Forecasting for remanufacturing: the effects of serialization

Thanos Goltsos, Aris Syntetos and Erwin van der Laan

Remanufacturing operations rely upon accurate forecasts of demand and returned items. Return timing and quantity forecasts help estimate net demand (demand minus returns) requirements. Based on a unique dataset of serialized transactional issues and returns from the Excelitas Group and one of their defense contractors, Qioptiq, we assess the empirical performance of some key methods in the area of returns forecasting. We extend their application (for net demand forecasting), by considering that demand is also subject to uncertainty and thus needs to be forecast. Information on remanufacturing costs allows for an evaluation of the inventory implications of such forecasts under various settings. A foray into the literature on information technologies enables a discussion on the interface between information availability and forecast accuracy and utility. We find that serialization accounts for considerable forecast accuracy benefits, and that the accuracy of demand forecasts is as important as that of returns. Further, we show how the combined returns and demand forecast uncertainty affects the inventory performance. Finally, we identify opportunities for further improvements for the operations of Qioptiq, and for remanufacturing operations in general.

Keywords: Remanufacturing; Forecasting; Simulation; Empirical data.

1. Introduction and motivation

In the product recovery hierarchy, remanufacturing is placed towards the top tier as it encourages the re-use of both parts and products with minimal additional input of raw materials, thus playing a prominent role in the circular economy. It may be described as "*the transformation of used products* (referred as cores)*, consisting of components and parts, into products that satisfy exactly the same quality and other standards as new products*" (Guide and Jayaraman, 2000, p. 3780). Its value is estimated at \$43 billion in the US (USITC, 2012) and EURO 30 billion in the EU (ERN, 2015). While many operations are similar to those of traditional manufacturers, remanufacturers have to deal with extra uncertainties stemming from the core acquisition and transformation process (Guide and Jayaraman, 2000; Goltsos et al., 2018).

In traditional manufacturing, the emphasis is on forecasting the (independent) demand timing and quantity for final products, and then 'exploding' such forecasts into (dependent) demand requirements for materials and parts through a *predetermined/static* Bill of Materials (BOM), but also on a set of largely *constant* operations, through a similarly static Bill of Operations (BOO).

In remanufacturing however, the emphasis shifts to forecasting *net* demand, which is the difference between demand and returns (i.e. that part of the demand that cannot be satisfied through the remanufacturing of returned cores). Net demand then drives replenishment for new items (from Original Equipment Manufacturers, OEMs), although it can also be met by cores (from core brokers, that still need to be remanufactured), that together with the remanufactured returns ultimately service total (customer) demand. That is, although demand forecasting remains equally relevant, the timing and quantity[1] of the returns also needs to be forecast, compounding to a dual-source uncertainty.

Unlike demand, the stream of returned items is not an independent random variable, but rather correlates with variables such as past sales (Goh and Varaprasad, 1986), the installed base (number of items in the market), lead time sales and returns (Kelle and Silver, 1989a), and at times the stage in a product's life cycle (van der Laan and Salomon, 1997). It is this relationship of returns to past sales that is integral to returns forecasting. This means that univariate approaches, such as those traditionally used in time series demand forecasting, are susceptible to underperformance (Kiesmüller and van der Laan, 2001). It also means that the performance of the returns forecasting is subject to the level of detail and accuracy of information available about this relationship.

---

[1] For both demand and returns, the rate (quantity per time unit) should be distinguished from its constituent elements, i.e. the timing and quantity per occurrence, when the relevant series are intermittent in nature, as it often happens in practice. Two streams with the same rate may have quite different patterns in terms of timing and quantity per occurrence. As such, it is the constituent elements that are of interest for characterising the relevant series and forecasting purposes, as opposed to the rate itself.

In addition, products are of unknown condition (quality) when eventually returned, which is an important issue in its own right (Guide and Jayaraman, 2000). This implies that the probability of returns needs to be used in conjunction with the probability of cores being remanufacturable, or else the assumption that returned cores can be used towards satisfying demand does not hold. It is the probability of a *remanufacturable return* that is of interest to remanufacturing operations. Any quality induced variability in the remanufacturing lead times, while being problematic, opens up scheduling opportunities for distributing operations. For example, a planner would like to know which of the returns over the (forthcoming) lead time can be used (remanufactured in time) towards satisfying lead time demand. A first step towards that is to move away from aggregate and into per period lead time forecasts, as we later do.

In is evident then, that in the context of remanufacturing operations, forecasting entails the simultaneous challenges of (see also Guide, 2000): 1. forecasting of timing and quantity of demand, 2. forecasting of timing and quantity of returns, and 3. forecasting of the quality status of the returns (at the very minimum how many returns are useable, i.e. the number of remanufacturable returns). Some research, reviewed in the next section, has been conducted to address the second problem. In particular, four methods proposed by Kelle and Silver (1989a) have attracted attention both in theoretical and empirical domains. The methods assume varying degrees of available information regarding the relationship between demand and returns: from no information to serialized matching of a stock keeping unit's (SKU) individual issues and returns (and thus the time each spent with the customer). It is this latter level of information that has been missing in previous empirical studies, and this constitutes the main point of departure of our analysis from previously offered results. The availability of such information can be achieved through a number of ways as we will later see.

An integral part of the methods' application is the distribution of the time taken for an item to return after being sold to a customer (time-to-return distribution). Such methods are not specific

to remanufacturing but rather they are relevant in any circular economy or reverse supply chain context. Important as they are, they rely on knowing with certainty the probability that an item will eventually return, the time-to-return distribution, and the demand rate, with quality taken into account only implicitly. That is, the four methods of Kelle and Silver (1989a) ignore challenges 1 and (to an extent) 3 above, whilst challenge 2 is addressed assuming some 'perfect' information. The same is, up to a great extent, true for follow up work conducted on these methods. Further, empirical evidence on their practical validity and utility is lacking. Some notable exceptions and how these contributions differ from our work, are discussed in the next section of the paper. The focus of this paper is on challenges 1 and 2. What is of particular interest is the impact of serialization on challenge 2: how can serialized data be conductive to characterizing the relationship between past sales and returns, and what benefits does this translate to (in terms of accuracy and utility)?

1.1 Contribution and organization of the paper

We assess the empirical behavior of returns and characterize the time-to-return distributions (never addressed before in the literature) by means of using a unique dataset of serialized transactional issues and returns from the Excelitas Group and one of their defense contractors, Qioptiq. The company is a defense and aerospace integrated logistics supplier (managing flows and repairs), operating within various state military supply chains globally. The product line we are concerned with consists of 15 electronic night vision, thermal (infrared, IR) and image enhancing visual aid equipment for the dismounted soldier (head/weapon-mounted, handheld).

The dataset allows us to be the first to test the methods of Kelle and Silver (1989a) in empirical terms, but also to explore the effect of serialization. We extend the methods' application (for net demand forecasting) by considering that demand is also subject to uncertainty and thus needs to be forecast, and by means of relaxing some further unrealistic assumptions of the forecasting

procedures. Information on remanufacturing costs allows for an evaluation of the inventory implications of the forecasts. A foray into the literature on information technologies enables a discussion on the interface between information availability and forecast accuracy and utility. As we will see in the literature review, information availability is integral to forecasting returns.

Our evaluation is based on simulating ex ante the behavior of various forecasting methods, and where appropriate their inventory implications. That is, we assess what would have happened if the methods under concern were to be applied on Qioptiq's data. Beyond the dataset, this is an extremely 'rich' case, containing information on the actual structure and flows of the supply chain, current and future contractual arrangements, and a complete specification of operations. This enables the contextual interpretation of our empirical findings and their logical association with what might have happened in other industries.

We find that serialization accounts for sizable forecast accuracy benefits, both for low and high volume items, and that the accuracy of demand forecasts is as important as that related to the returns. Further, we show how the combined returns and demand forecast uncertainty affects the inventory performance. Finally, we identify opportunities for further improvements for the operations of Qioptiq, and for remanufacturing operations in general.

The remainder of the paper is organized as follows. In §2 we review the literature on returns forecasting and information availability, and point out the gaps that are addressed in our work. §3 presents the case organization, along with the data available for the purposes of this research and an empirical characterization of the returns' behavior and time-to-return distributions. The simulation is presented in §4 along with a detailed description of the forecast methods and inventory management process we consider, and the scenarios we evaluate. The results are analyzed and discussed in §5, where insights are offered to remanufacturing managers. We close in §6 with conclusions and extensions.

2. Theoretical background

Literature on forecasting in a remanufacturing context is scarce, although there are some publications that deal with returns in general. Such work can be used (or adapted) to forecast the rate of returns in a remanufacturing context (Wei et al., 2015).

The returns forecasting literature largely focuses on net demand forecasts of items whose restoration process is, or is often assumed to be, trivial. The re-manufacturability of a return is represented as a binary possibility: it can either happen (remanufacturable return) or not (lost or beyond-economical-repair, BER). Further, demand itself is assumed known and largely not subjected to estimation error, ignoring one end of the demand-returns dual-source uncertainty. Net demand forecasts then inform the procurement of new items (from OEMs), and in some cases cores (from core brokers), to cover the demand that cannot be served from remanufacturable returns. As such, the majority of relevant studies' main focus is on estimating the net demand over some (constant) OEM manufacturing lead times.

Our independent review agrees with two literature review papers conducted by Govindan et al. (2015) and Wei et al. (2015). The former conducted a comprehensive review of relatively recent (2007-2013) reverse logistics and Closed-Loop Supply Chains (CLSC) publications. They reported forecasting to be a 'missing subject', and went on to suggest that only 3 papers deal with this issue out of 382 papers reviewed.

A similar survey was conducted by Wei et al. (2015) but focusing on core acquisition management in remanufacturing in particular. The authors reviewed a set of 88 papers up to 2014 and categorized them in 5 classes (the returns forecasting class containing only Clottey et al., 2012). They mentioned that there are papers dealing with forecasting returns in general that could be adopted in a remanufacturing setting, but opted to exclude them for lack of

remanufacturing considerations. They agreed with Govindan et al. (2015) by concluding that forecasting the returns of cores is under-researched.

In the remainder of this section, we independently (to the above studies) review the returns forecasting literature, before we consider some other adjacent approaches close to our research. We then move to discuss information availability and serialization, in an attempt to gain insights into their interface with returns forecasting. We close with a summary of our literature review where we point out the gaps that are addressed in this work.

2.1 Forecasting returns: methods

One of the first attempts to forecast returns by correlating them to past issues/sales can be attributed to Goh and Varaprasad (1986), who computed life-cycle parameters of returned containers. They modelled returns as a transfer function of past sales and employed a Box-Jenkins approach to estimate the parameters of the distribution of returns, to forecast the lead time returns. They used an empirical dataset of soft drink bottles monthly issues and returns to assess the implications of their proposed method.

Criticism of the above work includes that it only produces forecasts (with no consideration of forecast errors or forecast utility) and is associated with no visibility of whatever might happen during the lead time (items sold and returned within the lead time, Kelle and Silver, 1989a). Another problem stems from the fact that the approach is very demanding in terms of past data requirements, and also not very easy to update in light of new data (Clottey et al., 2012).

The last criticism also stands for the work conducted by Kelle and Silver (1989a), who proposed four procedures to forecast lead time multinomial returns for reusable containers, each being associated with different levels of information. The availability of such information acts as a constraint to the employment of the methods (in industrial and academic settings alike), an issue discussed later in this section.

Method 1 (M1) employs only the mean and variance of the demand, along with an overall probability of eventual return, assuming in effect that every sale is probabilistically accompanied by a return. Method 2 (M2) requires the past issues per period (whatever that time bucket might be) and the time-to-return distribution, somehow derived from managerial judgement or by tracking a subset of the items in circulation. In addition to the requirements associated with M2, Method 3 (M3) also requires the per-period per past issue realized returns. M3 is utilizing the most detailed information, by tracking the yet-to-return items and updating their return probabilities, and acts as a benchmark. Finally, Method 4 (M4), is presented as a more practical alternative approach to M3; it requires per period issues and per period realized returns to derive the time-to-return distribution. Toktay (2003) defined the requirements of M4 (per period issues and returns volume) as *period-level information* and those of M3 (sales and returns tracked on an individual serial number basis) as *item-level information*. M3 improves on M2 by taking into account the true number of in-circulation cores. M4 improves on M2 by a term depending on the difference between observed and expected values of the recent aggregate returns, in effect providing a way to measure the returns distribution when only period-level information is available. We circumvent the problem of estimating the distribution for M2/M4 via directly fitting various distributions proposed in the literature to our empirical returns data, which enables us to assess the validity of the relevant assumptions.

Kelle and Silver (1989a) compared the different methods by means of mean absolute deviation and mean cost penalties (inventory holding and backorder costs in a base stock policy) by simulating performance on the dataset of Goh and Varaprasad (1986). Importantly, Kelle and Silver (1989a) were the first to recognize (and model) that some issues might return sooner than the lead time for a new product[2]. In a subsequent technical note, Kelle and Silver (1989b)

---

[2] Imagine a forecasting lead time of two periods, and a non-zero probability of an item returning one period after it has been issued. It follows that some of the lead time expected returns for the second period will originate from

determined an optimal purchasing policy under the assumption of normally distributed lead time net demand, a common assumption in much of the literature and a very convenient one since net demand may indeed assume negative values.

The four methods of Kelle and Silver (1989a) were revisited by de Brito and van der Laan (2009) who tested them under the presence of the more realistic scenario of imperfect information, with simulated data. Similar to Kelle and Silver (1989a), they employed a base stock model to measure cost penalties under varying levels of imperfect information. The authors reported that under such circumstances the most detailed method (M3) does not always outperform the more robust one (M2). They also suggested improvements for M1 and M2 that lead to an overall better performance of those methods. We also discuss, in the results analysis section, how performance of M2 may be improved, and the circumstances under which M3 performs best.

Toktay et al. (2000) built on Kelle and Silver (1989a) to develop a geometric and a negative binomial distributed lag model (DLM) with Bayesian estimation to capture the returns of Kodak single-use cameras. They distinguished between period-level information and item-level information, and constructed a 'heuristic dynamic aggregate base-stock policy' to make use of dynamically updated information of customer use.

Parameters used in the simulation were set as to be '*roughly representative*' of the Kodak environment. The empirical dataset at their disposal consisted of 22 months of 'period-level' (aggregate sales and returns) information of a single-use Kodak camera. Unknown initial conditions were estimated through the maximum likelihood method. We also face this problem, but the length of our dataset and full characterization of the empirical returns distribution in

---

expected issues at the first. It also follows that the accuracy of the returns forecast for that second period is subject to the accuracy of the demand forecast for the first.

conjunction with the simulation approach employed, allow for sufficiently long initialization/ warm-up periods that diminish the effects of the unknown initial conditions.

Toktay et al. (2000) reported that longer lead times have a negative impact on the precision of the models, and that serialized information is particularly helpful for items associated with low demand volumes (such as the majority of items considered for the purposes of our research). We return to this issue in §5, to show that our findings support the first claim, but also the fact that serialized information accounts for considerable benefits for both low and high volume items.

Clottey et al. (2012) further built on the above to develop an exponential delay DLM with Bayesian updating, to forecast the returns of an electronics hybrid remanufacturer. The formulation of their DLM is consistent with a single period model such as the newsvendor. Various cost and other parameters were either informed from real data or set to match that of the company through discussions with management, and run simulations on theoretically generated data (both matched and mismatched with the model specifications). They discussed where their method outperformed that of Toktay et al. (2000) in terms of total costs and mean absolute percentage errors (MAPE). Clottey and Benton (2014) similarly developed a Gamma representation in a method accommodating for long return lags.

It is worth mentioning here, that Kelle and Silver's (1989a) M4 and its adaptations by Toktay et al. (2000), Clottey et al. (2012) and Clottey and Benton (2014) are estimators of the parameters of the returns distributions assumed in the respective studies. They do that by employing a covariance matrix, the size of which is equal to the square of the longest number of periods an item has spent with a customer. In our case, employing these procedures is computationally infeasible due to much longer times-to-return (time the items spent with a customer). Nevertheless, the richness of our dataset renders these computations redundant since it allows

us to fully characterize the empirical distributions (as opposed to estimating parameters of hypothesized ones). This enables us to elegantly overcome problems of application and discuss the relevance of these assumptions to our dataset.

Finally, Carrasco-Gallego and Ponce-Cueto (2009) employed a dynamic regression model to forecast the returns in a reusable containers' CLSC in the Liquid Petroleum Gas (LPG) industry. They concluded that forecasting returns is of limited value under a direct replacement policy (i.e. when every sale is accompanied by a return), which is expected.

2.2 Forecasting returns: further approaches

Kumar and Yamaoka (2007) employed a system dynamics approach to study the Japanese automotive industry CLSC. They represented returns as a simple delay function of past sales (11 years' time lag). Hanafi et al. (2008) employed a 'Fuzzy colored Petri Net' forecasting method taking into account income, population, age, and past sales to estimate a return rate and location, and then suggested an optimal collection program for end of life electronics.

Zhou and Disney (2006) assumed correlation between stationary independent identically distributed (iid) demand and returns, when the time-to-return is exponentially distributed. They highlighted the positive effect of reverse logistics operations on inventory variance and the bullwhip effect when compared to the equivalent OEM operations. Their findings agree with those by Kiesmüller and van der Laan (2001), who investigated the effect of neglecting the dependency between demands and returns, in that returns should be modelled as a function of past sales. Conversely, Muckstadt and Isaac (1981) modelled returns as a Poisson process, independent from the demand generating process, and reported the reverse operations to have an adverse effect on the total variance of their system. Hosoda et al. (2015) investigated information sharing with regards to product returns from the remanufacturer to the OEM in a hybrid CLSC setting. They modelled demand and returns as correlated white noise iid processes.

They distinguished between the probability of a return and the probability of remanufacturable returns by introducing a random yield (independent of time or return volume) in the remanufacturing process. They showed that (in most cases) sharing the return and yield (i.e. remanufacturable returns) information may be beneficial to the manufacturer, especially in the case of longer lead times.

Zhou et al. (2016) attempted to approach the problem of remanufacturing forecasting holistically, and mentioned that they might have been the first to do so. They employed the 'Graphical Evaluation and Review Technique' (GERT) to develop a method to estimate the probability of return, time and quantity for cores, salvageable parts, components and materials.

2.3 Information availability and serialization

Item-level, serialized data can be acquired in a number of ways. At one end, a barcoded or manual input information system can allow for the tracking of items at the serial number level. At the other, the 'Internet of Things (IoT)', which refers to a network structure of uniquely identifiable products (things), enables automatic monitoring and tracking (Ondemir and Gupta, 2014). This can be facilitated by the use of tags such as 'RFID' (Radio Frequency IDentification), whereby product information can be stored in radio antennae bearing circuits. These circuits are attached to conveyances and ultimately to boxes, pallets or individual items, enabling the reading (and often storing) of information as they move through the supply chain. In their most advanced form, 'Product Embedded Information Devices' (PEID) include sensors that can track a number of parameters of the product's use and its environment (Kiritsis, 2011).

This information can be of great benefit for the management of the dual-source uncertainty faced by remanufacturers. Such technologies are integral to the shifts towards 'Industrial IoT' in the US, and Industry 4.0 in Germany (Fang et al., 2016). As a result of such initiatives, the serialized data we are concerned with in this work are becoming increasingly available, and the need for

an empirical evaluation of their utility is now pressing. Further, and although serialized information can be acquired in a straightforward manner nowadays, it is dangerous to assume (even in the case of PEID systems) perfect knowledge; data inaccuracies and changes over time mean that estimation of relevant parameters is still appropriate.

Luttropp and Johanson (2010) explored the impact of RFID on a number of beyond-economical-repair electrical and electronic items in recycling reverse supply chains. Based on numerous case studies and projects in such supply chains, they suggested an array of valuable information that should be stored on RFID tags, to help identify residual value in such items. Parlikad and McFarlane (2010) conducted a similar exercise in remanufacturing of end of life vehicles, to enable cost effective auto part recovery. Fang et al. (2016) developed an IoT (PEID) enabled model, showing how such information can be used for the optimization of the integrated procurement, production and recovery, and acquisition problem, across the different life cycle stages of a modular remanufacturable product.

Lee and Ozer (2007) developed an RFID-enabled reverse supply chain model of a remanufacturer. They distinguished between three scenarios: the 'naïve', where the returns channel is ignored, the 'smart', where properties of the returns channel are statistically derived, and the 'RFID', where the returns channel is visible. Karaer and Lee (2007) employed three very similar scenarios, called 'no visibility-naïve', 'no visibility-enlightened' and 'full visibility'. They assumed demand and returns to be normally distributed and reported the cost reductions from moving from the naïve to the RFID-enabled paradigm. They concluded that "*the smarter the [re]manufacturer is or less volatile the return flow is, the less benefit visibility is likely to bring to the system*, p. 640". This is the logical equivalent of one of our own recommendations which we discuss later.

Of course, visibility provides item-level information that can link returns to past sales as M3 requires. The no visibility scenarios can at worst (naïve) be the equivalent of M1 or applying a simple time series forecasting method on returns, thus ignoring the relationship between past sales and returns. At best (enlightened/advanced), they could be seen as the equivalent of M2, with a (somehow) well selected and fitted returns distribution. We return to this point in our discussion of our forecasting accuracy results.

2.4 Summary

'*The challenges for forecasting of product returns are mainly from two sources: lack of quality/credible data and unproven assumptions*' (Liang et al., 2014, p. 3). It is safe to conclude that returns forecasting remains open for contributions. Most work in this area relies on some convenient assumptions. None of the above studies have employed real item-level data, and most use parameters set through '*discussion with managers*' for a very limited number of SKUs. The current is the first study, to the best of our knowledge, that relies upon actual empirical distributions of returns, thereby assessing the validity of assumptions made in previous studies.

Moreover, although the need to (simultaneously) forecast the associated demand has been explicitly recognized, the impact of the relevant forecast error has not been addressed. We argue that although remanufacturing companies should never disregard the estimation of returns, the importance of demand forecasting is such that it deserves equal consideration.

The literature on information availability argues for the value of visibility in reverse supply chains, lending itself naturally to our discussion about the level of information employed in returns forecasting. As such, we believe our results are also relevant to this literature. Visibility then can be seen as the enabler of the Kelle and Silver (1989a) methods: 'no visibility-simple' using no information (M1), 'no visibility-advanced' using period-level information (M2), and 'full visibility' using item-level information (M3).

3. Case organization and descriptive empirical analysis

Excelitas is a global parent company of various defense and commercial contractors. Qioptiq, one of their contractors, is a global defense and aerospace Integrated Logistics Supplier (managing flows and repairs), operating within various commercial and state military supply chains (SCs) globally. We focus on their remanufacturing operations for one of such state military SCs, distinct of any others, which makes the operating scenario that of a pure remanufacturer.

The company deals with 15 electronic night vision, thermal (IR) and image enhancing visual aid equipment for the dismounted soldier (head-mounted, weapon-mounted and handheld). An as-good-as-new product is issued to a Ministry of Defense (MoD) facility, with very short delivery lead times. Because of these very short delivery times, the company operates in a (re)make-to-stock setting. That product can be subsequently sent to a number of other units (without visibility from the company), before eventually returning as a used core, essentially operating as a mix of first and second-hand-market (Zhou et al., 2016).

As depicted in Figure 1 the used products (cores) will upon arrival sit in an inventory of returned cores (1), until picked up to be remanufactured. Replenishment orders for the cores inventory are placed to an outside core broker who delivers cores of unknown but similar quality. This describes a pull system with core (broker) replenishment. The core broker is assumed to have infinite supply. Remanufacturing may be represented as a 3-stage process of assessment, repair and/or cleaning, and testing. Upon successful completion of the process (i.e. when the item is brought to an as-good-as-new state), it will sit in the serviceables inventory (2). If the core is deemed BER, the item is written off and cannibalized for useful parts to be utilized in remanufacturing of other cores (3).
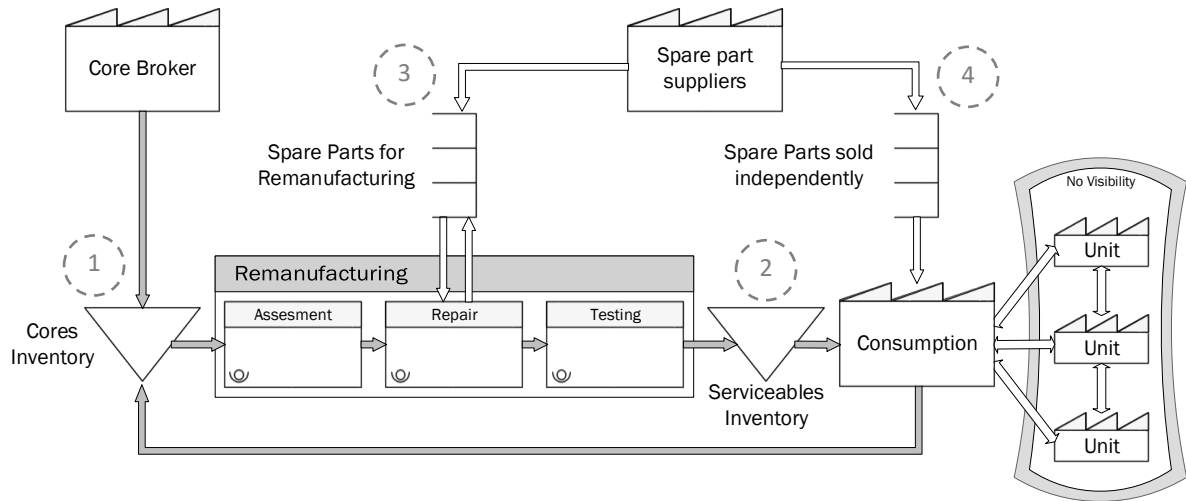
Figure 1: Simplified supply chain of the case organization. Focus of study represented by grayscale arrows.

There is a set of spare parts (procured from suppliers and salvaged from BER cannibalization) used for the remanufacturing of cores (3), a subset of which (not in physical terms but rather in SKU number terms), can be sold independently (4) for units to be able to carry out minor repairs in the field. The latter mostly consists of simple items such as caps, straps, etc., and does not affect the number of items that are with the end customer. We note that the spare part inventories (3) and (4) represent distinct inventories. However, important as they are for Qioptiq's operations, these are of no interest to this study (and are represented with white arrows to denote this, Figure 1).

Due to the military nature of the supply chain, the probability of a return taking place is close to 100%, if only to be written off in case of extensive damage (BER). The probability of a return being remanufacturable varies. It averages, across all 15 cores and their returns, to 95%. There are a number of industries that exhibit similarly high or higher (average) probabilities of returns, viz. automotive car parts (van der Laan and Salomon, 1997); reusable Coca-Cola soda bottles (Goh and Varaprasad, 1986); reusable liquefied petroleum gases containers (Carrasco-Gallego and Ponce-Cueto, 2009). It is easy to imagine leasing contracts such as those for airplane engines having comparable behavior.

Further, the company operates with targeted service levels of 95%. The current lead time for both the cores and serviceables inventories replenishment is 2 months (corresponding to the core broker and remanufacturing lead times, respectively). Due to forthcoming changes in their contracts, targeted service levels are to be revisited and core broker lead times may also increase. Importantly, and although the probability of returns will remain close to 1, probabilities of remanufacturable returns are also expected to alter (for a number of reasons, including changes in the very definition of what constitutes remanufacturability or BER). The above empirically motivates some important sensitivity and robustness analysis to be performed with 'control parameters' (lead time, service level and probability of remanufacturable returns) values other than the current, rendering our results relevant to other industries too.[3]

In terms of the company's inventory forecasting practices, they universally apply (to both demand and returns) a Simple Exponential Smoothing (SES) method, with a fixed smoothing constant value, $\alpha = 0.4$. The forecasts are then translated to inventory replenishment decisions through a periodic order-up-to level, (T, S), policy, where T is the inventory review period ($= 1$ month) and S the order-up-to level ($= 4$ x SES forecast).

The company provided us with a partially redacted raw output of their information system in the form of transactional data. They also provided instructions of various fields and data entry nomenclature, critically including identifiers that allow the separation of 'bulk' moves or deployments (e.g. training exercises) to 'one-for-one' returns (orders for replacement of broken equipment). We filtered the data to only include the latter, to avoid sector-specific phenomena (e.g. training exercises which appear as large spikes in both demand and returns and, if taken into account, would heavily skew the time-to-return distributions). This process led to the

---

[3] It is also important to note that the above do not collectively describe a dynamic environment whereby key parameters are subject to constant change. Rather, these are forthcoming one-off contractual changes, that are employed here as motivation for further exploration of realistic scenarios through the control parameters.

alteration of 18, out of a total of 1490, monthly demands (and respective returns) across the 15 equipment. This was done to strengthen the generalizability of our results to an expanded number of remanufacturing contexts.

In terms of the empirical dataset constructed for the purposes of this research, it contains from 4 up to 11 years (across the 15 equipment) of transactional data that include:

- Demand (and actual respective issues) of serviceable (remanufactured) end products (good as new equipment);

- Core (used product) returns, linked with the issues at a serial number level, i.e. timestamped/item-level information;

- Remanufacturing dates at a serial number level, distribution of failure rates, costs and lead times, allowing us to assess inventory implications.

3.1 Characterization of the demand and return process

In this section, we provide some empirical evidence on the behavior of the demand and return processes, by means of presenting some key descriptive statistics and identifying appropriate statistical distributions for the respective data.

3.1.1 Demand and return arrivals and occurrences

Both the demand and returns series are intermittent in nature, i.e. demand (returns) occurring periods are interspersed by a varying number of periods with no demand (returns) at all. Further, demand (returns) when they occur there may be associated with a variable size, leading to what is often termed as 'lumpy' behavior[4] (Cattani et al., 2011).

---

[4] Batching can be a contributory factor to lumpiness, especially in the returns process. In our case, batching of core returns is less of an issue since it happens across different equipment (not just the ones we deal with for the purposes of this research), diminishing its effects.

There is a growing body of literature in the area of intermittent demand forecasting. Unlike fast demand items, intermittent demand SKUs are very difficult to forecast. This is because they are associated with a dual source of uncertainty: demand arrivals (how often demand occurs) and demand sizes (how large and variable demand is, when it does occur) (Syntetos et al., 2005).

Intermittent demand patterns are known to be very common in military settings, but also in other industries, such as oil, telecom and electronics. Intermittent demand SKUs also monopolize the stock bases in any environment where demand is generated because of failures, such as the after sales industry; spare parts have been repeatedly shown in empirical studies to be (almost invariably) intermittent in nature (Lengu et al., 2014).

The intermittent behavior of the returns series is an interesting finding. Although it is conditioned to the particular context, it is not unreasonable to expect that returns will occur in an intermittent fashion, along the lines of the failure-driven demand generating process for spare parts. However, it seems that the returns are less intermittent than their respective demand time series, especially for bigger demand sizes. Importantly, this is (to the best of our knowledge) the first attempt appeared in the literature to empirically characterize the behavior of returns.

In Table 1 we present some key descriptive statistics for demand. Non-disclosure agreements specify that we may not present such information on an individual equipment basis. Rather, summary information is presented across all 15 equipment. The mean and standard deviation of the demand rate, inter-demand intervals and demand sizes (on the periods where demand occurs) are calculated over the entire history for each series individually, and then summarized through some key quantities across all series. We do the same for the returns series in Table 2. Please note that in both tables, figures read horizontally might or might not refer to the same SKU and a value in Table 1 might or might not refer to the same SKU as the relevant value in Table 2.

	Demand per period		Demand intervals		Demand sizes	
	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
min	*0.34*	*0.74*	*1.00*	*0.00*	*1.52*	*0.81*
0.25%	*2.39*	*4.17*	*1.13*	*0.38*	*4.29*	*4.82*
median	*5.57*	*11.89*	*1.57*	*0.94*	*10.20*	*12.09*
0.75%	*14.59*	*20.63*	*2.03*	*1.45*	*19.05*	*22.21*
max	*115.29*	*117.90*	*4.46*	*4.43*	*115.46*	*117.96*

Table 1: Summary descriptive statistics of (monthly) demand across 15 equipment.

	Returns per period		Returns intervals		Returns sizes	
	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
min	*0.33*	*0.59*	*1.00*	*0.00*	*1.19*	*0.49*
0.25%	*2.43*	*1.76*	*1.00*	*0.00*	*2.76*	*1.63*
median	*5.53*	*3.45*	*1.04*	*0.18*	*5.76*	*3.43*
0.75%	*14.38*	*4.62*	*1.15*	*0.39*	*14.37*	*4.63*
max	*114.18*	*21.80*	*3.64*	*3.22*	*114.19*	*21.82*

Table 2: Summary descriptive statistics of (monthly) returns (including BER) across 15 equipment.

It is important to note that despite the prominence of intermittence, three of the equipment are actually fast, with the remaining twelve following a slow, intermittent demand behavior.

Although our work emphasizes extrapolation rather than empirical fitting, it is worthwhile to investigate possible return intervals distributions, because little empirical evidence exists. We do so in Figure 2 where we plot squared skewness versus kurtosis for our data and candidate distributions. The exponential distribution is a popular choice for modelling return inter-arrival times (de Brito and Dekker 2003), or the geometric distribution as its discrete time analogue, but this does not seem to be a likely candidate for our data (see Figure 2). For some SKUs the negative binomial may work, but the (discretized) Beta distribution seems more appropriate as it has more flexibility in terms of skewness and kurtosis.[5]

---

[5] This is confirmed by distribution fitting analysis, the presentation of which is beyond the scope of this paper.
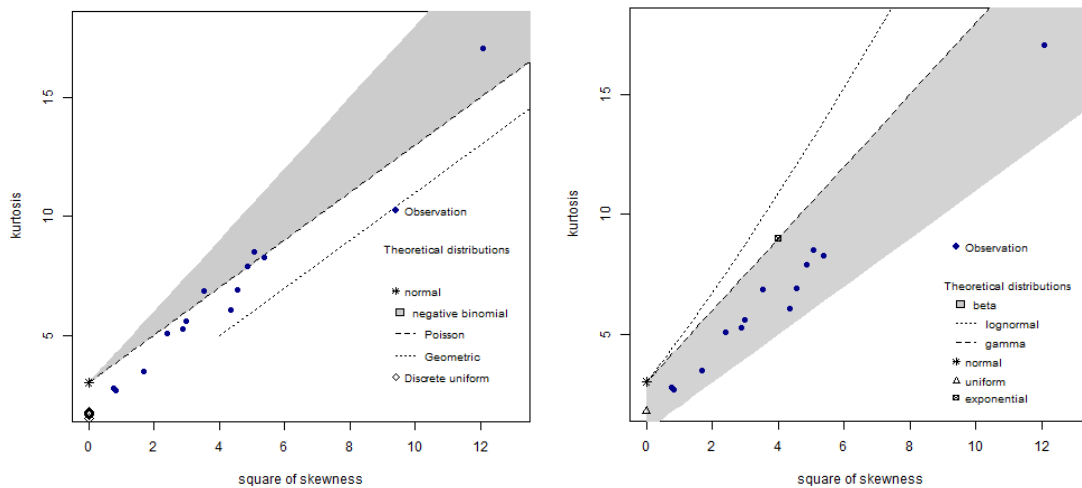
Figure 2: Squared skewness versus kurtosis for return intervals. The exponential distribution does not seem a likely candidate. The Beta distribution seems plausible. Plot is adapted from the Cullen and Frey graph generated with the R package "fitdistRplus". The two fastest moving SKUs are excluded from the graphs as all but a couple of intervals are 1 (and therefore exhibit very high kurtosis and skewness).

3.1.2 Time-to-return distribution

Because for some of the returns forecasting procedures it is key to have a reasonable estimate of the time-to-return distribution, here we attempt to determine which distributions are good candidates. In the literature, some suggestions as to what distributions to use are made, but not much empirical assessment has been conducted so far on this issue. Kelle and Silver (1989a) use uniform and some non-standard distributions to illustrate their forecasting methods, de Brito and van der Laan (2009) use uniform and geometric, whereas Toktay et al. (2000) find some empirical evidence for the geometric distribution. A geometric distribution reflects the probability of needing a certain number of Bernoulli trials before success. An extension of the geometric distribution is the negative binomial distribution, sometimes referred to as the 'kth order inter-arrival time of a Bernoulli process'. There is no theoretical foundation for Poisson distributed times to return, but we use it for benchmarking purposes. It may also be worthwhile to benchmark with some continuous distributions. Figure 3 shows that our data generally displays high skewness together with thick tales (kurtosis). The figure suggests that the Beta and, less so, the Gamma distribution may be good alternatives to the discrete ones. The negative

binomial distribution seems less appropriate, but we have to be careful as we are dealing with truncated data (there are no observations beyond a certain time horizon).
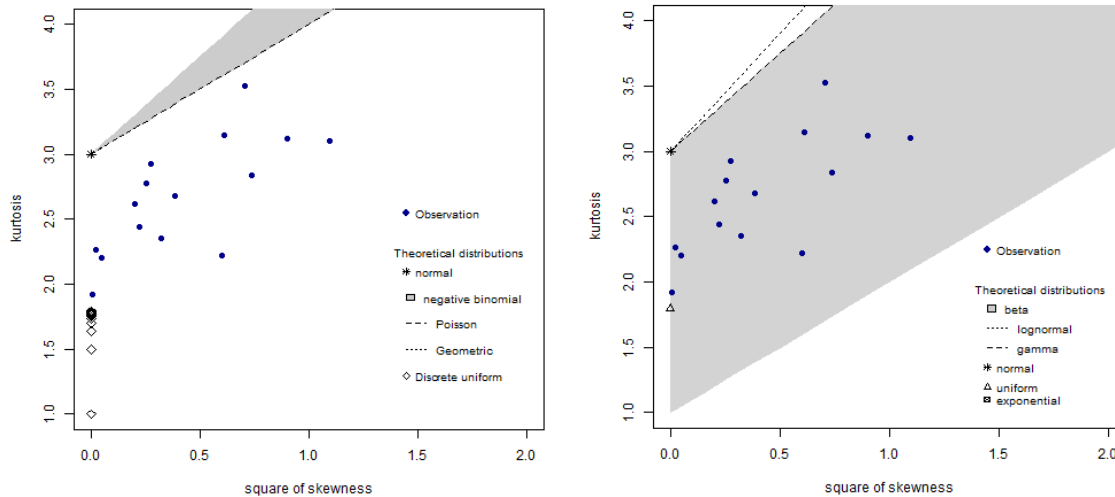

Figure 3: Squared skewness versus kurtosis for time to return. The empirical distributions are generally heavy tailed and skewed. The lines and area indicate the skewness and kurtosis range of some theoretical distributions. Plot is adapted from the Cullen and Frey graph generated with the R package "fitdistRplus".

The abovementioned range of distributions is fitted by maximum likelihood to each SKU[6]. Per SKU per fit, we calculate the maximum absolute deviation (MAD) between the empirical and theoretical distribution and rank the MAD scores. Averaging the MAD scores and summing the rank scores across SKUs gives us information on which distributions fit well in general.

Table 3 indicates that overall the Gamma and Beta distributions fit best according to the MAD criterion. The Beta distribution performs best with respect to total rank. The geometric and uniform distributions do not perform well because these are not flexible enough to account for the heavy skewness and kurtosis in many of the series. The negative binomial may be a candidate after all, but the Beta distribution is most appropriate in terms of fit and general applicability;

---

[6] Please note that we have truncated data, so we fit truncated distributions whenever the support of the distribution is infinite (such as Gamma, geometric, negative binomial). This is not straightforward, because expressions for higher moments of the distribution are complex and often not available. Hence, for this analysis we resort to numerical methods, but these may not be convenient in practice. For the Beta distribution we first normalize the data, such that the support of the empirical distribution is [0,..,1], see appendix.

its base is finite, which is a necessary condition for the forecasting methods, and there is no need to fit a truncated distribution, which is much less cumbersome.

	Average MAD	Total rank
Beta	**0.088**	**26**
Gamma	**0.087**	**34**
Negative Binomial	**0.096**	**38**
Geometric	**0.135**	**53**
Uniform	**0.295**	**80**
Poisson	**0.345**	**84**

Table 3: Average MAD and total rank across SKUs indicate fitting performance (lowest is best).

4. Simulation

The mathematical-simulation model relationship is well established. Less is known, though, about the value of empirical-simulation approaches to research, although evidence suggests that it is through such approaches that recent important propositions in the area of inventory forecasting have been established. Cattani et al. (2011), for example showed, through simulation on an extensive empirical dataset, how some very common inventory assumptions often fall short. It is also through simulation combined with empirical investigations, that an important (disturbing) fact has been revealed. Improved forecast accuracy does not necessarily translate to inventory benefits, because of the interactions between the forecasting and inventory control process – that may not be captured analytically. That is, the fact that method x performs better than method y in terms of forecast accuracy (however this is judged), may be reversed when it comes to inventory performance (see, e.g., Syntetos et al., 2010). This has challenged traditional thinking in the supply chain forecasting community, calling for a clear distinction between forecast *accuracy* and forecast *utility* (i.e. the implications of the forecasts for whatever function, inventory in our case, forecasting serves).

Simulation models that are not empirically based may be criticized as shortcuts to analytical problems that are intractable, or as abstract constructs with no insightful outcomes. However,

when simulation is used in conjunction with empirical data it allows reaching insights not enabled by the empirical data alone. That is, the extremely rich case we use for the purposes of this research would never reveal by itself any insights into inventory forecasting performance. Simulation though is the means to expanding the capability offered by the data by realistically structuring parts of the inventory forecasting process to allow us to make more than what the data offers. The empirical data and context alongside their statistical analysis inform our simulation model. The simulation model allows us to reach conclusions as to what would have happened if the forecast methods and distributional assumptions we experiment with were to be used on the empirical data.

In the remainder of this section, we first discuss the logical systematic process of the development of our simulation model, followed by its structure, scenarios and control parameter values we have employed. We close with a discussion on performance evaluation.

4.1 Simulation process

A number of logical systematic processes have been designed to support the development of a simulation model (e.g. Watson and Blackstone, 1989; Law and Kelton, 2000). One such process was developed by Hoover and Perry (1989) and has guided the structure of this work. The steps of the process are as follows: *Formulating the problem and planning the study* (§1); *collecting the data* (§3); *validation* (we confirmed that our model is a credible representation of the system we wish to simulate and tested face validity of the model by reviewing the results, with Qioptiq)*; construction of the conceptual model* (§4.2); *verification* (we have verified that the computer coding of the simulation model corresponds to the model logic by manually generating in Excel 'short runs' for every method and control parameter combination we have considered); *determining the control parameters of the simulation* (§4.2, §4.3)*; and analyzing the output data* (§4.4, §5).

4.2 Simulation structure and scenarios

We use 15 series (equipment) containing from 4 up to 11 years' worth of data. Although data is available at the most granular level (timestamped transactions), monthly time buckets are utilized to match the decision-making processes employed by the company. We introduce a data augmentation process whereby k new blocks of data (the blocks being equal to the original size of each series) are reconstructed for each series based on bootstrapping (randomly, with replacement; Efron, 1979) the series' characteristics (issues and times to return). Please note that we draw from the actual empirical distributions rather than fitted ones. We generate, and run each simulation on, at least $k = 1,000$ blocks, unless sampling estimation requirements necessitate the generation of more blocks – this issue is further discussed later in this section. We initialize necessary conditions in the first 100 blocks, and evaluate forecasting and inventory control performance ex-ante on the remaining (at least 900) blocks. Our initialization structure ensures that we avoid any problems resulting from returned items that were issued before the start of the simulation. Note that none of the simulation inputs and control parameters are hypothesized. In all cases they are calculated or estimated from the data. That is, we anchor our simulation on the actual empirical data.

In terms of extrapolation, we work as follows. Forecasts of returns and demand are produced at the very end of a time period (month), meaning that all returns and demand that have occurred during that very period are fully observable, and are used as the latest actual observations. Forecasts are produced in a rolling fashion: at the end of every month, new forecasts are generated. For both returns and demand, at the end of every time period forecasts are produced over a lead time + 1 month, to reflect the periodic nature of the system[7]; the same principle applies later, in inventory control. The current lead time for core procurement in Qioptiq is 2

---

[7] Periodic inventory control applications necessitate the inflation of the lead time by one (inventory review) period to reflect the extra uncertainty associated with reviewing inventory positions periodically, rather than continuously.

months, but forthcoming changes in their contracts (see §3) imply that they will increase in some cases to 3 or 4 months; as such, the (total) lead times we test for are 3, 4 and 5 months.

For returns forecasting, three of the four methods suggested by Kelle and Silver (1989a) (KS hereafter) are used. We simulate and report results for M1, M2 and M3. As mentioned before, M4 is found to be computationally prohibitive (due to large matrix inversions). As it is also known to perform worse than M3 and contributes little to M2 (if other means exist to characterize the returns distribution, as is the case here) we do not consider it further, but we use the M2 in conjunction with relevant distributions (fitted to the actual empirical ones) as a 'perfect' proxy. By *perfect*, we mean that using the true (relevant) moments of the empirical time-to-return distributions enables us to compare and discuss the performance of the various adaptations of M4 found in the literature against our proposed (Beta) formulation.

The KS methods provide aggregate lead time forecasts. However, we argue for the importance of utilizing point (per period) forecasts in the context of remanufacturing, since those may be more informative in conjunction with quality related information. We have re-written the KS analytical expressions for point forecasts[8]. Forecasts are generated based on those expressions; their aggregate outcome (e.g., the sum of 1-step and 2-step ahead point forecasts, for a lead time equal to 2 periods) equals the relevant (aggregate) KS forecasts which was numerically validated.

M1 is simulated with a fixed probability of remanufacturable returns equal to 95% (please see §2). M2 is simulated with the assumed time-to-return distribution being Uniform, and the return probability being determined based on the average maximum time to return (N) across SKUs: $1/(N+1)$. Following discussions with the company, this was judged to be the most realistic

---

[8] Please note that these are more or less straightforward reformulations of the known equations, which we suppress here due to page limit constraints. The relevant expressions are available upon request from the corresponding author.

representation of how they would anticipate returns. In addition, the Uniform is a very easy distribution to apply. M2 is also simulated in conjunction with the Beta distribution (with its parameters updated at the end of every month), which was found to be the best performing one in the previous section. M3 is simulated based on three scenarios: using the empirical time-to-return distribution up to the very point in time where a (new) forecast is generated, but also by assuming a Uniform and Beta distribution, as discussed above. Finally, the method employed by the company to forecast returns (SES) is also simulated as a simple time series benchmark[9], which allows us to discuss the impact of disregarding the dependence of returns to past issues. SES is employed with a fixed value of $\alpha = 0.4$, as used by the company.

Next, we extend the KS methods' application, for net demand forecasting, by means of introducing appropriate forecasts for the end product demand itself, and thus showing the effect of demand forecast error on the methods' performance when it comes to net demand. Decomposition of the demand series into their constituent components (level, trend, seasonal factors) is a key exercise in forecasting (e.g. Roodman, 1986). However, in the case of intermittence the scarcity of observations prohibits the identification of components other than the level. Interestingly, SES has been shown in many studies to perform very well in an intermittent context, despite the fact that it has been developed for fast, rather than intermittent, demand items (Syntetos et al., 2005). In periodic applications, such as the one considered here, SES is unbiased (Croston, 1972) and associated with a very robust performance (Syntetos et al., 2015). As such, and in order to retain the relevance with the forecast method the case organization is using[10], demand forecasts are generated using SES. We also consider the method with the most empirical evidence in its support in an intermittent demand context: the Syntetos-

---

[9] Although returns forecasting should be explicitly linked to past demands and time to return distributions, we know of many companies that rely upon simple univariate time series methods to produce such forecasts.

[10] Strictly speaking, the SES forecasts are occasionally subject to judgmental adjustments. Judgmental interventions into statistical forecasts has long been recognized as a recurring practice (Sanders and Ritzman, 1995).

Boylan Approximation (SBA) (Syntetos and Boylan, 2005). The SBA builds forecasts from constituent elements (demand intervals and demand sizes). The SES and SBA results are contrasted with those obtained from assuming that we do know mean demand (as per Kelle and Silver, 1989a).

Finally, forthcoming changes in the company's contracts and stricter definitions as to what constitutes remanufacturability (as opposed to a salvaged/scrapped beyond-economical-repair return) will lead to a reduced (estimated by the company) average probability from the current 95% to 80%. Exploring the resulting effects could be of value to other industries that operate under similar circumstances.

4.3 Inventory management

It is beyond the scope of the paper to simulate the recovery process in all its complex detail. Instead we focus on the core level and simulate the higher level processes as depicted in Figure 4. Please note that all lead times discussed below are inflated by 1, as described in §4.2 (see also footnote 7), to reflect the periodic nature of the system.
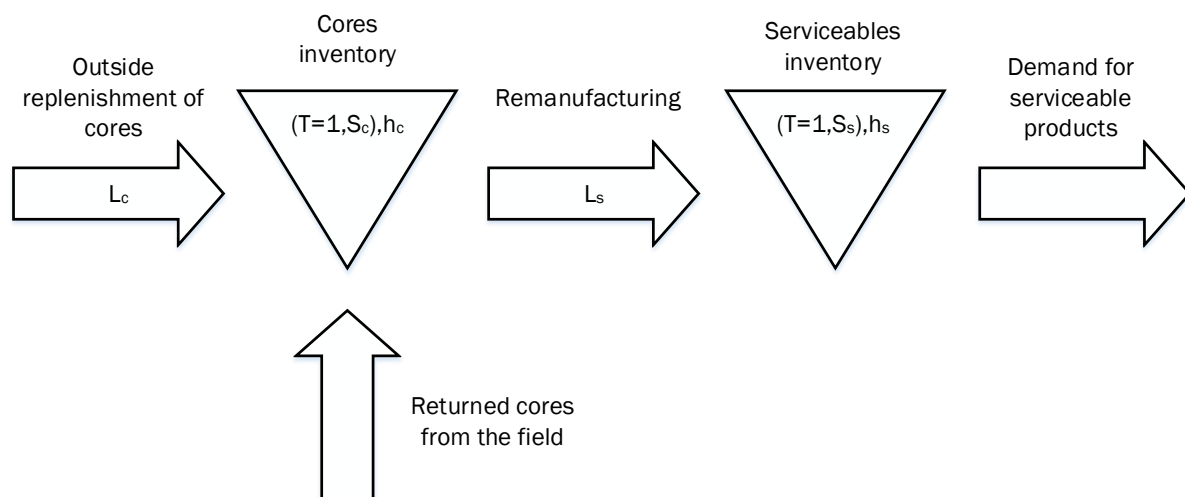

Figure 4. Inventory system.

Demand from the field for serviceable items is satisfied from the serviceables inventory. Demand that cannot be satisfied immediately is backordered, that is, it will be delivered as soon as it is available.[11] This inventory is managed by a (T=1,S_s) base stock policy, meaning that every month the serviceables inventory position I_s is inspected and if found to be smaller than the serviceables base stock level S_s, a remanufacturing order is placed of size S_s - I_s. These orders appear in the serviceables inventory (subject to the availability of cores) after L_s months, which is equal to the remanufacturing lead time. The base stock level S_s is determined every month by forecasting the expected **demand** over the lead time L_s and applying the formula (Silver et al. 2016, p. 259):

$$S_s = E(lead\ time\ demand) + k_s \sigma_D,$$

where σ_D is either the standard deviation of lead time demand, or the standard deviation of the lead time demand forecast errors. The safety factor k_s is set by management.

In real world applications, demand variance is not known but rather is forecasted through some forecast error variance updating procedure, typically the smooth Mean Absolute Deviation (MAD) or smooth Mean Squared Error (MSE) approach. The latter has been found to, generally, perform better (Bretschneider, 1986), and is used here in an intuitively appealing adapted version for lead time demand forecasts (Syntetos and Boylan, 2006).

Orders for remanufactured items are fulfilled from the inventory of returned cores. If sufficient cores are not available, the missing part of the order is backordered. The inventory of cores is managed by a (T=1,S_c) policy, meaning that every month the inventory is inspected and if the inventory position I_c is smaller than S_c an order of size S_c-I_c is placed at the outside supplier of cores. These orders appear in the cores inventory (in full, i.e. we assume that the broker has enough cores to cover net demand requirements) after L_c months, which is equal to the core

---

[11] In practice, if an outside order (demand) cannot be satisfied the order is adjusted. The missed demand is then rescheduled for a later time. In fact, this is very similar to backordering.

broker's lead time. The order-up-to level S_c is determined every month by forecasting the expected **net demand** over the lead time L_c and applying the formula:

$$S_c = E(lead\ time\ net\ demand) + k_c\sigma_{ND},$$

where σ_{ND} is either the standard deviation of lead time net demand, or the standard deviation of the lead time net demand forecast errors. The safety factor k_c is set to achieve a high service level of 99%. This is to reflect the fact that the serviceables inventory is not subjected to supply uncertainty (something that closely resembles the operation of the company).

Kelle and Silver (1989a) derived the lead time net demand variance expressions resulting from their proposed forecast methods, under the assumption that the mean demand is known. We assess the effects for inventory control when using these expressions versus the realistic scenario of estimating the variance of lead time net demand through the variance of the lead time net demand forecast error (as discussed above for the case of demand).

The financial consequence of keeping items in stock are out-of-pocket (unit costs that occur due to the physical presence of the item) and financial investment costs. The latter are typically much higher than the former. Most out of pocket costs are related to the overall infrastructure of the warehouse, but are rather independent of the number of items on stock. If we assume that the financial flows coincide exactly with the physical flows then the financial consequence of the decision to acquire one core from the outside supplier is ρc_a, where c_a is the unit acquisition cost and ρ is the opportunity cost of capital (typically 10-20%, Silver et al., 2016). Similarly, the financial consequence of deciding to move one unit from the cores inventory to the serviceables inventory is ρc_r where c_r is the unit remanufacturing cost. Note that the returning of cores is an autonomous process that is independent of our inventory decisions, so there are no relevant financial consequences to consider here. Hence, we set $h_c = \rho c_a$ and $h_s = \rho(c_a + c_r)$ as the holding costs per item per time unit (for the cores and serviceables inventories,

respectively). Please also note that we do not profess optimality of the proposed system, but rather employ it as a case inspired, straightforward heuristic to explore forecasting utility.

We note that preliminary analysis shows that the simple heuristic suggested above leads to better inventory-service trade-offs than the ad hoc inventory policy employed by the company. In the interests of space economy, the relevant results are not reported here.

4.4 Performance measurement

Forecast accuracy measurement for intermittent series is not a straightforward exercise (see, e.g., Van der Laan et al., 2016). Popular with management is the Mean Absolute Percentage Error (MAPE) metric, where the forecast error is divided by the actual demand. However, this metric cannot be defined in our context of intermittent demand and returns, since the denominator can often be zero. The symmetric MAPE (sMAPE) could be a viable option for demand forecasting (van der Laan et al., 2016). Although its asymmetric properties are desirable in an inventory context (over-penalizing under-forecasting) its application becomes problematic when the series contain zeroes interspersed with both positive and negative values, which is the case for net demand.

In a multi-SKU setting, such as the one considered here, where accuracy needs to be summarized across series, scale independence is the main criterion for selecting an accuracy measure. The Mean Absolute Scaled Error (MASE), proposed by Hyndman and Koehler (2006), has been put forward as a robust such measure, with many desirable statistical properties and particular relevance to intermittence.

For non-seasonal series and 1-step ahead forecasts, the (signed) forecast error e_t produced in every period is scaled based on the in-sample Mean Absolute Error (MAE) from the Naïve forecast method, to give a scaled error:

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$$

where n is the history of data (in-sample) available and Y_i the actual observation in period i.

The result is independent of the scale of the data. A scaled error is less than 1 if it arises from a better forecast than the average 1-step naïve forecast computed in sample. For 2-, 3- or further step ahead forecast errors, as well as for measuring aggregate, across lead time, forecast errors, relevant adjustments are needed for the in-sample naïve forecast and corresponding MAE calculations. The scaled errors can be summarized across time (and then across series) as: $MASE = mean(|q_t|)$.

As previously discussed, every simulation runs for at least 900 blocks (with each block containing, depending on the SKU, from 48 to 132 periods/months). The simulation stops as soon as the relative error in the MASE is less than 1%. The relative error is computed by constructing 95% confidence intervals $\left[\bar{x} - t_{0.975} \, s /\sqrt{k}, \bar{x} + t_{0.975} \, s/\sqrt{k}\,\right]$, where $\bar{x}$ and s are the mean and standard deviation of MASE based on k runs, and $t_{0.975}$ is the 0.975 percentile of the t-distribution with k-1 degrees of freedom. The relative error then is $t_{0.975} \, s /\left(\bar{x}\,\sqrt{k}\right)$.

Results are first recorded per SKU/series, and then analyzed across series. Pair-wise comparisons are conducted between any two forecasting methods. A reasonable way to assess statistical significance in a parametric manner is to rely upon a t-test (for matched pairs), with 28 degrees of freedom (2 x 15 (series) – 2). The critical values are 2.467 and 1.701 for 1% and 5% significance level, respectively[12].

Finally, by fixing the stock control model (as discussed in the previous sub-section) and varying the forecasting related input in the model, the inventory implications of various forecasting

---

[12] Degrees of freedom, t-statistic calculations and critical values are adjusted when evaluating performance within subsets of the 15 series.

approaches may be evaluated and compared. This takes place by reporting inventory investments and achieved service levels, and contrasting them through the construction of efficiency curves (Gardner and Dannenbring, 1979).

5. Results and discussion

5.1 The effect of serialization

The results show that M3 (in conjunction with the empirical time-to-return distribution) performs best across all 15 equipment. Although the MASE differences with M2 (used in conjunction with the Beta distribution) are not particularly marked, they are in fact statistically significant, when employing the t-test ($MASE_{M3\text{-}Empirical} = 0.374$; $MASE_{M2\text{-}Beta} = 0.448$)[13],[14].

Further analysis on an individual equipment basis reveals some interesting results. Tables 1 and 2 demonstrate the skewed behavior of the demand and returns rate across equipment. As previously discussed, while most equipment are associated with a rather low volume of demand, three of them are distinctly different in that high volumes are encountered on average. In Table 4, we present the MASEs for SES, M1, M2-Beta and M3-Empirical, across all control parameter combinations, and by also grouping the results into two (distinct) product categories: high volume (3 items); low volume (12 items). The t-statistic values for all pair-wise comparisons, and across all equipment, are presented in Table 5 (all are statistically significant).

	SES	M1	M2-Beta	M3-Empirical
All equipment	0.670	0.921	0.448	0.374
High volume (3 items)	0.676	1.430	0.438	0.359
Low volume (12 items)	0.668	0.793	0.451	0.378

Table 4: MASEs of lead time returns forecasts results across all control parameter combinations.

---

[13] The subscript associated with the MASE refers to the *method-time to return distribution* combination under discussion, and the relevant value is an average across all relevant control parameter combinations.

[14] Is important to note that statistical significance is also reached when employing non-parametric tests. In particular, the Wilcoxon (1949) rank sum test does support the fact that *M3-Empirical* performs best across all equipment. Given the volume of our simulation output, additional non-parametric analysis becomes cumbersome to consider for all control parameter combinations, but in this occasion is worthwhile referring to. The non-parametric tests are in full accordance with the t-tests, not only for the comparisons involving the M3-Empirical but for all comparisons reported here.

	SES	M1	M2-Beta	M3-Empirical
SES	-	2.79	17.45	19.42
M1	-	-	5.28	6.09
M2-Beta	-	-	-	6.70

Table 5: *t*-test statistic pair-wise comparison values of MASE results across all equipment.

The forecast accuracy of M2-Beta and M3-Empirical increases when we move from the low to the high volume category, with M3-Empirical benefiting marginally more. Their comparative performance (relative to the other methods) also improves, since the $MASE_{SES}$ and $MASE_{M1}$ deteriorate for the high volume items.

In summary, M3-Empirical compares favorably to the other methods for both high and low volume items. Thus, our findings partly support those of Toktay et al. (2000) who found that benefits of serialization are focused on slow moving items only: "*significant benefits can be gained from using the additional information obtained by time stamping the product. As [we move from low volume to high volume items], the relative benefit of collecting additional information diminishes and a less information-intensive method becomes adequate*, p. 1423". M1 is found to perform rather poorly, and cannot be recommended for practical returns forecasting applications. The same is the case for SES, as expected. Returning to the discussion related to the IoT-RFID literature, SES and M1 ignore the correlation of past sales to returns ('no visibility-simple'), the M2-Beta correctly specifies the distribution and its moments ('no visibility-advanced') and the M3-Empirical requires item-level data such as the ones that can be obtain through ('full visibility') 'reading' technologies.

An interesting observation, explicitly linked with the effects of serialization, is how M2 and M3 perform in the presence of imperfect information. In Table 6, we can see how M3 is much more susceptible to forecasting inaccuracies, in agreement with de Brito and Van der Laan (2009). This can happen in two ways: correctly characterizing the wrong distribution, and incorrectly characterizing the correct distribution. This might be attributed to either problematic

assumptions, wrong methods of extrapolation or simply an environment of constant change. Serialization then does not only enable the use of M3 (by tracking outstanding returns), but its benefits transcend method selection in enabling/updating the correct characterization of the correct distribution. In any case, it should be recommended that in the presence of higher uncertainties, the robust M2 should be preferred to the over-sensitive M3. The interactions of the level of inaccuracies (between the estimated and the real time-to-return distribution) and the performance of the methods perhaps merit a deeper exploration.

	M2	M3
Empirical	0.403	0.374
Beta	0.448	0.420
Uniform	0.525	1.069

Table 6: MASEs of lead time returns forecasts for M2 and M3 combined with the Empirical, Beta, and Uniform distributions, exploring method performance in the presence of increasingly (top to bottom) imperfect information.

Interestingly, the differences between M2-Empirical and M3-Beta are the only ones that are not statistically significant, further suggesting that Beta approximates our data well. We can see that when moving to a distributional assumption that does not work e.g. as demonstrated here via the Uniform distributional assumption (the reader is reminded the MAD scores of Table 3: Beta, 0.088; Uniform, 0.295), M2 handles itself much better than M3.

Although some of these results were admittedly expected based on our analysis in §3, they do give rise to an important observation. An important issue related to returns forecasting is how demand forecasting takes place.

As mentioned before, (expected) lead time returns can originate from (observed) past issues that have not returned yet, but also from (expected) lead time issues that may still be returned within that lead time (Kelle and Silver 1989a). In that respect, demand forecasting does have some impact on returns forecasting. We find that SBA performs marginally worse than assuming perfect knowledge about the mean lead time demand, while SES leads to higher errors, in

accordance with theoretical expectations. The MASE results for SES, SBA and the case when the mean lead time demand is assumed to be known are offered in Table 7. Two of the three pair-wise differences are statistically significant at the 1% level, while the difference between SBA and the mean demand assumed known is not particularly marked.

	Mean demand assumed known	SES	SBA
All equipment	0.556	0.736	0.566

Table 7: MASEs of demand forecasts, across all control parameter combinations.

The impact of demand forecasting increases considerably for net demand forecasting, and this is discussed in more detail in §5.2, below. The impact of the lead time length is further considered in §5.3 and §5.4.

5.2 The effect of forecasting demand

Net demand forecasts are severely affected by the interactions between the returns and demand process, such that it becomes difficult to isolate their individual contributory effects on the final forecasts. Because of its scaling, MASE cannot convey changes in the absolute forecast error when we move from returns to net demand forecasting[15], and as such it cannot attest to how important it is to consider appropriate demand forecasting methods. This is why we consider here the MAE instead to show differences. (Incidentally, the MAE relates explicitly to the inventory task, which is addressed in §5.4.)[16]

---

[15] As the MASE denominator (naïve forecast MAE) changes, comparative performance insights may be lost. For example, naïve may result in high MAE for high volume items, therefore flattering the performance of the forecasting method in consideration in terms of MASE. Conversely, naïve has been shown to perform well for low volume items by 'guessing correctly' all the zero demands (optimizes on the median rather than the mean), resulting in low MAEs that could inflate the MASE thereby concealing any improvements of the method under consideration. The parallel can be drawn for lead time forecasts where a comparatively more severe deterioration of the naïve forecasts (as opposed to say SBA forecasts) might produce smaller MASEs in increasing lead times.

[16] The ratio of the MAE to the standard deviation of the forecast error (based on which safety stock decisions are determined) is (within sampling error) almost constant for many distributions, including the normal, negative exponential and rectangular probability density function.

Table 8 shows these differences for SES, M1, M2-Beta and M3-Empirical. Please note that the table is not intended for use to compare methods (i.e. horizontally). Instead, its value lies in allowing us to interrogate the uncertainty increase as we move from returns to net demand forecasting (i.e. vertically), the latter of which compounds the forecast errors from both returns and demand[17]. Referring to the table, the increase in MAEs displays a level of uncertainty that calls for increased attention to demand forecasting in a remanufacturing (or other circular economy) context.

		SES	M1	M2-Beta	M3-Empirical
Lead time returns forecasts	**All equipment**	9.296	16.240	6.063	4.982
	High volume	24.620	52.530	15.735	12.913
	Low volume	5.465	7.167	3.645	3.000
Lead time net demand forecasts	**All equipment**	43.230	49.699	42.545	42.125
	High volume	145.931	168.596	141.958	141.585
	Low volume	16.977	18.955	16.137	15.975

Table 8: MAEs of lead time returns and net demand forecasts, across all control parameter combinations, and across all demand forecasting methods.

5.3 The effect of lead time length

Point, rather than aggregate lead time, returns (and net demand) forecasts are of particular relevance in remanufacturing operations since they can (in conjunction with remanufacturing lead time information) be appropriately linked to what may or may not be remanufactured over the lead time (meaning that varying degrees of quality need different times to be restored to as-good-as-new state). This is very important for tactical capacity planning, allowing for remanufacturers to schedule time consuming operations at the beginning of the lead time, and exploit opportunities of expediting production by scheduling the remanufacturing of good quality cores when facing demand peaks.

---

[17] Given the extent to which the MAE is used in forecasting studies to compare alternative methods, we report the comparative results here (for information only). The lead time return forecast MAE differences are not statistically significant when considered across all 15 SKUs (except for M1 against M2-Beta and M3-Empirical, at 5%). They are however, at 1%, when considered within their class of high/low volume items (except for M2-Beta and M3-Empirical, at 5%). The net demand forecast MAE differences are not statistically significant.

The longer the planning horizon though, the greater the inaccuracy of the point forecasts. Our findings provide empirical evidence in support to those offered by de Brito and van der Laan (2009) and Clottey (2016), in that point forecasts deteriorate the further along into the future these forecasts refer to. We show this in Table 9, for both returns and net demand forecasting, across all returns and demand forecast methods. Please note that MASEs do not convey information on comparative performance if considered across lead time lengths, since the performance of the Naïve method changes (deteriorates) with the lead time[15]. This is why MAEs are considered here to demonstrate performance implications across lead times. We also exclude the M1 due to its (as seen above) very poor performance. For information, the differences are statistically significant for both returns forecasts (at 5% for high and 1% for low volume items) and net demand forecasts (at 5% for both high and low volume items), when considered horizontally.

	Lead time = 3	Lead time = 4	Lead time = 5
Returns forecasts	5.652	6.758	7.930
Net forecasts	35.141	42.029	47.996

Table 9: MAEs of returns and net demand forecasts averaged for SES, M2-Beta and M3-Empirical, across all control parameter combinations, over a lead time = 3, 4, 5 periods.

5.4 Forecast utility

We now consider how (whether) the forecast accuracy results discussed above translate to inventory benefits. Inventory investments are calculated by considering the actual acquisition cost of the 15 equipment we deal with and the average (per equipment) cost of the relevant cores. The opportunity cost is set as: $\rho = 12\%$ (after consultation with the company). We cannot disclose actual cost figures and, therefore, we normalize the presentation of the results, in a scale of 0 to 1 (with 1 representing the maximum investment in the scenario under consideration).

In Figure 5, we show the achieved cycle service level (CSL) – inventory investment trade-offs for M2-Beta, M3-Empirical and the method employed by Qioptiq (SES). For conciseness, we

omit presentation of the results related to M1, since this method performs very poorly – as previously shown. The demand forecasts have been generated using the SBA, found in §5.1 to be better than SES and not statistically different than the 'known mean' approach.
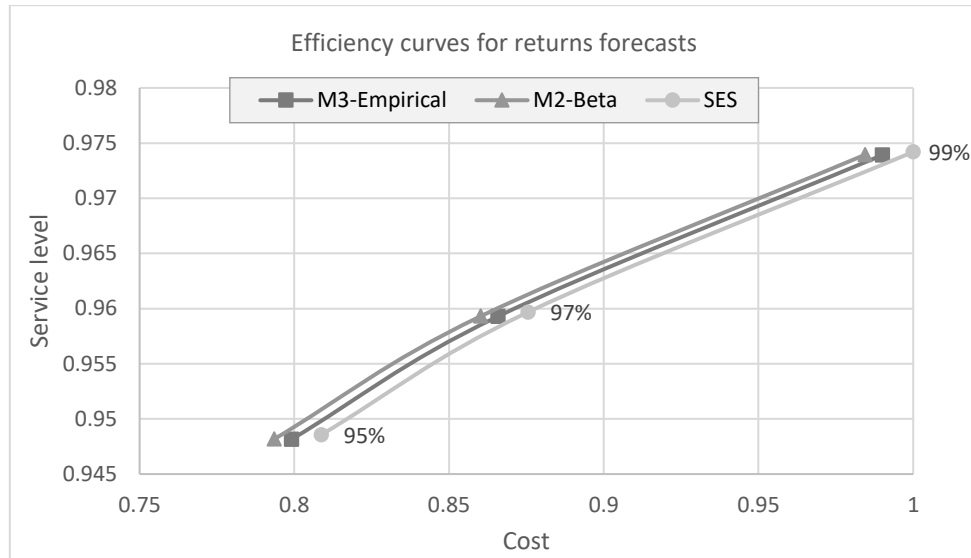

Figure 5. Efficiency curves for M2-Beta, M3-Empirical and SES (0.4), across all lead times, and varying targeted service level = 95%, 97%, 99%, demand forecasting through SBA, and empirical forecast variance calculation.

We have simulated inventory and service level performance for targeted service levels of 95%, 97% and 99%. The intermittent behavior of the demand and the approximate nature of our error distributional assumption (normal; see Syntetos and Boylan, 2008) implies that the higher the service level target is the more difficult it becomes (for any method) to achieve it (Syntetos and Boylan, 2006). All three forecast methods behave similarly in terms of achieved service levels, very close to the 95% target, but do 'struggle' when the target raises to 97% and especially 99%. In terms of cost, SES performs worse (as expected) and the performance of M2-Beta and M3-Empirical is very similar. The resulting graph can be read horizontally (left to right, down), linking resulting CSL with inventory investment, and vertically (bottom up, left) for the inverse.

In Figure 6, we show the effect of using an appropriate forecast error variance procedure, as opposed to relying upon the theoretical variance expressions advocated by Kelle and Silver (1989a). It is intuitive, but also strongly supported by our results, that the theoretical expressions

developed under the known and constant demand assumption fail to deliver good inventory performance in the reality of unknown demand.
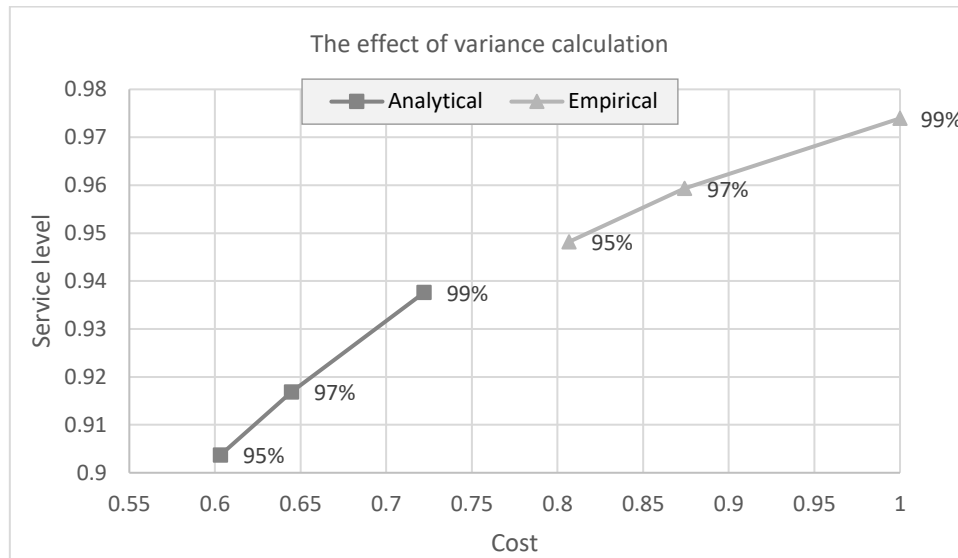

Figure 6: Efficiency curves averaged for M3-Empirical and M2-Beta, targeted service level = 95%, 97%, 99% and variance estimated through the theoretical expression (Kelle and Silver, 1989a) or through the empirical approach.

In Figure 7, we show the effect of manufacturability and lead time length on inventory performance. Empirical inventory findings (Syntetos and Boylan, 2006) suggest that inventory performance deteriorates as the lead time increases. This was demonstrated (indirectly) in Table 9 where we show the changes in MAE as a function of the lead time. Please recall that the MAE relates directly to the standard deviation of the lead time forecast error and thus to the determination of safety stocks needed to sustain a targeted service performance. We see that these results carry over into inventory control, albeit their impact is less severe than expected. This is, at least partly, attributable to the high probabilities of remanufacturable returns. It can be seen from the efficiency curves, becoming marginally steeper (and lower) when moving from probability of a remanufacturable return = 95% to the one of 80%. Indeed, as the probability of remanufacturable returns drops, the uncertainty in the supply of cores becomes more problematic.
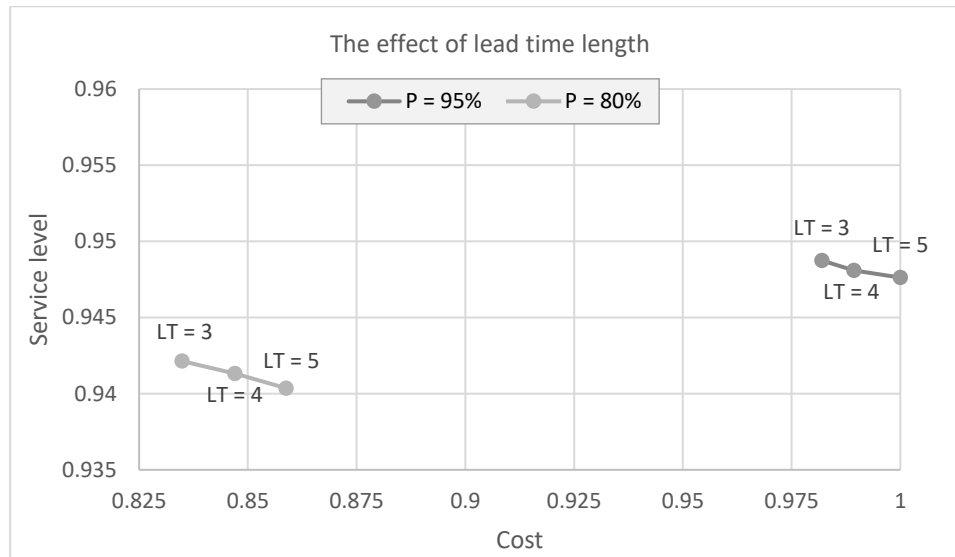
Figure 7: Efficiency curves for M3-Empirical, across lead times (= 3, 4, 5) and targeted service level = 95%, variance empirical approach for 95% and 80% probabilities of remanufacturable return.

6. Conclusion

This study has provided an in-depth empirical analysis (via simulation) of the performance of some key forecast methods in a remanufacturing environment. A unique dataset of serialized transactional issues and returns from the Excelitas Group and one of their defense contractors, Qioptiq, was used for that purpose. Our study is the first to present such detailed information on forecasting for remanufacturing and the first to present such large empirical forecast performance analysis. We have characterized the empirical behavior of the returns distribution, as well as both the demand and returns series, and constructed a detailed simulation exercise to assess, *ex ante*, the performance of various forecast methods associated with differing degrees of information intensiveness. The performance, both in terms of accuracy and utility, has been discussed against the backdrop of information technologies and data availability literature. Beyond utilizing empirical data, the simulation exercise was conducted, in its entirety, by employing real (as opposed to hypothesized) values for all control parameters. Variation of the control parameters and sensitivity analysis of the results helped assess the robustness of the methods, expanding the relevance of our findings to a number of circular SCs and remanufacturers that replenish cores through brokers. Core brokers represent the most regular

source of cores for remanufacturers after their own customers (at 10% in the US; Guide, 2000). Examples of industries relying on core brokerage include car remanufacturing where modules (e.g., gearboxes, power trains) are shared across models and manufacturers, circuit board and electronics (e.g., Clottey et al., 2012).

We suggest that serialization is something worthwhile pursuing for low volume products, especially if they are expensive. This makes a lot of sense from an investment perspective, since the relevant serial numbers are very few. However, we provide evidence that such benefits also expand in the case of high volume items (as is found through our analysis). Importantly, the benefits of serialization not only enable the implementation of the more complex M3, but also the accurate characterization of the returns process thus also benefiting the more robust M2. In terms of distributions, we recommend either the employment of the actual empirical or a Beta representation of it, though the latter is admittedly conditional to our dataset and more empirical studies are needed to validate this finding. In terms of methods, M3 performs best in stable environments while M2's robustness is preferable otherwise. We believe that these findings are relevant to both the returns forecasting and the information availability literature streams.

We also argue for the importance of demand forecasting. Previous studies seem to be preoccupied with returns forecast accuracy, neglecting the effect that demand forecast errors may have on net demand forecasts. Although this is understandable, up to a certain extent, it shifts attention away from the real problem, which is the dual-source uncertainty. That is, net demand series are associated with a dual source of variation: demand and returns.[18] The interactions between them are not straightforward for at least two reasons: i) demand obviously influences returns, as demand forecasts also influence the returns forecasts; ii) the resulting net demand series may be statistically very difficult to handle (both in terms of analysis and

---

[18] In the scenario that both series are intermittent, i.e. when the constituent series (demand and returns) are themselves associated with a dual source of uncertainty, arrivals and sizes, things are obviously even more involved.

measurement of resulting forecast errors) because they are intermittent, interspersed with both positive and negative observations.

Causal forecasting developments would be something natural to consider in this context. Information on fleet size or consumer behaviour (maybe in classes) is naturally expected to possess great explanatory power. Incorporation of such information could take two forms: i) introduce regression-type models with appropriate dummy variables and lags to exploit explanatory variables present in a dataset – which, to some extent, has been addressed already in the literature; ii) using such information to 'better' characterise the time-to-return distributions. Given the importance of the latter for the methods currently available in the literature, further work on better informing the distributions under concern would appear to be merited.

As discussed in the introduction, forecasting for remanufacturing entails the three challenges of forecasting demand, returns, and their quality. Therefore, to holistically address forecasting for remanufacturing, research would need to move away from the assumption of uniform quality of returns. A complementary *quality of the returns* distribution (ranging from beyond-economical-repair to almost new) needs to be estimated and applied on the distribution of returns, to drive part procurement and operations planning. We could describe this as an unknown '*Bill of Remanufacturing*' (BOR) that needs to be estimated. This concept could be expanded into a '*Bill of Remanufacturing Operations*' (BORO) taking into account the relevant operations, similarly to how the 'Bill of Operations' works for manufacturing. Accurate estimations of BOR and BORO can help production levelling efforts and planning.

We have seen in §2.3 how increasingly available technological innovations can enable the accurate estimation of the return rate, at a serial level, as soon as the core enters the reverse supply chain (and in cases even before then). Such forward signaling may delimit the temporal

scope of returns forecasting (from sale to return, to end-of-life/product failure to return) and thus greatly aid forecast accuracy. The proliferation of RFID and IoT technologies can provide remanufacturers with forward information to alleviate the dual-source uncertainties they face. Information storage frameworks (i.e. what information is to be stored in RFID-type tags) are integral to forecasting returns by providing the data needed for either forecasting methods such as the M3-empirical (full visibility), causal forecasting, and indeed the BOR and BORO (Parlikad and McFarlane, 2007). In scenarios where demand is majorly driven by product failures, such data may simultaneously help forecast demand, returns, and their quality.

References

Bretschneider, S., 1986. Estimating forecast variance with exponential smoothing: some new results. International Journal of Forecasting 2, 349-355.

Carrasco-Gallego, R., Ponce-Cueto, E., 2009. *Forecasting the returns in reusable container's closed-loop supply chains. a case in the LPG industry*. 3rd International Conference on Industrial Engineering and Industrial Management, September 2nd-4th, 2009, Barcelona, Spain.

Cattani, K.D., Jacobs, F.R., Schoenfelder, J., 2011. Common inventory modelling assumptions that fall short: Arborescent networks, Poisson demand, and single echelon approximations. Journal of Operations Management 29, 488-499.

Clottey, T., 2016. Development and evaluation of a rolling horizon purchasing policy for cores. International Journal of Production Research 54, 2780-2790.

Clottey, T., Benton, W. C, 2014. Determining core acquisition quantities when products have long return lags. IIE Transactions 46, 880-93.

Clottey, T., Benton, W.C., Srivastava, R., 2012. Forecasting product returns for remanufacturing operations. Decision Sciences 43, 589-614.

Croston, J.D., 1972. Forecasting and stock control for intermittent demands. Operational Research Quarterly 23, 289-303.

de Brito, M.P., Dekker, R., 2003. *A framework for reverse logistics*. In R. Dekker et al. (Eds.), Reverse logistics: quantitative models for closed loop supply chains. Springer Verlag, Berlin, 3-27.

de Brito, M.P., van der Laan, E., 2009. Inventory control with product returns: the impact of imperfect information. European Journal of Operational Research 194, 85-101.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7, 1-26.

European Remanufacturing Network (*ERN*), 2015. Remanufacturing market study (for Horizon 2020). https://www.remanufacturing.eu/wp-content/uploads/2016/01/study.pdf

Fang, C., Liu, X., Pardalos, P.M. and Pei, J., 2016. Optimization for a three-stage production system in the Internet of Things: procurement, production and product recovery, and acquisition. The International Journal of Advanced Manufacturing Technology 83, 689-710.

Gardner, E.S. Jr, Dannenbring, D.G., 1979. Using optimal policy surfaces to analyze aggregate inventory tradeoffs. Management Science 25, 709-720

Goh, T.N., Varaprasad, N., 1986. A statistical methodology for the analysis of the life-cycle of reusable containers. IIE Transactions 18, 42-47

Goltsos, T.E., Ponte, B., Wang, S., Liu, Y., Naim, M.M. and Syntetos, A.A., 2018. The boomerang returns? Accounting for the impact of uncertainties on the dynamics of remanufacturing systems. International Journal of Production Research, in press, doi: 10.1080/00207543.2018.1510191.

Govindan, K., Soleimani, H., Kannan, D., 2015. Reverse logistics and closed-loop supply chain: a comprehensive review to explore the future. European Journal of Operational Research 240, 603-626.

Guide, V.D.R. Jr, 2000. Production planning and control for remanufacturing: industry practice and research needs. Journal of Operations Management 18, 467-483.

Guide, V.D.R. Jr, Jayaraman, V., 2000. Product acquisition management: current industry practice and a proposed framework. International Journal of Production Research 38, 3779-3800.

Hanafi, J., Kara, S., Kaebernick, H., 2008. Reverse logistics strategies for end-of-life products. International Journal of Logistics Management 19, 367-388.

Hoover, S.V., Perry, R.F., 1989. *Simulation: a problem solving approach*. Addison-Wesley, New York.

Hosoda, T., Disney, S.M., Gavirneni, S., 2015. The impact of information sharing, random yield, correlation, and lead times in closed loop supply chains. European Journal of Operational Research 246, 827-836.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679-688.

Karaer, Ö. and Lee, H.L., 2007. Managing the reverse channel with RFID-enabled negative demand information. Production and Operations Management 16, 625-645.

Kelle, P., Silver, E.A., 1989a. Forecasting the returns of reusable containers. Journal of Operations Management 8, 17-35.

Kelle, P., Silver, E.A., 1989b. Purchasing policy of new containers considering the random returns of previously issued containers. IIE Transactions 21, 349-354.

Kiesmüller, G.P., van der Laan, E., 2001. An inventory model with dependent product demands and returns. International Journal of Production Economics 72, 73-87.

Kumar, S., Yamaoka, T., 2007. System dynamics study of the Japanese automotive industry closed loop supply chain. Journal of Manufacturing Technology Management 18, 115-138.

Law, A.M., Kelton, W.D., 2000. *Simulation modelling and analysis*. Third edition, McGraw-Hill, New York.

Lee, H. and Özer, Ö., 2007. Unlocking the value of RFID. Production and Operations Management 16, 40-64.

Lengu, D., Syntetos, A.A., Babai, M.Z., 2014. Spare parts management: linking distributional assumptions to demand classification. European Journal of Operational Research 235, 624-635.

Liang, X., Jin, X., Ni, J., 2014. Forecasting product returns for remanufacturing systems. Journal of Remanufacturing 4, 1-18.

Luttropp, C. and Johansson, J., 2010. Improved recycling with life cycle information tagged to the product. Journal of Cleaner Production 18, 346-354.

Muckstadt, J.A., Isaac, M.H., 1981. An analysis of single item inventory systems with returns. Naval Research Logistics Quarterly 28, 237-254.

Ondemir, O. and Gupta, S.M., 2014. Quality management in product recovery using the Internet of Things: An optimization approach. Computers in Industry 65, 491-504.

Parlikad, A.K. and McFarlane, D., 2007. RFID-based product information in end-of-life decision making. Control engineering practice 15, 1348-1363.

Parlikad, A.K., & McFarlane, D., 2010. Quantifying the impact of AIDC technologies for vehicle component recovery. Computers & Industrial Engineering 59, 296-307.

Roodman, G.M., 1986. Exponentially smoothed regression analysis for demand forecasting. Journal of Operations Management 6, 485-497.

Sanders, N.R., Ritzman, L.P., 1995. Bringing judgment into combination forecasts. Journal of Operations Management 13, 311-321.

Sarac, A., Absi, N. and Dauzère-Pérès, S., 2010. A literature review on the impact of RFID technologies on supply chain management. International Journal of Production Economics 128, 77-95.

Silver, E.A., Pyke, D.F., Douglas, J.T., 2016. *Inventory and production management in supply chains*. Fourth edition, CRC Press, Boca Raton, Florida.

Syntetos, A.A., Babai, M.Z., Gardner, E.S. Jr, 2015. Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. Journal of Business Research 68, 1746-1752.

Syntetos, A.A., Boylan, J.E., 2005. The accuracy of intermittent demand estimates. International Journal of Forecasting 21, 303-314.

Syntetos, A.A., Boylan, J.E., 2006. On the stock control performance of intermittent demand estimators. International Journal of Production Economics 103, 36-47.

Syntetos, A.A., Boylan, J.E., 2008. Demand forecasting adjustments for service-level achievement. IMA Journal of Management Mathematics 19, 175-192.

Syntetos, A.A., Boylan, J.E., Croston, J.D., 2005. On the categorisation of demand patterns. Journal of the Operational Research Society 56, 495-503.

Syntetos, A.A., Nikolopoulos, K., Boylan, J.E., 2010. Judging the judges through accuracy-implication metrics: the case of inventory forecasting. International Journal of Forecasting 26, 134-143.

Toktay, L.B., 2003. *Forecasting product returns*. In V.D.R. Guide, Jr. & L.N. van Wassenhove (Eds.), Business aspects of closed-loop supply chains. Carnegie Mellon University Press, Pittsburgh, 203-209.

Toktay, L.B., Wein, L.M., Zenios, S.A., 2000. Inventory management of remanufacturable products. Management Science 46, 1412-1426.

US International Trade Commission (*USITC*), 2012. Remanufactured goods: an overview of the US and global industries, markets, and trade. www.usitc.gov/publications/332/pub4356.pdf

van der Laan, E., Salomon, M., 1997. Production planning and inventory control with remanufacturing and disposal. European Journal of Operational Research 102, 264-278.

van der Laan, E., van Dalen, J., Rohrmoser, M., Simpson, R., 2016. Demand forecasting and order planning for humanitarian logistics: an empirical assessment. Journal of Operations Management 45, 114-122.

Watson, H.J., Blackstone, J.H., 1989. *Computer simulation*. Second edition, John Wiley & Sons, Inc. Chichester.

Wei, S., Tang, O., Sundin, E., 2015. Core (product) acquisition management for remanufacturing: a review. Journal of Remanufacturing 5, 1-27.

Wilcoxon, F., 1949. *Some rapid approximate statistical procedures*. American Cyanamid Company, New York.

Zhou, L., Disney, S.M., 2006. Bullwhip and inventory variance in a closed loop supply chain. OR Spectrum 28, 127-149.

Zhou, L., Xie, J., Lin, Y., Ieromonacou, P., Zhang, X., 2016. Forecasting return of used products for remanufacturing using graphical evaluation and review technique (GERT). International Journal of Production Economics 181, 315-324.

Appendix. Distribution fitting

The observations with respect to the time to return are fitted with the Beta distribution as follows. Let N be the maximum of the observations X, so the empirical distribution has support $[1,2,..,N]$. The Beta distribution has support $[0..1]$, so we normalize through dividing by N and fit $\left(X - \frac{1}{2}\right)/N$ based on matching the sample mean and variance (which in this case is much more convenient than maximum likelihood estimation). If the underlying random variable is Y then we calculate the discretized probabilities as Prob($Y=i$)=Beta(i)-Beta(i-1), $i=1,2,..,N$. If μ is the sample mean and σ is the sample variance, then the parameters α and β of the Beta distribution are calculated as

$$\alpha = \left(1 - \frac{\mu}{\sigma^2} - \frac{1}{\mu}\right)\mu^2$$

$$\beta = \left(\frac{1}{\mu} - 1\right)\alpha,$$

so that the Beta distribution has the same moments as the sample data.