Working paper - February 2023

Data-Driven Failure Time Estimation in a Consumer Electronics Closed-Loop Supply Chain

Andre P. Calmon
Operations Management Area - Georgia Institute of Technology, andre.calmon@gatech.edu

Stef Lemmens
Technology and Operations Management, Rotterdam School of Management, s.lemmens@rsm.nl

Stephen C. Graves
A. P. Sloan School of Management, Massachusetts Institute of Technology, sgraves@mit.edu

Problem definition: We examine and analyze a strategy for forecasting the demand for replacement devices in a large Wireless Service Provider (WSP) that is a Fortune 100 company. The Original Equipment Manufacturer (OEM) refurbishes returned devices that are offered as replacement devices by the WSP to its customers, and hence the device refurbishment and replacement operations are a closed-loop supply chain.
Academic/practical relevance: We introduce a strategy for estimating failure time distributions of newly launched devices that leverages the historical data of failures from other devices. The fundamental assumption that we make is that the hazard rate distribution of the new devices can be modeled as a mixture of historical hazard rate distributions of prior devices.
Methodology: The proposed strategy is based on the assumption that different devices fail according to the same age-dependent failure distribution. Specifically, this strategy uses the empirical hazard rates from other devices to form a basis set of hazard rate distributions. We then use a regression to identify and fit the relevant hazard rates distributions from the basis to the observed failures of the new device. We use data from our industrial partner to analyze our proposed strategy and compare it with a Maximum Likelihood Estimator (MLE).
Results: To evaluate our forecasting strategies, we use the Kolmogorov-Smirnov (KS) distance between the estimated Cumulative Distribution Function (CDF) and the true CDF, and the Mean Absolute Scaled Error (MASE). Our numerical analysis shows that both forecasting strategies perform very well. Furthermore, our results indicate that our proposed forecasting strategy also performs well (i) when the size of the basis is small and (ii) when producing forecasts early in the life cycle of the new device.
Managerial implications: A forecast of the failure time distribution is a key input for managing the inventory of spares at the reverse logistics facility. A better forecast can result in better service and less cost (see Calmon and Graves (2017)). Our general approach can be translated to other settings and we validate our hazard rate regression approach in a completely different application domain for Project Repat, a social enterprise that transforms old t-shirts into quilts.

Key words: Closed-Loop Supply Chains, Reverse-Logistics, Forecasting, Sustainability.
History: Working Paper - March 2019.

1. Introduction

The management of product failures, product replacements, and warranty claims is a significant challenge in the consumer electronics industry where warranty-related costs commonly exceed 2% of product revenues (Apple Inc. 2017, HP Inc. 2017, Samsung 2018). We propose and analyze a strategy for forecasting the demand for replacement devices in a large Wireless Service Provider (WSP), a Fortune 100 company that sells smartphones and was our partner for this research. More specifically, this WSP offers a warranty to their customers (usually 12 months in length), and customers covered by the warranty are entitled to receive a replacement from the WSP if and when their device fails. When a customer's device fails under warranty, the customer files a warranty claim and receives a replacement (usually a refurbished device) that is shipped overnight from the WSP's reverse logistics facility. After receiving the replacement device, the customer ships their failed device to the WSP (usually within one or two weeks), which then proceeds to refurbish/repair the device (if possible) and stores it in inventory in order to use it as a replacement device in the future. If the WSP finds that, at some point, it has too many refurbished devices in inventory, excess devices can be sold through a side-sales channel.

As mentioned in Calmon et al. (2020), our partner WSP faces three operational challenges when managing this warranty system: (*i*) forecasting the hazard rates of new products; (*ii*) deciding how many devices should be kept in inventory to serve the warranty claims; and (*iii*) assigning devices in inventory to incoming customer warranty requests.

Although these problems are intertwined, we study them separately. In this paper, we address challenge (*i*) and introduce two strategies for forecasting the failure time distribution of new products. In Calmon and Graves (2017) we address challenge (*ii*), namely the inventory management at the WSP's reverse logistics facility. The forecasts produced by the methods in this paper were used as input for the inventory management models in Calmon and Graves (2017). In Calmon et al. (2020) we address challenge (*iii*), which we call *the warranty matching problem*, where we take the inventory management policy as a given.

Forecasting failure times at the beginning of the life-cycle of a device is challenging since failure observations are limited in number, and any empirical distribution will be censored or truncated. For example, t weeks after a device's launch there cannot be observations of failures where a device's age is greater than t weeks. Nevertheless, the WSP has a large amount of historical data available about sales and failure times from different devices that it sells. In fact, since our partner WSP is one of the largest players in this market, it has data from millions of customer purchases and failures. Leveraging this information will play a key role in the estimation strategies that we develop. We note that the WSP launches around a dozen to two dozen new devices per year, each with life cycles between one and two years. As the warranty period is usually a year, the WSP is managing replacement requests for two to three years for each device, and hence is supporting around 40 different devices at any point in time.

The proposed strategy assumes that we can identify for any new device a set of existing devices that have similar age-dependent hazard rate as will the new device. An element of the set might be the entire population of devices for a particular device model, e.g., the iPhone 11 Pro, or possibly an element of a subset like the iPhone 11 Pro sold in August to customers in California.

For our proposed strategy from the set of existing devices we obtain a set of empirical hazard rate distributions as a basis. We then use regression to identify and fit the relevant hazard rates distributions from the basis to the observed hazard rates for the new device. Hence we call this a *hazard rate regression*. In addition, we show how our forecasting strategy can be applied to other settings. Namely, we validate the hazard rate regression using data from Project Repat, a social enterprise that transforms old t-shirts into quilts.

1.1. Snapshot of the WSP's data

We motivate the development of our forecast methods with the observation that the hazard rate distributions of the devices sold by the WSP have a similar temporal shape. For example, in Figure 1 we plot the normalized empirical hazard rate distributions from three devices of three different manufacturers sold by the WSP. Despite the differences between the devices, their hazard rates are somewhat similar, resembling scaled versions of each other. Note there is a sharp decrease after one year[1]. Our forecasting strategy leverages these similarities. Thus, when a new device is introduced into the market, the "prior information" available from other devices will be used to estimate the right-tail of the hazard rate distribution of the new device, for which there may be no observations.

The remainder of this paper is structured as follows. In Section 2 we present a literature review. In Section 3 we formalize our estimation problem and in Section 4 we present our estimation strategy. In Section 5 we present a few numerical experiments including one utilizing data from our partner WSP and also discuss how this strategy was used on data from Project Repat.

2. Literature Review

The most popular non-parametric approach for estimating hazard rate distributions with censored data is the Kaplan-Meier (KM) estimator (Kaplan and Meier 1958). For the estimation problem faced by our partner WSP, however, the KM estimator is not very useful for estimating the hazard rate distribution of a new device. When a new device is launched to market, all failure observations are truncated, since the oldest device sold is no older than the time past since the launch date.

[1] Our computation of the hazard rates beyond one year is inaccurate, as it should only include the population of devices with extended warranties. However, for the WSP, we do not have additional information available on the extended warranties, and hence are using the entire device population in this calculation.

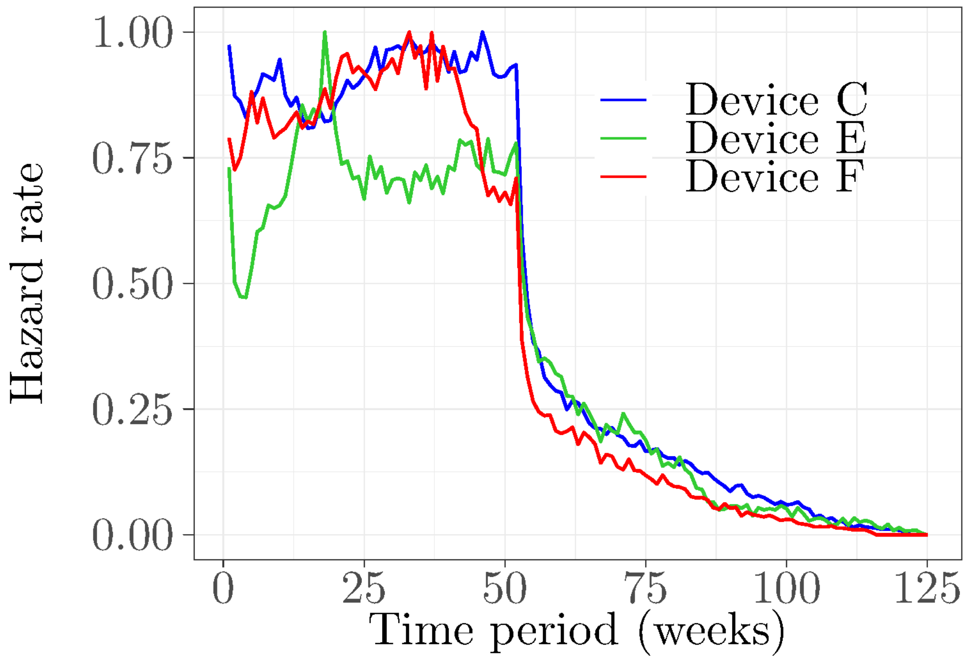


Figure 1 **Hazard Rates of three different devices sold by the WSP. The age of failures is given in weeks. The maximum of the curves was normalized to one to preserve data confidentiality.**

Thus, if the KM estimator were applied to this case, it would not be possible to estimate the right tail of the hazard rate distribution.

There are also many parametric strategies for estimating hazard rate distributions. One classic parametric approach is the Cox Proportional Hazards Model (Cox and Oakes 1984), where the hazard rate is assumed to be a scaled version of some baseline hazard function, and the scaling parameter is a function of certain covariates. This approach is commonly used in medical applications when analyzing the survival rate of a patient as a function of characteristics the patient has, such as genetic factors or pre-existing conditions. Our approach eschews device-specific features and, instead, assumes that the hazard rate of a new device can be approximated by a linear combination of the hazard rates of existing devices to “quickly” learn the new device’s hazard rate. Other parametric strategies involve the Expectation-Maximization (EM) type algorithms that have been used to estimate parameters of hazard rate distributions (McLachlan and Krishnan 2007). An EM approach was investigated in Calmon (2015) and found to perform less well compared to hazard rate regression.

Taylor (1995) introduces a semi-parametric approach for maximum-likelihood information with censored data when failures are exponentially distributed. In the operations management literature, Karim and Suzuki (2005) present a comprehensive literature review on statistical methods for analyzing warranty claim data. Hazard rate models based on warranty data have been developed for reliability studies in the automotive industry by Rai and Singh (2003) and Zhou et al. (2012, 2017). Kumar et al. (2017) propose hazard rate models to estimate the return delay distribution of remanufacturing parts in this industry.

Based on an extensive literature search, Krapp et al. (2013) and Govindan et al. (2015) conclude that forecasting in closed-loop supply chains has been understudied. Kelle and Silver (1989) propose four forecasting methods, dependent upon the available data, to estimate the returns and (net) demand during the lead time for reusable containers. For these four methods, de Brito and van der Laan (2009) investigate the impact of information with respect to the return process on inventory management. Toktay et al. (2000) use Bayesian methods to estimate parameters for the distribution of product returns and rely on a distributed lag model to capture the dependence of production returns on sales. Clottey et al. (2012) elaborate on the distributional assumptions of such lags. Tsiliyannis (2018) forecast product returns for remanufacturing based on Markov chain modeling of stock and flows. Recently, Cui et al. (2019) use four machine learning methods to predict the returns of automotive accessories. A neuro-fuzzy approach and an adaptive network based fuzzy inference system for forecasting returns have been proposed by Marx-Gómez et al. (2002) and Kumar et al. (2014) respectively.

Similar to Baardman et al. (2019) and Hu et al. (2019), we generate forecasts of new products by leveraging the information of existing and past products. However, to the best of our knowledge, our work is the first to estimate the hazard rate distribution with a regression that uses a basis of hazard rate distributions.

A detailed analysis of the importance and challenges related to forecasting warranty claims at our partner WSP's closed-loop supply chain is explored in Petersen (2013). This paper benefited from the same partnership, discussions, and data as Petersen (2013) and, because of this, shares many of its core ideas. However, while Petersen (2013) presents results and strategies tailored to the WSP, this paper frames the discussion in more generic terms, and we believe our approach has a wider range of applications.

3. Problem Set-Up

The study and development of methods for estimating failure or survival distributions of products, machines, and subjects in clinical trials have a long history, dating back to seminal work of Greenwood and others (1926) in the early 20$^{\text{th}}$ century. These methods attempt to build a distribution for the occurrence time of an event (such as age of failure of a device or death of a patient) based on a set of observations of the said event. As in most of the reliability literature, we will call the time at which an event occurs the *failure time*. In many practical settings, these observations can be *censored*, i.e., there is no information available on the exact time that a failure occurs, only that it is outside of some interval. For example, if we are trying to estimate the failure time distribution of electronic devices sold at different times during the last few months, we have censored observations in that we have yet to observe the failure times for the devices that have yet to fail. All we can say is that their failure times are at least as long as the devices' current ages. The age at which the device first fails is the *failure age* and we assume that each device has an age of 0 when sent to the customer.

We consider a discrete-time model where there is a maximum failure age T for devices, such that, for all practical purposes, devices of age larger than T will never fail. This comes from the fact that the WSP offers a warranty of at most two years, and customers that are not covered by a warranty are not entitled to receive a replacement device. In practice, the choice of T depends on the context of the estimation problem. We assume that the age at which a device first fails[2] can be described by a (initially unknown) discrete failure time distribution, $\boldsymbol{p} = (p_1, \ldots, p_T)$ where

$$\Pr(\text{failure of device at age } t) \triangleq p_t.$$

It is also useful to describe the failure process in terms of hazard rates. The hazard rate at age t is the probability that, conditioned on surviving until the beginning of age t, the device fails at age t. Thus, let $\boldsymbol{h} = (h_1, \ldots, h_T)$ be the hazard rates where

$$\Pr(\text{failure at age } t \mid \text{survived beyond age } t-1) \triangleq h_t.$$

The relationship between the hazard rate and the failure distribution is $p_1 = h_1$ and, for $t > 1$,

$$h_t = \frac{p_t}{1 - \sum_{k=1}^{t-1} p_k} \text{ and } p_t = h_t \cdot \prod_{k=1}^{t-1} (1 - h_k). \tag{1}$$

We denote the complementary cumulative distribution function (CCDF) of the failure age by $\bar{F}_t$. Then, $\bar{F}_t = \Pr(\text{failure} > \text{age } t)$ and the relationship between the hazard rate and the CCDF is

$$\bar{F}_t = \prod_{k=1}^{t} (1 - h_k).$$

To develop the forecast we assume that we have two sets of observations in hand: (i) a set of uncensored failure observations $\boldsymbol{y} = (y_1, \ldots, y_T)$, where y_t is the number of devices that failed at age t, and (ii) a set of censored failure observations $\boldsymbol{z} = (z_1, \ldots, z_T)$, where z_t is the number of devices that have yet to fail and that are of age t. The total number of observations is $\sum_{i=1}^{T} (y_i + z_i)$. Another critical assumption is that *failure age and censoring are independent*. In the WSP case, this is equivalent to assuming that the sales date and failure date are independent, since all elements of a devices fail according to the same hazard rate distribution.

[2] Similar to Calmon et al. (2020) we do not consider subsequent failures that might occur for repaired devices.

We denote the estimate of the hazard rate given a set of uncensored and censored failure observations by a vector $\hat{\boldsymbol{h}}(\boldsymbol{y},\boldsymbol{z}) = (\hat{h}_1(\boldsymbol{y},\boldsymbol{z}),\ldots,\hat{h}_T(\boldsymbol{y},\boldsymbol{z}))$. The estimate of the discrete failure time distribution from the observations y and z is given by

$$\hat{p}_t(\boldsymbol{y},\boldsymbol{z}) = \hat{h}_t(\boldsymbol{y},\boldsymbol{z}) \cdot \prod_{k=1}^{t-1}(1-\hat{h}_k(\boldsymbol{y},\boldsymbol{z})). \tag{2}$$

The corresponding estimate of the CDF is then

$$\hat{F}_t(\boldsymbol{y},\boldsymbol{z}) = \sum_{i=1}^{t} \hat{p}_i(\boldsymbol{y},\boldsymbol{z}).$$

Let $\{\boldsymbol{h}^1,\ldots,\boldsymbol{h}^m\}$ be a collection of m different hazard rate distributions from existing devices, such that $\boldsymbol{h}^j = (h_1^j,\ldots,h_T^j)$ would typically represent the hazard rates for some device j. We view this collection as a basis set for modeling the population of possible hazard rate distributions. Furthermore, we allow for each element j of the basis to be scaled by some non-negative parameter w_j, such that $w_j\boldsymbol{h}^j = (w_j h_1^j,\ldots,w_j h_T^j)$. We use the set of distributions as the basis in a mixture model for the estimation of the true hazard distribution $\boldsymbol{h}$. Namely, we assume that $\boldsymbol{h}$ can be expressed as a mixture of the scaled hazard rate distributions. That is,

$$\Pr(\text{failure at age } t|\text{survived beyond age } t-1) = h_t = \sum_{j=1}^{m} w_j \cdot h_t^j, \tag{3}$$

where $w_j \geq 0$ is the weight for the scaled basis on the basis element j and, for all t, $\sum_{j=1}^{m} w_j \cdot h_t^j \leq 1$, so that the resulting hazard rates are meaningful.

By using (1) the probability of failure of a device at age t in this case is

$$p_t = \left(\sum_{j=1}^{m} w_j h_t^j\right) \cdot \prod_{k=1}^{t-1}\left(1-\left(\sum_{j=1}^{m} w_j h_k^j\right)\right). \tag{4}$$

Hence, our goal is to estimate $\boldsymbol{w} = (w_1,\ldots,w_m)$. From a practical standpoint, $\boldsymbol{w}$ allows a practitioner to identify if the failure distribution of a newly launched device is a more intense or subdued version of the hazard rate distributions in the basis.

We use $\hat{\boldsymbol{w}}(\boldsymbol{y},\boldsymbol{z}) = (\hat{w}_1(\boldsymbol{y},\boldsymbol{z}),\ldots,\hat{w}_m(\boldsymbol{y},\boldsymbol{z}))$ to denote the estimate for $\boldsymbol{w}$. We consider an estimation strategy to be effective if, given the observations $\boldsymbol{y}$ and $\boldsymbol{z}$, we have

$$\hat{w}_j(\boldsymbol{y},\boldsymbol{z}) \to w_j, \forall j \text{ almost surely as } \sum_{t=1}^{T} y_t \to \infty \text{ and } \sum_{t=1}^{T} z_t \to \infty.$$

Although this is a parametric approach, we make no explicit assumptions on the underlying shape of the failure time distribution. This is a departure from other models, such as the Cox Proportional Hazards model, that assume a specific underlying distribution. Nonetheless, our approach requires the identification of a reasonable basis such that it will be representative of the actual hazard rate distribution to be estimated. Our approach is similar to the additive hazards model proposed is Lin and Ying (1994) and further described in Klein and Moeschberger (2006). However, while the additive model in Lin and Ying (1994) assumes a base hazard rate distribution that is "shaped" by a linear combination of covariates, we explicitly assume that the hazard rate distribution we wish to estimate is in the convex cone of a set of basis hazard distributions.

In the next section, we introduce and discuss our estimation strategy.

4. Maximum Likelihood Estimator and Hazard Rate Regression

With the basis $\{\boldsymbol{h}^1,\dots,\boldsymbol{h}^m\}$ as defined in the prior section we are ready to introduce our strategy for estimating the hazard rates of a new device. We assume that the failure age of a new device has a finite discrete support $[1,T]$ and that observations are truncated at some age $\tau \le T$, such that $y_t = z_t = 0$ for $t > \tau$.

We present our estimation strategy in three parts. In the first part we describe the Kaplan-Meier (KM) Estimator, which is the non-parametric Maximum-Likelihood Estimator (MLE) for hazard rates. The usefulness of the KM estimator for the WSP's problem is limited since we cannot observe failure ages greater than τ. To overcome this, we assume that $\boldsymbol{h}$ can be expressed as a mixture of historical hazard rates as in Equation (3). We consider two ways to determine the mixture. In the second part, we introduce the MLE that assumes that the hazard rates are a combination of the basis $\{\boldsymbol{h}^1,\dots,\boldsymbol{h}^m\}$. One way is to find the mixture that maximizes the likelihood function under the assumption that $\boldsymbol{h}$ is of this functional form. The resulting MLE involves solving a concave optimization problem. The third part approximates this concave optimization problem through in a simpler problem which we denote *hazard rate regression*. In this way we choose the mixture to match the KM estimator as closely as possible.

4.1. The Kaplan-Meier Estimator

The likelihood, L, of some sample $(\boldsymbol{y},\boldsymbol{z})$ with hazard rates $\boldsymbol{h}$ is

$$L(\boldsymbol{h};\boldsymbol{y},\boldsymbol{z}) = \Pr(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{h}) = \prod_{t=1}^{\tau}\left(h_t \cdot \prod_{k=1}^{t-1}(1-h_k)\right)^{y_t} \cdot \left(\prod_{k=1}^{t}(1-h_k)\right)^{z_t}.$$

By rearranging the products and defining $r_t = \sum_{k\ge t} y_k + z_k$ we obtain the log-likelihood function

$$\log(L(\boldsymbol{h};\boldsymbol{y},\boldsymbol{z})) = \sum_{t=1}^{\tau} y_t \cdot \log h_t + (r_t - y_t)\log(1-h_t).$$

If we make no assumptions on the functional form of $\boldsymbol{h}$, the non-parametric estimator for the hazard rates solves

$$\max_{\boldsymbol{h}} \quad \sum_{t=1}^{\tau} y_t \cdot \log(h_j) + (r_t - y_t)\log(1-h_j)$$

The solution to the optimization problem above, which we denote by $\boldsymbol{h}^{KM} = (h_1^{KM}, \ldots, h_\tau^{KM})$, is commonly known as the Kaplan-Meier non-parametric estimate of the hazard rates. For $t \leq \tau$ we can write h_t^{KM} in closed form as

$$h_t^{KM}(\boldsymbol{y}, \boldsymbol{z}) = \frac{y_t}{r_t}, \tag{5}$$

While this approach yields a simple estimator, we cannot use this approach to estimate hazard rates for periods after τ. As a result, the value of this estimator for WSP's estimation problem is limited, in particular early in a device's life-cycle when most failure observations are truncated.

4.2. Parametric Maximum Likelihood Estimator

We now assume that a device's hazard rate is a weighted combination of elements of the basis, i.e. we introduce a variable h_t such that $h_t = \sum_{j=1}^m w_j h_t^j$. The likelihood of the weights $\boldsymbol{w}$ given observations $(\boldsymbol{y}, \boldsymbol{z})$ is

$$\begin{aligned} L(\boldsymbol{w}; \boldsymbol{y}, \boldsymbol{z}) = \Pr(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{w}) &= \prod_{t=1}^{\tau} \left(h_t \cdot \prod_{k=1}^{t-1} (1 - h_k) \right)^{y_t} \cdot \left(\prod_{k=1}^{t} (1 - h_k) \right)^{z_t} \\ &= \prod_{t=1}^{\tau} \left[\left(\sum_{j=1}^m w_j h_t^j \right) \cdot \prod_{k=1}^{t-1} \left(1 - \sum_{j=1}^m w_j h_k^j \right) \right]^{y_t} \cdot \left[\prod_{k=1}^{t} \left(1 - \sum_{j=1}^m w_j h_k^j \right) \right]^{z_t}. \end{aligned}$$

If $\tau > 1$ the log-likelihood then becomes

$$\begin{aligned} \log(L(\boldsymbol{w}; \boldsymbol{y}, \boldsymbol{z})) &= \sum_{t=1}^{\tau} y_t \cdot \left(\log \left(\sum_{j=1}^m w_j h_t^j \right) + \sum_{k=1}^{t-1} \log \left(1 - \sum_{j=1}^m w_j h_k^j \right) \right) + z_t \cdot \sum_{k=1}^{t} \log \left(1 - \sum_{j=1}^m w_j h_k^j \right) \\ &= \sum_{t=1}^{\tau} y_t \cdot \log \left(\sum_{j=1}^m w_j h_t^j \right) + z_t \log \left(1 - \sum_{j=1}^m w_j h_t^j \right) + (y_t + z_t) \sum_{k=1}^{t-1} \log \left(1 - \sum_{j=1}^m w_j h_k^j \right). \end{aligned}$$

By rearranging the summations and using $r_t = \sum_{k \geq t} y_k + z_k$ we obtain

$$\log(L(\boldsymbol{w}; \boldsymbol{y}, \boldsymbol{z})) = \sum_{t=1}^{\tau} y_t \cdot \log \left(\sum_{j=1}^m w_j h_t^j \right) + (r_t - y_t) \log \left(1 - \sum_{j=1}^m w_j h_t^j \right).$$

The function above is concave in $\boldsymbol{w}$ since the logarithm of an affine function is concave. The Maximum-Likelihood Estimator of the weights, which we denote by $\boldsymbol{w}^{ML}$, is the optimizer of the concave optimization problem

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \sum_{t=1}^{\tau} y_t \cdot \log \left(\sum_{j=1}^m w_j h_t^j \right) + (r_t - y_t) \log \left(1 - \sum_{j=1}^m w_j h_t^j \right) \\ \text{s.t.} \quad & \sum_{j=1}^m w_j h_t^j \leq 1, \; t = 1, \ldots, T, \\ & \boldsymbol{w} \geq 0. \end{aligned} \tag{6}$$

The constraints above ensure that the resulting estimates are meaningful hazard rates.

Although the optimization problem in Equation 6 is convex (since we maximize a concave objective with linear constraints), it still requires a specialized numerical optimization solver. Next, we introduce a simpler approach.

4.3. Hazard Rate Regression

The hazard rate regression approach has two steps: (i) calculate the empirical quantiles with the available data using a Kaplan-Meier estimator and (ii) use a regression to calculate the weights in our model. We motivate this approach using both the MLE and the Kaplan-Meier estimator. To do so, we re-write the optimization problem in (6) as

$$\begin{aligned} \max_{\boldsymbol{w},\boldsymbol{h}} \quad & \sum_{t=1}^{\tau} y_t \cdot \log(h_t) + (r_t - y_t)\log(1-h_t) \\ \text{s.t.} \quad & \sum_{j=1}^{m} w_j h_t^j = h_t,\, t = 1,\ldots,\tau \\ & \sum_{j=1}^{m} w_j h_t^j \le 1,\, t = 1,\ldots,T, \\ & \boldsymbol{w} \ge 0. \end{aligned} \tag{7}$$

The objective function of the optimization problem above is the same as in Equation (6). Thus, if there exists a $\tilde{\boldsymbol{w}} = (\tilde{w}_1, \ldots, w_n)$ that satisfies the system of constraints

$$\begin{aligned} & \sum_{j=1}^{m} \tilde{w}_j h_t^j = h_t^{KM},\, t = 1,\ldots,\tau \\ & \sum_{j=1}^{m} \tilde{w}_j h_t^j \le 1,\, t = 1,\ldots,T, \\ & \tilde{\boldsymbol{w}} \ge 0, \end{aligned} \tag{8}$$

then $\tilde{\boldsymbol{w}} = \boldsymbol{w}^{ML}$. This is because the KM estimator maximizes the objective function in (7); hence if we can express the KM estimator as a mixture of basis elements, then it is an optimal solution to (7). However, there may be no feasible solution to (8). To address this we can relax the equality constraint in (8) and minimize the "relaxation error" by solving

$$\begin{aligned} \min_{w} \quad & \left\| \sum_{j=1}^{m} w_j \boldsymbol{h}^j - \boldsymbol{h}^{KM}(\boldsymbol{y}, \boldsymbol{z}) \right\|_{p,\tau} \\ \text{s.t.} \quad & \sum_{j=1}^{m} w_j h_t^j \le 1, t = 1,\ldots,T, \\ & w_j \ge 0, j = 1,\ldots,m, \end{aligned} \tag{9}$$

where $\|\cdot\|_{p,\tau}$ is the p-norm of the first τ components of the vector. Namely, for some vector $\boldsymbol{x}$,

$$\|\boldsymbol{x}\|_{p,\tau} = \left(\sum_{i=1}^{\tau} |x_i|^p \right)^{1/p}.$$

Let $\boldsymbol{w}^{HR}$ be the optimal solution to the problem in (9). Then, the hazard rate regression approach sets $\hat{\boldsymbol{w}} = \boldsymbol{w}^{HR}$ and $\hat{\boldsymbol{h}} = \sum_{j=1}^{m} \hat{w}_j \boldsymbol{h}^j$. We have $\boldsymbol{w}^{HR} = \boldsymbol{w}^{ML}$ when the system of equations in (8) has a solution. Furthermore, when $p = 1$ or $p = 2$, any linear or quadratic numerical optimizer solves (9).

Simply put, the hazard rate regression approach finds a point in the cone generated by the basis of hazard rate vectors that is a feasible hazard rate distribution and minimizes the distance to the KM estimates. If we assume that the original samples were drawn from the mixture model, and if $\tau = T$, $\hat{\boldsymbol{h}}$ will converge to $\boldsymbol{h}$ as the number of samples goes to infinity. We prove this result regarding the convergence of $\hat{\boldsymbol{h}}$ in Appendix A.

Finally, the optimization in Equation (9) is useful for *model selection*, i.e., for identifying which of the basis elements are the most relevant based on the values of $\boldsymbol{w}$. In the next section we will analyze the hazard regression approach and the MLE approach through numerical experiments.

5. Numerical Experiments

We now examine the performance of the MLE and hazard rate regression through a set of numerical experiments. The first set of experiments corresponds to an artificial set-up where we compare, in a controlled setting, the performance of hazard rate regression to the parametric MLE approach from Section 4.2, and to a "naive" estimator, where we assume failures ages follow a Weibull distribution. The second set of experiments uses data from our partner WSP and the goal is to forecast the amount of weekly customer warranty claims it receives for six different devices. The third set of experiments uses data from Project Repat, a company that transforms old t-shirts into quilts.

A first performance metric that we use is comparing the maximum distance between the estimated Cumulative Distribution Function (CDF) and the true CDF. Given a hazard rate distribution estimate $\hat{\boldsymbol{h}}$ we denote the CDF estimate by $\hat{\boldsymbol{F}} = (\hat{F}_1, \ldots, \hat{F}_T)$, where $\hat{F}_t$ is

$$\hat{F}_t = 1 - \prod_{k=1}^{t} (1 - \hat{h}_k).$$

If the true CDF is $\boldsymbol{F} = (F_1, \ldots, F_T)$, we have that the maximum distance between the true and estimated CDFs is

$$\max_t |F_t - \hat{F}_t|.$$

We call this distance the *Kolmogorov-Smirnov (KS) distance*, since it has the same form of the Kolmogorov-Smirnov statistic for an empirical distribution.

The second performance measure is a measure of the forecast accuracy of each estimation method. We assume that sales of devices happen in some interval $[1, T_s]$, such that there are no sales after time T_s. Let the sales in each period be denoted by s_τ for $\tau = 1, \ldots, T_s$, and $\tau = 1$ is when the device is launched. For our experiments, we assume that these sales are known. We denote the estimate of the hazard rate distribution τ periods after launch by $\boldsymbol{\hat{h}^\tau} = (\hat{h}_1^\tau, \ldots, \hat{h}_T^\tau)$.

In this setup, we expect the last period in which failures occur to be $T + T_s$. Hence, we define $T_{max} = T + T_s$ to be the length of time that the WSP has to manage warranty claims for a device. We recall the assumption that a device fails at most once and that we do not model the possible subsequent failures of refurbished devices. Let the age distribution of devices be $\boldsymbol{x}(\tau) = (x_1(\tau), \ldots, x_T(\tau))$ where $x_t(\tau)$ represents the number of surviving devices of age t at the beginning of period τ. Then, we have that $\boldsymbol{x}(1) = (s_1, 0, \ldots, 0)$ and $x_1(\tau) = s_\tau, \forall t$ and that

$$E[x_{t+1}(\tau+1)|\boldsymbol{h}] = (1 - h_t) \cdot x_t(\tau), \forall t = 2, \ldots, T; \forall \tau = 1, \ldots, T_{max-1}.$$

Furthermore, let the number of failures during period τ be given by a vector $\boldsymbol{f}(\tau) = (f_1(\tau), \ldots, f_T(\tau))$, where $f_t(\tau)$ is the number of failures of age t in period τ. Then, given $\boldsymbol{h}$ we have

$$E[f_t(\tau)|\boldsymbol{h}] = h_t \cdot x_t(\tau), \forall t = 1, \ldots, T; \forall \tau = 1, \ldots, T_{max-1}, \tag{10}$$

and the expected total number of failures in period τ will be $\sum_t h_t \cdot x_t(\tau)$.

We leverage the expectations above to create a forecast of the number of failures as follows. Assume we are in period τ and that $\boldsymbol{\hat{h}}^\tau$ is the current hazard rate distribution estimate. For each $k \in [\tau+1, T_{\max}]$, the forecast of surviving devices, which we denote by $\boldsymbol{\hat{x}}(k)$, and the forecast of total failures, which we denote by $\hat{f}(\tau)$ are given by the equations

$$\hat{x}_{t+1}(k) = \begin{cases} (1 - h_t) \cdot x_t(\tau), \text{ if } k = \tau + 1, \\ (1 - h_t) \cdot \hat{x}_t(k-1), \text{ if } k \in [\tau+2, T_{\max}], \end{cases} \qquad \text{for } t = 1, \ldots, T,$$
$$\hat{f}(k) = \sum_t h_t \cdot \hat{x}_t(k).$$

At time τ, for each $k \in [\tau+1, T_{\max}]$, We denote the true number of failures at time τ by $f(\tau)$ and the forecasting error at time τ by $e_\tau = \hat{f}(\tau) - f(\tau)$. We evaluate forecast performance using the Mean Absolute Scaled Error (MASE) introduced in Hyndman and Koehler (2006). Given that we observe the true number of failures until time τ, the MASE for the remaining forecasting horizon $[\tau+1, T_{\max}]$ is

$$MASE(\tau) = \frac{\frac{1}{T_{max}-\tau} \sum_{k=\tau+1}^{T_{max}} |e_k|}{\frac{1}{T_{max}-\tau} \sum_{k=\tau+1}^{T_{max}} |f(k) - f(k-1)|}.$$

The denominator above is the average forecasting error of a "naive" one-step estimator that sets $\hat{f}(\tau+1) = f(\tau)$.

5.1. Experiments with simulated data

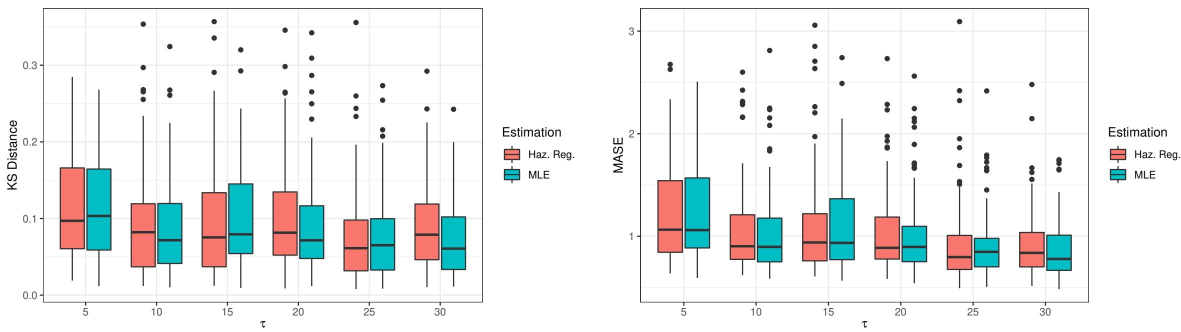
The first set of experiments consists of comparing the MLE estimator, the hazard rate regression, and a "naive" estimator that fits a Weibull distribution to the data in a controlled setting. A time period in the simulation corresponds to one week. Additionally, we assume that $T = 100$ and that all devices are sold at time $t = 0$, and $t = 1$ is the first period in which a device can fail. We consider a basis set of 30 hazard rate distributions that will be used to estimate the hazard rate of a new device. The basis elements are generated as follows

1. For each basis element, we sample integers a and b from a discrete uniform distribution with parameters $[1, T]$. Furthermore, we sample a value p from a uniform distribution with parameters $[0, 1]$;
2. We model the failure time for the basis element as a mixture of a discrete uniform random variable with parameters $[1, a]$ and an exponential random variable with mean b. The mixture probability is p;
3. We draw 100 samples from the distribution in the previous step, which represent the failure time of 100 devices. Namely, with probability p we draw a device's failure time according to a discrete uniform with $[1, a]$ and with probability $1 - p$ we draw according to the exponential distribution;
4. We censor the samples in the previous step. For each failure time observation x, we sample a censoring random variable y from a uniform distribution on the interval $[0, T]$. We observe a failure of age x when $x \leq y$, and we observe a censored observation, namely a device that has yet to fail, when $x > y$;
5. We use these 100 observations with the KM estimator to estimate the hazard rate distribution. This estimate of the hazard rate distribution becomes an element of the basis.

The above procedure is repeated 30 times to generate the basis elements. Furthermore, this reflects the data available to the WSP. In practice, the WSP does not have access to a set of "real" hazard rates, it only has the hazard rates estimated from past device failures. For the new device, the target of the estimation problem, we also generate failure times observations for 100 devices (potentially censored) in the same fashion as described above.

5.1.1. Comparing the Different Forecasting Methods We evaluate the performance of the two estimation strategies for different times that the new device is released. We run our experiments for different values of τ, which is the number of periods since the launch of the new device. For a given τ, we modify the failure time observation for of the new devices to be $\min(x, y, \tau)$ and we observe a censored observation when $\tau = \min(x, y, \tau)$.

We then estimate for each value of τ the hazard rate distribution of the new device using both the regression approach and the MLE. For each value $\tau \in \{5, 10, 15, 20, 25, 30\}$, we run 100 test cases.



(a) KS Distance (b) MASE

Figure 2 **Boxplot of the KS distance and MASE between the true hazard rate and the estimated hazard rate for different values of τ. The boxes range from the first quartile to the third quartile and include the median. The boundaries of the whiskers (lines) are based on 1.5 times the interquartile range and all other points represent outliers.**

For each test case we generate a new device and a basis set of 30 hazard rate distributions using the sampling strategy described in the beginning of this subsection. As τ increases, the data availability on the true number of failures of the new device on the interval $[0,\tau]$ increases. Furthermore, a larger τ implies producing less forecasts as the forecasting horizon $[\tau+1,T]$ becomes smaller.

The results for the KS distance metric and the MASE are summarized in Figure 2a. For $\tau = 5$, the hazard regression method produces a KS distance of 0.10 on average. Thus, this means that the hazard regression approach produces already a reasonable estimate of the failure distribution at 5 periods after launch. For the MLE, we obtain an average KS distance which is comparable to the hazard regression method for every τ. In Figure 2b, we display the results for the MASE calculations. For the median values, both forecasting methods perform better than a one-step naive forecast when $\tau \geq 10$. Based on these results, we have similar findings as for the KS distance: both methods perform comparably according to the average MASE for every τ. This means that both forecasting methods converge to the right weights as we increase τ under the assumption that the true hazard rate is a mixture of the true basis elements.

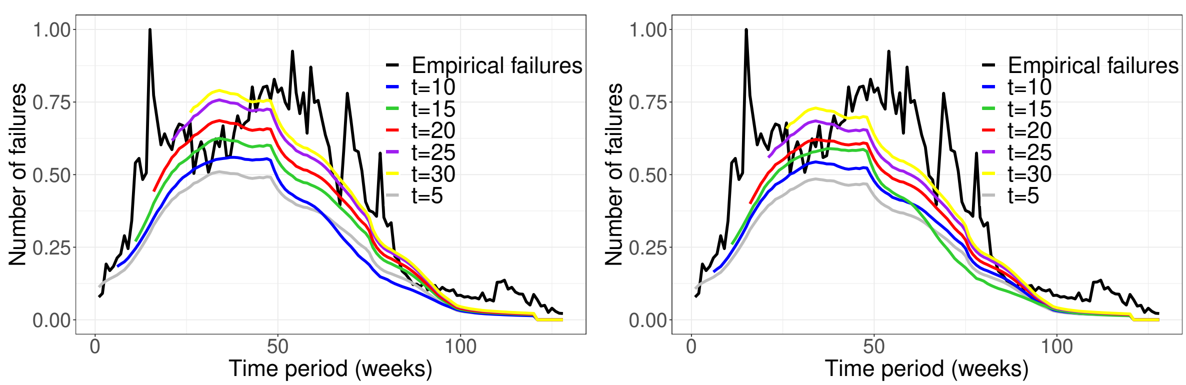
For the hazard regression method, none of the observations in the numerical tests used more than 13 basis elements and the average number of elements selected was 5.0. For the MLE, we observed a maximum of 12 basis elements and an average of 5.3 elements selected.

5.1.2. Sensitivity analysis Table 1 shows the sensitivity to a change in the number of basis elements for 100 replications. For sensitivity analysis we vary the size of the basis and find that the hazard regression method and the MLE keep on performing comparably according to the KS distance as well as the MASE. We observe that these performance measures improve when the number of basis elements increases from 10 to 30 while increasing the number of basis elements from 30 to 50 does not a guarantee an improved performance.

Forecasting method	τ	Performance measure	10 basis elements	30 basis elements	50 basis elements
Hazard regression	5	KS distance	0.13	0.10	0.10
		MASE	1.23	1.09	1.07
	10	KS distance	0.11	0.09	0.09
		MASE	1.11	1.00	1.04
	15	KS distance	0.10	0.09	0.09
		MASE	1.07	0.96	1.02
	20	KS distance	0.11	0.07	0.07
		MASE	1.05	0.88	0.88
	25	KS distance	0.09	0.07	0.07
		MASE	0.97	0.80	0.86
	30	KS distance	0.09	0.06	0.06
		MASE	0.88	0.78	0.79
MLE	5	KS distance	0.11	0.10	0.10
		MASE	1.12	1.02	1.07
	10	KS distance	0.10	0.09	0.09
		MASE	1.11	1.00	1.04
	15	KS distance	0.11	0.08	0.08
		MASE	1.08	0.95	1.08
	20	KS distance	0.10	0.08	0.07
		MASE	1.05	0.88	0.88
	25	KS distance	0.10	0.07	0.07
		MASE	0.98	0.82	0.88
	30	KS distance	0.08	0.07	0.07
		MASE	0.86	0.75	0.88

Table 1 **Median of the KS distance and MASE of each forecasting method for an increase or decrease of 20 basis elements and for $\tau \in \{5, 10, 15, 20, 25, 30\}$.**

5.2. Forecasting failures at the WSP



(a) Number of weekly failures - HR approach (b) Number of weekly failures - MLE approach

Figure 3 **Number of forecasted and true failures per week for different τ for device E. The maximum of the curves was normalized to one to preserve data confidentiality.**

	Hazard regression					
Device	$\tau = 5$	$\tau = 10$	$\tau = 15$	$\tau = 20$	$\tau = 25$	$\tau = 30$
A	0.16	0.13	0.11	0.10	0.09	0.07
B	0.12	0.11	0.10	0.09	0.09	0.08
C	0.09	0.08	0.07	0.07	0.06	0.06
D	0.08	0.08	0.07	0.06	0.06	0.05
E	0.15	0.15	0.14	0.13	0.12	0.10
F	0.15	0.14	0.12	0.12	0.10	0.10
	MLE approach					
Device	$\tau = 5$	$\tau = 10$	$\tau = 15$	$\tau = 20$	$\tau = 25$	$\tau = 30$
A	0.16	0.13	0.11	0.10	0.09	0.07
B	0.12	0.11	0.10	0.09	0.09	0.08
C	0.09	0.08	0.07	0.07	0.07	0.07
D	0.08	0.08	0.03	0.06	0.06	0.06
E	0.15	0.14	0.13	0.12	0.11	0.10
F	0.15	0.14	0.13	0.13	0.12	0.11

Table 2 **KS distance for forecasting failures at the WSP for $\tau \in \{5, 10, 15, 20, 25, 30\}$**

	Hazard regression					
Device	$\tau = 5$	$\tau = 10$	$\tau = 15$	$\tau = 20$	$\tau = 25$	$\tau = 30$
A	1.25	1.21	1.12	1.06	0.88	0.85
B	1.04	0.99	0.97	0.92	0.87	0.88
C	1.11	0.99	0.92	0.92	0.89	0.85
D	1.16	1.04	1.02	0.94	0.96	0.84
E	1.40	1.33	1.22	0.94	0.87	0.86
F	1.29	0.90	0.85	0.83	0.76	0.69
	MLE approach					
Device	$\tau = 5$	$\tau = 10$	$\tau = 15$	$\tau = 20$	$\tau = 25$	$\tau = 30$
A	1.26	1.22	1.16	1.07	0.88	0.84
B	1.04	1.00	0.96	0.95	0.90	0.88
C	1.12	0.99	0.93	0.93	0.85	0.83
D	1.14	1.04	1.03	0.95	0.84	0.83
E	1.37	1.25	1.17	0.95	0.88	0.85
F	1.33	0.90	0.86	0.85	0.79	0.71

Table 3 **MASE for the number of forecasted failures $\tau \in \{5, 10, 15, 20, 25, 30\}$**

We consider the problem of estimating the weekly number of failures for six device models sold by our partner WSP. We display the number of sold and failed devices for each model in Figure 6 in Appendix B. These devices were made by four different manufacturers. We take the hazard rate distributions estimated from five other devices as the set of basis hazard rate distributions in our estimation. We estimate the true hazard rate distribution for each device from the entire failure history for the device using the KM estimator.

We set T as 100 weeks and $T_{\max}$ as 150 weeks, since more than 95% of sales happen during the first 50 weeks from launch. We index the time periods $\tau = 1$ being the launch time period. As of any time τ, we estimate the hazard rate distribution h^τ, based on having observed all of the sales and failures up until that time. We then can use the hazard rate distribution to forecast all future failures until time period T, using Equation (10) and the assumption that we have a perfect forecast of weekly sales for time periods $\tau + 1, \ldots, T$.

The results for the hazard regression and for the MLE for all devices are displayed in Table 2 and in Table 3. After receiving 25 weeks of failure data, both estimators produce good results, having a maximum KS distance of 0.12 and maximum MASE of about 0.90. Note that 25 weeks is still at the beginning of the warranty life-cycle of the device, which is usually more than 100 weeks long.

Figure 3 displays the normalized number of failures by week for device E and the forecasts for different weeks from initial launch. The yearly failure probability of device E is 8.9%. In general, as the amount of information increases, both forecasting strategies improve to capture the empirical failure distribution and both forecasting strategies keep on performing comparably. When forecasting in an early stage of the life cycle, the number of failed devices that will need to be handled by the WSP is underestimated for both strategies.

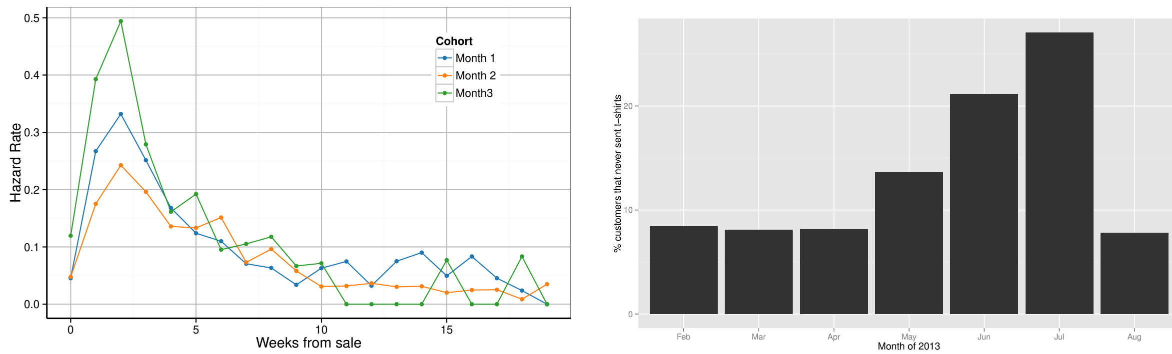
In many applications, such as at our partner WSP, estimating a hazard rate distribution as a mixture of the hazard rate distribution of other devices is also useful to identify which manufacturers and/or features lead to large number of failures. For example, this estimation strategy can help identify if devices with similar operating systems have similar hazard functions. Additionally, this strategy can quickly help identify if a recently launched device has an unusually high (or low) failure rate.

A tailored version of the estimation strategy was implemented at our partner WSP, and the implementation is described in Petersen (2013). In addition, a plug-in was developed for Microsoft Outlook and Excel that allowed managers at the reverse logistics facility to forecast the amount of failures and also provided an estimate of inventory needs.

5.3. Forecasting returns at Project Repat

The third set of numerical experiments uses data from Project Repat, a social enterprise in the Boston area that transforms old t-shirts into quilts. Their product is popular among college students and recent graduates that want to preserve their college (or fraternity/sorority) t-shirts, and among athletes, particularly runners, who collect t-shirts from races. Depending on the season of the year, Project Repat can sell from hundreds to thousands of quilts per week.

The dynamics of Project Repat's customer-facing operation is as follows: (i) customers "purchase" and pay for a quilt on Project Repat's website; (ii) Project Repat registers the order and sends the



(a) Hazard Rate Distribution (b) Customers that never sent their t-shirts

Figure 4 **Panel (a) is the hazard rate distribution of the time until the customer sends the t-shirts. Panel (b) is the fraction of customers that never sent their t-shirts.**

customer a pre-paid envelope; (iii) the customer puts old t-shirts in the envelope and sends it to Project Repat; (iv) the t-shirts are received, cut, sewn into a quilt; and (v) the quilt is shipped to the customer.

Project Repat puts a high value on the social and environmental impact of their work. Besides being a company that *upcycles* old t-shirts, Project Repat contracts all of its sewing to textile plants in the United States as an attempt to "repatriate the textile industry". Finally, this company actively works with NGOs that employ individuals with disabilities and that have limited employment opportunities.

A major issue in Project Repat's operations is forecasting the volume of envelopes with t-shirts that they receive from the customers that purchased a quilt online. More specifically, they use these forecasts to decide how many working-hours they should contract from textile plants; and if the volume of work needed exceeds the amount contracted, they have to pay overtime.

In this context, we use the hazard rate regression strategy to estimate the lead time between Project Repat mailing an envelope to the customer, and the customer sending back their old t-shirts. We model this lead time as a random variable, having a similar interpretation as the failure time. Previously we were estimating the failure time of devices, here we are estimating the *customer return time*, i.e., how many days (or weeks) after purchasing a product do customers send in their t-shirts.

The hazard rate by week for the time until customers send in their t-shirts is depicted in Figure 4a for three sample months of customer purchases. Note that these hazard rates appear to be scaled versions of each other. Also, from the return data, we have that customers take between 2 and 3 weeks on average to send their t-shirts, if they send it in at all. However, the lead time is heavy tailed, and over 20% of customers that eventually send their t-shirts take more than 5 weeks to send them.

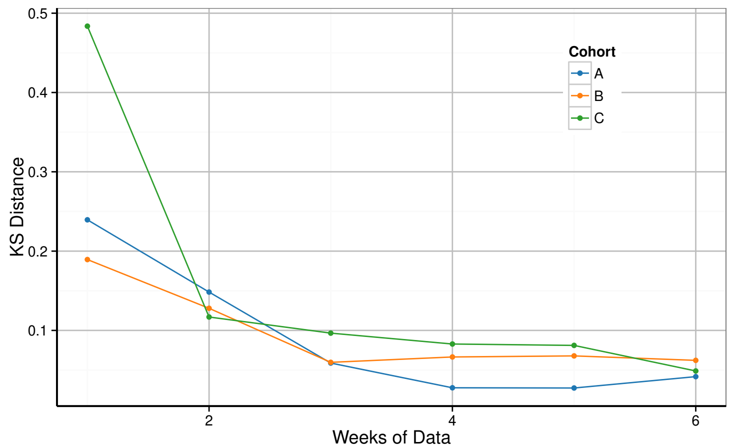


Figure 5 **KS distance for estimates from different cohorts for different amounts of information available.**

In Figure 4b, we have the fraction of customers that never sent their t-shirts. This seems to be the driving factor that makes the hazard rate distributions dissimilar. The fraction of customers that never send t-shirts is driven by two factors: (i) seasonality effects - quilts purchased as gifts have a lower percentage of returns; (ii) the promotion in place - coupons and discounts attract customers with a lower return rate.

For our estimation, we define a cohort as the customers that purchase quilts in a given week. The goal is to estimate the weekly hazard rate of a customer sending in the t-shirts. In our experiment, we use $T = 24$ weeks and use the empirical hazard rate from 35 prior weeks as the basis set. Each basis element corresponds to the empirical hazard rate associated with sales in a particular week. We chose 3 cohorts (A, B, C) to estimate; each cohort is from a different month, in order to ensure that they are not too similar, and they were all chosen from time periods long after the cohort weeks in the basis.

We observe the performance of the estimation strategies for different amounts of information available, i.e., different values of τ. As shown in Figure 5, for all 3 cohorts, with two weeks of information (out of a horizon of 24 weeks) the KS distance is less than 0.15 and, with three weeks of information, the distance is less than 0.10. Since all customers in a cohort purchase a quilt at the same time, the KS distance is the same as the maximum error of the cumulative estimate of the number of t-shirts sent per week. The estimation procedure also led to sparse representations, and for all cohorts and values of τ no estimate used more than seven elements of the basis.

This estimation procedure was built into a cloud-based forecasting tool that was given to Project Repat. Through the tool, the company can forecast the volume of t-shirts received given open pending orders (analogous to devices of different ages that have yet to fail) of each cohort in the system. Also, by analyzing the basis selected by the estimation procedure, Project Repat can identify which weeks best represent a new cohort, and use this to try to identify what influences the customer lead time.

6. Conclusion

We propose a method for estimating time-based distributions for product and service events, in contexts where there is relevant historical data on these distributions from comparable events for related products or services. The proposed strategy can account for censored observations and the forecasts improve as more observations are collected. This may lead to early information about how a new product is similar or not to prior products.

This estimation strategy, called hazard rate regression, uses a model selection method, where we assumed a basis set of hazard rate distributions determined from historical data. We use hazard rate regression to identify and fit the relevant hazard rates distributions from the basis to the observed events (failures) from the new cohort.

We compare hazard rate regression with a Maximum Likelihood Estimator (MLE) to estimate the parameters of a mixture model. Here, the fundamental underlying assumption is that the hazard rates of a new product can be modeled as a mixture of scaled hazard rate distributions built from historical data.

We examine both estimation strategies through a series of numerical experiments using hypothetical data as well as data from our partner WSP, and data from Project Repat, a Boston-based social enterprise that transforms old t-shirts into quilts. We introduce and apply different metrics to measure the quality of the forecast in these experiments. We also describe how hazard rate regression can be used to forecast the volume of warranty requests received by our partner WSP.

These two different applications provide evidence that the hazard rate regression approach is suitable to forecast failure times of new products and return lead times for online rental businesses. A fundamental requirement is a common time-based pattern in how failures or returns occur. For the WSP, discovering this pattern for the new offering is key to forecast the load and capacity requirements for conducting repair services, so as to do operational planning. Additionally, we foresee that our proposed method is also applicable to products or services with a seasonal or cyclic demand. Finally, we believe that our method can also be extended to entirely different application areas such as forecasting mortality rates of a new epidemic given reliable data of survival curves on prior epidemics.

There are a few open problems that we have yet to examine. First, a more thorough theoretical characterization of the hazard rate regression procedure, and an analysis of its connections with other estimation strategies may lead to a deeper understanding of its advantages and disadvantages. A second problem is the connection between estimation and operational decisions such as inventory management. For example, in the original setting for the WSP, it is not clear if the presence of censored information leads to policies that oversell items, or a policy that undersells items. Finally, investigating how the basis in the estimation impacts overall estimation quality can lead to a more precise guideline for defining cohorts and selecting the basis used for estimation.

References

Apple Inc (2017) Apple inc., form 10-k 2017. http://investor.apple.com/secfiling.cfm?filingID=320193-17-70&CIK=320193.

Baardman L, Levin I, Perakis G, Singhvi D (2019) Predicting product return volume using machine learning methods. *Available at SSRN. https://ssrn.com/abstract=3086237* URL http://dx.doi.org/10.2139/ssrn.3086237.

Calmon AP (2015) *Reverse logistics for consumer electronics: Forecasting failures, managing inventory, and matching warranties*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Calmon AP, Graves SC (2017) Inventory management in a consumer electronics closed-loop supply chain. *Manufacturing & Service Operations Management* 19(4):568–585, URL http://dx.doi.org/10.1287/msom.2017.0622.

Calmon AP, Graves SC, Lemmens S (2020) Warranty Matching in a Consumer Electronics Closed-Loop Supply Chain. *Manufacturing & Service Operations Management* ISSN 1523-4614, URL http://dx.doi.org/10.1287/msom.2020.0889, publisher: INFORMS.

Clottey T, Benton Jr QC, Srivastava R (2012) Forecasting product returns for remanufacturing operations. *Decision Sciences* 43(3):589–614, URL http://dx.doi.org/10.1111/j.1540-5915.2012.00362.x.

Cox DR, Oakes D (1984) *Analysis of survival data*, volume 21 (CRC Press), URL http://books.google.com/books?hl=en&lr=&id=Y4pdM2soP4IC&oi=fnd&pg=PR9&dq=cox+hazards+model&ots=hKOigEcQmk&sig=3gKfxZkoQ5Dy3d3yd4li8BQwSco.

Cui H, Rajagopalan S, Ward AR (2019) Predicting product return volume using machine learning methods. *European Journal of Operational Research. Accepted for Publication* URL http://dx.doi.org/10.1016/j.ejor.2019.05.046.

de Brito MP, van der Laan EA (2009) Inventory control with product returns: The impact of imperfect information. *European Journal of Operational Research* 194(1):85–101, URL http://dx.doi.org/10.1016/j.ejor.2007.11.063.

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(3):603–626, URL http://dx.doi.org/10.1016/j.ejor.2014.07.012.

Greenwood M, others (1926) A report on the natural duration of cancer. *Reports on Public Health and Medical Subjects. Ministry of Health* (33), URL http://www.cabdirect.org/abstracts/19272700028.html.

HP Inc (2017) Hp inc., form 10-k 2017. https://www.sec.gov/Archives/edgar/data/47217/000004721716000093/hp-103116x10k1.htm.

Hu K, Acimovic J, Erize F, Thomas DJ, Van Mieghem JA (2019) Forecasting new product life cycle curves: Practical approach and empirical analysis. *Manufacturing & Service Operations Management* 21(1):66–85, URL http://dx.doi.org/10.1287/msom.2017.0691.

Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. *International journal of forecasting* 22(4):679–688.

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53(282):457–481, URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452.

Karim MR, Suzuki K (2005) Analysis of warranty claim data: a literature review. *The International Journal of Quality & Reliability Management* 22(7):667–686, ISSN 0265671X, URL http://search.proquest.com.libproxy.mit.edu/docview/197620953/abstract?accountid=12492.

Kelle P, Silver EA (1989) Forecasting the returns of reusable containers. *Journal of Operations Management* 8(1):17–35, URL http://dx.doi.org/10.1016/S0272-6963(89)80003-8.

Klein JP, Moeschberger ML (2006) *Survival analysis: techniques for censored and truncated data* (Springer Science & Business Media).

Krapp M, Nebel J, Sahamie R (2013) Forecasting product returns in closed-loop supply chains. *International Journal of Physical Distribution & Logistics Management* 43(8):614–637, URL http://dx.doi.org/10.1108/IJPDLM-03-2012-0078.

Kumar A, Chinnam RB, Murat A (2017) Hazard rate models for core return modeling in auto parts remanufacturing. *International Journal of Production Economics* 183:354–361, URL http://dx.doi.org/10.1016/j.ijpe.2016.07.002.

Kumar D, Soleimani H, Kannan G (2014) Forecasting return products in an integrated forward/reverse supply chain utilizing an anfis. *International Journal of Applied Mathematics and Computer Science* 24(3):669–682, URL http://dx.doi.org/10.2478/amcs-2014-0049.

Lin D, Ying Z (1994) Semiparametric analysis of the additive risk model. *Biometrika* 81(1):61–71.

Marx-Gómez J, Rautenstrauch C, Nürnberger A, Kruse R (2002) Neuro-fuzzy approach to forecast returns of scrapped products to recycling and remanufacturing. *Knowledge-Based Systems* 15(1):119–128, URL http://dx.doi.org/10.1016/S0950-7051(01)00128-9.

McLachlan G, Krishnan T (2007) *The EM algorithm and extensions*, volume 382 (John Wiley & Sons), URL http://books.google.com/books?hl=en&lr=&id=NBawzaWoWa8C&oi=fnd&pg=PR3&dq=em+algorithm+censored&ots=tn79RN0wxQ&sig=JTJjFstak1Tc-Qc7MWpl4wjmewQ.

Petersen BJ (2013) *Reverse supply chain forecasting and decision modeling for improved inventory management*. Thesis, Massachusetts Institute of Technology, URL http://dspace.mit.edu/handle/1721.1/80988.

Rai B, Singh N (2003) Hazard rate estimation from incomplete and unclean warranty data. *Reliability Engineering & System Safety* 81(1):79–92, URL http://dx.doi.org/10.1016/S0951-8320(03)00083-8.

Samsung (2018) Samsung electronics co., ltd., 2018 consolidated financial statement. https://images.samsung.com/is/content/samsung/p5/global/ir/docs/2017_con_quarter04_all.pdf.

Taylor JM (1995) Semi-parametric estimation in failure time mixture models. *Biometrics* 899–907, URL http://www.jstor.org/stable/2532991.

Toktay LB, Wein LM, Zenios SA (2000) Inventory management of remanufacturable products. *Management Science* 46(11):1412–1426, URL http://dx.doi.org/10.1287/mnsc.46.11.1412.12082.

Tsiliyannis CA (2018) Markov chain modeling and forecasting of product returns in remanufacturing based on stock mean-age. *European Journal of Operational Research* 271(2):474–489, URL http://dx.doi.org/10.1016/j.ejor.2018.05.026.

Zhou C, Chinnam RB, Dalkiran E, Korostelev A (2017) Bayesian approach to hazard rate models for early detection of warranty and reliability problems using upstream supply chain information. *International Journal of Production Economics* 193:316–331, URL http://dx.doi.org/10.1016/j.ijpe.2017.07.020.

Zhou C, Chinnam RB, Korostelev A (2012) Hazard rate models for early detection of reliability problems using information from warranty databases and upstream supply chain. *International Journal of Production Economics* 139(1):180–195, URL http://dx.doi.org/10.1016/j.ijpe.2012.04.007.

Appendix A: Convergence of the Hazard Rate Regression Estimator

We now discuss the convergence of $\hat{\boldsymbol{h}}$ to $\boldsymbol{h}$ as the number of samples $(\boldsymbol{y},\boldsymbol{z})$ increases. For this discussion, assume that $\boldsymbol{h}=\sum_{j=1}^m w_j \boldsymbol{h}^j$ for some non-negative $\boldsymbol{w}=(w_1,\dots,w_m)$. Recall that $r_t=\sum_{k\geq t} y_k+z_k$. A bound on the tail of $\left\|\sum_{j=1}^m \hat{w}_j \boldsymbol{h}^j-\boldsymbol{h}\right\|_{p,\tau}$ is given in the next Proposition. The proposition states that as the number of observations (censored or not) for each age t increases, $\hat{\boldsymbol{h}}$ will converge to $\boldsymbol{h}$ at an exponential rate for ages that are less or equal to τ.

Proposition 1. *For a given $\boldsymbol{r}=(r_1,\dots,r_\tau)$, we have that, for any $\epsilon>0$,*

$$Pr\left(\left\|\boldsymbol{h}-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{1,\tau}\geq \epsilon|\boldsymbol{r}\right)\leq 4\sum_{t=1}^{\tau}\exp\left(-2r_t\frac{\epsilon^2}{m^2}\right).$$

Proof. If samples are drawn independently then, conditional on a given $r_t=\sum_{k\geq t} y_k+z_k$

For some $\tau\leq T$ we use the triangle inequality to obtain

$$\left\|\sum_{j=1}^m \hat{w}_j\boldsymbol{h}^j-\boldsymbol{h}\right\|_{p,\tau}\leq\left\|\sum_{j=1}^m \hat{w}_j\boldsymbol{h}^j-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}+\left\|\boldsymbol{h}-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}.$$

We will bound the first term in the right-hand side above. To simplify notation, let $\mathcal{W}=\{\boldsymbol{w}|\sum_{j=1}^m w_j h_t^j\leq 1,\boldsymbol{w}\geq 0\}$. Then,

$$\begin{aligned}\left\|\sum_{j=1}^m \hat{w}_j\boldsymbol{h}^j-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}&=\min_{w\in\mathcal{W}}\left\|\sum_{j=1}^m w_j\boldsymbol{h}^j-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}\\&\leq\min_{w\in\mathcal{W}}\left\|\sum_{j=1}^m w_j\boldsymbol{h}^j-\boldsymbol{h}\right\|_{p,\tau}+\left\|\boldsymbol{h}-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}\\&=\left\|\boldsymbol{h}-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}\end{aligned}$$

The first equality above is from the definition of $\hat{\boldsymbol{w}}$. The inequality is a result of the triangle inequality. The last equality uses the fact that $\boldsymbol{h}$ is within the cone of $(\boldsymbol{h}^1,\dots,\boldsymbol{h}^m)$ and thus $\min_{w\in\mathcal{W}}\left\|\sum_{j=1}^m w_j\boldsymbol{h}^j-\boldsymbol{h}\right\|_{p,\tau}=0$. As a result,

$$\left\|\sum_{j=1}^m \hat{w}_j\boldsymbol{h}^j-\boldsymbol{h}\right\|_{p,\tau}\leq 2\left\|\boldsymbol{h}-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{p,\tau}.$$

We now use the inequality above to obtain bounds on the tail distribution of $\hat{\boldsymbol{h}}$. We will do the derivation for $p=1$, i.e. the 1-norm. The derivation for other norms is the same, but with slightly more notation.

As discussed in Section 4, for a sample $(\boldsymbol{y},\boldsymbol{z})$ drawn according to the hazard rate distribution $\boldsymbol{h}$, the Kaplan-Meier estimator for the hazard rate of age t is given by Equation (5). If samples are drawn independently then, conditional on a given $r_t=\sum_{k\geq t} y_k+z_k$, h_t^{KM} will be a Binomial random variable with r_t trials and success probability h_t. Hence, for a given $\epsilon>0$ and r_t, we can use Hoeffding's Inequality to bound the tail of the distribution of the difference of h_t^{KM} and h_t. Namely,

$$\Pr\left(|h_t^{KM}-h_t|\geq\epsilon|r_t\right)\leq 2\exp(-2\epsilon^2 r_t).$$

Using the Union Bound and the inequality above, we have

$$\Pr\left(\left\|\boldsymbol{h}-\boldsymbol{h}^{KM}(\boldsymbol{y},\boldsymbol{z})\right\|_{1,\tau}\geq\epsilon|\boldsymbol{r}\right)\leq\sum_{t=1}^{\tau}\Pr\left(|h_t-h_t^{KM}|\geq\frac{\epsilon}{m}|\boldsymbol{r}\right)$$

$$= \sum_{t=1}^{\tau} \Pr\left(|h_t - h_t^{KM}| \geq \frac{\epsilon}{m} | r_t\right)$$
$$\leq 2 \sum_{t=1}^{\tau} \exp\left(-2 r_t \frac{\epsilon^2}{m^2}\right).$$

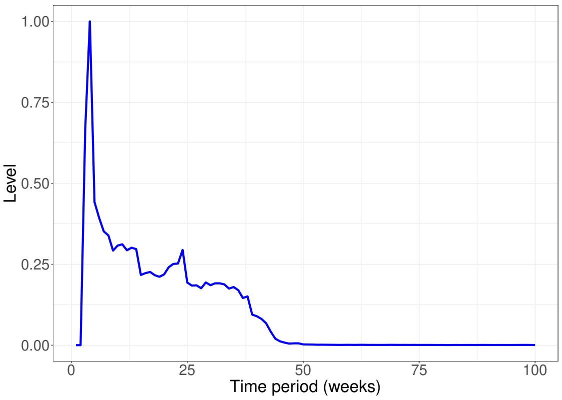
As a result,

$$\Pr\left(\left\|\boldsymbol{h} - \boldsymbol{h}^{KM}(\boldsymbol{y}, \boldsymbol{z})\right\|_{1,\tau} \geq \epsilon | \boldsymbol{r}\right) \leq 4 \sum_{t=1}^{\tau} \exp\left(-2 r_t \frac{\epsilon^2}{m^2}\right).$$

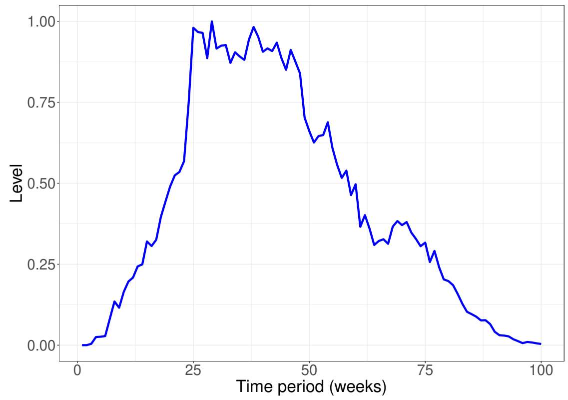
□

The proposition above, however, does not guarantee convergence of the hazard estimates for ages above the truncation period τ. Indeed, if τ is small there might be combinations of the basis that lead to a precise estimate for $\boldsymbol{h}$ for ages less than τ and potentially large errors for ages greater than τ.

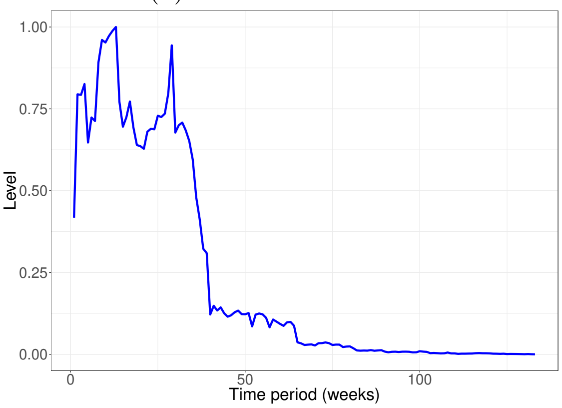
Appendix B: Sales and failures for different device models sold by WSP



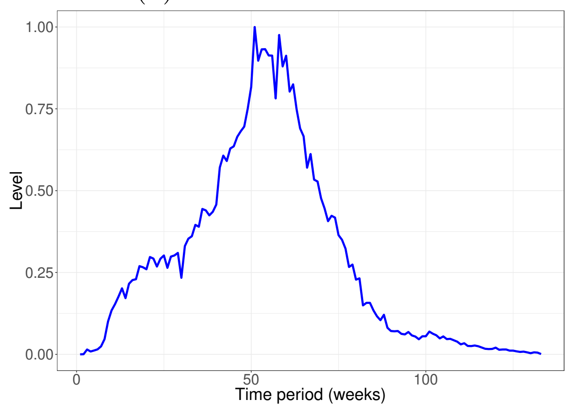
(a) Sales of model A



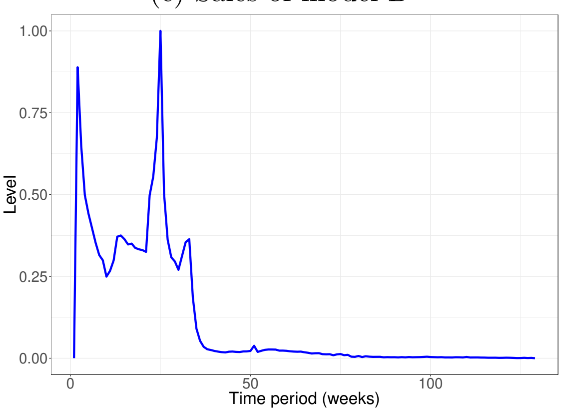
(b) Failures of model A



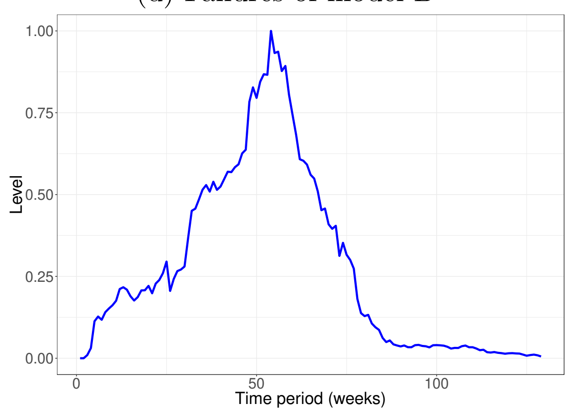
(c) Sales of model B



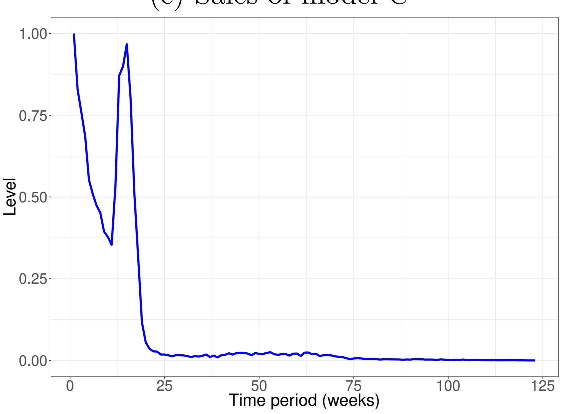
(d) Failures of model B



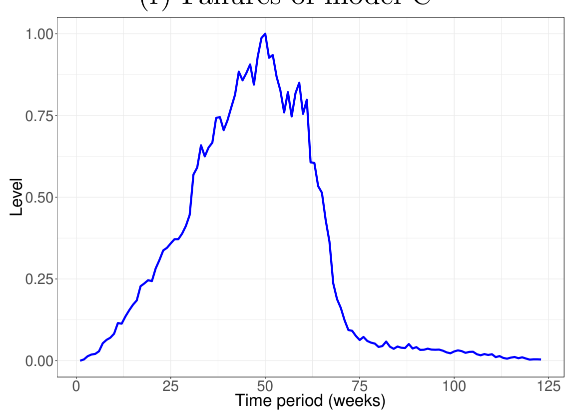
(e) Sales of model C



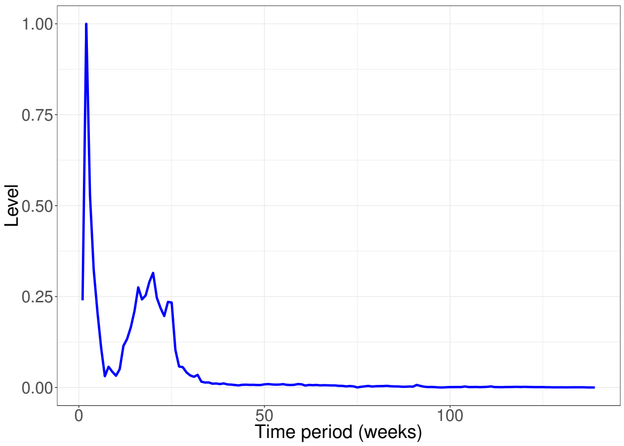
(f) Failures of model C



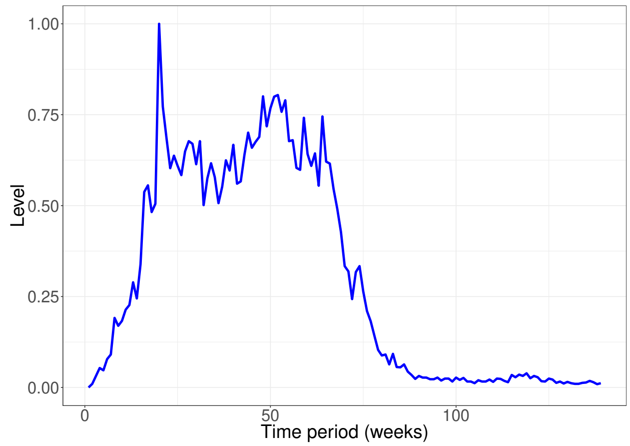
(g) Sales of model D



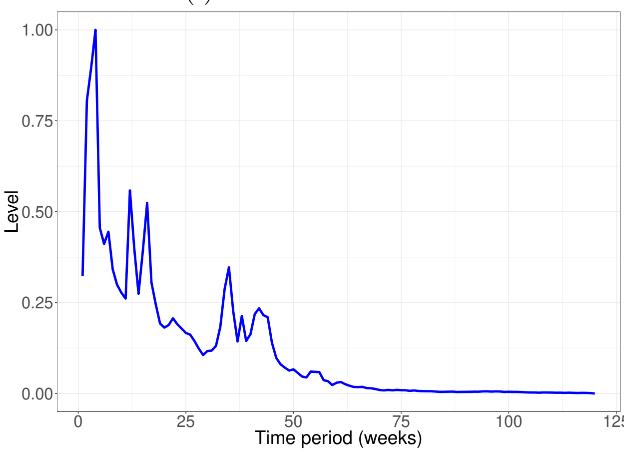
(h) Failures of model D



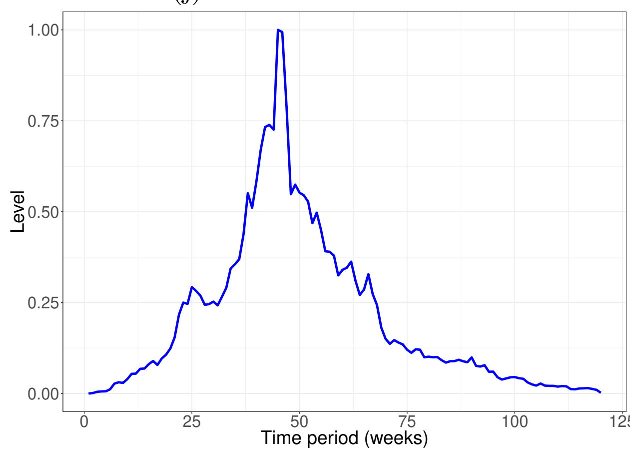
(i) Sales of model E



(j) Failures of model E



(k) Sales of model F



(l) Failures of model F

Figure 6 Sales and failures for different device models. The maximum of the curves was normalized to one to preserve data confidentiality.