RANDOM PROPORTIONAL WEIBULL HAZARD MODEL: APPLICATION TO LARGE-SCALED INFORMATION SYSTEMS

Kiyoyuki KAITO[1], Kengo OBAMA[2], Kiyoshi KOBAYASHI[3], Kazuya AOKI[4], Koji YAMAMOTO[5]

ABSTRACT. In this study, aiming for asset management of large-scaled information systems that support infrastructures, the failure generation process of wear-out failure component, whose failure rate changes with time, is formulated with a Weibull hazard model. Information systems are composed of many devices. In order to consider the heterogeneity of the hazard rate of each device, the random proportional Weibull hazard model, which expresses the heterogeneity of the hazard rate in random variables is proposed. Furthermore, the authors develop a methodology that expresses the heterogeneity of hazard rates in gamma distribution, as well as estimates unknown parameters and the heterogeneity of hazard rates contained in the hazard model. Finally, using the traffic control system of expressways, the failure rate for wear-out failure component is estimated from actual failure history data, and the validity of the methodology is investigated through a case study.

INTRODUCTION

In large-scaled infrastructures, with an aim for efficiency of operation and speedy provision of information to users, information systems composed of various monitoring sonsors and processing/output devices are established. Asset management of information systems is as important a problem as the infrastructures. Points to be noted for asset management of information systems are, 1) that the information system is a large-scaled system formed from an enormous amount of components, 2) that the system has a hierarchical structure and a failure in each device can possibly develop into a functional failure of the whole system, and 3) that the deterioration of the information system's function, such as obsolescence, as well as the physical deterioration, is an important management point.

---

[1] Assoc. Prof., Frontier Research Center, Osaka University, Japan, e-mail:kaito@ga.eng.osaka-u.ac.jp
[2] Ph.D. candidate, Department of Urban Management, Kyoto University, Japan, e-mail: k.obama@psa.mbox.media.kyoto-u.ac.jp
[3] Prof., Graduate School of Management, Kyoto University, Japan, e-mail: kkoba@psa.mbox.media.kyoto-u.ac.jp
[4] Assistant Manager, R & D Center, PASCO CORPORATION, Tokyo, Japan, e-mail: kazuya_aoki@pasco.co.jp
[5] Sub-Chief Engineer, Facilities Development Management Team, Yokohama Regional Office, Central Nippon Expressway Co.Ltd., Yokohama, Japan, e-mail: k.yamamoto.af@c-nexco.co.jp

When performing asset management of information systems it is necessary to consider the differences in the management levels of the 1) component level, 2) system level, and 3) function level. Regarding the perspective of asset management, the latter points require a more integrated consideration. Among these, the authors develop a random proportional Weibull hazard model to carry out a fault analysis at the component level, of the components in large-scaled information systems. Of course, in order to perform asset management on information systems, a fault analysis at the level of each component is not enough, and it is necessary to develop a methodology that simultaneously achieves the above three points. The fault analysis model for the component level proposed in this study can be a basic analysis tool for constructing integrated asset management systems of large-scaled information systems.

Information systems are composed of various types of components. These components can be divided into two groups: accidental failure equipment, whose failure rate does not depend on time, and wear-out failure equipment whose failure rate increases over time. Generally, regarding the fault process of accidental failure equipment, the hazard rate expressed by the moment's failure rate density can be formulated by an exponential hazard model that does not depend on time (or a Poisson based model of failure events). On the other hand, with wear-out failure equipment, a non-homogeneous hazard model that takes into consideration the hazard rate's dependency on time is necessary. With wear-out failure equipment, there is the characteristic that the failure rate increases as time progresses after installing the system. Therefore, in order to make decisions about renewals and replacements, information regarding the deteriorating process of wear-out failure equipment is needed. For this, the authors focus on wear-out failure equipment, and analyze the mechanism of the failure rate increasing with time with a Weibull hazard model, known as a representative non-homogeneous hazard model. However, large-scaled information systems comprise many components, and different types of components might have different hazard rates. Therefore, when analyzing faults of information systems that comprise various types of devices and components, it is important to consider the heterogeneity of the hazard rates that exist between the different types of components. In this study, with this in consideration, the random proportional Weibull hazard model, whose heterogeneity of hazard rates are in accord with a gamma distribution, is formulated and a methodology that estimates the failure rate of various components that compose an information system is proposed.

BASIC IDEAS OF THIS STUDY

Overview of the Existing Study

In traditional hazard analysis (Lancaster, 1990), the targeted system is assumed to be constructed of elements of the same quality, and the aim is to express a model of failure generation randomly attained according to a certain hazard function. With hazard analysis, the random failure generation process is modeled, and a deterministic model called as a hazard function is employed. However, large-scaled information systems, such as the one targeted for actual analysis in this study, have complex structures that are composed of enormous types of components. When managing and operating a large-scaled information system, considering the strategy of when to change each component among the many and considering the storing policy is an important theme. However, the failure rate of all of the components might not be expressed by the same hazard rate. Rather, it is natural to assume that the hazard rate differ for each type of component. As the method to express the heterogeneity of the

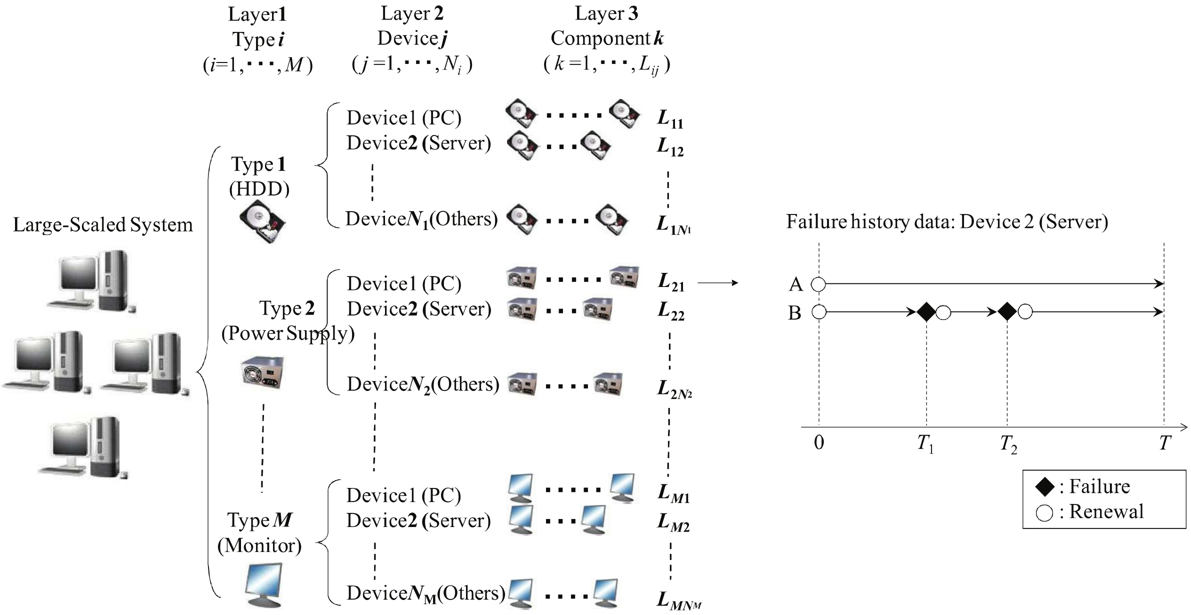
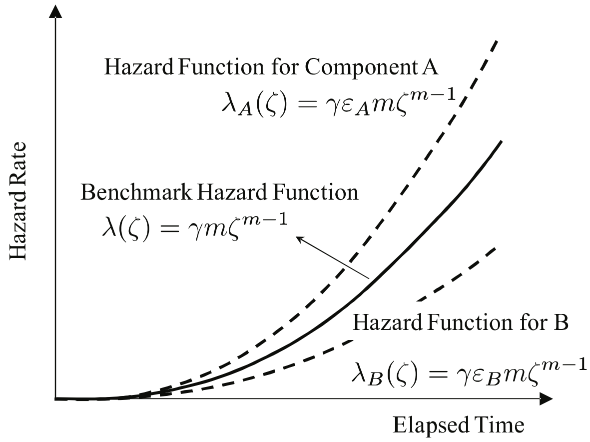


Figure 1. Large-Scaled Infromation System and History Data of Failure Occurrences

hazard rates of different types of components, there is the 1) Method to express the difference of the component characteristics in dummy variables and 2) Method of considering the probability distribution of the hazard rate. The former method is simple and easy to understand. However, as the types of components increase, the dummy variables needed to express the components' characteristic also increase, and the efficiency of the model's estimation results decline considerably. In fact, in large-scaled information systems, even components of the same type, categorized by type and device, have different deterioration characteristics depending on how it is used and where it is located, and these systems are structures with extremely divided deterioration characteristics. It is not practical to use dummy variables for deterioration characteristics sub-divided in this way in hazard models. Therefore, in this study the authors chose the latter method, using a random proportional Weibull hazard model that expresses the heterogeneity of hazard rates of different components in probability distribution, and formulate a model for the failure process of component groups that compose information systems.

Basic Frame of Modeling

This is a model for the occurrence and process of failure events in information systems. As Figure 1. shows, components of information systems are in three layers: 1) Type, 2) Device, and 3) components. Type-layer contains hard disk drive (HDD), power supply, and processing and monitoring equipment. An information system is composed of M-unit Type components and each component is represented by suffix i ($i=1, \ldots, M$). Type i components are used for N_i-unit Device and each Device is represented by suffix j ($j=1, \ldots, N_i$). In a traffic control system, for example, each Type of components is used as different Devices such as a personal computer (PC), a server and so on. Since each Device has its own application of components and therefore its failure probability is different from others'. For Device j ($j=1, \ldots, Ni$), L_{ij}-unit Type i components are used and each component is represented by suffix k



Note: The baseline model applies to ε=1. Also, The hazard function of component B is a function multiplied by $\varepsilon_B > 1$ on the baseline model, and the hazard function shifts upwards. While component is $\varepsilon_A < 1$ and the hazard rate proportionally shifts downwards.

Figure 2. Heterogeneity of Hazard functions

($k=1, \ldots, L_{ij}$). Components in each Type and each Device are considered to have different hazard rates. But failure process in components in each Device is considered to be described by using the same hazard rates. Here, an infinitely continuous time axis starting from time point t=0 is used. If the existing information systems as a whole are renewed at t=0, deterioration of each component starts from t=0. When a component has a failure, it is immediately replaced. The new one is supposed to have the same performance that the old one had. Now, take a look at $t=T$ where a certain period of time has passed. Then, a failure history is obtained as shown in Figure 2., which cites an example of a failure history of Device 2 (server). Device 2 is composed of L_2-unit components. Among them, component A has no failure from t=0. Time of the use of component A is T and the lifetime of component A is considered to be longer than T. On the other hand, component B had a failure twice at T_1 and T_2. The first lifetime $\zeta = T_1$ and the second $\zeta = T_2 - T_1$.

Here, it is supposed that each type of component has the failure characteristics of wear-out failure component. With wear-out failure component, as shown in Figure 2., the generation rates of failures (hazard rates) grow as the elapsed time moves farther from the nearest time of renewal. This type of lifespan distribution for wear-out failure component is often used to express the components' time-dependent deterioration. It is assumed that this is subject to the Weibull distribution. Furthermore, the hazard rates of different types of components can be expressed in time functions as are shown in Figure 2. Functions like this that express the hazard rate's temporal change are called hazard functions. The hazard functions of each component are in an expansion or contraction by a factor of a baseline hazard function. A model that expresses proportionally expanded or contracted functions is called as proportional hazard models. If the failure process of each type of component that compose each device can be mutually expressed with a proportional hazard model, the heterogeneity in hazard rates can be expressed by the probability distribution of proportionality constants of hazard functions. Information systems comprise many devices, but in most cases, the number of the same type of components in each device is not that large. By estimating the parameter of the standard proportional Weibull hazard function and the parameter of the probability distribution that expresses the heterogeneity of the proportionality constant between the types, the random proportional Weibull hazard model can easily express the heterogeneity of the hazard rates between types and components. On the other hand, if the heterogeneity of the Weibull hazard rate cannot be expressed by the proportional hazard model, it becomes necessary to estimate the Weibull hazard model for the different types and devices. However, if the number of components of the same type that compose each device is small, it becomes difficult to estimate the Weibull hazard model. If the above points are considered, the random proportional Weibull hazard model proposed in this study is very effective for expressing the failure process of information systems that have sub-divided component structures. However, in order to use the random proportional Weibull hazard model, it is a prerequisite that the heterogeneity of the failure rates between the types are in a proportional relation. the hypothesis test of the proportional relation is carried out in the later section.

RANDOM PROPORTIONAL WEIBULL HAZARD MODEL

Formulation of Random Proportional Weibull Hazard Model

Large-scaled information systems are, as shown in Figure 1., composed of M types of components, and the component at number i (i = 1, …, M) is used in a total of N_i devices. Furthermore, the total number of type i components used in device j is L_{ij}. Among the type i components, the component at number k (k = 1, …, L_{ij}) out of the components that compose device j (j = 1, …, Ni) is picked up. The elapsed time since this component has been renewed will be expressed as t_{ij}^k. The arrival rate of each component's failure occurrence is in accord with the random proportional Weibull hazard model.

$$\lambda_{ij}\left(t_{ij}^k\right)=\varepsilon_{ij}\gamma_i m\left(t_{ij}^k\right)^{m-1} \tag{1}$$

However, γ_i is the parameter expressing the arrival density of type i, and m is the acceleration parameter. Equation (1) is the general Weibull hazard function (Aoki *et al*, 2007; Tsuda *et al*, 2006) with the parameter ε_{ij} (hereinafter, heterogeneity parameter) expressing the heterogeneity (Maher, 1996) of the hazard rates of type i and device j. The heterogeneity parameter expresses the heterogeneity of the hazard rates between the components of different devices and types. Especially, in the case that ε_{ij} = 1, the random proportional Weibull hazard function (1) matches the general Weibull hazard function. This kind of hazard function is called as a baseline hazard function (See Figure 2.). However, for the same component used in the same device, the heterogeneity parameter has a common value. The heterogeneity parameter takes a deterministic value in reality, but is a parameter that is impossible to observe. Also, the probability density function $f_{ij}\left(t_{ij}^k\right)$ and survival probability $\overline{F}_{ij}\left(t_{ij}^k\right)$ of the lifespan of type i component k in device j are as follows.

$$f_{ij}\left(t_{ij}^k\right)=\varepsilon_{ij}\gamma_i m\left(t_{ij}^k\right)^{m-1}\exp\left\{-\gamma_i\varepsilon_{ij}\left(t_{ij}^k\right)^m\right\} \tag{2a}$$

$$\overline{F}_{ij}\left(t_{ij}^k\right)=\exp\left\{-\gamma_i\varepsilon_{ij}\left(t_{ij}^k\right)^m\right\} \tag{2b}$$

The value of the heterogeneity parameter is a probabilistic variable which is distributed according to a certain probability distribution. The random proportional Weibull hazard model (1) has the same deterioration acceleration parameter m for all components, but the hazard arrival density $\varepsilon_{ij}\gamma_i m$ expresses a proportionately different deterioration characteristic for each type and device. Regarding the hypothesis test of the homogeneity of the acceleration parameter (hereinafter, proportionality), that is considered in the later section. In this study, a Weibull hazard model in which the hazard arrival density has a proportional distribution for each type and device (if there are sub-divisions in how they are used and where they are located, these last categories are included), is defined as a random proportional Weibull hazard model. Depending on the hazard arrival density, different random proportional Weibull hazard models can be formulated. In other words, as shown in Table 1., two expressions are possible: 1) The case in which the hazard arrival density is different for each type of component (model 1), and 2) The hazard arrival density is the same, without

Table 1. Model's Parameters

Model	Arrival Density	Acceleration Rate	Variance
1	$\gamma_1 \neq \cdots \neq \gamma_M$	m	ϕ^{-1}
2	$\gamma_1 = \cdots = \gamma_M$	m	ϕ^{-1}

depending on the type of component i (model 2). However, the heterogeneity parameter is thought to follow the same proportional distribution.

Here, the heterogeneity parameter ε_{ij} is subject to the gamma distribution. Furthermore, focus on model 1, and use the case in which there are different averages of the heterogeneity parameter for each type. Gamma distribution as a special form contains exponential distribution, and can express the exponential probability density function family defined by $[0,\infty)$. Also, it has the merit that it is easy to handle analytically. Here, the parameter γ_i expresses the average hazard arrival density of the type i component, and that the heterogeneity parameter ε_{ij} is a probability error term that is subject to the gamma distribution of average 1, variance ϕ^{-1}. The gamma distribution is defined by $[0,\infty)$, and regarding the arbitrary explanatory variable and probability error term, the right side of equation (1) is guaranteed to have a positive value. Generally, the probability density function $g(\varepsilon_{ij} : \alpha, \beta)$ of the gamma distribution $G(\alpha,\beta)$ can be defined as follows.

$$g(\varepsilon_{ij} : \alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \varepsilon_{ij}^{\alpha-1} \exp\left(-\frac{\varepsilon_{ij}}{\beta}\right) \tag{3}$$

The average of gamma distribution $G(\alpha,\beta)$ is $\mu = \alpha\beta$, and the variance is $\sigma^2 = \alpha\beta^2$. Also, $\Gamma(\cdot)$ is the gamma function. Furthermore, the probability density function $\bar{g}(\varepsilon_{ij} : \phi)$ of the gamma distribution of average 1, variance ϕ^{-1} can be expressed as follows.

$$\bar{g}(\varepsilon_{ij} : \phi) = \frac{\phi^{\phi}}{\Gamma(\phi)} \varepsilon_{ij}^{\phi-1} \exp(-\phi\varepsilon_{ij}) \tag{4}$$

When expressing the probability distribution of the heterogeneity parameter with a standard gamma distribution, the difference of the two models can be listed as shown in Table 1.

Estimating Method of the Model

Without losing generality, the model 1 for the random proportional Weibull hazard model is employed. Model 2 is merely a simplified form of model 1, with the added conditions as shown in Table 1. The random proportional Weibull hazard model (model 1) has unknown parameters such as arrival density parameters γ_i (i=1,…, M) for each type, an acceleration parameter m, different heterogeneity parameters ε_{ij} (i = 1, …, M ; j = 1, …, N_i) for each type and device and the variance parameter ϕ of the heterogeneity parameter. Normally with Weibull hazard models, the parameters γ_i and m can be estimated from failure history data. However, with the random proportional Weibull hazard model, it is necessary to find the variance parameter ϕ of the heterogeneity parameter, and the heterogeneity parameters ε_{ij} (i = 1, …, M ; j = 1, …, N_i) for each type and device, as well.

Now the database of the failure history of the targeted system is available. All the information regarding the time that each component failed (was exchanged) since the time of installation, of the targeted system, is stored in the database. Then, the failure history of the component is expressed as $\Xi = (\xi_1,\ldots,\xi_M)$, and $\xi_i = (\xi_{i1},\ldots,\xi_{iNi})$ is the failure history of type i component. Also, ξ_{ij} is the failure history for the type i component of device j, and is $\xi_{ij} = \{(\delta_{ij}^1, t_{ij}^1),\ldots,(\delta_{ij}^{Lij}, t_{ij}^{Lij})\}(i=1,\ldots,M; j=1,\ldots,Ni)$. Also, δ_{ij}^k is the dummy variable which takes the value 1 if the type i component k (k =1, …, L_{ij}) of device j fails, and the value 0 if it does not. t_{ij}^k is the in-service time (or lifespan) of the type i component k of device j. Therefore, when δ_{ij}^k=0, t_{ij}^k is the length of time since installation to the present. On the other hand, when δ_{ij}^k=1, t_{ij}^k is the lifespan. Here it is assumed that the heterogeneity parameter $\bar{\varepsilon}_{ij}$ is given. The conditional likelihood $\ell_{ij}\left(\xi_{ij} : \gamma_i, m, \bar{\varepsilon}_{ij}\right)$ with observed data concerning faults of type i components of device j can be expressed as follows.

$$\ell_{ij}\left(\xi_{ij} : \gamma_i, m, \bar{\varepsilon}_{ij}\right) = \prod_{k=1}^{Lij}\left\{\bar{F}_{ij}\left(t_{ij}^k : \gamma_i, m, \bar{\varepsilon}_{ij}\right)\right\}^{\left(1-\delta_{ij}^k\right)}\left\{f_{ij}\left(t_{ij}^k : \gamma_i, m, \bar{\varepsilon}_{ij}\right)\right\}^{\delta_{ij}^k} \tag{5}$$

However, with the above equation, the probability density function of the lifespan distribution $f_{ij}\left(t_{ij}^k : \gamma_i, m, \varepsilon_{ij}\right)$ and the survival function $\bar{F}_{ij}\left(t_{ij}^k : \gamma_i, m, \varepsilon_{ij}\right)$ are explicitly expressed as the parameter $\gamma_i, m, \varepsilon_{ij}$ function. Here, when the heterogeneity parameter ε_{ij} is subject to the standard gamma distribution $\bar{g}\left(\varepsilon_{ij} : \phi\right)$, the likelihood function of the observed data ξ_{ij} can be expressed as follows.

$$L_{ij}\left(\xi_{ij} : \theta_i\right) = \int_0^\infty \prod_{k=1}^{L_{ij}} \left\{\widetilde{F}_{ij}\left(t_{ij}^k : \gamma_i, m, \varepsilon_{ij}\right)\right\}^{\left(1-\delta_{ij}^k\right)} \left\{f_{ij}\left(t_{ij}^k : \gamma_i, m, \varepsilon_{ij}\right)\right\}^{\delta_{ij}^k} \bar{g}\left(\varepsilon_{ij} : \phi\right) d\varepsilon_{ij}$$

$$= \frac{\phi^\phi}{\Gamma(\phi)} \prod_{k=1}^{L_{ij}} \left\{\gamma_i m \left(t_{ij}^k\right)^{m-1}\right\}^{\delta_{ij}^k} \int_0^\infty \varepsilon_{ij}^{s_{ij}+\phi-1} \exp\left\{-\left(\phi+\gamma_i\tau_{ij}\right)\varepsilon_{ij}\right\} d\varepsilon_{ij} \tag{6}$$

However, $\theta = \left(\gamma_i, m, \phi\right)$. Also, $s_{ij} = \sum_{k=1}^{L_{ij}} \delta_{ij}^k, \tau_{ij} = \sum_{k=1}^{L_{ij}} \left(t_{ij}^k\right)^{m-1}$. With the above equation, the heterogeneity parameter ε_{ij} of all type i components of device j takes the same value. Notice that to express this, the authors define the likelihood function $L_{ij}(\xi_{ij} : \theta_i)$ as the expected value regarding the probability variable ε_{ij} of the conditional likelihood $\ell_{ij}\left(\xi_{ij} : \gamma_i, m, \varepsilon_{ij}\right)$. Hence, by the transformation of the variable $x_{ij} = \varepsilon_{ij}(\phi + \gamma_i t_{ij})$, the following is obtained

$$L_{ij}\left(\xi_{ij} : \theta_i\right) = \frac{\phi^\phi}{\Gamma(\phi)} \prod_{k=1}^{L_{ij}} \left\{\gamma_i m \left(t_{ij}^k\right)^{m-1}\right\}^{\delta_{ij}^k} \int_0^\infty \left(\frac{x_{ij}}{\phi+\gamma_i\tau_{ij}}\right)^{s_{ij}+\phi-1} \exp\left(-x_{ij}\right) \frac{dx_{ij}}{\phi+\gamma_i\tau_{ij}}$$

$$= \frac{\phi^\phi}{\Gamma(\phi)} \frac{\Gamma\left(s_{ij}+\phi\right)}{\left(\phi+\gamma_i\tau_{ij}\right)^{s_{ij}+\phi}} \prod_{k=1}^{L_{ij}} \left\{\gamma_i m \left(t_{ij}^k\right)^{m-1}\right\}^{\delta_{ij}^k} \tag{7}$$

Therefore, the logarithmic likelihood function with the observed data $\Xi = (\xi_1, ..., \xi_M)$ can be expressed as:

$$\ln L(\Xi, \theta) = \sum_{i=1}^{M} \sum_{j=1}^{Ni} \ln L_{ij}\left(\xi_{ij} : \theta_i\right) = N\phi \ln\phi - \sum_{i=1}^{M} \sum_{j=1}^{Ni} \left(s_{ij}+\phi\right) \ln\left(\phi+\gamma_i\tau_{ij}\right)$$

$$+ \sum_{i=1}^{M} \sum_{j=1}^{Ni} \sum_{k=0}^{s_{ij}-1} \ln(\phi+k) + \sum_{i=1}^{M} \sum_{j=1}^{Ni} \sum_{k=1}^{L_{ij}} \delta_{ij}^k \left\{\ln\gamma_i + \ln m + (m-1)\ln t_{ij}^k\right\} \tag{8}$$

However, each element of $\theta = (\theta_1, ..., \theta_{M+2})$ are expressed as $(\theta_1, ..., \theta_M) = (\gamma_1, ..., \gamma_M)$, $\theta_{M+1} = m$, $\theta_{M+2} = \phi$. Also, $N = \sum_{i=1}^{M} N_i$, and with the third item of the right side of equation (8), when s_{ij} =0, $\sum_{k=0}^{s_{ij}-1} \ln(\phi+k) = 0$ is defined. Also, when s_{ij} =1, $\sum_{k=0}^{s_{ij}-1} \ln(\phi+k) = \ln\phi$. Furthermore, the logarithmic likelihood function when using model 2 will be as follows.

$$\ln L(\Xi, \theta) = N\phi \ln\phi - \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(s_{ij}+\phi\right) \ln\left(\phi+\gamma\tau_{ij}\right)$$

$$+ \sum_{i=1}^{M} \sum_{j=1}^{N_i} \sum_{k=0}^{s_{ij}-1} \ln(\phi+k) + \sum_{i=1}^{M} \sum_{j=1}^{N_i} \sum_{k=1}^{L_{ij}} \delta_{ij}^k \left\{\ln\gamma + \ln m + (m-1)\ln t_{ij}^k\right\} \tag{9}$$

Also, the maximum likelihood estimator of the parameter value $\boldsymbol{\theta}$ which maximizes the logarithmic likelihood function (8) is given as $\hat{\theta} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_{M+2}\right)$ which simultaneously satisfies

$$\frac{\partial \ln L\left(\Xi, \hat{\theta}\right)}{\partial \theta_i} = 0, \quad \left(i = 1, \ldots, M+2\right) \tag{10}$$

Furthermore, the estimated value $\hat{\Sigma}\left(\hat{\theta}\right)$ of the asymptotic covariance matrix of the parameter can be expressed as follows.

$$\hat{\Sigma}\left(\hat{\theta}\right) = \left[\frac{\partial^2 \ln L\left(\Xi, \hat{\theta}\right)}{\partial \theta \partial \theta'}\right]^{-1} \tag{11}$$

However, the inverse matrix of the right side of the above equation is the inverse matrix of the Fisher information matrix after (M+2)x(M+2) with elements $\partial^2 \ln L\left(\hat{\theta}, \Xi\right) / \partial \theta_i \partial \theta_j$. The parameter's maximum likelihood estimator can be obtained by solving the non-linear simultaneous equation (10) of M+2 dimension. In this study, the maximum likelihood estimator is estimated by the Newton-Raphson method. If the maximum likelihood estimator $\hat{\theta}$ is obtained, the covariance matrix estimated value $\hat{\Sigma}\left(\hat{\theta}\right)$ is employed to estimate the t-test statistics.

Next, with the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of the parameter vector as postulate, the maximum likelihood estimator of the heterogeneity parameter ε_{ij} (i=1, …, M; j =1, …, N_i) is obtained. Here, the partial likelihood function is defined as follows.

$$L_{ij}^{\circ}\left(\xi_{ij}, \varepsilon_{ij} : \hat{\theta}_i\right) = \frac{\hat{\phi}^{\hat{\phi}}}{\Gamma\left(\hat{\phi}\right)} \prod_{k=1}^{L_{ij}} \left\{\hat{\gamma}_i \hat{m}\left(t_{ij}^k\right)^{\hat{m}-1}\right\}^{\delta_{ij}^k} \varepsilon_{ij}^{s_{ij}+\hat{\phi}-1} \exp\left\{-\left(\hat{\phi} + \hat{\gamma}_i \hat{\tau}_{ij}\right)\varepsilon_{ij}\right\} \tag{12}$$

However, $\hat{\tau}_{ij} = \sum_{k=1}^{L_{ij}} \left(t_{ij}^k\right)^{\hat{m}}$. At this time, the maximum likelihood estimator of the heterogeneity parameter ε_{ij} can be obtained as $\hat{\varepsilon}_{ij}$ that satisfies

$$\frac{\partial \ln L_{ij}^{\circ}\left(\xi_{ij}, \varepsilon_{ij} : \hat{\theta}_i\right)}{\partial \varepsilon_{ij}} = 0 \tag{13}$$

The maximum likelihood estimator of the heterogeneity parameter obtained in this way is the estimated value obtained with the parameter $\hat{\theta}_i = \left(\hat{\gamma}_i, \hat{m}, \hat{\phi}\right)$ as postulate. To express this explicitly, the solution of equation (13) is expressed as $\hat{\varepsilon}_{ij}\left(\hat{\theta}_i\right)$. Finally, from equation (12), (13), the following equation can be obtained.

$$\hat{\varepsilon}_{ij}\left(\hat{\theta}_i\right) = \frac{s_{ij} + \hat{\phi} - 1}{\hat{\phi} + \hat{\gamma}\hat{\tau}_{ij}} \tag{14}$$

EMPIRICAL STUDY

Overview of Empirical Study

Targeting the traffic control system managed by the Yokohama Branch of the Central Nippon Expressway Company Limited, the random proportional Weibull hazard model is estimated. The traffic control system of this branch is a system that has been sequentially renewed from the old system since 1990, and has been in operation continuously. The traffic control system comprises 9 central station systems (hereinafter, stations), manages the conditions of expressways, and provides appropriate real-time information to users. The operational

Table 2. Estimation Results

Parameter		Model 1	Model 2
γ	γ_1	1.251E-5 (-5.104E6)	5.834E-6 (-1.800E6)
	γ_2	1.631E-6 (-2.311E7)	
	γ_3	5.293E-6 (-9.182E6)	
m		2.174 (49.031)	2.216 (8.768)
ϕ		1.193 (2.182)	0.632 (2.801)
Logarithmic likelihood		-402.441	-407.405

condition of the traffic control system is also under real-time surveillance, and in the case of faults, the failure-generated component will be specified and the time and content of the failure will be recorded.

In this study, out of all the components that compose the traffic control system, the authors targeted the component group that, in case of a fault, has the possibility of developing into a serious fanctional failure of the whole system, and estimated the random proportional Weibull hazard model. After investigating the failure history database of the traffic control system and interviewing the system manager, the component group to target for model estimation is selected. As a result, the authors finally decided on three types, HDD (Hard disk drive), power supply and processing device. In the current traffic control system, there are 177 HDD, 306 power supply, and 180 processing device. From the same research and interview, HDD and processing device are used as three different devices: PC (used as monitors and terminals), servers (used as servers and processing equipment), and other devices (other uses not included in the former two). Furthermore, these components are used in 9 different stations, each type are categorized in 27 categories.

Estimation Results

In this study, 3 types of components: HDD, power supply and processing device are used. Therefore, in model 1 3 arrival density parameters γ_i (i=1,2,3) are introduced. Also, assuming the hazard rates of the PCs, servers and other devices in the 9 stations are heterogeneous, heterogeneity parameters ε_{ij} (i=1,2,3; j=1, …, N_i) are defined for each device. Therefore, the unknown parameters to be estimated are 1) in the case of model 1, arrival density parameter $\gamma_i(i=1,2,3)$, acceleration parameter m, heterogeneity dispersion parameter ϕ, ε_{ij}. 2) in the case of model 2, γ, m, ϕ and ε_{ij}. The random proportional Weibull hazard model estimated by the method proposed is shown in Table 2. Here, the results of the two models as probability distribution of heterogeneity parameters are compared. The value in parentheses shows the t-value, and the t- value of either parameter, as a result, the null hypothesis that they have no explanatory power for each explanatory variable model, are rejected at significance level 0.95. The model 1, in which the heterogeneity parameter average of each component was established, has a bigger logarithmic likelihood value. Moreover, by using model 1, it is possible to show more clearly the difference in hazard rates for each type of component. Therefore, the estimation results of model 1 will be employed for the following analysis. As shown in Table 2., the maximum likelihood estimator of the acceleration parameter of model 1 is $\hat{m}=2.174$. From Equation (2b), it can be seen that the survival probability of each type of component gradually decreases as the in-service time increases. Generally, if $\hat{m}=1.00$, it would be an accidemtal failure component whose failure rate does not depend on time, but all of the components used in this study have characteristics of wear-out failure component.

Testing of Proportional Hypothesis

With the random proportional Weibull hazard model, the proportional hypothesis that all types of components have the same acceleration parameter $\hat{m}$ is a prerequisite. However, depending on the type of component, there can be cases in which it is hard to think that the components have the same acceleration parameter. In these cases, it becomes necessary to divide the types of components into groups with the same acceleration parameter, and estimate the random proportional Weibull hazard model for each group. Here, the authors will first perform hypothesis evaluation to consider whether or not the proportional hypothesis works for failure processes of all types of components.

Now, the hypothesis testing model to consider the proportional hypothesis for type *i* (*i*=1, …, *M*) is,

$$\begin{cases} H_i^0 : m = \hat{m} \quad \text{and} \quad \hat{\gamma}_i, \hat{\phi} \\ H_i^1 : m \neq \hat{m} \quad \text{and} \quad \hat{\gamma}_i, \hat{\phi} \end{cases} \tag{15}$$

Here, the maximum estimator of parameter θ_i is expressed as $\hat{\theta}_i^0 = \left(\hat{\gamma}_i, \hat{m}, \hat{\phi}\right)$. The partial likelihood function of type i components with the maximum estimator $\hat{\theta}_i^0$ as postulate is defined as:

$$L_i\left(\xi_i : \hat{\theta}_i^0\right) = \frac{\hat{\phi}^{\hat{\phi}}}{\Gamma\left(\hat{\phi}\right)} \frac{\Gamma\left(s_{ij} + \hat{\phi}\right)}{\left(\hat{\phi} + \hat{\gamma}_i \hat{\tau}_{ij}\right)^{s_{ij} + \hat{\phi}}} \prod_{k=1}^{L_{ij}} \left\{\hat{\gamma}_i \hat{m}\left(t_{ij}^j\right)^{\hat{m}-1}\right\}^{\delta_{ij}^k} \tag{16}$$

Next, the maximum of the partial likelihood function (16) with parameter $\hat{\gamma}_i, \hat{\phi}$ as postulate is

$$L_i\left(\xi_i : \tilde{\theta}_i\right) = \max_m \left\{ \frac{\hat{\phi}^{\hat{\phi}}}{\left(\hat{\phi} + \hat{\gamma}_i \hat{\tau}_{ij}\right)^{s_{ij} + \hat{\phi}}} \frac{\Gamma\left(s_{ij} + \hat{\phi}\right)}{\Gamma\left(\hat{\phi}\right)} \prod_{k=1}^{L_{ij}} \left\{\hat{\gamma}_i m \left(t_{ij}^k\right)^{m-1}\right\}^{\delta_{ij}^k} \right\} \tag{17}$$

At this time, the likelihood-ratio test statistics for evaluating the hypothesis testing model (15) can be expressed as:

$$LR_i = 2\left\{\ln\left[L_i\left(\xi_i : \tilde{\theta}_i\right)\right] - \ln\left[L_i\left(\xi_i : \hat{\theta}_i^0\right)\right]\right\} \tag{18}$$

$\ln\left[L_i\left(\xi_i : \tilde{\theta}_i\right)\right]$ shows the partial likelihood when there is no limitation by the null hypothesis H_i^0, and $\ln\left[L_i\left(\xi_i : \hat{\theta}_i^0\right)\right]$ is the partial likelihood under limitation by the null hypothesis H_i^0. Because there is one parameter that is influenced by the limitation of the null hypothesis H_i^0, the degree of freedom of the likelihood-ratio test statistics is 1. Therefore, when the test statistics LR_i satisfy the rejected region $LR_i \geq \chi^2_{(100-\alpha)}(1)$, the null hypothesis H_0^i will be rejected at significance level α%. However, $\chi^2(1)$ is the probability variable subjected to χ^2 distribution of 1 degree of freedom, and $\chi^2_{(100-\alpha)}$ is the critical value of significance level α%. Therefore, $\mathrm{Prob}\left\{\chi^2(1) \geq \chi^2_{(100-\alpha)}(1)\right\} = 0.01 \times (100 - \alpha)$ is true. Using the proportional hypothesis testing model, let's consider whether the failure rates of all types of components with one random proportional Weibull hazard model can be expressed. When the proportional hypothesis is rejected, it becomes necessary to divide the types of components in several groups and estimate the random proportional Weibull hazard model for each group.

Step 1 Estimate the random proportional Weibull hazard model on a database that stores all types of components (hereinafter, original database)

Table 3. Likelihood-ratio Test Statistics

	LR_i
HDD (i=1)	2.9237E-2
Power-supply parts (i=2)	7.9481E-2
Processing parts (i=3)	1.7850E-5

Step 2 evaluate proportional hypothesis by likelihood-ratio test on each type of component. Create a derived database that stores the types of components rejected by the proportional hypothesis. Remove the samples in the derived database from the original database, and use the reduced database as the original database.
Step 3 Estimate the random proportional Weibull hazard model for the original database and derived database.
Step 4 On the random proportional Weibull hazard model estimated from the derived database, evaluate the proportional hypothesis on the component groups in the database. If necessary, create a second derived database. If the proportional hypothesis rejects, continue to divide the derived databases. When the proportional hypothesis does not reject, continue to Step 5.
Step 5 Estimate the maximum estimator $\hat{\varepsilon}_{ij}\left(\hat{\theta}_i\right)$ of the heterogeneity parameter for each type and device using equation (14).

This case study uses 3 types of components. When using model 1 for the probability distribution of the heterogeneity parameter, the likelihood-ratio test statistics $LR_i(i=1,2,3)$ are as shown in Table 3. Here, if the significance level is α =95%, then $\chi^2_{(100-\alpha)}(1)=3.84$, so the null hypothesis H_i^0 that all types have the same acceleration parameter $\hat{m}$ is not rejected. Therefore, the same acceleration parameter can be used for the 3 types of components.

Analysis Results

The random proportional Weibull hazard model is appllied to obtain each type's arrival density parameter γ_i (i=1,2,3), acceleration parameter m, and the maximum estimator of the heterogeneity variance ϕ. With the random proportion Weibull hazard model, the proportional hypothesis that all components possess the same acceleration parameter $\hat{m}$ is a prerequisite. With this prerequisite, the heterogeneity of Weibull hazard functions of each type of component can be expressed with γ_i, ϕ and heterogeneity parameter value ε_{ij} (i=1,2,3; j=1, …, 27). Also, in the previous section, the proportional hypothesis that all types of components have the same acceleration parameter $\hat{m}$ is evaluated. As a result it is confirmed that the null hypothesis H_i^0 of the hypothesis testing model (15) is not rejected. In this section the heterogeneity parameters of the hazard rates of components of each type and device are estimated.

The authors categorized the 3 types of components focused on in this study (HDD, power supply, processing device) by which of the 9 stations of the traffic control system they are installed in, and further subdivided them by the 3 types of usage, 1) PC (used as monitors, terminals), 2) Server (used as servers, processing equipment) and 3) Other devices (Used in other ways other than described in the former two). As shown in Table 4., this creates 27 categories for each type. Using equation (14), the maximum estimator of heterogeneity parameters for each category can be obtained. However, some categories (i,j) have no applicable components, and the number of heterogeneity parameters to be estimated are: 21 for HDD, 9 for power supply, and 26 for processing device. Table 4. shows each estimated

Table 4. Heterogeneity Parameters

		HDD	Power supply	Processing device
Station 1	(PC)	0.154 (1.536)	0.006 (2.401)	0.154 (1.398)
	(Server)	0.148 (1.744)	-	0.148 (1.484)
	(Others)	-	-	0.008 (3.694)
Station 2	(PC)	0.123 (2.973)	0.120 (2.586)	0.770 (8.94E-02)
	(Server)	2.208 (0.541)	-	0.125 (1.897)
	(Others)	0.161 (1.321)	-	0.008 (3.313)
Station 3	(PC)	0.146 (1.799)	0.007 (5.549)	0.146 (1.507)
	(Server)	0.669 (3.357)	-	0.008 (3.715)
	(Others)	0.133 (2.337)	-	0.860 (7.62E-02)
Station 4	(PC)	1.437 (5.38E-02)	0.004 (3.174)	0.688 (0.190)
	(Server)	0.768 (2.136)	-	0.674 (0.213)
	(Others)	0.006 (11.207)	-	0.833 (7.82E-02)
Station 5	(PC)	0.753 (0.479)	0.008 (1.983)	0.113 (2.170)
	(Server)	0.600 (1.500)	-	0.628 (0.303)
	(Others)	-	-	-
Station 6	(PC)	0.134 (2.310)	0.008 (1.954)	0.132 (1.752)
	(Server)	0.114 (3.365)	-	0.802 (6.24E-02)
	(Others)	-	-	0.142 (1.579)
Station 7	(PC)	0.147 (1.779)	0.007 (2.090)	0.147 (1.500)
	(Server)	1.304 (0.246)	-	0.136 (1.674)
	(Others)	-	-	0.009 (3.070)
Station 8	(PC)	5.400 (12.044)	1.360 (3.519)	0.481 (0.830)
	(Server)	1.833 (0.181)	-	1.508 (0.325)
	(Others)	-	-	0.581 (0.424)
Station 9	(PC)	0.844 (0.178)	0.632 (3.44E-02)	0.416 (1.260)
	(Server)	0.138 (2.140)	-	0.948 (3.43E-03)
	(Others)	-	-	0.123 (1.937)

heterogeneity parameters $\hat{\varepsilon}_{ij}(\hat{\theta})$ and the number of samples of each component of each type. The same table shows that the maximum estimators of heterogeneity parameters are distributed variously depending on the device, and it can be seen that in order to express the deterioration characteristic of components in an information system it is necessary to consider the heterogeneity of hazard rates. Also, because the component structure is extremely subdivided, it can be surmised that the estimation efficiency will fall if the different component characteristics with dummy variables are expressed. By using the maximum estimator of heterogeneity parameters, we can obtain a deterministic Weibull hazard model that expresses the deterioration characteristic of each device. However, as shown in Table 4., there are some device categories with a very small amount of samples. Therefore, there is the possibility of a problem occurring with the reliability of the Weibull hazard models obtained for each device category. In fact, of the heterogeneity parameters shown in Table 4, there are some for which the null hypothesis that it has no explanatory power for implementing the heterogeneity parameter could not be rejected, at significance level 95%. Therefore, by grouping the device categories, aggregative Weibull hazard models that express the average deterioration characteristic for each group can be obtained. Let's define the heterogeneity parameters for each type and device ε_{ij}, as aggregative average heterogeneity parameters for each type $E\varepsilon_i$ (i=1, …, M). The aggregative partial likelihood functions for all devices are defined as,

$$L_i^{\circ\circ}\left(\xi_i, E\varepsilon_i : \hat{\theta}_i\right) = \frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \prod_{j=1}^{N_j} k=1 \prod^{L_{ij}} \left\{\hat{\gamma}_i \hat{m}\left(t_{ij}^k\right)^{\hat{m}-1}\right\}^{\delta_{ij}^k} E\varepsilon_i^{s_{ij}+\hat{\phi}-1} \exp\left\{-\left(\hat{\phi}+\hat{\gamma}_i \hat{\tau}_{ij}\right) E\varepsilon_i\right\} \qquad (19)$$

The maximum estimator of the heterogeneity parameter $E\varepsilon_i$ satisfies

Table 5. Heterogeneity parameters

	HDD	Power supply	Processing device
$E\hat{\varepsilon}_i(\hat{\theta})$	0.923 (9.746)	0.205 (8.737)	0.431 (20.086)
$E\hat{\varepsilon}_{i1}(\hat{\theta})$	1.302 (0.022)	-	0.366 (8.344)
$E\hat{\varepsilon}_{i2}(\hat{\theta})$	0.900 (6.403)	-	0.527 (3.973)
$E\hat{\varepsilon}_{i3}(\hat{\theta})$	0.095 (14.552)	-	0.410 (8.243)

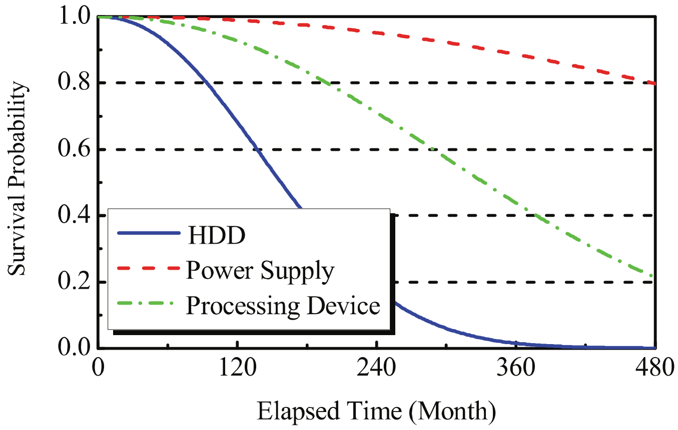


Figure 3. Survival Probabilities of Each Type

$$\frac{\partial \ln L_i^{\circ\circ}\left(\xi_{ij}, E\varepsilon_i : \hat{\theta}_i\right)}{\partial E\varepsilon_i} = 0 \tag{20}$$

And can be expressed as:

$$E\hat{\varepsilon}_i\left(\hat{\theta}_i\right) = \frac{\Sigma_{j=1}^{N_i} s_{ij} + \hat{\phi} - 1}{\Sigma_{j=1}^{N_i} \hat{\phi} + \hat{\gamma}\hat{\tau}_{ij}} \tag{21}$$

Similarly, the maximum estimators $E\hat{\varepsilon}_{il}\left(\hat{\theta}_i\right)$ of heterogeneity parameters $E\varepsilon_{il}$ (i,l=1,2,3) aggregated for each type and usage (PC, server, others) is,

$$E\hat{\varepsilon}_{il}\left(\hat{\theta}_i\right) = \frac{\Sigma_{j\in_l} s_{ij} + \hat{\phi} - 1}{\Sigma_{j\in_l} \hat{\phi} + \hat{\gamma}\hat{\tau}_{ij}} \tag{22}$$

However, l is the usage of the components, and l=1 means PC, l=2 means servers, and l=3 means others. Also, ω_l is the set of devices of the usage l. The maximum estimators of the average heterogeneity parameters aggregated with the above idea are shown in Table 5. If the average heterogeneity parameters aggregated by type are compared, $E\hat{\varepsilon}_1\left(\hat{\theta}\right) > E\hat{\varepsilon}_3\left(\hat{\theta}\right) > E\hat{\varepsilon}_2\left(\hat{\theta}\right)$ is true. Furthermore, the same relation applies to the arrival density parameter, so the hazard rate of HDD is highest, and that of power supply is the smallest. By obtaining the average heterogeneity parameter $E\hat{\varepsilon}i\left(\hat{\theta}\right)$, which is the heterogeneity parameters of each type of component aggregated, the Weibull hazard model that expresses the average deterioration characteristic of each type of component can be obtained. By using the average heterogeneity parameter $E\hat{\varepsilon}i\left(\hat{\theta}\right)$ (i=1,2,3) aggregated from each type, the average survival function of each type is obtained, and these are shown in Figure 3. Of all the samples, the shown survival probabilities are the relative ratio of the samples that survived in the targeted period. The lifespan of components is generally evaluated as service life. The same figure can be illustrated to show the service life according to the survival probability. Therefore, by using the survival probability as a management indicator, the service life can be evaluated by an

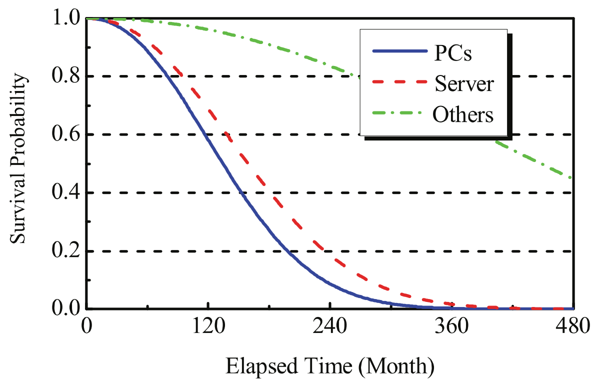


Figure 4. Survival Probabilities of HDD

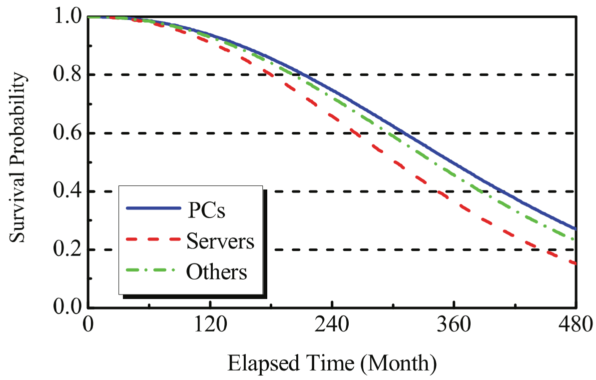


Figure 5. Survival Probabilities of Processing Device

arbitrary management indicator. However, the management indicator should be prepared with the importance of components in consideration. From Figure 4., it can be seen, for example, the use period (service life) at which the survival probability is 50% is 158 months for HDD, 804 months for power supply, and 332 months for processing device. Furthermore, the survival probability of power supply that are used for 120 months is 98.9%, and it is 95% at 240 months. For processing device, it is 92.7% at 120 months and 71.1% at 240 months. It can be seen that the failure rate increases, as the use time grows longer, for power supply and processing device as well as HDD. However, among the 3 types, HDD has the most rapid incline, while the power supply have the gentlest incline. Furthermore, the survival functions of HDD and processing device obtained by aggregating heterogeneity parameters by usage is shown in Figure 4. and Figure 5. Figure 4. shows the survival function of HDD. Even for the same HDD, the use environment is different depending on the usage. As seen from the figure, among the 3 different usages (PC, server, others), failures happen the earliest for PCs. PCs have the highest heterogeneity parameter, and the survival probability is 50% after using 135 months. On the other hand, devices of other uses have the smallest heterogeneity parameter, and the survival probability is 50% at 303 months. In Figure 5., of processing parts, the hazard rate when used as servers is the highest, with a survival probability of 50% at 303 months. The hazard rate is smallest for PCs, with a survival rate of 50% at 356 months.

CONCLUSIONS

In this study, a deterioration failure estimation model for components in information systems is proposed, aiming for asset management at the component level of large-scaled information systems which support infrastructures. The authors focus on the point that information systems are composed of many types of components, and point out that a fault analysis model that can express the heterogeneity of hazard rates of different types is necessary. To operationally express the heterogeneity of failure rates, the Weibull hazard model is employed as a base and a random proportional Weibull hazard model that expressed the proportional heterogeneity of hazard rates with a standard gamma distribution is formulated. Furthermore, through a case study using a traffic control system for expressways, the validity of the proposed hazard model is empirically verified. When applying the random proportional Weibull hazard model to asset management, there are some issues left to study in the future.

The first is that it is necessary to develop an asset management methodology for components, using the proposed hazard model. Especially, there are cases in which it is difficult to acquire the necessary equipment, if an information component goes out of production during the operation of an information system. In this case, substituting with a different component is expensive. Therefore, to deal with out of stock components, it is necessary to store replacement components. Or, to perform preventative maintenance of information equipment, it is necessary to logically decide on the renewal time of components. When performing asset management at the component level, the hazard model proposed in this study can be used. Secondly, fault analysis at the function level of information systems is necessary. With asset management at the function level, it is necessary to focus on the seriousness of the effect a failure in one component or component group can have to the functional level of the whole system, and it is necessary to consider the maintenance strategy of each component and component group. The authors are in the process of developing a methodology to analyze the system's dynamic failure characteristics, which expresses the failure process of each component using the hazard model proposed in this study, and expresses the impact of each failure on the function of the whole system using a fault-tree (Aoki, 2008). Thirdly, it is necessary to work on asset management at the system level of information systems. For this, it is necessary to simultaneously consider the technical obsolescence of the information system, delay in processing time, and the dynamic failure process of the information system, and to develop a real-time option model to determine the best timing for renewing the information system. Fourth, the proposed method is applied to an asset management of infrastructures, for instance pavement, since the effect of heterogeneity of individual section would be larger than that of information system.

ACKNOWLEDGEMENTS

The authors would like to express our gratitude to Traffic Control Team, Facilities Team, and Facilities Maintenance Team of the Traffic Control Center, Yokohama Branch, Central Nippon Expressway Company Limited, and Highway Facilities Engineering Department, Expressway Technology Center. In addition, part of this study (by Kiyoyuki Kaito) was conducted at the Frontier Research Base for Global Young Researchers, Graduate School of Engineering, Osaka University, supported by Special Coordination Funds for Promoting Science and Technology, Ministry of Education, Culture, Sports, Science and Technology.

REFERENCES

Aoki, K., Yamamoto, K. and Kobayashi, K. (2007). "Optimal Inspection and Replacement Policy using Stochastic Method for Deterioration Prediction", *The 11th World Conference on Transportation Research*, Berkeley, USA

Aoki, K., Yamamoto, K., Kaito, K. and Kobayashi, K. (2008). "Dynamic Fault Analysis of Large-Scaled Information Systems with Reference to Component's Deterioration", *10th International Conference on Application of Advanced Technologies in Transportation*, Greece, Athens (abstract accepted).

Lancaster, T. (1990). "The Econometric Analysis of Transition Data", *Cambridge University Press*.

Maher, M.J. and Summersgill, I. (1996). "A comparative methodology for the fitting predictive accident models", *Accident Analysis & Prevention*, Vol.28, pp.281-296.

Tsuda, Y., Kaito, K., Aoki, K. and Kobayashi, K. (2006). "Estimating Markovian Transition Probabilities for Bridge Deterioration Forecasting", *Journal of Structural Eng./Earthquake Eng., JSCE*, Vol.23, No.2, pp.241s-256s.