to 5.5%.

Now suppose that the average residual value is 80% of production cost, calculated by averaging a high revenue for the first few units with much lower revenue for lower units. In follow-on research, Wager and de Treville (2013) showed that the cost-differential frontier increases not only in the average salvage loss, but also in its stochasticity. Assuming a constant salvage-value heuristic will lead a firm to systematically undervalue the marginal value of time even if the constant value assumed is equal to the expected salvage value.

Extension: increase in target fill rate. Until now, we have assumed that the order quantity maximized profit. The algorithm underlying the cost-differential frontier begins with the make-to-order cost and the resulting newsvendor critical fractile. As lead time increases, the required cost differential increases, which then increases the critical fractile for given price and salvage value. In Figs. 3 and 4, we used this evolving critical fractile to determine the order quantity.

As we apply the cost-differential frontier in practice, however, we are often informed that the company bases its order quantity on a target fill rate that tends to be higher than what would maximize profit. For example, suppose that Nissan were to target a 99.4% fill rate, which corresponds to a 99.1% service level for $\sigma = 0.3$. Referring back to Fig. 3, we observe that the long-lead-time cost differential with $\sigma = 0.3$ is 7%. The profit-maximizing service level for a 7% cost differential is $100 - 50(1 - 0.07)/100 - 40 = 0.89$, for a fill rate of 93.5%, lower than the 99.4% target.

In Fig. 5 we show how the Nissan Europe cost-differential frontier at 30% volatility increases when we increase the target fill rate to 99.4% (i.e., a service level of 99.1% at 30% volatility). We hold constant price, make-to-order cost, and residual value. As expected, the cost differential is minimized when the order quantity corresponds to the newsvendor critical fractile: Deviations from the critical fractile increase the marginal value of time.

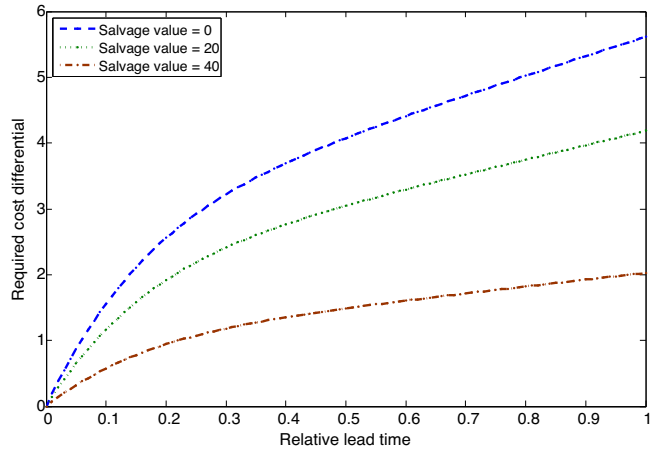

Fig. 4. The cost-differential frontier is decreasing in salvage value: demand-volatility exposure cost may be low when the salvage or residual value covers most of the production cost, but increases as incentives and holding cost reduce the salvage value. The above curves assume a 7% volatility.

5. Risk of losing a tender: GSK vaccines

Heath and Jackson's martingale model of forecast evolution follows Black–Scholes option pricing in assuming that the increase in volatility added from a one-instant increase in lead time – the instantaneous volatility – is constant. As we described in Section 3, under a constant instantaneous volatility the marginal demand density for any lead time follows a lognormal distribution whose variance parameter increases linearly with lead time. The assumption that information arrives in a steady flow, formalized in the martingale model of forecast evolution as a constant instantaneous volatility, is, however, often violated in practice.

Let us now consider what happens to the marginal value of time when the possibility exists that demand would drop to zero. GSK Vaccines faces such a risk due to the tender structure that applies too much of the company's business. The GSK Vaccines supply chain is extended across several countries and runs at a high utilization, resulting in a 10-month average lead time. The company bids on an order, and learns as late as 2 months before the delivery date whether the tender was won or lost. If the tender is lost, the

demand forecast drops to zero and remains there. GSK thus faces two sources of demand uncertainty: volatility and tender risk.

Senior management estimates the long-term average probability of winning a tender to be around 50%. Because the company has a long lead time that forces it to begin production well before knowing whether the tender was won, then it will end up discarding all advance production 50% of the time. Ten months before delivery, management begins production with no information except this average probability. If management could reduce the lead time to 9 months, they would either have learned that they had already lost the tender and could use production capacity otherwise, or would begin production with a slightly higher estimated probability of winning. If lead time is reduced to 3 months before delivery, management's estimate of the probability of winning given that the tender has not yet been lost begins to approach 1. Shortening lead times to allow the company to delay beginning production until the estimated probability of winning the tender has risen to, say, 75% avoids much unnecessary production.[3]

We structure the problem using the distribution of the actual-to-median ratio of demand independent of whether the tender is won. This makes it easy to see the difference between demand volatility and tender-loss risk. Price is normalized to 100, the make-to-order cost is 70, and salvage value is 0. Vaccines produced for one market cannot be sold elsewhere because of testing and packaging restrictions, and shelf life is short enough that vaccines cannot be held over until a future tender. We assume a constant instantaneous volatility of 20%.

Adding the risk of losing a tender to normal volatility exposure is analogous to asset-default risk, addressed in the finance literature through the "jump-to-default" model (described in Gatheral, 2006). The risk of losing the tender is captured via a Markov jump term that reduces demand to zero if a jump occurs. With this model, the probability of winning a tender decays as $e^{-\lambda \times RLT}$ for some "default intensity" $\lambda \geq 0$ that describes the probability of losing the tender in any given instant, and for a relative lead time $RLT = 1 - t$.

The current lead time is around 10 months. The company informed us that they definitely know whether or not they have won the tender by 2 months before delivery at the latest (relative lead time $RLT = 0.2$). Thus, we set the default intensity to 0 for $RLT \leq 0.2$. Writing λ for the default intensity for lead times longer than 2 months, the probability of losing the tender at $RLT > 0.2$ conditional on not yet having lost it is $1 - e^{-\lambda \times (RLT - 0.2)}$. Given management's estimate that 10 months before delivery the probability of winning a given tender is around 50%, we first consider a default intensity of $\lambda = 0.8$. At a lead time of 5 months (relative lead time of 0.5), the probability $e^{-0.8 \times (0.5 - 0.2)}$ of winning the tender increases to 79%. In Fig. 6 we vary λ from 0 (no tender-loss risk) to 0.8. With $\lambda = 0.8$, beginning production 10 months in advance (a relative lead time of 1 in the figure) requires a cost differential of over 70%. More generally, we observe that even modest tender-loss risk dramatically increases the required cost differential.

These results have led GSK Vaccines to begin a company-wide investment in lead-time reduction. Management had previously assumed lead-time reduction to be expensive, and the marginal value of time to be relatively low and avoidable through forecasting. Also, management faced temptation to consider incremental lead-time reduction. Incorporating forecast evolution into decision

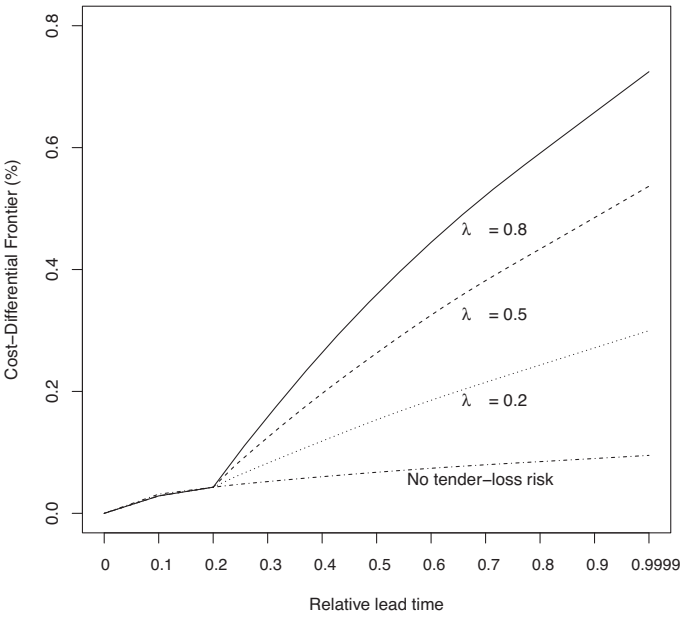

Fig. 6. The required cost differential increases in tender-loss risk. The probability of losing the tender given that we have not lost it so far is $1 - e^{-\lambda \times (RLT - 0.2)}$ for $RLT \geq 0.2$. We vary λ from 0 (no tender-loss risk) to 0.8. The tender-loss risk is eliminated for relative lead times below 0.2. Volatility is 20%.

making has made the marginal value of time obvious, and has clearly demonstrated that lead-time reduction activities must target reducing lead time enough to eliminate tender-loss risk.

This led to another important observation. As soon as decision makers began to explore why lead times were long, they immediately found ways to reduce them at a cost that is justified many times over by avoiding demand-volatility exposure. Much of the lead time turns out to be caused by a small number of bottlenecks. The company is now in the process of adding capacity to these process steps, which is expected to reduce lead times dramatically. Long lead times have also resulted from the supply-chain tactic of moving production between factories in search of the highest capacity utilization. It is recognized that the cost savings from such high utilization are completely dominated by the cost of supply–demand mismatch risk. Lead times can be radically reduced by avoiding such supply-chain extensions.

6. Stochastic volatility: Nestlé Switzerland

Managers from the Demand and Supply Planning Department at Nestlé Switzerland observed a conundrum. A long-shelf-life product believed to have low underlying demand volatility – the archetypal functional product (Fisher, 1997) – had exceptionally high salvage losses. Around half of what was produced had to be salvaged. The long shelf life and low demand volatility at the end-consumer level would normally protect against salvage losses. In this case, however, product demand was characterized by clustered information arrivals arising from promotional campaigns and order batching. The marginal demand density at a given lead time thus depended not only on the time until the delivery date, but also on whether demand was being shifted forward or backward in response to promotions and batching. As we will show, this volatility clustering increased the demand-volatility exposure for the product to a level that – in combination with a high required fill rate – led to high enough product inventory that even the long shelf life did not protect against salvage losses.

One way to account for such volatility clustering is to use a stochastic volatility model. Heston (1993) proposed that the impact

[3] In both cases, GSK wins the same number of tenders on average. In the first case, the company begins production before all possible tenders, and then wins half of the competitions it participates in. With the shorter lead time example, it would only start manufacturing in anticipation of 2/3 of tender competitions but win 75% of those in which it participates. The key idea is that because the company starts producing later, it can avoid wasting effort on tenders that have been awarded to someone else.

Table 1
Parameter values under constant volatility.

	Make to order	Full lead time
Price	100	100
Cost	44	39.6
Residual value	N/A	36.5
Service level	100%	96%
Fill rate at 20% constant volatility	100%	99.4%
Fill rate at 53% constant volatility	100%	98.7%

of stochastic volatility on the price of an option could be modeled through the volatility of volatility and the rate at which instantaneous volatility reverts to the long-run average volatility. de Treville et al. (2014) showed how to incorporate Heston's (1993) stochastic volatility model into the cost-differential frontier.

The company had 113 weeks of demand data available. The average volatility of the weekly log returns was 151%. We used a thought experiment to get a rough estimate of what the volatility of demand for the product would be without amplification. If the very high and very low values of weekly demand result from demand that is moved forward or backward as a result of promotional campaigns, the middle area should show demand under unamplified volatility. We sorted the data and identified a middle range around a median of 20 cases where weekly demand varies from 11 to 36 cases, representing the 18th to the 62nd quantile. The ratio of the upper value to the median (or the median to the lower value) is 1.8. If the upper and lower values represent three standard deviations from the median, then $\sigma = \log(1.8/3) = 0.2$. In what follows, we thus use a constant volatility of 20% to represent the no-amplification case.

For this product Nestlé used an order-up-to model with a weekly review period and a delivery lead time that averaged 3 weeks. Each week Nestlé compared the inventory that was in stock, in the pipeline, and back ordered to the demand that was expected to occur over the next 4 weeks; placing an order that would bring total inventory to the target service level. Under these conditions the service level that maximizes profit remains the newsvendor critical fractile, with the cost of overage a weighted average of inventory holding and spoilage costs. The average volatility of log returns over the 110 4-week demand periods was 53%, quite high for a functional product whose demand volatility would normally be expected to be relatively low.

The price was normalized to 100, and the make-to-order production cost to 44. We assumed a long-lead-time cost of 39.6, yielding a 10% long-lead-time cost differential. Because the product is functional, product not sold during the demand period can be carried over to the following period as long as it does not hit its expiration date. In this case, the newsvendor model is again carried out based on the residual value rather than the salvage value, with the inventory value carried forward decreased by the inventory holding cost and the increased risk of obsolescence.

We can enrich the example by focusing on the case in which Nestlé's usual target fill rate of 99.4% corresponds to the service level that maximizes profit in the absence of volatility amplification. A target fill rate of 99.4% corresponds to a service level of 96% at a constant volatility of 20%. We can estimate the residual value to be 36.5 under these conditions (96 % = 100 − 39.6/100 − 36.5). It is interesting to note in passing that an increase in constant volatility from 20% to 53% reduces the fill rate corresponding to this 96% service level to 98.7%: a first casualty of increased volatility even before adding in stochasticity. For the make-to-order case, the service level and fill rate are 100%. Table 1 summarizes the above parameter values.

Stochastic volatility in the Heston model is operationalized via four parameters: (1) the long-run average volatility, (2) the volatility of volatility, (3) how quickly volatility reverts to the long-run average after a shock, and (4) the correlation between increments of demand and increments of volatility (whether increases in demand are expected to be accompanied by increases or decreases in volatility).

We estimated the volatility of volatility and the mean-reversion rate from the time-series demand data using the square-root GARCH model proposed by Heston and Nandi (2000) that converges to the Heston (1993) closed-form model. We used a percentile parametric bootstrap method to derive confidence intervals for the volatility of volatility and the mean-reversion rate. The derivation of parameter values is described in Appendix A. The volatility of volatility was estimated (based on 199 bootstrap replications) to be 1.41 (141%) with 95% confidence interval [1.35; 1.47]. The mean-reversion rate was estimated to be 0.74 with 95% confidence interval [0.53; 0.99]. The volatility of the log returns unconditioned on time was 1.51, rising to 2.00 when conditioned on time with 95% confidence interval [0.055; 2.23].

Much of the above volatility was eliminated because demand was pooled over 4 weeks through its order-up-to process. As mentioned above, for weeks 4 through 113 we aggregated data from that week and the previous 3 weeks to capture the volatility of 4-week demand faced by the company when placing a given week's order.

A moving sum is not a memoryless process, so the Heston and Nandi (2000) model does not provide an unbiased estimate of the volatility of volatility for aggregated data. To resolve this issue, we cleaned the residuals from autocorrelation (the process is described in Appendix A), then applied the Heston and Nandi (2000) model to the residuals to obtain the volatility of volatility and mean-reversion rate for the aggregated demand. The volatility of volatility was 1.16 with 95% confidence interval [1.11; 1.20]. The mean-reversion rate was estimated to be 0.95 with 95% confidence interval [0.45 ; 0.99]. The long-run volatility conditioned on time was 1.34, with 95% confidence interval [0.043 ; 1.42]. In computing the cost-differential frontier we used the more conservative 0.53 average volatility estimated from the log returns (not conditioned on time) rather than the higher Heston–Nandi estimate of 1.34. The unconditioned estimate lay in the confidence interval and was more conservative, so that the cost-differential frontier as calculated can be understood as a lower bound.

The cost-differential frontier is shown in Fig. 7. Under constant and unamplified volatility, the marginal value of the 4-week horizon is less than 5%. Promotional campaigns and other sources of volatility amplification raise the average volatility to 53%, which increases the required cost differential to around 9%. Let us now incorporate the fact that volatility is stochastic. The long-run volatility remains 53%, but information arrivals are clustered. The stochastic volatility results are shown in the top curve in Fig. 7. There are two important managerial insights from this curve. The required cost differential approaches 30% for full lead time, and would increase to 40% were we to use the much higher long-run volatility estimated from the data by the Heston–Nandi model. Second, the constant and stochastic volatility curves corresponding to 53% volatility start together at a relative lead time of 0, but the stochastic volatility frontier increases at a faster rate.

Nestlé has been committed to making production both lean and responsive, balancing the objective of flexibility against the reality of the need to fully deploy capital investments. Our results demonstrate how demand volatility amplification increases the challenge of achieving such a balance. When the demand volatility of a functional product is not amplified it is possible to be reasonably flexible while remaining lean. As volatility amplification increases mismatch costs – especially in combination with ambitious fill-rate goals – there is an increasing trade-off between leanness and flexibility.

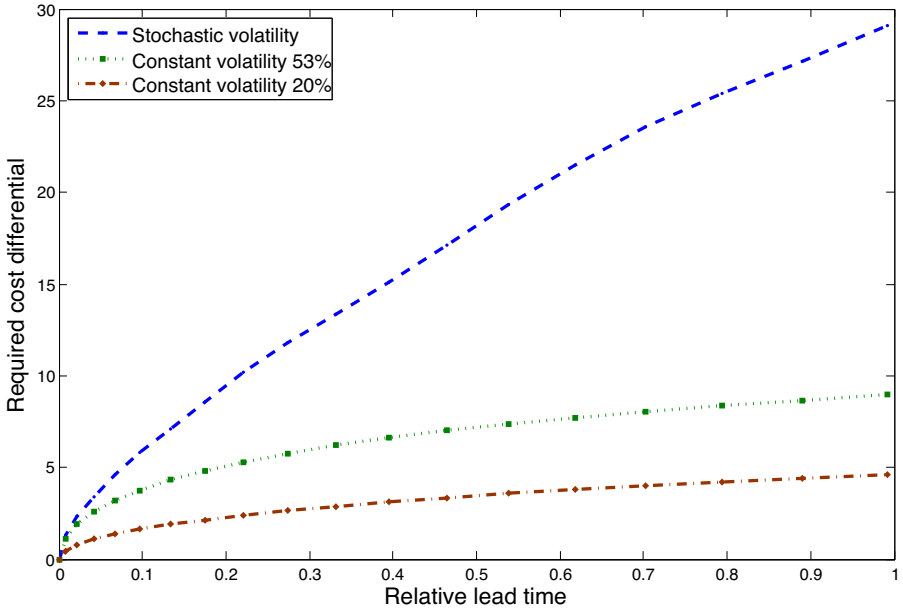

Fig. 7. The bottom curve shows the cost of demand-volatility exposure under 20% constant volatility, typical of a functional product without volatility amplication. The middle curve from the bottom assumes constant volatility of 53%. The top curve assumes an average long-run volatility of 53%, a volatility of volatility of 1.16, and a mean-reversion rate of 0.95. The assumed service level is 96%, corresponding to the 99.4% Nestlé target fill rate at a 20% constant volatility.

7. Summary and conclusions

Blackburn (2012) argued for the importance of incorporating the marginal value of time in supply-chain decision making. The model proposed by Blackburn (2012) estimated the marginal value of time under predictable demand with a forecast distribution that did not evolve in time. We capture the marginal value of time by the cost differential required to compensate for lead time under various forecast-evolution regimes. The projects described here have allowed us to work through the implementation of the theoretical results proposed by de Treville et al. (2014), Bicer et al. (2013), and Wager and de Treville (2013). Application of the cost-differential frontier to three very different supply chains that cover the gamut of demand forecast evolution types has allowed us to respond to Blackburn's call for better understanding of the marginal value of time for products that are neither purely functional nor purely innovative.

We first described the cost-differential frontier to estimate the marginal value of time for a Nissan Europe production line. Lacking historical demand data we were able to obtain a good enough estimate of demand volatility from managers' intuition to establish that the marginal value of time was higher than the benefits of freezing the production schedule 8 weeks in advance. Working on this project brought home to us the benefits of working with demand forecast evolution rather than trying to estimate the marginal demand density for a given lead time. Our results from Nissan Europe provided an intriguing new perspective on standard lean production theory concerning demand smoothing through techniques such as heijunka.

The Blackburn model estimated the marginal value of time in the absence of perishability or obsolescence. We used the cost-differential frontier to gain insight into how increases in per-unit overstock costs (salvage or residual value decreases) affect the marginal value of time. Inventory theory is built around the concept of salvage value, where a product that does not sell during the demand period is sold below cost. In the companies that we studied, overstock costs were more likely to appear as reductions in residual value, with the value of the item held in inventory decreasing according to the inventory holding cost and increased risk of obsolescence. As expected, reductions in salvage or resideual value increased the required cost differential. This exploration aided in fleshing out the middle ground described by Fisher (1997) concerning products that appear to be functional yet generate high mismatch costs. As lead times increase, higher volatility exposure may increase left-over inventory enough to dramatically reduce residual value, causing a previously functional product to experience the mismatches expected from an innovative product.

Bicer et al. (2013) extended the cost-differential frontier to cover jumps. We applied this extension to GSK's order structure based on tenders to show that a firm required to commit production before knowing whether the tender was won because of long production lead times may face a high marginal value of time. This tender structure required that we consider demand forecast evolution along two dimensions: volatility and jumps. Again we observed that we could transform managers' intuition about demand forecast evolution into parameter values that gave a good enough picture of the marginal value of time to convince managers to give serious thought to cutting lead times.

The cost-differential frontier assumes a profit-maximizing service level. As we worked with companies, we were informed that the service level prescribed by the newsvendor model was insufficient, and that the companies set a higher fill rate. We used the properties of the lognormal distribution to calculate the service level corresponding to a target fill rate at a given volatility. This allowed us to estimate the increase in the marginal value of time from setting a higher service level than that specified by a newsvendor analysis.

The cost-differential frontier begins with an ideal world where the order quantity maximizes profit, and the only source of supply-chain risk is demand volatility. This means, for example, that lead times are constant and known, there is no supply risk, loss of innovation, or loss of intellectual property. In each of the projects, the required cost differential exceeded the cost of reducing lead time. We were informed of many other sources of supply risk. The cost-differential frontier thus has served as a lower bound for the marginal value of time, with other supply risks making the marginal value of time even higher.

Acknowledgements

The authors would like to acknowledge financial support for this research from Nissan Europe and GSK Vaccines. Many people at the three companies described in this research provided invaluable support, including Szymon Walus, Maria Vaccaro, Bryan Barr, and Christopher Benardis at Nissan Europe; Mauro Bernuzzi and Frederic Mahieu at GSK Vaccines; and Daniel Costa and François Facchin at Nestlé Switzerland. We also would like to thank Joe Blackburn and co-Editor-in-Chief Dan Guide for very helpful insights and comments on earlier versions of this work.

Appendix A. Manual calculation of the cost differential under constant volatility

In this section, we demonstrate how to calculate the cost differential for the Nissan constant-volatility example. To make the results as universally applicable as possible, we will work with the actual-to-forecast ratio distribution, which is assumed to follow a lognormal distribution with parameters μ and $\sigma^2(T-t)$ for a production commitment made at time t for a delivery date T. For simplicity, let us consider the case where the forecast corresponds to the median of the demand distribution, so that $\mu = 0$. In other words, we normalize demand by its median. The volatility for the full lead time $T-t = 1$ is σ. We here reproduce the value for $\sigma = 0.3$ shown in Fig. 3. Price is normalized to 100 and the residual value is 40% of the price. The short-lead-time cost is 50% of the price. The cost differential required to compensate for the volatility is derived in the following steps:

1. Estimate the expected profit for the make-to-order case: The expected value of the actual-to-forecast ratio will be $e^{\mu+\sigma^2/2} = e^{0.09/2} = 1.05$. Each unit will make a profit of $100 - 50 = 50$, so the expected profit per unit forecast will be $50 * 1.05 = 52$.
2. Unless a fill rate is targeted that exceeds the profit-maximizing critical fractile the order quantity is set using the newsvendor critical fractile $100 - c_L/100 - 40$. Let's begin by setting $c_L = 49$ to correspond to the case where freezing the production schedule by 8 weeks yields a 2% cost reduction as estimated by senior management. This yields a critical fractile of 0.85, which is 1.04 geometric standard deviations above the median, yielding an order quantity of $e^{\mu+\sigma z} = e^{0.3\times 1.04} = 1.36$ times the forecast.
3. The fill rate for the lognormal distribution for z standard deviations is calculated using the formula fill rate $= \Phi(z-\sigma) + e^{z\sigma-\sigma^2/2}(1-\Phi(z)) = 96\%$. (The derivation of the fill rate for the lognormal distribution is given in a later section of the Appendix A.)
4. We apply the fill rate to expected demand as calculated in the first step to determine expected sales of 1.01 times the forecast.
5. The expected left-over inventory is the difference between the order quantity and the expected sales, so $1.36 - 1.01 = 0.36$ times the forecast.
6. The expected profit per unit forecast is $100 - 49 = 51$ for each unit of expected sales, less a $49 - 40 = 9$ loss for each unit of leftover inventory $= 1.01 \times 51 - 0.36 \times 9 = 48$ per unit forecast. We observe that this is lower than the 52 per unit forecast for the make-to-order case, so that at a 30% volatility and full lead time a 2% cost differential does not suffice to cover the demand-volatility exposure.
7. By binary search we obtain the long-lead-time cost that yields the same per unit of forecast profit. This value is 46.5, representing a 7% cost differential relative to the make-to-order cost.
8. For lead times less than 1, the analysis is repeated, adjusting volatility from σ to $\sigma\sqrt{T-t}$.

Table A.2
Calculation of the required cost differential for the Nissan-Europe case with 30% volatility.

Nissan calculations for 30% volatility		
	Make-to-order	Full lead time
Expected demand per unit forecast	1.05	1.05
Profit-maximizing service level	100%	85%
Order quantity per unit forecast	1.05	1.4
Fill rate	100%	96.5%
Expected sales per unit forecast	1.05	1.01
Expected left-over inventory per unit forecast	0	0.36
Expected profit per unit forecast with 2% cost differential	52	48
Required cost differential	0%	7%
Required cost differential if fill rate 99.6% (service level 98%)	0%	9%
Required cost differential if service level 50%	0%	10%

9. The same analysis can be done beginning with a fixed long-lead-time cost to determine the maximum cost *premium* justified by elimination of costs related to demand volatility exposure. In this case, if we start with a long-lead-time cost of 49, these calculations establish that a 10% cost premium is justified by the elimination of demand-volatility exposure.
10. Use of the actual-to-forecast ratio allows us to separately incorporate changes in the forecast such as occur with the tender losses experienced by GSK Vaccines, or after observing early sales.
11. We can test the impact of requiring a service level that exceeds the newsvendor critical fractile of 85%. If company policy is to maintain a fill rate of 99.6% (which corresponds to a service level of 98%), the required cost differential increases to 9%.
12. We have not encountered companies that target a service level that is less than the newsvendor critical fractile. Were a company to reduce the target service level from 85% to 70%, the required cost differential would rise to 10%.

Table A.2 summarizes the results of these calculations.

Appendix B. Capturing stochastic volatility parameters from time-series data

We use the square-root GARCH model proposed by Heston and Nandi (2000) to estimate parameter values, assuming the following demand process $D(t)$ over time steps of length Δ:

$$\log D(t) = \log D(t-\Delta) + r + \lambda h(t) + \sqrt{h(t)}z(t) \tag{B.1}$$

where r is the continuously compounded rate for the time interval Δ, $z(t)$ is a standard normal disturbance, and the conditional variance $h(t)$ is explained by

$$h(t) = \omega + \beta h(t-\Delta) + \alpha\left[z(t-\Delta) - \gamma\sqrt{h(t-\Delta)}\right]^2 \tag{B.2}$$

If α and β are zero, the process coincides with the discrete-time geometric Brownian motion found in the constant volatility model. The long-run variance can be evaluated as

$$E[h(t)] = \frac{\alpha+\omega}{1-\beta-\alpha\gamma^2}$$

The mean reversion rate is equal to $\beta + \alpha\gamma^2$.

The Nestlé data consists of 113 weeks of demand data aggregated over 4 weeks. The log returns demonstrate autocorrelation, which we removed using a moving average model with order 4 (MA(4)). We then applied the Heston–Nandi model to the residuals. Classical ARMA (autoregressive moving average) processes are

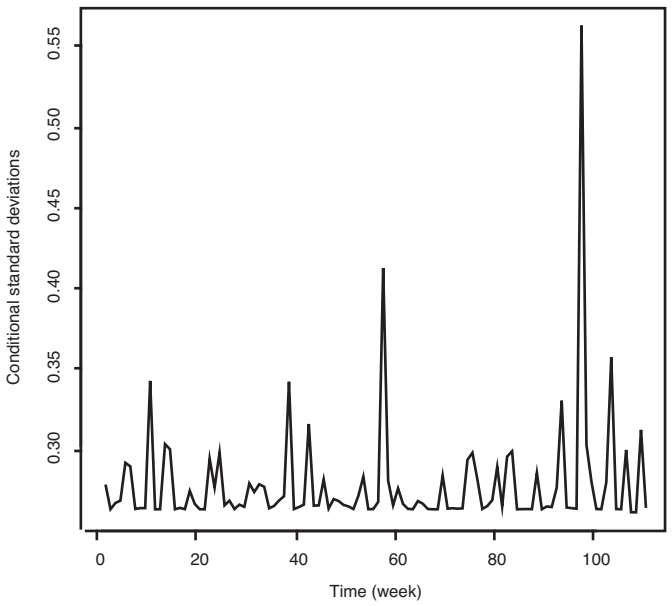

Fig. 8. Estimated conditional standard deviation $\hat{h}(t)$ over time.

constructed from white noise. Let $(\varepsilon_t)_{t\in\mathcal{Z}}$ be a white noise process with mean zero and finite variance σ_ε^2. The (ε_t) form the *innovations* that drive the ARMA process. A MA(q) process is defined as the linear sum of the noise (ε_t), with (X_t) following a MA(q) process if

$$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t.$$

Fig. 8 shows the estimated conditional standard deviation $\hat{h}(t)$ that nicely imitates the volatility behavior of the underlying process.

Appendix C. Calculating the fill rate under lognormal demand

Demand D is a lognormal random variable with parameters (μ,σ^2), with $E(D) = e^{\mu+\sigma^2/2}$. Let $Q = e^{\mu+z\sigma}$ denote the order quantity that is z geometric standard deviations above the median, corresponding to an in-stock probability of $\Phi(z)$.

An order quantity $Q = e^{\mu+z\sigma}$ yields expected sales $S(z)$:

$$\begin{aligned}
S(z) &= E[\min(Q,D)] = \int_0^Q Df(D)dD + (1-F(Q))Q \\
&= e^{\mu+\sigma^2/2}\Phi\left(\frac{\log(Q)-\mu}{\sigma} - \sigma\right) \\
&\quad + e^{\mu+z\sigma}\left(1-\Phi\left(\frac{\log(Q)-\mu}{\sigma}\right)\right) \\
&= E(D)\Phi\left(\frac{\log(Q)-\mu}{\sigma} - \sigma\right) \\
&\quad + E(D)e^{z\sigma-\sigma^2/2}\left(1-\Phi\left(\frac{\log(Q)-\mu}{\sigma}\right)\right) \\
&= E(D)\Phi\left(\frac{\log(e^{\mu+z\sigma})-\mu-\sigma^2}{\sigma}\right) \\
&\quad + E(D)e^{z\sigma-\sigma^2/2}\left(1-\Phi\left(\frac{\log(e^{\mu+z\sigma})-\mu}{\sigma}\right)\right) \\
&= E(D)(\Phi(z-\sigma) + e^{z\sigma-\sigma^2/2}(1-\Phi(z))). \qquad \text{(C.1)}
\end{aligned}$$

The fill rate $\Psi(z)$ is the ratio of expected sales to expected demand:

$$\Psi(z) = \frac{S(z)}{E(D)} = \Phi(z-\sigma) + e^{z\sigma-\sigma^2/2}(1-\Phi(z)). \qquad \text{(C.2)}$$

References

Abernathy, F., Dunlop, J., Hammond, J., Weil, D., 1999. A Stitch in Time: Lean Retailing and the Transformation of Manufacturing – Lessons from the Apparel and Textile Industries. Oxford University Press, USA.

Allon, G., Van Mieghem, J., 2010. Global dual sourcing: tailored base-surge allocation to near-and offshore production. Manage. Sci. 56 (1), 110.

Bicer, I., de Treville, S., Hagspiel, V., 2013. The Value of Reducing Lead Time under Non-stationary Demand. University of Lausanne, Working Paper.

Blackburn, J., 1991. Time-based Competition: The Next Battleground in American Manufacturing. Business One Irwin, Homewood, IL.

Blackburn, J., 2012. Valuing time in supply chains: Establishing limits of time-based competition. J. Oper. Manage. 30 (5), 396–405.

de Treville, S., Schürhoff, N., Trigeorgis, L., Avanzi, B., 2014. Optimal sourcing and lead-time reduction under evolutionary demand risk. Prod. Oper. Manage., http://dx.doi.org/10.1111/poms.12223, 15 pages.

Fisher, M., Hammond, J., Obermeyer, W., Raman, A., 1997. Configuring a supply chain to reduce the cost of demand uncertainty. Prod. Oper. Manage. 6 (3), 211–225.

Fisher, M., Raman, A., 1996. Reducing the cost of demand uncertainty through accurate response to early sales. Oper. Res. 44 (1), 87–99.

Fisher, M.L., 1997. What is the right supply chain for your products? Harv. Bus. Rev. 75 (2), 105–116.

Gallego, G., Özer, Ö., 2001. Integrating replenishment decisions with advance demand information. Manage. Sci. 47 (10), 1344–1360.

Gatheral, J., 2006. The Volatility Surface, 1st ed. Wiley-Finance, Hoboken, NJ.

Graves, S., Meal, H., Dasu, S., Qui, Y., 1986. Two-stage production planning in a dynamic environment. In: Axter, S., Schneeweiss, C., Silver, E. (Eds.), Multi-stage Production Planning and Control, Lecture Notes in Economics and Mathematical Systems. Vol. 266. Springer-Verlag, Berlin, pp. 9–43.

Hausman, W., 1969. Sequential decision problems: a model to exploit existing forecasters. Manage. Sci. 16 (2), 93–111.

Heath, D., Jackson, P., 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. IIE Trans. 26 (3), 17–30.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6 (2), 327.

Heston, S.L., Nandi, S., 2000. A closed-form GARCH option valuation model. Rev. Financ. Stud. 13 (3), 585–625.

Hopp, W.J., Spearman, M.L., 1996. Factory Physics, 1st ed. Irwin, Chicago, IL.

Hull, J.C., 1996. Options, Futures, and Other Derivatives, 3rd ed. Prentice-Hall, Upper Saddle River, New Jersey.

Iyer, A., Bergen, M., 1997. Quick response in manufacturer-retailer channels. Manage. Sci. 43 (4), 559–570.

Lee, H., Billington, C., 1995. The evolution of supply-chain-management models and practice at hewlett-packard. Interfaces 25 (5), 42–63.

Liker, J., 2004. The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer. McGraw-Hill, New York.

Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. Bioscience 51 (5), 341–352.

Milner, J., Kouvelis, P., 2005. Order quantity and timing flexibility in supply chains: the role of demand characteristics. Manage. Sci. 51 (6), 970–985.

Schweitzer, M., Cachon, G., 2000. Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. Manage. Sci., 404–420.

Suri, R., 1998. Quick Response Manufacturing. Productivity Press, Portland, OR.

Wager, S., de Treville, S., 2013. Constant Salvage Value Models: A Source of Systematic Bias in Predicting the Value of Lead-time Reduction. http://ssrn.com/abstract=2202422

Wang, Y., Tomlin, B., 2009. To wait or not to wait: Optimal ordering under lead time uncertainty and forecast updating. Nav. Res. Logist. 56 (8), 766–779.

Womack, J., Jones, D., Roos, D., Carpenter, D.S., 1990. The Machine That Changed the World. Simon & Schuster, Inc., New York, NY (2007).