Chapter 2

Advanced Failure Intensity Models

2.1 Introduction

Advanced failure intensity models are in this thesis defined as mathematical representations of failure processes that require more than standard distributions or 2-parameter counting process models to capture their characteristics. This chapter deals with advanced failure intensity models found in the literature.

Chapter 2 starts off with a discussion of the concept of *intensity* with specific reference to non-repairable and repairable situations. The importance of the difference between these situations cannot be overemphasized even though it is frequently ignored in statistical failure analysis. A clear notation with regards to intensities is defined in Section 2.2 and used throughout this thesis. Deviations from the notation are explicitly indicated. Different model classes are identified and relevant models are discussed. Some acclaimed applications of advanced failure intensity concepts are also considered . For most models, the likelihood or partial likelihood are derived or presented without describing the estimation of regression parameters. Parameter estimation techniques are considered in Chapter 3.

The chapter ends with a summary of the advantages and disadvantages of the models considered.

2.2 Intensity Concepts

The concept of intensity was introduced briefly in Section 1.2.3.2. In this section, the concept is explained in detail since all reliability models discussed in this chapter strive to represent the intensity of a certain failure process. It is assumed throughout the thesis that all failure processes considered are orderly, i.e. simultaneous failures cannot occur on the same item.

This is a reasonable assumption according to most authors, e.g. Hokstad (1997) and Lawless (1987), and not much generality is sacrificed.

Let $N(t)$ denote the number of failures an item has experienced in the interval $(0, t]$. The *unconditional* intensity (i.e., the rate of failure events) of the process at any instant in time, t, is then given by

$$\begin{aligned} \iota_u(t) &= \lim_{\Delta t \to 0} \frac{\Pr[\text{Failure occurs in } [t, t+\Delta t)]}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{E[\Delta N(t)]}{\Delta t} \end{aligned} \tag{2.1}$$

where $\Delta N(t)$ represents the increment $N(t+\Delta t) - N(t)$. Because it is assumed that the process is orderly, the following basic relation for counting processes applies:

$$\iota_u(t) = \frac{\mathrm{d}M(t)}{\mathrm{d}t} \tag{2.2}$$

with $M(t) = E[N(t)]$. The time derivative of the expected number of failures as in (2.2), is referred to as the rate of occurrence of failure (ROCOF) and will often be denoted by $\rho(t)$ (instead of $\iota_u(t)$), for convenience. From (2.2) it follows directly that the cumulative number of failures up to time t is equal to the cumulative unconditional intensity, i.e.

$$M_u(t) = E[N(t)] = \int_0^t \iota_u(u)\mathrm{d}u \tag{2.3}$$

Additional information about the failure process is often recorded with the times to failure. The additional information is referred to as the *history*, H_t, or *filtration* of the process. History is recorded in the form of covariates and could be any quantification of an influence on the failure process. From Martingale theory (see Hokstad (1997)) it follows that H_t is the σ-algebra generated by $N(s)$, $s \leq t$, starting from a probability space (Ω, H_t, P) that defines the stochastic process, $N(t) = N(t, \omega)$, with $\omega \in \Omega$. Hence it is possible to define the *full intensity* (also referred to simply as *intensity* or *conditional intensity*), $\iota(t, H_t) = \iota(t|H_t)$, which is the conditional rate of occurrence of events, given the state of H_t. Thus, $\iota(t, H_t)\Delta t$ is the probability of an event to occur in $[t, t+\Delta t)$, i.e.

$$\iota(t, H_t) = \lim_{\Delta t \to 0} \frac{E[\Delta N(t)|H_t)}{\Delta t} \tag{2.4}$$

The complete intensity as defined in (2.4) provides a general framework for modeling failure event processes because the effect of maintenance activities can be recorded in H_t. Conventional failure process modeling concepts such as the FOM and ROCOF are also special cases of (2.4).

Similar to (2.3), it is possible to define a cumulative intensity process, i.e.

$$M(t, H_t) = \int_0^t \iota(u, H_t)\mathrm{d}u \tag{2.5}$$

where $M(t, H_t)$ is the *compensator* in Martingale theory. Both $\iota(t, H_t)$ and $M(t, H_t)$ are denoted as *predictable* which means that for a given H_t, the values of $\iota(t, H_t)$ and $M(t, H_t)$ are known but the value of $N(t)^*$ not yet.

It is important to note that $\iota_u(t)$ is a mean function of $\iota(t, H_t)$, averaged over all possible sample paths. Suppose $N(t, \omega)$ is a specific realization of the process of $N(t)$ where $\omega \in \Omega$ in the probability space (Ω, H_t, P). Here, N is not only a function of ω for a fixed value of t but also a function of t for a fixed ω (called the sample path of N). Taking the the mean over the sample space, Ω, yields

$$E[\Delta N(t)] = \int_{\Omega} \mathrm{d}N(t, \omega)\mathrm{d}P(\omega) = i_u(t) \tag{2.6}$$

Similarly, $E[\Delta N(t)|H_t]$, and thus $\iota(t, H_t)$, is found as the conditional mean.

The last intensity concept to define is that of *average intensity*. Average intensity is simply the average of $M_u(t)$ or $M(t, H_t)$ over an interval $[0, \tau]$, i.e. $\iota_{u_\tau} = M_u(\tau)/\tau$ or $\iota_\tau = M(\tau)/\tau$. The concept of average intensity is not encountered frequently in the literature but is not without interest. Bodsberg and Hokstad (1995) have shown that the average intensity concept is very useful in modeling dormant failures.

Table 2.1: Summary of failure intensity concepts

	Failure Intensity Concept		
	Intensity	**Mean Intensity**	**Average Intensity**
Alternative term	Conditional intensity	Unconditional intensity	-
Symbol	ι	ι_u	ι_τ
Definition	$\lim_{\Delta t \to 0} \frac{E[\Delta N(t)\|H_t)}{\Delta t}$	$\lim_{\Delta t \to 0} \frac{E[\Delta N(t)]}{\Delta t}$	$t^{-1} \cdot E[N(t)]$
Non-repairable case	$h_X(x)$, truncated at time of failure	$f_X(x)$	$x^{-1} \cdot (1 - R_X(x))$
Repairable case	A sequence of truncated FOMs (defined in local time)	ROCOF or $v(t)$	Average ROCOF, i.e. AROCOF

In Table 2.1, a concise summary (adapted from Hokstad (1997)) of the failure intensity concepts discussed in this section is presented. Note that local time, denoted by x, is used as time scale for the non-repairable case, consistent with the terminology introduced in Section 1.2.1.

*$N(t)$ has right continuous sample paths.

2.3 Literature survey on advanced failure intensity models

There are countless publications on advanced failure intensity models attempting to represent the intensity concepts outlined in Table 2.1 as part of practical statistical failure analysis exercises. Most of these publications consider variations on a small number of fundamentally different approaches. The fundamentally different approaches are referred to as model classes and are listed below:

(i) Multiplicative intensity models
(ii) Additive intensity models
(iii) Models with mixed or modified time scales
(iv) Marginal regression analysis
(v) Competing risks
(vi) Frailty or mixture models

These model classes are discussed in Section 2.3. Publications that consider combinations of two or more model classes are discussed as part of the model class where it makes the most significant contribution. At the end of this section, noteworthy extensions of the listed model classes are also discussed.

2.3.1 Multiplicative Intensity Models

Multiplicative intensity models represent the intensity of a failure process as the product of a baseline intensity, that is a function of time only, and a functional term, that may be a function of both time and covariates. Covariates are allowed to be time-independent or time-dependent.

2.3.1.1 Proportional Hazards Model (PHM)

Survival data analysis underwent a revolution with the introduction of the PHM by Cox (1972). The model was originally intended for biomedical applications but was soon applied in reliability engineering. As the name implies, this model represents the FOM, i.e. the failure intensity of non-repairable items, as a proportion of different FOMs.

The PHM is constructed as the product of a totally arbitrary and unspecified baseline FOM, $h_0(x)$, and a functional term $\lambda(x, \boldsymbol{z})$, where $\boldsymbol{z}$'s dependence on time is not important, i.e. †

$$h(x, \boldsymbol{z}) = h_0(x) \cdot \lambda(x, \boldsymbol{z}(x)) \tag{2.7}$$

†The subscript x is dropped here for notational convenience.

There are several possible forms for the functional term. Some are: the exponential form, $\exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}(x))$; the logarithmic form, $\log(1+\exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}(x)))$; the inverse linear form, $1/(1+\boldsymbol{\gamma}\cdot\boldsymbol{z}(x))$; or the linear form, $1+\boldsymbol{\gamma}\cdot\boldsymbol{z}(x)$, where $\boldsymbol{\gamma}$ is a vector of regression coefficients associated with a particular data set. The exponential form of the functional term is used most often in reliability applications and results in the following PHM:

$$h(x,\boldsymbol{z}) = h_0(x)\cdot\exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}(x)) \tag{2.8}$$

The model assumes the following:

(i) Event times are IID.
(ii) All influential variables are included in the model.
(iii) The ratio of any two hazard rates as determined by any two sets of time-independent covariates $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ associated with a particular item has to be constant with respect to time, i.e. $h_X(x,\boldsymbol{z}_1) \propto h_X(x,\boldsymbol{z}_2)$. (This assumption is not valid for time-dependent covariates).

The biggest advantage of the PHM, as defined in (2.8) in its semi-parametric form, is that no assumption needs to be made about the baseline FOM when fitting the model. This is a result of partial likelihood theory developed by Cox (1975). Kalbfleisch and Prentice (1980) explain partial likelihood in detail. Partial likelihood only yields relative risks but can be very useful in gross analyses.

Suppose m items are under observation and n events have occurred up to time x. Let $\mathbb{F}(x_i)$ be a risk set of the events up to time x_i and let l be the number of events yet to occur. The partial likelihood of (2.8) is then given by

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}_i)}{\sum\limits_{l\in\mathbb{F}(x_i)} \exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}_l)} \tag{2.9}$$

In the case where relatively few ties, d_i, are present, the following relation holds:

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}_i)}{\left[\sum\limits_{l\in\mathbb{F}(x_i)} \exp(\boldsymbol{\gamma}\cdot\boldsymbol{z}_l)\right]^{d_i}} \tag{2.10}$$

It is also possible to stratify the PHM into different strata, i.e.

$$h(x,\boldsymbol{z}) = h_{0_j}(x)\cdot\exp(\boldsymbol{\gamma}_j\cdot\boldsymbol{z}(x)) \tag{2.11}$$

with partial likelihood given by,

$$L(\boldsymbol{\gamma}) = \prod_{j=1}^{r}\prod_{i=1}^{k_j} \frac{\exp(\boldsymbol{\gamma}_j\cdot\boldsymbol{z}_{ij})}{\sum\limits_{l,j\in\mathbb{F}(x_{ij})} \exp(\boldsymbol{\gamma}_j\cdot\boldsymbol{z}_{lj})} \tag{2.12}$$

where r denotes the number of strata and k_j is the number of events in the j^{th} stratum. Ascher, Kobbacy, and Percy (1997) applied the stratified PHM successfully.

If absolute risks are required, a fully parameterized PHM is required. A distribution often used to perform the parameterization is the Weibull distribution because of its flexibility. Substitution of the Weibull distribution in (2.8) yields

$$h(x, \boldsymbol{z}) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} \cdot \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}(x)) \tag{2.13}$$

where β and η are the shape and scale parameters of the Weibull distribution respectively.

The parameters PHM in (2.13) can be calculated by constructing the full likelihood as,

$$\begin{aligned} L(\beta, \eta, \boldsymbol{\gamma}, \boldsymbol{z}) &= \prod_{i=1}^{n} h(x_i, \boldsymbol{z}) \cdot \exp\left(-\int_0^{x_i} h(x, \boldsymbol{z})\mathrm{d}x\right) \\ &= \prod_{i=1}^{n} \frac{\beta}{\eta} \cdot \left(\frac{x_i}{\eta}\right)^{\beta-1} \cdot e^{\boldsymbol{\gamma}\cdot\boldsymbol{z}_i} \cdot \exp\left(-\int_0^{x_i} \frac{\beta}{\eta} \cdot \left(\frac{x}{\eta}\right)^{\beta-1} \cdot e^{\boldsymbol{\gamma}\cdot\boldsymbol{z}(x)}\,\mathrm{d}x\right) \end{aligned} \tag{2.14}$$

The solution of (2.14) is complex if $\boldsymbol{z}$ is dependent on time. Press et al. (1993) discuss some numerical techniques with which an economic solution can be obtained.

It is also possible to stratify the fully parametric PHM. Usually, either the baseline FOM or the regression coefficients are stratified, not both. This is done to limit the number of parameters in the model and to obtain synergy amongst different strata.

A useful extension of the PHM is Aalen's Regression Model as discussed in Aalen (1980) and Aalen (1989). This model can be used to test time dependence of covariates in the PHM and adds significant value to PHM analysis. In this model, the vector $\boldsymbol{h}(x; \boldsymbol{z})$ of the FOMs $h_j(x; \boldsymbol{z})$ for $j = 1, 2, .., n$, is given by:

$$\boldsymbol{h}(x; \boldsymbol{z}) = \boldsymbol{Y}(x) \cdot \boldsymbol{\alpha}(x) \tag{2.15}$$

Here $\boldsymbol{Y}(x)$ is an $n \times (q+1)$ matrix whose rows at time x_j consist of those vectors,

$$\boldsymbol{z}^j = \left[1, z_1^j(x), ..., z_q^j(x)\right] \tag{2.16}$$

where $z_i^j(x)$, $i = 1, 2, ..., q$ are covariate values, corresponding to those failure times that have not occurred up to time x_j. In the vector,

$$\boldsymbol{\alpha}(x) = [\alpha_0(x), \alpha_1(x), ..., \alpha_q(x)] \tag{2.17}$$

$\alpha_0(x)$ is the baseline parameter function, while $\alpha_i(x)$, $i = 1, 2, ..., q$ are called regression functions, defining the effects of covariates. The effect of a covariate is represented by the cumulative regression function, $\boldsymbol{A}(x)$, defined as:

$$\boldsymbol{A}_i(x) = \int_0^x \alpha_i(s)\mathrm{d}s \tag{2.18}$$

for $i = 0, 1, ..., q$. To study the time-varying effect of the i^{th} covariate, an estimate of the i^{th} cumulative regression function should be plotted against the failure times. There are 4 possible outcomes:

(i) Straight line with an incline m. The effect is independent of time.
(ii) Constant line at value y. Indicates no effect at all.
(iii) Increasing at a decreasing rate. Indicates a decreasing effect over time.
(iv) Increasing at an increasing rate. Indicates an increasing effect over time.

Aalen's approach is particularly useful in analyzing condition monitoring data since condition monitoring data is almost always time-dependent.

2.3.1.2 Proportional Mean Intensity Models (PMIM)

Proportional Mean Intensity Models or Proportional ROCOF Models are constructed by the product of a baseline ROCOF multiplied with a functional term, dependent on time and covariates. PMIMs are very similar to PHMs as far as construction and estimation is concerned but they are based on fundamentally different representations of the intensity of failure processes. In the literature, the terminology for these concepts are often inconsistent, e.g. Kumar (1996) investigated the use of "Proportional Hazards Modeling" on repairable systems while he was actually using PMIMs.

Suppose the PMIM is constructed as the product of a baseline ROCOF, $\iota_{u_0}(t)$, and a functional term $\lambda(t, \boldsymbol{z}(t))$, where $\boldsymbol{z}$ may or may not depend on time, i.e.

$$\iota_u(t, \boldsymbol{z}) = \iota_{u_0}(t) \cdot \lambda(t, \boldsymbol{z}(t)) \tag{2.19}$$

As before, it is possible to estimate the semi-parametric model in (2.19) without making any assumptions about $\iota_{u_0}(t)$ by using partial likelihood theory. Let m denote the number of items under observation and let n represent the total number of failures that have occurred. Let $\mathbb{F}(t_i)$ be the risk set of the failure events and let l represent the number of events yet to occur at time. The partial likelihood is then given by

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_i)}{\sum\limits_{l \in \mathbb{F}(t_i)} \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_l)} \tag{2.20}$$

If the number of ties, d_i, in the data set is small, the following relation holds

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_i)}{\left[\sum\limits_{l \in \mathbb{F}(t_i)} \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_l)\right]^{d_i}} \tag{2.21}$$

If the PMIM is stratified into r strata, i.e. $\iota_{u0j}(t,\boldsymbol{z}) = \iota_{u0j}(t)\exp(\boldsymbol{\gamma}_j \cdot \boldsymbol{z}(t))$, the partial likelihood becomes,

$$L(\boldsymbol{\gamma}) = \prod_{j=1}^{r}\prod_{i=1}^{k_j} \frac{\exp(\boldsymbol{\gamma}_j \cdot \boldsymbol{z}_{ij})}{\sum_{l,j \in F(t_{ij})} \exp(\boldsymbol{\gamma}_j \cdot \boldsymbol{z}_{lj})} \tag{2.22}$$

where r is the number of strata and k_j is the number of of events in the j^{th} stratum.

If an absolute mean intensity is required, the PMIM can be parameterized. The log-linear representation of a NHPP is often used to perform the parameterization, i.e. $\iota_{u_0}(t) = \exp(\alpha_0 + \alpha_1 \cdot t)$. The full likelihood becomes,

$$\begin{aligned} L(\alpha_0, \alpha_1, \boldsymbol{\gamma}, \boldsymbol{z}) &= \prod_{i=1}^{n} \iota_{u_0}(t) \cdot \exp\left(-\int_0^{Tn} \iota_u(t, \boldsymbol{z}(t))\mathrm{d}t\right) \\ &= \prod_{i=1}^{n} \left(e^{\alpha_0 + \alpha_1 \cdot T_i} \cdot e^{\boldsymbol{\gamma} \cdot \boldsymbol{z}(T_i)}\right) \cdot \exp\left(-\int_0^{Tn} e^{\alpha_0 + \alpha_1 \cdot t} \cdot e^{\boldsymbol{\gamma} \cdot \boldsymbol{z}(t)} \mathrm{d}t\right) \end{aligned} \tag{2.23}$$

As in the case of the parametric PHM, it is difficult to maximize (2.23) if the covariates are time-dependent.

If the fully parametric PMIM is stratified, usually, either the baseline ROCOF or the regression coefficients are stratified, not both. This is done to limit the number of parameters in the model and to obtain synergy amongst different strata.

2.3.1.3 Proportional Odds Model (POM)

The proportional odds model originated from epidemiological studies and was introduced by Bennet (1983) for use in biomedicine. This model is structurally similar to the PHM, but not a direct extension. It models the odds of an event occurring and unlike the PHM, the effect of covariates in the POM model diminishes as time approaches infinity. This diminishing property of the covariates means that the model is suitable for situations where an item adjusts to factors imposed on it or the factors only operate in early stages.

For this model the *odds* of a failure occurring is defined in terms of the survivor function as,

$$\frac{F_X(x)}{R_X(x)} = \frac{1 - R_X(x)}{R_X(x)} \tag{2.24}$$

This definition of odds is used to introduce the POM:

$$\frac{1 - R(x, \boldsymbol{z})}{R(x, \boldsymbol{z})} = \psi \cdot \frac{1 - R_X(x)}{R_X(x)} \tag{2.25}$$

Equation (2.25) states that the odds for a failure to occur under the influence of covariates are ψ times higher than the odds of a failure without the effects of covariates. If ψ increases, so does the probability of a shorter life time. Differentiation of (2.25) with respect to time leads to,

$$\frac{h(x,\boldsymbol{z})}{R(x,\boldsymbol{z})} = \psi \cdot \frac{h_X(x)}{R_X(x)} \tag{2.26}$$

after using the coefficient rule. By rearranging the terms in (2.26) and re-using (2.25), a FOM ratio can be obtained:

$$\frac{h(x,\boldsymbol{z})}{h_X(x)} = \psi \cdot \frac{R(x,\boldsymbol{z})}{R_X(x)} = \frac{1-R(x,\boldsymbol{z})}{1-R_X(x)} \tag{2.27}$$

Inspection shows that $\psi|_{x=0} = \psi$ and $\psi|_{x=\infty} = 1$, from there the diminishing effect of the covariates.

Bennet (1983) derives the full likelihood for the model in his original paper to estimate the model parameters. Research done by Shen (1998) provides more efficient estimation methods and methods to enable the model to handle suspended observations.

A special case of the POM arise when it assumed that event times are distributed according to a log-logistic distribution. Kalbfleisch and Prentice (1980) describe this special case in detail. The FOM of an item with event times following a log-logistic distribution is given by:

$$h(x;\boldsymbol{z}) = \frac{\delta}{x \cdot (1 + x^{-\delta} \cdot \exp(-\boldsymbol{\gamma} \cdot \boldsymbol{z}(x)))} \tag{2.28}$$

where δ is a measure of precision. The FOM is assumed to be increasing first and then decreasing with a change at time

$$x = \{(1-\delta)\exp(-\boldsymbol{\gamma} \cdot \boldsymbol{z}(x))\}^{1/\delta} \tag{2.29}$$

If $x \to \infty$, $x^{-\delta} \cdot \exp(-\boldsymbol{\gamma} \cdot \boldsymbol{z}(x)) \to 0$ (see (B.4)) and subsequently covariates will influence the FOM less and less as the item ages.

2.3.2 Additive Intensity Models (AIMs)

Additive Intensity Models represent the intensity of a failure process as the sum of a baseline intensity and a functional term containing covariates. Pijnenburg (1991) deals with AIMs in completely general terms. Newby (1993) compare this type of model, for the case where the FOM is used as intensity, to various other regression models. Authors often refer to AIMs incorrectly as Additive *Hazard* Models (AHM) in reliability modeling literature. This section describes AIMs in general terms.

Suppose two items are in series, S_1 and S_2. Suppose S_1 represents a repairable system and S_2 is an item representing the influence of covariates. Let T_i be be the time at which the i^{th} system failure occurs and X_i the system's i^{th} interarrival time, i.e. $X_i = T_i - T_{i-1}$. The system is supposed to have a survival time $X_{1,1}$ and FOM $h_1(\cdot)$ and the item representing the covariates has a survival time of $X_{2,1}$ with a FOM, $\lambda(t_0, \boldsymbol{z}_0)$. For the moment $\lambda(t_0, \boldsymbol{z}_0)$ is defined as constant in-between interarrival times, i.e. constant covariates, but variable over successive lifetimes, i.e. dependent on time.

After the first failure at system level at time, T_1, i.e. $T_1 = X_1 = \min(X_{1,1}, X_{2,1})$ both components are replaced, such that:

(i) S_1 is renewed by an identical component, also called S_1, with lifetime $X_{1,2}$ and FOM $h_1(\cdot)$.

(ii) S_2 is replaced by a component with lifetime $X_{2,2}$ and FOM $\lambda(t_1, \boldsymbol{z}_1)$.

In general terms it means that after the i^{th} failure on system level at time T_i, i.e. $T_i = T_{i-1} + \min(X_{1,i}, X_{2,i})$:

(i) S_1 is replaced by an identical new component with lifetime $X_{1,i+1}$ and FOM $h_1(\cdot)$. The lifetimes $X_{1,k}$ are assumed to be IID.

(ii) S_2 is replaced by a component with a lifetime $X_{2,i+1}$ and FOM $\lambda(t_i, \boldsymbol{z}_i)$. The lifetimes $X_{2,k}$ are assumed to be statistically independent.

It is also assumed that the survival times $X_{1,k}$ and $X_{2,k}$ are mutually independent.

For the various FOMs, $\lambda(t_i, \boldsymbol{z}_i)$, it is assumed that the covariates, $\boldsymbol{z}_i$, are constant in $[T_i, T_{i+1})$ but may change for different lifetimes. Higher covariate values, generally represent more severe environmental stresses and $x_{2,i} = \min(x_{1,i}, x_{2,i})$ should be interpreted as a system failure due to these higher environmental stresses. Pijnenburg (1991) suggests a few forms for $\lambda(t_i, \boldsymbol{z}_i)$. The simplest form for $\lambda(t_i, \boldsymbol{z}_i)$ is a linear function,

$$\lambda(t, \boldsymbol{z}) = \boldsymbol{\gamma} \cdot \boldsymbol{z} = \sum_{i=1}^{p} \gamma_i \cdot z_i \tag{2.30}$$

for p covariates. A linear term can be included in (2.30) by simply specifying $z_1 = 1$. In the case where higher order terms are present in the polynomial, $\lambda(t_i, \boldsymbol{z}_i)$ can be specified as,

$$\lambda(t, \boldsymbol{z}) = \sum_{i=1}^{p} \sum_{j=0}^{m} \gamma_{ij} \cdot z_i^j \tag{2.31}$$

If covariates appear to interact, $\lambda(t_i, \boldsymbol{z}_i)$ can be chosen as,

$$\lambda(t, \boldsymbol{z}) = \boldsymbol{\gamma}_0 + \sum_{i=1}^{p} \sum_{j=1}^{i-1} \gamma_{ij} \cdot z_i \cdot z_j \tag{2.32}$$

A suitable form to handle both higher order terms and interaction can be,

$$\lambda(t, \boldsymbol{z}) = \sum_{i=1}^{p} \sum_{j=1}^{i} \sum_{k=0}^{r} \sum_{l=0}^{r} \gamma_{ijkl} \cdot z_i^k \cdot z_j^l \tag{2.33}$$

Data limitations often cause that only (2.30) is practical.

Following the argument above, the AIM is completely generalized by allowing covariates to dependent on time, i.e.

$$\iota(t, \boldsymbol{z}) = \iota_0(t) + \lambda(t, \boldsymbol{z}(t)) \tag{2.34}$$

The AIM can be stratified and parameterized in the same manner as the PHM and PMIM. Pijnenburg (1991) also allows the AIM to have modified time scales. These extensions are discussed in sections to follow. Crowder, Kimber, Smith, and Sweeting (1991) and Newby (1993) derive the full likelihood to fit AIMs[‡].

2.3.3 Models with mixed or modified time scales

Modified or mixed time scales in intensity models can be interpreted as an additional covariate in data sets to provide more flexibility. Modified time scales increase or decrease the modeled intensity of a failure process by either accelerating or decelerating the actual age of an item. Mixed time scale models incorporate local and global time in the same model to utilize the advantages of both long and short term history. Newby (1993) refers to the result of these concepts as the *virtual age* of an item since the actual survival time differs from the survival time used in models.

2.3.3.1 The Prentice Williams Peterson (PWP) model

Prentice, Williams, and Peterson (1981) published the so-called PWP model after research done by Williams (1981). This model is generally considered as the most significant extension of the PHM by Cox (1972) according to Ascher and Feingold (1984). Two versions of the PWP model were proposed, both of the stratified Proportional Hazards type, which means this discussion would also be applicable in Section 2.3.1 where multiplicative models were considered but it is believed that this model made a more significant contribution to models with modified or mixed time scales.

The PWP model is specifically directed towards the analysis of situations where only a small number of observations is available on an item but where a large number of items is studied.

[‡]Partial likelihood can not be used because of the summation of terms in the model and relative risks are thus not possible

Specific items are also allowed to experience multiple failures. This makes the PWP very attractive in reliability modeling where data sets are often limited in size.

The model is constructed as follows. Let $\boldsymbol{z} = [z_1(t), ..., z_p(t)]$ denote a vector of covariates of a specific item, part of the covariate process, $Z(t)$. Also, let $N(t)$, denote the counting process of the the number of failures, $n(t)$, on an item up to time t. The counting process, $N(t)$, is equivalent to the random failure times $T_1 < .... < T_{n(t)}$ in $[0,t)$. Prentice, Williams and Peterson then define the intensity of a failure process as,

$$\iota(t|N(s), Z(s), s \leq t) = \lim_{\Delta t \to 0} \frac{\Pr[t \leq T_{n(t)+1} \leq t + \Delta t | N(s), Z(s), s \leq t]}{\Delta t} \tag{2.35}$$

Some special cases of (2.35), in the absence of covariates, are:

(i) $\iota(t|N(t)\} = \iota_u(t)$ for some $\iota_u(\cdot) \geq 0$ is the unconditional intensity function of a NHPP.
(ii) $\iota(t|N(t)) = n(t)\iota_u(t)$ specifies a nonhomogeneous pure birth process.
(iii) $\iota(t|N(t)) = \iota_k(t - t_k)$ for an arbitrary $\iota_k(\cdot) \geq 0$ $(k = n(t) + 1 = 1, 2, ...)$ gives a semi-Markov process.
(iv) A further restriction on the semi-Markov process that $\iota_k(\cdot) = \iota_0(\cdot)$ for all k gives an ordinary renewal process.

Prentice, Williams and Peterson suggest two models based on (2.35), both of a stratified Proportional Hazards type,

$$\text{PWP Model 1}: \ \iota(t|N(t), Z(t)) = \iota_{0_s}(t) \cdot \exp(\boldsymbol{\gamma}_s \cdot \boldsymbol{z}(t)) \tag{2.36}$$

$$\text{PWP Model 2}: \ \iota(t|N(t), Z(t)) = \iota_{0_s}(t - t_{n(t)}) \cdot \exp(\boldsymbol{\gamma}_s \cdot \boldsymbol{z}(t)) \tag{2.37}$$

The stratification variable $s = s\{N(t), Z(t), t\}$ may change as a function of time for a given item, e.g. $s = n(t) + 1$ and the subject moves to stratum k immediately following its $(k-1)^{\text{th}}$ failure and remains there until the k^{th} failure. More refined stratum conditions can easily be constructed. For Model 1, it is possible to define $s = 2 \cdot n(t) + \Lambda\{N(t)\}$, where $\Lambda\{N(t)\} = 1$ if the time since the last failure, $t - t_{n(t)}$, is less than some specified value and $\Lambda\{N(t)\} = 2$, otherwise. In the case of Model 2, it is possible to define $s = 2 \cdot n(t) + \Delta(t)$, where $\Delta(t) = 1$ if t is less than some value and $\Delta(t) = 2$, otherwise.

The PWP formulations differ from Andersen (1985) in two aspects: (a) the risk sets of the $(k+1)^{\text{th}}$ recurrences are restricted to the individuals who have experienced the first k recurrences; and (b) the underlying intensity functions and regression parameters are allowed to vary amongst distinct recurrences. Gail, Santner, and Brown (1980) published a two-sample special case of Model 2 with strata defined at least as finely as $s = n(t) + 1$. Clifton and Crowley (1978) considered a special case of Model 1 without covariates and with $s = 1$ if $n(t) = 0$ and $s = 2$ if $n(t) \geq 1$.

Partial likelihood can be used to estimate relative risks with both Models 1 and 2. For Model 1 the partial likelihood is,

$$L(\gamma) = \prod_{s\geq 1} \prod_{i=1}^{d_s} \frac{\exp(\boldsymbol{\gamma}_s \cdot \boldsymbol{z}_{si}(t_{si}))}{\sum\limits_{l\in \mathbb{R}(t_{si},s)} \exp(\boldsymbol{\gamma}_s \cdot \boldsymbol{z}_l(t_{si}))} \tag{2.38}$$

where t_{si} denotes the failure time of item i in stratum s, $\boldsymbol{z}_{si}(t_{si})$ refers to the covariate vector of item i at time t_{si} and d_s denotes the total number of events in stratum s.

The partial likelihood for Model 2 is,

$$L(\gamma) = \prod_{s\geq 1} \prod_{i=1}^{d_s} \frac{\exp(\boldsymbol{\gamma}_s \cdot \boldsymbol{z}_{si}(t_{si}))}{\sum\limits_{l\in R(u_{si},s)} \exp(\boldsymbol{\gamma}_s \cdot \boldsymbol{z}_l(\zeta_l + u_{si}))} \tag{2.39}$$

where u_{si} are the interarrival times of the different items in various stratums and ζ_l is the last failure time on item l prior to entry into stratum s ($\zeta_l = 0$ if no prior failure on the item).

Prentice et al. also extend the PWP model to multivariate failure time applications where there will be more than one type of failure. Let $J \in \{1, 2, ..., m\}$ denote m mutually exclusive failure type classes. Analogous to (2.35) it is possible to define type-specific intensity functions at time t by,

$$\iota_j(t|N(t), Z(t)) = \lim_{\Delta t\to 0} \frac{\Pr[t \leq T_{n(t)+1} \leq t + \Delta t, J = j|N(t), Z(t)]}{\Delta t} \tag{2.40}$$

where, in this case, $N(t) = \{N_1(t), ..., N_m(t)\}$ is the counting process for each of the m types of failure and $n(t) = n_1(t) + ... + n_m(t)$. This leads to the following extensions of (2.36) and (2.37):

$$\iota_j(t|N(t), Z(t)) = \iota_{0_{sj}}(t) \cdot \exp(\boldsymbol{\gamma}_{sj} \cdot \boldsymbol{z}(t)) \tag{2.41}$$

$$\iota_j(t|N(t), Z(t)\} = \iota_{0_{sj}}(t - t_{n(t)}) \cdot \exp(\boldsymbol{\gamma}_{sj} \cdot \boldsymbol{z}(t)) \tag{2.42}$$

Prentice et al. applied their models on a data set from Atkinson et al. (1979) and generally achieved better results than with an ordinary PHM. In another example, Ascher (1983) used PWP Model 2 on marine gas turbine failure data by using indicative covariates, i.e. 0's and 1's, with good results. Ascher believes the "custom tailoring" allowed by the PWP models, is essential in failure data analysis.

2.3.3.2 Accelerated Failure Time Models (AFTM)

Pike (1966) introduced the AFTM and it is often a useful alternative to the PHM in many reliability modeling situations, according to Newby (1988). This model incorporates the effect of covariates by allowing for changes in the time scale of, for example, the reliability function.

Let the probabilistic reliability function be given by $R_X(x)$ and the accelerated reliability function be denoted by $\widehat{R_X(x)}$, due to environmental stresses, i.e.

$$\widehat{R_X(x)} = R[(x-c)/b] \tag{2.43}$$

where c is a location parameter and b is a scale parameter. The model is similar to regression models which assume that $(x-c)/b$ is distributed according to a known parametric form. Some density functions often used, are,

Weibull:

$$\widehat{f_X(x)} = \frac{k}{b}u^{k-1}\exp(-u^k), \quad u = (x-c)/b \tag{2.44}$$

Gamma:

$$\widehat{f_X(x)} = \frac{1}{b\Gamma(k)}u^{k-1}\exp(-u), \quad u = x/b \tag{2.45}$$

Log-normal:

$$\widehat{f_X(x)} = \frac{1}{buk\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left[\frac{\ln(u)}{k}\right]^2\right\}, \quad u = (x-c)/b \tag{2.46}$$

Inverse Gaussian:

$$\widehat{f_X(x)} = \frac{1}{b}\sqrt{\frac{k}{2\pi}}u^{-3/2}\exp\left\{-\frac{-k(u-1)^2}{2u}\right\}, \quad u = x/b \tag{2.47}$$

where k is a shape parameter.

A logical variation on the AFTM is where the underlying distribution is only a function of x and k, $f_X(x,k)$, but the accelerated distribution is $f_X\widehat{(x/b_i,k)}$, with $b_i = b(\boldsymbol{z}_i;q)$ where $\boldsymbol{z}_i$ is a vector of covariates and q is a flexibility parameter. Commonly used models for b are:

(i) Constant

(ii) Linear as a function of stress, e.g. $b = \boldsymbol{z}_i \cdot a$

(iii) Exponential, e.g. $b_i = \exp(\boldsymbol{z}_i \cdot a)$

(iv) Inversely exponential, e.g. $b_i = \exp(-1/(\boldsymbol{z}_i \cdot a))$ (Arrhenius model)

Constructing the likelihood for the mentioned models is similar to simple two-parameter likelihood construction. Many authors have discussed the likelihood construction, including Smith and Naylor (1987), Cheng and Amin (1983) and Cheng and Isles (1987).

Solomon (1984) has shown that, in the absence of censoring, the relative effect of covariates are identical in the AFTM and the PHM. Great care should thus be taken in such cases that either one of the two models is not misspecified.

A very popular application of the AFTM is fatigue crack growth in the field of structural mechanics. The acceleration-property of the model is used to estimate fatigue crack growth rates. Many examples of this kind can be found in the literature, including Crowder, Kimber, Smith, and Sweeting (1991) and Newby (1988).

Ciampi and Etezadi-Amoli (1985) and Etezadi-Amoli and Ciampi (1987) combined the PHM and the AFTM in the so-called Extended Hazard Regression Model (EHRM), i.e.

$$h(x;\boldsymbol{z}) = h_0\left(x \cdot \psi_1(\boldsymbol{z}(x)\cdot\alpha)\right)\psi(\boldsymbol{z}(x)\cdot\beta) \tag{2.48}$$

where $\psi_1(\boldsymbol{z}(x)\cdot\alpha)$ and $\psi(\boldsymbol{z}(x)\cdot\beta)$ are positive functions equal to 1 when all covariate values are equal to 0. When $\alpha = 0$, the model in (2.48) becomes an ordinary PHM and when $\alpha = \beta$, the corresponding model is an AFTM.

2.3.3.3 Proportional Age Setback (PAS)

In this approach, introduced by Martorell, Munoz, and Serradell (1996), each maintenance action is assumed to shift the origin of time from where the age of the component is evaluated. Let every maintenance action reduce the age of a component, just before maintenance, proportionally by a factor ε, where ε lies in $[0,1]$. If $\varepsilon = 0$, the PAS produces the BAO situation and if $\varepsilon = 1$, the GAN situation results. Thus, the virtual age, τ, of an item after it has undergone its first maintenance action[§] is given by:

$$\tau_1^+ = (1-\varepsilon_1)\cdot\lambda(\boldsymbol{z}_1)\cdot T_1 \tag{2.49}$$

In (2.49), $\lambda(\cdot)$ is a functional term containing covariates in the vector $\boldsymbol{z}_1$. Covariates could be time-dependent or time-independent. The superscript "+" indicates that the virtual age is applicable shortly after the event at T_1. After the second maintenance action the virtual age is,

$$\tau_2^+ = (1-\varepsilon_2)\cdot[\tau_1 + \lambda(\boldsymbol{z}_2)\cdot(T_2 - T_1)] \tag{2.50}$$

Substitution of the above yields:

$$\tau_2^+ = (1-\varepsilon_2)\cdot[(1-\varepsilon_1)\cdot\lambda(\boldsymbol{z}_1)\cdot T_1 + \lambda(\boldsymbol{z}_2)\cdot(T_2 - T_1)] \tag{2.51}$$

If m denotes the maintenance number, the virtual age of a component is generally given by,

$$\tau_m^+ = \sum_{k=0}^{m-1}\lambda(\boldsymbol{z}_{m-k})\cdot\prod_{r=0}^{k}(1-\varepsilon_{m-r})\cdot(T_{m-k} - T_{m-k-1}) \tag{2.52}$$

Martorell, Sanchez, and Serradell (1999) simplify the virtual age model in (2.52) by assuming that,

[§] *Maintenance action* could be interpreted here as renewal, minimal repair or imperfect repair.

(i) the effectiveness of each maintenance action is equal to some constant value ε, i.e. $\varepsilon_k = \varepsilon$.

(ii) constant operating conditions apply, i.e. $\boldsymbol{z}_k = \boldsymbol{z}$.

This leads to a simplification of (2.52), i.e.

$$\begin{aligned}\tau_m^+ &= \lambda(\boldsymbol{z}) \cdot \left[\sum_{k=0}^{m-1} (1-\varepsilon)^{k+1} \cdot (T_{m-k} - T_{m-k-1})\right] \\ &= \lambda(\boldsymbol{z}) \cdot (t_m - \Delta\tau_m)\end{aligned} \tag{2.53}$$

where

$$\Delta\tau_m = \sum_{k=0}^{m-1} (1-\varepsilon)^k \cdot k \cdot T_{m-k} \tag{2.54}$$

This simplification is useful in cases where information in data sets are limited or where a first approximation suffices.

Atwood (1992) used an approach similar to the PAS, specifically on the ROCOF of items, i.e. $\rho(t) = \rho_0 \cdot g(t;\beta)$. Here, ρ_0 is a constant multiplier and $g(t;\beta)$ is the portion of the expression that determines the shape of $\rho(t)$. Three models for the ROCOF are proposed (exponential, linear and power law):

$$\rho(t) = \begin{cases} \rho_0 \exp[\beta(t-t_0)] \\ \rho_0[1+\beta(t-t_0)] \\ \rho_0(t/t_0)^{\beta}] \end{cases} \tag{2.55}$$

The value of t_0 can be selected for convenience. In the first two cases, if t_0 is set to zero, $t - t_0$ is the time measured from the system's installation. In the third case, t_0 normalizes the scale in which time is measured. In all three models, ρ_0 has units of 1/time. If $\beta > 0$, $\rho(t)$ is increasing, $\beta = 0$, $\rho(t)$ is constant or $\beta < 0$, $\rho(t)$ is decreasing. The value of ρ_0 is the value of $\rho(t)$ at time $t = t_0$. Atwood uses a Bayesian approach to fit the models, i.e. a preliminary analysis is done first based on the conditional likelihood of the models given in (2.55) whereafter the full likelihood is constructed and the values of parameters are estimated.

2.3.3.4 Proportional Age Reduction (PAR)

Malik (1979) introduced the PAR model, where the virtual age is based on the survival time of the most recent lifetime. This differs from the PAS approach where the virtual age is based on the entire history.

Let ε be the efficiency factor, as before, that lies within $[0,1]$. The virtual age of an item, after it has undergone its first maintenance action, in the PAR model is given by:

$$\tau_1^+ = (1-\varepsilon_1) \cdot \lambda(\boldsymbol{z}_1) \cdot T_1 \tag{2.56}$$

The functional term, $\lambda(\cdot)$, incorporates covariates and the superscript "+" denotes applicability of the τ_1 shortly after event T_1 occurred. After the second maintenance action the virtual age is,

$$\tau_2^+ = \tau_1 + (1 - \varepsilon_2) \cdot \lambda(\boldsymbol{z}_2) \cdot (T_2 - T_1) \tag{2.57}$$

Immediately after maintenance action m, the virtual age is given by:

$$w_m^+ = \sum_{k=m}^{m} (1 - \varepsilon_k) \cdot \lambda(\boldsymbol{z}_k) \cdot (T_m - T_{m-1}) \tag{2.58}$$

If ε and $\boldsymbol{z}$ are fixed, (2.58) simplifies to,

$$\tau_m^+ = (1 - \varepsilon) \cdot \lambda(\boldsymbol{z}) \cdot T_m \tag{2.59}$$

This simplified estimation of the PAR was applied by, amongst others, Malik (1979) and Shin, Lim, and Lie (1996). Shin, Lim, and Lie, for example, implemented the PAR concept on two models namely the power-law intensity function (Weibull) and log-linear intensity function. From here the PAR model is defined as

$$\iota_{k+1}(t) = \iota_k(t - \zeta \cdot \tau_k),\ t > \tau_k \tag{2.60}$$

where ζ is an improvement factor or factor of rejuvenation and $0 \leq \zeta \leq 1$. This particular model was only used on a single item under observation but it can be extended to handle counts of multiple system copies.

2.3.4 Marginal regression analysis

Marginal regression analysis has been used with success in the field of biomedicine to represent multiple-event time data. See for example Pepe and Cai (1993) and Wei, Lin, and Weissfeld (1989). The approach of marginal regression analysis is similar to the stratified PHM approach with vaguely defined strata. This approach has the attractive attribute that no explicit model needs to be formulated for the probabilistic association between failures of the same individual. Wei et al. also allow for k different failure types, censoring and missing observations, which could be very useful in reliability modeling.

Let X_{ki} be the failure time of the i^{th} subject $(i = 1, ...n)$ that experiences the k^{th} type of failure $(k = 1, ..., K)$. In some instances a bivariate vector, $\widetilde{\boldsymbol{X}}_{ki}$, is observed consisting of $[X_{ki}, \Delta_{ki}]$, where $X_{ki} = \min(\widetilde{\boldsymbol{X}}_{ki}, C_{ki})$ and C_{ki} is the censoring time. Let $\Delta_{ki} = 1$ if $X_{ki} = \widetilde{\boldsymbol{X}}_{ki}$ and $\Delta_{ki} = 0$ otherwise. If $\widetilde{\boldsymbol{X}}_{ki}$ is missing, $C_{ki} = 0$, which implies that $X_{kk} = 0$ and $\Delta_{ki} = 0$, since $X_{ki} = \widetilde{\boldsymbol{X}}_{ki}$ is positive. Let $\boldsymbol{z}_{ki} = [z_{1ki}, ..., z_{1ki}]$ be a vector of p covariates for the i^{th} subject with respect to the k^{th} type of failure. Conditional on $\boldsymbol{z}_{ki}$, the failure vector $\widetilde{\boldsymbol{X}}_i = [\widetilde{X}_{1i}, ..., \widetilde{X}_{Ki}]$ and the censoring vector $\boldsymbol{C}_1 = [C_{1i}, ..., C_{Ki}]$ $(i = 1, ..., n)$ are assumed to be independent. For the k^{th} type of failure of the i^{th} subject, the FOM $h_{ki}(x)$ is assumed to take the form,

$$h_{ki}(x) = h_{k0}(x) \cdot \exp[\boldsymbol{\gamma}_k \cdot \boldsymbol{z}_{ki}(x)] \tag{2.61}$$

where $h_{k0}(x)$ is an unspecified baseline FOM and $\boldsymbol{\gamma}_k$ is a vector of failure-specific regression coefficients. If $\mathbb{F}_k(x) = \{l : X_{kl} \geq x\}$ is defined as the set of subjects at risk just prior to time x with the respect to the k^{th} type of failure, the k^{th} failure-specific partial likelihood is,

$$L_k(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_{ki}(X_{ki}))}{\sum\limits_{l \in \mathbb{F}_k(X_{ki})} \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_{kl}(X_{ki}))} \right]^{\Delta_{ki}} \tag{2.62}$$

Pepe and Cai (1993) considered a simplification of the approach above by defining a FOM $h^F(x)$ for individuals at risk at time x but not previously infected and a FOM $h^R(x)$ for individuals at risk and previously infected. It is possible to decompose $h^R(x)$ into further components, i.e. $\{h^{2/1}, h^{3/2}, ...\}$, where $h^{k/(k-1)}(x)$ is the FOM of individuals with the k^{th} infection amongst those who have already experienced $k-1$ infections. Ascher, Kobbacy, and Percy (1997) proposed a similar approach by specifying a different FOM for items following either corrective maintenance or preventive maintenance. Every model has its own baseline and regression coefficients.

The approach of marginal regression analysis as outlined above requires large data sets - something that is not common in reliability. The failure type specific regression coefficients is an important attribute however, since machines rarely fail repeatedly because of the same type of failure.

2.3.5 Competing risks

Crowder (1991) believes the principle of competing risks is best explained by an example from the field of biomedicine. Suppose the time to recurrence of a specific type of cancer in a group of patients is modeled. Patients not only run the risk of the recurrence of cancer but also, for example, of dying before recurrence or developing a different disease before recurrence. This problem is defined as competing risks in data and is common in reliability problems.

Competing risk models have two interpretations: (1) it describes the lifetime of a system subject to several potential causes of failure; and (2) it describes the lifetime of a system consisting of a series of components which fails as soon as one of the components fail. The occurrences of potential failures can be regarded as a vector of random variables $\boldsymbol{X} = [X_1, ..., X_n]$, so that the actual stopping time is at the smallest element of $\boldsymbol{X}$, say X_i. If the random variables in

$\boldsymbol{X}$ are independent, the system reliability is given by,

$$R_{sys}(x) = \prod_{i=1}^{n} R_i(x) \tag{2.63}$$

with FOM

$$h_{sys}(x) = \sum_{i=1}^{n} h_i(x) \tag{2.64}$$

Competing risks situations arise naturally in reliability problems, particularly where series systems are considered. Equations (2.63) and (2.64) are, for example, directly applicable in the "weakest-link" argument of Blanchard and Fabrycky (1990). Lewis (1987) considered an approach similar to that of competing risks, called the "β-factor" method. This method analyzes a system as a series of (i)subsystems of independent components; and (ii) common-cause components. Crowder (1991) derives the likelihood for competing risks models in general terms.

2.3.6 Frailty or Mixture Models

The concept of frailty or mixture is used in two ways in reliability models: (1) as a way of introducing an idea of heterogeneity into the construction of a model; and (2) as an object of interest in itself. Mixture models are more applicable in reliability than frailty models.

Frailty in this context is an unobservable random effect shared by subjects in a group. It is defined by Vaupel, Manton, and Stallard (1979) rather like the PHM, but differs in that the relative risk factor is a random variable in this case. The frailty, ξ, is defined in terms of the FOMs of individuals in a population, i.e.

$$h(x|\xi) = \xi \cdot h_0(x) \tag{2.65}$$

is the FOM of an individual with frailty ξ and baseline FOM, h_0. If the frailty at time x has a density $f_x(\cdot)$, the average FOM at time x is,

$$\begin{aligned} \overline{h}(x) &= \int_0^{\infty} h(x|\xi) \cdot f_x(\xi) \mathrm{d}\xi \\ &= h_0(x) \int_0^{\infty} \xi \cdot f_x(\xi) \mathrm{d}\xi \\ &= \overline{\xi} \cdot h_0(x) \end{aligned} \tag{2.66}$$

If the frailty decreases with time, so will $\overline{\xi}$ (since the weakest die young if no fatal external influences are present). This leads to a situation where the average FOM is declining more rapidly than the FOM for individuals.

Many authors have also used frailties in regression models. See for example Klein and Moeschberger (1990). The inclusion of frailties overcome the limiting assumption of most regression models that survival times of distinct subjects are independent of each other. This assumption is for instance not valid for a study on litter mates that share the same genetic makeup or married couples that share the same, unmeasured environment.

Klein and Moeschberger (1990) present two types of frailty models (both based on Cox's PHM). For the first it is assumed that the FOM of the j^{th} subject in the i^{th} group, given the frailty, to be

$$h_{ij}(x) = h_0(x) \cdot \exp(\sigma w_i + \boldsymbol{\gamma} \cdot \boldsymbol{z}_{ij}) \tag{2.67}$$

where $w_1, ..., w_G$ are frailties. It is assumed that the w's are an independent sample from some distribution with mean 0 and variance 1. The second model is given by

$$h_{ij}(x) = h_0(x) \cdot u_i \cdot \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_{ij}) \tag{2.68}$$

where the u_i's are an independent and identically distributed sample from a distribution with mean 1 and some unknown variance. Common models proposed in the literature for the random effect are the one-parameter gamma distribution, the inverse Gaussian distribution and the log normal distribution.

Lawless (1987) introduced frailties in a Poisson process model by including a variable α_i which accounts for unobservable random effects for each subject, i.e. $\rho(t, \boldsymbol{z}) = \alpha_i \rho_0(t) \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_i)$. The α_i's are independent and identically distributed random variables, independent of the z_i's with some distribution $G(\alpha)$. The likelihood for subject i's event history over $(0, T_i]$ is

$$L_i(\boldsymbol{\theta}) = \int_0^{\infty} \prod_{j=1}^{n_i} \alpha_i \rho(t_{ij}) \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_i) \exp\left\{- \int_0^{T_i} \alpha_i \rho(t) \exp(\boldsymbol{\gamma} \cdot \boldsymbol{z}_i) \mathrm{d}t\right\} \mathrm{d}G(\alpha_i) \tag{2.69}$$

Mixture models arise naturally in reliability according to, amongst others, Lancaster (1990) and Littlewood and Verrall (1973). These models are expressed as a conditional FOM, $h(x|\boldsymbol{z}, \sigma)$, where σ is a random variable with density ω. The conditional density and survivor functions for x are,

$$\begin{aligned} f(x|\boldsymbol{z}) &= \int f(x|\boldsymbol{z}, \sigma)\omega(\sigma)\mathrm{d}\sigma \\ &= \int h(x|\boldsymbol{z}, \sigma) \exp[-H(x|\boldsymbol{z}, \sigma)]\mathrm{d}\sigma \end{aligned} \tag{2.70}$$

and

$$\begin{aligned} R(x|\boldsymbol{z}) &= \int R(x|\boldsymbol{z}, \sigma)\omega(\sigma)\mathrm{d}\sigma \\ &= \int \exp[-H(x|\boldsymbol{z}, \sigma)]\mathrm{d}\sigma \end{aligned} \tag{2.71}$$

Note that,

$$h(x|\boldsymbol{z}) = \frac{f(x|\boldsymbol{z})}{R(x|\boldsymbol{z})} \neq \int h(x|\boldsymbol{z},\sigma)w(\sigma)\mathrm{d}\sigma = \overline{h}(x|\boldsymbol{z}) \tag{2.72}$$

since the FOM defined in terms of frailty is not the FOM of the unconditional distribution.

Mixture models are also often interpreted as Bayesian models with prior $\omega(\sigma)$ for a parameter σ. See for example Lancaster (1990) and Ridder (1990).

2.3.7 Noteworthy extensions of intensity concepts

On a few occasions authors published extensions of the failure intensity concepts described above that are beneficial to this study. These extensions are mostly integrations of different approaches to suit particular applications.

2.3.7.1 A point-process model incorporating renewals and time trends, with application to repairable systems

Lawless and Thiagarajah (1996) presented a family of models that incorporates both Poisson[¶] and renewal behavior although multiple system copies are not considered. The authors studied models of the form,

$$\iota(t,\boldsymbol{z}) = e^{\boldsymbol{\gamma}\cdot\boldsymbol{z}(t)} \tag{2.73}$$

where $\boldsymbol{z}(t) = [z_1(t), ..., z_p(t)]$ and $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_p]$. This model is a special case of that considered by Berman and Turner (1992). Two important Poisson processes can be modeled by (2.73) by specifying the covariates intelligently:

(i) $\rho_1(t) = \exp(\alpha + \beta t)$ by letting $\boldsymbol{z}(t) = [1, t]$ and $\boldsymbol{\gamma} = [\alpha, \beta]$
(ii) $\rho_2(t) = \alpha t^{\beta}$ by letting $\boldsymbol{z}(t) = [1, \log t]$ and $\boldsymbol{\gamma} = [\log\alpha, \beta]$

Renewal processes are obtained by taking $\boldsymbol{z}(t)$ as a function of the backward recurrent time, $B(t)$, as defined in Section 1.2.1. For example, $\boldsymbol{z}(t) = [1, \log B(t)]$ and $\boldsymbol{\gamma} = [\log\alpha, \beta]$ produce a renewal process with a Weibull distribution and FOM $h_X(x) = \alpha x^{\beta}$. Models with $\boldsymbol{z}(t) = [1, g_1(t), g_2(B(t))]$, where g_1 and g_2 are specified functions, incorporate both renewal and time trend behavior.

[¶]See Section A.3.3 for details on Poisson processes.

Anderson, Borgan, Gill, and Keiding (1993) and Berman and Turner (1992) have shown that the maximum likelihood of (2.73) is given by,

$$L(\gamma) = \prod_{i=1}^{n} \rho(t_i) \cdot \exp\left\{-\int_0^T \rho(t)\mathrm{d}t\right\} \tag{2.74}$$

for Poisson processes and for renewal processes by,

$$L(\gamma) = \prod_{i=1}^{n} f(x_i) \cdot \overline{F}(x_{n+1}^*) \tag{2.75}$$

where x_i is defined as before, x_{n+1}^* is a suspension time and $F_X(\cdot)$ is the survivor function.

In an example where Prochan's[‖] "famous" airplane air-conditioning data is modeled, the usefulness of this approach is illustrated. The general model with $g_1(t)$ and $g_2(B(t))$ as covariates was fitted on the data. After evaluation of the significance of the covariates by means of the Wald test statistic it was clear that only $g_1(t)$ was significant, i.e. the the data was more suitable for repairable systems theory because an underlying trend was present in the data. Laplace's trend test (see De Laplace (1773)) and the test by Cox and Lewis (1966) confirmed this result.

Calabria and Pulcini (2000) presented a special case of the model by Lawless and Thiagarajah (1996) where the model determines the characteristics of the failure process during fitting procedures. The two most popular NHPPs (Power-Law Process (PLP) and Log-Linear Process (LLP)) are considered in terms of (2.73), together with the Weibull Renewal Process (WRP). The proposed models are as follows:

(i) The Power-Law Weibull Renewal process with an intensity of,

$$\iota(t|H_t) = \frac{\beta+\delta-1}{\theta^{\beta+\delta-1}} t^{\beta-1}[u(t)]^{\delta-1} \tag{2.76}$$

where $\theta > 0$, $\beta + \delta > 1$ for $0 < t \leq T_1$ and $u(t) = t - t_{N(t)}$. The intensity up to the first failure time T_1 is $\iota(t) = \gamma t^{\beta+\delta-2}$ where $\gamma = (\beta + \delta - 1)/\theta^{\beta+\delta-1}$. This is a power law function which does not depend on H_t or the maintenance policy. If minimal repair was done after each failure, the failure process should evolve on the basis of the intensity. Thus, the ratio of (2.76) and the intensity up to the first failure gives a measure of the improvement or worsening introduced by the actual maintenance policy (minimal repair), i.e. $[u(t)/t]^{\delta-1}$. An indication of the departure from minimal repair is thus given by δ. For example, if $\delta > 1$, the ratio is less than 1 for any $t > T_1$ and, at a given distance $B(t)$ from the most recent failure, it becomes smaller and smaller as the number of occurred failures increases, thus indicating a repeated beneficial effect of maintenance actions on the equipment reliability.

‖ See Prochan (1963).

The parameter β in (2.76) measures the departure from perfect maintenance. If $\beta = 1$, then (2.76) reduces to Weibull renewal. When $\beta > 1$, reliability degradation is experienced and if $1 - \delta < \beta < 1$, reliability improvement is experienced.

(ii) The Log-Linear Weibull Renewal process with intensity function,

$$\lambda(t|H_t) = \delta \exp(\theta + \beta t)[u(t)]^{\delta - 1} \tag{2.77}$$

with $-\infty < \theta,\ \beta < \infty$ and $\delta > 0$ for $0 < t \leq T_1$. Up to the first failure, $\iota(t|H_t) = \exp(\gamma + \beta t)t^{\delta-1}$, where $\gamma = \theta + \ln \delta$, which is exactly the intensity of an ordinary Poisson process. When $\beta = 0$, (2.77) does not depend on global age but only on $B(t)$ which implies perfect maintenance, i.e. Weibull renewal. If $\delta = 1$ the process intensity does not depend on local time and reduces to a log-linear process.

The value of δ has the same physical meaning here as in the Power-Law Weibull Renewal model. But, for the value of β, if $\beta > 0$ a reliability deterioration of the equipment with the operating time is described. The more β differs from 0, the bigger the time trend. Finally, if $\beta = 0$ and $\delta = 1$, the Log-Linear Weibull Renewal process reduces to the HPP.

Likelihood construction for the above models is trivial and will not be discussed here.

2.3.7.2 Simple and robust methods for the analysis of recurrent events in repairable systems

Lawless and Nadeau (1995) considered some robust methods to estimate the behavior of point process data based on the Poisson model. This is an extension of the techniques described by Nelson (1982) for IID data.

Suppose k systems are observed and system i is under consideration over a time period $[0; T_i]$. Let $N_i(t)$ be the number of events up to time t. It follows that the Cumulative Mean Function (CMF) is $M_i(t) = E[N_i(t)]$. In the continuous sense $m_i(t) = M_i^{'}(t)$, which is the ROCOF.

To estimate the common CMF in a discrete sense, let $n_i(t) > 0$ be the number of events that occur to system i at time t. This means $m(t) = E[n_i(t)]$ and hence $M(t) = \sum_{s=1}^{t} m(s)$. System i is observed over $[0; T_i]$ and we define $\delta_i(t) = 1$ if $t \leq T_i$ and $\delta_i(t) = 0$ if $t > T_i$ to indicate whether i is observed at t. The total number of events is given by $n.(t) = \sum_{i=1}^{k} \delta_i n_i(t)$ and the total number of systems observed at t is $\delta.(t) = \sum_{i=1}^{k} \delta_i(t)$. Further, assume the k systems under observation are mutually independent. Then, if the $n_i(t)$'s are independent Poisson random variables with means $m(t)$, the MLE's of $m(t)$ is given by:

$$\hat{m}(t) = \frac{n.(t)}{\delta.(t)} \tag{2.78}$$

Similarly for $M(t)$ we have,

$$\hat{M}(t) = \sum_{s=0}^{t} \frac{n.(t)}{\delta.(t)} \tag{2.79}$$

The authors present a valve seat replacement example as well as an automobile warranty claim example to illustrate the above concepts.

This publication is very useful for this study since it is robust and simple. It is ideal for a preliminary analysis of data. The assumption that the end observation times τ_i are independent of the event process may be somewhat unrealistic in reliability problems. For example, if system failures are studied and systems with many failures are withdrawn from service earlier, the estimates of $M(t)$ or the regression coefficients could be badly biased.

2.4 Conclusion

Chapter 2 covered the majority of advanced failure intensity model classes in the literature as well as a few noteworthy extensions of model classes. Kumar and Westberg (1996b) compiled a summary of advanced failure intensity models for non-repairable systems and how these models are interrelated. This summary was broadened and generalized for both non-repairable and repairable systems and is shown in Figure 2.1 on the next page.

The theory, models and concepts discussed in this chapter are used in chapters to follow to achieve the objectives set in Section 1.6.

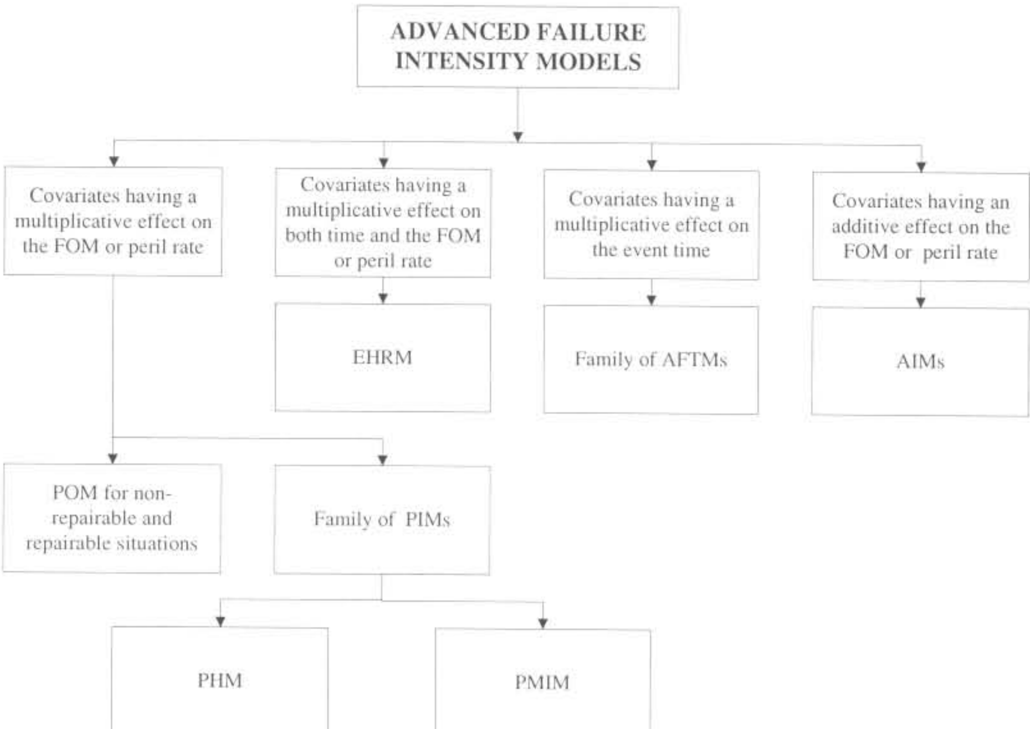

Figure 2.1: Summary of different advanced failure intensity models