A Parametric Modeling of Factors Associated With Teachers' Exit from Service in Kenya

Jerita Jemimah Mwambi[1], Leonard Kiti Alii[2]

[1]Pwani University, School of Humanities and Social Sciences, P.O Box 195, Kilifi - 80108, Kenya

Abstract: *The Survival analysis methods have been widely applied in studying factors associated with duration in employment. This study implemented an accelerated failure time parametric Weibull model to determine the factors associated with teachers' exit from service in Kenya; using data from the Teachers' Service Commission of 178,063 teachers who were still working in public schools and institutions up to October 2014 and 28,403 teachers who had already exited from the profession. The accelerated failure time Weibull and log-logistic models were fitted to study data. To find the best fitting model we used graphical methods and the Akaike's Information Criterion (AIC). The results indicated that the accelerated failure time Weiibull model was the best fitting model to the study data, with the least AIC value. Results of the Weibull model revealed that age, gender, Job group, province or region where the teacher was working and type of exit were the factors that have an effect on the survival time of teachers in service. Therefore accelerated failure time models can be applied to a survival data with even better and more accurate estimates.*

Keywords: Survival time, Accelerated Failure Time model, Time to exit.

1. Introduction

Survival analysis methods are concerned with analysis of times from specific time origin until some critical time event. These methods have been applied in various areas including medicine, biology, engineering, biostatistics, socials sciences and many others [11]. In the employment context, this may be time to exit from employment and this could be through resignation, retirement, dismissal, death and so on.

Due to censoring and lack of normality of survival data, normal statistical methods like linear regression and logistic regression cannot be used. Methods used in the analysis of survival data include non-parametric, semi-parametric proportional hazards and parametric methods including Accelerated Failure Time (AFT) models. AFT parametric models allow the use of an acceleration factor or time ratio which can easily be interpreted as compared to a ratio of two hazards [3].

Teacher attrition and factors associated with it has been a subject of interest to educational researchers. Teacher attrition and retention is a worldwide problem facing educational planners and administrators in both developing and developed countries. Attrition has been a problem according to research since 1970s, for example [4], [13] and [14].

[18] was among the first researcher to use the concept of survival rates to teaching; he examined the survival rates of 937 public school teachers in Missouri from 1951 to 1953. [4] also applied survival rates to study 2,024 teachers who begun working in Oregon in 1962. They both used a 5-year time frame; their general results were very similar in that a high proportion of those who join teaching in any given year do not come back the following year. They found a high rate of attrition during the first years of teaching, then it slowed down and eventually the process of attrition became stable. [13] examined the survival rates of public school teachers from 1968 to 1976. They looked at the survival behavior of

each cohort of new entrants to the teaching profession from 1968 to 1975. They were able to examine how survival rates change over a period of time and still found a downward sloping survival curve just as Charter's but the curve had moved steadily upward with time. According to their findings, 64.7% of the teachers survived beyond one year for the 1965 entrants while 83.6% survived for the 1973 entrants.

In a similar study to update the 1978 study, [12] examined the lengths of employment of 14,827 teachers who were hired by districts in St. Louis, Missouri area from 1969 – 1982. They were able to examine how survival rates change over a period of time.

[8] used the National longitudinal study which included data of teachers of 14 years experience and found that attrition rates were high particularly in the first 3 or 4 years of service.

Research reveals that personal characteristics of teachers, school working conditions contribute to teachers' decision to exit from service [2], [9], [7] and [6].

[15] employed non-parametric methods and the semi-parametric Cox Proportional hazard (PH) model to estimate the time to exit of teachers from service in Kenya.

[1] also employed the application of the Cox proportional hazards model to the study of explanatory factors influencing the survival of elementary teachers in Texas.

[16] implemented the AFT and proportional hazards models and found the AFT model to be more appropriate in the influenza study. The log-logistic was the best fitting model.

The aim of this paper is to implement the AFT model to the study data to determine factors associated with teachers' exit from service.

2. Materials and Methods

This study used data from the Teachers Service Commission (TSC) of all teachers employed by the government through the TSC. The data consisted of 278,063 teachers employed by the TSC in all public schools and institutions and 28,403 teachers who had left the teaching profession. [15]; however teachers with missing or wrongly entered information, were removed from the data. For every teacher information concerning their age, date of hire, date of exit and reason for exit (for those who had exited), Job group, salary, gender, and province or region where the teacher was working was availed. The number of years the teacher had worked for under the TSC, which is our response variable was also included as the survival time, along with the censoring indicator (0) for teachers who were still working (censored) and (1) for those who had already exited.

This study considered the explanatory variables age (coded as 0 for those 35 years and below, as 1 for those aged above 35 years old); gender coded as M for male teachers and F for female teachers. Job group covariate initially had ten levels from group F to R and were reduced to four levels (A for combined group F, G and H; B for groups J, K and L; C for M and N; D for groups P, Q and R); salary in Kenya shillings was allocated to each teacher using the job groups; we however used the log (salary) in our model. Wprovince represented the region where the teacher was working, we used the eight provinces in the country; covariate exit represents the reasons for exit, 0 denoting death, 1 for retire mandatory and contract expiry, 2 for dismissal and retire on public interest and 3 for early exit such as retire on medical grounds, transfer service, resignation, retire at 50 and voluntary early retirement.

3. Methodology

Parametric AFT models can provide precise estimates of the quantities of interest and are more robust as compared to the proportional hazards model, provided the model is chosen correctly. They model the effect of the covariates directly on the survival time.

Under AFT models, the survival times are assumed to follow a parametric distribution which must be specified; for example the Weibull, the exponential, log-logistic and lognormal. The AFT model is usually written in log-linear form with respect to time as

$$\log T_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots + \beta_p x_{ip} + \sigma \varepsilon_i \quad (1)$$

where ε_i is the residual, a measure of variability in the survival times following some distribution. $\alpha \ and \ \sigma$ are the intercept and scale parameters respectively.
$x_1, x_2 \ldots\ldots, x_{ip}$ is set covariates of a group of teachers and $\beta_1, \beta_2 \ldots\ldots, \beta_{ip}$ are the unknown regression parameters of the covariates. Mathematically the AFT model can also be written as

$$S(t) = S_0(\phi t) \quad (2)$$

where $S_o(t)$ is the baseline survivor function and ϕ is the acceleration factor which can also be expressed as

$$\varphi = \exp\ (\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots + \beta_p x_{ip})\ [5]$$

In this study we implemented the Weibull AFT and the Log-logistic AFT approaches to model the covariates associated with each teacher with the survival time in service. The best model was selected using the Akaike's information criterion (AIC).

4. Model Checking

The first approach used to check whether an AFT model is appropriate is to use a quantile-quantile (QQ) plot .Plotting the percentiles of the Kaplan-Meier (KM) estimated survival function from one group against another should produce a straight line passing through the origin with slope ϕ .

We also checked the distribution assumption for the Weibull and log-logistic model by using univariate data. We first transformed the cumulative hazard function H of the assumed distribution such that the transformed H that is, $\psi\big(H(t)\big)$ is a linear function of some time scale $l(t)$. We then plotted $\psi\big(\hat{H}(t)\big)$ against $l(t)$, the log of time, where $\hat{H}$ is the Nelson-Aalen estimator of H .

The statistical analysis was done using R 3.1.3 software [17].

5. Results

Figure 1 shows the Q-Q plot of the Weibull distribution, the plot is approximately a straight line through the origin.

The univariate analysis for each covariate was carried out using the AFT Weibull model. All the variables were significant at P-value < 0.05 and were used in fitting the Weibull model. The StepAIC function of the Mass library was used for variable selection. The weibull model and Log-logistic models were fitted to the data.
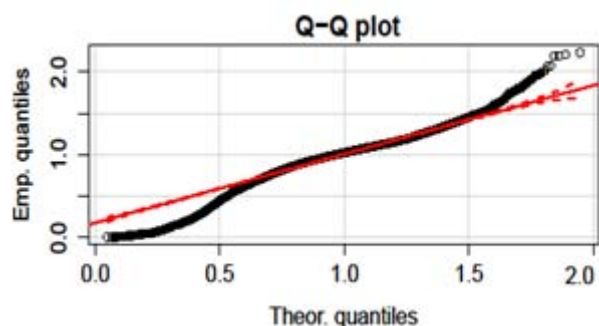

Figure 1: Quantile-Quantile plot for Weibull AFT distribution

Figures 2 shows the plots of the transformed cumulative hazard functions, against some time scale for checking the Weibull distribution assumption. Plot produces approximately straight parallel lines for two groups of male

Paper ID: NOV151421

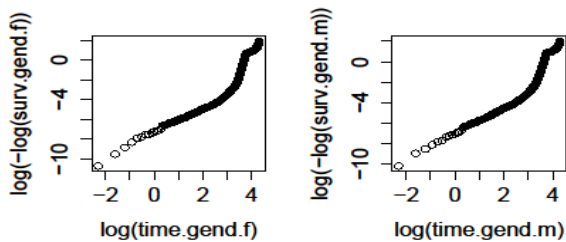1650

and female teachers.

Figure 2: Weibull assumption: transformed cumulative hazard function against log time.
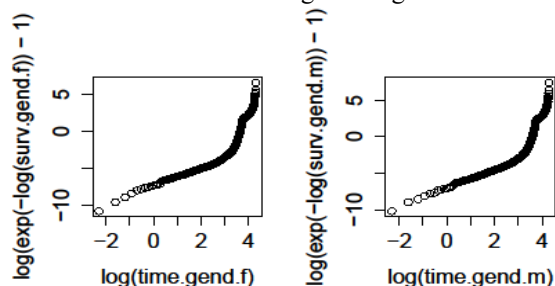

Figure 3: Log-logistic assumption: transformed cumulative hazard function against log time.

Figure 3 shows the log-logistic diagnostic plot. The lines of the plot are curved at some point.

Results from the AIC values of the AFT Weibull model and log-logistic model are 198985.8 and 202611.2 respectively.

Table 1: Results of AFT Weibull model

variable	estimate	Time Ratio	95% C.I	P value
Intercept	-0.0612		-0.229,0.107	0.474
Age				
<= 35 yrs	ref	1		
>35 yrs	1.313	3.718	1.297,1.330	<0.001
gender				
F (female)	ref	1		
M(male)	-0.049	0.952	-0.057,-0.041	1.60e-33
log(salary)	0.209	1.233	0.192,0.226	6.84e-125
Job group				
A	ref	1		
B	-0.239	0.788	-0,252,-0.226	1.63e-274
C	-0.282	0.754	-0.263,-0.301	2.06e-187
D	-0.315	0.730	-0.271,-0.359	8.03e-45
wprovince				
Central	ref	1		
Coast	0.003	1.003	-0.015,0.021	7.82e-01
Eastern	-0.004	0.996	-0.017,0.009	5.33e-01
Nairobi	-0.015	0.985	-0.033,0.003	0.108
N.Eastern	-0.083	0.920	-0.125,-0.041	1.23e-04
Nyanza	0.027	1.028	0.014,0.041	6.81e-05
Rift valley	-0.009	0.991	-0.021,0.003	0.140
Western	-0.017	0.983	-0.029,-0.004	8.66e-03
Exit types				
exit0(death)	ref	1		
exit 1	0.283	1.327	0.273,0.292	<0.001
exit 2	-0.155	0.856	-0.174,-0.137	1.22e-59
Exit 3	0.095	1.099	0.084,0.105	1.03e-74
Log scale	-1.166		-1.175,-1.157	<0.001
Scale	0.312			

Table 1 above shows the results of the multiple analyses of covariates using AFT Weibull model. Most of the covariates are highly significant, gender (P value = 1.60e-33, Time ratio (TR) = 0.9524), Job group B for combined groups J,K,L (P value = 1.63e-274, TR= 0.7876) Job group D for groups P,Q,R (P value 8.03e-45, TR = 0.7300), log(salary) had P value 6.84e-125, TR= 1.233,however , only North Eastern, Nyanza, and Western   provinces were   significantly associated with survival time in service of teachers.
The types of exits are significant as shown in table 1 above.

6. Discussion

The Q-Q plot in figure 1 of the Weibull model produced approximately a straight line passing through the origin; hence the AFT Weibull assumption was met. The survival data can be modeled using a parametric model if the data follows a certain distribution instead of using the semi-parametric Cox model and the Weibull Model is the most common AFT model [10].
The univariate analysis of the AFT Weibull model indicated that variables age, salary, gender, job groups, province, and types of exit are significant [6] and [9]. They were therefore included in the model.

Comparing the transformed cumulative hazard function against log of time of Weibull model, with that of the Log-logistic model indicates that the Weibull model is more appropriate for this study than the log-logistic model (figure 2).

The Weibull model had the lowest AIC value (198985.8) compared to that of the   log-logistic model of 202611.2. This also indicates that the Weibull model is more appropriate for this study data.

Analysis from Table 1 indicates that age greater than 35 years accelerates the average survival time of a teacher 3.72 fold compared to teachers aged 35 years or less; where as being a male teacher shortens the survival time by a factor of 0.95 as compared to the female teachers.
Younger teachers who join the teaching profession tend to leave early during their career [1], [15].

Job group B for groups J, K and L decelerates the survival time of teachers by 0.79 fold to that of the lower job group.
Job group D for groups P, Q and R shortens the average survival time by 0.7300 fold   to that of the lower job group. This could be attributed to the fact that, teachers with higher job groups may be in a better position to take up other non teaching jobs within the public service or other private institutions; also most of the teachers in these job groups have already worked for many years and are almost in the retirement bracket.

Salary is an important predictor to teachers' time to exit as evidenced by past studies [7], [15]. In this study salary increased the average survival time in service by 1.233 fold.
The working conditions of the teacher also influence the teacher's decision to leave or exit. Working in North Eastern Province of Kenya, shortens the average survival time of a

Paper ID: NOV151421

1651

teacher by 0.9204 holding other factors constant, while working in Nyanza province increases the survival time by 1.028 fold. North Eastern province is an arid and semi-arid region in addition to insecurity problem which may trigger the decision of teacher to leave the profession especially when a transfer to another school is not forth coming.

According to this study, holding other factors constant, dismissal of teachers have a higher impact on reducing the survival time of teachers.

7. Conclusion

A parametric AFT model like the Weibull model can be applied to a survival data provided the assumption has been met with even better and more accurate estimates [3]. In this study we implemented the AFT Weibull model to the study data to model the factors associated with teachers exit from service. The AFT model makes it possible to interpret results in terms of the survival time which is very applicable to this study. From this study it is evident that personal characteristics of teachers greatly influence the survival time of teachers in service as well as the compensation and conditions where the teacher is working. These findings can be of benefit to the policy makers and educational planners to come up with well informed strategies to counter the low retention of teachers in the country.

More work could be done in this area with the availability of more detailed data. There is need of more detailed data on teacher attrition, where teachers are tracked down from the time they enter the profession to the time of exit, along with their subject specialties, type of school, reasons of exit and their new station, as already stated in [15].

Further work could be done on the comparison between the Semi-parametric Cox proportional hazard model and the parametric AFT models.

8. Acknowledgement

References

[1] Adams, G.J. "Using a Cox Regression Model to Examine Voluntary Teacher Turnover," *Journal of Experimental Education,* 64, 267-285, 1996.

[2] Bowman, G.D., & Dowling, N.M. "Teacher attrition and retention: A meta-analytic and narrative review of the research," *Review of Educational Research*, **78**, 367 – 411, 2008.

[3] Bradburn MJ, Clarke TG, Love S, Altman DG,"Survival Analysis. Part 11: Multivariate data analysis- an introduction to concepts and methods." *Br J cancer* 89(3):431-436, 2003.

[4] Charters, W.W.jr, "Some factors affecting survival in school districts," *American Educational Research Journal,* Vol.7, No.1, pp.1-27, 1970.

[5] Collet, D .*Modelling Survival data in medical research*, London: Chapman and Hall, 2003.

[6] Crandell, R.M., & Howell, S.J. "An Analysis of teacher Choices: Transfers, Attrition and Retirement," California State University, 2009. [Online] Available: http://www.csus.edu/indiv/h/howellj/papers/TeacherAttrition_Jan2009.pdf (Accessed: October 24, 2013).

[7] Guarino, C. M., Santibanez, L., and Daley, G.A., & Brewer.D. "A Review of the research literature on teacher recruitment and retention," (TR-164- EDU). Santa Monica, CA: RAND, 2004.

[8] Heyenes, B. Educational defectors: "A first look at teacher attrition in the NLS-72," *Educational Researcher*, **17**(3): 24-32, 1988.

[9] Ingersoll, R.M. "Teacher Turnover and Teacher Shortages: An organizational analysis". *American Educational Research Journal,* fall 2001, Vol. 38, (3), 499- 534, 2001.

[10] Klein.J.P & Moeschberger .M.L, Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer, 2003.

[11] Kleinbaum D.G and Klein M.M, Survival Analysis: A Self-learning Text, second edition, Springer, New York, 2005.

[12] Mark J.H & Anderson B.D, "Teacher Survival Rates in St. Louis, 1969-1982.,"American Educational Research Journal, 22, 413-421, 1985.

[13] Mark, J.H., and Anderson, B.D. "Teacher survival rates – A current look," *American Educational Research Journal, **15**, 379-383,* 1978.

[14] Murnane, R .J,"Teacher Mobility revised," *The Journal of Human Resources*, 16(1), 3-19, 1981.

[15] Mwambi J.J & Alii L.K, "A Semi-Parametric Estimation of Time to Exit from Service of Teachers in Kenya", *The International Institute for Science, Technology and Education,* Vol. 5, (18), 2225- 0484, 2015.

[16] Patel K., kay R., Rowell L., "Comparing proportional hazards and Accelerated Failure Time Models: An Application in Influenza,"Pharmaceut Statist 5; 213-224, 2006.

[17] R Development Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[18] Whitener J, E (1965), "An Actuarial Approach to Teacher Turnover." Unpublished Doctoral Dissertation. St. Louis, MO: Washington University.