Statistical Inference of Generalized Time-Dependent Logistic Hazard Model

Ahmed A. El-Sheikh[1], Amaal Mubarak[2] and Fatma G. Abdelmaksoud[3]

[1] *Department of Statistics, Institute of statistical studies and Research, Cairo University, Egypt.*

[2] *Department of Statistics, Faculty of Commerce, Damietta University, Egypt.*

[3] *Department of Statistics, Faculty of Commerce, Al- Azhar University, Egypt.*

Abstract

In this paper, we extend the generalized time-dependent logistic hazard model (GTDL) proposed by Mackenzie (1996) by using the hazard function of Weibull distribution instead of the baseline hazard function as no form was assumed in this model. Thus a modified form called "generalized time dependent-Weibull logistic hazard (GTD-WL) model". The maximum likelihood method used to estimate the unknown parameters of the proposed model. Also, we obtained some special cases from this model. A simulation study examines the bias and the mean squared errors of the maximum likelihood estimates for the proposed model comparing with three other models.

1. Introduction

Survival data analyses often consist of a response variable that measures the duration of time until a specific event occurs and a set of variables associated with the event time variable. These data arise in many applied fields, such as medicine, biology, hygienic, engineering, economics, criminology and demography. For these data it is reasonable to presume that the hazard function is time-dependent, thereby accommodating crossing hazards. Such dependency can be modelled directly by introducing a time-dependent term in the model for the hazard function. Accordingly, we utilize a generalized time-dependent logistic (GTDL) hazard model which can accommodate non-proportional hazards data. MacKenzie (1996, 2002) proposed the generalized time-dependent logistic hazard regression (GTDL) model as a wholly parametric competitor for the Proportional Hazards(PH) model. An advantage of the GTDL model is its generalization of the relative risk in the PH model of Cox (1972) to time-dependent form.

The survival function is defined using the hazard function $h(t \mid \theta)$ as the generalized time-dependent logistic hazard family which is given by,

$$h(t \mid \alpha, \beta, x) = h_0(t) \frac{exp(t\alpha + x'\beta)}{1 + exp(t\alpha + x'\beta)} \tag{1}$$

[Mackenzie, 1996]

Where, $h_0(t)$ is the baseline hazard function.

Note that, this hazard function does not assume any functional form for the baseline hazard function.

The main objective of our study is present the generalized time dependent- Weibull logistic (GTD-WL) hazard model which has a specific baseline hazard function where, $h_0(t)$ is defined using the hazard function of Weibull distribution. Then, the statistical inference of the (GTD-WL) model is presented based on the maximum likelihood method to obtain the estimators of the unknown parameters. In addition, some models are derived from our model as special cases. The proposed model is illustrated by a simulation study compared with some other models.

The paper is organized as follows. Section 2 presents the formulations of the GTDL and GTD-WL models. Maximum likelihood method is illustrated to estimate the unknown parameters of the proposed model in section 3. In section 4, a simulation study is conducted to evaluate the performance of the proposed model compared with some other models. Finally, the conclusions are in section 5.

2. Model Formulation

In this section, we introduce the GTDL hazard model and the GTD -WL hazard model based on Type-I generalized logistic model. The hazard, the cumulative hazard, the survival and the probability density functions of the two models are considered. Some special cases are derived from the proposed model.

2.1 The GTDL hazard model

The GTDL model, which is a non-proportional hazard model, is defined by the hazard function given in (1) by Mackenzie, (1996). Then, this hazard function (1) can be written as follows,

$$h(t|\alpha,\beta) = h_0(t)\frac{1}{1+exp\left\{-\left(\alpha t+x'\beta\right)\right\}} \qquad (2)$$

Where, t is a non-negative random variable representing the univariate survival times of the units, $h_0(t)$ is an arbitrary function of time called "baseline hazard function", α is a measure of the time effect and, $\beta' = \left(\beta_1,\beta_2,\ldots,\beta_p\right)$ is a vector of p unknown parameters measuring the influence of the p covariates $x' = \left(x_1,x_2,\ldots,x_p\right)$.

The hazard function (2) can be generalized by replacing its logistic distribution function by a generalization of the logistic distribution called Type-I generalized logistic which is given by,

$$F(Y) = \frac{1}{\left(1+exp\left(-Y\right)\right)^{b}}, \ -\infty<Y<\infty, \ b>0. \qquad (3)$$

By using this form, and assuming that $t=1$ in the logistic distribution in equation (2), we can define the hazard function as follows,

$$h\left(t|\,\alpha,\beta,b\right)=\begin{cases} h_0(t)g(\alpha,\beta,b) & ,b>0 \\ h_0(t) & ,b=0 \end{cases} \qquad (4)$$

Where,

$$g\left(\alpha,\beta,b\right)=\frac{1}{\left[1+exp\left\{-\left(\alpha+x'\beta\right)\right\}\right]^{b}}=\left[1+exp\left\{-\left(\alpha+x'\beta\right)\right\}\right]^{-b}$$

And the survival function corresponding to the hazard function in (4) is,

$$S\left(t|\,\alpha,\beta,b\right)=exp\left[-H(t|\,\alpha,\beta,b)\right] \qquad (5)$$

Where,

$H(t|\,\alpha,\beta,b)=\int_0^t h(t|\,\alpha,\beta,b)dt$ is the cumulative hazard function.

Thus, we get the probability density function of the (GTDL) model based on a generalization of the logistic distribution from (4) and (5) as follows,

$$f\left(t|\,\alpha,\beta,b\right)=S\left(t|\,\alpha,\beta,b\right)h\left(t|\,\alpha,\beta,b\right)$$

Then,

$$f\left(t|\,\alpha,\beta,b\right)=exp\left\{-H\left(t|\,\alpha,\beta,b\right)\right\}h\left(t|\,\alpha,\beta,b\right) \qquad (6)$$

2.2 The GTD-WL hazard model

The baseline hazard in the hazard function (4) is not specified and it's unknown functional form (i.e., it does not make any assumptions about its form), so the model (6) is a semi-parametric model because the hazard function $h\left(t \mid \alpha,\beta,b\right)$ has parametric and nonparametric components [the parametric component is a type-I generalized logistic $g\left(\alpha,\beta,b\right)$ and the nonparametric component is the baseline hazard $h_0\left(t\right)$].

If we assume that this baseline hazard function is specified up to a few unknown parameters, we can use a specific parametric distribution such as the Weibull distribution that describe the times of failure. In this case, we get a generalized time-dependent Weibull-Logistic (GTD-WL) hazard model with the following hazard function,

$$h\left(t \mid \lambda,\gamma,\alpha,\beta,b\right) = \lambda\gamma t^{\gamma-1} g\left(\alpha,\beta,b\right) \quad ,t>0 \qquad (7)$$

Where, the baseline hazard function is, $h_0\left(t\right) = \lambda\gamma t^{\gamma-1}$, therefore, the baseline cumulative hazard function is, $H_0\left(t\right) = \int_0^t \lambda\gamma t^{\gamma-1} dt = \lambda t^{\gamma}$.

The cumulative hazard, survival and density functions of the (GTD-WL) model corresponding to the hazard function (7) are given by,

- The cumulative hazard:

$$H\left(t|\lambda,\gamma,\alpha,\beta,b\right)=\int_0^t h\left(t|\lambda,\gamma,\alpha,\beta,b\right)\mathrm{d}t$$
$$=\int_0^t \lambda\gamma t^{\gamma-1}g\left(\alpha,\beta,b\right)\mathrm{d}t=\lambda\gamma g\left(\alpha,\beta,b\right)\frac{t^{\gamma}}{\gamma}$$

Then,

$$H\left(t|\lambda,\gamma,\alpha,\beta,b\right)=\lambda g\left(\alpha,\beta,b\right)t^{\gamma} \tag{8}$$

- The survival function:

$$S\left(t|\lambda,\gamma,\alpha,\beta,b\right)=exp\left\{-H\left(t|\lambda,\gamma,\alpha,\beta,b\right)\right\}$$

Then,

$$S\left(t|\lambda,\gamma,\alpha,\beta,b\right)=exp\left\{-\lambda g\left(\alpha,\beta,b\right)t^{\gamma}\right\} \tag{9}$$

- The density function:

$$f\left(t|\lambda,\gamma,\alpha,\beta,b\right)=h\left(t|\lambda,\gamma,\alpha,\beta,b\right)S\left(t|\lambda,\gamma,\alpha,\beta,b\right)$$

Then,

$$f\left(t|\lambda,\gamma,\alpha,\beta,b\right)=\lambda\gamma t^{\gamma-1}g\left(\alpha,\beta,b\right)exp\left\{-\lambda g\left(\alpha,\beta,b\right)t^{\gamma}\right\} \tag{10}$$

where, $t>0$ is a positive variable represents the survival times of units, $\lambda,\gamma,\alpha>0$ are scalar values, β is a vector of unknown parameters and b is the shape parameter of type-I generalized distribution.

2.3 Special Cases

Some special cases can be derived from the probability density function (10) of the proposed model (GTD-WL) such as, the generalized time dependent-exponential logistic (GTD-EL) model and the generalized time dependent-Rayleigh (GTD-RL) model.

- **The (GTD-EL) model**

If we put $\gamma = 1$ in the p.d.f (10), we obtain the (GTD-EL) model as follows,

$$f\left(t|\lambda,\alpha,\beta,b\right) = \lambda g\left(\alpha,\beta,b\right) exp\left\{-\lambda g\left(\alpha,\beta,b\right)t\right\} \quad (11)$$

- **The (GTD-RL) model**

If we put $\gamma = 2$ in the p.d.f (10), we obtain the (GTD-RL) model as follows,

$$f\left(t|\lambda,\alpha,\beta,b\right) = 2\lambda t g\left(\alpha,\beta,b\right) exp\left\{-\lambda g\left(\alpha,\beta,b\right)t^{2}\right\} \quad (12)$$

Thus, the model in (10) is reduced to four parameters.

3. Estimation of the (GTD-WL) model

The maximum likelihood method will be used to estimate the unknown parameters $\left(\lambda,\gamma,\alpha,\beta,b\right)$ for this model. Consider a sample of n independent individuals with data $\left(t_i, x_i, \delta_i\right)$ where, $\delta_i = 1$ for an event and 0 otherwise, for $i = 1,\ldots,n$.

Let, $\theta = \left(\lambda,\gamma,\alpha,\beta,b\right)$ be denote the vector of unknown parameters. Then, the likelihood function for θ is given by,

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} h\left(t_i\,|\,\theta\right)^{\delta_i} S\left(t_i\,|\,\theta\right)$$

(Lawless, 2003)

Thus, the likelihood function for the (GTD-WL) hazard model is given by,

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} \left[\lambda\gamma t_i^{\gamma-1} g\left(\alpha,\beta,b\right)\right]^{\delta_i} \left[exp\left\{-\lambda g\left(\alpha,\beta,b\right)t_i^{\gamma}\right\}\right] \quad (13)$$

Then, the ln-likelihood function $\ell(\theta)=\ln\left[\mathrm{L}(\theta)\right]$ becomes,

$$\ell(\theta)=\sum_{i=1}^{n}\left[\delta_i\ln\left\{\lambda\gamma g(\alpha,\beta,b)t_i^{\gamma-1}\right\}-\left\{\lambda g(\alpha,\beta,b)t_i^{\gamma}\right\}\right] \quad (14)$$

To obtain the maximum likelihood estimates (MLEs), the ln-likelihood function in (14) must be maximized by using a procedure for constrained optimization. Then, in order to maximize (14), we obtain its first derivatives with respect to all the unknown parameters as follows,

$$\left.\begin{aligned}
\frac{\partial\ell(\theta)}{\partial\lambda}&=\sum_{i=1}^{n}\left[\frac{\delta_i}{\lambda}-g(\alpha,\beta,b)t_i^{\gamma}\right]\\
\frac{\partial\ell(\theta)}{\partial\gamma}&=\sum_{i=1}^{n}\left[\delta_i\left(\ln(t_i)+\frac{1}{\gamma}\right)-\lambda g(\alpha,\beta,b)t_i^{\gamma}\ln(t_i)\right]\\
\frac{\partial\ell(\theta)}{\partial\alpha}&=\sum_{i=1}^{n}\left[\frac{b}{1+exp(\alpha+x'\beta)}\left(\delta_i-\lambda t_i^{\gamma}\left[1+exp\left\{-(\alpha+x'\beta)\right\}\right]^{-b}\right)\right]\\
\frac{\partial\ell(\theta)}{\partial\beta}&=\sum_{i=1}^{n}\left[\frac{bx}{1+exp(\alpha+x'\beta)}\left(\delta_i-\lambda t_i^{\gamma}\left[1+exp\left\{-(\alpha+x'\beta)\right\}\right]^{-b}\right)\right]\\
&\text{And,}\\
\frac{\partial\ell(\theta)}{\partial b}&=\sum_{i=1}^{n}\left[\ln\left(1+exp\left\{-(\alpha+x'\beta)\right\}\right)\left\{\lambda t_i^{\gamma}\left[1+exp\left\{-(\alpha+x'\beta)\right\}\right]^{-b}-\delta_i\right\}\right]
\end{aligned}\right\} \quad (15)$$

4. Simulation Study

In this section, a simulation study has been constructed to obtain the bias, the percentage of relative bias (RB%), mean squared error (MSE) and the relative mean squared error (RMSE) of the MLEs for all the unknown parameters of the (GTD-WL) hazard model compared with three other models, namely, weibull generalized logit-link proportional hazard (WGL-

PH) model, weibull proportional hazard (W-PH) model and generalized time-dependent logistic (GTDL) hazard model.

1000 replications have been generated for each sample size (n = 20, 50, 100 and 200). The simulation study is conducted by assuming the following initial values of the parameters $(\lambda = 0.5,\ \gamma = 0.5,\ \alpha = 0.1,\ b = 0.5)$ as a part of many other values tested in the program and,

$\beta' = (\beta_1 = 0.5, \beta_2 = -0.5, \beta_3 = 1, \beta_4 = -1)$ based on [Khan & Khosa, 2016]. A dummy covariate values are generated from a Binomial distribution with $p = 0.5$. The n random numbers are generated from the Uniform distribution (0, 1). For each sample size we compute the bias, (RB %), MSE and (RMSE). The Bias and RB% for θ are, respectively, defined as;

$$\text{Bias}(\theta) = \overline{\left(\hat{\theta}\right)} - \theta \quad \text{and} \quad \text{RB\%} = \frac{Bias(\theta)}{\theta} \times 100$$;[Ha, and Mackenzie, 2010]

where, $\overline{\left(\hat{\theta}\right)}$ is the mean of $\hat{\theta}$ ($\hat{\theta}$ is the ML estimate of θ).

And MSE and (RMSE) for θ are, respectively, defined as;

$$\text{MSE}(\theta) = \text{E}\left[\left(\hat{\theta} - \theta\right)^2\right] \quad \text{and} \quad \text{RMSE}(\theta) = \frac{\text{MSE}(\theta)}{\text{MSE}^*(\theta)}$$

[Shcherbakov, , et al, 2013]

where, $\text{MSE}^*(\theta)$ is the (MSE) of the (GTD-WL) model.

All of the calculation in this section were done using MATHCAD program version 2007.

The RB% and (RMSE) of the parameters of all models are presented in Table 1 for small and moderate sample size (n = 20, 50) and Table 2 for the large sample sizes (n = 100, 200).

For parameter λ, the WGL-PH model has the least (RMSE) and the W-PH has the least RB%. With respect to parameter γ, it does not exist in the GTDL model, the W-PH model has the least RB% and the GTD-WL model has the least (RMSE). For the parameter α, it exists only in the GTD-WL and GTDL models. So, the results show that the GTD-WL model has the least RB% and the GTDL model has the least (RMSE). With respect to parameter b, it only exists in the GTD-WL and WGL-PH models. The results showed that the WGL-PH model has the least RB% and (RMSE). For the vector of the parameters β', it can be concluded that, β_1 and β_3 have the same behavior and the same signal of the RB%. The GTD-WL model has the least RB% and (RMSE) in almost cases. Similarly, for β_2 and β_4.

5. Conclusion

In this paper, we consider the GTD-WL model [based on Type-I generalized logistic distribution] which assume that the baseline hazard function is specific with the hazard function of the Weibull distribution, finding its hazard, cumulative hazard, survival and probability density functions. We use the ML method to estimate the unknown parameters of the proposed model. Some models can be derived from the GTD-WL model as special cases. The simulation study shows that, the proposed model has the least RB% and (RMSE) in almost cases compared with the other three models.

Table 1. The Bias, RB%, MSE and RMSE of the parameters for the four models.

n	θ	GTD-WL				WGL-PH				W-PH				GTDL			
		Bias	RB%	MSE	RMSE	Bias	RB%	MSE	RMSE	Bias	RB%	MSE	RMSE	Bias	RB%	MSE	RMSE
20	λ	-0.189	-18.9	4.669	1	-0.605	-60.54	0.452	0.1	-1.576	-157.6	3.189	0.68	-0.79	-79.00	0.64	0.14
	γ	-0.447	-44.7	0.213	1	-0.451	-45.1	0.217	1.02	-1.079	-215.9	1.471	6.91	—	—	—	—
	α	-0.478	-95.5	6.752	1	—	—	—	—	—	—	—	—	0.49	490.35	5.081	0.75
	β_1	-0.299	-59.85	0.112	1	-0.171	-34.2	0.043	0.38	-0.105	-21.05	0.026	0.23	-0.138	13.8	0.035	0.31
	β_2	0.2	-39.9	0.05	1	0.257	-51.3	0.096	1.92	0.158	-31.58	0.059	1.18	0.207	-41.4	0.078	1.56
	β_3	-0.599	-59.85	0.449	1	-0.342	-34.2	0.171	0.38	-1.342	-134.2	0.105	0.23	-0.276	-27.6	0.138	0.31
	β_4	0.399	-39.9	0.2	1	0.513	-51.3	0.385	1.93	0.316	-31.58	0.237	1.19	0.414	-41.4	0.311	1.56
	b	-0.715	-143	15.2	1	-0.525	-105	0.832	0.05	—	—	—	—	—	—	—	—
50	λ	-0.576	-57.59	0.45	1	-0.63	-63.03	0.416	0.92	-0.985	-98.52	0.973	2.16	-0.807	-80.71	0.656	1.46
	γ	-0.49	-49.01	0.244	1	-0.491	-49.15	0.245	1	-0.647	-129.4	0.452	1.85	—	—	—	—
	α	-0.289	-57.77	4.174	1	—	—	—	—	—	—	—	—	0.024	23.892	0.783	0.19
	β_1	-0.291	-58.2	0.109	1	-0.172	-34.4	0.043	0.39	-0.174	-34.8	0.044	0.4	-0.166	16.6	0.042	0.39
	β_2	0.194	-38.8	0.049	1	0.258	-51.6	0.097	1.98	0.261	-52.2	0.098	2	0.249	-49.8	0.093	1.9
	β_3	-0.582	-58.2	0.437	1	-0.344	-34.4	0.172	0.39	-1.239	-123.9	0.174	0.4	-0.332	-33.2	0.166	0.38
	β_4	0.388	-38.8	0.194	1	0.516	-51.6	0.387	1.99	0.522	-52.2	0.392	2.02	0.498	-49.8	0.374	1.93
	b	-0.554	-110.7	3.809	1	-0.486	-97.25	0.384	0.1	—	—	—	—	—	—	—	—

Table 2. The Bias, RB%, MSE and RMSE of the parameters for the four models.

n	θ	GTD-WL				WGL-PH				W-PH				GTDL			
		Bias	RB%	MSE	RMSE	Bias	RB%	MSE	RMSE	Bias	RB%	MSE	RMSE	Bias	RB%	MSE	RMSE
100	λ	-0.623	-62.27	0.424	1	-0.641	-64.07	0.418	0.99	-0.969	-96.91	0.941	2.22	-0.813	-81.25	0.662	1.56
	γ	-0.503	-50.29	0.254	1	-0.503	-50.34	0.255	1	-0.556	-111.1	0.323	1.27	—	—	—	—
	α	-0.239	-47.75	2.717	1	—	—	—	—	—	—	—	—	-0.102	-102.2	0.084	0.03
	β_1	-0.274	-54.9	0.103	1	-0.17	-34.05	0.043	0.42	-0.226	-45.25	0.057	0.55	-0.183	18.3	0.046	0.45
	β_2	0.183	-36.6	0.046	1	0.255	-51.08	0.096	2.09	0.339	-67.88	0.127	2.76	0.274	-54.9	0.103	2.24
	β_3	-0.549	-54.9	0.412	1	-0.341	-34.05	0.17	0.41	-1.161	-116.1	0.226	0.55	-0.366	-36.6	0.183	0.44
	β_4	0.366	-36.6	0.183	1	0.511	-51.08	0.383	2.09	0.679	-67.88	0.509	2.78	0.549	-54.9	0.412	2.25
	b	-0.533	-106.6	1.269	1	-0.501	-100.3	0.315	0.25	—	—	—	—	—	—	—	—
200	λ	-0.639	-63.92	0.416	1	-0.643	-64.29	0.418	1.005	-0.898	-89.77	0.807	1.94	-0.814	-81.41	0.664	1.6
	γ	-0.506	-50.59	0.257	1	-0.506	-50.61	0.257	1	-0.428	-85.52	0.183	0.71	—	—	—	—
	α	-0.198	-39.60	1.332	1	—	—	—	—	—	—	—	—	-0.123	-122.8	0.017	0.01
	β_1	-0.257	-51.45	0.096	1	-0.168	-33.55	0.042	0.44	-0.25	-49.95	0.062	0.65	-0.189	18.875	0.047	0.49
	β_2	0.171	-34.3	0.043	1	0.252	-50.33	0.094	2.19	0.331	-66.28	0.113	2.63	0.283	-56.63	0.106	2.47
	β_3	-0.514	-51.45	0.386	1	-0.336	-33.55	0.168	0.44	-1.169	-116.9	0.25	0.65	-0.377	-37.75	0.189	0.49
	β_4	0.343	-34.3	0.172	1	0.503	-50.33	0.377	2.19	0.663	-66.28	0.454	2.64	0.566	-56.63	0.425	2.47
	b	-0.502	-100.5	0.669	1	-0.5	-99.96	0.284	0.42	—	—	—	—	—	—	—	—

References

Balakrishnan, N., Pardo, M.C. and Avendano, M.L. (2013). Generalized Logit-Based Proportional Hazards Models and Their Applications in Survival and Reliability Analyses. In Stochastic Reliability and Maintenance Modeling. Springer, London, pp.1-16.

Cox, D.R. (1972). Regression Models and Life-Tables (with Discussion) JRSS B.34, pp.187-220.

Ha, I.D. and Mackenzie, G. (2010). Robust Frailty Modelling using Non-Proportional Hazards Models. In: Statistical Modelling, Vol.10. No.3, pp. 315-332.

Khan, S.A. and Khosa, S. (2016). Generalized Log-Logistic Proportional Hazard Model with Applications in Survival Analysis. Journal of Statistical Distributions and Applications, Vol.3, No.16.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Mackenzie, G. (1996). Regression Models for Survival Data: The Generalized Time-Dependent Logistic Family. The Statistician, Vol.45, No.1, pp.21-34.

Mackenzie, G. (2002). A Logistic Model for Survival Data. In: Proceedings of the 17th International Workshop in Statistical Modeling, pp.431-438.

Shcherbakov, M.V., Brebels, A., Shcherbakova, N.L., Tyukov, A.P., Janovsky, T.A. and Kamaev, V.A. (2013). A Survey of Forecast Error Measures. World Applied Sciences Journal, Vol.24, pp.171-176.