Accelerated Failure Time Models: An Application in Insurance Attrition

Abdul-Fatawu Majeed

Université de Pau et des Pays de l'Adour, Pau, France
mangoamana@gmail.com

Received: May 21, 2020
Revised: August 4, 2020
Accepted: August 5, 2020

Abstract

Despite the seeming power of survival analysis over popular binary models in insurance attrition analysis, its consideration is now growing in the literature. Besides, studies have only considered the Kaplan-Meier estimator and the Cox proportional hazards model. To our knowledge, no single study has modeled insurance attrition using the accelerated failure time model. This study presents some parametric models in survival analysis, specifically, the accelerated failure time model. Furthermore, we investigate the applicability of this model in analyzing insurance attrition using life insurance data. We show for the first time that the accelerated failure time model offers an attractive alternative to the Kaplan-Meier estimator, and the Cox proportional hazards model in estimating insurance attrition. Based on the Akaike information criterion, the generalized gamma model provides the best fit for the data. This work will serve as the basis for the consideration of parametric survival models in estimating insurance attrition, deepen knowledge in retention analysis, and broaden the scope of survival analysis.

Keywords: Insurance Attrition, Survival Analysis, Accelerated Failure Time Model, Proportional Hazards Model.

1 Introduction

The growing need to include covariates in the analysis of time-to-event data has brought forth the two popular regression models: the Cox proportional hazards model (PH model) and the accelerated failure time (AFT) model. The Cox proportional hazards model (D. R. Cox, 1972) is the most popular and widely used regression model in survival analysis. Beyond its extensive use in biostatistics, it is also receiving attention in finance, insurance, labor market research, and political science (Xue & Schifano, 2017). The PH model assumes that the effect of covariates is multiplicative on the hazard and that for any two individuals with fixed covariates, the hazard ratio is constant. Despite the benefits of the PH model, it has been met with some criticisms. Notably, the proportional hazard assumption is seldom met, and this problem is more acute if many predictors in a multivariate analysis were to meet the assumption (Khanal, Sreenivas, & Acharya, 2014).

Parametric regression models offer an attractive alternative to make predictions beyond the last survival time and when hazard functions or relative survival times are of primary interest or measure of association (C. Cox, Chu, Schneider, & Munoz, 2007). George, Seals, and Aban (2014) identified three classes of parametric models: parametric proportional hazards, the additive hazards model, and the AFT model. Here, we focus on the last class, the AFT model, a case similar to conventional linear regression. The AFT model explains a linear relationship between the logarithm of the survival time and the covariates. Furthermore, the model is comparatively easier to implement and interpret. Also, unlike the proportional hazards model, the AFT model assumes a multiplicative effect of covariates on survival time (Swindell, 2009). Even though there are clear cut benefits of AFT models, they come with strong assumptions about the distribution of the event-time. Also, the choice of the appropriate parametric distribution is not a straight forward task. Nonetheless, the AFT model describes the evolution of the time-to-event better when the assumptions of the underlying distribution are met. Besides, in implementing the AFT model, one does not necessarily have to check, a prior, the proportional hazards assumption. For some comparative discussion and application of the PH and AFT models, (see, for example Bradburn, Clark, Love, & Altman, 2003; Clark, Bradburn, Love, & Altman, 2003).

The application of survival analysis is widespread and not new, from criminology to epidemiology. However, unlike the PH model, parametric models are not a common choice in lifetime analysis. That notwithstanding, they are widely used in modern medical statistics and actuarial work (Richards, 2011) and are considered in different applications. In retention analysis, specifically in insurance, survival analysis has shown an added advantage over popular binary regression models and thus receiving consideration in the literature. Nevertheless, up to our knowledge, no single study has modeled insurance attrition using the AFT model. In regards, this study demonstrates that the AFT model is an efficient alternative to the Kaplan-Meier (KM) and PH methods in estimating insurance attrition. Consequently, we propose that consideration be given to its application in analyzing insurance attrition.

The rest of the paper is organized as follows: Section 2 presents some useful concepts, the accelerated failure model, model selection, parameter estimation, and a brief discussion of the PH model. The application of the AFT model in insurance attrition is discussed in Section 3. Data analysis and results are presented in Section 4, and we conclude in Section 5.

2 Methods

2.1 Parametric model

In this model, the survival time is assumed to follow a known distribution. Parametric survival models are known for their consistency with theoretical survival, completeness (hazard and survival are specified), time-quantile prediction, and simplicity (Dätwyler & Stucki, 2011). Some distributions commonly used for modeling survival time are exponential, Weibull, log-logistic, log-normal, gamma, and generalized gamma. For extensive discussion and illustrations, see Kleinbaum and Klein (2010), Marshall and Olkin (2007) and Klein and Moeschberger (2006). These models have been assessed in various applications in the literature for their fit to time-to-event data (see, for example Abadi, Amanpour, Bajdik, & Yavari, 2012; George et al., 2014; Montaseri, Charati, & Espahbodi, 2016). The Weibull distribution is known to be the most commonly used survival model. This has been attributed to the flexibility of its hazard function. However, where the underlying hazard function is a bathtub or unimodal shaped, the Weibull distribution does not provide a reasonable parametric fit (Barriga, Louzada-Neto, & Cancho, 2008). In the following, we discuss the AFT model and then, briefly, these commonly used survival models. Further, we investigate the applicability of the AFT model in insurance attrition analysis.

2.2 Some useful concepts

Let T denote the survival time of the event of interest. The survival function, $S_T(t)$ is the probability that the event occurs later than some time t and defined as

$$S_T(t) = P(T > t), \quad \forall\, t \geq 0.$$

The lifetime of T can also be characterized by; the probability density function (pdf)

$$f_T(t) = -S'_T(t),$$

the hazard function

$$h_T(t) = f_T(t)/S_T(t) = -S'_T(t)/S_T(t) = -\frac{\mathrm{d}}{\mathrm{d}t}[\ln\left(S_T(t)\right)],$$

the cumulative hazard function

$$H_T(t) = \int_0^t h_T(u)\mathrm{d}u = -ln[S_T(t)]$$

and the mean time to event

$$\mu_T(t) = \left(\frac{1}{S_T(t)}\right) \int_t^\infty S_T(u)\mathrm{d}u.$$

Mainly, $f(t)$, $S(t)$ and $h(t)$ are fundamental, and specifying one allows the other two to be derived. In particular, $S(t)$ and $h(t)$ are more interesting and better explains the evolution of T.

2.3 The accelerated failure time (AFT) model

For a given survival time T and a vector of covariates $X \in \mathbb{R}^p$ with corresponding parameters $\beta \in \mathbb{R}^p$, the accelerated failure time model can be formulated on the log-scale (similar to linear regression) as

$$Y = \beta_0 + \beta' X + \varepsilon, \tag{1}$$

where $Y = \log(T)$, ε is a random error term assumed to follow some parametric distribution and β_0 is the intercept. For some distributions (e.g. Weibull) there is an additional parameter σ, which scales ε. In this case, the AFT model becomes;

$$Y = \beta_0 + \beta' X + \sigma\varepsilon. \tag{2}$$

The AFT model assumes that covariates have a multiplicative effect on the survival time and an additive effect (see equations 1 and 2) on $\log(T)$. From equation (2), the former implies

$$T = \exp\left(\beta_0 + \beta' X + \sigma\varepsilon\right) = \exp\left(\beta_0\right) \times \exp\left(\beta' X\right) \times \exp\left(\sigma\varepsilon\right).$$

Also, the survival function of T can be expressed in terms of the survival function of ε. Given the set of covariates, the survival function of T denoted as $S_{T|X}(t|x)$ is derived as follows:

$$\begin{aligned}
S_{T|X}(t|x) &= P\left(T > t|X = x\right) = P\left(e^Y > t|X = x\right) \\
&= P\left(e^{\beta_0+\beta' X+\sigma\varepsilon} > t|X = x\right) = P\left(\beta_0 + \beta' X + \sigma\varepsilon > \log t|X = x\right) \\
&= P\left(\varepsilon > \frac{1}{\sigma}\left(\log t - (\beta_0 + \beta' X)\right)|X = x\right) \\
&= S_\varepsilon\left(\frac{1}{\sigma}\left(\log t - (\beta_0 + \beta' x)\right)\right) \qquad (3) \\
&= S_\varepsilon\left(\log\left(\frac{t}{e^{\beta_0+\beta' x}}\right)^{\frac{1}{\sigma}}\right) \\
&= S_\varepsilon\left(y\right), \qquad \text{where} \quad y = \log\left(\frac{t}{e^{\beta_0+\beta' x}}\right)^{\frac{1}{\sigma}}.
\end{aligned}$$

In the following, we look at some common distributions of the error term ε and for each, the associated AFT model of T.

Exponential

The exponential distribution has the survival function, $S_T(t) = e^{-\lambda t}$ for all $t \geq 0,\ \lambda > 0$ and a constant hazard function $h_T(t) = \lambda$. If the lifetime T is exponential, then ε follows a Gumbel distribution with the survival function $S_\varepsilon(y) = \exp(-e^y)$, and from equation (1) we obtain

$$S_{T|X}(t|x) = \exp(-\lambda t), \qquad \text{where} \quad 1/\lambda = \exp(\beta_0 + \beta' x).$$

The exponential distribution is both a PH model and an AFT model with different parametrization, where in the former $\lambda = \exp(\beta_0 + \beta' x)$.

Appropriateness Check (Graphical method)
The AFT assumption holds for the exponential model if a plot of $\log[-\log \hat{S}(t)]$ against $\log(t)$ yields a straight line with a unit slope, where $\hat{S}(t)$ is the KM (Kaplan & Meier, 1958) survival estimate.

Weibull

The Weibull distribution with shape and scale parameters α and λ respectively, have the survival function

$$S_T(t) = e^{-(\lambda t)^\alpha}$$

and hazard function

$$h_T(t) = \alpha \lambda^\alpha t^{\alpha - 1}$$

for all $t \geq 0$ and $\lambda, \alpha, > 0$. The hazard function is increasing if $\alpha > 1$ and decreasing if $\alpha < 1$. When $\alpha = 1$, the Weibul model reduces to the exponential model. There are several ways of parameterizing the Weibull distribution. Similar to the exponential model, the Weibull model is also related to the Gumbel distribution. In this case, we have;

$$S_{T|X}(t|x) = \exp(-(\lambda t)^\alpha), \quad \text{where } 1/\lambda = \exp(\beta_0 + \beta' x) \text{ and } \alpha = 1/\sigma.$$

We can see that the exponential model is a special case of the Weibull model with a shape parameter equal to 1. Both models are also candidates for the PH assumption.

Appropriateness Check(Graphical method)
The AFT assumption holds for the Weibull model if a plot of $\log[-\log \hat{S}(t)]$ against $\log(t)$ yields a straight line, where $\hat{S}(t)$ is the KM survival estimate. For covariates with two or more levels, the AFT assumption holds if the lines are straight and parallel, otherwise, it is violated.

Log-logistic

The lifetime T follows a log-logistic distribution if ε is logistically distributed with survival function

$$S_\varepsilon(y) = 1/(1 + e^y).$$

In regards, the log-logistic AFT model has a survival function given as

$$S_{T|X}(t|x) = \frac{1}{1+(\lambda t)^{\alpha}}, \quad \text{where } 1/\lambda = \exp\left(\beta_0 + \beta' x\right),\ \alpha = 1/\sigma.$$

Unlike the Weibull model, the log-logistic AFT model provides a non-monotonic hazard function. The hazard decreases monotonically over time when $\alpha \leq 1$ and it is unimodal when $\alpha > 1$.

Proportional Odds (PO) model
In a PO model, the ratio of the odds of survival does not depend on time (i.e constant). The odd of survival is the ratio of the probability of surviving beyond time t to the probability of failure at time t. The inverse of survival odds gives the failure odds, the odds of failure at time t. The log-logistic AFT model is a PO model as shown below.
Let's consider two levels of a covariate, namely x_1 and x_2. The survival odds and the survival odds ratio of the two levels are obtained as follows.

$$\left(S_{T|X}(t|x_1)\right) / \left(1 - S_{T|X}(t|x_1)\right) = 1/\left(\lambda_1 t\right)^{\alpha}, \quad \text{where } \lambda_1 = 1/\left(\exp\left(\beta_0 + \beta x_1\right)\right), \alpha = 1/\sigma$$

$$\frac{S_{T|X}(t|x_1)/\left(1 - S_{T|X}(t|x_1)\right)}{S_{T|X}(t|x_2)/\left(1 - S_{T|X}(t|x_2)\right)} = \frac{(\lambda_2 t)^{\alpha}}{(\lambda_1 t)^{\alpha}} = \frac{\lambda_2^{\alpha}}{\lambda_1^{\alpha}}.$$

Appropriateness Check (Graphical method)
The log-logistic assumption holds if a plot of $\log[\hat{S}(t)/(1-\hat{S}(t))]$ or $\log[(1-\hat{S}(t))/\hat{S}(t)]$ against $\log(t)$ yields a straight line of slope $-\alpha$ and α respectively. For covariates with two or more levels, the AFT assumption holds if the lines are straight (log-logistic) and parallel (PO).

Log-normal

If ε follows the standard normal distribution (i.e $\varepsilon \backsim \mathcal{N}(1,0)$), then $S_{\varepsilon}(y) = 1 - \Phi(y)$, where $\Phi(y)$ is the cumulative density function of the standard normal distribution. In this case, T follows the log-normal distribution and has the survival function given by

$$S_{T|X}(t|x) = 1 - \Phi\left(\frac{\log t - (\beta_0 + \beta' x)}{\sigma}\right).$$

We can see that $T \backsim \mathcal{N}\left(\beta_0 + \beta' x,\ \sigma\right)$. The log-normal AFT model does not model accurately, most time-to-event distributions. Its hazard function initially increases from 0 to reach a maximum and then afterward decreases monotonically, approaching 0 as $t \to \infty$.
The log-normal and log-logistic models yield similar results. However, the PO property of the later differentiates the two.

Appropriateness Check (Graphical method)
For the log-normal distribution, the AFT assumption holds if a plot of $\Phi^{-1}[1 - \hat{S}(t)]$ versus $\log(t)$ is linear.

Gamma

The probability density function of the gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\lambda > 0$ is given by

$$f_T(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} \quad \forall\, t > 0, \text{ where } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \mathrm{d}t \quad \text{is the gamma function.}$$

The survival function and the hazard function do not have a closed form expressions which are given by

$$S_T(t) = \int_t^\infty f_T(u)\mathrm{d}u = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_t^\infty u^{\alpha-1} e^{-\lambda u} \mathrm{d}u = 1 - I(\lambda t, \alpha),$$

where $I(.)$ is the incomplete gamma function.

$$h_T(t) = f_T(t) / \left(\int_t^\infty f_T(u)\, \mathrm{d}u \right) = \lambda / \left(1 + (\alpha - 1) \int_1^\infty s^{\alpha-2} e^{-\lambda t(s-1)} \mathrm{d}s \right).$$

The hazard of the gamma model is increasing if $\alpha > 1$, decreasing if $\alpha < 1$ and constant if $\alpha = 1$. In the last case, the gamma model reduces to exponential, and as $t \to \infty$, the hazard is equal to λ (constant). The gamma model has properties that are somewhat similar, though not mathematically tractable, to the Weibull model as observed above. From equation (1), T has a gamma distribution if ε has a negatively skewed distribution with skewness decreasing with increasing α, and a pdf defined as (Kalbfleisch & Prentice, 2011).

$$f_\varepsilon(y) = \frac{e^{\alpha y - e^y}}{\Gamma(\alpha)} \quad \forall\, y > 0,\ \alpha > 0$$

and the AFT model has the pdf

$$f_{T|X}(t|x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \text{ where } \lambda = 1/\exp(\beta_0 + \beta' x).$$

Generalized gamma

This model extends the gamma distribution by adding the parameter σ, which scales the error term ε as in equation (2), where ε has the pdf defined above for the gamma distribution. This gives the pdf of T for the generalized gamma AFT model as

$$f_{T|X}(t|x) = \frac{\gamma \lambda^{\gamma\alpha}}{\Gamma(\alpha)} t^{\gamma\alpha-1} e^{-(\lambda t)^\gamma}, \text{ where } \lambda = 1/\exp(\beta_0 + \beta' x),\ \gamma = 1/\sigma.$$

It is easy to check that the generalized gamma distribution includes in special cases, the exponential when $\alpha = \sigma = 1$; the gamma when $\sigma = 1$; the Weibull when $\alpha = 1$, and approximates the log-normal distribution as $\alpha \to \infty$.

2.4 Parametric model fit checks

In addition to the graphical method, there are other metrics for checking the fitness of the underlying distribution of the AFT model. One popular method for comparing AFT models with different underlying distributions is the Akaike (Akaike, 1969) information criterion (AIC), defined as $AIC = -2l + 2p$, where l is the log-likelihood, and p is the number of parameters in the model (model-specific and covariates). Another tool similar to the AIC is the Bayesian (Schwarz et al., 1978) information criteria (BIC) computed as $BIC = -2l + p\log(n)$, where n is the number of observations. In both methods, the decision is to consider the model with the smallest AIC or BIC. Also, the likelihood ratio (LR) test can be used to compare nested models (Qi, 2009). Other methods are the quantile-quantile plot and the Cox- Snell (D. R. Cox & Snell, 1968) residuals plot.

2.5 Parametric model estimation

The maximum likelihood method can be used to estimate the parameters of the AFT model. This is based on the assumption that event times are independent and there are no competing risks. Also, the method can accommodate right, left, or interval-censored data (Kleinbaum & Klein, 2010). However, this study is restricted to the right-censored case. To be precise, Type I right censoring. Suppose we have n observations from a time-to-event study, and the event time for each observation is either censored or observed and independent of other event times. Let $\theta = (\theta_1, ...\theta_p)$ for $j \in (1, ...p)$ be the set of parameters to estimate. The maximum likelihood for the AFT model is formulated as follows.

$$L\left(\theta|obs, \delta\right) = \prod_{i=1}^{n} \left(f(t_i|x_i, \theta)\right)^{\delta_i} \left(S(t_i|x_i, \theta)\right)^{1-\delta_i},$$

where $f(.)$ and $S(.)$ are the density and survival functions of the distribution of T, and for $1 \leq i \leq n$, δ_i is the censoring indicator such that $\delta = \begin{cases} 1 & \text{if} \quad T \leq C, \\ 0 & \text{if} \quad T > C, \end{cases}$ where C is the study end date.

The maximum likelihood estimate of each parameter in θ is obtained by solving the score function

$$\frac{\partial \log L\left(\theta|obs, \delta\right)}{\partial \theta_j} = 0, \qquad \text{for } j = 1, 2, ..., p.$$

Solving this score function can lead to painful expressions, near impossible. Estimates can then be obtained easily by numerical approximations. One possible consideration is the popular Newton-Raphson method.

2.6 Cox proportional hazards (PH) model

The hazard function of the PH model is given by

$$h_{T|X}(t|x) = h_0(t; \theta)e^{\beta' x} \quad \forall\, t \geq 0,$$

where $h_0(t;\theta)$ is a baseline hazard function (to be specified) with unknown parameter(s) θ. The effect of the covariates is multiplicative on the hazard. For two levels of the covariate X, namely x_1 and x_2, the proportion

$$\frac{h(t|x_1)}{h(t|x_2)} = \frac{h_0(t;\theta)e^{\beta' x_1}}{h_0(t;\theta)e^{\beta' x_2}} = \frac{e^{\beta' x_1}}{e^{\beta' x_2}} = e^{\beta'(x_1 - x_2)}$$

is a constant, called the hazard ratio (HR). The survival, cumulative hazard and the probability density function of the PH model is derived accordingly as

$$\begin{aligned}
H_{T|X}(t|x) &= \int_0^t h_{T|X}(u|x)\mathrm{d}u = \int_0^t h_0(u;\theta)e^{\beta' x}\mathrm{d}u = e^{\beta' x}\int_0^t h_0(u;\theta)\mathrm{d}u = H_0(t;\theta)e^{\beta' x} \\
S_{T|X}(t|x) &= e^{-H_{T|X}(t|x)} = e^{-H_0(t;\theta)e^{\beta' x}} = [e^{-H_0(t;\theta)}]^{e^{\beta' x}} = [S_0(t;\theta)]^{e^{\beta' x}} \quad \forall\, t \geq 0 \\
f_{T|X}(t|x) &= \left[h_{T|X}(t|x)\right]\left[S_{T|X}(t|x)\right] = h_0(t;\theta)e^{\beta' x}[S_0(t;\theta)]^{e^{\beta' x}},
\end{aligned}$$

where $S_0(t;\theta)$ and $H_0(t;\theta)$ are the baseline survival and cumulative hazard functions respectively.

3 Application in insurance attrition

Customer retention is one of the main contributors to business profit and growth, and it is an essential part of marketing, pricing, and customer service initiatives. Retention is highly crucial in a purely customer-oriented business where projected profits are tied to the loyalty of customers. As customer needs become more sophisticated in a very competitive industry, winning a new policy contract is relatively more expensive than retaining an existing one. It is, therefore, imperative for insurance providers to have a better understanding of the evolution of customer attrition for effective implementation of customer service initiatives. Before proceeding, let's differentiate between customer retention and attrition. In simple terms, retention relates to customers who stay after a specified period. In contrast, attrition relates to those customers who leave (or defect) after a specified period. It is also popularly referred to as customer churn. Retention and attrition are mostly measured in terms of rate, thus retention rate and attrition rate, and are complements.

Retention analysis is popular in marketing research and practice, banking, insurance, and telecommunications industry. The focus is often on whether or not attrition will occur after a specified duration and the key drivers of attrition. Whereas in banking, this could be whether a customer will default a loan, in insurance, the interest is whether a policyholder will renew or cancel his policy after, say a year. Conventional methods in retention analysis are the generalized linear models (GLM), specifically logistic and probit regressions, decision trees, neural networks and random forest (Goonetilleke & Caldera, 2013; Hosseni, Tarokh, et al., 2011; Smith, Willis, & Brooks, 2000; Spiteri & Azzopardi, 2018; Su, Cooper, Robinson, & Jordan, 2009). These methods are easy to understand and

interpret, and very useful when the interest is to determine whether attrition will occur or not. However, they only identify the attrition status of a customer, but they do not answer the question of when he will leave or defect (Banasik, Crook, & Thomas, 1999; Lu, 2002). To overcome this limitation, alternative methods (in particular), survival analysis have since been adopted in the recent literature. Survival analysis does not only model the attrition status of a customer but also when attrition will occur. Furthermore, unlike binary models, survival analysis can accommodate time-varying macroeconomic variables and differentiate attrition into non-renewal and cancellation (Fu & Wang, 2014).

Despite the seeming power of survival analysis over conventional models in insurance attrition analysis, its consideration is now growing in the literature. Besides, studies have only considered the KM estimator and the PH model (see for example, Aziz & Razak, 2019; Brockett et al., 2008; Fu & Wang, 2014; Guillén, Nielsen, Scheike, & Pérez-Marín, 2012; Harrison & Ansell, 2002; Hasanthika & Jayasekara, 2017; Pérez Marín, 2006; Spiteri & Azzopardi, 2018). To the best of our knowledge, no single study has modeled insurance attrition using the accelerated failure time (AFT) model. In this study, we propose the AFT model as an alternative to the KM estimator and the PH model in estimating insurance attrition. Further, the applicability of the model in this context is demonstrated using life insurance data. We expect that this work will unveil the need to consider parametric survival models in insurance attrition analysis, and possibly in retention analysis.

3.1 Risk factors associated with life insurance attrition

Empirical studies have shown that policy and policyholder characteristics, as well as economic and the investment environment are risk factors associated with customer attrition in life insurance.

Most findings from the literature showed interest rates and (or) unemployment as significant risk factors of customer attrition in life insurance (Dar & Dodds, 1989; Kuo, Tsai, & Chen, 2003; Outreville, 1990) based on standard time series regressions and cointegration techniques. Also, Kagraoka (2005) observed that the surrender of personal accident insurance contracts is explained by changes in the unemployment rate. However, Kim (2005a), Kim (2005b), Samuel and Lin (2006) and Kiesenbauer (2012) demonstrated that beyond the interest rate and unemployment rate, policyholder attrition behavior depends on additional exogenous factors such as policy age, GDP growth, surrender charge, buyer confidence and company age, based on a broad class of generalized linear models.

Also, Pinquet, Guillén, and Ayuso (2011) modeled the attrition of long-term insurance contracts, using proportional hazard models and found that policyholder's age, health history and method of premium payments are risk factors of customer attrition. Similarly, Milhaud and Dutang (2018) established an association of policyholder's smoking status, age, and payment frequency of premium with life insurance attrition. Furthermore, the findings of Renshaw and Haberman (1986), Cerchiara, Edwards, and Gambini

(2008), Milhaud, Loisel, and Maume-Deschamps (2010) and Eling and Kiesenbauer (2014) suggest that policy duration and policy type are essential drivers of life insurance attrition. Other risk factors associated with life insurance attrition include; policyholder's age and gender (Cerchiara et al., 2008; Eling & Kiesenbauer, 2014; Milhaud & Dutang, 2018).

In summary, the empirical evidence outlines the importance of economic indicators (such as interest rate and unemployment rate), policyholder characteristics (such as age, gender, and health-related habits) and policy characteristics (such as duration, type, and premium payment scheme) in estimating customer attrition in life insurance.

4 Data Analysis

4.1 Data

We investigate the applicability of the AFT model in estimating insurance attrition using the life insurance data used in Milhaud and Dutang (2018). The data is a portfolio of 29,317 Whole Life policies from anonymous life insurers in the United States sold from the tied-agent channel between January 1995 and December 2008. We retrieved it from the R package "CASdatasets" (version 1.0-10) by Dutang and Charpentier (2019). The data consists of both categorical and continuous covariates, which are the characteristics of a policy, the policyholder, and the investment environment. Milhaud and Dutang (2018) differentiated attrition into surrender and other causes (e.g death), which they collectively defined as a lapse. Also, they modeled contract lifetime by a competing risk approach. Again, in their analysis, they compared a nonparametric regression model and a proportional hazards model (Fine & Gray, 1999) and found that the later is quite efficient in accurately predicting policy lifetime. However, in this study, we do not consider such differentiation, and we model attrition by a cause-specific approach using an accelerated failure time model (AFT). Also, we consider additional data on the unemployment rate (see Eling & Kochanski, 2013) for the same period taken from IMF (2020). In life insurance, lapse and surrender are two related technical terms used to somewhat refer to attrition. Here, attrition is used to refer to both surrender and lapse.

4.2 Results

4.2.1 Preliminary analysis

In Table 1 and Table 2, we present the descriptive statistics of the insurance portfolio. The time duration (in quarters) in Table 1 indicates the survival time of a policy, in this case, the time from the onset of the contract until attrition occurs. On average, the lifetime of a policy in the portfolio is 30.26 quarters. The average annual premium (on the original scale) for all payment streams is \$560.88. Also, the mean unemployment rate for the follow-up period (1995 - 2008) is 5.18%. The Dow Jones Index (DJIA) quarterly

variation is standardized on the original scale. Later the normalized values of the DJIA quarterly variation and the unemployment rate will be considered.

Furthermore, the binary censoring indicator in Table 2 describes the attrition status of a policy (observed or censored). The data seem balanced for the gender covariate (50.05% male and 49.95% female policyholders). Invariably, more male policyholders (1.44%) than females have experienced attrition. Also, a higher proportion of non-smokers, young and non-accidental death riders are observed compared to their counterparts. Further, more than half (74.78%) of the policyholders in the portfolio live in other addresses other than East coast and West coast. More so, over 50% of the policies have infra annual premium payment frequency. Accordingly, higher than half (33.54% out of 61.37%) of these policies experienced attrition. In the following, we study the attrition lifetime using a survival model.

Table 1: Policy and environment characteristics

Covariate	Time duration (in quarters)	Annual premium	Dow Jones Index (DJIA) variation	Unemployment rate
Min	0.01	-1.07	-4.53	3.6
mean	30.26	$560.88	0	5.18
max	62.09	12.13	2.43	7.1
Std dev.	18.78	$526.59	0.05	0.68

Note: The correlation values between each pair of the covariates are 0.05, 0.04 and 0.01

Table 2: Policy and policyholder characteristics (categorical)

Covariate		Percentage (%)	Censoring indicator	
			Censored (%)	Observed (%)
Gender	Male [0]	50.05	23.98	26.07
	Female [1]	49.95	25.32	24.63
Payment frequency	infra annual(monthly, quarterly, semi-annual) [0]	61.37	27.83	33.54
	annual [1]	23.44	12.68	10.77
	other (supra annual) [2]	15.19	8.8	6.39
Risk state	smoker [0]	36.99	19.03	17.95
	non smoker [1]	63.01	30.27	32.75
Underwriting age	young (0 to 34 years old) [0]	47.46	23.38	24.08
	middle (35 to 54 years old) [1]	34.04	15.61	18.43
	old (55 to 84 years old) [2]	18.5	10.31	8.19
Living place	East coast [0]	20.62	9.99	10.63
	West coast [1]	4.6	2.41	2.19
	other [2]	74.78	36.9	37.89
Accidental death rider	Rider [0]	16.42	9.1	7.33
	Non rider [1]	83.58	40.2	43.37
Censoring indicator	0 : censored	49.3		
	1 : observed (attrition)	50.7		

In Figure 1, we first check if the attrition lifetime differs across various subgroups. To achieve this, we adopt a graphical approach. For the gender covariate, the survival curve for females lies closely above that of males after period 10 (in quarters). This implies the lifetime of females differs significantly from males. Also, there appear some differences in the survival curves for risk state, premium frequency, and underwriting. Furthermore, the mean survival times of the different levels of these covariates differ significantly. However, the plots for the living place are overlapping and thus appear not very different. This implies that living place may not be a risk factor of customer attrition. However, we will still consider this covariate in the model to further investigate its significance in insurance attrition.

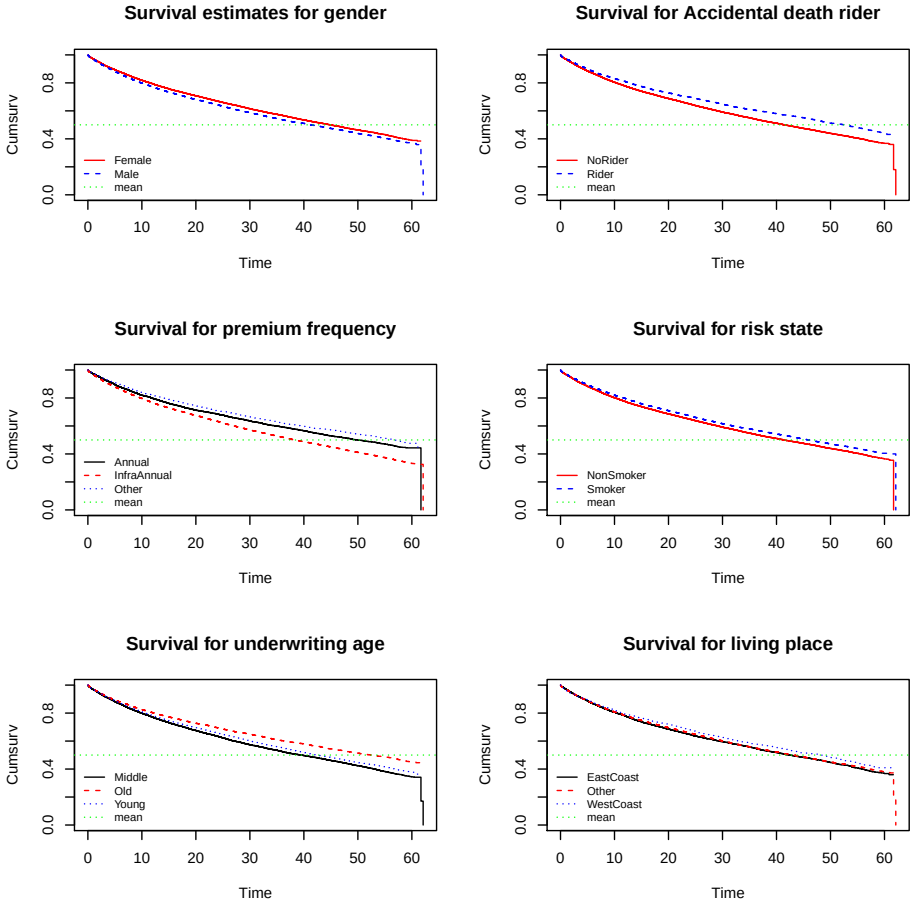

Figure 1: *Kaplan-Meier survival curves for subgroups of categorical covariates*

PH assumption

Before we take the analysis further to the AFT model, we want to check, not necessarily a prior, the proportional hazard assumption. The goodness of fit test in

Table 3 gives a significant global p-value. Therefore the global null hypothesis that the proportionality assumption holds is rejected, and the PH model is not appropriate here. In the following, we proceed to model the attrition lifetime data using the AFT model.

Table 3: Goodness of fit testing for PH assumption

Covariate	chisq	df	p
Gender	2.719	1	0.0990
Payment frequency	25.878	2	0.0000
Risk state	0.152	1	0.6960
Underwriting age	19.243	2	0.0001
Living place	0.459	2	0.7950
Accidental death rider	4.66	1	0.0310
Annual premium	3.643	1	0.0560
Dow Jones Index variation	86.839	1	0.0000
Unemployment rate	1570.603	1	0.0000
GLOBAL	1750.317	12	0.0000

4.2.2 AFT model fitting

To start with, we first check the assumption of the AFT model for the underlying distributions. For the exponential, Weibull, log-logistic, and log-normal models, we use the graphical approach. By comparing the plots in Figure 2, we can see that they approximate fairly, straight lines. However, the Weibull presents a slightly better straight line through the origin, which gives the exponential if it has a unit slope. Thus, the Weibull and exponential assumptions hold better than that of the log-logistic and log-normal models. Also, the plot for the log-logistic appears more straight and close to the origin than the log-normal, which bends towards the right. The generalized gamma approximates these models in special cases except for the log-logistic. Therefore, we will investigate its suitability for the attrition lifetime data.

Next, we now model the data using the exponential, Weibull, log-logistic, log-normal, and generalized gamma AFT models. In each case, first, we fit the model for each covariate in the univariate setting and then all the covariates in the multivariate case. In both the univariate and multivariate AFT models, all the covariates are statistically significantly associated with time to attrition except living place at an error level of 5%, which is consistent with Figure 1. The results, quite similar, for the exponential, Weibull, log-logistic, and log-normal models are presented in the appendix.

In Table 4, we compare the performance of all the models using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The model with the smallest AIC or BIC is considered to provide the best fit. Based on the two criteria, the generalized gamma model is the appropriate AFT model for the attrition data, although

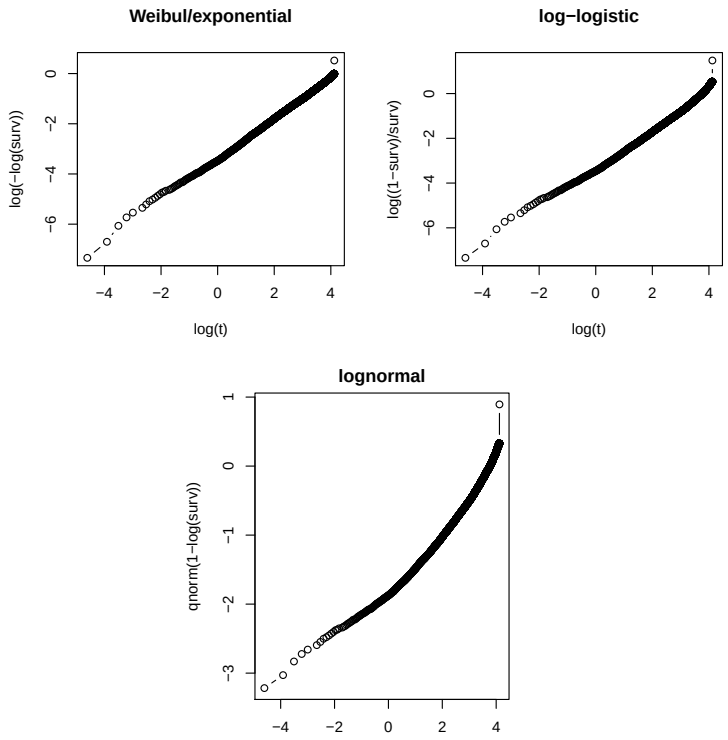

Figure 2: *Graphical check of AFT assumption for exponential/Weibull, log-logistic and log-normal distributions*

it is only slightly better than the Weibull model. Also, the log-normal and log-logistic models perform poorly, which is consistent with the results in Figure 2.

Table 4: Akaike and Bayesian information criteria (AIC and BIC) for the AFT models

Model	No. of parameters (p)	Log-likelihood (L)	AIC	BIC
Exponential	13	-70993.70	142013.30	142045.473
Weibull	14	-70978.00	141984.10	142018.54
Log-logistic	14	-72147.70	144323.30	144357.94
log-normal	14	-72974.00	145976.00	146010.54
Generalized gamma	15	-70583.40	141196.8	141233.807

Note: p = model-specific parameters + no. of betas for the covariates

Overall goodness-of-fit

We assess the goodness of fit of the generalized gamma model using the Cox-Snell residuals plot. Furthermore, we compare the survival estimates from the parametric

model to the non-parametric (KM) estimates. In Figure 3, the survival curve of the fitted AFT model approximates fairly the baseline KM survival, although there appear some departures at early through to mean time duration. Also, the residuals plot approximates the straight line through the origin in the beginning and then afterward shifts for higher values. The assignable cause could be due to covariates that are subject to seasonal variations, in particular, the DJIA quarterly variation and unemployment rate.

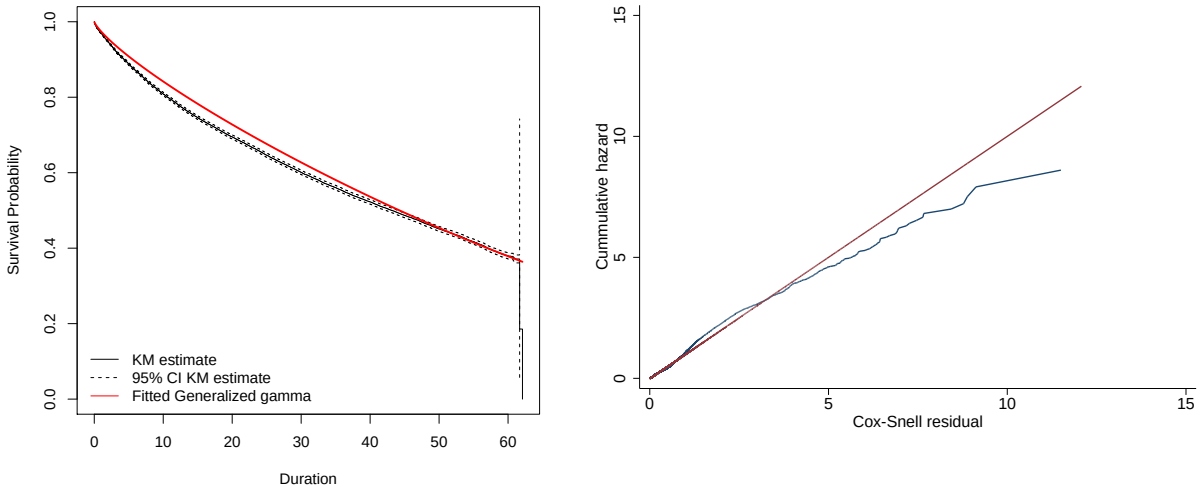

Figure 3: *KM/fitted survival plot (left) and residual plot (right) of Generalized gamma for all covariates*

To understand the shift in the estimates of the model, we drop the DJIA quarterly variation and the unemployment rate covariates and then fit a new generalized gamma model for the remaining covariates. Interestingly, the survival plot (Figure 4) suggest a very good fit of the model. Thus, we can attribute the shift in the full model to possible seasonal variations in the DJIA variation and unemployment rate. Accordingly, we are confident that the generalized AFT model provides a good fit for the data. However, in the presence of the DJIA variation and unemployment rate, the model tends to overestimate for some time duration.

Interpretation of the results

The results of the generalized gamma AFT model fitted to all the covariates are presented in Table 5. The effect of a covariate is to accelerate or decelerate the attrition lifetime. To understand this better, a time ratio (TR), also called the acceleration factor is estimated. The acceleration factor for a given covariate is the (natural) exponent of the corresponding coefficient (i.e $\exp(\beta)$). A positive coefficient means the effect of the covariate is to prolong the survival time while a negative coefficient is to shorten the time to attrition. Relatively, a time ratio greater than 1 implies the effect of the covariate increases the survival time and otherwise decreases ("speeds up") the time to attrition.

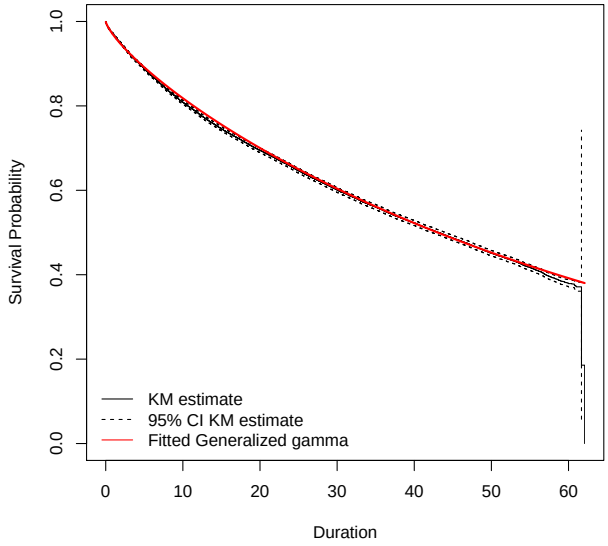

Figure 4: *KM estimate and survival of fitted Generalized gamma without DJIA and unemployment*

The TR of 0.9667 for males relative to females implies gender is a significant predictor of customer attrition, and men are at a slightly higher risk compared to women. This may be attributed to the general conception that women are naturally risk-averse, and thus less likely to be observed for attrition earlier, compared to male policyholders. The implication of this to the insurer is that gender should be considered in attrition risk profiling of an insurance portfolio. Also, compared to policies with annual payment frequency, infra annual policies (TR =0.8869) are at high risk of attrition while that of supra annual policies (TR =1.0554) are at low risk. While in practice, short term payment frequency may put premium within the means of potential customers, over time, the insured may become stressed out, especially when their short term expenses go up without a corresponding increase in income. The insurer may take into consideration the source of funding of a policyholder and adjust the payment frequency accordingly while keeping premium at an affordable level. Potential customers with multiple sources of finances may not be overstretched financially with monthly payments, coupled with increasing expenses overtime.

Interestingly, smokers (TR=1.0512), the old (TR=1.1806), and accidental death riders (TR=1.1074) have longer survival times compared to their counterparts. Naturally, these groups have higher exposure to the risk of death, and therefore, may be willing to keep their policies compared to other policyholders in the same portfolio. Accordingly, these risk factors should be considered when profiling policies for the risk of attrition. Middle-aged policyholders may have demand for funds to own property and possibly,

invest or save towards retirement compared to young policyholders. Hence, they are at high risk of diverting their premiums to other alternative investment products. Also, after meeting health-related costs, old-aged policyholders may not have an immediate demand for funds, hence at a lower risk of attrition compared to middle-aged and young policyholders. Similar to the results in Figure 1, the living place covariate is not significant. Hence it does not explain the attrition evolution of a policyholder.

Conversely, the coefficients for the annual premium, DJIA quarterly variation, and unemployment rate covariates are negative. Accordingly, the effect of these covariates is to decrease survival time. An increase in the unemployment rate will shorten the survival time and thus increase the risk of attrition. This result is expected as a loss of income will inevitably affect the capacity of policyholders to honor their premiums. Also, an increase in premium may trigger customers who do not see it justifiable to terminate their policies since they may feel being cheated. Further, policyholders who are well informed about the investment environment (or have advisors) may react accordingly to increases in the DJIA quarterly variation and thus are at high risk of attrition. Policyholders should therefore be well informed early enough of any possible increase in premiums by the insurer to sustain their trust. However, the DJIA and unemployment rate tend to inflate the estimates at early to mean durations, which tends to explode the risk of attrition.

Table 5: Generalized gamma AFT model

Covariate		β	TR (exp(β))	P-value	95% CI (TR)	
Gender	Male [0]	-0.0339	0.9667	0.0160	0.9404	0.9937
	Female [1]					
Payment frequency	infra annual [0]	-0.1200	0.8869	0.0000	0.8556	0.9183
	annual [1]					
	other (supra annual) [2]	0.0539	1.0554	0.0408	1.0023	1.1118
Risk state	smoker [0]	0.0499	1.0512	0.0007	1.0211	1.0821
	non smoker [1]					
Underwriting age	young (0 to 34 years old) [0]	0.0307	1.0312	0.0464	1.0005	1.0628
	middle (35 to 54 years old) [1]					
	old (55 to 84 years old) [2]	0.1660	1.1806	0.0000	1.1320	1.2312
Living place	East coast [0]					
	West coast [1]	0.0119	1.0120	0.7525	0.9396	1.0900
	other [2]	0.0064	1.0064	0.7100	0.9730	1.0411
Accidental death rider	Rider [0]	0.1020	1.1074	0.0000	1.0634	1.1537
	Non Rider [1]					
Annual premium		-0.0597	0.9420	0.0000	0.9314	0.9529
Dow Jones Index variation		-0.5740	0.5633	0.0000	0.5532	0.5735
Unemployment rate		-0.3810	0.6832	0.0000	0.6730	0.6935

5 Conclusion

This study proposes the accelerated failure time (AFT) model as an alternative to the popular PH model in estimating insurance attrition. Further, we have fitted this

model for the exponential, Weibull, log-logistic, log-normal, and the generalized gamma using life insurance attrition data. Using the Akaike information criterion (AIC), the generalized gamma is the best AFT model for the data. Also, we assessed the overall goodness-of-fit of the model using the Cox-Snell residuals. Besides, we compared the survival curve of the fitted parametric model with the nonparametric (KM) estimates. In both these methods, the model showed a good fit. Furthermore, it is comparatively easier to interpret and allows analysts to make predictions beyond the last survival time. However, in the presence of time series covariates that are subject to seasonal variations, the model tends to overestimate for early to mean lifetimes.

Acknowledgements

The author thanks all reviewers of this article for their insightful suggestions and comments. This research was carried out while the author was enrolled as a master's student at the Université de Pau et Pays de l'Adour (UPPA), Pau, France. Thanks to E2S UPPA for providing the grant for the program. All the views expressed here and any errors are exclusively that of the author.

References

Abadi, A., Amanpour, F., Bajdik, C., & Yavari, P. (2012). Breast cancer survival analysis: Applying the generalized gamma distribution under different conditions of the proportional hazards and accelerated failure time assumptions. *International Journal of Preventive Medicine*, *3*(9), 644.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, *21*(1), 243–247.

Aziz, N., & Razak, S. A. (2019). Survival analysis in insurance attrition. *AIP Conference Proceedings*, *2184*(1).

Banasik, J., Crook, J. N., & Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, *50*(12), 1185–1190.

Barriga, G. D., Louzada-Neto, F., & Cancho, V. G. (2008). A new lifetime distribution with bathtube and unimodal hazard function. In *AIP conference proceedings* (Vol. 1073, pp. 111–118).

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436.

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J., & Perez-Marin, A. M. (2008). Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection? *Journal of Risk and Insurance*, *75*(3), 713–737.

Cerchiara, R. R., Edwards, M., & Gambini, A. (2008). Generalized linear models in life insurance: decrements and risk factor analysis under solvency II. In *18th international AFIR colloquium* (pp. 1–18).

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232–238.

Cox, C., Chu, H., Schneider, M. F., & Munoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, *26*(23), 4352–4374.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, *30*(2), 248–265.

Dar, A., & Dodds, C. (1989). Interest rates, the emergency fund hypothesis and saving through endowment policies: some empirical evidence for the UK. *Journal of Risk and Insurance*, 415–433.

Dätwyler, C., & Stucki, T. (2011). *Parametric survival models.* Retrieved from `https://stat.ethz.ch/education/semesters/ss2011/seminar/contents/handout_9.pdf`

Dutang, C., & Charpentier, A. (2019). *Casdatasets: Insurance datasets (official website).* Retrieved from `http://cas.uqam.ca/`

Eling, M., & Kiesenbauer, D. (2014). What policy features determine life insurance lapse? An analysis of the German market. *Journal of Risk and Insurance*, *81*(2), 241–269.

Eling, M., & Kochanski, M. (2013). Research on lapse in life insurance: what has been done and what needs to be done? *The Journal of Risk Finance*, *14*(4), 392–413.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*(446), 496–509.

Fu, L., & Wang, H. (2014). Estimating insurance attrition using survival analysis. *Variance*, *8*(1), 55–72.

George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, *21*(4), 686–694.

Goonetilleke, T. O., & Caldera, H. (2013). Mining life insurance data for customer attrition analysis. *Journal of Industrial and Intelligent Information*, *1*(1), 52–58.

Guillén, M., Nielsen, J. P., Scheike, T. H., & Pérez-Marín, A. M. (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Systems with Applications*, *39*(3), 3551–3558.

Harrison, T., & Ansell, J. (2002). Customer retention in the insurance industry: using survival analysis to predict cross-selling opportunities. *Journal of Financial Services Marketing*, *6*(3), 229–239.

Hasanthika, N., & Jayasekara, L. (2017). Analyzing the customer attrition using survival techniques. *International Journal of Statistics and Probability*, *6*(6), 85–91.

Hosseni, M. B., Tarokh, M. J., et al. (2011). Customer segmentation using clv elements.

Journal of Service Science and Management, *4*(3), 284.

IMF. (2020). *International financial statistics.* Retrieved from http://data.imf.org/

Kagraoka, Y. (2005). *Modeling insurance surrenders by the negative binomial model.* Retrieved from https://www.researchgate.net/publication/228481596_Modeling_Insurance_Surrenders_by_the_Negative_Binomial_Model

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data* (Vol. 360). New Jersey: John Wiley & Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2014). Accelerated failure time models: an application in the survival of acute liver failure patients in india. *International Journal of Science and Research*, *3*, 161–166.

Kiesenbauer, D. (2012). Main determinants of lapse in the german life insurance industry. *North American Actuarial Journal*, *16*(1), 52–73.

Kim, C. (2005a). Modeling surrender and lapse rates with economic variables. *North American Actuarial Journal*, *9*(4), 56–70.

Kim, C. (2005b). *Report to the policyholder behavior in the tail subgroups project.* Retrieved from https://www.researchgate.net/publication/237378532_Report_to_the_Policyholder_Behavior_in_the_Tail_Subgroups_Project

Klein, J. P., & Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Berlin: Springer Science & Business Media.

Kleinbaum, D. G., & Klein, M. (2010). *Survival analysis* (Vol. 3). Berlin: Springer.

Kuo, W., Tsai, C., & Chen, W. K. (2003). An empirical study on the lapse rate: The cointegration approach. *Journal of Risk and Insurance*, *70*(3), 489–508.

Lu, J. (2002). Predicting customer churn in the telecommunications industry — an application of survival analysis modeling using SAS. In *SAS user group international (SUGI27) proceedings* (pp. 114–127).

Marshall, A. W., & Olkin, I. (2007). *Life distributions* (Vol. 13). Berlin: Springer.

Milhaud, X., & Dutang, C. (2018). Lapse tables for lapse risk management in insurance: a competing risk approach. *European Actuarial Journal*, *8*(1), 97–126.

Milhaud, X., Loisel, S., & Maume-Deschamps, V. (2010). *Surrender triggers in life insurance: classification and risk predictions.* Retrieved from https://www.researchgate.net/publication/41118079_Surrender_Triggers_in_Life_Insurance_Classification_and_Risk_Predictions

Montaseri, M., Charati, J. Y., & Espahbodi, F. (2016). Application of parametric models to a survival analysis of hemodialysis patients. *Nephro-Urology Monthly*, *8*(6).

Outreville, J. F. (1990). Whole-life insurance lapse rates and the emergency fund hypothesis. *Insurance: Mathematics and Economics*, *9*(4), 249–255.

Pérez Marín, A. M. (2006). *Survival methods for the analysis of customer lifetime duration in insurance* (Unpublished doctoral dissertation). Universitat de Barcelona.

Pinquet, J., Guillén, M., & Ayuso, M. (2011). Commitment and lapse behavior in long-term insurance: a case study. *Journal of Risk and Insurance*, *78*(4), 983–1002.

Qi, J. (2009). *Comparison of proportional hazards and accelerated failure time models*

(Unpublished master's thesis). University of Saskatchewan.

Renshaw, A., & Haberman, S. (1986). Statistical analysis of life assurance lapses. *Journal of the Institute of Actuaries*, *113*(3), 459–497.

Richards, S. J. (2011). *Survival models for actuarial work.* Retrieved from http://www.richardsconsulting.co.uk/Survival%20Models%20for%20Actuarial%20Work.pdf

Samuel, C. H., & Lin, Y. (2006). *Annuity lapse rate modeling: Tobit or not tobit.* Retrieved from http://cbafiles.unl.edu/public/cbainternal/researchlibrary/Annuity%20Lapse%20Rate%20Modeling%20-%20Tobit%20or%20Not%20Tobit0.pdf

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Smith, K. A., Willis, R. J., & Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: A case study. *Journal of the Operational Research Society*, *51*(5), 532–541.

Spiteri, M., & Azzopardi, G. (2018). Customer churn prediction for a motor insurance company. In *Thirteenth international conference on digital information management* (pp. 173–178).

Su, J., Cooper, K., Robinson, T., & Jordan, B. (2009). Customer retention predictive modeling in healthcare insurance industry. In *Southeast SAS users group conference proceedings* (pp. 1–8).

Swindell, W. R. (2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental Gerontology*, *44*(3), 190–200.

Xue, Y., & Schifano, E. D. (2017). Diagnostics for the cox model. *Communications for Statistical Applications and Methods*, *24*(6), 583–604.

Appendix

Table 6: Exponential AFT model

Covariate		β	TR (exp(β))	Sig.	95% CI (TR)	
Gender	Male [0]	-0.0487	0.9524	0.003	0.9223	0.9836
	Female [1]					
Payment frequency	infra annual [0]	-0.1424	0.8673	0.0000	0.8328	0.903
	annual [1]					
	other (supra annual) [2]	0.0809	1.0843	0.0056	1.024	1.148
Risk state	smoker [0]	0.0691	1.0715	0.0001	1.0361	1.1085
	non smoker [1]					
Underwriting age	young (0 to 34 years old) [0]	0.0595	1.0613	0.001	1.0243	1.0995
	middle (35 to 54 years old) [1]					
	old (55 to 84 years old) [2]	0.1904	1.2097	0.0000	1.1526	1.27
Living place	East coast [0]					
	West coast [1]	0.0275	1.0278	0.5266	0.9441	1.1185
	other [2]	0.0073	1.0073	0.7206	0.968	1.0481
Accidental death rider	Rider [0]	0.1251	1.1333	0.0000	1.0825	1.1865
	Non Rider [1]					
Annual premium		-0.0839	0.9195	0.0000	0.9062	0.933
Dow Jones Index variation		-0.6592	0.5172	0.0000	0.5086	0.5262
Unemployment rate		-0.4095	0.664	0.0000	0.6531	0.675

Table 7: Weibull AFT model

Covariate		β	TR (exp(β))	Sig.	95% CI (TR)	
Gender	Male [0]	-0.0503	0.951	0.0033	0.9196	0.9834
	Female [1]					
Payment frequency	infra annual [0]	-0.1477	0.8627	0.0000	0.8273	0.8996
	annual [1]					1
	other (supra annual) [2]	0.0843	1.088	0.0055	1.0251	
Risk state	smoker [0]	0.0716	1.0742	0.0001	1.0372	1.1125
	non smoker [1]					
Underwriting age	young (0 to 34 years old) [0]	0.062	1.0639	0.001	1.0254	1.1039
	middle (35 to 54 years old) [1]					
	old (55 to 84 years old) [2]	0.197	1.2177	0.0000	1.1582	1.2804
Living place	East coast [0]					
	West coast [1]	0.029	1.0294	0.5205	0.9423	1.1247
	other [2]	0.0077	1.0077	0.7148	0.9669	1.0503
Accidental death rider	Rider [0]	0.1297	1.1385	0.0000	1.0855	1.1941
	Non Rider [1]					
Annual premium		-0.087	0.9167	0.0000	0.9029	0.9308
Dow Jones Index variation		-0.6789	0.5072	0.0000	0.4977	0.5168
Unemployment rate		-0.4172	0.6589	0.0000	0.6472	0.6708

Table 8: Log-logistic AFT model

Covariate		β	TR (exp(β))	P-value	95% CI (TR)	
Gender	Male [0]	-0.0668	0.9354	0.0008	0.8994	0.9727
	Female [1]					
Payment frequency	infra annual [0]	-0.1715	0.8424	0.0000	0.8027	0.8840
	annual [1]					
	other (supra annual) [2]	0.1140	1.1207	0.0008	1.0483	1.1981
Risk state	smoker [0]	0.0892	1.0933	0.0000	1.0496	1.1388
	non smoker [1]					
Underwriting age	young (0 to 34 years old) [0]	0.0837	1.0873	0.0002	1.0410	1.1357
	middle (35 to 54 years old) [1]					
	old (55 to 84 years old) [2]	0.2199	1.2460	0.0000	1.1762	1.3199
Living place	East coast [0]					
	West coast [1]	0.0475	1.0487	0.3613	0.9469	1.1613
	other [2]	0.0136	1.0137	0.5828	0.9656	1.0642
Accidental death rider	Rider [0]	0.1493	1.1610	0.0000	1.0997	1.2258
	Non Rider [1]					
Annual premium		-0.1168	0.8898	0.0000	0.8728	0.9071
Dow Jones Index variation		-0.6670	0.5132	0.0000	0.5023	0.5244
Unemployment rate		-0.4159	0.6598	0.0000	0.6459	0.6740

Table 9: Log-normal AFT model

Covariate		β	TR (exp(β))	Sig.	95% CI (TR)	
Gender	Male [0]	-0.0721	0.9305	0.0017	0.8896	0.9732
	Female [1]					
Payment frequency	infra annual [0]	-0.1941	0.8236	0.0000	0.7795	0.8702
	annual [1]					
	other (supra annual) [2]	0.1500	1.1618	0.0001	1.0766	1.2537
Risk state	smoker [0]	0.1066	1.1124	0.0000	1.0617	1.1656
	non smoker [1]					
Underwriting age	young (0 to 34 years old) [0]	0.0819	1.0853	0.0013	1.0324	1.1409
	middle (35 to 54 years old) [1]					
	old (55 to 84 years old) [2]	0.2367	1.2670	0.0000	1.1865	1.3530
Living place	East coast [0]					
	West coast [1]	0.0602	1.0621	0.3116	0.9451	1.1935
	other [2]	0.0079	1.0080	0.7797	0.9534	1.0657
Accidental death rider	Rider [0]	0.1771	1.1937	0.0000	1.1221	1.2698
	Non Rider [1]					
Annual premium		-0.1393	0.8700	0.0000	0.8510	0.8894
Dow Jones Index variation		-0.6942	0.4995	0.0000	0.4893	0.5099
Unemployment rate		-0.3825	0.6822	0.0000	0.6668	0.6979