Valuing Supply-Chain Responsiveness under Demand Jumps

Işık Biçer[1], Verena Hagspiel[1], Suzanne de Treville[1],

[a]*Rotterdam School of Management, Erasmus University*
Rotterdam 3000 DR, The Netherlands
[b]*Norwegian University of Science and Technology*
Department of Industrial Economics and Technology Management
7491 Trondheim, Norway
[c]*University of Lausanne*
Faculty of Economics and Business
1015 Lausanne, Switzerland

Abstract

As the time between the decision about what to produce and the moment when demand is observed (the *decision lead time*) increases, the demand forecast becomes more uncertain. Uncertainty can increase gradually in decision lead time, or can increase as a dramatic change in median demand. Whether the forecast evolves gradually or in *jumps* has important implications for the value of responsiveness, which we model as the cost premium worth paying to reduce the decision lead time (the *justified cost premium*). Demand uncertainty arising from jumps rather than from constant volatility increases the justified cost premium when an average jump increases median demand, but decreases the justified cost premium when an average jump decreases median demand. We fit our model to two data sets, first publicly available demand data from Reebok, then point-of-sale data from a supermarket chain. Finally, we present two special cases of the model, one covering a sudden loss of demand, and the other a one-time adjustment in median demand.

1. Introduction

Being able to postpone an order quantity decision until demand is known—operating with a null *decision lead time*—eliminates demand-risk exposure.[1] As the decision lead time increases, demand-

---

[*]Corresponding author: suzanne.detreville@unil.ch; Tel.:+41 21 692 3448.
[1]A decision lead time of zero means that the production decision can be postponed enough to permit working with firm demand data. Even with a decision lead time of zero, the delivery lead time may well be positive. It need only be short enough that the production can be decided after demand is observed.

risk exposure is likely to increase: The resulting stockouts or overstocks generate mismatch costs. The ability to postpone the decision about what to order so that the order quantity can be based on better demand information can be conceptualized as a real option (de Treville and Trigeorgis 2010), and that option's value can be estimated using quantitative-finance methods. Being able to quantify the value of reducing demand-uncertainty exposure that arises from an increase in the decision lead time transforms time into a decision variable.

The first step in estimating option value is to specify the *forecast-evolution process*: how demand uncertainty increases in decision lead time.[2] The simplest case is the random-walk assumption that underlies the Black-Scholes option-pricing model (Black and Scholes 1973). Each instant that the decision lead time increases, demand uncertainty increases by a minute amount following a geometric Brownian motion. When this constant-instantaneous-volatility process holds, demand is lognormally distributed with volatility increasing in the square root of the decision lead time. This assumption underlies the Cost-Differential Frontier decision tool proposed by de Treville et al. (2014b) that estimates the cost differential that must be offered by a long-lead-time supplier to compensate for the increase in demand-uncertainty exposure resulting from an increase in decision lead time.[3]

In practice, changes in demand may occur suddenly as a change in median demand (*jump*) rather than as an instantaneous increase in volatility. Demand is frequently subject to jumps: The World Economic Forum in its 2012 report on supply-chain risk attributed 44% of supply-chain disruptions to demand shocks (World Economic Forum 2012).[4] In finance, the limitations of the Black-Scholes model are well known, but the model is generally used as a reasonable approximation (e.g., Bakshi et al. 1997). When the true forecast-evolution process is subject to jumps but the mismatch cost is estimated assuming that all demand uncertainty emerges from a constant-volatility process, how bad is the error? Does the constant-volatility version of the model give a good enough approximation of the mismatch cost for practical purposes, or does the error impact decision making enough to warrant

---

[2]For a stock option, decision lead time translates into the time until the stock price is known.

[3]The tool also calculates the cost premium worth paying to reduce decision lead time.

[4]A sudden shift in median demand represents a common type of demand shock, see `https://www.investopedia.com/terms/d/demandshock.asp`.

the use of a more complex model?

To address this question we extend the Cost-Differential Frontier decision tool to include jumps following the classic model proposed by Merton (1976). We use publicly available demand data from Reebok to gain insight into how the choice of model impacts supply-chain decision making. Parsons (2004) studied the cost of demand-risk exposure faced by Reebok in the context of the exclusive license held during the period 2000-2010 to produce replica jerseys with the National Football League (NFL, see also Graves and Parsons 2005).[5] Available published Reebok data includes the mean and standard deviation of annual demand for replica jerseys for New England Patriots fans; price, cost, and residual value; and a qualitative description of the many types of demand jumps observed by Reebok. Parsons (2004,pp. 74-75) concluded his analysis of Reebok data by proposing that "perhaps the single greatest opportunity for Reebok is to improve its ability to respond to shifts in demand through shorter lead times." This data is used by Parsons (2004), Graves and Parsons (2005), Parsons and Graves (2005) to demonstrate the value of postponement. Cattani et al. (2008) cited this work as exemplifying the importance of analyzing the value of responsiveness (see also Uppala 2016). The decision by the authors of the Reebok study to make their data and analysis publicly available made it possible for us to build directly on their work and use our model to extend their analysis.

Our first result is that the impact of jumps on the cost premium worth paying to reduce decision lead time depends on whether a jump is expected to increase or decrease median demand. If a jump is expected to increase median demand, then treating demand uncertainty as though it came from a constant-volatility process results in an understatement of the justified cost premium. If, however, a jump is expected to reduce median demand, then assuming a constant-volatility process will lead to an overstatement of that cost premium. This result arises from how the jump changes the skewness of the marginal demand density. Jumps that are expected to increase median demand will increase skewness as long as they occur relatively rarely.[6] The resulting increase in right-tail weight increases the value of the option to postpone the production commitment. A jump that reduces median demand reduces

---

[5]The license was won by Nike in 2010.
[6]As jumps become less rare, their impact on the marginal density moves from the tail to the body.

skewness, making the postponement option less valuable. Managers with whom we have reviewed this result have found it counterintuitive, as they experience more concern about being stuck with excess inventory if a negative jump occurs than about stocking out following a positive jump.

In order to make the analysis as useful as possible to practitioners, we explore two special cases of jumps that are frequently encountered in practice. The first special case models the risk that demand would be completely lost. In the Reebok case this corresponds to a change of team jersey that reduces the value of inventory of the old model to zero. We show that adding any reasonable risk of demand loss to a constant-volatility process substantially increases the justified cost premium. The second special case models a one-time update of median demand such as occurs when decision makers obtain early-sales data, which we use to quantify the impact of a possible Super-Bowl win on the cost premium worth paying to reduce decision lead time. These results are not surprising in their direction, but they are striking in their magnitude. When the jumps that everyone knows to exist are explicitly considered in setting the decision lead time, the company is likely to much more aggressively reduce decision lead time.

A question that arose during the research project was whether demand jumps are experienced in supply chains. To address this question, we randomly selected two products from a supermarket chain and analyzed 100 observations of daily demand from point-of-sale data, then counted how many observations had standardized residuals more than three standard deviations from zero. The first product had four such outliers, indicative of demand jumps, and the second had none. We then considered what would happen if we forced a jump model on a product where it seemed like a constant-volatility assumption would suffice. By moving the threshold defining outliers from the usual three down to 2.52 standard deviations, the number of outliers for the second product increased from zero to four. Interestingly, treating these four points as jumps rather than normal variation for the second product substantially increased the cost-premium frontier. Which representation is correct? Our model cannot say. But, the fact that the option value of responsiveness is quite sensitive to when outliers are assumed to represent demand jumps indicates that this is an area that managers should be pondering.

2. Literature review

When we fit a normal or lognormal distribution to demand data, this is consistent with assuming a constant instantaneous-volatility process: The increase in demand uncertainty as the decision lead time increases can be modeled by a Brownian motion. Assuming that demand follows a normal (lognormal) distribution implies an underlying process that follows an arithmetic (geometric) Brownian motion. The standard deviation of demand (log demand) increases with the square root of the decision lead time. Modeling the evolution of a demand forecast as a constant instantaneous-volatility process is not new in supply-chain research. Hausman (1969) demonstrated that a forecast may plausibly evolve according to a geometric Brownian motion. This research formed the basis for the Martingale Model of Forecast Evolution (e.g., Heath and Jackson 1994, Milner and Kouvelis 2005). Oh and Özer (2013) emphasized the importance of accurately capturing the forecast-evolution process when deciding about investing in lead-time reduction, giving as an example the case where manufacturer and supplier have asymmetric information. De Treville et al. (2014b) demonstrated the use of a constant-volatility process to transform decision lead time into a decision variable, showing that an apparently compelling cost differential offered by a long-lead-time supplier may result in an increase in mismatch cost that eliminates the cost advantage when volatility is high and the residual value of the item being acquired is low. Including the value of responsiveness into decision making will frequently change the production-location decision. A summary of the calculations that underlie the Cost-Differential Frontier is given in Appendix E. The literature on the value of lead time is summarized in de Treville et al. (2014b) and de Treville et al. (2014a), so we refer readers to those papers. It is generally agreed that lead-time reduction makes companies better able to respond to demand uncertainty, whether from general randomness or demand shocks, which reduces the mismatch cost (Fisher and Raman 1996, Iyer and Bergen 1997, Milner and Kouvelis 2005, Lutze and Özer 2008). De Treville et al. (2014a) applied the insights arising from the addition of forecast evolution in three industrial settings: one with constant instantaneous volatility, one with stochastic instantaneous volatility due to a bullwhip effect, and one with a simple jump-diffusion process in which there was a risk of demand suddenly dropping to zero. The demand forecasts considered in de Treville et al. (2014b) and here are assumed

to eventually converge to the true value of demand. They are also assumed to be unbiased, so that the expected value of any given forecast update is zero (this is explained in further detail in de Treville et al. 2014a). De Treville et al. (2014b) considered instantaneous volatility that was both constant and stochastic, showing that stochasticity in the instantaneous volatility increases the value of lead time, especially when comparing longer decision lead times. Thus, demand-risk exposure can be a source of profit for a responsive firm. Similar results are emerging in the field of marketing concerning the use of demand clumpiness to increase profit (e.g., Zhang et al. 2014).

The ability to extract value from responsiveness has turned out to be a key piece of the puzzle to policy makers that recognize the importance of manufacturing to the local economy, and that are seeking to identify what kind of manufacturing has the best chance of being competitive in a high-cost economy.[7]

Demand jumps are described as a source of operational inefficiency in the economics and the operations management literatures (Lorenzoni 2009, Albuquerque and Bronnenberg 2012, Kesavan and Kushwaha 2014, Kesavan et al. 2016). Empirical research supports the idea that shorter lead times reduce the negative consequences of jumps (Kesavan et al. 2016, Tokar et al. 2014). Our models and analysis contribute to this literature in making it possible to quantify the value of responsiveness when demand uncertainty includes jumps.

3. The impact of jumps on the justified cost premium

3.1. Model presentation and application to Reebok demand data

In this section we present a model that estimates the cost premium worth paying to reduce the relative decision lead time when jumps are added to a constant-volatility forecast-evolution process. Not only does demand uncertainty incorporate the variance of log demand that increases linearly in

---

[7]A decision tool called the Cost-Differential Frontier emerged from this analysis. It addresses the question: What cost differential does a supplier have to offer to compensate for the increase in demand-volatility exposure arising from an increase in decision lead time, that is, the time between when the order must be placed and when demand is known with certainty? The Cost-Differential Frontier has now become part of the U.S. Department of Commerce toolbox to aid decision makers in valuing lead time (`acetool.commerce.gov/inventory`), and it has also been made available by the Swiss State Secretariat for Economic Affairs (`https://goo.gl/rjvQEO`).

lead time, but median demand might shift up or down.

Figure 1 compares sample paths from three processes that have the same constant instantaneous volatility, and that begin from a normalized demand forecast. The dashed (middle) line illustrates the case where noise is added in very small amounts each instant, captured by a geometric Brownian motion without drift. The solid (top) line combines geometric Brownian motion with a jump-diffusion process, with an average jump increasing median demand by 60%. The dotted (bottom) line combines a geometric Brownian motion with the risk of complete loss of demand. Our objective is to quantify how the form of the demand forecast-evolution process affects the value of lead time.
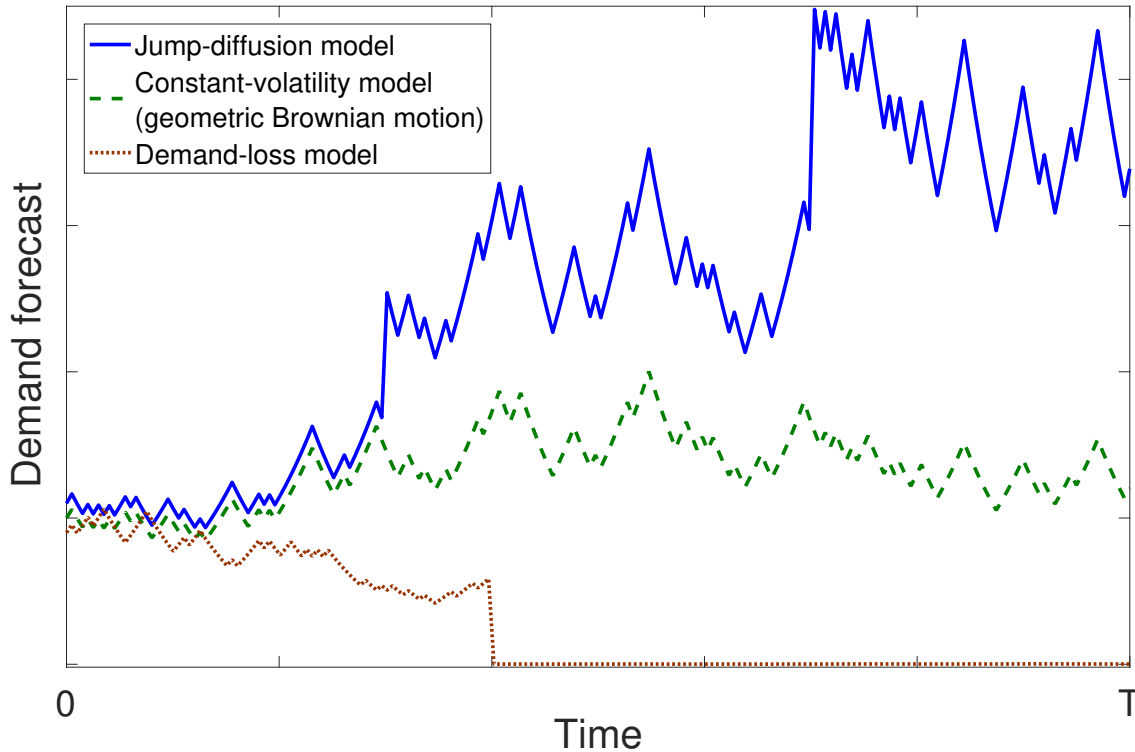

Figure 1: EVOLUTION OF DEMAND FORECASTS UNDER A CONSTANT-VOLATILITY (GEOMETRIC BROWNIAN MOTION), A JUMP-DIFFUSION, AND A DEMAND-LOSS MODEL

At time $t = 0$, the demand forecast is equal to actual demand. The constant instantaneous volatility σ is 0.8 for the three paths. The solid (top) line adds jumps that change median demand at the previous instant by an average of 60%. The dashed (middle) line denotes addition of noise to the forecast based solely on the geometric Brownian motion. The dotted (bottom) line adds the risk that at any moment demand may drop to 0.

We fit the model to publicly available demand data for Reebok replica NFL jerseys for the New England Patriots (Parsons 2004, Parsons and Graves 2005, Graves and Parsons 2005) to illustrate the impact of assumptions about the underlying forecast-evolution process on the cost premium worth paying to reduce decision lead time. Jerseys were produced by an Asian contract manufacturer, with the production decision made well before demand was known. Demand for jerseys evolved over the year based on team and player performance, also on player trades. As the Super Bowl approached, demand increased for teams that made it to the playoffs, with a dramatic increase in jersey demand for the winning team. We first analyzed this data using our models in early 2014, at which point the New England Patriots were estimated by various betting websites to have a 10% chance of winning the 2015 Super Bowl. That probability remained fairly constant throughout the fall of 2014. The Patriots won the 2015 Super Bowl and then won again in 2017, dramatically bringing our story to life. Such a win would be expected to increase expected demand several fold: Reaping the profit from being able to meet such a demand peak requires capacity rather than inventory. We calculate the justified cost premium arising from a possible Super Bowl win in Section 5.

Our objective in this section is to consider jumps that are less spectacular, such as occur during the season when teams perform above or below average or when players are traded. The probability of a jump occurring over a given relative lead time is captured by a Poisson process with jump intensity λ, following standard practice in quantitative finance (Merton 1976). For an order placed at time $t \in [0, 1]$ where demand will be known at time $t = 1$, the probability of a jump occurring is $1 - e^{-\lambda(1-t)}$. For an order placed at time $t = 0$—representing a full relative lead time—and a jump intensity of $\lambda = 0.2$ the probability of a jump occurring at some point is $1 - e^{-0.2} = 0.18$. If the order can be postponed until $t = 0.4$, the probability of a jump occurring between when the order is placed and when demand is observed is reduced to 0.11. The mathematical formulation of the general model is described in detail in Appendices A and B.

The results of our analysis are shown in Figure 2. The selling price for a blank jersey is \$21.60.[8]

---

[8]The player name could be added by the Asian contract manufacturer, or the blank jersey could be printed locally. We are only considering the demand-risk exposure for blank jerseys, and have adjusted the selling price accordingly: The published data gives the selling price for a printed jersey as \$24.00. We remove the cost of local printing of \$2.40

The residual value for a blank jersey—one that has not yet been printed with a player's name— left over after the season is $8.46. The cost of purchasing a blank jersey from an offshore contract manufacturer is $9.50. These values yield a critical fractile of 92.1%, which corresponds to a fill rate of 98.8% if log demand is normally distributed (de Treville et al. 2014a).

The bottom curve in Figure 2 shows the justified cost premium in the absence of jumps. We calibrate the volatility parameter by fitting a lognormal density to the demand data given in the case, assuming that the demand variation captured in case data has emerged from a constant instantaneous-volatility process. This yields a volatility parameter σ of around 0.22. We make the simplifying assumption that the published demand data does not include jumps: Although the case describes the kinds of jumps that are typically experienced in the industry, no jumps are mentioned with respect to the New England Patriots. If this assumption is incorrect and the historical data does include jumps, then model parameters should be adjusted to reduce the constant volatility parameter and increase the expected jump impact. The cost premium worth paying to eliminate demand-volatility exposure for these parameter values is less than 6%.

To this constant-volatility process we add jumps, following the qualitative description of jumps given in the published company information. Jumps increase demand uncertainty, so push up the cost-premium frontier. We consider three scenarios designed to have the same overall demand uncertainty: 1) jumps that are expected to increase expected demand when they occur, 2) jumps that are expected to reduce expected demand when they occur, and 3) demand uncertainty modeled as though coming from a constant-volatility process.

Again following standard practice in quantitative finance we assume that the jump follows a lognormal distribution (Merton 1976). We have programmed the model into a decision tool in R that is available at `https://goo.gl/ZnP35M` so that interested readers can test their own parameter values. We begin by considering the case where the location parameter is 0. This yields a median jump size of 1, so that a jump is equally likely to increase or decrease median demand. The result is shown in Figure 2 in the top curve, which describes a jump that is expected to increase median demand by

---

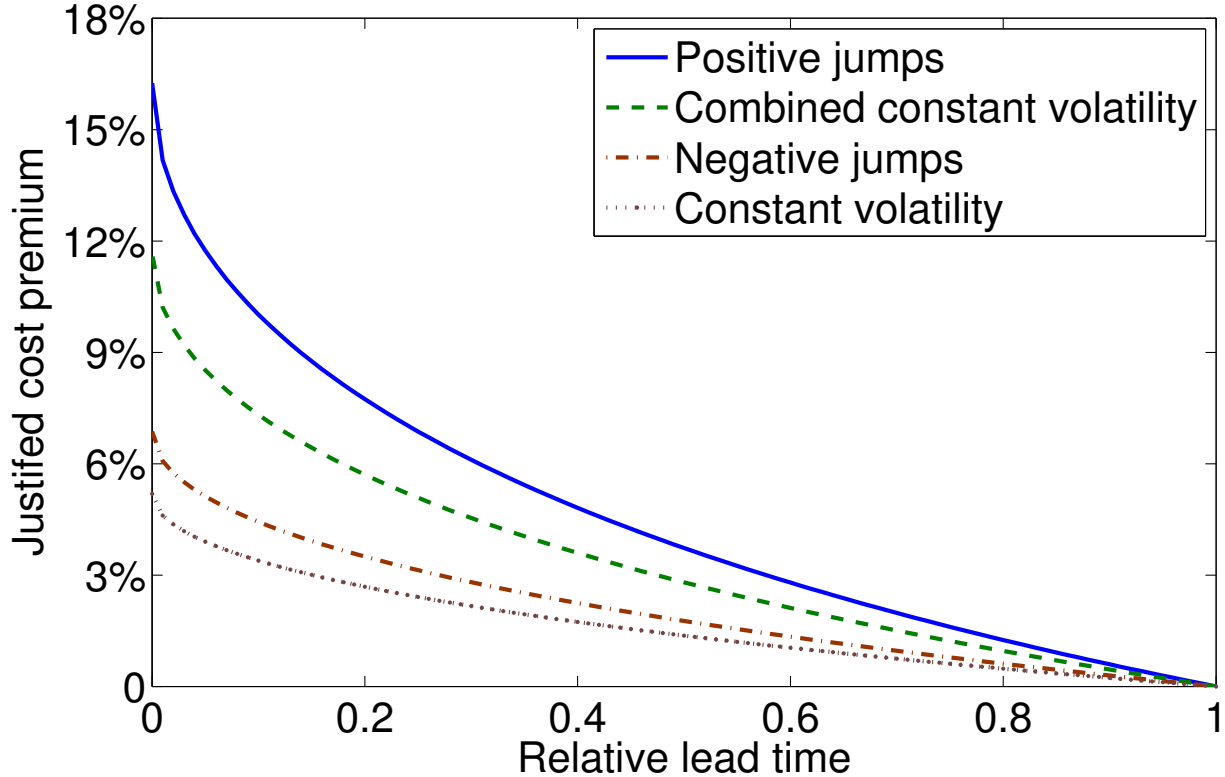to obtain our estimate of the unit price for a blank jersey of $21.60.

Figure 2: IMPACT OF JUMPS ON THE COST-PREMIUM FRONTIER

We set the selling price for a jersey to $21.60, the unit cost to $9.50, and the unit residual value to $8.46, yielding a critical fractile of 92.1%. The constant instantaneous volatility σ is set to 0.22, the jump intensity λ to 0.2, and $\bar{\sigma}$ to 0.43 for the modified constant-volatility model. The bottom curve shows the cost-premium frontier arising solely from constant volatility. Adding a jump to a constant-volatility process pushes up the cost-premium frontier because of the increase in demand uncertainty exposure. Positive jumps—shown in the top curve—are produced by a process Y_t that follows a lognormal distribution with parameters $\tau = 0$ and $\varsigma = 0.83$, so that a median jump leaves median demand unchanged, and the expected impact of a jump that occurs is $\mathbb{E}(Y_t) = 1.4$. The second curve from the top shows the cost-premium frontier with the same demand uncertainty as under positive jumps (shown in the top curve) but coming instead from a constant-volatility process. The third curve from the top shows the cost-premium frontier arising from negative jumps, where the expected impact of a jump is a 40% *reduction* in median demand, in contrast to the 40% *increase* represented in the top curve. The negative jump process follows a lognormal distribution with $\tau = -0.64$ and $\varsigma = 0.51$, which parameters were chosen by trial and error so that the negative jump curve represents the same demand uncertainty as the curves with the positive jump and the modified model.

40%, which corresponds to a jump volatility of 0.83.[9] For these parameter values the cost premium

---

[9]The mean of a lognormal density with location parameter τ and scale parameter ς is $e^{\tau + \varsigma^2/2}$. To consider a jump that increases median demand by 40%, we set $\log(1.4) = \tau + \sigma^2/2$. In Figure 2 we assume that a median jump does not change median demand, that is, the median jump size is 1 and $\tau = 0$. This allows us to solve for $\varsigma = 0.83$.

worth paying to eliminate demand uncertainty exceeds 15%.

Our next question is how the estimate of the justified cost premium would differ were we to correctly estimate the amount of demand uncertainty implied by the combination of constant volatility and the jump expected to increase median demand by 40% when it occurs, but incorrectly model it as coming from a constant-volatility process. Would this be close enough, or does the difference justify explicitly considering the two forecast-evolution processes? To answer this question we simulate demand values from the jump demand process and fit a lognormal distribution, which has a volatility of $\bar{\sigma} = 0.43$: A decision maker who interprets demand uncertainty from jumps as coming from a constant-volatility process will base her analysis on a volatility parameter of 0.43. The result is shown in the second curve from the top in Figure 2. If demand uncertainty from jumps is added but all demand uncertainty is treated as coming from a constant volatility process, the cost premium worth paying drops from 15% to around 10%. This simple illustration demonstrates the potential impact of correctly specifying the forecast-evolution process on the production-location decision: A 15% cost premium might suffice to make local production of blank jerseys feasible, but 10% might fall short.

Now let's consider a jump that is expected to reduce rather than increase demand by 40%.[10] The result is shown in the second curve from the bottom in Figure 2. The cost premium worth paying to eliminate demand uncertainty drops to around 7%, considerably below that for either positive jumps or the same demand uncertainty coming from a constant instantaneous process. The constant-volatility approximation thus overestimates the justified cost premium for negative jumps and underestimates it for positive jumps. This insight has the intriguing managerial implication that managers should be more willing to invest in reducing decision lead time in the face of positive than negative jumps. We have shared this result with several managers, all of whom have found it counterintuitive and contrary to how they manage risk. The reaction across the board has been that these managers find more worrisome—thus would be willing to pay a higher cost premium to avoid—jumps that reduce median demand by 40% when they occur than those that instead increase median demand by the

---

[10]We hold total demand uncertainty constant at $\bar{\sigma} = 0.43$. The expected jump size—reducing median demand by 40%–is 0.6. We used trial and error to find values of τ and ς such that $\bar{\sigma} = 0.43$ and $\log(0.6) = \tau + \varsigma^2/2$, which yielded $\tau = -0.64$ and $\varsigma = 0.51$.

same amount.

We lack an explicit form for the demand density arising from the jump-diffusion model. We can, however, use asymptotic analysis to gain initial understanding of how the cost-premium frontier differs when estimated under the modified rather than the jump model. For very small values of λ, τ, and ς we can approximate the density for the jump model using an Edgeworth expansion. Although the parameter values for which the Edgeworth series provides a reasonable approximation are much smaller than what we observe in practice, this asymptotic analysis allows us to formulate Propositions 1–3. These propositions are proved in Appendix A for the values for which the Edgeworth approximation holds. We then use numerical analysis to demonstrate that the propositions continue to hold for parameter values commonly encountered in a realistic supply-chain setting.

Proposition 1. *The cost premium worth paying to reduce decision lead time will be underestimated (overestimated) when demand uncertainty is approximated by a constant-volatility model and true demand uncertainty includes jumps that increase (decrease) the skewness of the marginal demand density.*

Proposition 2. *The skewness of the marginal demand density of the jump model increases in the median jump e^τ. For $\tau < 0$ the resulting decrease in skewness offsets the increase resulting from jump volatility, and may result in the marginal demand density of the jump model being lower than that of the constant-volatility approximation.*

Proposition 3. *The skewness of the marginal demand density of the jump model increases in the jump volatility ς both absolutely and relative to the constant-volatility approximation for $\tau \geq 0$ as long as jumps are relatively infrequent.*

The above calculations are based on the 92% critical fractile that emerges from the published Reebok information. The residual value for a blank jersey not sold during the demand period is described as 89% of the acquisition cost. This is a relatively high target service level, so even though the right tail of the marginal demand density for the full decision lead time under positive jumps is heavy, only 8% of the density lies to the right of what the order quantity should cover. Let's consider what happens if the residual value is lower than assumed, reducing the critical fractile. In Figure 3 we derive the Cost-Premium Frontier for positive jumps, varying the residual value as a percent of acquisition cost from 0 to 50 to the current 89 to 99. We observe that as the critical fractile decreases, the cost premium worth paying to reduce the decision lead time substantially increases.

If the critical fractile drops below median demand because either the cost of stocking out by one unit is less than the cost of one unit of overstock, or because the residual value is negative, the justified cost premium to eliminate decision lead time as calculated by our model will continue to increase. These cases represent a context that is very different from the one that we are considering (stocking out is costly and unsold units retain some value), so the results concerning the justified cost premium for very low critical fractiles should be interpreted with caution.
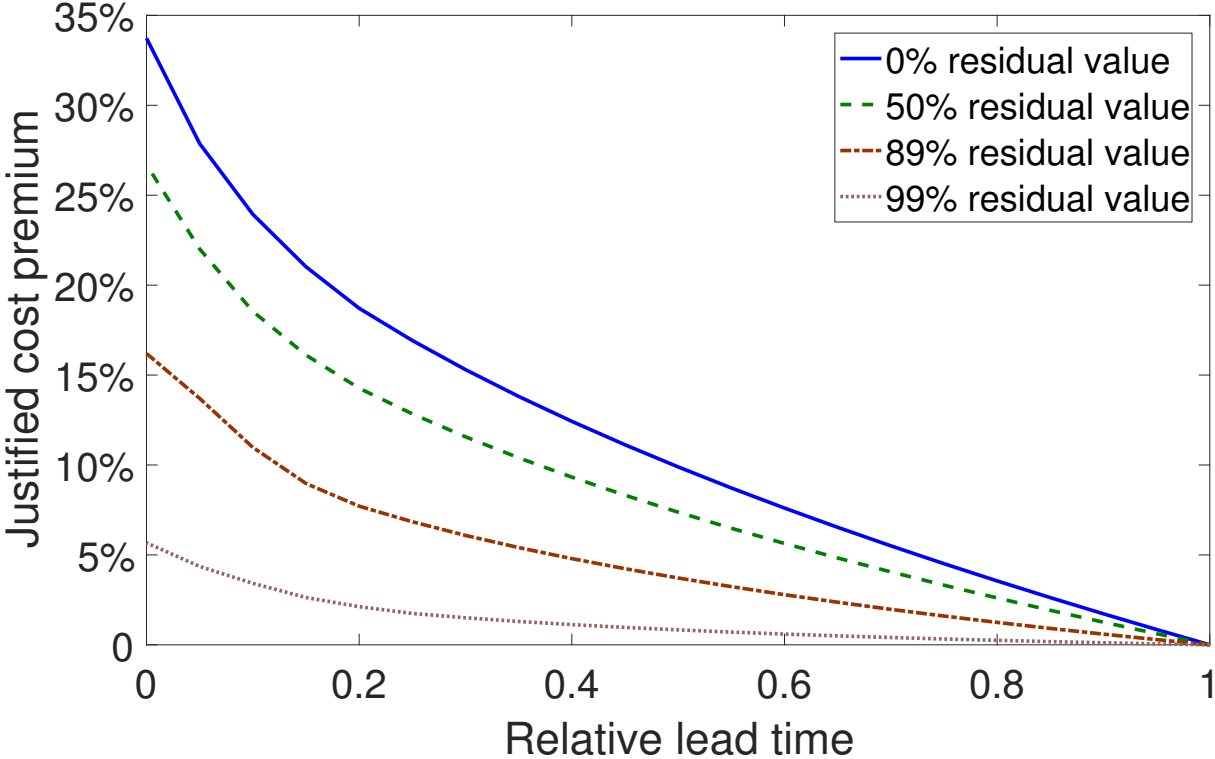

Figure 3: IMPACT OF RESIDUAL VALUE ON THE COST-PREMIUM FRONTIER

In Figure 3, we generate the Cost-Premium Frontier for positive jumps using the same Reebok parameter values except that we vary the residual value of an unsold unit of product as a percent of acquisition cost from 0 to 99%. The second curve from the bottom corresponds to the top curve in Figure 2. We see that as the residual value—hence critical fractile—decreases, the cost premium worth paying to reduced the decision lead time increases.

3.2. Estimating general-model parameter values from time-series data

In this section, we fit the general model to time-series data. We had access to point-of-sale data for 4000 products from a large supermarket retail chain, from which we chose two products at random

and considered daily sales over 100 days. Product 1 had four standardized residuals that were over three standard deviations from zero (4%). Product 2 had no standardized residuals more than three standard deviations from zero, but four that were at least 2.52 standard deviations from zero: The decision maker would need to use judgment in deciding whether these demand values arose from white noise in a constant-volatility process or indicated a jump. For both products, we fit a SARIMA model to the demand data, replaced the four outliers (potential outliers) by SARIMA forecast values, then used the outliers to estimate the parameters of a jump model. The next step was to compare the Cost-Premium Frontier under the two assumptions with critical fractiles of 99.5% and 70%. For both products, the Cost-Premium Frontier was considerably higher when outliers were treated as jumps rather than white noise when the critical fractile was very high. When the critical fractile was a more modest 70%, counting outliers as jumps had little impact on the frontier. In fact, because the revised SARIMA model provided a more accurate forecast for Product 1, the justified cost premium decreased slightly for the 70% critical fractile. The detailed calculations are provided in Appendix F.

Analysis of these two products suggests that when the service level is high, the value of responsiveness depends critically on how demand values that approach three-standard-deviation limits are interpreted. We thus propose:

Proposition 4. *For very high service levels, the threshold used to define outliers substantially impacts the premium worth paying to reduce the decision lead time.*

Our approach contrasts with more traditional treatment of outliers in demand data. That outliers arising from demand shocks can lead to an overstatement of underlying volatility is well understood. In demand forecasting, techniques such as winsorization have historically been used in practice to avoid such overestimation of volatility. Winsorization involves defining some maximum percentile (e.g., 95%), then replacing the demand value by the value at that threshold. Such approaches focus on the body of the marginal demand density, and do not make use of information provided by outliers. Our results suggest that—especially for high critical fractiles—the information provided by outliers may be of value.

4. Special case 1: The demand-loss model

In this section we develop the first special case of a jump-diffusion model in which there is a risk that demand would fall to zero. We continue our analysis of the Reebok case and explore the situation where a team suddenly changes its jersey, reducing the value of stock of the old jersey held by Reebok to zero.[11]
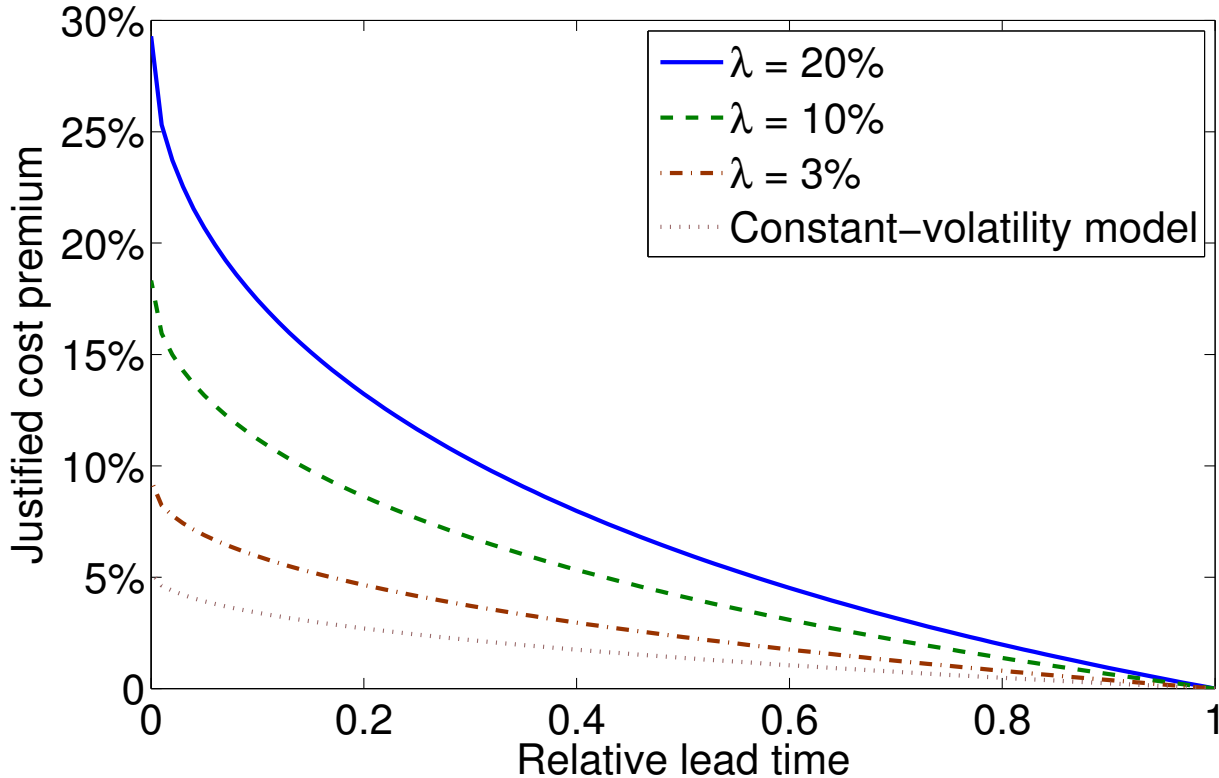

Figure 4: IMPACT OF THE RISK OF COMPLETE DEMAND LOSS ON THE VALUE OF LEAD-TIME REDUCTION

The unit price received for a blank jersey is $21.60, the unit cost to $9.50, and the unit residual value is $8.46. The constant instantaneous volatility without jumps $\sigma = 0.22$. We vary λ from 0% to 20%.

We now apply our model to Reebok's risk of jerseys losing their value following a jersey change. As in the previous section, demand volatility in the absence of jumps is $\sigma = 0.22$, the price received for the blank jersey is $21.60, the cost for production in Asia is $9.50, and the residual value is $8.46.

---

[11]The theoretical results developed in this section have been applied to tender-loss risk (as described in de Treville et al. 2014a, which cites an early version of this paper).

There are 32 NFL teams. Uniform changes are permitted every 5 years, suggesting a maximum jump intensity of 20%. In recent years, one to three teams have changed their uniform each year (`http://goo.gl/GbRvfk`), suggesting that jump intensities of 3% to 10% are worth incorporating into analysis of demand uncertainty. We vary the probability of a change of jersey λ from 0% (no risk of a jersey change, constant instantaneous volatility only) to 20%.

Figure 4 illustrates how the cost premium that Reebok should be willing to pay to avoid demand certainty increases in the probability of a sudden change in team jersey. Here, we are not comparing the jump model to a modified model as we did in the general case. Rather, we are demonstrating the understatement of the justified cost premium for reducing decision lead time that occurs when the risk of a complete loss of demand is not included in decision making. As the probability increases to 10% or 20%, the cost premium justified by the mismatch cost increases enough to make local production competitive. An R tool is available at `https://goo.gl/pLvw0m` to permit exploration of different jump intensity and constant-volatility combinations.

5. Special case 2: Making a single adjustment to the forecast

In this section, we consider a single adjustment to the forecast. If, for example, a team wins the Super Bowl, the demand forecast for replica jerseys will need to be revised. Another example comes from observing early sales, resulting in a forecast update as discussed by authors such as Fisher and Raman (1996) in considering how to allocate production between reactive and speculative capacity, and exemplified in the classic Sport Obermeyer case (Fisher and Raman 1996, Hammond and Raman 1995). De Treville et al. (2014b) extended analysis of the Sport Obermeyer case to consider lead time as a decision variable, modeling the observation of early sales either as a reduction of stochasticity in the instantaneous volatility, or as a one-time reduction in the right-tail heaviness of the marginal demand density. Here we consider a third alternative, modeling the observation of early sales as a jump in the forecast that causes the forecast to be multiplied by a factor that follows an independent lognormal distribution Y with parameters τ and ς. The median value of the factor is e^{τ} and the expected value is $e^{\tau+\varsigma^2/2}$. Demand uncertainty is thus composed of 1) the constant instantaneous

16

volatility σ that holds after the forecast is adjusted to take early sales into consideration, 2) the median e^τ of the factor by which the forecast is multiplied based on the early sales, and 3) the volatility ς of the factor.

Figure 5 illustrates estimation of the cost premium that Reebok should be willing to pay to postpone their production commitment incorporating the demand uncertainty arising from a possible Super-Bowl win. The dashed line shows the calculation assuming that all demand uncertainty comes from a constant-volatility model, and the solid line separates out the demand uncertainty coming from the Super Bowl. As the Super Bowl date approaches, some uncertainty will be resolved, but—as was observed in 2017—much remains until the last second of the game. Therefore, we will simplify the model to combine a constant-volatility process with a single update. The constant volatility is again 22%, the price remains at \$21.60, the long-lead-time unit cost at \$9.50, and the unit residual value at \$8.46.

We follow betting markets in assigning the probability of the New England Patriots winning the Super Bowl at 10% at the time when an order would need to be placed to the distant supplier. If a Super Bowl win is expected to increase demand by 200% and the probability of winning is estimated to be 10%, then the possible jump increases expected demand by 20%. The standard deviation changes by a factor of $\sqrt{(0.1 \times 3^2 + 0.9 \times 1^2) - 1.2^2} = 0.6$, so the coefficient of variation is 0.5, which corresponds to a volatility of $\sqrt{\log(0.5^2 + 1)} = 0.47$, justifying a cost premium of between 12 and 15%. An R tool that lets readers experiment with their own assessment of the probability of winning is available at `https://goo.gl/HvoZck`.

Managers are typically not able to estimate what the impact of a Super-Bowl win will be on demand. In our experience, managers are more comfortable answering the question: "Do you expect that a Super-Bowl win would at least double demand?" rather than trying to estimate directly by how much they expect that demand would increase. The model makes it possible to identify the demand impact that suffices to make a given cost premium required for local responsive production worth paying, making it easier for managers to quantify their intuition.
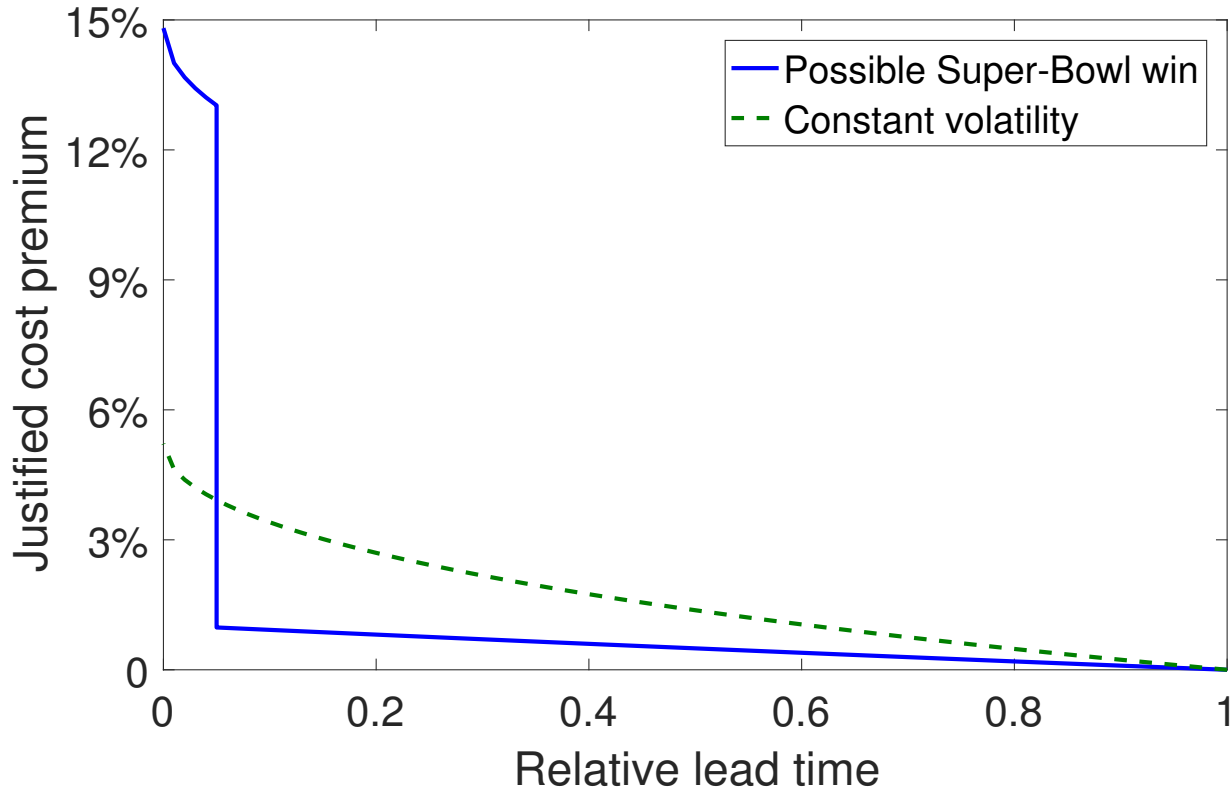
Figure 5: THE JUSTIFIED COST DIFFERENTIAL CONSIDERING DEMAND UNCERTAINTY FROM A POSSIBLE SUPER-BOWL
WIN
The cost differential worth paying to postpone the production decision enough to observe who wins the Super Bowl
under the jump and constant volatility assumptions. The probability of winning the Super Bowl at the time that an
offshore order would need to be placed is set to 10%, and demand for jerseys is expected to triple for the winning team.
If demand uncertainty is assumed to come from a constant-volatility process, then the cost premium worth paying is
less than 6%, but more than doubles when correctly modeled as coming from a jump process. The cost premium worth
paying to postpone the decision less than what is required to permit observing who has won is close to zero.

6. Conclusion

Although it is not traditional to include forecast evolution in calculating the supply-demand mis-
match cost, research is increasingly demonstrating the importance of including this dynamic in supply-
chain analysis. We contribute to this stream of research by exploring the impact of jumps on the cost
premium worth paying to reduce decision lead time. We ask: If forecast uncertainty evolves according
to a jump process but we make the simplifying assumption that the forecast evolves according to a
constant volatility, how much will our decision be affected? Option pricing has been used for many
decades in finance: Although the constant-volatility assumption is often violated and tools have been

developed that allow an accurate portrayal of the evolution process, results from models that assume a constant volatility are often good enough for practical purposes. In the supply-chain domain, does the same hold? Is correct incorporation of the forecast-evolution process likely to change key decisions about how much or where to produce, or whether to buffer with inventory or capacity?

We fit our general model to two data sets, first to published demand data for Reebok NFL replica jerseys, then to time-series data from a supermarket chain. We then model and explore using the Reebok data the risk that demand would drop to zero at some point during the decision lead time, then a one-time adjustment to median demand at an expected point in time. The empirical analysis demonstrates that consideration of jumps has a substantial impact on how much the company should be willing to invest in responsiveness when a jump is expected to increase median demand, but little impact when a jump is expected to reduce median demand. This in turn has far-reaching implications for the production-location decision.

The field of operations and supply-chain management has access to a wide variety of analytical models. Many of these models have the potential to be useful to decision makers, but are not used because the last mile of implementation is left to the users. And, many of these models have the potential to lead decision makers into error: It is only in observing what happens when we apply these models to a real problem that we see potential inconsistencies. Neglecting this last mile of implementation in research thus results in models that are less used and useful than they could be, and a sharper-than-needed divide between practice and theory. It is for these reasons that the *Journal of Operations Management* has called for empirically grounded analytical modeling `https://www.journals.elsevier.com/journal-of-operations-management/empirically-grounded-analytic-ega-models`, where model development is combined with empirically grounded parameterization. Our work exemplifies this category. Our primary contribution lies neither in the analytical model nor in its ability to generalize empirically, but rather in the interaction between model development and its impact on decision making in a typical application. For this reason, we focus in the body of the paper on the research question that lies at this intersection: Does the valuation of responsiveness change enough when using a model that accurately captures jumps in the forecast-evolution process

to compensate for increased model complexity? The model that we have developed to address that question is made available for users through online calculators. Model development is described in detail in the appendices, but is not given center stage as would be typical were the model the primary contribution. Similarly, although the two cases that we examine are realistic, we make no claims about the prevalence of jumps: Our focus is on how decision making changes when our model is used in an empirical setting where demand jumps are considered to be possible.

Appendix A   Development of the general model

The general model combines a compound Poisson and constant-volatility process to explore how jumps affect the mismatch cost, following the jump-diffusion option-pricing model developed by Merton (1976). We normalize lead time using the longest lead time under consideration: Actual demand is realized at $t = 1$ and the supplier with the longest decision lead time requires an order-quantity decision at relative lead time $t = 0$.

We normalize the demand forecast using median demand. The instantaneous volatility (assumed here to be constant) is denoted by σ. The standard Brownian motion process is denoted by Z_t. The jump intensity λ is captured by a Poisson process N_t. Consistent with jump models from quantitative finance (Merton 1976), we assume the jump Y_t to follow a lognormal distribution with parameters (τ, ς^2). A jump at time t changes the demand forecast from D_t to $D_t Y_t$, where $\mathbb{E}(Y_t) = e^{\tau + \varsigma^2/2}$. We offset the impact of jumps on the drift rate by adding the term $-\lambda(1 - \mathbb{E}(Y_t)) = -\lambda(e^{\tau + \varsigma^2/2} - 1)$.

Demand follows the process:

$$dD_t = -\lambda(e^{\tau + \varsigma^2/2} - 1)D_t dt + \sigma D_t dZ_t + (Y_t - 1)D_t dN_t, \tag{1}$$

Applying Ito's lemma, the log-demand process can be formulated as follows:

$$
\begin{aligned}
d(\log(D_t)) &= \left(-\lambda(e^{\tau + \varsigma^2/2} - 1) - \frac{1}{2}\sigma^2\right) dt + \sigma dZ_t + \log(Y_t)dN_t, \\
\log(D_t) &= \log(D_0) + \left(-\lambda(e^{\tau + \varsigma^2/2} - 1) - \frac{1}{2}\sigma^2\right) t + \sigma Z_t + \sum_{i=1}^{N_t} \log(Y_i).
\end{aligned}
\tag{2}
$$

The demand processes D_t^S and D_t^J have the same expected demand ($\mathbb{E}(D_1^J) = \mathbb{E}(D_1^S)$) and the same instantaneous volatility: D_t^S evolves according to a constant-volatility process, and D_t^J adds a jump-diffusion process to D_t^S.

$$dD_t^S = \sigma D_t^S dZ_t, \tag{3}$$

$$dD_t^J = -\lambda k D_t^J dt + \sigma D_t^J dZ_t + (Y_t - 1) D_t^J dN_t. \tag{4}$$

We simulate demand values from the jump process and fit a lognormal density to those values, defining a modified process D_t^M

$$dD_t^M = \overline{\sigma} D_t^M dZ_t, \tag{5}$$

with modified volatility $\overline{\sigma} > \sigma$ and where D_1^M and D_1^J have the same mean and variance.

We lack an explicit form for the demand density $f(\textbf{.})$ arising from the general jump-diffusion model given in Equation (4). When $\lambda^2 \approx 0$ and τ and ς are small, $f(\textbf{.})$ is close enough to a lognormal that it can be approximated using an Edgeworth expansion, as described in Jarrow and Rudd (1982). This makes it possible to prove Propositions 1 to 3 for the Edgeworth case. Numerical analysis using larger values suggests that the same relationships continue to hold.

Proposition 1: The cost premium worth paying to reduce decision lead time will be underestimated (overestimated) when demand uncertainty is approximated by a constant-volatility model and true demand uncertainty includes jumps that increase (decrease) the skewness of the marginal demand density.

Proof.

$$f(D) = \varphi(D) + \frac{(\kappa_2(f) - \kappa_2(\varphi))}{2!}\frac{\partial^2 \varphi(D)}{\partial D^2} - \frac{(\kappa_3(f) - \kappa_3(\varphi))}{3!}\frac{\partial^3 \varphi(D)}{\partial D^3} + o(D^3), \tag{6}$$

where $\varphi(D)$ is a lognormal density used as the approximating density of demand, and $o(D^3)$ is the third-order approximation error (Jarrow and Rudd 1982). The term $\kappa_i(h)$, the i'th cumulant for a

given probability distribution $h \in \{f, \varphi\}$, is given by:

$$\kappa_1(h) = \int_{-\infty}^{+\infty} Dh(D)dD = \mathbb{E}(D), \tag{7}$$

$$\kappa_2(h) = \int_{-\infty}^{+\infty} (D - \mathbb{E}(D))^2 h(D)dD, \tag{8}$$

$$\kappa_3(h) = \int_{-\infty}^{+\infty} (D - \mathbb{E}(D))^3 h(D)dD, \tag{9}$$

The expected mismatch cost at time t is

$$
\begin{aligned}
MC_t(Q) &= (p - c)\mathbb{E}(D) - V_t(Q), \\
&= \left((p - s) \int_Q^\infty (D - Q)f(D)dD + (c - s)(Q - \mathbb{E}(D)) \right),
\end{aligned}
\tag{10}
$$

where $V_t(Q)$ is given by Equation (28).

Combining Equations (6) and (10), the third-order approximation of the mismatch cost is formulated as

$$
\begin{aligned}
MC_f(Q) &= MC_\varphi(Q) + e^{-r(1-t)}(p - s)\left\{ \int_Q^\infty (D - Q)\frac{(\kappa_2(f) - \kappa_2(\varphi))}{2!}\frac{\partial^2 \varphi(D)}{\partial D^2}dD_1 \right. \\
&\quad \left. - \int_Q^\infty (D - Q)\frac{(\kappa_3(f) - \kappa_3(\varphi))}{3!}\frac{\partial^3 \varphi(D)}{\partial D^3}dD \right\} + o(Q^3),
\end{aligned}
\tag{11}
$$

where $MC_f(Q)$ is the true value of the mismatch cost, and $MC_\varphi(Q)$ the mismatch-cost value based on the assumption that demand has a lognormal distribution. As D_t^J and D_t^M are constructed to have the same variance, then $\kappa_2(f) = \kappa_2(\varphi)$, and the second term drops out of the equation.

Jarrow and Rudd (1982) showed that for a lognormal distribution:

$$\int_Q^\infty (D - Q)\frac{\partial^j \varphi(D)}{\partial D^j}dD = \frac{\partial^{j-2}\varphi(x)}{\partial x^{j-2}}\bigg|_{x=Q}, \qquad \text{for} \quad j \geq 2. \tag{12}$$

Combining Equations (11) and (12) we obtain:

$$MC_f(Q) = MC_\varphi(Q) + e^{-r(1-t)}(p-s)\left\{ -\frac{(\kappa_3(f) - \kappa_3(\varphi))}{3!} \frac{\partial\varphi(x)}{\partial x}\bigg|_{x=Q} \right\} + o(Q^3). \quad (13)$$

From Equation (13) we see that the mismatch cost from the jump model exceeds that of the modified model when:

$$\frac{\partial\varphi(Q)}{\partial D_1} < 0, \text{ and} \quad (14)$$

$$\kappa_3(f) > \kappa_3(\varphi). \quad (15)$$

The first derivative of a lognormal density with parameters η and γ is

$$\frac{\partial\varphi(x)}{\partial x}\bigg|_{x=Q} = \frac{1}{Q^2\sqrt{2\pi\gamma^2}} e^{-\frac{(\log(Q)-\eta)^2}{2\gamma^2}} \left(-\frac{\log(Q)-\eta}{\gamma^2} - 1 \right), \text{ so that} \quad (16)$$

$$\frac{\partial\varphi(Q)}{\partial D_1} < 0 \iff Q > e^{\eta-\gamma^2}. \quad (17)$$

As $\gamma = \bar{\sigma}$, the first derivative of the lognormal density is negative for an order quantity $Q = e^{-\bar{\sigma}^2}$ times median demand, corresponding to a service level of $\Phi(-\bar{\sigma})$. As this is below median demand, we can infer that the first condition holds whenever an effort is made to match supply to demand.

The second condition (15) states that the mismatch cost is higher for the jump model when the third cumulant—the skewness—is greater for the jump than the lognormal model ($\kappa_3(f) > \kappa(\varphi)$), thus completing the proof based on asymptotic analysis. ∎

That making a constant-volatility assumption—using the modified model—understates the justified cost premium for positive jumps and understates it for negative jumps stems from the impact of the two jump types on skewness. A positive jump increases the skewness of the demand density, and a negative jump decreases it. More generally, the skewness of the demand density is increasing in τ. We thus propose:

Proposition 2: The skewness of the marginal demand density of the jump model increases in the median jump e^τ. For $\tau < 0$ the resulting decrease in skewness offsets the increase resulting from

jump volatility, and may result in the marginal demand density of the jump model being lower than that of the modified model.

Proof. The second and third cumulants of the modified model are formulated as:

$$\kappa_2(\varphi|\tau) = e^{\bar{\sigma}^2} - 1, \tag{18}$$

$$\kappa_3(\varphi|\tau) = (e^{\bar{\sigma}^2} + 2)(e^{\bar{\sigma}^2} - 1)^2. \tag{19}$$

The marginal density for the jump model has the same variance as that of the modified model such that:

$$\kappa_2(\varphi|\tau) = \kappa_2(f|\tau), \tag{20}$$

$$e^{\bar{\sigma}^2} = \lambda e^{2\tau + \varsigma^2}(e^{\varsigma^2} - 1) + e^{\sigma^2}. \tag{21}$$

Integrating Equation (21) into Equation (19) yields:

$$\kappa_3(\varphi|\tau) = (e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^2 + 3\lambda e^{2\tau + \varsigma^2}(e^{2\sigma^2} - 1)(e^{\varsigma^2} - 1) \tag{22}$$

for $\lambda^2 \approx 0$. The relative skewness equates to:

$$\kappa_3(f|\tau) - \kappa_3(\varphi|\tau) = \lambda e^{3\tau + (3/2)\varsigma^2}(e^{\varsigma^2} + 2)(e^{\varsigma^2} - 1)^2 - 3\lambda e^{2\tau + \varsigma^2}(e^{2\sigma^2} - 1)(e^{\varsigma^2} - 1), \tag{23}$$

$$= \lambda e^{2\tau + \varsigma^2}(e^{\varsigma^2} - 1)\left(e^{\tau + \varsigma^2/2}(e^{\varsigma^2} + 2)(e^{\varsigma^2} - 1) - 3(e^{2\sigma^2} - 1)\right). \tag{24}$$

For $\tau \ll 0$, the first part in the parenthesis of Equation (24) reduces to zero, making the relative skewness negative. This result together with Equation (26) completes the proof of Proposition 2. ∎

We use numerical analysis to explore whether Propositions 1 and 2—proved by asymptotic analysis—continue to hold for parameter values like those that we observe in practice. In Table 1 we use the cost parameters from the Reebok case ($p = \$21.60$, $c = \$9.50$, $s = \$8.46$, yielding a critical fractile of 92.1%). The first nine runs assume a jump intensity of 0.05, rising to 0.2 for the second nine runs.

We vary the jump volatility ς from 0.05 to 0.8. From Table 1 we see that the results continue to hold for the parameter values tested. The modified model overstates the true justified cost premium CP_{true} for negative jumps, and understates it for positive jumps. From these results, it is also useful to observe that the Edgeworth approximation is reasonably accurate for small values of λ and ς, but is inaccurate as ς increases even though λ is limited to 0.05. In finance the Edgeworth approximation is often close enough for use in estimation, but our numerical analysis demonstrates that this does not hold for parameter values that are typical in the supply-chain context.

Run	σ	λ	τ	ς	$\mathbb{E}[Y]-1$	CP_{true}	$CP_{edgeworth}$	$\overline{\sigma}$	$\overline{CP}$
1	0.22	0.05	-0.64	0.05	-0.472	5.52%	5.22%	0.2602	6.33%
2	0.22	0.05	-0.64	0.1	-0.471	5.52%	5.22%	0.2638	6.44%
3	0.22	0.05	-0.64	0.2	-0.462	5.51%	5.24%	0.2664	6.51%
4	0.22	0.05	-0.64	0.3	-0.449	5.51%	5.28%	0.2707	6.63%
5	0.22	0.05	-0.64	0.4	-0.429	5.52%	5.37%	0.2796	6.89%
6	0.22	0.05	-0.64	0.5	-0.403	5.57%	5.58%	0.2834	7.00%
7	0.22	0.05	-0.64	0.6	-0.369	5.67%	6.04%	0.2932	7.29%
8	0.22	0.05	-0.64	0.7	-0.326	5.84%	7.15%	0.3046	7.63%
9	0.22	0.05	-0.64	0.8	-0.274	6.08%	9.91%	0.3177	8.03%
10	0.22	0.05	0	0.05	0.0013	5.22%	5.22%	0.2202	5.22%
11	0.22	0.05	0	0.1	0.005	5.25%	5.23%	0.2205	5.23%
12	0.22	0.05	0	0.2	0.0202	5.36%	5.31%	0.2243	5.30%
13	0.22	0.05	0	0.3	0.046	5.56%	5.50%	0.2298	5.48%
14	0.22	0.05	0	0.4	0.0833	5.86%	5.96%	0.2388	5.73%
15	0.22	0.05	0	0.5	0.1331	6.24%	7.10%	0.2475	5.98%
16	0.22	0.05	0	0.6	0.1972	6.70%	9.87%	0.2576	6.26%
17	0.22	0.05	0	0.7	0.2776	7.27%	16.80%	0.2695	6.60%
18	0.22	0.05	0	0.8	0.3771	7.94%	34.60%	0.2834	7.00%

Table 1: The actual justified cost premium is compared to the estimated justified cost premium first using the Edgeworth approximation and then using the modified model. The price, cost, and salvage values are taken from the Reebok case. We hold λ constant at 0.05, and σ constant at 0.22. The first nine lines of the table represent negative jumps with $\tau = -0.64$, and the bottom nine lines represent positive jumps with $\tau = 0$. The results are consistent with Propositions 1 and 2: The justified cost premium is understated by the modified model for positive jumps, and overstated by the modified model for negative jumps. The results in the table further illustrate that the Edgeworth approximation does not provide a good estimate of the justified cost premium for larger values of ς, even though λ is relatively small in this example.

When jumps are relatively infrequent ($\lambda \leq 0.3$), increases in ς lead to increases both absolutely and relative to the modified model. As jumps become more frequent ($\Lambda \geq 0.5$), they impact the body of the demand density as well as the tail, so their impact is better captured by the modified model. We thus propose:

Proposition 3: The skewness of the marginal demand density of the jump model increases in

the jump volatility ς—absolutely and relative to the modified model—for $\tau \geq 0$ as long as the jump intensity is such that jumps are relatively infrequent.

Proof. The second and third cumulants for the jump-diffusion model are formulated as:

$$\kappa_2(f|\tau) = \lambda e^{2\tau+\varsigma^2}(e^{\varsigma^2} - 1) + e^{\sigma^2} - 1, \tag{25}$$

$$\kappa_3(f|\tau) = (e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^2 + \lambda e^{3\tau+(3/2)\varsigma^2}(e^{\varsigma^2} + 2)(e^{\varsigma^2} - 1)^2. \tag{26}$$

Equation (26) completes the proof of Proposition 3. ∎

For our numerical analysis of Proposition 3, we use a higher critical fractile ($p = \$100.00$, $c = \$1.00$, and $s = \$0.00$ to yield a critical fractile of 99%) to better explore the region under the right tail. The results are shown in Table 2. This table only considers positive jumps, as negative jumps reduce skewness. We have removed the column with the Edgeworth approximation because our focus here is on parameters values that go outside of its useful range. We have added a column with the ratio of the true justified cost premium to that estimated using the modified model. The first nine runs have a jump intensity $\lambda = 0.05$, increasing to 0.2 and then 0.5 for the next two sets of nine runs. As predicted by our asymptotic analysis, the ratio of the true justified cost premium to that of the modified model increases in ς for all values of λ. As we increase λ from 0.2 to 0.5, we see that $\frac{CP_{true}}{CP}$ decreases. To permit further exploration of how skewness changes in jump volatility—absolutely and relative to the skewness of the modified model—for given constant volatility and jump intensity, we have created an R tool that is available at `https://goo.gl/xkIVS8`[12].

Appendix B Derivation of the newsvendor profit-maximizing order quantity and expected profit by Fourier transform

Because the demand density arising from the general model is too complex to be presented in an explicit form, in this appendix we demonstrate the use of a Fourier transform to calculate the newsvendor profit-maximizing order quantity (e.g., Carr and Madan 1999, Lee 2004, Bakshi and

---

[12]In this tool τ is set to 0, so that a median jump leaves median demand unchanged. We recall from Proposition 2 that skewness increases in τ, holding all else constant.

Run	σ	λ	τ	ς	$\mathbb{E}[Y]-1$	CP_{true}	$\overline{\sigma}$	$\overline{CP}$	$\frac{CP_{true}}{\overline{CP}}$
1	0.22	0.05	0	0.05	0.0013	76.00%	0.2202	75.86%	1.002
2	0.22	0.05	0	0.1	0.005	76.49%	0.2205	75.99%	1.006
3	0.22	0.05	0	0.2	0.0202	79.65%	0.2243	76.20%	1.045
4	0.22	0.05	0	0.3	0.046	87.84%	0.2298	79.07%	1.11
5	0.22	0.05	0	0.4	0.0833	102.60%	0.2388	84.19%	1.22
6	0.22	0.05	0	0.5	0.1331	124.54%	0.2475	88.16%	1.41
7	0.22	0.05	0	0.6	0.1972	154.02%	0.2576	92.85%	1.66
8	0.22	0.05	0	0.7	0.2776	191.61%	0.2695	98.50%	1.95
9	0.22	0.05	0	0.8	0.3771	238.19%	0.2834	105.29%	2.26
10	0.22	0.2	0	0.05	0.0013	76.39%	0.221	76.31%	1.001
11	0.22	0.2	0	0.1	0.005	78.31%	0.225	78.05%	1.003
12	0.22	0.2	0	0.2	0.0202	89.86%	0.237	83.38%	1.078
13	0.22	0.2	0	0.3	0.046	116.19%	0.257	92.57%	1.25
14	0.22	0.2	0	0.4	0.0833	157.83%	0.283	105.09%	1.50
15	0.22	0.2	0	0.5	0.1331	213.42%	0.314	120.92%	1.76
16	0.22	0.2	0	0.6	0.1972	283.85%	0.347	138.88%	2.04
17	0.22	0.2	0	0.7	0.2776	372.07%	0.382	159.27%	2.34
18	0.22	0.2	0	0.8	0.3771	482.18%	0.416	180.44%	2.67
19	0.22	0.5	0	0.05	0.0013	77.17%	0.221	76.30%	1.011
20	0.22	0.5	0	0.1	0.005	81.84%	0.234	82.03%	0.997
21	0.22	0.5	0	0.2	0.0202	106.81%	0.261	94.45%	1.13
22	0.22	0.5	0	0.3	0.046	154.36%	0.301	114.16%	1.35
23	0.22	0.5	0	0.4	0.0833	222.24%	0.350	140.58%	1.58
24	0.22	0.5	0	0.5	0.1331	312.51%	0.401	170.93%	1.83
25	0.22	0.5	0	0.6	0.1972	430.42%	0.475	220.59%	1.95
26	0.22	0.5	0	0.7	0.2776	583.09%	0.551	279.20%	2.08
27	0.22	0.5	0	0.8	0.3771	779.47%	0.613	333.24%	2.34

Table 2: The actual justified cost premium is compared to the estimated justified cost premium using the modified model. The price, cost, and salvage values are $p = \$100$, $c = \$1$, and $s = \$0$ to yield a critical fractile of 99%. We hold τ constant at 0, and σ constant at 0.22. The first nine lines of the table represent low jump intensity with $\lambda = 0.05$; the next nine lines represent moderate jump intensity with $\lambda = 0.2$; and the last nine lines represent high jump intensity with $\lambda = 0.5$. The results are consistent with Proposition 3. The actual justified cost premium relative to the modified model increases in jump volatility.

Madan 2000). The Fourier transform is widely used in finance to derive option-pricing formulas when asset prices follow complex, non-Gaussian processes. Applying the Fourier transform, an integral over a probability distribution is first transformed from probability space to Fourier space. The explicit form of the integral is then obtained by the inverse Fourier theorem.

Let D_t be a random variable denoting demand at time t, p the retail price of the product per unit, c the cost of the product per unit, and s the salvage value per unit. We denote by $f(.)$ the probability density function of demand. We normalize the total time length in that demand forecasts evolve over $t \in [0, 1]$ and actual demand is realized at time $t = 1$. The newsvendor profit function for an order

quantity Q is given by:

$$\pi(D_1, Q) = (p - c)Q - (p - s)E\left[\max(Q - D_1, 0)\right]. \tag{27}$$

The expected profit for the demand period is

$$V(Q) = (p - c)Q - (p - s)\int_0^Q (Q - D)f(D|D_t)dD, \tag{28}$$

where $f(D|D_t)$ is the density function of demand conditional on D_t.

We recall that demand follows the process:

$$dD_t = -\lambda(e^{\tau + \varsigma^2/2} - 1)D_t dt + \sigma D_t dZ_t + (Y_t - 1)D_t dN_t, \tag{29}$$

The log-demand process is then:

$$d(\log(D_t)) = \left(-\lambda(e^{\tau + \varsigma^2/2} - 1) - \frac{1}{2}\sigma^2\right)dt + \sigma dZ_t + \log(Y_t)dN_t,$$

$$\log(D_t) = \log(D_0) + \left(-\lambda(e^{\tau + \varsigma^2/2} - 1) - \frac{1}{2}\sigma^2\right)t + \sigma Z_t + \sum_{i=1}^{N_t}\log(Y_i). \tag{30}$$

We rewrite $\log(D_t)$ as the sum of four parts by defining $\mathscr{A} = \log(D_0)$, $\mathscr{B} = \left(-\lambda k - \frac{1}{2}\sigma^2\right)t$, $\mathscr{C} = \sigma Z_t$, and $\mathscr{D} = \sum_{i=1}^{N_t}\log(Y_i)$.

The characteristic function of $\log(D_t)$ is created from these four parts as follows:

$$\phi_{\log(D_t)}(\omega) = \phi_{\mathscr{A}}(\omega) \times \phi_{\mathscr{B}}(\omega) \times \phi_{\mathscr{C}}(\omega) \times \phi_{\mathscr{D}}(\omega), \tag{31}$$

with

$$\phi_{\mathscr{A}}(\omega) = D_0^{i\omega}, \quad \phi_{\mathscr{B}}(\omega) = e^{i\omega\left(-\lambda k - \frac{1}{2}\sigma^2\right)t}, \quad \phi_{\mathscr{C}}(\omega) = e^{-\frac{\sigma^2 t\omega^2}{2}}, \quad \phi_{\mathscr{D}}(\omega) = e^{\lambda t(\phi_{\tilde{y}}(\omega) - 1)},$$

where $\phi_{\tilde{y}}(\omega)$ is the characteristic function of $\log(Y_i)$, and $\phi_{\tilde{y}}(\omega) = e^{\left(i\omega\tau - \frac{\varsigma^2\omega^2}{2}\right)}$.

Theorem 1. *If demand forecasts follow a process with a characteristic function $\phi(.)$, the optimal order quantity Q^* satisfies the equation:*

$$\beta Q^* = \frac{e^{-\alpha \log(Q^*)}}{\pi} \int_0^{+\infty} Re\left(e^{-i\omega \log(Q^*)} \frac{\phi(\omega - (\alpha + 1)i)}{-(i\omega + \alpha + 1)} \right) d\omega, \tag{32}$$

where $\beta = \frac{p-c}{p-s}$ is the newsvendor critical fractile, $Re(.)$ is an operator that gives the real part of a complex number, $i = \sqrt{-1}$, and α is a damping factor required to permit square integrability as described in Carr and Madan (1999) (as explained below).

To apply the Fourier transform to the expected net present value of the profit, we rewrite Equation (28) by change of variables:

$$V_t(q) = e^{-r(1-t)}(p-c)e^q - (p-s)e^{-r(1-t)} \int_{-\infty}^q (e^q - e^{y_1}) f(y_1|y_t) dy_1, \tag{33}$$

where $y_1 = \log(D_1)$, and $q = \log(Q)$. Let $P(q)$ denote the integral in Equation (33).

$$P(q) = \int_{-\infty}^q (e^q - e^{y_1}) f(y_1|y_t) dy_1. \tag{34}$$

The term $P(q)$ in Equation (34) tends to $+\infty$ as $q \to +\infty$, thus $P(q)$ is not square-integrable.

Carr and Madan (1999) note that the sufficient condition for square integration is

$$\int_{-\infty}^{\infty} |P(q))|^2 dq < \infty. \tag{35}$$

and suggest to work instead with

$$p(q) = e^{\alpha q} P(q), \tag{36}$$

where $\alpha < -1$ serves as a damping factor that makes $p(q)$ square integrable.

The Fourier transform of $p(q)$ is:

$$\psi(\omega) = \int_{-\infty}^{+\infty} e^{i\omega q} e^{\alpha q} \int_{-\infty}^{q} (e^q - e^{y_1}) f(y_1|y_t) dy_1 dq, \tag{37}$$

$$= \frac{\phi(\omega - (\alpha+1)i)}{\alpha^2 + \alpha - \omega^2 + i\omega(2\alpha+1)}, \tag{38}$$

where $\phi(.)$ is the characteristic function of the random variable y_1. Using the inverse Fourier transform, Equation (34) becomes

$$P(q) = \frac{e^{-\alpha q}}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega q} \psi(\omega) d\omega, \tag{39}$$

where $\psi(\omega)$ is given by Equation (38). Since $P(q)$ is a real number without any imaginary part (Carr and Madan 1999), Equation (39) can be rewritten as

$$P(q) = \frac{e^{-\alpha q}}{2\pi} \int_{-\infty}^{+\infty} Re\left(e^{-i\omega q}\psi(\omega)\right) d\omega = \frac{e^{-\alpha q}}{\pi} \int_{0}^{+\infty} Re\left(e^{-i\omega q}\psi(\omega)\right) d\omega. \tag{40}$$

The expected net value of the profit, given by Equation (33), in terms of $q = \log(Q)$ becomes

$$V_t(q) = e^{-r(1-t)}(p-c)e^q - (p-s)e^{-r(1-t)}\left(\frac{e^{-\alpha q}}{\pi} \int_{0}^{+\infty} Re\left(e^{-i\omega q}\psi(\omega)\right) d\omega\right). \tag{41}$$

Equation (41) gives the expected profit function in terms of the characteristic function of demand. Let q^* denote the value of q that maximizes $V_t(q)$:

$$q^* = \arg\max_{q} V_t(q) \tag{42}$$

$$\frac{dV_t(q)}{dq} = (p-c)e^q - (p-s)\frac{dP(q)}{dq} = 0, \tag{43}$$

$$\beta e^{q^*} = \frac{e^{-\alpha q^*}}{\pi} \int_{0}^{+\infty} Re\left(e^{-i\omega q^*}\frac{\phi(\omega-(\alpha+1)i)}{-(i\omega+\alpha+1)}\right) d\omega. \tag{44}$$

The profit-maximizing order quantity for a full lead time is:

30

$$\beta_L e^{q_0^*} = \frac{e^{-\alpha q_0^*}}{\pi} \int_0^{+\infty} Re\left(e^{-i\omega q_0^*} \frac{\phi(\omega - (\alpha+1)i)}{-(i\omega + \alpha + 1)}\right) d\omega, \tag{45}$$

This yields the full-lead-time expected profit assuming a profit-maximizing order quantity:

$$V_0(q_0^*|c_L) = (p - c_L)e^{q_0^*} - (p - s)\left(\frac{e^{-\alpha q_0^*}}{\pi} \int_0^{+\infty} Re\left(e^{-i\omega q_0^*}\psi(\omega)\right) d\omega\right), \tag{46}$$

For lead time $T - t$, we iterate to find the unit cost that yields an equivalent maximum profit:

$$c_t = \{c|V_t(q_t^*|c) = V_0(q_0^*|c_L)\}. \tag{47}$$

The cost premium CP worth paying to obtain the reduced demand-risk exposure resulting from reducing the lead time from T to $T - t$ is then:

$$CP = c_t/c_L - 1. \tag{48}$$

Appendix C   Development of the Demand-Loss Model

Demand forecasting here has two components: a jump-to-zero component where demand or sales is lost, and constant volatility otherwise. We represent the demand-loss process using the stochastic differential equation given by Equation (49).[13] The demand forecast at time t is denoted by D_t, σ the instantaneous volatility (assumed to be constant), Z_t the standard Brownian motion process, and N_t a Poisson process with intensity λ. We again assume that the supplier with the longest lead time

---

[13]This equation is used in the quantitative finance literature to model the default risk of assets (for a complete review of this jump-to-default model, see Gatheral 2006).

31

requires that the order be placed at $t = 0$, with actual demand realized at $t = 1$.

$$dD_t \;=\; \lambda D_t dt + \sigma D_t dZ_t - D_t dN_t. \tag{49}$$

The drift rate in Equation (49) is adjusted by adding λ to offset the effects of the jumps on the drift, thereby making the absolute drift rate equal to zero. Under Equation (49), a jump reduces demand to zero, and its value remains at zero after the jump. If no jump occurs, demand forecasts follow a geometric Brownian motion with a drift rate of λ and an instantaneous volatility of σ. The log-demand process can be formulated as

$$
\begin{aligned}
d(\log(D_t)) \;&=\; \left(\lambda - \frac{1}{2}\sigma^2\right) dt + \sigma dZ_t + \log(0) dN_t, \\[2mm]
\log(D_t) \;&=\;
\begin{cases}
\log(D_0) + \left(\lambda - \frac{1}{2}\sigma^2\right) t + \sigma Z_t & \text{if } t < \widehat{t}, \\[2mm]
\log(0) & \text{otherwise},
\end{cases}
\end{aligned}
\tag{50}
$$

where $\widehat{t}$ is the time of the first jump. The density function of demand conditional on the demand forecast at time t can be written as follows:

$$f(D_1|D_t = 0) \;=\; \delta(D_1), \tag{51}$$

$$f(D_1|D_t > 0) \;=\;
\begin{cases}
1 - e^{-\lambda(1-t)}, & D_1 = 0 \\[2mm]
e^{-\lambda(1-t)} \dfrac{1}{D_1 \sigma \sqrt{(1-t)}\sqrt{2\pi}} e^{-\frac{(\log(D_1) - \log(D_t) - (\lambda - \sigma^2/2)(1-t))^2}{2\sigma^2(1-t)}}, & D_1 > 0
\end{cases}
\tag{52}$$

where $\delta(D_1)$ is the Dirac delta function:

$$\delta(D_1) \;=\;
\begin{cases}
\infty, & \text{if } D_1 = 0, \\[2mm]
0, & \text{otherwise.}
\end{cases}
\tag{53}$$

Equation (51) implies that $D_1 = D_t = 0$ if a jump occurs between 0 and t. Otherwise, the point density of demand at $D_1 = 0$ is equal to the probability of observing a jump between t and 1. If there is no jump occurring between t and 1, the value of demand would be larger than 0 ($D_1 > 0$), and the

32

demand follows a lognormal distribution as given in Equation (52).

Appendix D   Development of the Early-Sales Model

We combine the geometric Brownian motion process with a single jump at time t_{ES} as follows:

$$dD_t = (1 - e^{\tau + \varsigma^2/2})D_t dt + \sigma D_t dZ_t + (Y - 1)D_t d\Theta_t, \qquad (54)$$

where

Θ_t is a deterministic time, discrete-state process:

$$\Theta_t = \begin{cases} 1 & \text{if } t = t_{ES}, \\ 0 & \text{otherwise.} \end{cases} \qquad (55)$$

In Equation (54), we add a drift term with a rate of $1 - e^{\tau + \varsigma^2/2}$ to offset the effect of the jump on the drift. Demand forecasts now follow a geometric Brownian motion process with a drift rate of $1 - e^{\tau + \varsigma^2/2}$ and an instantaneous volatility of σ when $t < t_{ES}$. At $t = t_{ES}$, the demand forecast jumps from D_t to $e^{\tau + \varsigma^2/2} D_t$, and then follows the geometric Brownian motion with the parameters $1 - e^{\tau + \varsigma^2/2}$ and σ. The log-demand process can be written as:

$$\log(D_1) = \log(D_t) + \left((1 - e^{\tau + \varsigma^2/2}) - \frac{1}{2}\sigma^2 \right)(1 - t) + \sigma Z_{\{1-t\}} + (\tau + \varsigma^2/2)1_{\{t < t_{ES}\}}, \qquad (56)$$

where $t \in [0, 1]$, and $1_{\{t_{ES}\}}(t)$ is an indicator function such that

$$1_{\{t < t_{ES}\}} = \begin{cases} 1 & \text{if } t < t_{ES}, \\ 0 & \text{otherwise.} \end{cases}$$

The distribution of a demand forecast made at time $t \in [0, 1]$ and realized at time $t = 1$ will be

$$\log(D_1)|D_t \sim \mathcal{N}\left(\log(D_t) + (1 - e^{\tau + \varsigma^2/2}) - \sigma^2/2)(1 - t), \sigma^2(1 - t) + \varsigma^2 1_{\{t < t_{ES}\}} \right). \qquad (57)$$

33

From Equation (57) we can see that the log variance of the conditional distribution of demand before observing the early sales (i.e., $t < t_{ES}$) is simply the weighted average of the squared jump and constant volatilities, and the median is the product of the median under constant volatility times the expected value of the jump factor.

Appendix E Review of the Cost-Differential Frontier

Here we review the calculation of the cost differential that underlies the constant instantaneous-volatility version of the Cost-Differential Frontier (de Treville et al. 2014b,a). In its simplest form, the Cost-Differential Frontier represents evolution of demand forecasts from time $t = 0$ to $t = T$ by a geometric Brownian motion. Actual demand is realized at time $t = T$. The total time period T is normalized by the longest lead time under consideration, which is equivalent to setting $T = 1$. The volatility parameter σ is then calibrated using the volatility for the full lead time.

Let D_t denote the initial demand forecast at time t, σ the constant instantaneous volatility, μ the rate at which median demand approaches average demand as t approaches 1, and $(Z_t)_{t \geq 0}$ a standard Brownian motion. Then,

$$dD_t \;\; = \;\; \mu D_t dt + \sigma D_t dZ_t. \tag{58}$$

The demand forecast at time $t \in [0,1]$ follows a lognormal distribution with the following parameters:

$$\log(D_1)|D_t \sim \mathcal{N}\Big(\log(D_t) + (\mu - \sigma^2/2)(1-t), \sigma^2(1-t) \Big). \tag{59}$$

The expected value of the newsvendor profit for an order quantity Q is given by:

$$E[\pi(Q)|D_t] \;\; = \;\; (p-c)Q - (p-s)\int_0^Q (Q - D_1)f(D_1|D_t)dD_1, \tag{60}$$

where $f(\cdot|\cdot)$ denotes the conditional demand density that follows a lognormal distribution given by Equation (59) for a geometric Brownian motion process with a drift of μ and a volatility of σ. The

optimal value of Q maximizing Equation (60) is found by the critical-fractile approach:

$$Q^* = F^{-1}\left(\frac{p-c}{p-s}\right), \tag{61}$$

$$z^* = \frac{\log(Q^*/D_t) - (\mu - \sigma^2/2)(1-t)}{\sigma\sqrt{1-t}}. \tag{62}$$

Then,

$$
\begin{aligned}
E[\pi(Q^*)|D_t] &= (p-s)\int_0^Q D_1 f(D_1|D_t)dD_1, \\
&= (p-s)D_t e^{\mu(1-t)}\Phi(z^* - \sigma\sqrt{1-t}).
\end{aligned} \tag{63}
$$

Now, we consider a situation in which the decision maker places the profit-maximizing order quantity when $t = 0$ (i.e., full lead time) at a unit cost of c_l. Thus,

$$E[\pi(Q_l^*)|D_0] = (p-s)D_0 e^{\mu}\Phi(z_l^* - \sigma). \tag{64}$$

If the decision maker places the profit-maximizing order quantity when $t > 0$ (i.e., with a shorter lead time) at a unit cost $c_s > c_l$, the expected profit becomes

$$E[\pi(Q_s^*)|D_t] = (p-s)D_t e^{\mu(1-t)}\Phi(z_s^* - \sigma\sqrt{1-t}). \tag{65}$$

It follows from the equality of Equations (64) and (65):

$$z_s^* = z_l^* - \sigma(1 - \sqrt{1-t}). \tag{66}$$

As time t increases from 0 (the order is placed the full lead time before knowing demand) to 1 (the order is placed with full knowledge of demand), the volatility of demand will decrease from σ to 0. Expected demand is $D_0 e^{\mu}$ for all values of t because the process is a martingale. Median demand is $D_0 e^{\mu - \sigma^2/2}$ when $t = 0$, then increases to the mean as t approaches 1.

By normalizing the median of the initial demand forecast density to one by setting $D_0 = 1$ and

35

$\mu = \sigma^2/2$ we are able to work with expected demand, expected sales, the order quantity, and expected leftover inventory as multiples of median demand. Median demand approaches expected demand as t approaches 1 at a rate of $\sigma^2/2$, so median demand at t will be $e^{\sigma^2 t/2}$.

The quantity ordered at $t = 0$ (i.e., with a full lead time) is set to achieve a target service level θ, corresponding to $z = \Phi^{-1}(\theta)$ multiplicative standard deviations above the median, where $\Phi(\cdot)$ is the standard normal distribution function and $\Phi^{-1}(\cdot)$ is the inverse of $\Phi(\cdot)$.

The order quantity as a multiple of the median demand forecast at $t = 0$ then becomes:

$$Q = e^{\sigma z},$$

The fill rate will be (for a derivation of the fill rate, see de Treville et al. 2014a):

$$\Omega(z, \sigma) = \Phi(z - \sigma) + e^{z\sigma - \frac{\sigma^2}{2}}(1 - \Phi(z)). \tag{67}$$

Expected demand as a multiple of median demand at $t = 0$ will be:

$$\mathbb{E}[\text{Demand}] = e^{\sigma^2/2}. \tag{68}$$

Expected sales will be:

$$\mathbb{E}[\text{Sales}] = e^{\sigma^2/2}\Omega(z, \sigma). \tag{69}$$

Expected leftover inventory will be the difference between the order quantity and expected sales:

$$\mathbb{E}[\text{Leftover inventory}] = e^{z\sigma} - e^{\sigma^2/2}\Omega(z, \sigma). \tag{70}$$

If an order is placed based on exact information as to the quantity demanded, the expected profit

per unit of expected demand $(\pi_t)_{t=1}$ will be:

$$(\pi_t)_{t=1} = p - c_s, \tag{71}$$

where p denotes the retail price, and c_s is the short-lead-time cost of the product per unit. We then solve for the breakeven long-lead-time cost c_l:

$$p - c_s = (p - c_l)\Omega(z, \sigma) - (c_l - s)(e^{z\sigma - \sigma^2/2} - \Omega(z, \sigma)), \tag{72}$$

where s denotes the salvage value per unit. The cost differential that must be offered by the long-lead-time supplier to compensate for the mismatch cost is $1 - c_l/c_s$. Similarly, the cost premium a retailer is willing to pay at most to reduce lead time is $c_s/c_l - 1$. By definition, both cost differential and cost premium are positively correlated with the mismatch cost:

$$
\begin{aligned}
\mathbb{E}[\text{Mismatch cost}] &= (p - c_l)\mathbb{E}[\text{Lost sales}] + (c_l - s)\mathbb{E}[\text{Leftover inventory}], \\
&= (p - c_l)(\mathbb{E}[\text{Demand}] - \mathbb{E}[\text{Sales}]) + (c_l - s)\mathbb{E}[\text{Leftover inventory}], \\
&= (p - c_l)e^{\sigma^2/2}(1 - \Omega(z, \sigma)) + (c_l - s)(e^{z\sigma} - e^{\sigma^2/2}\Omega(z, \sigma)). \tag{73}
\end{aligned}
$$

Appendix F   Fitting a jump model from time-series data

In this section we apply prescriptive analytics to raw data to explore the kinds of jumps that are routinely encountered in analyzing time-series data. We randomly chose two products from a set of 4000 products sold by a large supermarket chain for which we have point-of-sale data. The first product exhibited behavior strongly suggestive of jumps, while the other was consistent with a constant-volatility assumption. We use the first product to illustrate how one estimates jump-model parameters from time-series demand data, then use the second to explore the implications of force fitting a jump model in the absence of obvious jumps. We only considered these two products. In future research, it would be interesting to observe the jumps that occur over the entire range of products for this data set and compare those values to other product types to gain an idea of the

range of jumps that one might expect to encounter across a wide variety of demand settings.

For the two products, we considered demand over 100 days. We followed the protocol recommended by Shumway and Stoffer (2011, following the protocol presented in Tables 3.1 and 3.3) to fit a model that combines seasonal (S), autoregressive (AR), integrated (I), and moving average (MA) components. The resulting SARIMA model is parameterized as $ARIMA(a, b, c)(x, y, z)_n$. The seasonality n is estimated by observing cycles that show up in the data. The first parameter values a and x specify the number of lags included in the autoregressive component (AR), first for the original values and then for those values after a lag of n periods, referred to as the seasonality. These autoregressive parameter values are estimated by observing when partial autocorrelations disappear. The second parameter values b and y specify the differencing (I) required to compensate for any trend in the data: When time-series data shows a trend, differencing the data points in sequence yields a stationary process ($\Delta x_t = x_t - x_{t-1}$). Differencing can be applied several times for non-linear trends (denoted by $\Delta^d x_t$ for d applications of differencing). The third parameter values c and z specify the number of lags included in the moving average (MA) component.

This SARIMA model allows us to replace extreme values of demand by a forecast. The extreme values are then used to fit the jump portion of the model.

F.1 Product 1

Figure 6 shows the autocorrelation (panel a) and partial autocorrelation (panel b) for the natural log of daily sales over 100 days for Product 1. The autocorrelation plot of the log of the daily sales shows continuous decay for daily and weekly autocorrelations. The weekly partial autocorrelations disappear after one lag, suggesting an AR1 model type. There does not appear to be a moving-average effect. These graphs suggest clear weekly seasonality, so we capture the seasonality (S) component of the SARIMA model by setting $n = 7$. As our time-series data does not show a trend we set $b = y = 0$. The resulting SARIMA model is of type $ARIMA(1, 0, 0)(1, 0, 0)_7$.

Figure 7 shows the standard output from the R package made available by Shumway and Stoffer (2011). The standardized residuals are shown in panel a. From panel b of Figure 7 we see that most autocorrelation values fall within 95% confidence intervals for a white-noise process, suggesting that
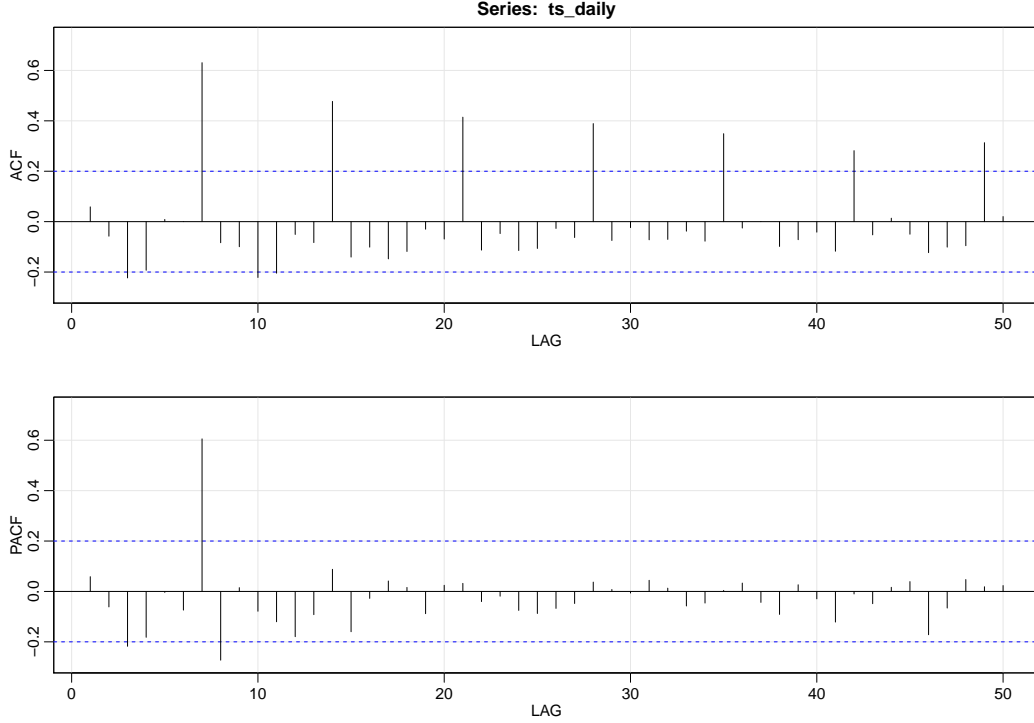
Figure 6: Autocorrelation and partial autocorrelation plots for the natural logarithm of daily sales for Product 1

The top panel (panel a) shows autocorrelations of the natural logarithm of daily sales. The weekly autocorrelations suggest a seasonality effect for a lag of 7 days. The daily autocorrelations do not show a consistent pattern. The bottom panel (panel b) shows the partial autocorrelations. Error bounds are set to $+ 2/\sqrt{n} = +0.2$, corresponding to a 95% confidence interval for both panels. The weekly partial autocorrelations cut off after one lag: A partial autocorrelation that exceeds the lower error bound on day 8, for example, is followed by one that is close to zero on day 9, so that $a = x = 1$. The time-series data shows neither a trend, nor a moving average, and has a seasonality of 7 days, so we fit a SARIMA model of type $ARIMA(1,0,0)(1,0,0)_7$.

the SARIMA model of type $ARIMA(1,0,0)(1,0,0)_7$ provides a good fit to the data.[14]

Although the model fit based on assuming that all variation in daily demand comes from a constant-volatility paradigm, the Normal Q-Q plot shown in panel c of Figure 7 shows kurtosis. This suggests that the demand data has some extreme values. The standardized residuals are shown in Figure 8. Four of 100 residuals were more than three standard deviations from zero, which corresponds to $\lambda = 0.04$ in our jump model. To explore the impact of these extremes on development of the

---

[14] As usually occurs, the Ljung-Box test p values—applied to data cleansed from seasonality and AR1 effects as shown in Figure 7, panel d—are too high to permit rejecting the null hypothesis that residuals are independent for most lags.
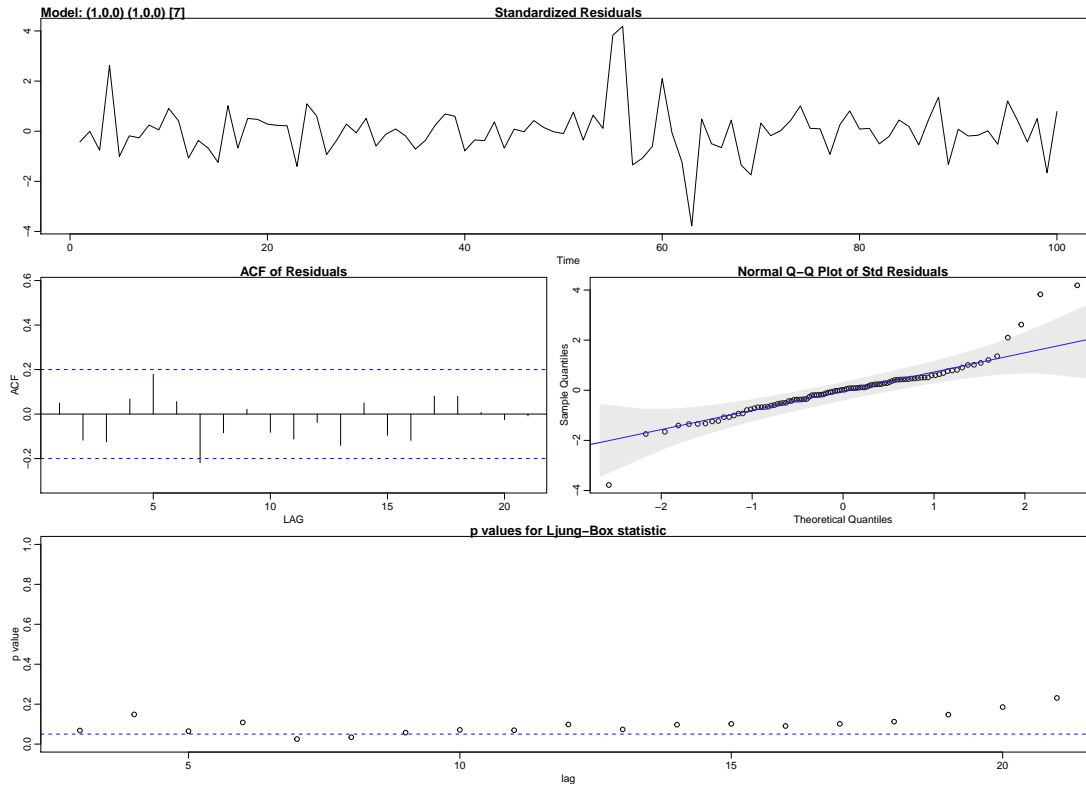
Figure 7: Fitting a time-series model for Product 1

In this figure we analyze the goodness of fit of the SARIMA model of type $ARIMA(1,0,0)(1,0,0)_7$. The top panel (panel a) shows the standardized residuals. The middle-left panel (panel b) shows the autocorrelation values for these residuals. The error bounds given in panel b correspond to a 95% confidence interval: The majority of autocorrelation values lie within these bounds, suggesting that the time-series model is a good fit with the sales data and that it explains all the autocorrelation in the original data. The middle-right panel (panel c) shows the Q-Q plot of residuals, which suggests that residuals follow a normal distribution for the kurtosis to the far right. The bottom panel (panel d) shows the p-values for the Ljung-Box test statistics. The null hypothesis of this test is that the residuals are not correlated for most lags. Panel d shows that the null hypothesis cannot be rejected.

cost-premium frontier, we fit a revised SARIMA model based on the assumption that such outliers represent jumps rather than normal volatility. The revised autocorrelation plot shown in Figure 9 (panel a) shows a cutoff point after 4 lags for daily sales, suggesting an MA4 model. The revised partial autocorrelation plot shown in Figure 9 (panel b) no longer shows a cutoff point for daily sales, and cuts off after one lag for weekly sales. The type of SARIMA model thus changes from $ARIMA(1,0,0)(1,0,0)_7$ to $ARIMA(0,0,4)(1,0,0)_7$ with the new interpretation of the four outliers. The SARIMA model based on the adjusted data performs better than the original SARIMA model. We thus note that the outliers had hidden a strong 4-day moving-average effect. The revised SARIMA
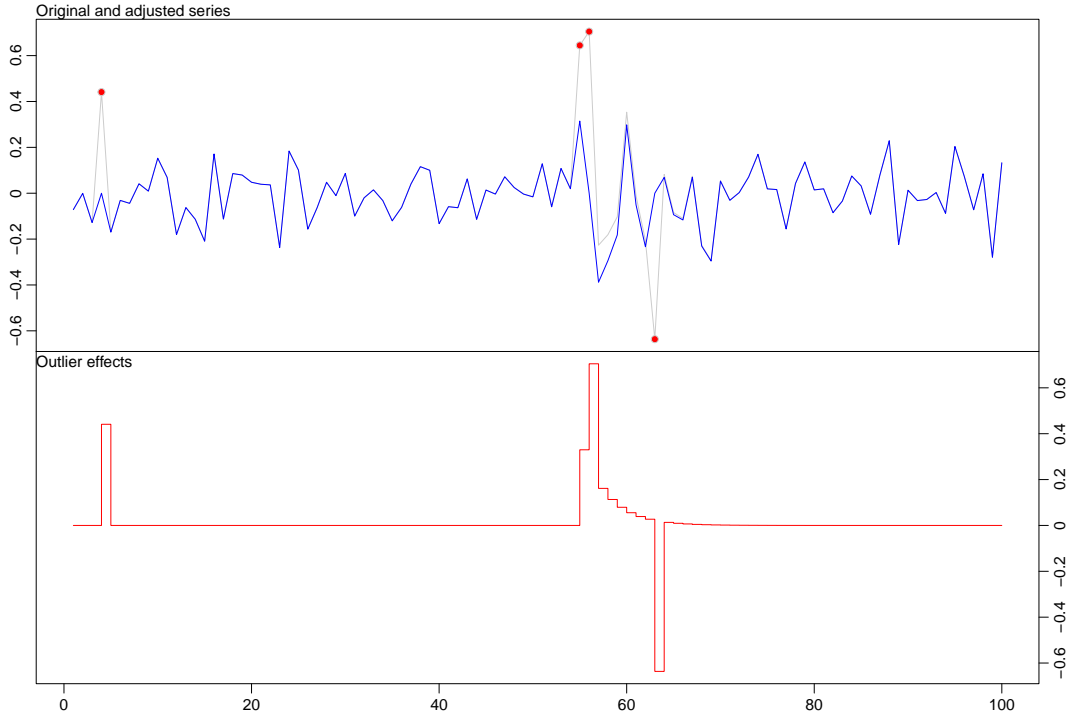
Figure 8: Outliers of the standardized residuals

The top panel (panel a) shows outliers that are more than three standard deviations from zero. The bottom panel (panel b) shows the impact over time of these outliers on the standardized residuals. The outlier that occurs right before day 60, for example, affects subsequent residuals for almost a week. Thus, interpreting an outlier as white noise may well reduce the forecasting ability of the model.

model performed better in forecasting daily demand than did the original model that included the extremes.

After replacing each outlier by the SARIMA forecast for that day, we analyzed the adjusted time-series data as shown in Figure 10. Note that this eliminates the kurtosis. The demand volatility was re-estimated using the adjusted data: Modeling the outliers as jumps rather than as white noise reduced the estimate of volatility from 0.169 to 0.11. The four outliers were used to estimate the parameters of the jump model, with $\tau = 0.28$ and $\varsigma = 0.54$. The parameters for the continuous volatility and jump versions of the SARIMA model are summarized in Table 3.

In Figure 11a, we present the cost-premium frontiers for a jump and a constant-volatility forecasting model using the parameters given in Table 3 assuming a high critical fractile (99.5%). Whether
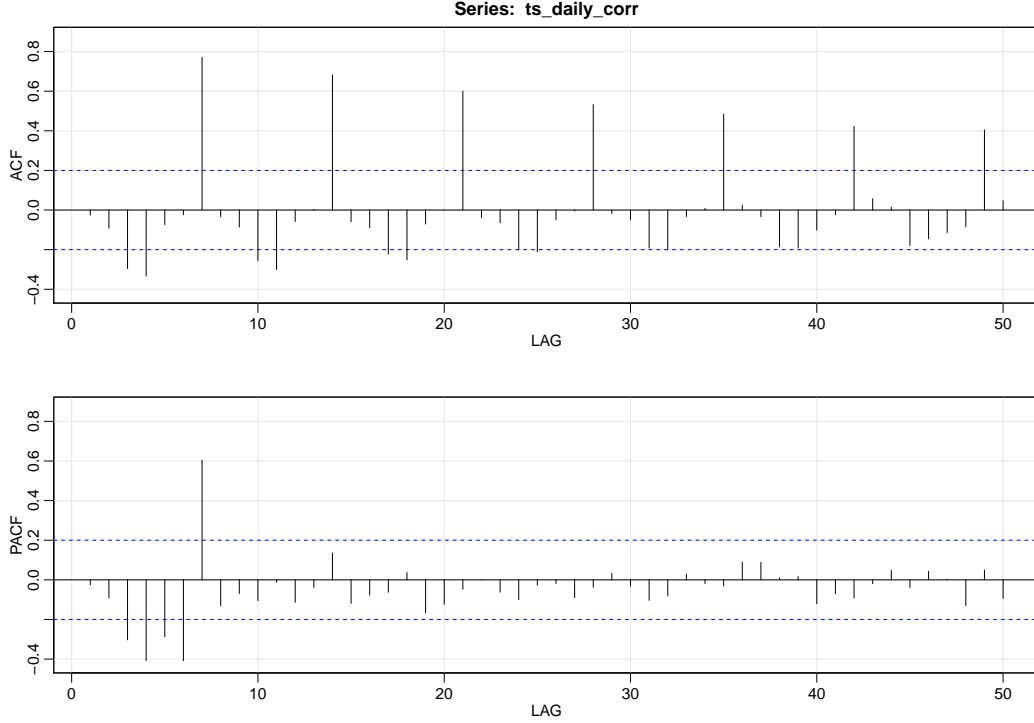
Figure 9: Autocorrelation and partial autocorrelation plots after correcting for the outliers for Product 1

The top panel (panel a) shows autocorrelations of the natural logarithm of daily sales for the adjusted model. The daily autocorrelations suggest seasonality effect of lag 7. The weekly autocorrelations decay continuously, whereas the daily autocorrelations cut off after lag 4. The bottom panel shows partial autocorrelations. The weekly partial autocorrelations cut off after one lag. We also observe that the daily partial autocorrelations do not show a regular pattern. Therefore, we selected a SARIMA model of type $ARIMA(0,0,4)(1,0,0)_7$.

Model	σ	λ	τ	ς	SARIMA parameters
Constant Volatility	0.169	N/A	N/A	N/A	$ARIMA(1,0,0)(1,0,0)_7$
Jump diffusion	0.11	0.04	0.28	0.54	$ARIMA(0,0,4)(1,0,0)_7$

Table 3: Estimation of constant volatility and jump parameters for Product 1

In this table, we summarize the parameters of the constant-volatility and jump models. We used the standardized residuals given in Figure 7 to estimate the σ value of the constant-volatility model. Then, we used the outliers given in Figure 8 and the standardized residuals given in Figure 10 to estimate the parameters of the jump model.

outliers are assumed to represent jumps or white noise has a substantial impact on the value of re-sponsiveness: The premium worth paying to postpone the production enough to obtain full demand information is around 23% in the case of jumps, and just over 5% in the case of a constant-volatility
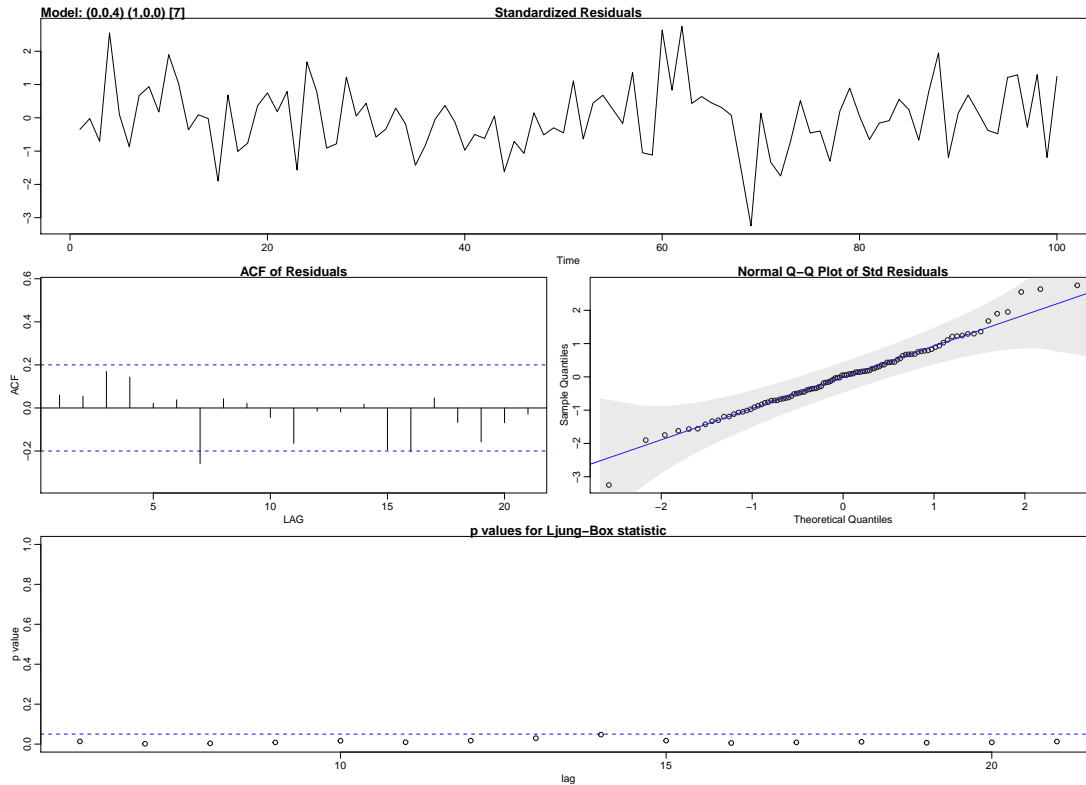
Figure 10: Fitting a time-series model after the outliers were replaced by the forecast

In this figure we analyze the goodness of fit of the SARIMA model of type $\mathrm{ARIMA}(0,0,4)(1,0,0)_7$ to the natural logarithm of daily sales data after the outliers were replaced by their forecasts. The top panel (panel a) shows the standardized residuals. The middle-left panel (panel b) shows the autocorrelation values for these residuals. The error bounds given in panel b correspond to a 95% confidence interval: The majority of autocorrelation values lie within these bounds, suggesting that the time-series model is a good fit with the sales data and that it explains all the autocorrelation in the original data. The middle-right panel (panel c) shows the Q-Q plot of residuals, which suggests that residuals follow a normal distribution. The bottom panel (panel d) shows the p-values for the Ljung-Box test statistics. The null hypothesis of this test is that the residuals are not correlated for most lags. The p values are less than 0.05 so that we can reject the null hypothesis. However, we saw from panel c that the selected model fits well to the data.

assumption.

Figure 11b shows the cost-premium frontiers for a jump and a constant-volatility forecasting model using the parameters given in Table 3 assuming a lower critical fractile (70%). Paradoxically, modeling the outliers as jumps reduces the justified cost premium when the target service level is lower: That the adjusted SARIMA model provides an improved forecast combines with the less-ambitious service level to reduce the premium that a decision maker should be willing to pay to reduce the decision

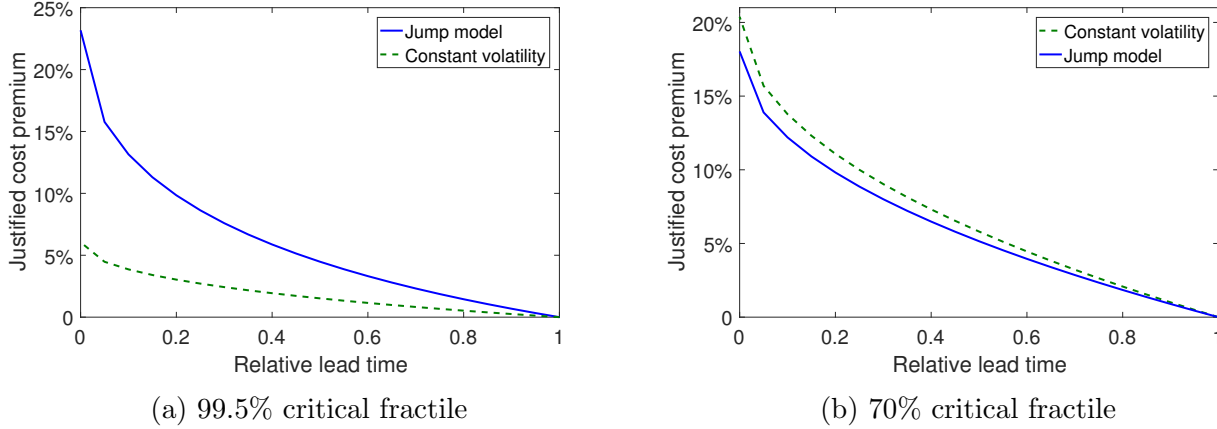(a) 99.5% critical fractile (b) 70% critical fractile

Figure 11: Cost-premium frontiers for Product 1, assuming high and moderate critical fractiles

Panel a shows the cost-premium frontier for a jump and a constant-volatility forecasting model using the parameters given in Table 3 and assuming a high critical fractile of 99.5%, obtained by normalizing the unit price to 100, then setting the cost of the unresponsive supplier to 5% of the unit price, and the residual value to 4.5% of the unit price. The analysis in panel a shows that for this example, Whether outliers are assumed to represent jumps or white noise has a substantial impact on the value of responsiveness: The premium worth paying to postpone the production enough to obtain full demand information is around 23% in the case of jumps, and just over 5% in the case of a constant-volatility assumption. Other combinations of price, cost, and residual that yield the same critical fractile result in a very similar cost-premium frontier. Panel b shows the cost-premium frontier for a jump and a constant-volatility forecasting model using the parameters given in Table 3, but now assuming a lower critical fractile of 70%, obtained by again normalizing the unit price to 100, then setting the cost of the unresponsive supplier to 30% of the unit price, and the residual value to 0. Here we see that the cost-premium frontiers are very close to each other, so treating the outliers as jumps or as white noise makes little difference. Other price, cost, and residual value combinations again result in very similar cost-premium frontiers. In fact, because the adjusted SARIMA model forecasts better than the original model, accounting for jumps actually slightly lowers the cost-premium frontier when the service-level target is modest.

lead time.

This example demonstrates how to estimate jump parameter values from time-series data in a case where a SARIMA model is an appropriate choice. It also shows that the extra step of fitting the jump model provides useful information to consider in estimating the value of responsiveness. This was the first product that we examined, chosen randomly from 4000 products available in our data base, which suggests that extreme values may be routinely encountered.

As with Product 1, we fit a SARIMA model to the natural logarithm of daily sales over 100 days. The autocorrelation plot shown in Figure 12, panel a indicates that daily and weekly autocorrelations decay continuously: The moving-average effect does not appear to be significant. This panel also shows clear weekly seasonality. The partial autocorrelations (Figure 12, panel b) disappear after one lag, consistent with a model of type AR1. Thus, we again fit a SARIMA model of type $ARIMA(1,0,0)(1,0,0)_7$.
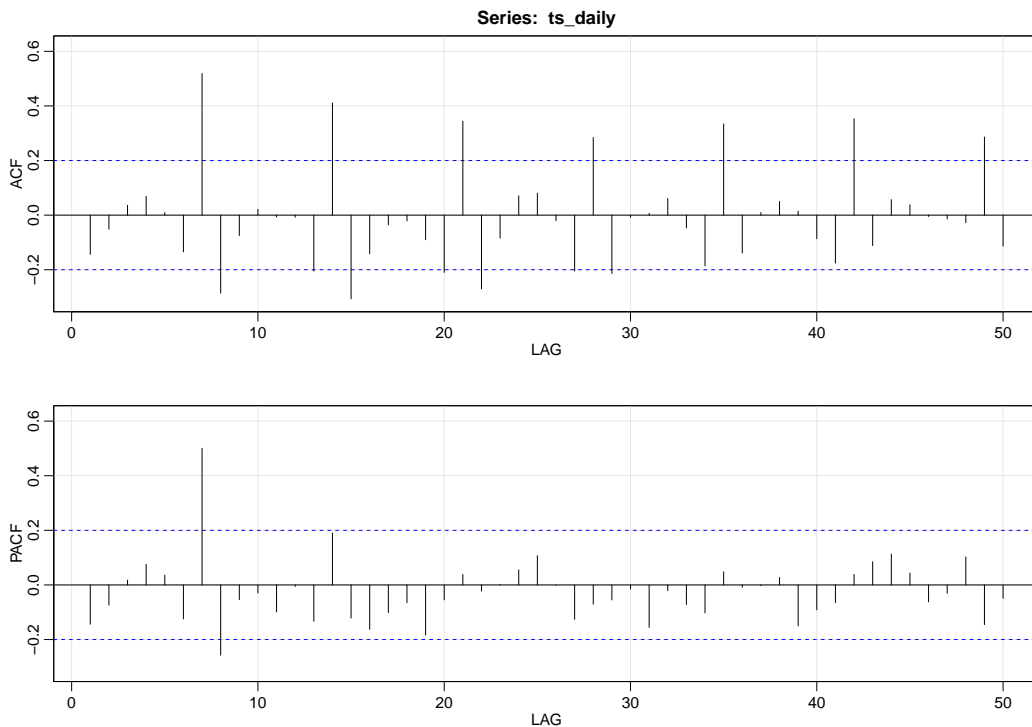

Figure 12: Autocorrelation and partial autocorrelation plots for Buttergipfel

The top panel (panel a) shows autocorrelations of the natural logarithm of daily sales. The weekly autocorrelations suggest a seasonality effect for a lag of 7 days. The daily autocorrelations do not show a consistent pattern. The bottom panel (panel b) shows partial autocorrelations. The weekly partial autocorrelations cut off after one lag. We also observe a strong partial autocorrelation on day 8, and it cuts off on day 9. Therefore, we selected the SARIMA model of type $ARIMA(1,0,0)(1,0,0)_7$. In both panels, error bounds correspond to a 95% confidence interval.

Figure 13, panel b shows that the autocorrelation of residuals is low for different lags.[15] Given

---

[15]As with the first product and as usually occurs, the Ljung-Box test statistics do not let us reject the null hypothesis

that the autocorrelation values for the first 20 lags are also very low the SARIMA model of type $ARIMA(1,0,0)(1,0,0)_7$ provides a good fit to the Product 2 data. All of the residuals lie within three standard deviations of zero and the Q-Q plot does not suggest kurtosis. In contrast to the first product analyzed, this demand data does not suggest jumps.

What happens to the Cost-Premium Frontier when an enthusiastic user is determined to fit a jump model even though the demand data is not suggestive of jumps? We explored this scenario by reducing the threshold for defining outliers from three to 2.52 standard deviations, which was the number of standard deviations that resulted in exactly four outliers, so that $\lambda = 0.04$ as for the first product analyzed. The revised threshold and resulting outliers are shown in Figure 14. We then used these outliers—replaced by the SARIMA forecast—to fit a jump model as described for Product 1. The revised time-series model is developed as shown in Figure 15. The parameter values for both versions of the SARIMA model are given in Table 4. The form of the SARIMA model remained $ARIMA(1,0,0)(1,0,0)_7$ whether the four outliers were treated as jumps or as white noise.

Figure 16a shows the cost-premium frontiers for a jump and a constant-volatility forecasting model using the parameters given in Table 4 assuming a high critical fractile (99.5%). As in Figure 11a, whether outliers are assumed to represent jumps or white noise has a substantial impact on the value of responsiveness. Interpretation thus calls for judgment on the part of decision makers as to whether the outliers should be interpreted as emerging from a constant-volatility process or as evidence of a jump when the critical fractile is very high.

Model	σ	λ	τ	ς	SARIMA parameters
Constant volatility	0.123	N/A	N/A	N/A	$ARIMA(1,0,0)(1,0,0)_7$
Jump diffusion	0.111	0.04	0.14	0.248	$ARIMA(1,0,0)(1,0,0)_7$

Table 4: Estimation of constant-volatility and jump parameters for Product 2

In this table, we summarize the parameters of the constant-volatility and jump models. We used the standardized residuals given in Figure 13 to estimate the σ value of the constant-volatility model. Then, we used the standardized residuals given in Figure 15 and the outliers given in Figure 14 to estimate the parameters of the jump model.

---

that the residuals are not correlated for most lags.

In Figure 16b, we show the cost-premium frontiers for a jump and a constant-volatility forecasting model using the parameters given in Table 4 assuming a lower critical fractile (70%). The cost-premium frontiers are very close to each other.
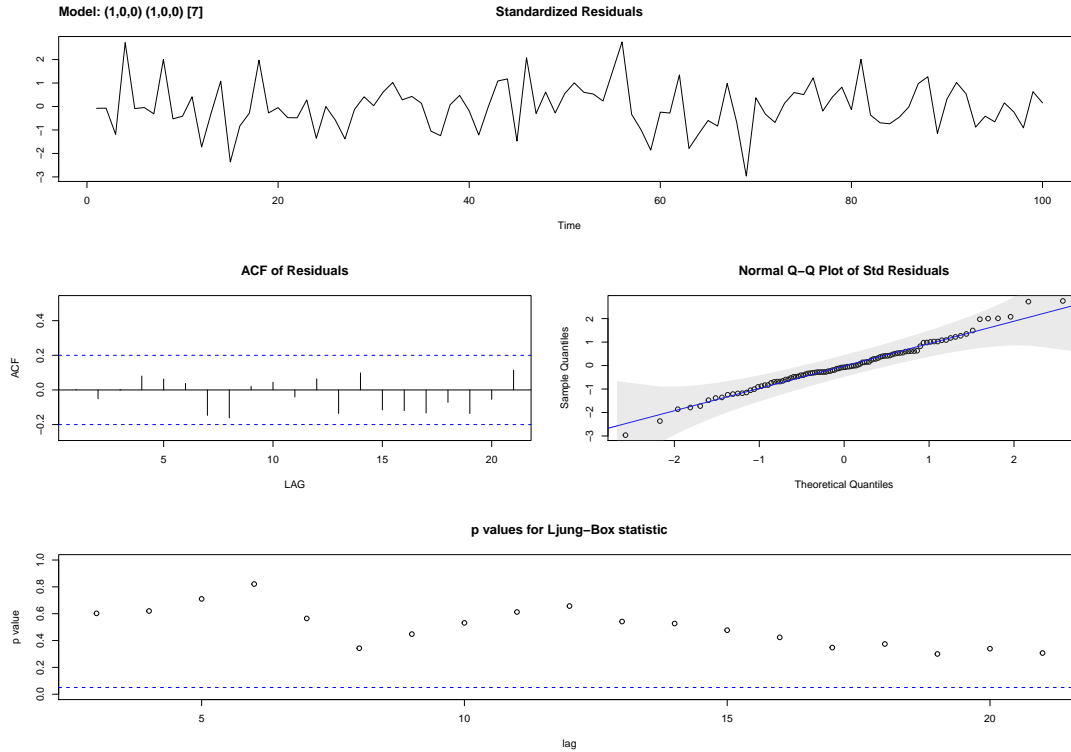

Figure 13: Fitting a time-series model to daily sales of Product 2

In this figure we analyze the goodness of fit of the SARIMA model of type $ARIMA(1, 0, 0)(1, 0, 0)_7$. The top panel (panel a) shows the standardized residuals. The middle-left panel (panel b) shows the autocorrelation values for these residuals. The error bounds given in panel b correspond to a 95% confidence interval: All of the autocorrelations lie within these bounds, suggesting that the time-series model is a good fit with the sales data and that it explains all the autocorrelation in the original data. The middle-right panel (panel c) shows the Q-Q plot of residuals, which suggests that residuals follow a normal distribution. The bottom panel (panel d) shows the p-values for the Ljung-Box test statistics. The null hypothesis of this test is that the residuals are independent for most lags. The p values are too high to allow us to reject the null hypothesis.
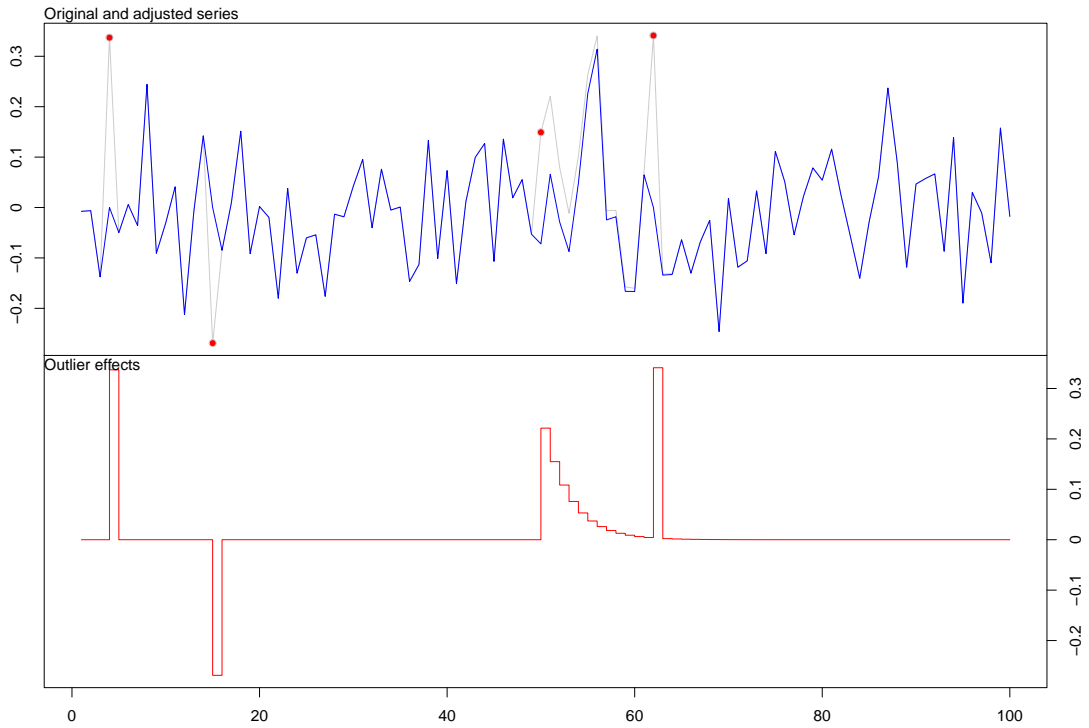
Figure 14: Outliers of the standardized residuals for Product 2

The top panel (panel a) shows outliers that are more than 2.52 standard deviations from zero. The bottom panel (panel b) shows the impact over time of these outliers on the standardized residuals. The outlier that occurs on day 50, for example, slightly affects subsequent residuals for almost a week.
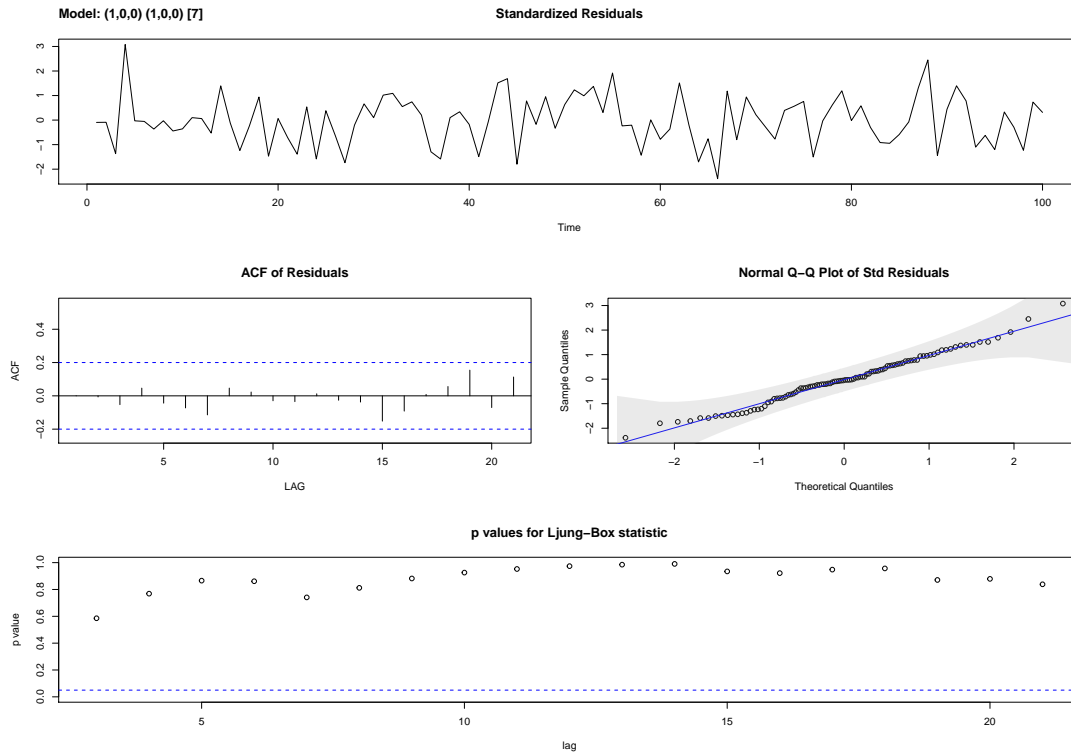
Figure 15: Fitting a time-series model after the outliers were replaced by the forecast for Product 2

In this figure we analyze the goodness of fit of the SARIMA model of type $\text{ARIMA}(1,0,0)(1,0,0)_7$ to the natural logarithm of daily sales data after the outliers were replaced by their forecasts. The top panel (panel a) shows the standardized residuals. The middle-left panel (panel b) shows the autocorrelation values for these residuals. The error bounds given in panel b correspond to a 95% confidence interval: All of autocorrelation values lie within these bounds, suggesting that the time-series model is a good fit with the sales data and that it explains all the autocorrelation in the original data. The middle-right panel (panel c) shows the Q-Q plot of residuals, which suggests that residuals follow a normal distribution. The bottom panel (panel d) shows the p-values for the Ljung-Box test statistics. The null hypothesis of this test is that the residuals are not correlated for most lags. The p values are too high to allow us to reject the null hypothesis. The results are not significantly different from those given in Figure 13.

(a) 99.5% critical fractile   (b) 70% critical fractile
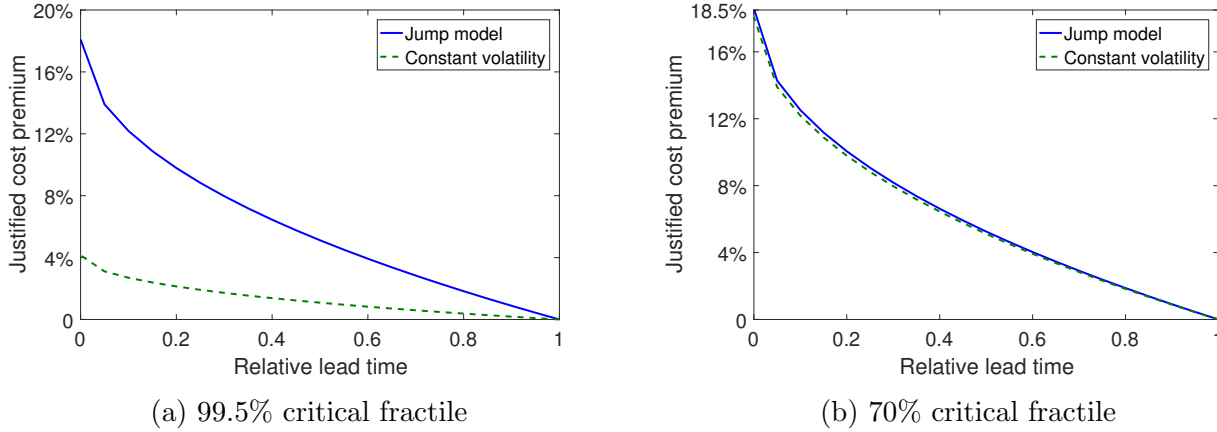
Figure 16: Cost-premium frontiers for Product 2

In panel A, we plot the cost-premium frontier for a jump and a constant-volatility forecasting model using the parameters given in Table 4 assuming a high critical fractile (99.5%). We normalize the unit price to 100. The cost of the unresponsive supplier is 5% of the unit price, and the residual value is 4.5% of the unit price. Whether outliers are assumed to represent jumps or white noise has a substantial impact on the value of responsiveness: The premium worth paying to postpone the production enough to obtain full demand information is around 18% in the case of jumps, and just around 4% in the case of a constant-volatility assumption. Other combinations of price, cost, and residual that yield the same critical fractile result in a very similar cost-premium frontier. In panel B, we plot the cost-premium frontier for a jump and a constant-volatility forecasting model using the parameters given in Table 4 assuming a lower critical fractile (70%). The unit price is normalized to 100, the cost of the unresponsive supplier is 30% of the unit price, and the residual value is 0. We see that the cost-premium frontiers are very close to each other. Other price, cost, and residual value combinations again result in very similar cost-premium frontiers.

7. References

Albuquerque, P., Bronnenberg, B. J., 2012. Measuring the impact of negative demand shocks on car dealer networks. Marketing Science 31 (1), 4–23.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. The Journal of Finance 52 (5), 2003–2049.

Bakshi, G., Madan, D., 2000. Spanning and derivative-security valuation. Journal of Financial Economics 55 (2), 205–238.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81 (3), 637–654.

Carr, P., Madan, D., 1999. Option valuation using the fast Fourier transform. Journal of Computational Finance 2 (4), 61–73.

Cattani, K. D., Dahan, E., Schmidt, G. M., 2008. Tailored capacity: Speculative and reactive fabrication of fashion goods. International Journal of Production Economics 114 (2), 416–430.

de Treville, S., Bicer, I., Chavez-Demoulin, V., Hagspiel, V., Schurhoff, N., Tasserit, C., Wager, S., 2014a. Valuing lead time. Journal of Operations Management 32 (6), 337–346.

de Treville, S., Schurhoff, N., Trigeorgis, L., Avanzi, B., 2014b. Optimal sourcing and lead-time reduction under evolutionary demand risk. Production and Operations Management 23 (12), 2103–2117.

de Treville, S., Trigeorgis, L., 2010. It may be cheaper to manufacture at home. Harvard Business Review 88 (10).

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

Gatheral, J., 2006. The Volatility Surface: A Practitioner's Guide. John Wiley & Sons, Hoboken, NJ.

Graves, S. C., Parsons, J. C., 2005. Using a newsvendor model for inventory planning of NFL replica jerseys. In: Proc. MSOM Conference.

Hammond, J., Raman, A., 1995. Sport Obermeyer Ltd., case 695-002. Harvard Business Publishing.

Hausman, W. H., 1969. Sequential decision problems: A model to exploit existing forecasters. Management Science 16 (2), B93–B111.

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

Iyer, A. V., Bergen, M. E., 1997. Quick response in manufacturer-retailer channels. Management Science 43 (4), 559–570.

Jarrow, R., Rudd, A., 1982. Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics 10 (3), 347–369.

Kesavan, S., Kushwaha, T., 2014. Differences in retail inventory investment behavior during macroeconomic shocks: Role of service level. Production and Operations Management 23 (12), 2118–2136.

Kesavan, S., Kushwaha, T., Gaur, V., 2016. Do high and low inventory turnover retailers respond differently to demand shocks? Manufacturing & Service Operations Management 18 (2), 198–215.

Lee, R. W., 2004. Option pricing by transform methods: extensions, unification and error control. Journal of Computational Finance 7 (3), 51–86.

Lorenzoni, G., 2009. A theory of demand shocks. The American Economic Review 99 (5), 2050–2084.

Lutze, H., Özer, Ö., 2008. Promised lead-time contracts under asymmetric information. Operations Research 56 (4), 898–915.

Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3 (1), 125–144.

Milner, J. M., Kouvelis, P., 2005. Order quantity and timing flexibility in supply chains: The role of demand characteristics. Management Science 51 (6), 970–985.

Oh, S., Özer, Ö., 2013. Mechanism design for capacity planning under dynamic evolutions of asymmetric demand forecasts. Management Science 59 (4), 987–1007.

Parsons, J. C. W., 2004. Using a newsvendor model for demand planning of NFL replica jerseys. Master's thesis, Massachusetts Institute of Technology.

Parsons, J. C. W., Graves, S. C., 2005. Reebok NFL replica jerseys: A case for postponement, MIT Sloan School of Management, SCM Case.

Shumway, R. H., Stoffer, D. S., 2011. Time Series Analysis and its Applications with R Examples. Springer, NY 10013, USA.

Tokar, T., Aloysius, J., Waller, M., Williams, B., 2014. Bracing for demand shocks: An experimental investigation. Journal of Operations Management 32 (4), 205–216.

Uppala, N. G., 2016. Modeling and optimizing of strategic/tactical production planning problems. Ph.D. thesis, Clemson University.

World Economic Forum, 2012. New models for addressing supply chain and transport risk: An initiative of the risk response network, Verified April 9, 2015. `http://www.accenture.com/us-en/Pages/insight-models-addressing-supply-chain-transport-risk.aspx`.

Zhang, Y., Bradlow, E. T., Small, D. S., 2014. Predicting customer value using clumpiness: From RFM to RFMC. Marketing Science 34 (2), 195–208.