IISE Transactions

Design & Manufacturing

Quality & Reliability Engineering

IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: www.tandfonline.com/journals/uiie21

On dynamically monitoring aggregate warranty claims for early detection of reliability problems

Chenglong Li, Xiaolin Wang, Lishuai Li, Min Xie & Xin Wang

To cite this article: Chenglong Li, Xiaolin Wang, Lishuai Li, Min Xie & Xin Wang (2020) On dynamically monitoring aggregate warranty claims for early detection of reliability problems, IISE Transactions, 52:5, 568-587, DOI: 10.1080/24725854.2019.1647477

To link to this article: https://doi.org/10.1080/24725854.2019.1647477

IISE TRANSACTIONS
2020, VOL. 52, NO. 5, 568–587
https://doi.org/10.1080/24725854.2019.1647477

Taylor & Francis
Taylor & Francis Group

On dynamically monitoring aggregate warranty claims for early detection of reliability problems

Chenglong Li[a,c], Xiaolin Wang[b], Lishuai Li[b,c], Min Xie[b,c], and Xin Wang[d]

[a]School of Management, Northwestern Polytechnical University, Xi'an, China; [b]Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong; [c]City University of Hong Kong Shenzhen Research Institute, Shenzhen, China; [d]Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore

ABSTRACT

Warranty databases managed by most world-leading manufacturers are constantly expanding in the big data era. An important application of warranty databases is to detect unobservable reliability problems that emerge at design and/or manufacturing stages, through modeling and analysis of warranty claims data. Usually, serious reliability problems will result in certain abnormal patterns in warranty claims, which can be captured by appropriate statistical methods. In this article, a dynamic control charting scheme is developed for early detection of reliability problems by monitoring warranty claims one period after another, over the product life cycle. Instead of specifying a constant control limit, we determine the control limits progressively by considering stochastic product sales and non-homogeneous failure processes, simultaneously. The false alarm rate at each time period is controlled at a desired level, based on which abrupt changes in field reliability, if any, will be detected in a timely manner. Furthermore, a maximum-likelihood-based post-signal diagnosis scheme is presented to aid in identifying the most probable time of problem occurrence (i.e., change point). It is shown, through in-depth simulation studies and a real case study, that the proposed scheme is able to detect an underlying reliability problem promptly and meanwhile estimate the change point with an acceptable accuracy. Finally, a moving window approach concerning only recent production periods is introduced to extend the original model so as to mitigate the "*inertia*" problem.

ARTICLE HISTORY
Received 24 June 2018
Accepted 18 July 2019

KEYWORDS
Warranty; reliability; product life cycle; statistical process monitoring; diagnostics

1. Introduction

The occurrence of quality and reliability problems for products in the field often incurs huge economic losses and reputational damages to manufacturers. The Toyota recall crisis (due to defective accelerator pedals) in 2010 and the Samsung Galaxy Note 7 battery explosion problem in 2016 are two extreme examples. Generally speaking, product reliability problems are often caused by one or some of the following issues: faults in product design, unanticipated failure modes, use of unreliable raw materials from external suppliers, harsher-than-expected operating condition, and improperly verified design changes (Wu and Meeker, 2002). Since many reliability problems arise at the design and manufacturing stages, lots of effort has been devoted, based on the modern philosophies of quality and reliability management, to building reliability into a product and conducting upfront reliability testing, so as to avoid serious reliability and warranty problems for the sold product in the field.

Unfortunately, despite these efforts, defective products inevitably enter the market (some of them even result in serious recall issues), since some emerging reliability problems in the product cannot be directly or readily observed. Consequently, many manufacturers have to regularly monitor their sold product units in the field to obtain more information for early detection of quality and reliability problems (Yashchin, 2012). One way such an early detection is attained could be through monitoring warranty claims made on the units in the field (Meeker and Hong, 2014; Hong *et al.*, 2018). For instance, it is not uncommon for automakers to recall a proportion of sold vehicles when the number of warranty claims regarding a specific component or failure mode rises to a certain level. In March 2017, Ford announced a new recall of 211 000 vehicles after receiving nearly 11 000 warranty claims related to door-latch failures on relevant models (Rosevear, 2017). Likewise, Harley-Davidson recalled over 27 000 motorcycles from its 2016 model after discovering the clutch may not disengage in certain conditions. The company recognized this reliability problem after observing a spike in relevant warranty claims (Thomas, 2016).

Nevertheless, due to the lack of systematic decision support on early detection of reliability problems, engineering and management decisions, such as product recall, design modification, and reliability improvement, made by manufacturers are often suboptimal—subjective, slow and even reactive. A major challenge faced by manufacturers is how to isolate and detect changes in field reliability when the concerned

CONTACT Xiaolin Wang xlwang28-c@my.cityu.edu.hk
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiie.

warranted base (i.e., the population of sold units that are still under warranty) varies over time. Traditional monitoring metrics, such as average warranty return rate, can hardly address such challenges. Therefore, developing effective early detection tools based on warranty claims data is of practical significance, as detecting a serious reliability problem months or even weeks earlier can reduce substantial economic losses and reputational damages for the manufacturers. However, until now this topic has only received limited attention in the literature.

In general, unobserved reliability problems can be traced by analyzing warranty claims, as they may be reflected by certain abnormal patterns in warranty claims data (Wu and Meeker, 2002; Lawless *et al.*, 2012). Statistical Process Monitoring (SPM), a prevailing tool for monitoring the stability of a process and detecting possible process changes (Montgomery, 2013), can be used to identify such abnormal patterns. This motivates us to develop appropriate SPM methods and investigate the detection and diagnosis of reliability problems through monitoring warranty claims over time.

In this work, we develop an integrated monitoring and diagnosis framework to track aggregate warranty claims to see if the claim rate changes along the reference line and, upon an alarm signal, to further estimate the most probable time of problem occurrence. More specifically, an online forecasting model of incremental (aggregate) warranty claims within each single period is first developed, in which features of the time-varying population over the product life cycle are incorporated. This model dynamically forecasts the incremental warranty claims one step ahead, since the production-sale process evolves over time. On this basis, a simple, but powerful, monitoring statistic is constructed. Using this statistic, a Shewhart-type control chart with dynamic probability control limits, which are progressively determined with a pre-set false alarm rate at each time period, is designed. In real applications, it is essential to know when the reliability problem actually emerged if a (true) signal is triggered. To this end, we further propose a maximum-likelihood-based post-signal diagnosis scheme to find out the most likely change point. In-depth simulation studies and a real case study are presented to demonstrate the effectiveness of the integrated monitoring and diagnosis framework.

The remainder of this article is organized as follows. Section 2 reviews the related literature. Section 3 describes the formulation of the dynamic monitoring and post-signal diagnosis scheme. Section 4 conducts in-depth simulation and comparison studies to assess its performance. In Section 5, an extension of the basic model is presented to deal with the "inertia" problem. After that, Section 6 demonstrates the integrated scheme by a real-life case study. Finally, Section 7 summarizes the contents and main conclusions of this article.

2. Literature review

As this article deals with early detection of reliability problems through the monitoring of aggregate warranty claims using SPM methods, there are two distinct streams of research related to this work, namely, warranty claims estimation/prediction and process monitoring with SPM methods.

The first stream of research focuses on estimating and/or predicting warranty claims for products in the field. Generally speaking, a warranty contract defines the coverage (components and services that are covered or not), the extent (length of period) and, in the event of item failures, the type of compensation (such as free repair/replacement and refund) offered by the seller or manufacturer (Shang *et al.*, 2018; Wang and Xie, 2018). In the warranty-oriented literature, warranty claims estimation and prediction have long been studied for various purposes, e.g., warranty cost estimation (Liu *et al.*, 2015, Zhao *et al.*, 2018; Wang, He, He, Li and Xie, 2019), maintenance planning (Chien 2019; Wang, Li and Xie, 2019), and spare parts management (Van der Heijden and Iskandar, 2013; Jin *et al.*, 2017). However, most existing studies on this topic focus on estimating or predicting the expected number of warranty claims per unit sold (i.e., for a single product unit). Recently, several works developed a new methodology to forecast aggregate warranty claims/costs by coupling a stochastic sales process with a product failure process (Xie and Ye, 2016; Wang *et al.*, 2017; Xie *et al.*, 2017), which provides the foundation for formulating the statistic in our monitoring scheme.

The dynamic warranty claims forecasting model developed in our work differs from those in the three papers above in the sense that they assume that the product sales process is preknown to the manufacturer based on market survey or previous sales data, whereas we consider a more realistic scenario in which actual sales data is progressively collected one period after another. In other words, our one-step-ahead warranty claims forecasting model can be implemented without making any unreliable predictions regarding future sales volumes, which improves prediction accuracy and thus facilitates subsequent monitoring of warranty claims.

In the warranty claims monitoring problem of interest, on the one hand, the number of warranty claims for a single unit in any time interval can usually be modeled by a Poisson distribution (Wu, 2012); on the other hand, the warranted base (or the *monitored population* in the SPM context), varies over the product life cycle, due to the dynamic nature of the sales of new units and the dropout of out-of-warranty units. As the warranty claims of all sold units are recorded in the warranty database, this is similar to performing "100% inspection". In this case, the sample is exactly the whole monitored population. As such, monitoring Poisson processes with time-varying sample (population) sizes is the second research area related to our work. This topic has been actively investigated in the SPM-related literature; however, existing research has focused predominately on industrial quality control and public health surveillance (Richards *et al.*, 2015). In this particular area, a number of works enhanced the ability of the classic Shewhart u-chart in detecting small shifts with the CUMulative SUM (CUSUM) and Exponentially Weighted Moving Average (EWMA) procedures; see, e.g., Dong *et al.* (2008), Ryan and Woodall (2010), Zhou, Zou, Wang and Jiang (2012), Shu *et al.* (2014), and the references therein.

It is important to note that all the literature (on varying sample sizes) above was built on a critical assumption that the distribution of sample sizes is known *a priori* before the control chart is initiated. In reality, however, knowledge about the time-varying sample sizes is rarely available in advance. Zhou, Zou, Wang and Jiang (2012) pointed out that traditional control limit design methods based on perfect knowledge of sample sizes are rather sensitive to the model of the sample sizes. As a result, the performance of such charts becomes significantly degraded if the actual sample size distribution deviates from the assumed one. To overcome this problem, Shen *et al.* (2013) applied the probability control limit approach in a dynamic manner, so as to monitor Poisson count data with time-varying sample sizes. Their control limits were determined so as to control the conditional false alarm rate at each step, with the latest sample size information. The dynamic design approach in Shen *et al.* (2013) no longer requires one to forecast future sample sizes, thus it can adapt to more complex situations than its constant counterpart. This idea was later adopted by Zhang and Woodall (2015), Huang *et al.* (2016), Shen *et al.* (2016), Yang *et al.* (2017), and Sogandi *et al.* (2019), among others.

Back to the warranty claims monitoring problem, things become much more complicated if the monitored population exhibits multiple time-varying features, not merely a varying size. Unlike traditional SPM problems, here the population of interest is highly heterogeneous. It is a mixture of units produced and sold in the past periods, that are still under warranty. The proportions of the units of different ages (time in service) and of different batches (production periods) in this population also vary over time. Such multiple time-varying features distinguish our problem from others and also make the modeling and analysis rather difficult.

To our knowledge, limited work can be found on monitoring warranty claims. Wu and Meeker (2002) made an early attempt to investigate this problem. They stratified product units by manufacturing periods, and then designed sequential hypothesis tests for claim counts at different product ages to detect arbitrary increases in the reference claim rate by controlling the overall false alarm probability at the desired value. To apply their method for detection purpose, a number of sequential test charts need to be operated simultaneously. Lawless *et al.* (2012) proposed the monitoring of the number of warranty claims recorded in each time period regardless of the production periods and product ages. They adopted CUSUM procedures with a constant signal limit, and used Markov chain calculations to determine monitoring plans with given false signal probabilities. Zhou, Chinnam and Korsteler (2012) adopted proportional hazard models to combine upstream supply chain quality/testing information (as explanatory covariates) with warranty claims data and, following Wu and Meeker (2002), developed a statistical method with superior performance for early detection of reliability problems. Zhou *et al.* (2017) further developed a Bayesian approach to proportional hazard models in order to reduce the need for the extensive warranty claim history in Zhou, Chinnam and Korsteler (2012). Recently, Shang *et al.* (2017) developed a nonparametric regression-based phase I scheme to detect the change point for profiles with binary data (e.g., warranty data), by integrating the change-point algorithm with a generalized likelihood ratio.

Compared with the existing literature, especially the two closest works (Wu and Meeker, 2002; Lawless *et al.*, 2012), the contributions of our work lie in the following three aspects:

1. They determine control limits before the control charts are activated, relying entirely on historical sales records (which are often unreliable and insufficient for accurately forecasting future sales volumes and would induce additional variability that degrades the detection ability of their control charts); whereas we consider a *dynamic design* of the control chart with probability control limits that are determined one step ahead, by including the latest sales information. In this manner, our proposed scheme is more suitable to "real-time" monitoring of warranty claims data.
2. The second contribution of our work is the *post-signal diagnosis* scheme, which is used to estimate the most probable change point. Diagnostics after a signal is indispensable for helping to identify the root cause(s) behind an abnormality, however, has received little attention in warranty claims monitoring problems.
3. Finally, the proposed monitoring and diagnosis scheme can be applied to the entire *life cycle* of a product, in which each stage has its unique characteristic. This is quite important since many reliability problems, some resulting in product recalls, will be apparent only after a relatively long time when a product is released, and they sometimes emerge later in the life cycle (Lawless *et al.*, 2012).

3. Problem formulation

3.1. Problem description

In this work, early detection of reliability problems is attained through dynamically monitoring aggregate warranty claims over time. The focus is on the detection of an abrupt change (increase) in the warranty claim rate of an unknown magnitude at an unknown time point τ. Let $\lambda_0(a)$, $a>0$, be the reference claim rate of a product that is designed or expected by the manufacturer and is assumed known *a priori*, and $\lambda(a|i)$, $a>0$, be the field claim rate for units manufactured in period i. Here, both $\lambda_0(a)$ and $\lambda(a|i)$ are functions of unit age a.

The early detection of reliability problems can then be formulated as a test of the hypothesis:

$$H_0 : \lambda(a|i) \le \lambda_0(a), a > 0, \tag{1}$$

versus

$$H_1 : \lambda(a|i) > \lambda_0(a), a > 0, \tag{2}$$

from a certain point of time onwards, i.e., $i > \tau$.

A parametric form of $\lambda_0(a)$, $a>0$, is adopted to provide the structure for the monitoring scheme. In practice, the parameters of $\lambda_0(a)$ can be estimated through the Maximum Likelihood Estimation (MLE) method, if historical warranty data of previous product generations or reliability test data of the current product are available (Wu, 2012, 2013; Chen and Ye, 2017). Otherwise, expert judgements, analytic hierarchy

Table 1. Illustration of the production-sale process.

Prod. (i) \ *Sales (j)*		*1*	*2*	...	*j*($j \geq i$)	...	*l*
1	V_1	$D_{1,1}$	$D_{1,2}$	...	$D_{1,j}$	...	$D_{1,l}$
2	V_2		$D_{2,2}$	...	$D_{2,j}$	...	$D_{2,l}$
3	V_3			...	$D_{3,j}$	...	$D_{3,l}$
...	...			...	...	...	...
i	V_i			...	$D_{i,j}$	...	$D_{i,l}$
...	...			...	...	...	...
p	V_p					...	$D_{p,l}$
	$\sum_{i=1}^{p} V_i$	$S_1 = \sum_{i=1}^{1} D_{i,1}$	$S_2 = \sum_{i=1}^{2} D_{i,2}$	...	$S_j = \sum_{i=1}^{j\wedge p} D_{i,j}$	...	$S_l = \sum_{i=1}^{l\wedge p} D_{i,l}$

process, or even customer requirement/competitor analysis are potential alternatives for specifying the model parameters.

To test the hypotheses above, sales data and warranty claims data of the product should be periodically collected and analyzed. In current practice, collected warranty data and sales data are often grouped. In such a discrete setting, a discrete formulation is appropriate and can be applied directly to the problem. The duration of a period could be arbitrary, e.g., ranging from a day to a quarter, which can be determined by balancing administration difficulty, data availability, and monitoring accuracy, among other factors.

In this study, a non-renewable free repair warranty policy of length w is assumed to be attached to each sold unit. Under this policy, for a unit sold at period j, an eligible failure at period $k \in (j, j+w]$ will be instantly rectified by the manufacturer at no cost to the customer. Then, the remaining warranty period becomes $w-(k-j)$.

Furthermore, in most practical cases (especially the make-to-stock case), product units may not be sold immediately after manufacturing. It is not uncommon that a *sales delay* between the production and sales can be observed (Wu, 2013; Ye and Ng, 2014). To capture this characteristic, a production-sale matrix is introduced, as shown in Table 1. Suppose that the manufacturer produces V_i units in the ith, $i = 1, 2, \ldots, p$, period, and these units are randomly sold in the current and subsequent periods. Let $D_{i,j}$ denote the number of units produced in period i and sold in period j, $j = i, i+1, \ldots, l$. In reality, the production process usually ceases earlier before the end of the sales, i.e., $p \leq l$. Thus, the total number of units sold in the jth period can be expressed as $S_j = \sum_{i=1}^{j\wedge p} D_{i,j}$, where $j\wedge p = \min\{j, p\}$. It is worth mentioning that $D_{i,j}$ might be zero for certain periods, especially when j is much larger than i or the discrete time interval is too short.

The product sales will terminate at the lth period (the first sales occurs at period one), beyond which the product will not be available in the market. In other words, the lth period is the end of the *product life cycle*. Notice that warranty claims would still be generated within the interval $(l, l+w]$, because the warranty of units sold exactly at the lth period will terminate at period $l+w$. In this context, our monitoring scheme can be extended beyond the product life cycle.

At the beginning of period k, the size of the monitored population (i.e., the warranted base) can be calculated by

$$\Omega_k = \sum_{j=1\vee(k-w)}^{(k-1)\wedge l} S_j = \sum_{j=1\vee(k-w)}^{(k-1)\wedge l} \sum_{i=1}^{j\wedge p} D_{i,j}, k = 2, 3, \ldots, \quad (3)$$

where $1\vee(k-w) = \max\{1, k-w\}$, and $\Omega_1 = 0$.

The expression of Ω_k confirms our statement before, namely, the warranted base varies over time; see Figure 1 for an illustration. It shows that: (i) Ω_k first increases, then fluctuates, and finally decreases as k increases; and (ii) the proportions of the units in different age groups also vary with k in a similar way. At the very beginning, the ages of most sold units are young (say, less than $w/4$), whereas at the very end, the ages of most units are near w. As mentioned earlier, these time-varying features make the warranty claims monitoring problem more complicated than others, which poses difficulty in the design of a monitoring scheme.

3.2. Statistic for monitoring warranty claims

For the uth unit produced in period i and sold in period j ($j \geq i$), let $X_{i,j,k}^{(u)}$ be a random variable denoting its cumulative number of warranty claims up to period k, where $k = j+1, j+2, \ldots$, and $u = 1, 2, \ldots, D_{i,j}$. In particular, in the period that the uth unit is sold (i.e., $k = j$), we consider relevant warranty claims do not appear till the next period, i.e., $X_{i,j,j}^{(u)} = 0$. In this work, a failed unit is assumed to be minimally rectified upon an eligible claim. In this case, $X_{i,j,k}^{(u)}$ can be modeled by a Non-Homogeneous Poisson Process (NHPP) with a claim rate $\lambda(k-j)$. Thus, we have:

$$E[X_{i,j,k}^{(u)}] = \int_0^{(k-j)\wedge w} \lambda(a|i)\mathrm{d}a. \quad (4)$$

For a specific k, let M_k denote the aggregate warranty claims up to this period ($M_1 = 0$). M_k includes eligible warranty claims generated by all units sold before, and including, period $k-1$. As such, M_k can be presented in a more detailed way; see Proposition 1 below.

Proposition 1. *Assuming that the sales volume $D_{i,j}$ in the jth period is available one after another and the cumulative number of warranty claims for the uth unit manufactured in period i and sold in period j up to period k is $X_{i,j,k}^{(u)}$, then the aggregate warranty claims M_k up to period k can be elaborated as*

$$M_k = \begin{cases} \sum_{j=1}^{k-1}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)}, & 1 \le k \le w+1, \\ \sum_{j=k-w}^{k-1}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)} + \sum_{j=1}^{k-w-1}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)}, & w+1<k \le l+1, \\ \sum_{j=k-w}^{l}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)} + \sum_{j=1}^{k-w-1}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)}, & l+1<k \le l+w, \\ \sum_{j=1}^{l}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)}, & k > l+w. \end{cases} \tag{5}$$

Proof. The proof is presented in Appendix A. □

Notice that $X_{i,j,k}^{(u)}, u = 1, 2, \ldots, D_{i,j}$, are independently distributed Poisson random variables, and M_k can be regarded as their summation. It is thus straightforward to know that the aggregate warranty claims M_k in Equation (5) also follows a Poisson distribution, with its mean $E[M_k]$ and variance $Var[M_k]$ being identical. Based on Equations (4) and (5), the mean (also variance) of M_k, i.e., $E[M_k]$, can be derived as

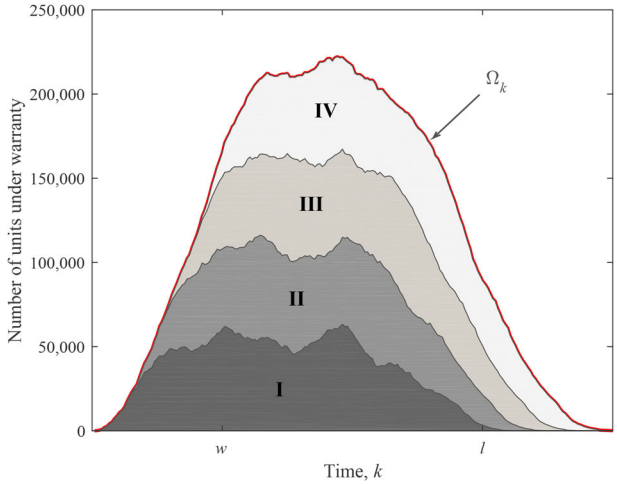

Figure 1. Illustration of warranted base over the product life cycle. The red curve represents the warranted base, Ω_k, and zones I, II, III, and IV show the number of units with age [0, $w/4$], ($w/4, w/2$], ($w/2, 3w/4$], and ($3w/4$, w], respectively. The production and sales process in this illustration follows from the scenario 1 reported in Section 4.1.1.

$$E[M_k] = \begin{cases} \sum_{j=1}^{k-1}\sum_{i=1}^{j\wedge p} D_{i,j}\int_0^{k-j}\lambda(a|i)\mathrm{d}a, & 1 \le k \le w+1, \\ \sum_{j=k-w}^{k-1}\sum_{i=1}^{j\wedge p} D_{i,j}\int_0^{k-j}\lambda(a|i)\mathrm{d}a + \sum_{j=1}^{k-w-1}\sum_{i=1}^{j\wedge p} D_{i,j}\int_0^{w}\lambda(a|i)\mathrm{d}a, & w+1<k \le l+1, \\ \sum_{j=k-w}^{l}\sum_{i=1}^{j\wedge p} D_{i,j}\int_0^{k-j}\lambda(a|i)\mathrm{d}a + \sum_{j=1}^{k-w-1}\sum_{i=1}^{j\wedge p} D_{i,j}\int_0^{w}\lambda(a|i)\mathrm{d}a, & l+1<k \le l+w, \\ \sum_{j=1}^{l}\sum_{i=1}^{j\wedge p} D_{i,j}\int_0^{w}\lambda(a|i)\mathrm{d}a, & k > l+w. \end{cases} \tag{6}$$

By the definition of M_k, the incremental (aggregate) warranty claims ΔM_k generated in the kth period is simply the difference of M_k at the two end-points, i.e., $\Delta M_k = M_k - M_{k-1}$, for $k = 2, 3, \ldots$, with $\Delta M_1 = 0$. Unlike the aggregate warranty claims, the incremental warranty claims in each period are independent of each other, due to the independent increment property of a Poisson process. The mathematical expression of ΔM_k in the kth period is given by

$$\Delta M_k = \begin{cases} \sum_{j=1}^{k-1}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}}\left(X_{i,j,k}^{(u)} - X_{i,j,k-1}^{(u)}\right), & 1 \le k \le w+1, \\ \sum_{j=k-w}^{k-1}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}}\left(X_{i,j,k}^{(u)} - X_{i,j,k-1}^{(u)}\right), & w+1<k \le l+1, \\ \sum_{j=k-w}^{l}\sum_{i=1}^{j\wedge p}\sum_{u=1}^{D_{i,j}}\left(X_{i,j,k}^{(u)} - X_{i,j,k-1}^{(u)}\right), & l+1<k \le l+w, \\ 0, & k > l+w. \end{cases} \tag{7}$$

Likewise, it is clear that ΔM_k also follows a Poisson distribution, and its mean (and variance) can be expressed as

$$E[\Delta M_k] = \begin{cases} \sum_{j=1}^{k-1}\sum_{i=1}^{j\wedge p} D_{i,j}\int_{k-j-1}^{k-j}\lambda(a|i)\mathrm{d}a, & 1 \le k \le w+1, \\ \sum_{j=k-w}^{k-1}\sum_{i=1}^{j\wedge p} D_{i,j}\int_{k-j-1}^{k-j}\lambda(a|i)\mathrm{d}a, & w+1<k \le l+1, \\ \sum_{j=k-w}^{l}\sum_{i=1}^{j\wedge p} D_{i,j}\int_{k-j-1}^{k-j}\lambda(a|i)\mathrm{d}a, & l+1<k \le l+w, \\ 0, & k > l+w. \end{cases} \tag{8}$$

One thing noteworthy is that Equations (5)–(8) can be expressed in a more concise manner. For example, $E[\Delta M_k]$ in Equation (8) is mathematically equivalent to $\sum_{i=1}^{(k-1)\wedge p}\sum_{j=i\vee(k-w)}^{(k-1)\wedge l} D_{i,j}\int_{k-j-1}^{k-j}\lambda(a|i)\mathrm{d}a$, $k = 1, 2, \ldots$. However, such concise expressions are less informative and less helpful for us to understand the differences in the

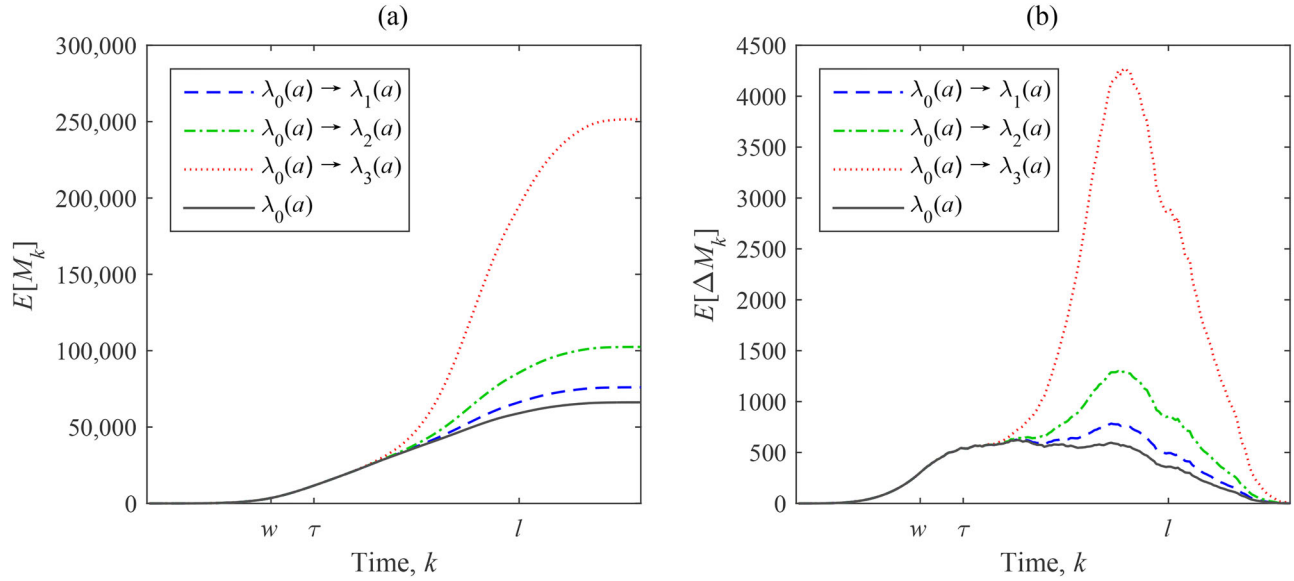

Figure 2. Illustration of the "*inertia*" problem. Four cases are demonstrated in each subfigure: the field claim rate remains at $\lambda_0(a)$ over the life cycle or is shifted from $\lambda_0(a)$ to $\lambda_1(a), \lambda_2(a)$, or $\lambda_3(a)$ from τ onwards, where $\lambda_3(a) > \lambda_2(a) > \lambda_1(a) > \lambda_0(a)$. The production and sales process in this illustration follows from the scenario 1 reported in Section 4.1.1.

behaviors of warranty claims at different stages of the product life cycle.

To detect an undesired change in $\lambda(a)$ as timely as possible, here it is suggested to adopt ΔM_k as the statistic, as M_k may suffer from a serious "inertia" problem, as illustrated in Figure 2. As can be observed, after a simulated change point τ of the claim rate, the curves $E[\Delta M_k]$ exhibit abnormal trends much earlier than $E[M_k]$, no matter if the change magnitude is slight, moderate, or significant. This inertia problem arises when the current aggregate warranty claims contain a mass of (historical) claims of those units that were produced a long time ago (Lawless *et al.*, 2012).

3.3. Dynamic monitoring scheme

In this study, ΔM_k plays a crucial role in the early detection of reliability problems during the product life cycle. To facilitate the design of control chart, we normalize ΔM_k as

$$Z_k = \frac{\Delta M_k - E[\Delta M_k]}{\sqrt{Var[\Delta M_k]}}, \tag{9}$$

where $E[\Delta M_k]$ and $Var[\Delta M_k]$ are, respectively, the mean and variance of ΔM_k under the null hypothesis.

Recall that ΔM_k are independent Poisson random variables with mean $E[\Delta M_k]$ and variance $Var[\Delta M_k]$ $(= E[\Delta M_k])$. The Central Limit Theorem suggests that Z_k follows an approximate standard normal distribution for sufficiently large values of $E[\Delta M_k]$ (say, $E[\Delta M_k] > 1000$). In reality, this condition could be satisfied in a certain stage of the product life cycle when the population of sold units is large enough to generate a fair amount of warranty claims.

The monitoring scheme is a Shewhart-type control chart based on the statistic Z_k. In particular, an upper one-sided control chart is adopted, as manufacturers are more concerned about an undesired increase in the claim rate. Due to the independence of the tests for different time periods, the upper control limit of the kth period, UCL_k, can be obtained by solving:

$$P(Z_k > UCL_k) = \alpha_k, \tag{10}$$

where α_k is the maximum desired false signal/alarm rate in period k.

That is to say, the control limits are specified in such a way that the signal probability in each period should never exceed its corresponding α_k when the claim rate has not shifted (increased). If manufacturers have no prior knowledge on the probability that a reliability problem will occur at different periods, then it is a common practice to simply control the false signal rate at each time period to be constant (i.e., $\alpha_k = \alpha$, for $k = 2, 3, \ldots$). Even so, the UCL_k are not invariant, since the Poisson population of interest changes over time, as mentioned earlier. This requires us to specify different signal limits for different periods. Since the sales volumes of future periods are not available in advance, we cannot obtain the whole set of signal limits for all periods at once. A dynamic monitoring scheme is thus needed, in which the control chart can be designed and implemented progressively; see the left part of Figure 3.

To be specific, as early as the end of the first period, the sales volume S_1 becomes available and thus we can design the upper control limit UCL_2 for the second period by using Equation (10). The monitoring procedure actually starts from the second period, as warranty claims (if any) do not start to appear until the second period. At the end of the second period, the statistic Z_2 is calculated with the incremental warranty claims ΔM_2 during this period, and its value is plotted against UCL_2, on the control chart. In a similar manner, UCL_3 and subsequent control limits can be derived one period after another, taking into account the latest sales information. By applying such dynamic control limits, one can obtain an in-control run length distribution that is approximately a geometric distribution for any production-sale process.

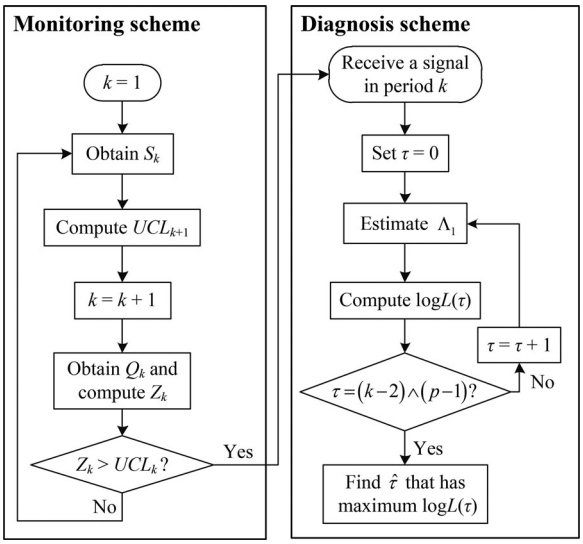

Figure 3. Integrated framework of the dynamic monitoring and post-signal diagnosis scheme.

For early detection purposes, the monitoring statistic Z_k is examined repeatedly in each period. A signal is triggered in the kth period if $Z_k > UCL_k$, dropping a hint that a certain reliability problem may have occurred. Then, the manufacturer must investigate possible causes of this problem. If the assignable cause(s) can be confirmed, the manufacturer should take actions to rectify it immediately. Otherwise, this signal will be regarded as a false alarm.

3.4. Post-signal diagnosis scheme

After receiving an out-of-control signal from the control chart, it is essential to know when the reliability problem emerged. This can help identify the root cause(s) if it is a true alarm. In this work, the post-signal diagnosis is formulated as a change-point estimation problem.

Suppose that an out-of-control signal is triggered in the gth period. Then, possible locations of the actual change point τ may range from the very beginning ($\tau=0$) to the time just before the most recent production period. Usually, the most recent production period that could lead to the signaling period g is the previous period $g-1$, thus, the reliability problem is believed to occur at least before period $g-1$ (i.e., $\tau \le g-2$). However, on the other hand, sometimes the signal might even occur later than the end of the production (i.e., period p). In this case, the problem occurs at least before the last production period p, i.e., $\tau \le p-1$. Therefore, depending on when the signal is released, the set of possible locations of the change point is $\{0, 1, \ldots, (g-2)\wedge(p-1)\}$. In this study, a maximum-likelihood-based diagnosis scheme is developed to estimate the unknown change-point by making full use of the production data, sales data, and warranty claims data up to period g. In other words, among these candidates of the change point location, the one with the maximum likelihood is considered the most probable one; see the right part of Figure 3.

Basically, the actual change-point τ splits the production process into two stages. The claim rate of the units manufactured before, and including, period τ is in control, but that of the units produced after period τ (and up to period $(g-1)\wedge p$) has an increase from $\lambda_0(a)$ to $\lambda_1(a)$. It should be noted that the change point τ only partitions the production process, whereas those units manufactured before τ may still be sold after τ and generate eligible warranty claims (consequently mixed with those claims from the units produced after τ). This feature makes our diagnosis problem different from classic change-point estimation problems (Amiri and Allahyari, 2012) and also poses difficulty in modeling and analysis.

Here, the likelihood function can be formulated as

$$L(\tau|Q_k, k=2,3,\ldots,g) = \prod_{k=2}^{\tau+1} \frac{\exp\{-E[\Delta M_k]\}E[\Delta M_k]^{Q_k}}{Q_k!} \times \prod_{k=\tau+2}^{g} \frac{\exp\{-E[\Delta M'_k]\}E[\Delta M'_k]^{Q_k}}{Q_k!}, \tag{11}$$

where Q_k is the observed number of warranty claims within the kth period, and $E[\Delta M'_k]$ represents the expected number of warranty claims recorded in the kth, $\tau+2 \le k \le g$, period generated from the mixture of the sold units manufactured before and after τ, which is

$$E[\Delta M'_k] = \sum_{i=1}^{\tau} \sum_{j=i\vee(k-w)}^{(k-1)\wedge l} D_{i,j} \int_{k-j-1}^{k-j} \lambda_0(a)\mathrm{d}a + \sum_{i=\tau+1}^{(k-1)\wedge p} \sum_{j=i\vee(k-w)}^{(k-1)\wedge l} D_{i,j} \int_{k-j-1}^{k-j} \lambda_1(a)\mathrm{d}a. \tag{12}$$

More often than not, $\lambda_1(a)$ is unknown and needs to be estimated from the field data. The Nelson–Aalen estimates can be used to approximate the expected number of warranty claims at different ages for a product (i.e., $\Lambda_1(a) = \int_{a-1}^{a} \lambda_1(t)\mathrm{d}t$), based on the claims data of the units manufactured after the change point. The Nelson–Aalen estimator is a well-known non-parametric estimator of the cumulative claim rate function and has been widely used in reliability engineering (Cook and Lawless, 2007). In the current context, the Nelson–Aalen estimate at age a can be formulated as

$$\widehat{\Lambda}_1(a) = \frac{\sum_{i=\tau+1}^{(g-a)\wedge p} \sum_{j=i}^{(g-a)\wedge l} N_{i,j}(a)}{\sum_{i=\tau+1}^{(g-a)\wedge p} \sum_{j=i}^{(g-a)\wedge l} D_{i,j}}, \tag{13}$$

where $N_{i,j}(a)$ is the observed number of warranty claims at age a, $1 \le a \le (g-j)\wedge w$, for units manufactured in period i and sold in period j. A parametric alternative to estimating the parameters of $\lambda_1(a)$ is the MLE method, which is presented in Appendix B for reference.

As such, $\int_{k-j-1}^{k-j} \lambda_1(a)\mathrm{d}a$ in Equation (12) can be estimated by $\widehat{\Lambda}_1(k-j)$, for $1 \le k-j \le w$. Generally, we shall have $Q_k = \sum_{i=1}^{(k-1)\wedge p} \sum_{j=i\vee(k-w)}^{(k-1)\wedge l} N_{i,j}(k-j)$, for $k = 2, 3, \ldots, l+w$. Substituting Equation (13) into Equation (12) and then Equation (12) into Equation (11), the estimated value of τ can be obtained by maximizing the log of the likelihood function (11), as follows:

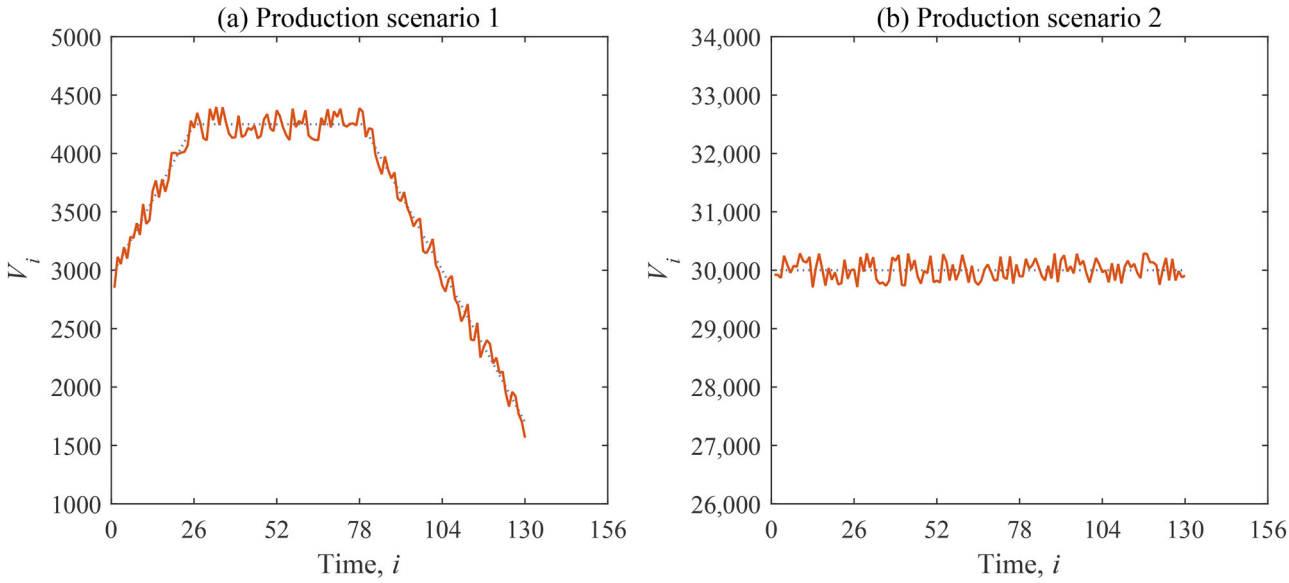

Figure 4. Two types of representative production scenarios of interest.

$$\hat{\tau} = \arg \max_{0 \le \tau \le (g-2) \wedge (p-1)} \log L(\tau | Q_k, k = 2, 3, ..., g).$$

It is quite difficult, if not impossible, to obtain the closed form of $\hat{\tau}$. Fortunately, we can easily obtain $\hat{\tau}$ through numerical search, as τ is an integer. In summary, the integrated framework of the dynamic monitoring and diagnosis scheme is illustrated in Figure 3.

4. Performance assessment

In this section, we assess the detection performance and diagnosis accuracy of the proposed scheme through in-depth simulation and comparison studies.

4.1. Simulation study

4.1.1. Simulation setting

We consider that a manufacturer produces and sells a type of product over time. All product units are sold with a free minimal repair warranty of length $w = 52$ weeks. The manufacturer periodically collects production data, sales data, and warranty claims data, and dynamically monitors warranty claims to detect potential reliability problems that might emerge at the design and manufacturing stages.

To start the simulation, the first task is to generate the production-sale matrix as in Table 1. To this end, two representative production scenarios are considered, i.e., scenario 1:

$$V_i = \begin{cases} 3000 + 50 \times (i-1) + v_{1i}, & 1 \le i \le 26, \\ 4250 + v_{1i}, & 27 < i \le 78, \\ 4250 - 50 \times (i-79) + v_{1i}, & 79 < i \le 130, \end{cases}$$

and scenario 2:

$$V_i = 30\,000 + v_{2i}, 1 \le i \le 130,$$

where $v_{1i} \sim U(-150, 150)$ and $v_{2i} \sim U(-300, 300)$ are uniform random integers representing the fluctuation of production volumes in each week.

The two production scenarios above are shown in Figure 4. In the first scenario, the production volumes first increase, then level off, and finally decrease; whereas the second scenario characterizes a relatively stable production mode with small fluctuations and large production volumes. Moreover, we assume that selling all units produced in a specific period often takes 10 to 30 weeks and the sales are randomly distributed across these weeks. We further assume that the production lasts for $p = 130$ weeks in total, and once the production terminates, the sales of remaining units will continue for an extra 26 weeks at most. Thus, the product life cycle is $l = 156$ weeks. It is noteworthy that the production-sale process above is defined for simulation purposes, whereas in real applications the production-sale matrix should be obtained, period by period, from actual production and sales processes.

The second task is to formulate the warranty claim rate. In fact, our model can adapt to a wide variety of parameterizations of the NHPP. In this study, the power law process is adopted to characterize the claim rate, due to its ability and flexibility in describing various types of warranty claim rates. In this setting, the claim rate function is given by

$$\lambda(a) = \frac{\beta}{\eta} \left(\frac{a}{\eta} \right)^{\beta - 1}, \tag{14}$$

where $\beta > 0$ is the shape parameter and $\eta > 0$ is the scale parameter. In this simulation, the parameters of $\lambda_0(a)$ are assumed to be $\beta_0 = 3$ and $\eta_0 = 100$ for the first production scenario, and $\beta_0 = 1$ and $\eta_0 = 1000$ for the second one. The first set of model parameters is suitable for products with an increasing claim rate, such as mechanical products, whereas the second one is appropriate for products with a constant claim rate, such as electronic products.

An assignable cause, which may occur unpredictably, would result in an undesired increase in the warranty claim rate, which is the underlying reliability problem with which this article is concerned. For simulation purposes, we prespecify the timing and magnitude of the change of the claim rate (but consider them as unknown information when

Table 2. Simulation results (monitoring part) under the production scenario 1.

	$\tau = 0$			$\tau \sim U(15,20)$			$\tau \sim U(75,90)$		
k (weeks)	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$
$(0, \tau+1]$	0	0	0	0.0176	0.0176	0.0176	0.1576	0.1576	0.1576
	(0)	(0)	(0)	(0.0035)	(0.0035)	(0.0035)	(0.0102)	(0.0102)	(0.0102)
$(\tau+1, \tau+2]$	0.0002	0.0004	0.0013	0.0017	0.0017	0.0017	0.0025	0.0025	0.0025
	(0.0002)	(0.0004)	(0.0014)	(0.0005)	(0.0005)	(0.0005)	(0.0001)	(0.0001)	(0.0001)
$(\tau+1, \tau+6]$	0.0037	0.0084	0.0598	0.0091	0.0093	0.0104	0.0126	0.0127	0.0128
	(0.0015)	(0.0038)	(0.0327)	(0.0011)	(0.0011)	(0.0014)	(0.0002)	(0.0002)	(0.0002)
$(\tau+1, \tau+11]$	0.0156	0.0690	0.7855	0.0201	0.0235	0.0722	0.0253	0.0259	0.0299
	(0.0033)	(0.0184)	(0.1077)	(0.0015)	(0.0020)	(0.0303)	(0.0003)	(0.0004)	(0.0016)
$(\tau+1, \tau+21]$	0.1498	0.9537	1.0000	0.0637	0.2643	1.0000	0.0553	0.0799	0.7024
	(0.0218)	(0.0299)	(0.0000)	(0.0050)	(0.0700)	(0.0001)	(0.0011)	(0.0071)	(0.1374)
$(\tau+1, \tau+31]$	0.7726	1.0000	1.0000	0.2742	0.9992	1.0000	0.1315	0.7145	1.0000
	(0.0433)	(0.0000)	(0.0000)	(0.0402)	(0.0022)	(0.0000)	(0.0100)	(0.0960)	(0.0000)
$(\tau+1, \tau+51]$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0004)	(0.0000)	(0.0000)

Table 3. Simulation results (diagnosis part) under the production scenario 1.

	$\tau = 0$			$\tau \sim U(15,20)$			$\tau \sim U(75,90)$		
	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$
$g-\tau$	27.1440	16.8102	9.9187	33.9751	23.1177	14.8390	38.4462	28.4837	19.8070
	(5.7582)	(3.3132)	(2.0173)	(7.2959)	(4.2506)	(2.3611)	(8.2822)	(5.4372)	(3.2501)
$\widehat{\tau}-\tau$	0.7871	0.5939	0.5104	−2.1215	−1.6678	−1.1727	−6.4820	−5.5546	−4.1970
	(1.3115)	(1.0446)	(0.8974)	(5.4782)	(4.7475)	(3.9683)	(16.8033)	(16.1734)	(14.7931)
[5%, 95%]	[0, 3]	[0, 3]	[0, 2]	[−15, 3]	[−14, 2]	[−12, 1]	[−45, 3]	[−44, 2]	[−42, 2]
$\widehat{\tau}^*-\tau$	0.5178	0.3665	0.2866	−0.9100	−0.5840	−0.3399	−1.7074	−1.1408	−0.6741
	(0.8931)	(0.6597)	(0.5462)	(3.1585)	(2.1279)	(1.3671)	(5.0261)	(3.5026)	(2.2374)
[5%, 95%]	[0, 2]	[0, 2]	[0, 1]	[−7, 3]	[−4, 2]	[−3, 1]	[−11, 3]	[−7, 2]	[−4, 2]

Table 4. Simulation results (monitoring part) under the production scenario 2.

	$\tau = 0$			$\tau \sim U(15,20)$			$\tau \sim U(75,90)$		
k (weeks)	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$
$(0, \tau+1]$	0	0	0	0.0375	0.0375	0.0375	0.1859	0.1859	0.1859
	(0)	(0)	(0)	(0.0043)	(0.0043)	(0.0043)	(0.0099)	(0.0099)	(0.0099)
$(\tau+1, \tau+2]$	0.0023	0.0066	0.0554	0.0026	0.0028	0.0036	0.0026	0.0027	0.0030
	(0.0014)	(0.0060)	(0.0733)	(0.0002)	(0.0004)	(0.0017)	(0.0001)	(0.0001)	(0.0005)
$(\tau+1, \tau+6]$	0.0315	0.2192	0.9679	0.0156	0.0258	0.1244	0.0144	0.0180	0.0386
	(0.0074)	(0.0781)	(0.0509)	(0.0012)	(0.0064)	(0.0782)	(0.0005)	(0.0020)	(0.0142)
$(\tau+1, \tau+11]$	0.1524	0.9165	1.0000	0.0479	0.1942	0.9575	0.0362	0.0800	0.5520
	(0.0254)	(0.0546)	(0.0000)	(0.0059)	(0.0605)	(0.0627)	(0.0023)	(0.0163)	(0.1521)
$(\tau+1, \tau+21]$	0.7578	1.0000	1.0000	0.3042	0.9974	1.0000	0.1766	0.8966	1.0000
	(0.0416)	(0.0000)	(0.0000)	(0.0389)	(0.0043)	(0.0000)	(0.0175)	(0.0501)	(0.0000)
$(\tau+1, \tau+31]$	0.9964	1.0000	1.0000	0.8405	1.0000	1.0000	0.6469	1.0000	1.0000
	(0.0015)	(0.0000)	(0.0000)	(0.0339)	(0.0000)	(0.0000)	(0.0331)	(0.0000)	(0.0000)
$(\tau+1, \tau+51]$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

monitoring). On the one hand, we assume that the occurrence instant of the reliability problem (i.e., the change-point τ) is uniformly distributed within τ_a and τ_b, instead of being a constant. On the other hand, we consider that the reliability problem shifts the scale parameter from η_0 to $\eta_1 = (1-\rho)\eta_0, 0<\rho<1$, with the shape parameter remaining unchanged. A decrease in the scale parameter corresponds to an increase in the claim rate, and ρ represents the magnitude of such a change.

More specifically, three cases of τ are considered in this simulation, i.e., $\tau = 0$, $\tau \sim U(15,20)$, and $\tau \sim (75,90)$, which correspond to the reliability problem emerging before mass production, within $(0, w]$, and within $(w, l]$, respectively. The three cases cover different stages of the life cycle. Moreover, we also consider three cases of ρ, viz., $\rho = 0.10, \rho = 0.25$, and $\rho = 0.50$, which correspond to slight, moderate, and significant changes, respectively. Furthermore, the maximum desired false signal rate α is set to 0.0027, which is the most commonly adopted in-control setup in the SPM literature. Other values of α can also be chosen. The results for different values of α are similar and thus reported in Appendix C.

4.1.2. Simulation results

Based on the setting above, the simulation study is conducted (with 100 000 random runs for each single case) and the results are summarized in Tables 2 to 5 for production scenarios 1 and 2, respectively. Specifically, Tables 2 and 4 are about the monitoring part and Tables 3 and 5 about the diagnosis part.

Tables 2 and 4 present the averages and standard deviations (in parentheses) of the probabilities of: (i) at least one false alarm triggered before and in the $(\tau+1)$th week; and (ii) at least one true alarm after the $(\tau+1)$th week up to a

Table 5. Simulation results (diagnosis part) under the production scenario 2.

	$\tau = 0$			$\tau \sim U(15, 20)$			$\tau \sim U(75, 90)$		
	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.10$	$\rho = 0.25$	$\rho = 0.50$
$g-\tau$	17.4349	8.3113	4.3601	24.5995	14.0927	8.6099	28.0262	17.1782	10.9800
	(5.6477)	(2.3535)	(1.1846)	(7.1229)	(3.2995)	(1.8628)	(7.6180)	(3.8853)	(2.2438)
$\widehat{\tau}-\tau$	0.6754	0.2959	0.1286	−1.4340	−1.2854	−1.0046	−4.4906	−3.4893	−2.4793
	(1.1198)	(0.6118)	(0.3632)	(5.0137)	(4.0726)	(3.4248)	(14.9106)	(13.1004)	(11.2629)
[5%, 95%]	[0, 3]	[0, 2]	[0, 1]	[−14, 4]	[−11, 2]	[−9, 1]	[−38, 5]	[−31, 3]	[−11, 2]
$\widehat{\tau}^*-\tau$	0.5238	0.2658	0.1414	−0.4854	−0.3285	−0.1941	−0.8267	−0.5285	−0.3256
	(0.8907)	(0.5258)	(0.3656)	(2.8494)	(1.6890)	(1.0963)	(3.8567)	(2.2792)	(1.4479)
[5%, 95%]	[0, 2]	[0, 1]	[0, 1]	[−5, 4]	[−3, 2]	[−2, 1]	[−8, 5]	[−4, 3]	[−3, 2]

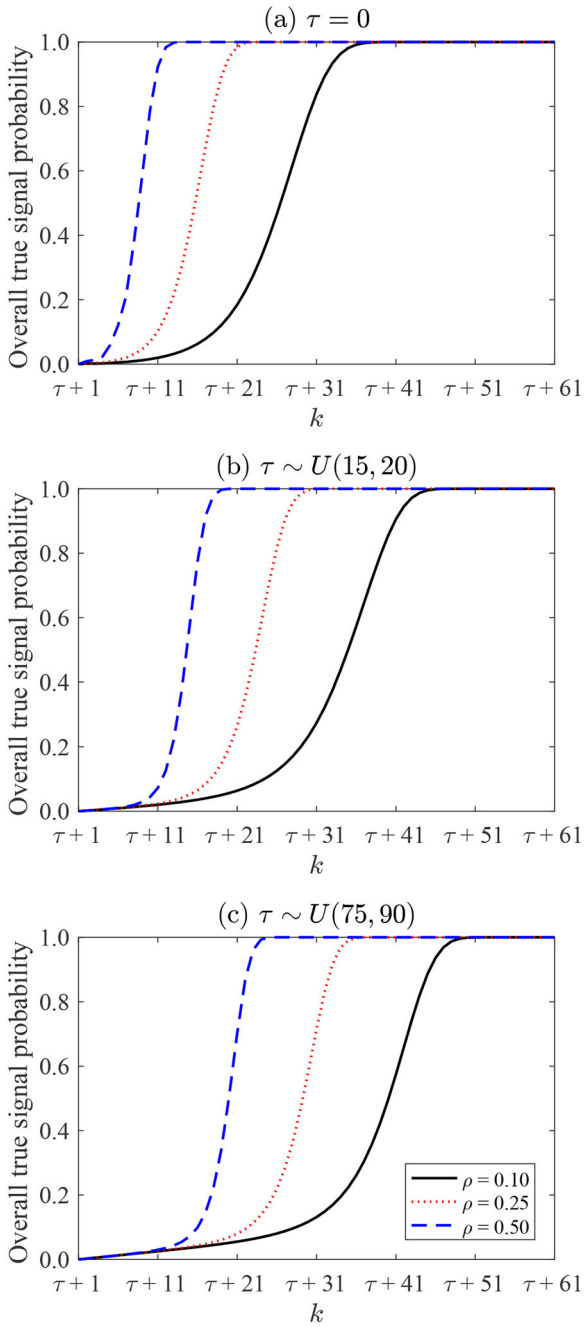

Figure 5. Overall true signal probability after the claim rate change under the production scenario 1. A faster growth in the overall true signal probability indicates a higher detection power.

number of selected time periods, over the 100 000 random runs. These probabilities are also known as the overall false signal probability and the overall true signal probabilities, respectively. The latter represents the detection ability of the control chart after the change point. The two types of signal probabilities usually serve as important performance indicators to assess an early detection scheme. Figures 5 and 6 illustrate the overall true signal probability over time under the production scenarios 1 and 2, respectively. The curves exhibit similar trends and relations under the two production scenarios. From these tables and figures, the following findings can be drawn:

1. Due to the discreteness of the Poisson distribution, it is typically impossible to exactly achieve the maximum desired false signal rate α when designing the control limits. The actual false signal rate in each week could be significantly lower than α when the expected number of claims per week is small, especially at the early stage of the life cycle. From Tables 2 and 4, it is evident that the overall false signal probabilities are lower than the corresponding values calculated in terms of a geometric distribution with parameter 0.0027 (see Section 3.3). Similarly, in some cases we even observe that the signal probabilities in the $(\tau + 2)$th week are lower than 0.0027. This implies that higher-than-expected protection against false alarms can be attained, while at the same time, a degree of detection sensitivity is sacrificed at the very beginning.
2. It is also observed that in most cases, the overall true signal probability after the change remains relatively low for a certain period of time. This observation is mainly due to: (i) only a small proportion of units manufactured in each week after τ have entered the market during this period of time, due to the sales delay; and (ii) in our simulation setting, the alteration of the claim rate stemming from the shift in scale parameter η_0 is tiny when the age a is small. After certain periods, the overall signal probability rapidly increases to and then stays at a value of one. In particular, under a specific production scenario and a specific change point case, the overall signal probability increases more rapidly as ρ increases. This is to be expected, as it would be easier and quicker to detect a larger magnitude of change in the claim rate.
3. When the underlying change occurs later in the life cycle, the overall signal probability increases more slowly, and it thus takes more weeks for this probability to reach a value of one (or a half). This is because a late occurrence of the reliability problem would lead to mixed warranty claims from units manufactured both

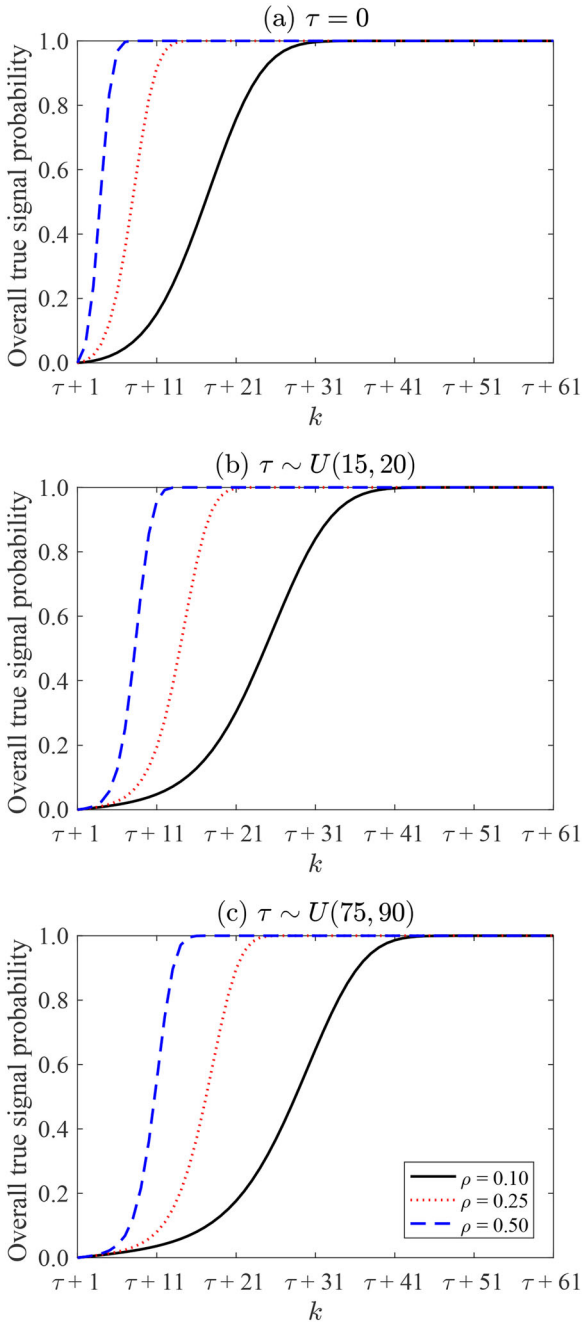

Figure 6. Overall true signal probability after the claim rate change under the production scenario 2.

before and after τ (the inertia issue), which in turn reduces the power of anomaly detection. This means that it is more difficult to detect a reliability problem occurring at the later stage of the life cycle, compared with one emerging earlier.

4. The overall signal probabilities under scenario 2 increase earlier and also more rapidly than those under scenario 1. This can be principally attributed to the fact that the production and sales volumes under scenario 2 are much larger (moreover, the units have a higher claim rate than those under scenario 1 when a is not too large), which can accumulate more warranty claims within a given time horizon. The detection power regarding a given rate change depends on the expected number of warranty claims in each week. On the other hand, as discussed previously, it also affects the attained false signal rate. It is observed that the overall false signal probabilities in Table 2 (for scenario 1) are smaller than those in Table 4 (for scenario 2). This indicates that the false signal rates under scenario 1 are generally lower than those under scenario 2, which further widens the gap in detection power between the two production scenarios.

In addition to the curve of overall true signal probability, the expected signal delay, measured by $E[g-\tau]$, which is analogous to the concept of *average run length* in classic SPM research, can also be used to indicate the detection power. Tables 3 and 5 show the averages and standard deviations (in parentheses) of the signal delays as well as some diagnostic indicators over the 100 000 random runs under production scenarios 1 and 2, respectively. From the signal delays, we can draw the same findings as those from the signal probability curves: The detection power is higher (i.e., the signal delay is shorter) when the underlying reliability problem occurs earlier, the magnitude of the change of the claim rate is larger, and/or the production and sales volumes are larger.

Furthermore, upon an out-of-control signal, the post-signal diagnosis scheme should be initiated to estimate the change point as accurately as possible. The diagnosis accuracy can be measured by the expected diagnosis error $E[\widehat{\tau}-\tau]$; see Tables 3 and 5. The following findings can be drawn:

1. Similar to the detection power, the diagnosis accuracy is higher (i.e., the average and standard deviation of the diagnosis errors are smaller) when τ is smaller, ρ is larger, and/or the production and sales volumes are larger.
2. In Tables 3 and 5, we introduce a benchmark (ideal) case in which $\lambda_1(a)$ is assumed pre-known, and the estimated change point with perfect knowledge of $\lambda_1(a)$ is $\widehat{\tau}^*$. As can be observed, when $\lambda_1(a)$ has to be estimated, the diagnosis accuracy is worse than when $\lambda_1(a)$ is pre-known. This is to be expected, as the estimation of $\lambda_1(a)$ would introduce an additional degree of inaccuracy. Nevertheless, the accuracy of the proposed diagnosis scheme is acceptable, especially when τ is small.
3. We also notice that the distribution of the diagnosis errors is skewed to the left (see the 5% and 95% quantiles of the diagnosis errors under the "known" and "unknown" cases of $\lambda_1(a)$) when the underlying change occurs later in the product life cycle, in which case both $E[\widehat{\tau}-\tau]$ and $E[\widehat{\tau}^*-\tau]$ are less than zero. This means that, on an average, the estimated change point is always earlier than the actual one. When the change occurs very early (say, $\tau = 0$), the situation is reversed. Thus, for higher diagnosis accuracy, further refinements and enhancements to the diagnosis scheme are worth studying in the future.

Beyond Tables 3 and 5, in each single simulation, say, r, we notice an interesting dilemma between $g_r-\tau_r$ and $\widehat{\tau}_r-\tau_r$

under the "unknown" cases of $\lambda_1(a)$. If the signal delay $g_r-\tau_r$ is small (resp. large), upon a signal one has less (resp. more) historical out-of-control data for estimating $\widehat{\Lambda}_1(a)$, which in turn results in a large (resp. small) diagnosis error $\widehat{\tau}_r-\tau_r$. This implies that the signal delay (representing the detection power) and the diagnosis error (indicating the diagnosis effectiveness) cannot be simultaneously reduced.

4.2. Comparison study

In this subsection, we compare the proposed dynamic Shewhart-type scheme with the CUSUM scheme in Lawless *et al.* (2012) to demonstrate its effectiveness in terms of early detection power. The comparison is conducted under the simulation setting in Section 4.1.1.

Lawless *et al.* (2012) similarly monitor warranty claims with "ΔM_k" but adopt a CUSUM scheme obtained from a likelihood ratio statistic for specified alternatives. Here, ΔM_k is captured in quotation marks, because its exact definition differs slightly from ΔM_k in our work; they additionally incorporate a moving window approach by using only the incremental warranty claims from units manufactured in recent periods. For a fair comparison, here we do not include this additional adjustment to the CUSUM scheme and our scheme, simply using ΔM_k (in Section 5, we will study the moving window approach in detail). In particular, the CUSUM statistic of Lawless *et al.* (2012) under our simulation setting is of the following form:

$$W_k = \max\{0, W_{k-1} + \Delta M_k - \psi E[\Delta M_k]\}, k = 1, 2, \ldots,$$

where $W_0 = 0$, and $\psi = [1-(1-\rho)^{-\beta_0}]/[\beta_0 \log(1-\rho)]$.

Identical control limits (i.e., $UCL_k = c$) are used in their CUSUM scheme to simplify the choice of a monitoring plan. Hence, the CUSUM scheme is designed by specifying c and ψ, so as to give specific signal probabilities under given scenarios. The explicit expression of ψ does not mean that we can easily derive its exact value. This is because ψ is a function of ρ that is usually unknown. Using a CUSUM statistic with improperly specified ρ (and thus ψ) will reduce its power in detecting the actual change. In this comparison, we assume that the CUSUM statistic can be correctly specified with the exact value of ψ, as the benchmark of all possible CUSUM statistics. To conduct the comparison fairly, we compare the two monitoring schemes for which the false signal probabilities within a given time horizon are the same. Thus, for the CUSUM scheme, c is chosen to (approximately) have the false signal probability during the first 26 weeks (about half a year after the production and sales of the product) identical to that of our scheme.

One issue with the use of the CUSUM scheme in Lawless *et al.* (2012) is that both c and ψ are determined before monitoring warranty claims, as noted earlier. This design requires full knowledge of the sales data of each future period. Lawless *et al.* (2012) implicitly assume that future sales information is pre-known, but this is scarcely possible in reality. Nevertheless, to facilitate the comparison, we allow the CUSUM scheme to be designed with full knowledge in each run of the simulation. Then, the Markov chain method is used to compute signal probabilities as in Lawless *et al.* (2012). The technical details are omitted for the sake of brevity.

Figure 7 displays the overall true signal probability over time for the two schemes under production scenario 1. It is observed that the proposed monitoring scheme exhibits overwhelming superiority when τ and ρ are larger. In other cases, the two schemes perform similarly or the CUSUM scheme slightly outperforms our scheme. Bear in mind that the CUSUM scheme considered in this comparison has always been specified with an ideal value of ψ (that is designed to be effective against the specified alternative) and assumed to have perfect knowledge of future sales volumes. If these ideal conditions cannot be satisfied (which is common in reality), the detection ability of the CUSUM scheme would become worse and the advantage of our scheme would be more obvious.

Although it is well recognized that CUSUM-type schemes are more advanced than, and superior to, Shewhart-type schemes, in this particular comparison we do not come to the same conclusion. The use of constant control limits heavily detracts from the detection ability of the CUSUM scheme in Lawless *et al.* (2012), whose conditional false alarm rate varies at each time period. In particular, its conditional false alarm rate generally stays at a very low level over a long period of time after the change. From Figure 7, we observe that the overall true signal probability of the CUSUM scheme is very close to zero for certain periods, especially when τ and ρ are larger. After that, it sharply increases to one. Note that the comparison result under production scenario 2 is consistent with that under scenario 1, and thus omitted due to space considerations.

In summary, the proposed dynamic monitoring scheme is superior to the CUSUM scheme in Lawless *et al.* (2012), not only for faster detection, but also for adaptive control chart design (no need of prior knowledge about the rate change and future sales volumes). In this sense, developing an improved CUSUM scheme with dynamic probability control limits is an interesting future research topic. However, this design will become much more complicated, and it is left for future research.

5. Model extension: Moving window approach

As discussed in Section 3.2, after the reliability problem occurrs, M_k contains a mass of (historical) claims of the units that were produced a long time ago. This leads to an "inertia" problem. This problem cannot be completely eliminated with the use of ΔM_k. In this section, motivated by Lawless *et al.* (2012), we attempt to further mitigate this problem by considering the warrantied units that are manufactured during the B ($B \geq 1$) recent periods. That is to say, in the kth period, only sold units manufactured within periods $[1\vee(k-B), k-1]$ are considered as the population under monitoring. In this case, ΔM_k in Equation (7) should be modified to

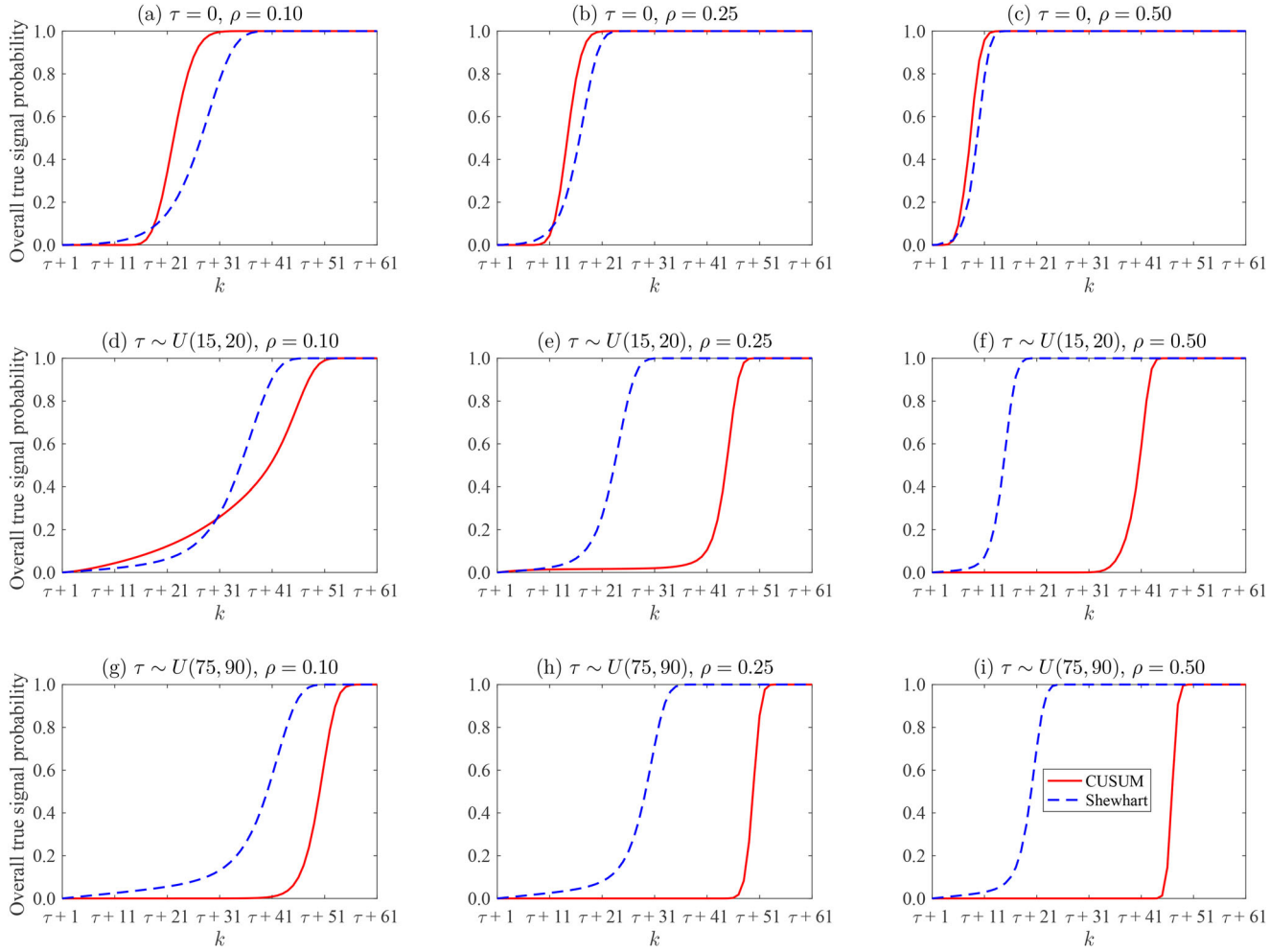

Figure 7. Comparison of the proposed scheme with the CUSUM scheme under different combinations of change point τ and change magnitude ρ under the production scenario 1.

$$\widehat{\Delta M}_k = \begin{cases} \sum_{j=1\vee(k-B)}^{k-1} \sum_{i=1\vee(k-B)}^{j\wedge p} \sum_{u=1}^{D_{i,j}} \left(X_{i,j,k}^{(u)} - X_{i,j,k-1}^{(u)} \right), & 1 \le k \le w+1, \\ \sum_{j=(k-w)\vee(k-B)}^{k-1} \sum_{i=1\vee(k-B)}^{j\wedge p} \sum_{u=1}^{D_{i,j}} \left(X_{i,j,k}^{(u)} - X_{i,j,k-1}^{(u)} \right), & w+1 < k \le l+1, \\ \sum_{j=(k-w)\vee(k-B)}^{l} \sum_{i=1\vee(k-B)}^{j\wedge p} \sum_{u=1}^{D_{i,j}} \left(X_{i,j,k}^{(u)} - X_{i,j,k-1}^{(u)} \right), & l+1 < k \le l+w, \\ 0, & k > l+w. \end{cases} \tag{15}$$

Then, Z_k in Equation (9) can be computed by replacing ΔM_k with $\widehat{\Delta M}_k$, and the control limits can be obtained by Equation (10) in a similar way. While in the diagnosis part, all historical warranty claims are included without screening (i.e., we do not consider B) to increase the estimation accuracy of the abnormal field claim rate $\lambda_1(a)$. In other words, the maximum-likelihood-based post-signal diagnosis scheme does not need to be modified in the moving window approach.

One thing noteworthy is that Lawless *et al.* (2012) propose to incorporate B into their original model to reduce the inertia problem; however, they simply choose one particular value of B in their case study without any quantitative investigations or demonstrations. In fact, this moving window approach works at the expense of information loss to some extent, especially when B is small. Therefore, a sensible balance is required when choosing the value of B. In this study, the effect of B on the detection and diagnosis powers is investigated thoroughly, by directly following the simulation setting in Section 4.1.1. The evaluation result under production scenario 1 is shown in Figure 8. Note also that the result under production scenario 2 is quite similar to that of scenario 1, and thus omitted.

Figure 8 shows that the way that the signal probability curve changes with B for $\tau = 0$ is quite different from that for $\tau \sim U(15, 20)$ or $\tau \sim U(75, 90)$. When $\tau = 0$ (i.e., the underlying change occurs at the very beginning), all product units are manufactured under the out-of-control condition. Since all warranty claims are generated by the out-of-control product population, in this case it is surely better to make use of as much warranty claims data as possible for early detection purposes. This can be observed from the first row of Figure 8, where a steeper signal probability curve can be observed for a larger B, indicating a higher detection ability. For $\tau = 0$, the limiting case (i.e., $B = \infty$, which is the original model without regard to truncation by production periods) is the optimal case. This implies that if the underlying change is expected to occur very early (say, $\tau = 0$), the

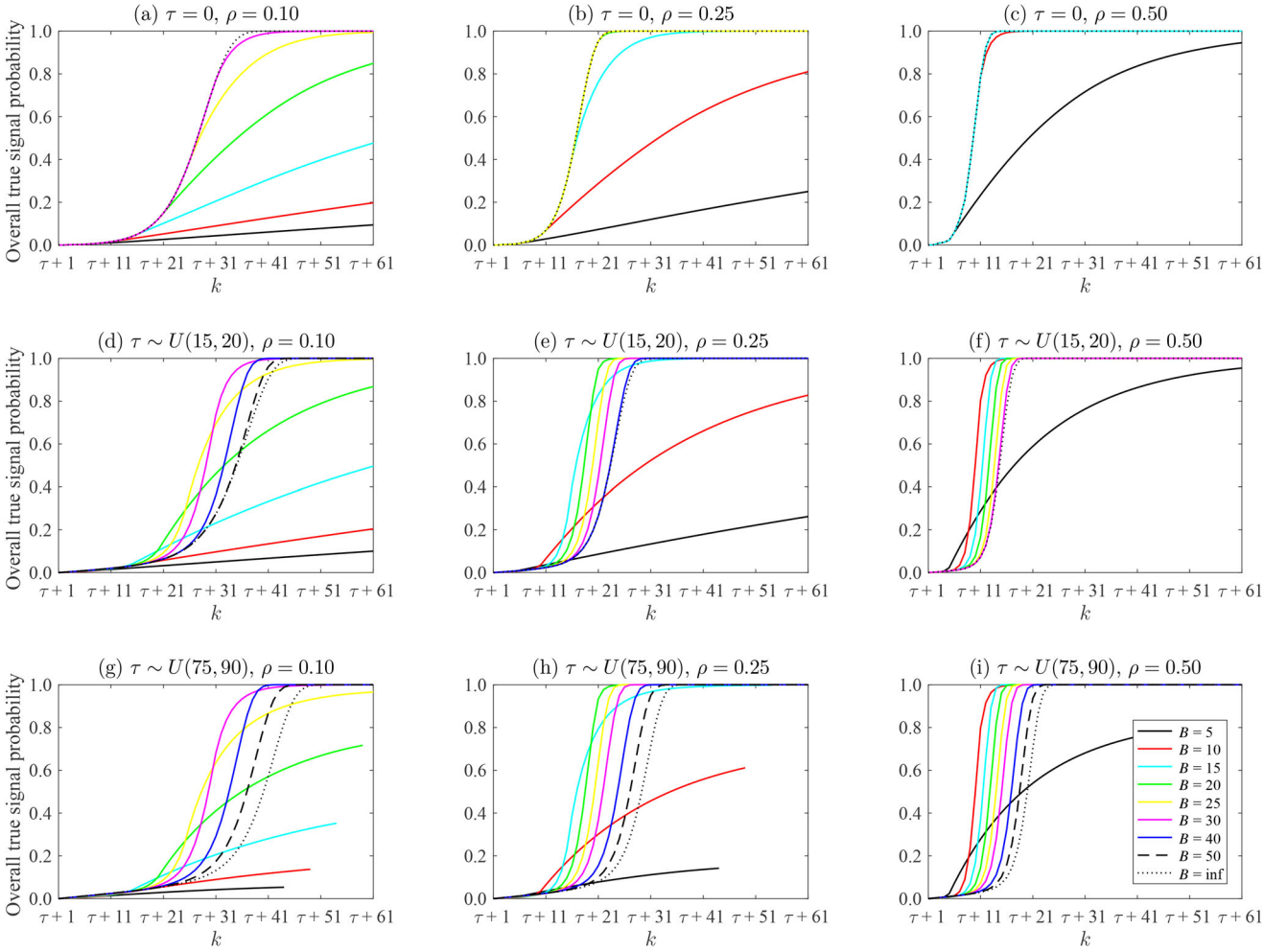

Figure 8. Effect of B on the true signal probability after the change under the production scenario 1.

manufacturer should include all warranty claims data, as in the original model.

However, if the underlying change occurs at a later stage of the product life cycle, the effect of B appears more complicated. Unlike the case of $\tau = 0$, no obvious optimal choice of B can be found for $\tau \sim U(15, 20)$ or $\tau \sim U(75, 90)$, in terms of the signal probability curve. The second and third rows of Figure 8 show that along with time, the curves dominate the field alternately (in ascending order of B). This is because a smaller B is able to exclude the units manufactured before τ more quickly from the monitored product population, leading to a better detection power at the early stage after the change; on the other hand, a small B also restricts the detection power at a low level at later stages, due to insufficient information being utilized. The situation is reversed for a larger B. Nevertheless, we can roughly see that small (resp. large) values of B are preferable for detecting large (resp. small) rate changes. It is worth noting that the limiting case is inferior to most cases. In summary, choosing a reasonable value of B in this case can improve the detection ability as long as B is not too small.

Moreover, the curves approach to the limiting case faster as ρ increases. Also notice that the monitoring process ceases earlier if a relatively smaller B is considered. This is why some of the curves in the case of $\tau \sim U(75, 90)$ are shorter. Furthermore, although the post-signal diagnosis scheme can be implemented without any model modifications, the diagnosis performance is still affected by the choice of B. The reason is in line with the discussion at the end of Section 4.1.2, i.e., the trade-off between the signal delay and diagnosis error.

Lawless *et al.* (2012) do not investigate the effect of B on the performance of their CUSUM scheme, but we find that they exhibit similar patterns. To make the investigation more thorough, a small-scale comparison between the proposed Shewhart scheme and the CUSUM scheme in Lawless *et al.* (2012) is performed when both of them take the moving window into account. Figure 9 shows the performance comparison of the two schemes for particular choices of B. It can be seen that the CUSUM scheme incorporating the moving window tends to be more competitive compared with the Shewhart scheme with the same choice of B, and the CUSUM scheme gains a clear advantage when $B = 10$. However, the superiority of the CUSUM scheme here is maintained just for small-to-moderate changes (say, $\rho = 0.25$), and would rapidly diminish if ρ increases. In addition, as we have mentioned in Section 4.2, the CUSUM scheme in Lawless *et al.* (2012) is an ideal version and could be unreliable in practical use.

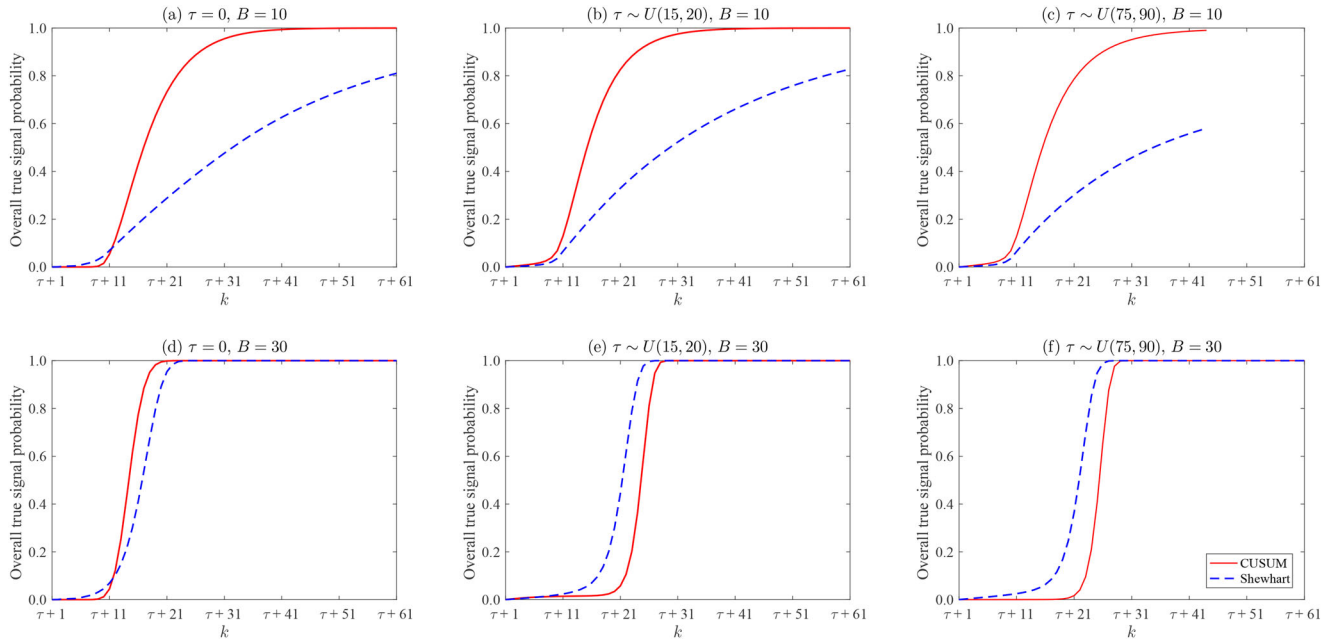

Figure 9. Comparison of the Shewhart scheme with the CUSUM scheme under the production scenario 1, when $\rho = 0.25$.

The findings above provide some guidelines on how to choose a reasonable value of B, which largely complement the introduction of the moving window method in Lawless *et al.* (2012). These are of importance to manufacturers who intend to improve early detection ability.

6. A real-data application

In this section, the integrated monitoring and diagnosis scheme described in Section 3 is demonstrated using real warranty claims data collected from a home appliance manufacturer.

The company manufactures and sells a home appliance product. The product is sold with a 52-week warranty, under which the company should provide free minimal repair services for eligible claims. The production data and associated warranty claims data of the units manufactured in 175 weeks, from March 2015 to July 2018, were collected (the data is preprocessed and scaled for confidential issues). In total, 8 349 101 units were produced and 808 820 claims were received until the collection date. As the claim rate is unacceptablly high, the company wants to find out if some quality and reliability problems have occurred during the course of design and manufacturing, through warranty data analysis.

To this end, we first fit the baseline warranty claim rate by using the standard MLE method. This is based on the front part of the dataset, specifically, the claims data of units manufactured within the first 30 weeks. It turns out that the claim rate approximately follows a power law process with shape parameter $\beta_0 = 3.053$ and scale parameter $\eta_0 = 113.300$ weeks, i.e., $\lambda_0(a) = 1.637 \times 10^{-6} \times a^{2.053}, a \geq 0$. It is necessary to note that detailed sales data is not fully available, as the company has no control over all sales channels. Instead, the sales data of an online direct-sales store is used to estimate the sales delay. It is evident that the sales delay is between 13 and 78 days, and its 95% confidence interval is [22, 54] days. Therefore, we assume that the units produced in a specific week will be uniformly sold from the third week to the eighth week since the production.

Then, the dynamic monitoring scheme is retrospectively implemented to the entire dataset. The maximum desired false signal rate α is set to 0.0027. Figure 10(a) shows the one-sided Shewhart-type control chart, based on the statistic Z_k in Equation (9), until the 250th week. We observe that no points fall above the control limit for a long time, until the first out-of-control signal is triggered at the 134th week. After that, Z_k fluctuates dramatically and more points fall beyond the control limit. This is a clear indication that the process becomes out of control. It is worth mentioning that the control limit in Figure 10(a) rapidly approaches and then stays around a constant (i.e., 2.7822). This demonstrates that Z_k (if in control) indeed follows an approximate standard normal distribution for sufficiently large values of $E[\Delta M_k]$, as stated before.

Upon the first out-of-control signal, the maximum-likelihood-based diagnosis scheme is applied to estimate the change point. For this purpose, the MLE method in Appendix B is adopted to estimate the out-of-control claim rate. According to Figure 10(b), the maximum likelihood is attained at $\widehat{\tau} = 95$, thus the reliability problem most possibly occurred just before the 96th week. Based on this diagnostic result, the company initiates an investigation on the possible root cause(s) that resulted in this abnormal pattern, around the 95th week. By checking the log book, it is found that there is a design modification around the 101-102th weeks, which is quite near to the estimated change point. Of course, a design modification does not necessarily imply an improvement in the product reliability. Such modifications

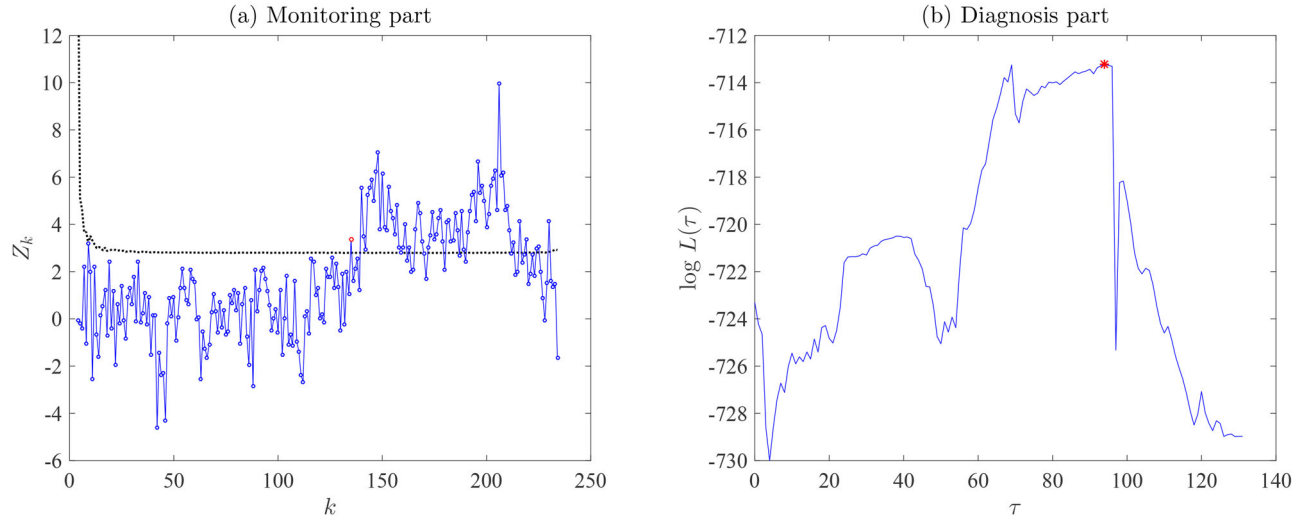

Figure 10. Control chart and post-signal diagnostic result for the real dataset. The first out-of-control signal occurs at the 134th week, and the most likely change point is the 95th week.

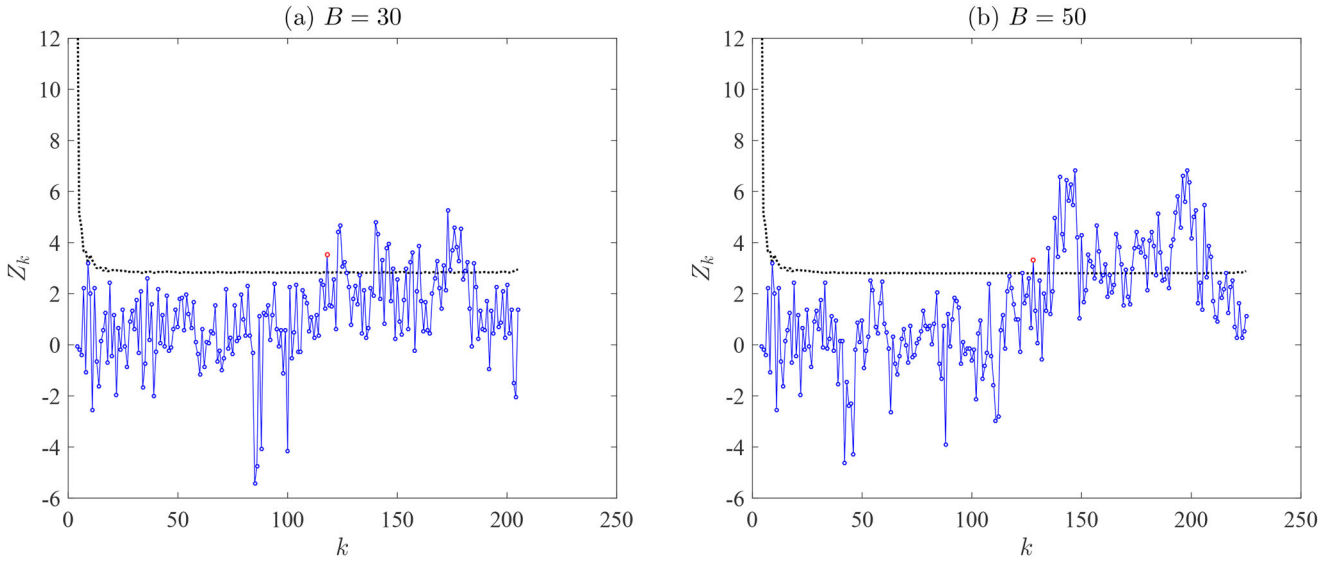

Figure 11. Control charts for two different choices of B. The first out-of-control signal occurs at the 117th week for $B = 30$, and at the 127th week for $B = 50$.

can be from a function perspective. Nevertheless, this means that it takes about 33 weeks to detect and diagnose the reliability problem. Such a detection power is generally unacceptable, even though the sales delay is present.

To improve the detection power, the moving window approach described in Section 5 is adopted and the corresponding control charts for two different choices of B are illustrated in Figure 11. It shows that the earliest out-of-control signal occurs at the 117th week for $B = 30$, and at the 127th week for $B = 50$, implying that the delays in detection are 16 and 26 weeks, respectively. The improvement in detection power is remarkable when considering only some recent production periods. Nevertheless, a small B, say $B = 30$, will result in significant information loss; a large proportion of the out-of-control signals is thus hidden; see Figure 11(a). Therefore, in this case, $B = 50$ is suggested since the resulting detection performance is good and the information loss is acceptable.

7. Conclusion

This article develops an integrated monitoring and diagnosis framework to detect reliability problems over the product life cycle, through monitoring aggregate warranty claims, and to estimate the most probable instant of problem occurrence, upon an alarm signal. The dynamic monitoring and diagnosis scheme is easy to design and implement in the sense that manufacturers only need to periodically collect production data, sales data and cumulative warranty claims data, just as many of them have been doing in practice.

Simulation and comparison studies, as well as a real-data application, show that the integrated scheme is able to detect the emerging reliability problem in a timely manner, and further estimate the change point with an acceptable accuracy. The main conclusions of this paper are summarized as follows:

1. It is shown that larger production and sales volumes contribute to earlier detection of reliability problems, since more warranty claims can be accumulated in each time period. In this regard, the method by Wu and Meeker (2002), which groups product units by manufacturing periods and tests for different ages separately, may not provide enough units for detecting slight or even moderate claim rate increases. This also supports our use of aggregate warranty claims.
2. Moreover, the proposed scheme is more powerful to detect and diagnose early reliability problems (say, change point $\tau = 0$). In fact, a majority of reliability problems emerge at the very early stage of the product life cycle (e.g., faults in design, unanticipated failure modes, use of unreliable raw materials, and unanticipated operating condition), which result in a high warranty claim rate. Product recalls can be predominately attributed to such reliability problems. In this sense, the scheme is expected to be useful for manufacturers.
3. Furthermore, the moving window approach is explored to mitigate the inertia problem and thus improve the detection power. It is shown that no optimal choice of B can dominate in all scenarios, and manufacturers should sensibly choose the value of B based on their detection objectives or expert experiences. For example, when a manufacturer pays more attention to detect reliability problems that emerged in the design phase, or previous experience shows that serious reliability problems mostly occur in the design phase, he/she should adopt the original model (without considering B); otherwise, choosing a reasonable value of B could improve the detection power as long as B is not too small.
4. More importantly, through comparing with the CUSUM scheme (with a constant control limit) in Lawless *et al.* (2012), it is found that the dynamic Shewhart-type control chart is superior in most cases, but tends to be less competitive when considering the moving window. Taking into account its other advantages, including easy design and implementation and no need of additional prior information (e.g., future sales volumes), the dynamic Shewhart-type scheme is recommended. In addition, our scheme considers a general alternative hypothesis $\lambda_1(a|i) > \lambda_0(a)$ such that it can maintain relatively robust detection power for a wide range of alternatives (of unknown form).

The scheme can be further extended in several directions to improve its applicability. For instance, the reference claim rate $\lambda_0(a)$ may be updated regularly, which is not uncommon for many multi-generation products. Our model can be modified to adapt to this scenario. Moreover, this article considered a Shewhart-type control chart, while dynamically-designed CUSUM and EWMA charts can be explored in the future. Finally, warranty claims data were exclusively used for early detection of reliability problems, while the wide availability of smart phones and social media makes online customer reviews (comments and/or complaints) become another source of reliability information to complement warranty claims data (Hong *et al.*, 2018). How to incorporate text analytics into the early detection model to improve its performance is an open question.

Acknowledgments

The authors are grateful to the Editors and the anonymous referees for their insightful comments that helped to significantly improve this article.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 71601166, 71532008, 71801179), the Research Grants Council of Hong Kong under Theme-based Research Fund (Grant No. T32-101/15-R) and General Research Fund (Grant Nos. CityU 11213116, CityU 11203519), and also the Singapore AcRF Tier 2 Funding (Grant No. R-266-000-125-112).

Notes on contributors

Chenglong Li is an associate professor in the School of Management at Northwestern Polytechnical University, China. He received his Ph.D. from Xi'an Jiaotong University and City University of Hong Kong in 2017. He received a B.E. in industrial engineering from Xi'an Jiaotong University in 2011. Currently, his research interests are mainly on quality engineering, statistical modeling and intelligent decision.

Xiaolin Wang is a Ph.D. candidate in industrial engineering at City University of Hong Kong. He received his B.E. and M.S. degrees in industrial engineering from Southeast University in 2013 and 2016, respectively. Currently, he is interested in applying machine learning, stochastic modeling, and optimization techniques in reliability, maintenance, and warranty management areas. He is a recipient of the prestigious Hong Kong Ph.D. Fellowship (HKPFS) from the Hong Kong Research Grants Council.

Lishuai Li is an assistant professor in the Department of Systems Engineering and Engineering Management at City University of Hong Kong. She received a Ph.D. and a M.S. in air transportation systems from Massachusetts Institute of Technology (MIT). She received a B.E. from Fudan University. She is interested in developing analytical methods and practical algorithms to improve air transportation systems with the use of large-scale data generated from real-world operations.

Min Xie is a Chair Professor in industrial engineering at City University of Hong Kong. He received his Ph.D. from Linkoping University, Sweden in 1987. He did his undergraduate study and received a M.S. at Royal Institute of Technology in Sweden in 1984. Dr. Xie joined the National University of Singapore in 1991 as one of the first recipients of the prestigious Lee Kuan Yew Research Fellowship. He has authored or co-authored numerous refereed journal papers and several books. He is a Department Editor of *IISE Transactions* and Editor of *Reliability Engineering & System Safety* and serves in a number of other international journals. He has organized many international conferences, and also 50 Ph.D. students have graduated under his supervision. He was elected a fellow of IEEE for his contribution to systems and software reliability.

Xin Wang is a research fellow in the Department of Industrial Systems Engineering and Management at National University of Singapore. She

received a Ph.D. in industrial engineering from National University of Singapore. She received a B.E. in international economy and trade from Harbin Institute of Technology. Her research interests include warranty management, resilience modeling and assessment, and reliability engineering.

ORCID

Xiaolin Wang http://orcid.org/0000-0003-0100-8154

References

Amiri, A. and Allahyari, S. (2012) Change point estimation methods for control chart postsignal diagnostics: A literature review. *Quality and Reliability Engineering International*, **28**(7), 673–685.

Chen, P. and Ye, Z.S. (2017) Estimation of field reliability based on aggregate lifetime data. *Technometrics*, **59**(1), 115–125.

Chien, Y.H. (2019) The optimal preventive-maintenance policy for a NHPBP repairable system under free-repair warranty. *Reliability Engineering & System Safety*, **188**, 444–453.

Cook, R.J. and Lawless, J. (2007) *The Statistical Analysis of Recurrent Events*, Springer Science & Business Media, New York, NY.

Dong, Y., Hedayat, A.S. and Sinha, B.K. (2008) Surveillance strategies for detecting changepoint in incidence rate based on exponentially weighted moving average methods. *Journal of the American Statistical Association*, **103**(482), 843–853.

Hong, Y., Zhang, M. and Meeker, W.Q. (2018) Big data and reliability applications: The complexity dimension. *Journal of Quality Technology*, **50**(2), 135–149.

Huang, W., Shu, L., Woodall, W.H. and Tsui, K.L. (2016) CUSUM procedures with probability control limits for monitoring processes with variable sample sizes. *IIE Transactions*, **48**(8), 759–771.

Jin, T., Taboada, H., Espiritu, J. and Liao, H. (2017) Allocation of reliability-redundancy and spares inventory under Poisson fleet expansion. *IISE Transactions*, **49**(7), 737–751.

Lawless, J.F., Crowder, M.J. and Lee, K-A. (2012) Monitoring warranty claims with CUSUMs. *Technometrics*, **54**(3), 269–278.

Liu, B., Wu, J. and Xie, M. (2015) Cost analysis for multi-component system with failure interaction under renewing free-replacement warranty. *European Journal of Operational Research*, **243**(3), 874–882.

Meeker, W.Q. and Hong, Y. (2014) Reliability meets big data: Opportunities and challenges. *Quality Engineering*, **26**(1), 102–116.

Montgomery, D.C. (2013) *Introduction to Statistical Quality Control*, seventh edition. John Wiley & Sons, New York, NY.

Richards, S.C., Woodall, W.H. and Purdy, G. (2015) Surveillance of nonhomogeneous Poisson processes. *Technometrics*, **57**(3), 388–394.

Rosevear. (2017) Ford's latest recalls will put a $295 million dent in profits. https://www.fool.com/investing/2017/03/29/fords-latest-recalls-will-put-a-295-million-dent-i.aspx. Accessed 14 April, 2018.

Ryan, A.G. and Woodall, W.H. (2000) Control charts for Poisson count data with varying sample sizes. *Journal of Quality Technology*, **42**(3), 260–275.

Shang, L., Si, S., Sun, S. and Jin, T. (2018) Optimal warranty design and post-warranty maintenance for products subject to stochastic degradation. *IISE Transactions*, **50**(10), 913–927.

Shang, Y., Wang, Z., He, Z. and He, S. (2017) Nonparametric change-point detection for profiles with binary data. *Journal of Quality Technology*, **49**(2), 123–135.

Shen, X., Tsui, K.L., Zou, C. and Woodall, W.H. (2016) Self-starting monitoring scheme for Poisson count data with varying population sizes. *Technometrics*, **58**(4), 460–471.

Shen, X., Zou, C., Jiang, W. and Tsung, F. (2013) Monitoring Poisson count data with probability control limits when sample sizes are time varying. *Naval Research Logistics*, **60**(8), 625–636.

Shu, L., Su, Y., Jiang, W. and Tsui, K-L. (2014) A comparison of exponentially weighted moving average-based methods for monitoring increases in incidence rate with varying population size. *IISE Transactions*, **46**(8), 798–812.

Sogandi, F., Aminnayeri, M., Mohammadpour, A. and Amiri, A. (2019) Risk-adjusted Bernoulli chart in multi-stage healthcare processes based on state-space model with a latent risk variable and dynamic probability control limits. *Computers & Industrial Engineering*, **130**, 699–713.

Thomas, A. (2016). Harley recalling 27,000 motorcycles for clutch problem. https://www.biztimes.com/2016/industries/manufacturing-logistics/harley-recalling-27000-motorcycles-for-clutch-problem/. Accessed 30 March, 2018.

Van der Heijden, M. and Iskandar, B.P. (2013) Last time buy decisions for products sold under warranty. *European Journal of Operational Research*, **224**(2), 302–312.

Wang, X., He, K., He, Z., Li, L. and Xie, M. (2019) Cost analysis of a piece-wise renewing free replacement warranty policy. *Computers & Industrial Engineering*, **135**, 1047–1062.

Wang, X., Li, L. and Xie, M. (2019) Optimal preventive maintenance strategy for leased equipment under successive usage-based contracts. *International Journal of Production Research* in press, DOI: 10.1080/00207543.2018.1542181.

Wang, X. and Xie, W. (2018) Two-dimensional warranty: A literature review. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **232**(3), 284–307.

Wang, X., Xie, W., Ye, Z.-S. and Tang, L.C. (2017) Aggregate discounted warranty cost forecasting considering the failed-but-not-reported events. *Reliability Engineering & System Safety*, **168**, 355–364.

Wu, H. and Meeker, W.Q. (2002) Early detection of reliability problems using information from warranty databases. *Technometrics*, **44**(2), 120–133.

Wu, S. (2012) Warranty data analysis: A review. *Quality and Reliability Engineering International*, **28**(8), 795–805.

Wu, S. (2013) A review on coarse warranty data and analysis. *Reliability Engineering & System Safety*, **114**, 1–11.

Xie, W., Shen, L. and Zhong, Y. (2017) Two-dimensional aggregate warranty demand forecasting under sales uncertainty. *IISE Transactions*, **49**(5), 553–565.

Xie, W. and Ye, Z.-S. (2016). Aggregate discounted warranty cost forecast for a new product considering stochastic sales. *IEEE Transactions on Reliability*, **65**(1), 486–497.

Yang, W., Zou, C. and Wang, Z. (2017). Nonparametric profile monitoring using dynamic probability control limits. *Quality and Reliability Engineering International*, **33**(5), 1131–1142.

Yashchin, E. (2012) Design and implementation of systems for monitoring lifetime data, in *Frontiers in Statistical Quality Control 10*. Springer, Berlin, pp. 171–195.

Ye, Z. and Ng, H.K.T. (2014) On analysis of incomplete field failure data. *The Annals of Applied Statistics*, **8**(3), 1713–1727.

Zhang, X. and Woodall, W.H. (2015) Dynamic probability control limits for risk-adjusted Bernoulli CUSUM charts. *Statistics in Medicine*, **34**(25), 3336–3348.

Zhao, X. and He, S. and Xie, M. (2018) Utilizing experimental degradation data for warranty cost optimization under imperfect repair. *Reliability Engineering & System Safety*, **177**, 108–119.

Zhou, C., Chinnam, R.B., Dalkiran, E. and 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.

Zhou, C., Chinnam, R.B. and 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.

Zhou, Q., Zou, C. Wang, Z. and Jiang, W. (2012) Likelihood-based EWMA charts for monitoring Poisson count data with time-varying sample sizes. *Journal of the American Statistical Association*, **107**(499), 1049–1062.

Appendix A: Proof of Proposition 1

It is noteworthy that the behavior of warranty claims varies in different stages of the product life cycle. To derive M_k, we need to split the time horizon into four sub-intervals: (i) $1 \le k \le w+1$, (ii) $w+1<k \le l+1$, (iii) $l+1<k \le l+w$, and (iv) $k > l+w$. Below, the four cases are discussed one by one according to the specific situation of M_k in each sub-interval.

Case 1: $1 \le k \le w+1$. In this case, all units sold before, and including, the $(k-1)$th period are under warranty, and are eligible to generate warranty claims. The manufacturer is therefore responsible for all claimed failures. Then, we have

$$M_k = \sum_{j=1}^{k-1} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)}, 1 \le k \le w+1, \tag{16}$$

where $D_{i,j}$ is the number of units manufactured in period i and sold in period j, and $X_{i,j,k}^{(u)}$ is the cumulative number of warranty claims for the uth unit manufactured in period i and sold in period j up to period k, as defined before.

Case 2: $w+1<k \le l+1$. The behavior of warranty claims within this interval is different from that of Case 1. In this case, the units sold within $[1, k-w-1]$ and $(k-w-1, k-1]$ should be considered separately, as follows.

i. For all units sold during $[1, k-w-1]$, their warranties have ceased by period k, and the cumulative warranty claims of an individual unit can be represented by $X_{i,j,j+w}^{(u)}$. Thus, we need to count the total warranty claims generated by each unit sold within this interval, i.e.,

$$\sum_{j=1}^{k-w-1} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)}.$$

ii. For units sold during $(k-w-1, k-1]$, their warranty coverage will not expire by period k. Analogous to Case 1, the warranty claims of the units sold in this interval is

$$\sum_{j=k-w}^{k-1} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)}.$$

Consequently, combining the two sources of warranty claims above yields

$$M_k = \sum_{j=1}^{k-w-1} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)} + \sum_{j=k-w}^{k-1} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)}, w+1<k \le l+1. \tag{17}$$

Case 3: $l+1<k \le l+w$. As mentioned earlier, the sales process terminates at period l. That is to say, no sales occur beyond this period, and the cumulative sales volume is thus $\sum_{j=1}^{l} S_j$. Among all of the sold units, for those sold before and in period $k-w-1$, they have already been out-of-warranty by period k; while those sold in the interval $(k-w-1, l]$ will still be protected by the warranty up to period k.

Following the similar derivation procedures in Case 2, the aggregate warranty claims can be expressed as

$$M_k = \sum_{j=1}^{k-w-1} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)} + \sum_{j=k-w}^{l} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,k}^{(u)}, l+1<k \le l+w. \tag{18}$$

Case 4: $k > l+w$. In this case, no new sales will occur and all of the sold units have run out of warranty. In other words, the manufacturer does not need to bear the repair cost of any further failures. Hence, we can obtain the total aggregate warranty claims, denoted by M_{total}, of all the sold units, as follows:

$$M_{total} = \sum_{j=1}^{l} \sum_{i=1}^{j\wedge p} \sum_{u=1}^{D_{i,j}} X_{i,j,j+w}^{(u)}, k > l+w. \tag{19}$$

As can be seen, when $k \to l+w$, (18) is equal to (19). □

Appendix B: MLE method for parameter estimation

The MLE method is a well-adopted approach to estimating the parameters of warranty claim rate function. For illustrative purposes, we assume that the out-of-control claim rate is also characterized by the power law process in Equation (14), i.e., $\lambda_1(a) = \beta a^{\beta-1}/\eta^{\beta}$. Let the random variable $\widetilde{\Delta M}_k$ denote the number of cumulative warranty claims in the kth period for the units manufactured from period $\tau+1$ up to period $(k-1)\wedge p$. It is clear that $\widetilde{\Delta M}_k$ follows a Poisson process with mean

$$E\left[\widetilde{\Delta M}_k\right] = \sum_{i=\tau+1}^{(k-1)\wedge p} \sum_{j=i\vee(k-w)}^{(k-1)\wedge l} D_{i,j} \left(\left(\frac{k-j}{\eta}\right)^{\beta} - \left(\frac{k-j-1}{\eta}\right)^{\beta} \right).$$

Accordingly, the likelihood function can be formulated as

$$L(\eta, \beta | \widetilde{Q}_k, k = \tau+2, \ldots, g) = \prod_{k=\tau+2}^{g} \frac{\exp\left\{-E\left[\widetilde{\Delta M}_k\right]\right\} E\left[\widetilde{\Delta M}_k\right]^{\widetilde{Q}_k}}{\widetilde{Q}_k!},$$

where $\widetilde{Q}_k$ is the observed number of cumulative warranty claims in the kth period for the units manufactured from period $\tau+1$ up to period $(k-1)\wedge p$. By maximizing $\ln L(\eta, \beta)$, the MLE estimates $\widehat{\eta}$ and $\widehat{\beta}$ can be obtained by

$$\widehat{\eta} = \left(\frac{\sum_{k=\tau+2}^{g} \widetilde{Q}_k}{\sum_{k=\tau+2}^{g} \Psi_k} \right)^{1/\widehat{\beta}}, \tag{20}$$

and

$$\sum_{k=\tau+2}^{g} \frac{\widetilde{Q}_k \Gamma_k}{\Psi_k} \sum_{k=\tau+2}^{g} \Psi_k = \sum_{k=\tau+2}^{g} \Gamma_k \sum_{k=\tau+2}^{g} \widetilde{Q}_k, \tag{21}$$

where $\Psi_k = \sum_{i=\tau+1}^{(k-1)\wedge p} \sum_{j=i\vee(k-w)}^{(k-1)\wedge l} D_{i,j}((k-j)\widehat{\beta} - (k-j-1)\widehat{\beta})$ and $\Gamma_k = \partial \Psi_k / \partial \widehat{\beta}$. Here, it should be pointed out that Equation (21) must be solved numerically. The direct search and bisection search techniques are good candidates for solving this non-linear equation.

Appendix C: Results for different values of α

In the simulation study above, we only present the results for $\alpha = 0.0027$ due to space limitations. In fact, the results for different values of α within a normal range are consistent. For demonstration purposes, Figure 12 presents the comparative results for different choices of α under the production scenario 1. In addition to $\alpha = 0.0027$, two more choices of α, one larger ($\alpha = 0.005$) and the other smaller ($\alpha = 0.002$), are considered. It shows that the results for different choices of α are very much alike, though the overall signal probability increases more rapidly for a larger α. It is notable that the curves for different choices of α come closer when ρ is larger. These findings confirm our earlier statement that the results for different choices of α are consistent.

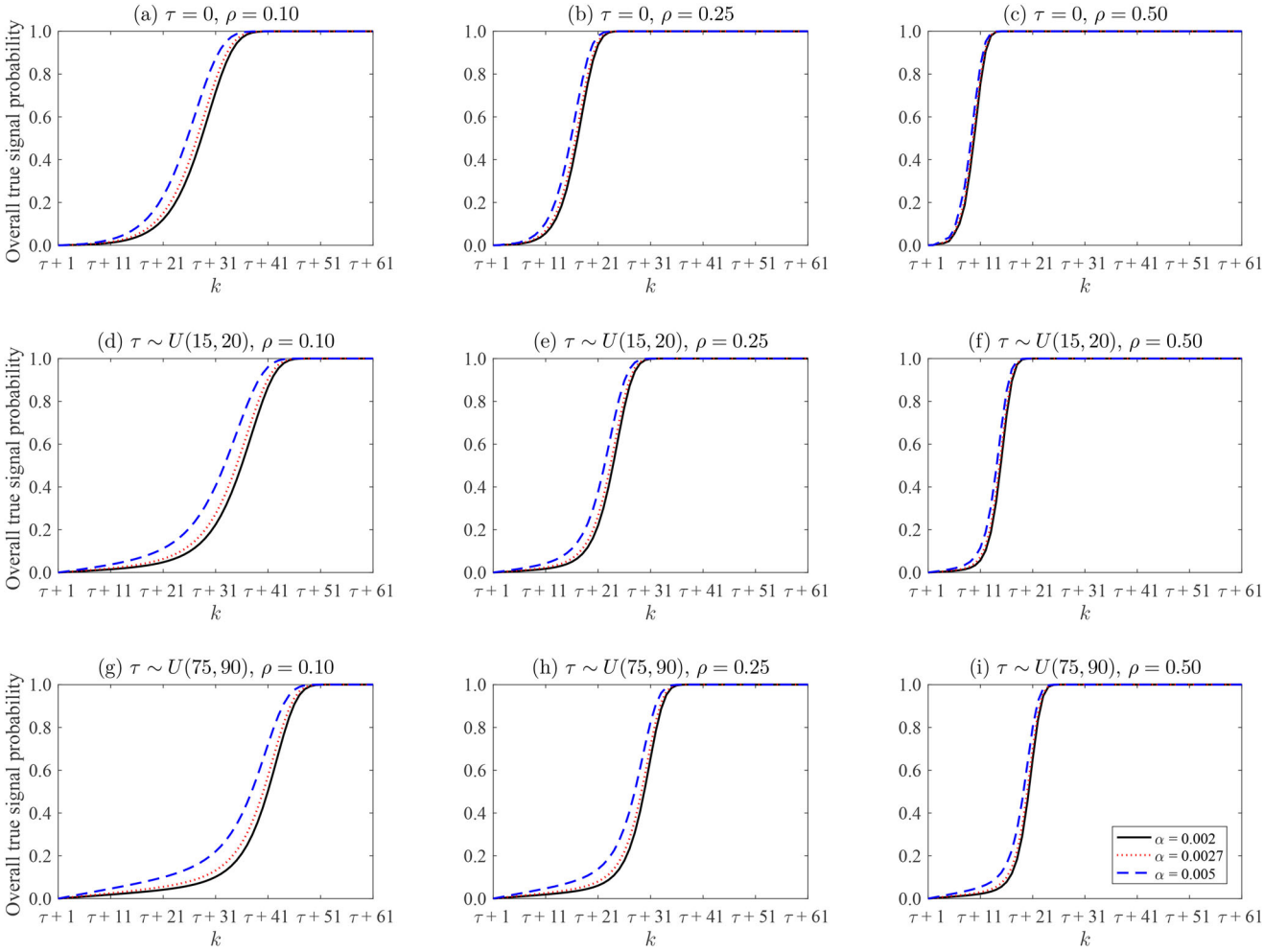

Figure 12. Overall true signal probability, after the claim rate change, of the Shewhart scheme under the production scenario 1 for different choices of α.