AALEN'S ADDITIVE HAZARD MODEL: AN APPLICATION IN ANALYZING JOB TENURE FACTORS IN PRIVATE AND PUBLIC COMPANIES IN ALBANIA

Lule Basha[1], Llukan Puka[2, 3]

[1, 2] Department of Applied Mathematics,

[1, 2] Faculty of Natural Science, University of Tirana, Tirana, Albania

[3] Canadian Institute of Technology, St. Andon Zako Çajupi, Zayed Center, Tirana 100, Albania

E-mail: lule.hallaci@fshn.edu.al; llukan.puka@fshn.edu.al; llukan.puka@cit.edu.al

Abstract

There are several statistical methods for time to event analysis, among which is the Cox proportional hazards model, that is most commonly used. In this model, the effect of the covariates is assumed to act multiplicatively on the baseline hazard rate and the ratio of the hazards is constant over survival time. However, when the proportionality assumption for the Cox proportional hazards model is not satisfied, an additive hazard regression model may be more appropriate.

In this study, we give an overview of this approach and then apply Aalen's additive hazard regression model to a data set which contain the job tenure for Albanian employees and variables that are thought of as important factors in the job tenure. The Aalen model allows for time-varying covariate effects and is used to assess the impact that age of the employee, the age at which he started the job, salary, gender, position and years of work in front of the current position may have in job termination.

Keywords: *Job tenure, Survival analysis, Cox proportional hazard model, Aalen's additive hazard regression model.*

1. Introduction

Survival models are frequently used in analysis of employment processes and estimation of factors that affect the job tenure. Job tenure is of paramount interest to workers and company since it can be interpreted as a measure of job stability [1]. Job tenure is the length of time an employee has been continuously employed by the same employer and in the same job position. Theodossiou and Zangelidis [2] estimate the relationship between job tenure and job satisfaction. They assess whether job tenure depends on career advancement opportunities. Ng and Feldman [3] evaluates the relationship between job tenure and job performance. Their results suggest that job tenure is largely unrelated to job performance. Grzenda and Buczyński [4] have studied the estimation of employee turnover with competing risks models. Jiang and Wang [5] examined the relationship between career adaptability and the job content plateau, focusing on the moderating roles of job tenure and job self-efficacy.

Proper analysis of processes related to employee job tenure is vital for both employee and employer. From the employee's point of view, the knowledge about individual employee's characteristics, such as the length of time, that this employee stay in the current job, give a competitive edge on the labour market is crucial. Nowadays, many companies struggle with problems related to the lack of sufficient information about the nature of job tenure processes.

Since the year 2000, the job tenure distribution has shifted away from jobs that have been held for a short duration toward jobs that have been held for a longer duration. From the U.S. Bureau of Labor Statistics [6], the median number of years that wage and salary workers had been with their current employer was 4.2 years in January 2018, unchanged from the median in January 2016.

The point at which half of all workers had more tenure and half had less tenure for men was 4.3 years. Among them, 30 percent of employees had a job tenure which is 10 years or more. The median tenure for men with less than a high school education was around 4.7 years and those with at least a college degree had median tenure of 5.2 years.

The median job tenure for women was around 4 years, where only 28 percent of women, had been in the same job and with the same employer for more than 10 years. The median tenure for women with less than a high school education was around 4.2 years and those with at least a college degree had median tenure of 5 years. For an employee in the public sector, his job tenure was 6.8 years, which is considerably higher than job tenure of an employee in the private-sector, with 3.8 years.

The data, which were used in this analysis, describe the employment duration (survival times) of 887 present and former employees from several different companies, in Albania. Dependent variable was defined as time in months of employment. The main aim of this study is modeling the duration of job tenure and evaluations of factors which are considered important in this time using survival analysis. In particular the Kaplan-Meier method [7] is used to estimate the survival function, and Aalen's additive hazard model [8] to assess the impact that independent variables have on survival time.

Having in mind the main aim of this study as well as the availability of the data, the use of the following independent variables has been proposed: the current age of the employee, the age at which he started the job, salary, gender, position, education, marital status and years of work in front of the current position. Given that some employees have not completed their tenure by the end of the sample, a hazard based analysis is useful to deal with right censoring.
Model validation and results were obtained through R package software, *survival(), timereg(), ggplot2()*.

2. Materials and methods

Survival analysis refers to the general set of statistical methods developed specifically to model the timing of events.
Let $Y_1,\ldots,Y_n$ be a variable of interest with density f and distribution function F, and let $C_1,\ldots,C_n$ be a censoring variable with continuous distribution function G.
We assume that Y is independent of C. Under random right censoring, the variable is not completely observed. One can only observe $(Ti,\delta i)$ where $T = min(Yi,Ci)$ and $\delta i = I(Yi \leq Ci)$ with $I(\cdot)$ being the indicator function [9].

2.1. Cox proportional hazards (Cox PH) model

A popular regression model for modeling the relationship of covariates to a survival time is Cox proportional hazards model (Cox PH model). This model has been proposed by Cox, 1972, [10] and hazard function relates with covariates as follows

$$h(t \mid Z) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j z_j\right) = h_0(t)\exp(\beta^T Z) \quad (1)$$

where $Z = (Z_1,\ldots,Z_p)^T$ is the vector of explanatory variables for a particular individual, $\beta = (\beta_1,\ldots,\beta_p)$ T is a parameter vector of regression coefficients and $h_o(t)$ is the baseline hazard function, which represents how the risk changes with time. Baseline hazard function can be interpreted as the hazard function when all covariates are ignored. This model assumes that covariates have a multiplicative effect on the hazard function and possesses the property that different individuals have hazard functions that are proportional.

2.2. Aalen's additive hazard regression model

In the Cox PH model, the effect of the covariates was to act multiplicatively on some unknown baseline hazard rate. Covariates which do not act on the baseline hazard rate in this fashion were modeled either by the inclusion of a time-dependent covariate, by stratification or by separate analyzes on disjoint time intervals.
An alternative model based on assuming that the covariates act in an additive manner on an unknown baseline hazard rate, is Aalen's additive hazard model. The model was first suggested by Aalen [8]. The unknown risk coefficients in this model are allowed to be functions of time so that the effect of a covariate may vary over time. Unlike the proportional hazards model which estimates hazard ratios, an additive model estimates the difference in hazards: the change in hazard function due to the exposure of interest or stated

more simply the absolute difference in the instantaneous failure rate per unit of change in the exposure variable.
The conditional hazard rate in Aalen's additive hazard model is:

$$h[t \mid Z(t)] = \beta_0(t) + \sum_{k=1}^{p} \beta_k(t) Z_k(t) \qquad (2)$$

where $Z(t) = (Z_1(t),...,Z_p(t))^T$ is the vector of, possibly, time-dependent covariates, for a particular individual, and $\beta_k(t)$, $k=1,...,p$ are unknown parametric functions to be estimated from the data [8]. Direct estimation of the $\beta_k(t)$ is difficult in practice, so we can directly estimate the cumulative risk function $B_k(t)$, defined by:

$$B_k(t) = \int_0^t \beta_k(u)du, \ k = 0,1,...,p \qquad (3)$$

To find the estimates of Bk(t) a least-squares technique can be used, by define an $n\times(p+1)$ design matrix, $X(t)$, where $Xi(t)=Yi(t)(1, Zj(t))$ Let $I(t)$ be the n×1 vector with ith element equal to 1 if subject i fail at t and 0 otherwise. The least-squares estimate of the vector $B(t)$ is:

$$\hat{B}(t) = \sum_{T_i \le t} [X^t(T_i)X(T_i)]^{-1} X^t(T_i)I(T_i) \qquad (4)$$

The variance-covariance matrix of $B(t)$ is:

$$\hat{D}(\hat{B}(t)) = \sum_{T_i \le t} [X^t(T_i)X(T_i)]^{-1} X^t(T_i)I^D(T_i)X(T_i)\{[X^t(T_i)X(T_i)]^{-1}\} \ (5)$$

here the matrix, $I^D(t)$ is the diagonal matrix with diagonal elements equal to $I(t)$ [11]. Vansteelandt Martinussen and Tchetgen [12] showed that Aalen's least square estimator is an unbiased estimator.

2.3. Counting process

Estimation in Aalen's additive model can be performed with counting process:

$$dN_i(t) = Y_i(t)\beta_0(t)dt + Y_i(t)\sum_{j=1}^{p} \beta_j(t)z_{ij}dt + dM_i(t) \quad (6)$$

where $dN_i(t)$ indicates event for individual i at time t, $Y_i(t)$ the at risk indicator for individual i at time t and Mi(t) the martingale increment [13]. This may be interpreted as a linear regression model at each time t with $dN_i(t)$ as responses; $\beta_j(t)dt$ as regression coefficients and $Y_i(t)z_{ij}$ as covariates. The problem of estimation, testing and model fitting were discussed in [14, 15].

The cumulative regression functions, $\hat{B}_k(t)$, estimation of $\hat{B}_k(t)$, with counting process, can be interpreted as: if $\hat{B}_k(t)$ is increasing in an interval, than we have higher risk with high value of z_k; if $\hat{B}_k(t)$ is decreases in an interval, than we have lower risk with high value of z_k and if $\hat{B}_k(t)$ is roughly constant in an interval, than we have little effect of z_k in the interval. An important contribution to the counting processes and martingale theory formulation is given by Aalen [13, 14, 15].

3. Study population

We have analyzed the length of time that employees have been in their current job or with their current employer for 887 employees, in Albania. With survival time we refer the time from the beginning of the relationship with the company, until the end of the term, where the time is in month. The sample contain employees who were still working at the time when the data was last updated, December 2018, 534 respondents were still working, while 353 of them had interrupted the employment relationship with the employer. For these employees we do not know the exact job tenure, but we cannot consider the time we have, as missing data because it contains information. Here, we have to deal with right censuring data, which is referred as the indicator variable, whether or not an employee is still working.

In this study, the survival function gives the probability that the employee will stay in a working relationship for a certain time. On the other hand, the hazard function gives us the potential risk that the employee will terminate the relationship with the company after a certain time. Aalen's additive hazard model is used to estimate the effect that some variables may have on job tenure. The factors studied are: the current age of the employee, the age at which he/she started the job, salary, gender (male, female), position (engineer, supervisory, specialist, financier and other positions (driver, cleaner, babysitter, etc.)) and years of work in front of the current position.

4. Results

There are in total 887 employees. Among the employees surveyed, 58 are financiers, 135 engineers, 108 supervisors, 265 specialists and 314 work in other positions. There are 254 women and 633 men. The average age is 45 years old, where the average age for women is 41 years old, while for men is 46.5 years old. The average salary is 68000 Lek. The shortest time that an employee has been in a current position is one month, while the the longest stay time is around 28 years, 26% of the employees have been in their current position for more than 10 years, where 77% are men, and 30% have been with the same employer for less than 1 years.

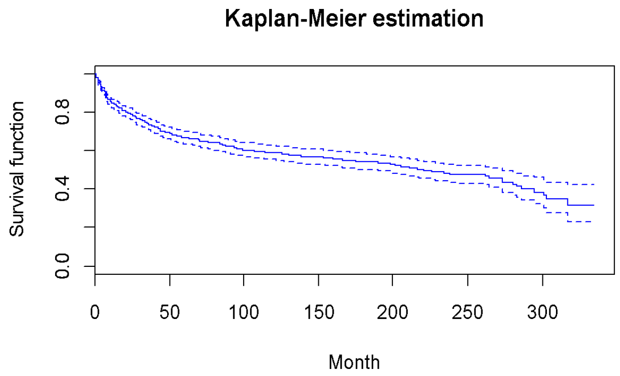

Figure 1. Estimated survival function with Kaplan-Meier method, for job tenure.

Figure 1 shows that for the employees, job survival is up to one year for 84.6%, up to 2 years for 80%, up to 3 years for 73%, up to 4 years for 69% and up to 5 years for 66%. An employee with tenure of less than 3 months has a median expected time in work of 2 months. Which means that 50% of the employees who have joined the company less than three months ago are expected to leave within 2 months of their joining. Similarly employees who have lasted for more than 3 months but haven't still completed 6 months, 50% of them will leave within 5 months. Thus if we somehow increase the percentage of employees staying beyond three months, we will have a higher percentage of people staying beyond the sixth month. An employee with tenure of 1-2 years has a median expected time in work of 21 months.

In the next step, in order to assess the impact of the factors in job tenure, the Aalen's In the next step, in order to assess the impact of the factors in job tenure, the Aalen's additive regression model is fitted to the data.

Table 1. Aalen's additive hazard regression model.

	Test for non-sig effects, p-value	Test for time invariant effects	
		Cramer von Mises test	p-value
(Intercept)	0.000	1.54e+03	0.005
Age	0.000	8.06e-01	0.000
Ywb	0.000	1.78e-04	0.515
Sal	0.000	6.46e-04	0.639
genderM	0.065	1.28e+02	0.000
Agebi	0.002	1.21	0.008
profFinancier	0.026	2.25e+01	0.025
Profother	0.000	2.39e+01	0.016
profSpecialist	0.000	2.86e+01	0.008
profsupervisory	0.000	5.97e+01	0.000

Table 1 shows that the covariates: the current age of the employee, the age at which he/she started the job, salary, position and years of work in front of the current position, turned to be statistically significant at the level of $\alpha = 0.05$. The covariate gender have no effects on the job tenure. Also this table shows the values taken from test for time invariant effects, where years of work in front of the current position and salary have a time-varying effect. Number of simulations in resampling is 1000.

The slope of an estimated cumulative regression function is positive when covariate increases and this fact correspond to an increasing hazard rate. On the other hand, if the slope is negative while the covariate increases, then this fact points to a decreasing hazard rate. If the slope of cumulative sums approaches zero then a covariate has no effect on the hazard.

The plots of estimated cumulative regression functions are obtained to see the effect of covariate over time. The estimated cumulative regression coefficients for covariates with 95% pointwise confidence intervals are shown in Figure 2.

Figure 2 indicates that the estimates of cumulative regression function for gender are constant at a level of zero. The regression function plot for age at which the employee started the job and salary have a significant and fairly constant increasing effect.

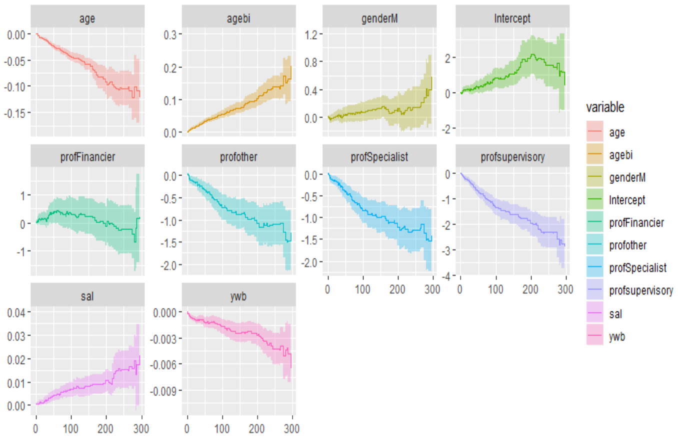

Figure 2. Estimated cumulative regression functions with 95% pointwise confidence intervals based on Aalen's additive model

The slope of estimated cumulative coefficient for age and years of work in front of the current position have a decreasing effect on hazard.

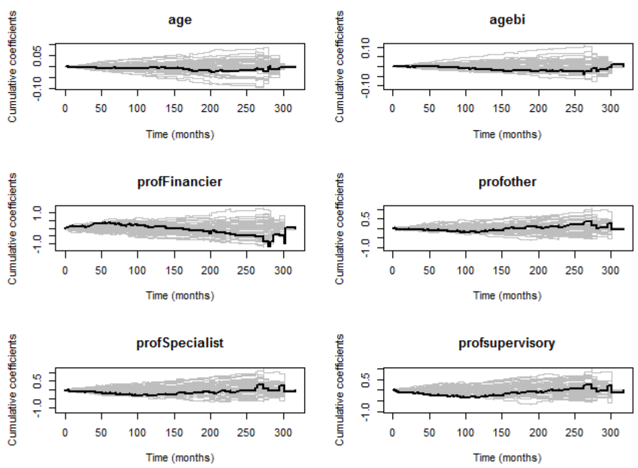

Figure 3. Score test for constant multiplicative effect for age, the age at which he/she started the job and position

The score processes to test proportionality for age, the age at which he/she started the job and position, are given in Figure 3 with 1000 random realizations under the null hypothesis of constant multiplicative effects. Figure 3 indicates that constant multiplicative effect is satisfied for these covariates.

From the final Aalen's additive regression model, based on the above results, that the gender has no effect on the hazard and years of work in front of the current position and salary have a time-varying effect, we conclude that one unit change in salary extends survival time by 1.01 times. One unit change in age shortens survival time by 0.86 times, while one unit change in age at the beginning of the current position in the company extends survival time by 1.16 times.

The estimated risk of dismissal for an employee, with age less than 30 years is approximately equal to 4.853 times higher than an employee older than 50 years old. Employees aged between 30 to 50 years at the time they start the job have 75% less risk of being dismissed than an employee over 50 years old at the moment he/she begin the job. If the job position of an employee is an engineer, this reduces the risk of leaving with 45% compared to a financier.

5. Conclusion

The objective of the study was to assess the impact of actual age, age at the beginning, salary, gender, years of work in front of the current position, in the length of time that employees have been in the current job or with their current employer, known as job tenure. Analysing job tenure and the factors which affect on it, is important for companies, as well as for employees. Sometimes employers, consider it as criteria for hiring new employees.

The analysis is made by Aalen's additive regression model, which allows for time-varying covariate effects. This model provides a graphical method to check on a time dependence of covariate effects and it may be used for the significance test of Cox's model.

The findings showed that gender has no effects on the job tenure, while years of work in front of the current position and salary have a time-varying effect. The average job tenure of an employee on a given company is approximately 187 months and job survival is up to one year for 84.6% of the employees. If we somehow increase the percentage of employees staying beyond three months, we will have a higher percentage of people staying beyond the sixth month in the same job.

References

1. *Neumark, D. (2000), Changes in job stability and job security: A collective effort to untangle, reconcile, and interpret the evidence, NBER Working Paper No. 7472, National Bureau of Economic Research, Cambridge, MA, USA.*
2. *Theodossiou, I., and Zangelidis, A. (2009). Career prospects and tenure–job satisfaction profiles: Evidence from panel data. The Journal of Socio-Economics, Vol 38, Issue 4, Aug 2009, pp 648-657. https://doi.org/10.1016/j.socec.2009.03.006.*
3. *Ng, Th., and Feldman, D. (2013). Does longer job tenure help or hinder job performance? Journal of Vocational Behavior. Vol 83, Issue 3, Dec 2013, pp 305-314, https://doi.org/10.1016/j.jvb.2013.06.012.*
4. *Grzenda, W., and Buczyński, M. (2015) Estimation of employee turnover with competing risks models. Folia Oeconomica Stetinensia. DOI: 10.1515/foli-2015-0035.*
5. *Jiang, Z., and Wang, Z. (2018). Career adaptability and plateaus: The moderating effects of tenure and job self-efficacy. Journal of Vocational Behavior.Vol 104, Feb 2018, pp 59-71, https://doi.org/10.1016/j.jvb.2017.10.006*
6. *Bureau of Labor Statistics, 2018 https://www.thebalancecareers.com/bureau-of-labor-statistics-bls-2059767*
7. *Kaplan, E., and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the Avmerican Statistical Association, vol. 53, pp. 457–481.*
8. *Aalen, O.O. 1980. A model for nonparametric regression analysis of counting processes. Lecture Notes in Statistics-2: Mathematical Statistics and Probability Theory, edited by Klonecki, W., Kozek, A. & Rosinski, J. New York: Springer. pp. 1-25.*
9. *Kleinbaum D., and Klein M, (2005). Survival Analysis. Statistics for Biology and Health, Springer, New York, NY, USA, 2nd edition.*
10. *Cox, D.R. (1972). Regression models and life tables, Journal of the Royal Statistical Society B, 34, 187-202.*
11. *Klein J., and Moeschberger M. (2003). Survival analysis techniques for censored and truncated data. 2nd Edition. Springer. ISBN 0-387-95399-X*
12. *Vansteelandt, S., Martinussen, T. and Tchetgen, E.J.T. 2014. On adjustment for auxiliary covariates in additive hazard models for the analysis of randomized experiment. Biometrika 101: 237-244.*
13. *Aalen, O.O., Borgan, Ø. and Gjessing, H. (2008). Event History Analysis: A Process Point of View. NewYork: Springer.*
14. *Aalen, O.O. 1989. A linear regression model for the analysis of life times. Statist. Med. 8: 907-925.*
15. *Aalen, O.O. 1993. Further results on the non-parametric linear regression model in survival analysis. Statist. Med. 12: 1569-1588.*