Exploring the nonlinear dynamics of the lost-sales order-up-to policy

Stephen M. Disney[a], Borja Ponte[b] and Xun Wang[c]

[a]Centre for Simulation, Analytics and Modelling, University of Exeter Business School, University of Exeter, UK. Email: *s.m.disney@exeter.ac.uk*,
[b]Department of Business Administration, Polytechnic School of Engineering, University of Oviedo, Spain. Email: *ponteborja@uniovi.es* (corresponding author),
[c]Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University, UK. Email: *wangx46@cardiff.ac.uk*.

ARTICLE HISTORY
Compiled June 29, 2020

ABSTRACT
With most inventory theory investigating linear models, the dynamics of nonlinear inventory systems is not well understood. We explore the dynamics of the order-up-to policy under lost-sales for the case of i.i.d. normally distributed demand and unit lead times. We consider the ideal minimum mean squared error forecast and two alternative scenarios: partial demand observation and dynamic demand forecasting, providing a broad understanding of the operational performance of lost-sales systems. In each scenario, we obtain analytical expressions for the order, inventory, and satisfied demand distributions. This allows us to quantify the Bullwhip and inventory variance amplification ratios as well as the fill rate and the inventory cover. We show the lost sales nonlinearity induces complex behaviors in inventory systems. Interestingly, lost sales smooth supply chain dynamics, significantly affecting the trade-off between service level and average inventory holding. We also reveal the inventory downsides of demand censoring and the production damages induced by dynamic forecasts. We identify a key parameter, the *relative safety margin*, that characterizes the performance of lost-sales systems. We finish by offering some prescriptive results for the optimal safety stock and capacity level in both a retail and a manufacturing lost-sales setting.

KEYWORDS
Inventory control, Supply chain dynamics, Bullwhip effect, Order-up-to policy, Lost sales.

1. Introduction

A research study by IHL (2015) estimated that globally inventory inefficiencies cost retailers $1,106 billion per year, a loss of 7.3% of annual revenue. This includes cost overruns related to both stock-outs (i.e. when customers want to buy the product but it is not available), of approx. $634 billion, and overstocks (i.e. holding more inventory than what is needed), of approx. $472 billion. These are two primary components of what the study calls the 'ghost economy'.

An earlier version of this manuscript was presented on 30 August 2019 in Berlin (Germany) at the 9th IFAC/IFIP/IFORS/IISE/INFORMS Conference on Manufacturing Modelling, Management and Control MIM 2019, where it received a Commended Paper Award. Please refer to Disney, Ponte, and Wang (2019).

In practice, inventory systems can be broadly categorised into *backlog systems* and *lost-sales systems* (Zipkin 2008a,b; Bijvank and Vis 2011; Gayon et al. 2016). In backlog systems, when facing an out-of-stock situation, customers are willing to (or must) wait until the product is available. This is the case in many B2B relationships and make-to-order products (e.g. cars and laptops), and also occurs frequently in online sales, where the customer may not even be aware that a stock-out has occurred. In contrast, customers in lost-sales settings are not willing to postpone the purchase after encountering a stock-out situation; for example, shopping for daily groceries in a local supermarket. Not only do lost sales define the generic retail environment for low-value products, they also apply to other scenarios such as box-office sales (e.g. cinema), face-to-face services (e.g. restaurants), and online sales of fixed-date services (e.g. hotel bookings).

It seems reasonable to argue that settings with lost-sales are more sensitive to stock-outs than settings with backlogs. Lost sales reduce revenue and also provide competitors with valuable opportunities; the same product could be purchased from another store or customers may choose a substitute product (Van Woensel et al. 2007). Becoming aware of the features of the alternative product, customers may then switch their default brand choice, causing long-term damage to market share. Stock-out instances can also affect the future in other ways; for example, an out-of-stock printer forgoes the sale of the original printer and also the future ink sales. Furthermore, out-of-stocks sometimes cause consumers to experience negative emotions (Kim and Lennon 2011), which adversely impacts customer loyalty and ultimately profitability through the 'service-profit chain' (Heskett et al. 1994).

The issue goes further if we take into account the operational consequences of out-of-stocks. These situations often provide managers with an inaccurate picture of their supply chains; under lost sales, actual sales may under-estimate the true demand, and this would lead to inaccurate forecasts (Gruen, Corsten, and Bharadwaj 2002). Furthermore, lost sales can lead to rationing and gaming. Both these phenomena induce the so-called *Bullwhip effect* (Lee, Padmanabhan, and Whang 1997) in supply chains. This effect creates a climate of instability that threatens efficient operation in all kinds of industries (Isaksson and Seifert 2016; Li and Disney 2017; Hu 2019).

This research seeks to provide an analytical understanding of the complex dynamics of supply chains governed by the rules of lost-sales inventory systems. As we will discuss later, the behaviour of such systems has received much less attention in the literature than that of backlog systems, probably due to the complexity of the analysis involved.

In the remainder of this section, we motivate and position our work, posing five research questions that encapsulate our contribution to the literature. Section 2 describes our analytical model and the key performance metrics. In Section 3, we explore the forces that shape the dynamics of the lost-sales order-up-to policy, providing exact expressions for the metrics. We consider three types of demand forecasting, accounting for different lost-sales scenarios in practice. Section 4 presents a numerical study of the long-term operational performance of these systems, offering detailed insights from the analytical results. Section 5 analytically investigates the economics of the lost-sales inventory systems, deriving closed-form expressions for the safety stock and capacity levels that minimise inventory and production costs. Finally, in Section 6, we conclude by answering the research questions posed, reflecting upon the managerial implications, and highlighting future research challenges.

1.1. Motivation: The need to investigate lost-sales systems

Two decades ago, Verbeke, Farris, and Thurik (1998) investigated the reaction of 1,750 customers when facing stock-outs. They found that only 20% of them would return to the same store to buy the same product. In a similar study, Corsten and Gruen (2003) surveyed more than 70,000 customers of retail products worldwide and observed that only 15% of customers facing stock-outs would wait until the item is available again. In contrast, 31% would look for it in another store, 45% would buy a substitute product, and 9% would abandon the purchase. Later, Van Woensel et al. (2007) analysed 3,800 responses to stock-outs in a Dutch grocery retailer, reporting that only 12% of regular customers and 6% of occasional customers would opt to delay a purchase.

These studies reveal that most retail customers facing a stock-out do not postpone the planned purchase. They also illustrate that many practical settings are driven by the principles of lost-sales inventory systems. Indeed, it seems reasonable to hypothesise that customers' tolerance to stock-outs decreases when more purchasing options are available (Dawn and Chowdhury 2011; Essila 2019).

Despite lost sales being practically relevant, with wide-ranging implications for organisations, the majority of inventory theory is based upon backlog systems (Bijvank and Vis 2011). That is, most inventory models assume unmet demand is accumulated in an order book and the product is delivered as soon as inventory becomes available. Indeed, Zipkin (2008a,b) argued that while we have a deep understanding of the behavior of inventory systems with backlogs, our comprehension of the lost-sales system is relatively limited.

The lack of exploration of lost-sales systems can be attributed to their complexity, which resides in their nonlinear nature. The assumption of backlogged demand keeps the system linear, as it provides a physical meaning for negative inventories, enabling the study of the system through analytical techniques (Ponte et al. 2017). In contrast, modelling lost sales requires a non-negative constraint to be placed on inventories, which seriously hampers its mathematical study. Indeed, complex dynamic behaviors emerge in nonlinear systems (Wang, Disney, and Wang 2014), which may even dominate the supply chain response (Nagatani and Helbing 2004).

Under such circumstances, configurations derived from the much better known backlog systems are sometimes used to manage real-world lost-sales scenarios. However, it is a risky practice. Zipkin (2008a) showed that the inventory costs could be inflated by up to 30% by using theory developed from linear backlogging models in nonlinear lost-sales settings. Also, Van Donselaar and Broekmeulen (2013) provided evidence that backorder models dramatically over-estimate the safety stock needed in lost-sales systems. Importantly, Huh et al. (2009) showed that the linear approximation of lost-sales inventory systems is only valid under strictly defined conditions. Overall, to reduce the disproportionate cost of inventory inefficiencies in retailers and other organisations, it is clear that we need a better understanding of lost-sales inventory systems.

1.2. Scope: The nonlinear dynamics of supply chains with lost sales

A noteworthy line of research has analytically investigated the optimality of traditional inventory policies in lost-sales environments. Interestingly, Karlin and Scarf (1958) showed that the optimal reorder quantity with backlogs can be derived from a single number (the sum of inventory-on-hand and on-order), but in the lost-sales systems this quantity is a relatively complex function of the on-hand inventory as well as the timing and quantity of all outstanding orders. Since then, several works have mined

this research vein, searching for (near-)optimal inventory policies and deriving bounds of optimal order quantities under various cost assumptions; including Morton (1969), Johansen (2001), Chen, Ray, and Song (2006), Bijvank and Vis (2012), Johansen (2013), Goldberg et al. (2016), Cardós, Guijarro, and Babiloni (2017) and Godichaud and Amodeo (2019). We refer readers to Bijvank and Vis (2011) for a detailed literature review of replenishment policies for lost-sales systems. They concluded that 'not much is known about an optimal replenishment policy when excess demand is lost' and that 'there is no structure for an easy-to-understand optimal replenishment policy which can be implemented in real-life applications' (Bijvank and Vis 2011, p.10-11).

Our approach differs from these prior works. We aim to investigate the impact of the lost-sales nonlinearity on the dynamic behaviour and stochastic performance of the supply chain under the industrially popular order-up-to (OUT) replenishment policy. In this way, we do not attempt to prove optimality; rather, we are concerned with how the, well-established in industry, OUT policy dynamically responds to the practically frequent lost sales. Furthermore, we not only focus on inventory performance, we also consider its fundamental interactions with efficient production by measuring its demand variability amplification caused by the Bullwhip effect.

This perspective positions our work within the discipline of *supply chain dynamics* (Towill 1991). This is a consolidated operational research domain (see e.g. Dejonckheere et al. 2003; Disney et al. 2006; Chen and Lee 2012; Hu 2019; Dolgui, Ivanov, and Rozhkov 2020) that investigates the time-varying forces among the various supply chain elements (replenishment policies, forecasting methods, lead times, and so on) by measuring the stability of the materials and information flows that collectively define the supply chain behaviour (Goltsos et al. 2019). By doing so, we quantify inventory variability and performance, as well as order variability, which has significant economic implications in supply chains[1] (Disney and Lambrecht 2008).

Specifically, we address the following research questions (RQs):

- *RQ1*: How do the dynamics of lost sales systems compare to the dynamics of the backlogging systems?
- *RQ2*: Can we use the same performance metrics for evaluating both backlogging and lost-sales inventory systems?
- *RQ3*: Under what circumstances can the dynamic behaviour of the nonlinear OUT policy with lost sales be approximated by the linear OUT policy with backlogs?
- *RQ4*: What is the impact of using dynamic forecasts, compared to static forecasts in lost-sales settings?
- *RQ5*: What are the consequences of observing the full demand, compared to observing only the satisfied demand?

The last question emerges from the problem of lost sales being intrinsically related to that of demand censoring (Agrawal and Smith 1996; Besbes and Muharremoglu 2013; Rudi and Drake 2014). That is, as discussed by Lariviere and Porteus (1999), real businesses may, or may not, be able to observe the lost sales. For instance, observable lost sales may be present in internet retailers who track browsing history or in face-to-face services like restaurants. On the other hand, unobservable lost sales may occur in general retail settings (supermarkets or vending machines, for example) with customers who, upon experiencing a stock-out, depart without leaving a trace of

[1]Indeed, meaningful interactions emerge between inventory and order variability, leading to a key concept in this area, the 'variability trade-off' (Disney, Towill, and Van de Velde 2004; Priore et al. 2019).

their disappointment. This may not only have implications on system performance, it could also create a gap between the actual, or theoretical, Bullwhip effect (based on the actual demand) and that perceived by the practitioners (based on the observed demand or sales), as discussed by Chen, Luo, and Shang (2017).

1.3. Contribution and methodological approach

To the best of our knowledge, no prior study in the supply chain dynamics field provides a clear understanding of our research questions. In line with previous discussions, most assume that excess demand is accumulated in an order book, and the product is delivered once it becomes available. While lost sales have been occasionally studied via simulation (e.g. see Potter and Disney 2010; Costas et al. 2015; Lin, Naim, and Spiegler 2019), this research stream has not focused on comprehensively exploring how the dynamics of supply chains are affected by the presence of lost sales[2].

Indeed, due to the ubiquity of lost sales in many practical environments, this has been often flagged as a key area for future research. For example, Holweg and Disney (2005, p.713) noted that 'surprisingly though, there are still a range of unsolved research questions [in supply chain dynamics]'. In this regard, they highlighted the 'nonlinearities found in real-life supply chains', including lost sales, that make mathematical analysis difficult. A decade later, in a contemporary review of the Bullwhip literature, Wang and Disney (2016, p.697) identified six primary research gaps, highlighting the study of nonlinearities as the first to explore 'more realistic supply chain models, such as [those based on] lost sales'.

To answer our research questions, we analytically derive the expressions for the Bullwhip (BW) and inventory variance amplification ratio (IVR) metrics maintained by the OUT policy, two common indicators in the literature that report on order and inventory variability, respectively. Also, we quantify the fill rate and inventory cover, thus providing in-depth understanding of the fundamental inventory trade-off between customer service level and inventory holding requirements. Due to the analytical complexity, our study is limited to the case of independent and identically distributed (i.i.d.) demand and unit lead times. In line with our research questions, we study three forecasting approaches that represent different types of lost-sales systems in practice:

(1) A static approach where the forecast always matches the mean demand. This is the *minimum mean squared error* (MMSE) forecast for the i.i.d. demand that implicitly assumes that lost sales are observed.
(2) A static approach that targets the mean demand, knowing its i.i.d. nature, but under-estimates it due to unobserved lost sales. This is a variant of the above that considers *partial demand observation* (PDO).
(3) A *dynamic demand forecasting* (DDF) approach based on the industrially popular exponential smoothing forecasting mechanism. Despite being sub-optimal, it seems to be a reasonable choice when the true demand process is unknown. Like the MMSE case, we assume the demand is fully observable in this scenario.

Finally, we use the analytical understanding of the operational dynamics induced by lost sales in OUT policies to explore the economics of the inventory system. From

[2]Due to the complexity of the analysis involved, simulation is the most common approach to explore nonlinear supply chains. Previous research efforts have also provided a fair understanding of other nonlinearities via simulation techniques, such as Ponte et al. (2017) for capacity constraints (an upper limit on orders) and Chatfield and Pritchard (2013) for forbidden returns (non-negative orders).

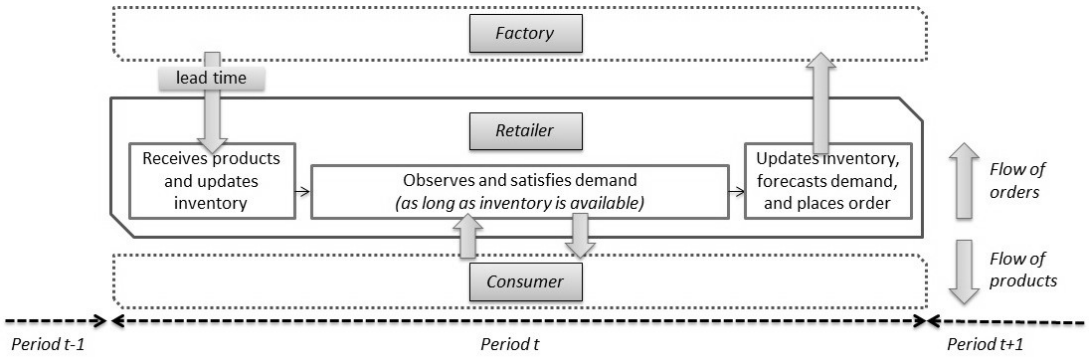


Figure 1.: The structure of, and the sequence of events, in the lost-sales OUT policy.

this viewpoint, we provide closed-form expressions to optimally adjust the safety stock and the production capacity in the MMSE case, which defines the efficient frontier in the lost-sales system under study.

2. The order-up-to inventory system with lost sales

We study a single-product production and distribution system formed by a customer placing demands on a retailer who is supplied by a factory that has enough capacity to always serve the retailer on time. We focus on a single echelon of a supply chain that we generally term the 'retailer' for ease of reference as this node frequently operates under lost-sales scenarios (Verbeke, Farris, and Thurik 1998; Gruen, Corsten, and Bharadwaj 2002; Van Woensel et al. 2007). However, the single echelon could also be a manufacturer. Indeed, we have experience of working with a automotive assembler who dual sourced a single product from two different suppliers. When one supplier was out of stock, demand was diverted to the other supplier. Here, the suppliers experienced lost sales, rather then a backlogging, when they had a stock-out.

Fig. 1 illustrates the sequence of events that governs the discrete-time operation in our model. The sequence of events consists of three stages in each time period t. At the beginning of t, the retailer receives the product from the factory. Then, during the course of t, the retailer satisfies the demand of consumers as on-hand inventory permits. When inventory is not available, the retailer incurs lost sales. At the end of t, the inventory levels are observed, a forecast of future demand is calculated, and finally the retailer releases an order to the factory calculated with the OUT replenishment policy. Table 1 introduces the notation of the main variables, parameters, metrics, and costs we use in our study.

2.1. Mathematical model and key assumptions

Our discrete-time system can be modelled through a set of difference equations. First, the retailer receives the products delivered by the factory at the start of t, r_t, corresponding to the orders, q_t, placed by the retailer in the previous period,

$$r_t = q_{t-1}. \tag{1}$$

This means that the factory is always able to deliver what was ordered in the previous period. That is, production and transportation constraints are not present in our study;

Table 1.: Notation of variables, parameters, metrics, and costs in our lost-sales inventory system.

Notation of variables		***Notation of parameters***	
d_t	customer demand	α	exponential smoothing constant
f_t	demand forecast	γ	demand's coefficient of variation
i_t	on-hand inventory level	δ	safety factor
q_t	order quantity	η	estimate of mean demand
r_t	receipts of new products	μ	mean of the demand
s_t	satisfied demand	σ	standard deviation of the demand
l_t	lost sales	k	capacity in regular working hours
Notation of metrics		***Notation of costs***	
BW	Bullwhip ratio	h	unit inventory holding cost
IVR	inventory variance ratio	p	unit penalty cost for lost sales
β	fill rate	u	unit production cost
τ	inventory cover	w	unit overtime cost

Note: When present, subscripts m, p, and d refer to the MMSE, PDO, and DDF systems, respectively.

moreover, we consider a unit lead time.

The consumer demand, d_t, is assumed to be an i.i.d. random variable, z_t, drawn from a normal distribution with mean μ and variance σ^2,

$$d_t = z_t : z_t \in \mathcal{N}[\mu, \sigma^2]. \tag{2}$$

Let γ be the demand's coefficient of variation, $\gamma = \sigma/\mu$. We adopt the normality assumption as this statistical distribution captures the behaviour of many demand patterns in the real world (Schneeweiss 1974; Disney et al. 2016). For example, normal demand may occur when demand that originates from many independent customers is aggregated into daily or weekly buckets, as the central limit theorem can be evoked. Our i.i.d. demand fluctuates randomly around a constant mean; that is, no trends, seasonal effects, or auto-correlation exists here.

At the end of each period, the on-hand inventory level, i_t, can be calculated as the accumulated difference between receipts and demand. Under lost sales, the inventory is constrained to be non-negative,

$$i_t = [i_{t-1} + r_t - d_t]^+, \tag{3}$$

where $[\cdot]^+ = \max\{\cdot, 0\}$ is the maximum operator; by truncating negative values it is here that we capture the influence of the lost sales. We highlight that i_t=0 indicates that stock-outs have occurred in t (except in the particular case when demand is exactly equal to the available inventory, i.e., d_t=i_{t-1}+r_t). We do not consider any other nonlinear effects, e.g. limited storage capacity or quality loss.

A key variable in our lost-sales system is the demand satisfied by the retailer, s_t. The satisfied demand is the minimum of the available inventory (at the start of t) and the demand,

$$s_t = (i_{t-1} + r_t) \wedge d_t, \tag{4}$$

where the operator $\wedge$ means 'take the minimum of both sides'.

We model three different forecasts in our study. The base forecast assumes the retailer can observe the full consumer demand, despite only satisfying a portion of it,

and recognises its i.i.d. nature. Thus, the retailer forecasts by minimising the mean squared error between the future demand and its forecast (Disney et al. 2016; Ponte et al. 2017), by using

$$f_t = \mu. \tag{5}$$

This represents an ideal scenario that we refer to as the MMSE system.

We consider two extensions to the base system. The first assumes that the retailer is not able to observe the lost sales, but only the satisfied demand, s_t. However, the retailer is aware of the i.i.d. nature of customer demand. Therefore, we assume the retailer employs a static forecast that tends to under-estimate the mean demand,

$$f_t = \eta : \eta < \mu. \tag{6}$$

We term this the PDO system. The only difference with the base system stems from the modelling of partial *vs.* full demand observation, which allows us to provide insights into the implications of demand censoring.

The second extension is aimed at investigating the impact of 'incorrectly' using exponential smoothing to forecast the i.i.d. demand via

$$f_t = \alpha d_t + (1 - \alpha) f_{t-1}, \tag{7}$$

where $0 \leq \alpha \leq 1$ is the exponential smoothing constant. This may occur if the retailer does not know the true demand distribution and wishes to use an 'adaptive' forecasting approach. This is the DDF system. We assume again that demand is fully observable, so that the only difference with the base MMSE system emerges from the static *vs.* dynamic forecasting approach.

Finally, orders are generated with the industrially prevalent OUT replenishment rule (Lalwani, Disney, and Towill 2006). For our lead time, the order quantity is the sum of the demand forecast and the discrepancy between the target and actual on-hand inventory. The target inventory represents the safety stock. Following common practice (Lin et al. 2017), the target inventory is the product of a decision parameter, the *safety factor* δ, and the forecasted demand, f_t. Thus, the order quantity is calculated with

$$q_t = f_t + (\delta f_t - i_t) = (1 + \delta) f_t - i_t. \tag{8}$$

Note, $(1 + \delta) f_t$ represents the OUT level (that is, the desired position of the on-hand inventory at the start of t), while i_t accounts for the actual inventory level.

2.2. Operational performance indicators

To measure performance, we employ two common approaches in the literature due to their industrial relevance. First, we analyse operational variability, a key source of inefficiencies in production and distribution systems. Indeed, popular management paradigms, like Lean Production, focus on mitigating such variability by tackling its sources; see Ohno (1988) and Womack and Jones (1997)[3]. Inventories and orders are two major, interrelated ways through which variability manifests itself in companies

[3]We highlight the Lean wastes of *mura* and *muri* are closely related to the Bullwhip effect.

(Disney and Lambrecht 2008). We measure order variability via the Bullwhip ratio,

$$BW = \frac{\mathbb{V}[q_t]}{\mathbb{V}[d_t]}, \tag{9}$$

where $\mathbb{V}[\cdot]$ is the variance operator. BW is directly related to capacity-related production costs (Disney and Lambrecht 2008).

To explore inventory variability, we employ the inventory variance ratio, IVR metric,

$$IVR = \frac{\mathbb{V}[i_t]}{\mathbb{V}[d_t]}. \tag{10}$$

In backlogging systems, IVR is known as the Net Stock Amplification ratio (NSAmp), and it is a very useful indicator as it is directly related to inventory costs (Disney and Lambrecht 2008). The term 'net stock' is used as the inventory can be both positive (indicating stock holding) and negative (indicating a backlog). A key trade-off for supply chain managers exists between BW and NSAmp under backlogging, as it is often possible to decrease one at the expense of increasing the other (Disney, Towill, and Van de Velde 2004). However, negative inventory (backlogs) do not exist in lost-sales settings; hence the change in moniker.

Second, we analyse inventory performance by looking at customer service and holding investment. This is another key business concern that may be expressed in the form of a trade-off; the service levels can be easily improved at the cost of holding extra inventory. We measure customer service through the fill rate, β. This is a popular metric in the fast-moving consumer goods industry, representing the proportion of demand satisfied immediately from stock[4] (Disney et al. 2015). The fill rate is the ratio of the mean positive satisfied demand to the mean positive demand (Sobel 2004; Disney et al. 2015),

$$\beta = \frac{\mathbb{E}[(s_t)^+]}{\mathbb{E}[(d_t)^+]}. \tag{11}$$

Here, $\mathbb{E}[\cdot]$ is the expectation operator. Note, the double-accounting-of-backlogs problem (Cachon and Terwiesch 2006; Disney et al. 2015), which may distort the fill rate measure in linear inventory systems, does not occur under the lost-sales condition.

In order to measure the inventory investment required to meet consumer demand, we consider the inventory cover, τ, which is the ratio of the mean inventory to the mean demand[5],

$$\tau = \frac{\mathbb{E}[i_t]}{\mathbb{E}[d_t]}. \tag{12}$$

[4]In lost-sales systems the definition simplifies to the proportion of customer demand that is satisfied, given that demand not satisfied immediately is lost.

[5]Note, in the linear backlog system it would be necessary to use $\mathbb{E}[(i_t)^+]$ in the numerator, so that τ becomes indicative of the mean inventory held by the retailer.

3. Analytical study of the lost-sales supply chain operation

We aim to provide closed-form expressions of the four operational performance metrics for our three lost-sales inventory systems. We first look to the statistical distributions of the on-hand inventory, satisfied demand, and orders, as these *state variables* determine the performance, see (9)-(12). For all three systems, the following relations can be obtained from (1)-(4) and (8):

$$i_t = [i_{t-1} + q_{t-1} - d_t]^+, \tag{13}$$
$$s_t = (i_{t-1} + q_{t-1}) \wedge d_t, \tag{14}$$
$$q_t = (1+\delta) f_t - i_t. \tag{15}$$

Before performing our analysis, we first make some introductory remarks on the statistical techniques we employ in our study.

3.1. Preliminary matters

We make extensive use of the normal distribution in our analysis. Let X be a random variable with support over $(-\infty, \infty)$ and realizations x. Let $\phi[x]$ and $\Phi[x]$ be the probability density function (pdf) and cumulative distribution function (cdf) of the standard normal distribution, $\mathcal{N}[0,1]$, with zero mean and unit variance. We may scale the pdf of a standard normal distribution, $\phi[x]$ into a specific normal distribution with mean μ and variance σ^2, $\mathcal{N}[\mu, \sigma^2]$, via

$$\phi[x|\mu,\sigma] = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right]. \tag{16}$$

Translations by a constant y, $X + y$, are incorporated into the mean,

$$\phi[x|\mu+y,\sigma] = \frac{1}{\sigma}\phi\left[\frac{x-(\mu+y)}{\sigma}\right]. \tag{17}$$

Reflections about the origin, $-X$, are dealt with via

$$\phi[x|(-\mu),\sigma] = \frac{1}{\sigma}\phi\left[\frac{x+\mu}{\sigma}\right]. \tag{18}$$

As discussed before, we make use of the maximum operator $[x]^+ = \max\{0, x\}$ to truncate negative values in our analysis. The pdf of $[X]^+$, is given by

$$\phi[x|\mu,\sigma]^+ = \theta[x]\frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] + \Delta[x]\int_{-\infty}^{0}\phi[x|\mu,\sigma^2]dx. \tag{19}$$

Here $\theta[x] = \{1, \text{ if } x \geq 0;\ 0, \text{otherwise}\}$ is the Unit Step function and $\Delta[x] = \{1, \text{ if } x = 0;\ 0, \text{otherwise}\}$ is the Dirac Delta function[6].

[6]Note, while we call this operation 'truncation of the pdf', our defined pdf is different to the pdf of the regular 'truncated normal distribution' (Wikipedia 2020).

The distribution of the minimum of two normally distributed, correlated random variables also occurs in our analysis. Cain (1994) and Nadarajah and Kotz (2010), building upon the work of Basu and Ghosh (1978) and Nagaraja and Mohan (1982), provided the following expression for the pdf of the minimum of two normally distributed, correlated random variables, a and b:

$$\phi_{a,b}[x] = \Psi[x] + \xi[x], \tag{20}$$

where

$$\Psi[x] = \frac{1}{\sigma_a}\phi\left[\frac{x-\mu_a}{\sigma_a}\right]\Phi\left[\frac{\rho(x-\mu_a)}{\sigma_a\sqrt{1-\rho^2}} - \frac{x-\mu_b}{\sigma_b^2}\right], \tag{21}$$

and

$$\xi[x] = \frac{1}{\sigma_b}\phi\left[\frac{x-\mu_b}{\sigma_b}\right]\Phi\left[\frac{\rho(x-\mu_b)}{\sigma_b\sqrt{1-\rho^2}} - \frac{x-\mu_a}{\sigma_a^2}\right]. \tag{22}$$

Here, $\{\mu_a, \mu_b\}$ is the mean of $\{a, b\}$ and $\{\sigma_a, \sigma_b\}$ the standard deviation of $\{a, b\}$, and ρ is the Pearson correlation coefficient between a and b, $-1 \leq \rho = \mathrm{cov}(a, b)/(\sigma_a\sigma_b) \leq 1$.

3.2. Lost sales with minimum mean squared error forecasts (MMSE)

3.2.1. Fundamental relations in the MMSE system

Using (5), an important property of this system emerges from (15); for all periods,

$$q_t + i_t = (1+\delta)\mu. \tag{23}$$

Substituting (23) into (13) and (14) leads to

$$i_t = [(1+\delta)\mu - d_t]^+, \tag{24}$$

and

$$s_t = (1+\delta)\mu \wedge d_t. \tag{25}$$

Using the relation $[a-b]^+ = a - (a \wedge b)$ in (24) reveals $i_t = (1+\delta)\mu - ((1+\delta)\mu \wedge d_t)$, which together with (23) leads to

$$q_t = s_t. \tag{26}$$

Importantly, the lost-sales OUT model here results in the retailer ordering every period exactly what they sold during the period, $q_t = s_t$. Notice the difference with the linear backlog model, where the retailer orders exactly what was demanded during the period, $q_t = d_t$, a 'pass-on-orders' strategy (Disney et al. 2016; Ponte et al. 2017).

3.2.2. Distribution of the MMSE state variables

The fundamental relations (24)-(26) show the state variables of the first lost-sales system as functions of constants and the demand. We now discuss their statistical distributions, which will allow us to derive the analytical expressions needed for our performance metrics.

Inspection of (24) reveals the inventory is a translated, scaled, and truncated demand distribution, see §2.1. These operations leads to the following pdf,

$$\phi_{i,m}[x] = \frac{\theta[x]}{\sigma}\phi\left[\frac{x}{\sigma} - \lambda_m\right] + \Delta[x]\Phi[-\lambda_m], \quad \text{where } \lambda_m = \frac{\delta\mu}{\sigma} = \frac{\delta}{\gamma}. \tag{27}$$

To derive (27) we exploit (16)-(19). The translation and scaling is dealt with by changing the mean and standard deviation of the normal distribution, while the Unit Step function attends to the truncation. The pdf at $x = 0$ is captured by the product of the Dirac Delta function and the cdf at $x = 0$. Lastly, we define the *relative safety margin* as the ratio of the safety factor to the coefficient of variation, $\lambda_m = \delta/\gamma$, to simplify our expressions[7].

Through a similar procedure, we can obtain the pdf of the satisfied demand and orders in this lost-sales system from (25), taking (26) into account. This is given by

$$\phi_{q,m}[x] = \phi_{s,m}[x] = \frac{1}{\sigma}\theta\left[\lambda_m - \frac{x-\mu}{\sigma}\right]\phi\left[\frac{x-\mu}{\sigma}\right] + \Delta\left[\lambda_m - \frac{x-\mu}{\sigma}\right]\Phi[-\lambda_m]. \tag{28}$$

Note, in this system, the pdf of the orders (and the satisfied demand) is a translated reflection of the pdf of the inventory, as per (23). This reveals an important insight: the variance of the orders is equal to the variance of the inventory levels, and

$$BW_m = IVR_m. \tag{29}$$

3.2.3. Deriving the MMSE performance metrics

From the distribution of the state variables, we can derive the expressions of the four operational metrics. First, we investigate the mean of the variables to consider the trade-off between the fill rate, β, and the inventory cover, τ. The inventory pdf can be used to obtain the inventory cover,

$$\tau_m = \frac{\mathbb{E}[i_t]}{\mu} = \frac{1}{\mu}\int_0^\infty x\phi_{i,m}[x]dx = \gamma(\phi\left[\lambda_m\right] + \lambda_m\Phi\left[\lambda_m\right]). \tag{30}$$

Denoting the pdf of the demand by $\phi_d[x]$, the fill rate is given by

$$\beta_m = \frac{\mathbb{E}[(s_t)^+]}{\mathbb{E}[(d_t)^+]} = \frac{\int_0^\infty x\phi_{s,m}[x]dx}{\int_0^\infty x\phi_d[x]dx} = \theta[1+\gamma\lambda_m]\left(1 - \frac{\phi[\lambda_m] + \lambda_m(\Phi[\lambda_m] - 1)}{\phi\left[\gamma^{-1}\right] + \gamma^{-1}\Phi\left[\gamma^{-1}\right]}\right). \tag{31}$$

[7] As *safety stock*$= \delta\mu$, λ_m can be interpreted as the safety stock divided by the standard deviation of demand.

Second, consider the variances. The variance of the inventories is given by

$$\begin{aligned}\mathbb{V}[i_t] = \sigma^2_{i,m} =& \int_0^\infty (x-\mu\tau_m)^2 \phi_i[x]dx \\ =& \sigma^2 \left(\left(\lambda_m^2+1\right) \Phi[\lambda_m] - (\lambda_m \Phi[\lambda_m] + \phi[\lambda_m])^2 + \lambda_m \phi[\lambda_m] \right). \end{aligned} \tag{32}$$

As discussed, the variances of the three state variables are identical in this system, $\mathbb{V}[q_t] = \mathbb{V}[s_t] = \mathbb{V}[i_t] = \sigma^2_{q,m} = \sigma^2_{s,m} = \sigma^2_{i,m}$. Thus, (32) allows us to obtain the IVR metric, and by (29) also the BW metric, in the MMSE system,

$$BW_m = IVR_m = \left(\lambda_m^2+1\right) \Phi[\lambda_m] - (\lambda_m \Phi[\lambda_m] + \phi[\lambda_m])^2 + \lambda_m \phi[\lambda_m]. \tag{33}$$

At this point, it is convenient to note that the equivalent MMSE linear system, with backlogs rather than lost sales, has $BW_m = IVR_m = 1$ (Disney et al. 2006). Note, this can be easily obtained by substituting[8] $\Phi[\lambda_m] = 1$ and $\phi[\lambda_m] = 0$ into (33).

3.3. Lost sales with partial demand observation (PDO)

We now assume the retailer is tempted to use a constant forecast of demand (knowing the i.i.d. nature of the demand) but cannot observe the lost sales. In such cases, the estimated mean demand, η, would tend to be less than the true mean demand, μ.

3.3.1. Fundamental relations in the PDO system

Using the forecast under the PDO scenario, (6), in (15) leads to

$$q_t + i_t = (1+\delta)\eta, \tag{34}$$

which highlights again in all periods the sum of the orders and inventory is a constant, $(1+\delta)\eta$. Through the same procedure as before, the following fundamental relationships for the state variables in this lost-sales supply chain can be found:

$$i_t = [(1+\delta)\eta - d_t]^+, \tag{35}$$

$$s_t = (1+\delta)\eta \wedge d_t, \tag{36}$$

$$q_t = s_t. \tag{37}$$

3.3.2. Distribution of the PDO state variables

Employing the same procedure used to study the MMSE system, the following expressions for the pdf of the inventory, satisfied demand, and orders can be derived from (35)-(37):

$$\phi_{i,p}[x] = \frac{\theta[x]}{\sigma}\phi\left[\frac{x}{\sigma} - \lambda_p\right] + \Delta[x]\Phi[-\lambda_p], \quad \text{where } \lambda_p = \frac{(1+\delta)\eta - \mu}{\sigma}, \tag{38}$$

[8]When δ is sufficiently high, the relations $\Phi[\lambda_m] = 1$ and $\phi[\lambda_m] = 0$ hold, lost sales do not occur and the nonlinear system behaves as the linear system would.

and

$$\phi_{q,p}[x] = \phi_{s,p}[x] = \frac{1}{\sigma}\theta\left[\lambda_p - \frac{x-\mu}{\sigma}\right]\phi\left[\frac{x-\mu}{\sigma}\right] + \Delta\left[\lambda_p - \frac{x-\mu}{\sigma}\right]\Phi[-\lambda_p]. \tag{39}$$

Notice the similarity in the form of the distributions of the MMSE and PDO systems; compare (27)-(28) to (38)-(39). They only differ in the definition of the *relative safety margin*, now called λ_p. As $(1+\delta)\eta$ is the initial position of the on-hand stock in every period, λ_p represents the protection level, $(1+\delta)\eta - \mu$, against shortages in relative terms to the standard deviation of demand σ. From λ_m and λ_p, a PDO system with δ_p behaves as a MMSE system with

$$\tilde{\delta}_m = \frac{\eta}{\mu}(1+\delta_p) - 1. \tag{40}$$

$\tilde{\delta}_m$ is an insightful parameter; it shows the equivalent safety factor if we were using the MMSE system (with δ_p being the safety factor used in the PDO system). Note, due to $\eta < \mu$, $\tilde{\delta}_m < \delta_p$; that is, the protection level is lower than δ_p suggests it should be. This shows that the target balance between fill rate and inventory cover in lost-sales inventory systems is strongly affected by the unobserved demand[9].

3.3.3. Deriving the PDO performance metrics

As the pdf's of the state variables differ only by the definition of λ, the performance metrics for the PDO case can be readily obtained from the MMSE case by making the substitution $\lambda_m \to \lambda_p$ (indicated by $\cdot|_{\lambda_m \to \lambda_p}$). The inventory cover is

$$\tau_p = \tau_m|_{\lambda_m \to \lambda_p} = \gamma(\phi\left[\lambda_p\right] + \lambda_p \Phi\left[\lambda_p\right]). \tag{41}$$

The fill rate is given by

$$\beta_p = \beta_m|_{\lambda_m \to \lambda_p} = \theta[1+\gamma\lambda_p]\left(1 - \frac{\phi[\lambda_p] + \lambda_p(\Phi[\lambda_p]-1)}{\phi\left[\gamma^{-1}\right] + \gamma^{-1}\Phi\left[\gamma^{-1}\right]}\right). \tag{42}$$

IVR and BW are also identical in the PDO system. These metrics are given by

$$\begin{aligned} BW_p = IVR_p &= BW_m|_{\lambda_m \to \lambda_p} = IVR_m|_{\lambda_m \to \lambda_p} \\ &= \left(\lambda_p^2 + 1\right)\Phi[\lambda_p] - \left(\lambda_p \Phi[\lambda_p] + \phi[\lambda_p]\right)^2 + \lambda_p \phi[\lambda_p]. \end{aligned} \tag{43}$$

Again, the equivalent linear system, characterised by $\Phi[\lambda_m] = 1$ and $\phi[\lambda_m] = 0$, results in $BW_p = IVR_p = 1$. This might seem surprising, but this is how the linear system reacts; the linear system is indifferent to mean values.

3.4. Lost sales with dynamic demand forecasts (DDF)

We now study a variant of the MMSE system where exponential smoothing forecasts are used within the OUT policy. Again, we consider the lost sales to be fully observ-

[9]For instance, consider a safety factor of $\delta_p = 0.2$, and $\eta/\mu = 0.9$. With (40), we obtain $\tilde{\delta}_m = 0.08$, a 60% reduction to the 20% target.

able, so that the differences with the base case emerge only from the use of dynamic forecasts.

3.4.1. Fundamental relations in the DDF system

In the DDF system, the sum of inventory and orders is no longer a constant. From (15), we see that

$$i_t + q_t = (1+\delta) f_t. \tag{44}$$

Using (44) in (13) and (14) results in the following expressions for the inventory levels and satisfied demand,

$$i_t = [\underbrace{(1+\delta) f_{t-1} - d_t}_{x_1}]^+, \tag{45}$$

$$s_t = d_t \wedge \underbrace{(1+\delta) f_{t-1}}_{x_2}. \tag{46}$$

The previous forecast, f_{t-1} does not depend on the current demand, d_t. Thus, the satisfied demand, s_t, is the minimum of two *independent* normally distributed random variables, d_t and $(1-\delta) f_{t-1}$. With $[a-b]^+ = a - (a \wedge b)$, the inventory equation becomes $i_{t,d} = (1+\delta) f_{t-1} - ((1+\delta) f_{t-1} \wedge d_t)$. Using this relation in (15), we can express the order quantity as the minimum of two *correlated* normally distributed random variables,

$$\begin{aligned} q_t =& (1+\delta) f_t - (1+\delta) f_{t-1} + ((1+\delta) f_{t-1} \wedge d_t) \\ =& (1+\delta)(f_t - f_{t-1}) + ((1+\delta) f_{t-1} \wedge d_t) \\ =& (\underbrace{d_t + (1+\delta)(f_t - f_{t-1})}_{x_3}) \wedge \underbrace{(1+\delta) f_t}_{x_4}. \end{aligned} \tag{47}$$

Notice, we have defined four auxiliary variables, $\{x_1, x_2, x_3, x_4\}$, whose study will allow us to understand the dynamics of the DDF system.

3.4.2. Distribution of the DDF state variables

In order to derive the distribution of the state variables, we first need to identify the mean and variance of the x variables. Under the assumption of normally distributed demand, these four variables are also normally distributed.

Consider first x_1. Expectation provides its mean, μ_1. As $\mathbb{E}[d_t] = \mathbb{E}[f_{t-1}] = \mu$,

$$\mu_1 = \mathbb{E}[(1+\delta) f_{t-1} - d_t] = \delta\mu. \tag{48}$$

The variance of x_1, $\mathbb{V}[x_1] = \sigma_1^2$, can be obtained by noting that d_t and f_{t-1} are independent. The variance of the difference (or sum) of independent random variance is the sum of the variances of the random variables. By definition the variance of the demand is $\mathbb{V}[d_t] = \sigma^2$. The stationary variance of f_{t-1} is the same the stationary variance of f_t. From (7), the following expression exists,

$$\mathbb{V}[f_t] = \alpha^2\sigma^2 + (1-\alpha)^2 \mathbb{V}[f_t] \tag{49}$$

Table 2.: The mean and variance of the x variables

Auxiliary variable	$\mathbb{E}[x]$	$\mathbb{V}[x] = \sigma_x^2$
$x_1 = (1+\delta)f_{t-1} - d_t$	$\mu_1 = \delta\mu$	$\sigma_1^2 = \sigma^2\left(1 + \frac{\alpha(\delta+1)^2}{2-\alpha}\right)$
$x_2 = (1+\delta)f_{t-1}$	$\mu_2 = (1+\delta)\mu$	$\sigma_2^2 = \sigma^2\left(\frac{\alpha(\delta+1)^2}{2-\alpha}\right)$
$x_3 = d_t + (1+\delta)(f_t - f_{t-1})$	$\mu_3 = \mu$	$\sigma_3^2 = \sigma^2\left(\frac{\alpha(2\delta(\alpha(1+\delta)+2)+3)+2}{2-\alpha}\right)$
$x_4 = (1+\delta)f_t$	$\mu_4 = (1+\delta)\mu = \mu_2$	$\sigma_4^2 = \sigma^2\left(\frac{\alpha(\delta+1)^2}{2-\alpha}\right) = \sigma_2^2$

which can be re-arranged for the variance of an exponential smoothing forecast of i.i.d. demand; $\mathbb{V}[f_t] = \frac{\sigma^2\alpha}{2-\alpha}$. The variance of $\mathbb{V}[(1+\delta)f_t] = \frac{\sigma^2\alpha(1+\delta)^2}{2-\alpha}$, which leads to the following expression for the variance of x_1:

$$\sigma_1^2 = \sigma^2\left(1 + \frac{\alpha(\delta+1)^2}{2-\alpha}\right). \tag{50}$$

We can obtain the expressions for the mean and variance of the other x variables using exactly the same mechanism as for x_1. To save space, we simply state their values in Table 2 (which also details x_1 for completeness and ease of reference).

The mean and variance of $\{x_1, x_2, x_3, x_4\}$ allow us to determine the distribution of our three state variables, i_t, s_t, and q_t. First, the pdf of the distribution of the inventory, $\phi_{i,d}[x]$, is a translated, scaled, and truncated x_1 distribution, and can be obtained in much the same way as the distributions for the MMSE and PDO systems.

$$\phi_{i,d}[x] = \frac{\theta[x]}{\sigma_1}\phi\left[\frac{x-\delta\mu}{\sigma_1}\right] + \Delta[x]\Phi\left[\frac{-\delta\mu}{\sigma_1}\right]. \tag{51}$$

The pdf's of the distribution of s_t and q_t require a different approach. Eq. (46) shows that the satisfied orders are the minimum of two independent normally distributed random variables, d_t and x_2. Using (20)-(22), and considering the independence ($\rho = 0$), the pdf of the distribution of s_t is given by

$$\phi_{s,d}[x] = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right]\Phi\left[-\frac{x-\mu_2}{\sigma_2}\right] + \frac{1}{\sigma_2}\phi\left[\frac{x-\mu_2}{\sigma_2}\right]\Phi\left[-\frac{x-\mu}{\sigma}\right]. \tag{52}$$

Eq. (47) shows that the orders, $q_{t,d}$, are also minimum of two normally distributed random variables, x_3 and x_4. However, now they are correlated random variables. Thus to determine the distribution of the the orders, $\phi_{q,d}$, we need the Pearson correlation coefficient, $\rho_{3,4}$, between x_3 and x_4. This can be obtained with the relation $\rho_{3,4} = \left(\sum_{t=0}^{\infty}(\tilde{x}_3\tilde{x}_4)\right)/(\sigma_3\sigma_4)$, where $\{\tilde{x}_3, \tilde{x}_4\}$ are the impulse responses[10] of $\{x_3, x_4\}$. The impulse response is the output of a system when the demand is given by $\tilde{d}_0 = 1$ and $\tilde{d}_t = 0$, $\forall t > 0$. We can determine the impulse response of x_3 and x_4 from the impulse response of the exponential smoothing forecast $\tilde{f}_t$. By applying the relation in (7) recursively under an impulse demand it is easy to see $\tilde{f}_t = \alpha(1-\alpha)^t$. From this, the required impulse responses for $\tilde{x}_3$ and $\tilde{x}_4$ can be easily obtained,

$$\tilde{x}_3 = (1+\delta)(\alpha(1-\alpha)^t - \alpha(1-\alpha)^{t-1}\theta[t-1]) \tag{53}$$

[10]We use the tilde to highlight when the impulse response is taken.

$$\tilde{x}_4 = (1+\delta)\alpha(1-\alpha)^t. \tag{54}$$

By means of (53) and (54), we may obtain the correlation coefficient

$$\rho_{3,4} = \frac{\sigma_4(\alpha\delta+2)}{\sigma_3(\delta+1)}, \tag{55}$$

which allows us to derive the pdf of the orders, $\phi_{q,d}[x]$. From (20)-(22),

$$\begin{aligned}\phi_{q,d}[x] =& \frac{1}{\sigma_4}\phi\left[\frac{x-\mu_4}{\sigma_4}\right]\Phi\left[\frac{\rho(x-\mu_4)}{\sigma_4\sqrt{1-(\rho_{3,4})^2}\sigma_4} - \frac{x-\mu_3}{\sigma_3\sqrt{1-(\rho_{3,4})^2}\sigma_3}\right] + \\ & \frac{1}{\sigma_3}\phi\left[\frac{x-\mu_3}{\sigma_3}\right]\Phi\left[\frac{\rho(x-\mu_3)}{\sigma_3\sqrt{1-(\rho_{3,4})^2}\sigma_3} - \frac{x-\mu_4}{\sigma_4\sqrt{1-(\rho_{3,4})^2}\sigma_4}\right].\end{aligned} \tag{56}$$

3.4.3. Deriving the DDF performance metrics

With $\phi_{i,d}$, $\phi_{s,d}$, and $\phi_{q,d}$, we can obtain expressions for the four performance metrics. First, the inventory cover is given by

$$\tau_d = \frac{\mathbb{E}[i_t]}{\mu} = \frac{1}{\mu}\int_0^\infty x\phi_{i,d}[x]dx = \delta(\Phi[\lambda_d] + \phi[\lambda_d]/\lambda_d), \text{ where } \lambda_d = \frac{\delta\mu}{\sigma_1}. \tag{57}$$

Note, the definition of a *relative safety margin*, $\lambda_d = \delta\mu/\sigma_1$, also simplifies our expressions in the DDF system. Using (52) the fill rate, β, is given by

$$\beta_d = \frac{\mathbb{E}[(s_t)^+]}{\mathbb{E}[(d_t)^+]} = \frac{\int_0^\infty x\left(\frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right]\Phi\left[-\frac{x-\mu_2}{\sigma_2}\right] + \frac{1}{\sigma_2}\Phi\left[-\frac{x-\mu}{\sigma}\right]\phi\left[\frac{x-\mu_2}{\sigma_2}\right]\right)dx}{\mu\Phi\left[\gamma^{-1}\right] + \sigma\phi\left[\gamma^{-1}\right]}, \tag{58}$$

which provides an exact solution, but it is not possible to find a closed-form expression for the integral in the numerator. However, it is easy to numerically evaluate with the Excel Add-in detailed in Appendix A of Disney et al. (2015)[11].

Next we look to the variances. The inventory variance is given by

$$\begin{aligned}\mathbb{V}[i_t] =& \sigma_{i,d}^2 = \int_0^\infty (x-\mu\tau_d)^2\phi_i[x]dx \\ =& \Phi\left[\lambda_d\right]\left(\delta^2\mu^2\left(1-\Phi\left[\lambda_d\right]\right)+\sigma_1^2\right) - \sigma_1^2\phi[\lambda_d]^2 + \delta\mu\sigma_1\left(1-2\Phi\left[\lambda_d\right]\right)\phi\left[\lambda_d\right].\end{aligned} \tag{59}$$

The metric IVR can be readily obtained from (59),

$$\begin{aligned}IVR_d = \frac{\mathbb{V}[i_t]}{\mathbb{V}[d_t]} = \frac{1}{\sigma^2}\Bigg(&\Phi\left[\lambda_d\right]\left(\delta^2\mu^2\left(1-\Phi\left[\lambda_d\right]\right)+\sigma_1^2\right) - \\ &\sigma_1^2\phi[\lambda_d]^2 + \delta\mu\sigma_1\left(1-2\Phi\left[\lambda_d\right]\right)\phi\left[\lambda_d\right]\Bigg).\end{aligned} \tag{60}$$

[11] This Excel Add-in is also available to download from http://www.bullwhip.co.uk/#shiny.

With $\Phi[\lambda_d] = 1$ and $\phi[\lambda_d] = 0$, we obtain $IVR_d = \sigma_1^2/\sigma^2$, the expression for IVR in the equivalent linear DDF system.

The variance of the orders can be found using the first and second moments of the distribution of the minimum of two normally distributed random variables. The first moment, $\mathbb{E}[q]$, and the second moment, $\mathbb{E}[q^2]$, were given in Nadarajah and Kotz (2008). Via $\mathbb{V}[q] = \mathbb{E}[q^2] - \mathbb{E}[q]^2$; they can be combined to yield the variance of q,

$$\mathbb{V}[q_t] = \sigma_{q,d}^2 = \left(\mu_3^2 + \sigma_3^2\right)\Phi\left[\frac{\mu_4 - \mu_3}{\psi}\right] + \left(\mu_4^2 + \sigma_4^2\right)\Phi\left[\frac{\mu_3 - \mu_4}{\psi}\right] - \tag{61}$$
$$\psi(\mu_3 + \mu_4)\phi\left[\frac{\mu_4 - \mu_3}{\psi}\right] - \left(\mu_4\Phi\left[\frac{\mu_3 - \mu_4}{\psi}\right] + \mu_3\Phi\left[\frac{\mu_4 - \mu_3}{\psi}\right] - \psi\phi\left[\frac{\mu_4 - \mu_3}{\psi}\right]\right)^2,$$

where $\psi = \sqrt{\sigma_3^2 + \sigma_4^2 - 2\rho_{3,4}\sigma_3\sigma_4} = \sigma_1$. That $\psi = \sigma_1$ is serendipity—there is no x_1 variable in this problem. However, it does allow us to define a compact notation as, considering $\frac{\mu_4 - \mu_3}{\theta} = \frac{\delta\mu}{\sigma_1} = \lambda_d$, we can arrive at the following BW expression,

$$BW_d = \frac{\mathbb{V}[q_t]}{\mathbb{V}[d_t]} = \frac{1}{\sigma^2}\Bigg(\left(\mu_3^2 + \sigma_3^2\right)\Phi\left[\lambda_d\right] + \left(\mu_4^2 + \sigma_4^2\right)\Phi\left[-\lambda_d\right] - \psi(\mu_3 + \mu_4)\phi\left[\lambda_d\right] -$$
$$\left(\mu_4\Phi\left[-\lambda_d\right] + \mu_3\Phi\left[\lambda_d\right] - \psi\phi\left[-\lambda_d\right]\right)^2\Bigg). \tag{62}$$

Finally, substituting $\Phi[\lambda_d] = 1$, $\Phi[-\lambda_d] = 0$, and $\phi[\lambda_d] = \phi[-\lambda_d] = 0$ into (62) results in the Bullwhip expression for the equivalent linear DDF system, $BW_d = \sigma_3^2/\sigma^2$.

4. Numerical analysis of the OUT policy under lost sales: Operational performance

This section analyses the response of our three lost-sales systems by numerically evaluating the metrics obtained in Section 3. We first consider the BW and IVR ratios; later we study the trade-off between the fill rate and inventory cover.

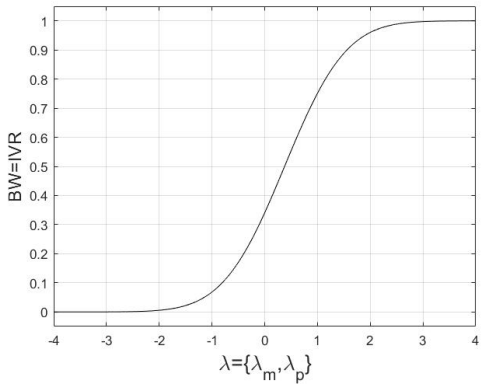
4.1. BW and IVR: Smoothing the operation of lost-sales systems

In the linear inventory system with backlogging, BW and IVR are generally used to characterise their dynamics. We now study the behaviour of the OUT policy with lost sales via these metrics, and assess their value in this nonlinear setting.

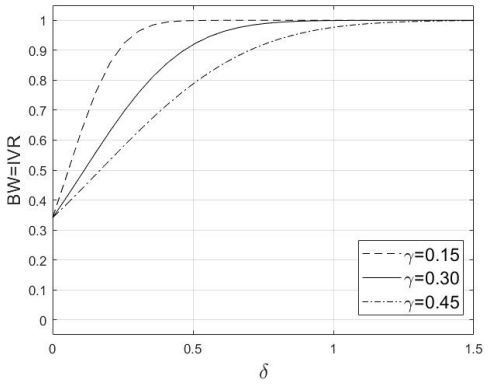
4.1.1. MMSE system

As discussed, the equivalent linear system with backlog has $BW = IVR = 1$. In our nonlinear lost-sales system, however, (33) reveals that BW and IVR depend upon λ_m. Due to $\lambda_m = \delta/\gamma$, the dynamic behaviour and stochastic performance of the lost-sales system with i.i.d. demand and MMSE forecasts is determined by the safety factor, δ, and the demand's coefficient of variation, γ.

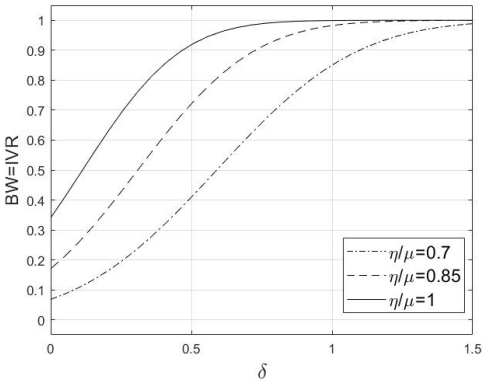
Fig. 2a displays the S-shaped relation between the variance ratios and λ_m. Notice, BW and IVR are both increasing in λ_m. When λ_m is sufficiently high, the retailer rarely experiences lost sales and the lost-sales system operates as a linear system would. To ensure $BW = IVR > 0.95$, we obtain $\lambda_m > 1.9$. That is, only when



(a) As a function of $\lambda = \{\lambda_m, \lambda_p\}$



(b) As a function of γ and δ under MMSE



(c) As a function of η/μ and δ under PDO

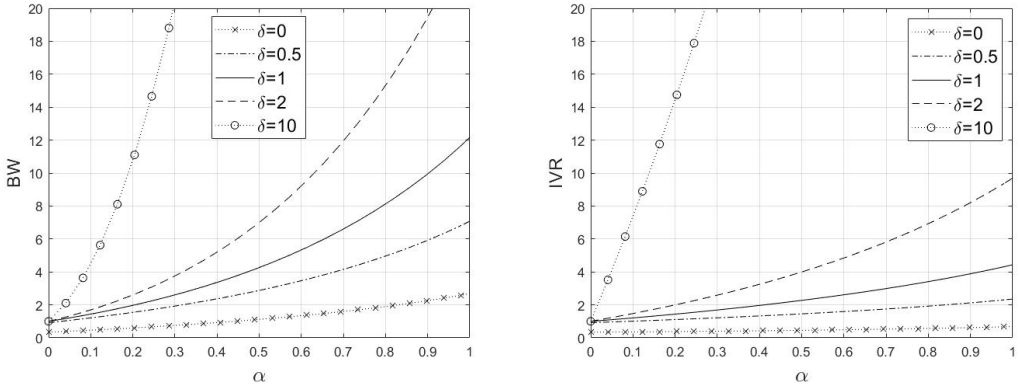
Figure 2.: BW and IVR in the MMSE and PDO systems.

$\delta > 1.9\gamma$ can the dynamics of the lost-sales inventory system be approximated by the backlog system (with less than 5% error). Smaller values of λ_m significantly reduce supply chain volatility, and the linear approximation does not hold.

Fig. 2b represents BW and IVR in the MMSE system as functions of δ for $\gamma \in \{15\%, 30\%, 45\%\}$, typical of retail time series (Dejonckheere et al. 2004). In line with our previous arguments, when δ becomes sufficiently high, the lost sales are negligible, and the nonlinear system behaves as the linear system does. However, for small to medium values of δ, the variability of orders and inventory decreases significantly. Importantly, larger γ requires a greater δ to ensure linear operation. γ has no impact upon the variance ratios for $\delta = 0$ (which is the value for $\lambda = 0$ in Fig. 2a).

4.1.2. PDO system

Due to the equivalence of the MMSE and PDO behaviours for λ_m and λ_p respectively, the relationship of BW and IVR with λ_p is also that shown in Fig. 2a. However, λ_p adopts a more complex form (compared to λ_m), being a function of $\{\mu, \sigma, \delta, \eta\}$. Note, $\lambda_m > 1.9$ now results in $(1 + \delta)(\eta/\mu) > 1 + 1.9\gamma$. This means that as more demand is unobserved, the conditions to mimic linear operation become more restrictive. For $\eta = \mu$ (MMSE), the relation leads to the previous condition $\delta > 1.9\gamma$. For $\eta/\mu = \{0.85, 0.7\}$, this becomes $\delta > 2.2\gamma + 0.2$ and $\delta > 2.7\gamma + 0.4$, respectively. That is, the linear approximation of lost-sales systems becomes more unrealistic under censored



(a) BW (b) IVR

Figure 3.: BW and IVR as a function of δ and α in the DDF scenario.

demand.

Consider now η, the static forecasting parameter that defines the difference between the PDO and the MMSE lost-sales systems. We do not study here the impact of γ, whose main effects were explored in the MMSE system; for this reason, we use $\gamma = 30\%$ (with $\sigma = 30$ and $\mu = 100$). Fig. 2c shows BW and IVR as a function of δ for $\eta/\mu \in \{1, 0.85, 0.7\}$. While $\eta/\mu = 1$ degenerates into the MMSE model (for $\gamma = 30\%$), $\eta/\mu \in \{0.85, 0.7\}$ represents the case when unobserved demand leads to biased forecasts. Fig. 2c shows that $BW = IVR$ is very sensitive to the the ratio η/μ and PDO leads to lower BW and lower IVR. That is, operational variability reduces as η/μ decreases. Note, unlike γ, η does impact the variance ratios for $\delta = 0$; this occurs because $\delta = 0$ does not result in $\lambda_p = 0$ under PDO. In line with prior discussions, Fig. 2c confirms that lower η requires a larger δ to ensure linear operation.

4.1.3. DDF system

Just as the PDO system may be interpreted as a generalisation of the MMSE system (for $\eta = \mu$), the DDF system is a different generalisation of the MMSE system (for $\alpha = 0$). Here we study the dynamics of the DDF system by exploring the impact of the exponential smoothing constant when $\alpha > 0$. As illustrated by (60) and (62), BW and IVR are complex functions of μ, γ, δ, and α. Also, the relation $BW = IVR$ no longer holds. For consistency, we again adopt $\mu = 100$ and $\sigma = 30$.

Fig. 3a and 3b represent BW and IVR as functions of α for $\delta \in \{0, 0.5, 1, 2, 10\}$. In general terms, we can see that as α increases, a large amount of BW and IVR is present; this is not a surprising result as in linear systems, MMSE is the inventory optimal forecasting method when used in the OUT policy, and the OUT policy with exponential smoothing is known to always generate Bullwhip (Dejonckheere et al. 2003).

There is a remarkable interaction between α and δ. Consider first BW (Fig. 3a). For low δ, the BW measure is relatively robust to variations in α. However, as δ grows, it becomes more sensitive to α. For instance, when $\delta = 0$, as α increases from 0.1 to 0.2, BW only increases from 0.5 to 0.6 (approx.); however, when $\delta = 1$, BW increases from 1.4 to 2 (approx.). Similar conclusions can be drawn about IVR (Fig. 3b); although IVR is generally less sensitive to α than BW (except for high δ).

Finally, δ impacts BW and IVR differently in the DDF system compared to the

previous cases. In linear MMSE systems, $\forall\delta$, $BW = IVR = 1$; in the DDF system, large δ generates $BW >> 1$ and $IVR >> 1$. This highlights an important drawback of using large dynamic safety stocks; while safety stocks are generally adjusted to balance holding costs with the stock-out risk in backlog systems, dynamic safety stocks can induce a large amount of BW and IVR in lost sales systems.

4.2. τ and β: Efficiently meeting customer demand in lost-sales settings

The variance study suggests the lost-sales nonlinearity creates some dynamic benefit. However, the lost-sales condition introduces other differences that should be taken into account. We emphasise that the expected order matches the expected demand in linear backlog systems (as unmet demand will be satisfied later). However, in a lost-sales system this equality does not hold. Indeed, the expected order is less than the expected demand, and a portion of the demand is not satisfied, $\mathbb{E}[o_t] = \mathbb{E}[s_t] < \mathbb{E}[d_t]$.

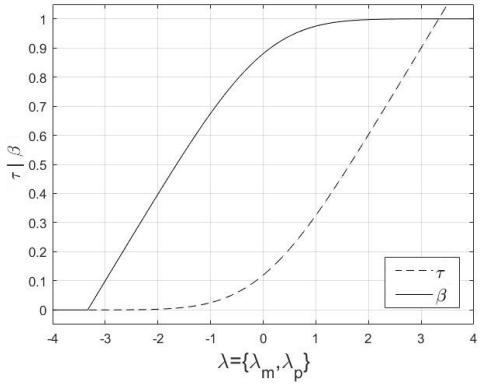
In light of this, BW may be reduced in our nonlinear system as a consequence of an increase in the lost sales; IVR may also decrease due to the unobserved demand[12]. This has important implications for theory and practice of lost-sales supply chains. While the simultaneous consideration of BW and IVR provides a complete picture of the performance of replenishment policies in linear supply chains, (Disney and Lambrecht 2008), to be meaningful in lost-sales scenarios they need to be considered alongside other metrics that account for the fact that $\mathbb{E}[o_t] < \mathbb{E}[d_t]$. Specifically, while IVR is a useful indicator in backlogged inventory systems for providing key information on the trade-off between service level and inventory investment, its physical meaning may be distorted in lost-sales systems due to the nonlinearity. Therefore, we also analyse inventory performance in our three systems by measuring the fill rate, β, and the inventory cover, τ.

4.2.1. MMSE system

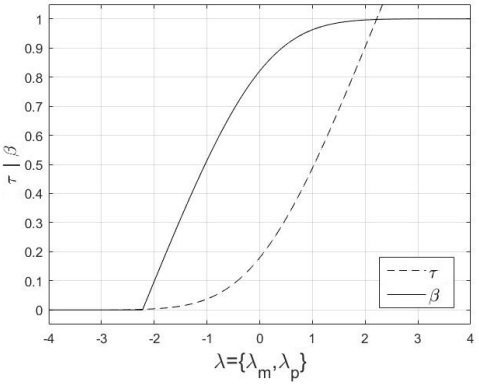
Here, β and τ depend on γ and λ_m, which in turn depend on $\{\sigma, \mu, \delta\}$, see (30)-(31). Fig. 4a represents the inventory metrics for $\gamma = 30\%$ (again, $\sigma = 30$, $\mu = 100$). For $\lambda_m << 0$, the fill rate $\beta = 0$ and the inventory cover $\tau = 0$. For larger λ_m, τ is increasing and convex in λ_m. In contrast, β is increasing and concave in λ_m, with $\beta = 1$ for $\lambda_m >> 0$. The impact of γ can be understood by comparing Fig. 4a to Fig. 4b, which shows the same information for $\gamma = 45\%$. Increasing γ shifts the β curve to the right and the τ curve to the left; the fill rate is reduced despite holding more stock when the co-efficient of variation of demand increases from 0.3 to 0.45.

In the MMSE system, $\lambda_m = \delta/\gamma$; Fig. 4c and 4d display respectively τ and β for the same three levels of demand variability considered before: $\gamma \in \{15\%, 30\%, 45\%\}$. Similar to a linear system, higher δ increases τ and also improves β. Note, for higher γ and small δ, τ deviates further from the linear slope $\tau = \delta$. Importantly, the fill rate, β, is the same as for the linear backlog version. This is a consequence of the unit lead-time assumption. Indeed, despite the difference between the orders ($q_t = d_t$ in the linear system; $q_t = s_t$ in the nonlinear system), the unit lead time ensures all backlogs are cleared in the next period and β measures the demand satisfied without delay, which is the same in both the linear and nonlinear systems.

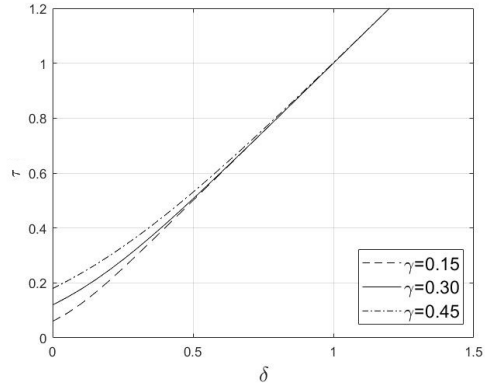
[12]When all demand is lost, $\forall t$, $i_t = 0$, $q_t = 0$, which explains why $BW = IVR = 0$ when $\lambda << 0$ (Fig. 2a).



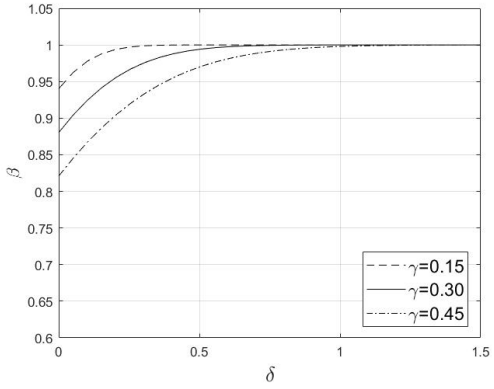
(a) As a function of $\lambda = \{\lambda_m, \lambda_p\}$ for $\gamma = 30\%$



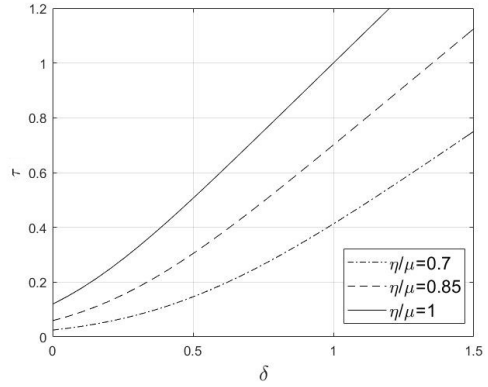
(b) As a function of $\lambda = \{\lambda_m, \lambda_p\}$ for $\gamma = 45\%$



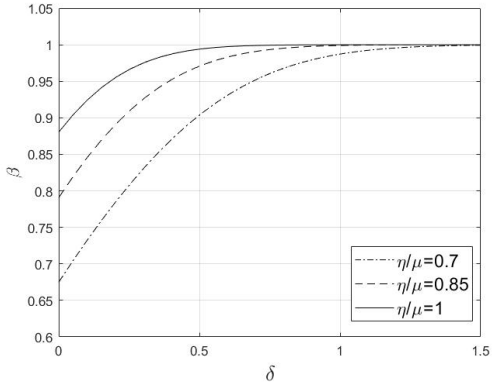
(c) τ as a function of γ and δ under MMSE



(d) β as a function of γ and δ under MMSE



(e) τ as a function of η/μ and δ under PDO



(f) β as a function of η/μ and δ under PDO

Figure 4.: τ and β in the MMSE and PDO systems.

4.2.2. PDO system

Fig. 4a and 4b are also valid under PDO, where $\lambda_p = [(1+\delta)\eta - \mu]/\sigma$. Recall, the PDO system may be interpreted as a MMSE system with a safety factor of $\tilde{\delta}_m$, see (40). From this perspective, the impact of unobservable lost sales on inventory performance is a reduction (determined by the ratio η/μ) of the 'effective' protection against stock-outs. In practical terms, for a selected safety factor, the partial observation of demand reduces the fill rate and the holding requirements, in relation to the target.

To better understand such effects, we focus again on a scenario with $\gamma = 30\%$. Fig. 4e and 4f represent the relation between τ and β with δ for three values of η/μ. In line with the previous discussion, not being able to observe consumer demand reduces β significantly, even with a relatively large δ. For example, if $\eta = \mu$, $\delta_p \approx 0.2$ is required to achieve 95% fill rate, while if $\eta = 0.7\mu$, $\delta_p \approx 0.7$ is needed to reach the same β ($\delta_p = 0.7$ implies $\tilde{\delta}_m = 0.19$). This highlights that using inventory configurations derived from linear models in scenarios with unobserved lost sales is particularly risky, as it results in dramatically decreased service levels.

Importantly, our analysis suggests the existence of a *vicious circle* that has not been modelled in our study but may be an interesting area of further study; reducing the percentage of observed demand, η/μ, reduces the fill rate β, which in turn will tend to reduce the portion of demand that is observed.

4.2.3. DDF system

To study the impact of using exponential smoothing forecasts on the inventory performance of lost-sales systems, we consider again the decision parameter α. Fig. 5a and 5b represent τ and β, respectively, in the DDF system as functions of the exponential smoothing constant, α, for different safety factors, $\delta \in \{0, 0.5, 1, 2\}$. Inspection of these figures reveal that increasing α has a relatively small impact on holding costs, indicated by τ/μ (Fig. 5a), but significantly affects the fill rate β (Fig. 5b). Also, the sensitivity of the metrics to α grows as δ decreases. Specifically, high α values reduce β significantly. When $\delta \geq 1$, the fill rate is barely affected as long as $\alpha < 0.5$; however, for $\delta \leq 0.5$, even small increases of α induce a considerable reduction in the fill rate. Increasing α, not only reduces the fill rate, but also tends to (slightly) increase inventory cover. From the perspective of α, our findings in these section are well aligned with the previous BW and IVR analysis; we emphasise the control parameters δ and α need to be set correctly to avoid inefficiencies originating from their interaction effects.

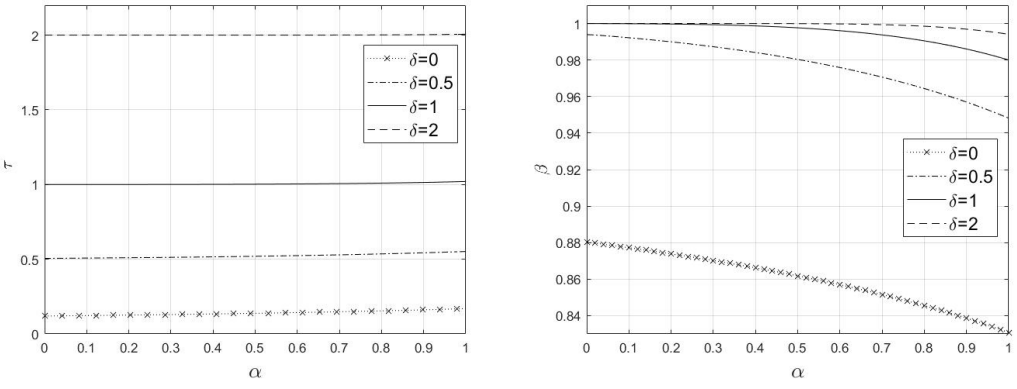
4.3. Summary of operational insights

To summarize, Table 3 provides an overview of the key conceptual findings related to the impact of the parameters of the lost-sales system on the key performance metrics under the three scenarios under consideration. The insights for the MMSE system are derived by comparing the lost-sales OUT policy to the linear (backlogging) OUT policy. The insights for the PDO and DDF systems are derived by comparing their behaviour to the MMSE system with lost sales.

Table 3.: Main effects of the parameters in the lost-sales order-up-to inventory systems.

	MMSE scenario	**PDO scenario**	**DDF scenario**
Parameter setting	$\eta/\mu = 1$ $\alpha = 0$	$\eta/\mu < 1$ $\alpha = 0$	$\eta/\mu = 1$ $0 < \alpha \leq 1$
Bullwhip ratio (BW)	Increases in δ/γ, with linear behavior ($BW = 1$) when $\delta/\gamma >> 0$	Reduces as η/μ lowers, with the reduction rate strongly influenced by δ/γ	Increases in α, with the increase rate strongly influenced by δ/γ
Inv. var. ratio (IVR)	$= BW$	$= BW$	$\neq BW$, but the main effects of α and δ/γ are similar
Fill rate (β)	Same as the linear system due to the unit lead-time	Reduces significantly as η/μ reduces, unless $\delta >> 0$	Decreases in α, unless $\delta >> 0$
Inv. cover (τ)	*Idem*	Reduces as η/μ reduces, but the trade-off β-τ is still efficient$^{(i)}$	Slightly increases in α, moving away from the efficient frontier

Note: $^{(i)}$ The efficient frontier in the β *vs.* τ trade-off is defined by the MMSE system. Varying η/μ deviates the real from the target trade-off, but does not make the system inefficient.



(a) τ (b) β

Figure 5.: τ and β as a function of δ and α in the DDF scenario.

5. Economic analysis of the lost sales OUT policy

The operational analysis in the previous section provided an in-depth understanding of our five research questions, which we will revisit in Section 6. Our analytical understanding of the lost sales inventory system also allows us to explore the economic optimisation of the lost sales OUT policy.

The economic analysis is here carried out for the MMSE system, as it defines the (cost-)efficient frontier in our lost sales inventory system. Note, it is not possible to provide prescriptive results for the PDO system as we need to know the mean demand[13]; by assumption, this is not known under PDO (indeed, knowing μ would lead us to the MMSE system). Furthermore, it would not be possible to do the same analysis for the DDF system as there is no closed-form expression for the fill rate in this system.

[13]For interested readers the 'optimised' PDO parameters and associated costs can be easily obtained from the following MMSE analysis using the relation in (40).

5.1. Cost models and lost-sales settings

In the economic study of the lost-sales inventory system, it is interesting to consider two practically relevant settings. First assume a pure retail setting where the OUT policy is designed to optimise the inventory trade-off between customer service level and holding investment. In this case, we assume that a per-unit inventory holding cost of h and a per unit penalty cost, p, for lost sales l_t is incurred in each period. Therefore, the inventory cost in t becomes

$$\pi_{i,t} = hi_t + pl_t, \tag{63}$$

where $l_t = [d_t]^+ - [s_t]^+$ measures the stock-out size in period t.

Using expectation, the average inventory is $\mathbb{E}[i_t] = \tau\mathbb{E}[d_t]$, see (12), where $\mathbb{E}[d_t] = \mu$, see (2), and if unit holding costs of h were applied, the expected, per period, inventory holding cost would be $h\tau\mu$.

The average lost sale can be determined from the fill rate expression in (11). The positive satisfied demand is given by $\mathbb{E}[[s_t]^+] = \beta\mathbb{E}[[d_t]^+]$. Using $\mathbb{E}[l_t] = \mathbb{E}[[d_t]^+] - \mathbb{E}[[s_t]^+]$, $\mathbb{E}[l_t]$ can be written as $\mathbb{E}[l_t] = (1-\beta)\mathbb{E}[[d_t]^+]$. With a unit penalty cost of p for each lost sale, the expected, per period, stock-out cost would be $p(1-\beta)\mathbb{E}[[d_t]^+]$. As the demand d_t is normally distributed, the expected value of (63) can be written as

$$\pi_i = \mathbb{E}[\pi_{i,t}] = h\tau\mu + p(1-\beta)(\sigma\phi\left[\gamma^{-1}\right] + \mu\Phi\left[\gamma^{-1}\right]). \tag{64}$$

Second, we consider an alternative scenario where order variability also creates costs. This could apply to many practical settings; but for the sake of simplicity we will call this the manufacturing setting. In this case, in addition to the aforementioned inventory costs, we may also incur capacity-related production costs.

We consider the practically common situation where a pre-installed capacity is present that can produce k units within a nominal working week at a unit production cost of u. If the weekly production requirement was less than k units, the labour is paid to stand idle for part of the week; labour is guaranteed their nominal weekly wage and a per period capacity cost of uk is incurred. If the weekly production requirements (q_t) is larger than the pre-installed capacity (k), the flexible labour is required to work over time (at a unit cost of $w > u$) to make-up the shortfall. This leads to the following production cost being incurred,

$$\pi_{q,t} = uk + w[q_t - k]^+; \tag{65}$$

see Disney, Gaalman, and Hosoda (2012) for further details behind the rationale of this economic model.

Taking expectation of (65), the expected, per period, production cost is given by

$$\pi_q = \mathbb{E}[\pi_{q,t}] = uk + w\mathbb{E}[q_t - k]^+. \tag{66}$$

Adding in the inventory holding and lost sales costs (64) to the production costs (66) gives a total cost in the manufacturing setting of

$$\pi_{q+i} = \mathbb{E}[\pi_q] + \mathbb{E}[\pi_i]. \tag{67}$$

In the following subsections we investigate the consequences of these two costs settings, and offer prescriptive advice on how to set to the safety factor δ and the production capacity k.

5.2. The economics of retailing

Consider the MMSE system with only inventory holding and lost-sales costs as given by (64) are present. Using τ_m and β_m from (30) and (31), we obtain the following inventory cost equation for the MMSE system,

$$\begin{aligned}\pi_i =& h\gamma(\phi\left[\lambda_m\right] + \lambda_m\Phi\left[\lambda_m\right])\mu + \\ & p\left(\sigma\phi\left[\gamma^{-1}\right] + \mu\Phi\left[\gamma^{-1}\right]\right)\left(1 - \theta[1+\gamma\lambda_m]\left(1 - \frac{\phi[\lambda_m] + \lambda_m(\Phi[\lambda_m]-1)}{\phi\left[\gamma^{-1}\right] + \gamma^{-1}\Phi\left[\gamma^{-1}\right]}\right)\right);\end{aligned} \tag{68}$$

recall $\lambda_m = \delta/\gamma$.

Assuming $\delta > -1$ (to eliminate the Heaviside function in (68)), the following derivative of π_i with respect to δ can be found

$$\frac{d\pi_i}{d\delta} = \mu\left((h+p)\Phi\left[\frac{\delta\mu}{\sigma}\right] - p\right). \tag{69}$$

Setting $d\pi_i/d\delta = 0$ and solving for $\delta = 0$ provides the first-order conditions for an optimal safety factor $\delta_i^\star$,

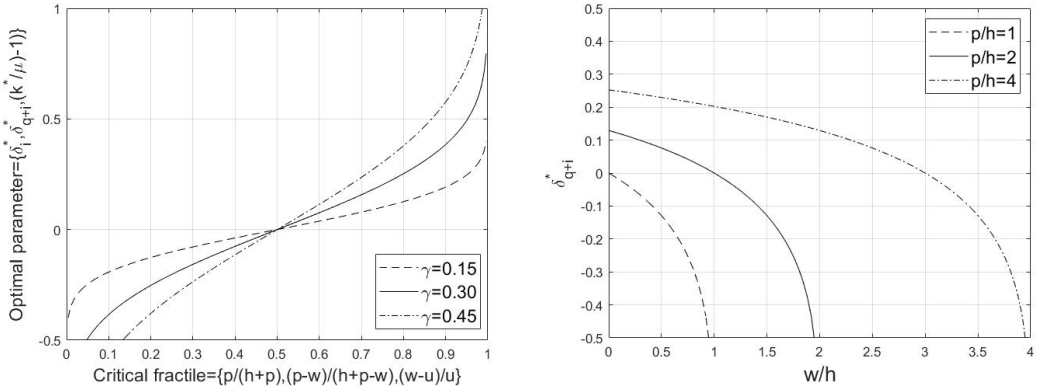
$$\delta_i^\star = \gamma\Phi^{-1}\left[\frac{p}{h+p}\right]. \tag{70}$$

That is, the optimal safety factor is proportional to the coefficient of variation of customer demand, γ, and to the inverse cdf at the critical fractile $p/(h+p)$. Fig. 6a illustrates how $\delta_i^\star$ is influenced by the costs (via the critical fractile $p/(h+p)$) and the demand's co-efficient of variation, γ. Note, for $p > h$, $\delta_i^\star > 0$ and $\delta_i^\star$ is increasing convex in the critical fractile and increasing in γ. When $p < h$, $\delta_i^\star < 0$ and $\delta_i^\star$ is increasing concave in the critical fractile and $\delta_i^\star$ is decreasing in γ.

Finally, using (70) in (68), the minimised sum of the inventory holding and lost-sales costs is given by

$$\pi_i^\star = \sigma(h+p)\phi\left[\Phi^{-1}\left[\frac{p}{h+p}\right]\right]. \tag{71}$$

The minimised inventory and lost sales costs are linear in the standard deviation of demand, σ. The critical fractal has a similar structure to the linear backlogging system, albeit with the backlog cost replaced with the lost sales penalty cost.



(a) The optimal parameters in our lost-sales system (b) The optimal δ in the manufacturing setting

Figure 6.: Optimisation of the safety factor in the MMSE lost-sales OUT policy.

5.3. The economics of manufacturing

From (66) and using the analytical understanding of $\mathbb{E}[q_t]$ in the MMSE case offered in Section 3.2, the expected production cost in the MMSE system is given by

$$\begin{aligned}\pi_q =& uk + w\Phi[-\lambda_m](\mu(1+\delta) - k) + w\int_k^{\mu(1+\delta)} \frac{(x-k)\phi\left[\frac{x-\mu}{\sigma}\right]}{\sigma}\, dx \\ =& uk - w\Bigg(\delta\mu\Phi[\lambda_m] - (\delta+1)\mu + \sigma\left(\phi[\lambda_m] - \phi\left[\frac{k-\mu}{\sigma}\right]\right) + \\ & (\mu-k)\Phi\left[\frac{k-\mu}{\sigma}\right] + k\Bigg).\end{aligned} \tag{72}$$

Using (68) and (72) in (67), taking derivative of π_{q+i} w.r.t. k yields

$$\frac{d\pi_{q+i}}{dk} = u + w\left(\Phi\left[\frac{k-\mu}{\sigma}\right] - 1\right), \tag{73}$$

which is independent of δ. This allows the optimal capacity level to be set independently from the safety stock. Setting the derivative to zero provides the first-order condition for the optimal capacity $k^\star$,

$$k^\star = \mu + \sigma\Phi^{-1}\left[\frac{w-u}{w}\right]. \tag{74}$$

We notice the optimal installed capacity in the MMSE lost-sales system is the same as the linear OUT policy, Hosoda and Disney (2012). Note, when $u < w < 2u$ then $k^\star < \mu$. The fundamental nature of $k^\star$ is the same as the for $\delta_i^\star$, as shown in Fig. 6a.

To determine the optimal safety factor in the manufacturing setting, denoted by $\delta_{q+i}^\star$, we now take the derivative of π_{q+i} w.r.t. δ,

$$\frac{d\pi_{q+i}}{d\delta} = \mu\left((h+p-w)\Phi\left[\frac{\delta\mu}{\sigma}\right] - (p-w)\right). \tag{75}$$

Setting the derivative to zero and solving for the first-order conditions yields the safety stock gain that minimizes total costs,

$$\delta^{\star}_{q+i} = \gamma\Phi^{-1}\left[\frac{p-w}{h+p-w}\right]. \tag{76}$$

Notice, $\delta^{\star}_{q+i}$ is different to $\delta^{\star}_{i}$ in (70). $\delta^{\star}_{q+i}$ is independent of k, but is a function of w, but not u (as the derivative (75) does not contain u). Moreover, while $k^{\star}$ exists whenever $w \geq u$, $\delta^{\star}_{q+i}$ only exists if $p \geq w \geq u$. That is, π_{q+i} has no minimum in δ when $p < w$. Indeed, as the order variability costs dominate, they drive $\delta^{\star}_{q+i} \to -\infty$ so as to reduce the order variability, and hence production costs, to zero.

To understand how $\delta^{\star}_{q+i}$ changes to the critical fractile we may use Fig. 6a again; the curves have the same form as $\delta^{\star}_{i}$. In addition, Fig. 6b shows how the ratios w/h and p/h affect $\delta^{\star}_{q+i}$ (as the influence of γ was explored in Fig. 6a, here $\gamma = 0.3$). $\delta^{\star}_{q+i}$ is decreasing in w/h, but increasing in p/h.

Finally, substituting in (74) and (76) into (67) provides the total minimised inventory holding, lost sales and production cost,

$$\pi^{\star}_{q+i} = \mu u + \sigma\left((h+p-w)\phi\left[\Phi^{-1}\left[\frac{p-w}{h+p-w}\right]\right] + w\phi\left[\Phi^{-1}\left[\frac{w-u}{w}\right]\right]\right). \tag{77}$$

Notice, the total costs are linear in the standard deviation of the demand, σ, but offset by μu.

6. Conclusions

The lack of literature on lost-sales inventory systems together with their relevance in real inventory settings motivated our research. We contribute to this area by providing an analytical understanding of the behaviour of lost-sales supply chains under the industrially popular OUT replenishment policy. We study three forecasting procedures under i.i.d. demand and unit lead times; this allowed us to investigate a wide spectrum of practical settings affected by lost sales.

Our insights have emerged from the derivation of expressions of the Bullwhip and IVR ratios, the fill rate, and the inventory cover in the three lost-sales settings; key performance metrics that report upon the stability of operations and the efficiency in satisfying consumer needs. The analysis answers the research questions posed earlier in the following ways:

- *RQ1: How do the dynamics of lost sales systems compare to the dynamics of the backlogging systems?* The lost-sales condition significantly influences the behaviour of supply chains in comparison to backlogging systems. In particular, lost sales help to mitigate the Bullwhip effect. Under full demand observation and MMSE forecasting, the degree of variability reduction only depends on the ratio of the safety factor δ to the demand's coefficient of variation γ, which we denote by *relative safety margin*, λ_m. This proves to be a key parameter for inventory managers; if $\delta \leq \gamma$ (i.e. $\lambda_m \leq 1$), BW can be reduced by at least 25%.
- *RQ2: Can we use the same performance metrics for evaluating both backlogging and lost-sales inventory systems?* High inventory variability is symptomatic of poor inventory control under backlog. Indeed, it determines one's ability to meet

a defined fill rate in a cost-effective manner. However, the non-negative constraint in the inventory distorts the value of IVR, which may wrong-foot managers and it cannot be assumed that reducing IVR always improves inventory control. Therefore, the order-inventory variability trade-off, of fundamental importance in backlog systems, should be reconsidered carefully in lost-sales environments.

- *RQ3: Under what circumstances can the dynamic behaviour of nonlinear OUT policy with lost sales be approximated by linear OUT policy with backlogs?* The dynamics of lost-sales inventory systems can only be approximated by backlogging systems when the safety stock is sufficiently high. Our research contributes to industrial practice by clarifying what 'sufficiently' means. In the ideal MMSE case, when customer demand can be fully observed, the linear approximation should be avoided when $\delta < 1.9\gamma$ (otherwise, the error is higher than 5%). As $\delta > 1.9\gamma$ generally results in very high fill rate, we may conclude the linear model is not a good approximation in settings with more humble fill-rate targets.
- *RQ4: What is the impact of using dynamic forecasts, compared to static forecasts in lost-sales settings?* Employing dynamic forecasts of i.i.d. demand has negative implications on the lost-sales supply chains. This applies to both production smoothing and inventory performance. For low values of the safety parameter, however, the system is robust to 'moderate' increases in the smoothing constant ($\alpha < 0.3$). In this sense, the impact of dynamic forecasts is similar in lost-sales and backlogging inventory systems. However, under lost sales, the BW increase is attenuated by λ, especially for low safety stocks. In light of this, managers should recognise that using simultaneously high α and high δ is potentially very harmful; e.g., $\alpha = 0.2$ and $\delta = 2$, BW increases from ≈ 1 (MMSE) to ≈ 4 (DDF).
- *RQ5: What are the consequences of observing the full demand, compared to observing only the satisfied demand?* Observing full demand allows for a more accurate control of lost-sales supply chains. When lost sales are unobservable, the effective protection against out-of-stocks decreases, and the fill rate achieved is significantly lower than the fill rate targeted. This may lead supply chain managers into a dangerous circle, with lower fill rates in turn resulting in a reduced observation of the actual demand. For this reason, when the demand cannot be fully observed, it may be reasonable (at least, temporarily or intermittently) to set safety stocks to moderately higher values than those calculated, given that under PDO, the safety stock needs to compensate for both the demand variability and the censored demand.

After discussing the operational characteristics of our four performance metrics, we combined three of them into two economic cost functions. One of the cost functions consisted of inventory holding and lost sales costs, relevant in retail settings. We were able to obtain prescriptive results for the optimal safety stock gain in this setting. The other cost function included the inventory holding and lost sales costs together with production costs that consisted of an installed capacity base for production with regular hours with a guaranteed wage for labour and a flexible overtime to meet large orders above the installed capacity. This cost function is relevant in production settings dominated by labour costs. Here we obtained prescriptive results for the safety stock gain and the required production capacity.

Finally, we underline important avenues for future research. We restricted our study to unit lead times in an OUT policy. Subsequent studies could be directed to understanding the general lead time case as well as the dynamics induced by other replen-

ishment policies. Also, starting from the base MMSE case, we considered censored demand and dynamic forecasts; however, the interaction between these characteristics makes the analysis considerably more difficult, and was not been studied herein. Similarly, we only modelled one source of nonlinearity in our study. Investigating the interactions of the lost-sales condition with other nonlinearities, such as capacity constraints or forbidden returns, is a important area for future research to increase our comprehension of real-world, nonlinear supply chains.

References

Agrawal, N., and S.A. Smith. 1996. "Estimating negative binomial demand for retail inventory management with unobservable lost sales." *Naval Research Logistics* 43 (6): 839–861.

Besbes, O., and A. Muharremoglu. 2013. "On implications of demand censoring in the newsvendor problem." *Management Science* 59 (6): 1407–1424.

Bijvank, M., and I.F.A. Vis. 2011. "Lost-sales inventory theory: A review." *European Journal of Operational Research* 215 (1): 1–13.

Bijvank, M., and I.F.A. Vis. 2012. "Lost-sales inventory systems with a service level criterion." *European Journal of Operational Research* 220 (3): 610–618.

Cachon, G., and C. Terwiesch. 2006. *Matching Supply with Demand: An Introduction to Operations Management.* McGraw-Hill.

Cardós, M., E. Guijarro, and E. Babiloni. 2017. "On the estimation of on-hand stocks for base-stock policies and lost sales systems and its impact on service measures." *International Journal of Production Research* 55 (16): 4680–4694.

Chatfield, D.C., and A.M. Pritchard. 2013. "Returns and the bullwhip effect." *Transportation Research Part E: Logistics and Transportation Review* 49 (1): 159–175.

Chen, Li, and Hau L Lee. 2012. "Bullwhip effect measurement and its implications." *Operations Research* 60 (4): 771–784.

Chen, Li, Wei Luo, and Kevin Shang. 2017. "Measuring the bullwhip effect: Discrepancy and alignment between information and material flows." *Manufacturing & Service Operations Management* 19 (1): 36–51.

Chen, Y., S. Ray, and Y. Song. 2006. "Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales." *Naval Research Logistics* 53 (2): 117–136.

Corsten, D., and T. Gruen. 2003. "Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks." *International Journal of Retail & Distribution Management* 31 (12): 605–617.

Costas, J., B. Ponte, D. de la Fuente, R. Pino, and J. Puche. 2015. "Applying Goldratt's Theory of Constraints to reduce the Bullwhip Effect through agent-based modeling." *Expert Systems with Applications* 42 (4): 2049–2060.

Dawn, S.K., and R. Chowdhury. 2011. "Electronic customer relationship management (E-CRM): Conceptual framework and developing a model." *International Journal of Business and Information Technology* 1 (1): 75–84.

Dejonckheere, J., S.M. Disney, M.R. Lambrecht, and D.R. Towill. 2003. "Measuring and avoiding the bullwhip effect: A control theoretic approach." *European Journal of Operational Research* 147 (3): 567 – 590.

Dejonckheere, J., S.M. Disney, M.R. Lambrecht, and D.R. Towill. 2004. "The impact of information enrichment on the bullwhip effect in supply chains: A control engineering perspective." *European Journal of Operational Research* 153 (3): 727–750.

Disney, S.M., I. Farasyn, M.R. Lambrecht, D.R. Towill, and W. Van de Velde. 2006. "Taming the bullwhip effect whilst watching customer service in a single supply chain echelon." *European Journal of Operational Research* 173 (1): 151–172.

Disney, S.M., G. Gaalman, and T. Hosoda. 2012. "Review of stochastic cost functions for

production and inventory control." In *17th International Working Seminar of Production Economics*, 117–128.

Disney, S.M., G.J.C. Gaalman, C.P.T. Hedenstierna, and T. Hosoda. 2015. "Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand." *International Journal of Production Economics* 170: 501–512.

Disney, S.M., and M.R. Lambrecht. 2008. "On replenishment rules, forecasting, and the bullwhip effect in supply chains." *Foundations and Trends in Technology, Information and Operations Management* 2 (1): 1–80.

Disney, S.M., A. Maltz, X. Wang, and R.D.H. Warburton. 2016. "Inventory management for stochastic lead times with order crossovers." *European Journal of Operational Research* 248 (2): 473–486.

Disney, S.M., B. Ponte, and X Wang. 2019. "The nonlinear dynamics of order-up-to inventory systems with lost sales." *IFAC-PapersOnLine* 52 (13): 2291–2296.

Disney, S.M., D.R. Towill, and W. Van de Velde. 2004. "Variance amplification and the golden ratio in production and inventory control." *International Journal of Production Economics* 90 (3): 295–309.

Dolgui, A., D. Ivanov, and M. Rozhkov. 2020. "Does the ripple effect influence the bullwhip effect? An integrated analysis of structural and operational dynamics in the supply chain." *International Journal of Production Research* 58 (5): 1285–1301.

Essila, J.C. 2019. "E-Business supply chains drivers, metrics, and ERP integration." In *Advanced Methodologies and Technologies in Business Operations and Management*, 989–1002. IGI Global.

Gayon, J.P., G. Massonnet, C. Rapine, and G. Stauffer. 2016. "Constant approximation algorithms for the one warehouse multiple retailers problem with backlog or lost-sales." *European Journal of Operational Research* 250 (1): 155–163.

Godichaud, M., and L. Amodeo. 2019. "EOQ inventory models for disassembly systems with disposal and lost sales." *International Journal of Production Research* 57 (18): 5685–5704.

Goldberg, D.A., D.A. Katz-Rogozhnikov, Y. Lu, M. Sharma, and M.S. Squillante. 2016. "Asymptotic optimality of constant-order policies for lost sales inventory models with large lead times." *Mathematics of Operations Research* 41 (3): 898–913.

Goltsos, T.E., B. Ponte, S. Wang, Y. Liu, M.M. Naim, and A.A. Syntetos. 2019. "The boomerang returns? Accounting for the impact of uncertainties on the dynamics of remanufacturing systems." *International Journal of Production Research* 57 (3): 7361–7394.

Gruen, T. W, D. S Corsten, and S. Bharadwaj. 2002. *Retail Out of Stocks: A Worldwide Examination of Extent, Causes, and Consumer Responses*. Technical Report. Washington, DC, US: Grocery Manufacturers of America.

Heskett, J.L., T.O. Jones, G.W. Loveman, W.E. Sasser, and L.A. Schlesinger. 1994. "Putting the service-profit chain to work." *Harvard Business Review* 72 (2): 164–174.

Holweg, M., and S.M. Disney. 2005. "The evolving frontiers of the bullwhip problem." In *EUROMA Conference Proceedings*, 707–716.

Hosoda, T., and S.M. Disney. 2012. "On the replenishment policy when the market demand information is lagged." *International Journal of Production Economics* 135 (1): 458 – 467.

Hu, Qingwen. 2019. "Bullwhip effect in a supply chain model with multiple delivery delays." *Operations Research Letters* 47 (1): 36–40.

Huh, W.T., G. Janakiraman, J.A. Muckstadt, and P. Rusmevichientong. 2009. "Asymptotic optimality of order-up-to policies in lost sales inventory systems." *Management Science* 55 (3): 404–420.

IHL. 2015. *Retailers and the ghost economy: $1.75 trillion reasons to be afraid*. Technical Report. Franklin, TN, US: IHL Group.

Isaksson, O.H.D., and R.W. Seifert. 2016. "Quantifying the bullwhip effect using two-echelon data: A cross-industry empirical investigation." *International Journal of Production Economics* 171: 311–320.

Johansen, S.G. 2001. "Pure and modified base-stock policies for the lost sales inventory system with negligible set-up costs and constant lead times." *International Journal of Production*

Economics 71 (1-3): 391–399.

Johansen, S.G. 2013. "Modified base-stock policies for continuous-review, lost-sales inventory models with Poisson demand and a fixed lead time." *International Journal of Production Economics* 143 (2): 379–384.

Karlin, S., and H. Scarf. 1958. "Inventory models and related stochastic processes." *Studies in the Mathematical Theory of Inventory and Production* 1: 319.

Kim, M., and S.J. Lennon. 2011. "Consumer response to online apparel stockouts." *Psychology & Marketing* 28 (2): 115–144.

Lalwani, C.S., S.M. Disney, and D.R. Towill. 2006. "Controllable, observable and stable state space representations of a generalized order-up-to policy." *International Journal of Production Economics* 101 (1): 172–184.

Lariviere, M.A., and E.L. Porteus. 1999. "Stalking information: Bayesian inventory management with unobserved lost sales." *Management Science* 45 (3): 346–363.

Lee, H.L., V. Padmanabhan, and S. Whang. 1997. "Information distortion in a supply chain: The bullwhip effect." *Management Science* 43 (4): 546–558.

Li, Q., and S. M. Disney. 2017. "Revisiting rescheduling: MRP nervousness and the bullwhip effect." *International Journal of Production Research* 55 (7): 1992–2012.

Lin, J., M.M. Naim, L. Purvis, and J. Gosling. 2017. "The extension and exploitation of the inventory and order based production control system archetype from 1982 to 2015." *International Journal of Production Economics* 194: 135–152.

Lin, J., M.M. Naim, and V.L.M. Spiegler. 2019. "Delivery time dynamics in an assemble-to-order inventory and order based production control system." *International Journal of Production Economics* in press.

Morton, T.E. 1969. "Bounds on the solution of the lagged optimal inventory equation with no demand backlogging and proportional costs." *SIAM Review* 11 (4): 572–596.

Nadarajah, S., and S. Kotz. 2008. "Exact distribution of the max/min of two Gaussian random variables." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 16 (2): 210–212.

Nagatani, T., and D. Helbing. 2004. "Stability analysis and stabilization strategies for linear supply chains." *Physica A: Statistical Mechanics and its Applications* 335 (3-4): 644–660.

Ohno, T. 1988. *Toyota Production System: Beyond Large-Scale Production.* Productivity Press, Portland.

Ponte, B., X. Wang, D. de la Fuente, and S.M. Disney. 2017. "Exploring nonlinear supply chains: the dynamics of capacity constraints." *International Journal of Production Research* 55 (14): 4053–4067.

Potter, A., and S.M. Disney. 2010. "Removing bullwhip from the Tesco supply chain." In *Production and Operations Management Society Annual Conference*, 19pp.

Priore, P., B. Ponte, R. Rosillo, and D. de la Fuente. 2019. "Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments." *International Journal of Production Research* 57 (11): 3663–3677.

Rudi, N., and D. Drake. 2014. "Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior." *Management Science* 60 (5): 1334–1345.

Schneeweiss, C.A. 1974. "Optimal production smoothing and safety inventory." *Management Science* 20 (7): 1122–1130.

Sobel, M.J. 2004. "Fill rates of single-stage and multistage supply systems." *Manufacturing & Service Operations Management* 6 (1): 41–52.

Towill, D.R. 1991. "Supply chain dynamics." *International Journal of Computer Integrated Manufacturing* 4 (4): 197–208.

Van Donselaar, Karel H, and Rob ACM Broekmeulen. 2013. "Determination of safety stocks in a lost sales inventory system with periodic review, positive lead-time, lot-sizing and a target fill rate." *International Journal of Production Economics* 143 (2): 440–448.

Van Woensel, T., K. Van Donselaar, R. Broekmeulen, and J. Fransoo. 2007. "Consumer responses to shelf out-of-stocks of perishable products." *International Journal of Physical Distribution & Logistics Management* 37 (9): 704–718.

Verbeke, W., P. Farris, and R. Thurik. 1998. "Consumer response to the preferred brand out-of-stock situation." *European Journal of Marketing* 32 (11/12): 1008–1028.

Wang, X., and S.M. Disney. 2016. "The bullwhip effect: Progress, trends and directions." *European Journal of Operational Research* 250 (3): 691–701.

Wang, X., S.M. Disney, and J. Wang. 2014. "Exploring the oscillatory dynamics of a forbidden returns inventory system." *International Journal of Production Economics* 147: 3–12.

Wikipedia. 2020. "Truncated normal distribution." https://en.wikipedia.org/wiki/Truncated_normal_distribution, May.

Womack, J.P., and D.T. Jones. 1997. *Lean thinking: Banish waste and create wealth in your corporation*. Simon and Schuster UK Ltd.

Zipkin, P. 2008a. "Old and new methods for lost-sales inventory systems." *Operations Research* 56 (5): 1256–1263.

Zipkin, P. 2008b. "On the structure of lost-sales inventory models." *Operations Research* 56 (4): 937–944.