International Journal of Business Management and Commerce; Vol. 4 No. 2; June 2019
International Journal of Business Management and Commerce
Published by Center for Contemporary Research

Estimation of Employment Survival Probabilities with Firm-Specific Data from the Trucking Industry

Dr. Steve Mullins
Professor of Economics
Breech School of Business Administration
Drury University
900 N. Benton Avenue
Springfield MO, 65802, USA

Dr. Sergey Borodich
Associate Professor of Mathematics
Drury University
900 N. Benton Avenue
Springfield MO, 65802, USA

Abstract

It is well established in the forensic economics literature that workplace attrition or job tenure should be considered when calculating the economic damages associated with lost pay resulting from wrongful failure to hire or employment termination. (Baum, 2013; Trout, 1995; White, Tranfa-Abboud, & Holt, 2003) When available, firm specific data should be used in specific cases dealing with employment in industries where employment attrition is high because reliance on aggregate data would likely understate employment attrition, overstate the duration of employment and consequently overestimate any economic damages associated with forgone employment opportunities. This is especially the case in an industry such as long-haul trucking where employee turnover is high and employment durations are low relative to most other industries. In this paper, employment records for more than 6000 applicants who entered a training program to become truck drivers between 2009 and 2013 are used to illustrate alternative statistical techniques that can be used with firm-specific data to determine the probabilities of continued employment and expected employment durations that are required to properly estimate the expected value of forgone earnings in cases where workplace attrition adjustments are warranted. The paper contrasts the Binary Logistic (Logit) model and the Kaplan – Meier right-censored survival model that can be used with this type of firm-specific data.

I. Introduction

Forensic economists must in many cases consider workplace attrition or job tenure when calculating the economic damages associated with lost pay resulting from wrongful failure to hire or employment termination. (Baum, 2013; Trout, 1995; White, Tranfa-Abboud, & Holt, 2003) When available, firm specific data is preferable over aggregate data in cases dealing with employment in specific industries where employment attrition is high because reliance on aggregate data would likely understate employment attrition, overstate the duration of employment and consequently overestimate any economic damages associated with forgone employment opportunities.

This paper utilizes employment records for more than 6000 applicants who entered a training program to become truck drivers between 2009 and 2013[1] to illustrate alternative statistical techniques that can be used with firm-specific data to determine the probabilities of continued employment and expected employment durations. These data are required to properly estimate the expected value of forgone earnings in cases where workplace attrition adjustments are warranted. The truck driver data clearly demonstrates the importance of attrition adjustments for that particular industry. The firm's data suggests that that long-haul trucking experiences high rates of employee turnover and relatively low employment durations. This paper uses this firm-specific data to compare some alternative procedures that are available to the forensic economist, and focused on the use of a common right-censored survival model to estimate the employment probabilities and durations that are crucial to economic damage calculations in cases where employment attrition is an issue.

Section II of the paper discusses some important contributions to the literature regarding adjustments for employment attrition. Section III demonstrates how the binary logistic (Logit) model suggested by Baum (2013) can be used with the trucking data. The limitations associated with this parametric model applied to the firm-specific data are also discussed. Section IV uses hypothetical data to demonstrate the Kaplan – Meier Procedure, a popular right-censored survival, and Section V uses Kaplan – Meier to estimate cumulative employment survival probabilities and expected durations of employment using the trucking data.

II. Recent Literature on Employment Attrition

In their 2003 *Journal of Forensic Economics* article, White, Tranfa-Abboud, and Holt demonstrate the importance of considering workplace attrition in discrimination cases. "In discrimination cases…..the forensic economist must consider how long the plaintiff would have worked with that employer. When reliable data are available from the employer, the economist is able to examine the attrition rates or typical tenure of similarly situated employees."[2] The authors provide a case study to demonstrate an appropriate methodology.

A hypothetical employee is assumed to be wrongfully terminated after 7 years of employment. In the absence of the termination, the employee could have worked at the firm until retiring 23 years later. The employee was earning $57,000 annually prior to termination, and it is assumed he could mitigate his damages somewhat by obtaining alternative employment earning $45,000 per year. Hypothetical retention rates ranging from .99 for the first year of post-termination employment to .74 for the 23rd year are used to adjust the earnings that could have been expected if it were not for the wrongful termination. If no adjustment for attrition is made, the present value of 23 years of earnings losses is $393,144. Adjusting for attrition reduces the damages to $231,087 or about 60% of the unadjusted loss.[3]

While the use of hypothetical attrition rates can demonstrate the need to account for employee attrition, the forensic economist requires estimates of relevant attrition rates in order to apply White's methodology in specific cases. Fortunately, other authors have obtained estimates of future employment probabilities using different sources of employment data. Three of these studies are considered here.

Trout used a sample of 63,163 individuals obtained from the Bureau of Labor Statistics Current Population Survey to estimate a logistic regression model that allowed him to determine the probability that an employee of a given age, income, education and employment duration would remain on the same job.[4] He found that a hypothetical 40 year old employee with a high school education who earns $25,000 per year has 72% and 55% probabilities of being with the same employer at ages 44 and 48 respectively.

---

[1] The data was made available to one of the authors as a result of his participation as an expert witness in the damages portion of a multiple claimant gender discrimination case, Equal Employment Opportunity Commission v. New Prime, INC. He was retained by counsel for the defendant firm, new Prime Trucking, Inc., one of the nation's largest trucking companies.

[2] See White et.al (2003), page 209.

[3] White's hypothetical results are likely to understate the importance of adjusting for attrition because he used hypothetical attrition rates (about 1% per year) that are much lower than one would expect based on estimates of employment attrition using BLS, NLSY, or my firm-specific data.

[4] Although Trout (1995) discussed the importance of considering retention in wrongful termination cases, the same economic logic applies in wrongful failure to hire cases as well.

The impact of applying these probabilities to the calculation of economic damages is significant. This 40 year old would experience employer-specific earnings of $416,000 over her expected work life if no accounting for probable employment changes is made. The estimate falls by almost 50% to $215,000 when employment probabilities are considered. If a ten year employment horizon were considered instead, the present value of future earnings is reduced by about one-third (from $216,000 to $160,000) when the earnings are adjusted with his employment survival probabilities.[5]

The BLS data used by Trout included occupational information for many of the respondents. This allowed him to demonstrate the need for forensic economists to consider occupational data (if available) because of the significant differences in employee attrition that may be observed across occupations. He found, for example, that his hypothetical 40 year-old manager would have a .65 probability of continued employment at age 50 compared to a probability of .57 for a sales person. (Trout, 1995, p. 175)

In his 2013 article, Baum used estimates of industry-specific attrition obtained with data from the National Longitudinal Survey of Youth (NLSY) to obtain adjustments for attrition. The NLSY tracked over 12,000 individuals from 1979 through 2010 and collected detailed employment status information that allowed him to determine when employment spells with specific employers ended. Baum used a binary-logistic (or Logit) model to estimate conditional probabilities that an employment spell with a particular employer would end in any year.[6] These conditional probabilities of terminated employment (or hazard rates) were used to calculate the cumulative probability of remaining with the same employer for different durations. (Baum, 2013, p. 54) Baum's employment survival functions derived from all of the employment spells experienced by the males and females in the NLSY dataset are shown in Figure 1.

Baum also used his results to demonstrate the significant impact that adjustment for the likelihood of continued employment can have on damage estimates. He considers a hypothetical female high school graduate born in 1960 who is wrongfully terminated after 5 years of employment from an office and administrative support position paying $20 per hour or $40,000 per year . The present value of damages over an 18 year expected work life are reduced by over 40% when adjustments for cumulative employment survival probabilities are used.[7] It is clear that using employment survival probabilities obtained from the NLSY can have a substantial impact on a forensic economist's estimate of the income losses experienced by the victim of employment discrimination.

Skoog and Ciecka (2014) used data obtained from the *Twenty-Fifth Actuarial Valuation* published by the Railroad Retirement Board's Bureau of the Actuary to estimate cumulative employment probabilities and work life expectancies (WLE) for railroad employees. Competing risks/multiple decrement and railroad Markov process models were used with data on mortality, disability retirements, age retirements and other sources of labor force exodus to generate estimates of WLE and employment duration probabilities for railroad employees of different ages and years of prior service.[8] Their estimates of cumulative employment probabilities for forty year-old employees with 0 and 15 years of prior service are reproduced in Figure 2. Of note is the important impact that years of prior service was found to have on future employment probabilities. Forty year-olds with 15 years of prior service with the railroad were found to be much more likely to stay employed in the industry for an additional 20 years than were new employees. After 20 years, the probability of additional years of employment plummets below those for workers of the same age with less prior service.

III. The Logit Model

As discussed in the previous section, one statistical procedure that can be used to estimate the probability of continued employment is the binary-logistic or Logit model.

---

[5] Trout used a 3% net discount rate to bring the $25,000 annual earnings loss to present value. He considers no actual earnings offsets.

[6] The Logit is a parametric model that is estimated using maximum likelihood estimation. A complete explanation of the model is beyond the scope of this paper, however a discussion of the basic assumptions and mathematics of the Logit model is provided in Appendix 1.

[7] The victim of discrimination is assumed to have experience annual losses beginning at $5000 for 18 years until retirement. Income growth and discount rates of 5% and 3% are used. See (Baum, 2013, pp. 54-55) for more details.

[8] See pages 110 to 112 of Ciecka and Skoog (2014) for a discussion of the relevant theory regarding these models.

This parametric model requires data that contains employment status expressed as a 1 or 0 (signifying employment was terminated or not) and one or more explanatory variables upon which the probability of continued employment is assumed to depend. We applied Logit analysis to our trucking sample of 543 applicants who completed training and were employed as new truck drivers in 2009. Three variables were created for each driver, Yi = 0 or 1 depending upon whether or not the driver's employment spell was terminated over the five year period for which we had data, EMPDURi equal to the number of days driver i was employed at the firm, and LEADSEATi equal to 0 or 1 based on a driver rank variable that classified each employee as either a "Second Seat" or a "Lead Seat" driver. Those drivers who had been promoted to lead seat status earned greater compensation.[9]

The Logit model uses Maximum Likelihood Estimation to find the values of the parameters β_0, $\beta_{1\ldots}\beta_k$ in equation 1 that best estimates the natural log of the "odds ratio".[10] The odds ratio is defined as the probability p_i that employee i continues working to the probability $1 - p_i$ that he or she terminates working during the time period under consideration.

$$\ln\frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}. \qquad (1)$$

The $\ln\frac{p_i}{1-p_i}$ is assumed to be a linear function of one or more independent variables. Since our firm-specific data set contained very limited data regarding personal driver characteristics (sex and age information for example were not available) our independent variables were limited to only the two described above……each driver's employment duration and seat status. The Logit procedure resulted in the flowing estimated relationship between employment duration and seat status (our independent variables) and the log of the odds ratio:

$$\ln\frac{p_i}{1-p_i} = 26.16 - .016EMPDUR - 4.68LEADSEAT \qquad (2)$$

Equation 2 was then used to calculate Pi, the probability of continued employment as a function of employment duration and driver seat status. The results are shown in Figure 3.

As seen in Figure 3 the predicted probability of continued employment stays close to one for approximately 1000 days of work and then sharply falls to close to zero after 1500 days. These results are clearly inconsistent with even a casual glance at the data. Over 80% of drivers in our dataset had employment durations shorter than 1000 days. In fact, over one-half were employed less than one year. Our results provide an example of one important shortcoming of the Logit model….it imposes on the data an assumed logistic probability distribution (like that shown in appendix 1) that, in this case at least, is clearly an inappropriate functional form.

IV. The Kaplan – Meier Procedure

The Kaplan-Meier (KM) procedure is used to study the duration of time that can be expected to elapse between an originating event and a subsequent occurrence that is the object of interest. It is specifically designed to deal with situations where some time-duration observations are incomplete, as is the case with our driver employment data. (Kaplan & Meier, 1958) The procedure is commonly used in medical drug trials and other health-care scenarios where researchers are interested in understanding the impact of medical treatments or disease on the probability of patient survival, but it is also applied to a multitude of cases where researchers are interested in modeling the expected duration of an event and/or the probability of an event occurring as a function of time. (Laerd Statistics)

KM is well suited for estimating firm-specific employment duration expectations and probabilities if data containing employment durations for a large number of employees are available. The data requirements and methodology are described here with a hypothetical example. Readers interested in a more rigorous explanation of the statistics behind the KM procedure should see Appendix 2.

[9] The average "Lead Seat" driver in the sample under discussion earned \$850 per week or 30% more than the \$650 per week earned by their "Second Seat" counterparts.

[10] The mathematics behind the logit model are explained in greater detail in Appendix 1.

The KM procedure has the following minimum data requirements.[11]

Ei = the duration of employment for employee i. For subjects whose employment ended during the study period Ei equals the elapsed time between the beginning date and termination date of the subject's employment at the firm.

Subjects who were still employed at the end of the study period have Ei set equal to the duration of employment from their employment date to the end of the study period.

Si = the status indicator. Si = 1 for those subjects whose employment ended during the study period. Si is set to 0 for censored observations where the subject remained employed beyond the end of the study.

The construction and use of this data to obtain KM estimates of employment survival probabilities are illustrated here with hypothetical employment data shown in Figure 4. The duration of this hypothetical study is one year with each of the ten subject's employment spells measured in days. The duration of each employment event is indicated by the length of the horizontal line and those terminated with a dot reflect an employment spell that ended before the year elapsed. Those without the terminal dots - the third, fifth and tenth observations - are censored cases because the employment spells extended beyond the end of the 365 day observation period. The Ei and Si values for each of the ten subjects are shown in the columns on the right-hand side of the Figure 4.

Table 1 uses the Ei and Si data from Figure 4 to demonstrate the KM methodology.[12] The data are sorted in ascending order by the length of each employment spell with the difference in the duration of each consecutive employment spell determining the time intervals over which employment probabilities are calculated. The first row (Time Interval 1) shows the probability of employment is .9 over the 50 to 100 day interval because the shortest employment spell ended at 50 days leaving 9 of 10 subjects employed from that time until the next employment spell ends at 100 days. Over the next interval (100 to 125 days) 8 of the 9 subjects employed at the beginning of that interval remain employed for a conditional employment probability of .89. The cumulative probability of employment over the 100 to 125 day period is .80, the product of .89 and the .9 cumulative probability of employment to 100 days. Subsequent cumulative probabilities are obtained in the same fashion. Employment past 125 days is associated with a cumulative probability of .70, the product of the cumulative probability for the previous interval (.80) and the .88 conditional probability of continued employment between 125 and 150 days when 7 of 8 subjects remained employed.

The probability calculations for those employment durations where no employment termination takes place, i.e. the censored cases, are treated differently. Row 5 illustrates the first of the three such cases. That subject (the fifth employee in Figure 3) remained employed beyond the 165 days that were observed, consequently no termination of employment occurred at 165 days so the conditional and cumulative employment probabilities are unaffected. However, each censored case does reduce the number of remaining cases that are used to determine subsequent conditional and cumulative survival probabilities.

At 200 days (the beginning of interval 6) another employment spell ends and the conditional probability of employment falls to .8 since four of the five remaining subjects are employed beyond that point. Consequently, the cumulative survival probability falls to .8*.6 = .48 for the 200-225 day interval. The censored case that occurs at 215 days has no impact on the employment probabilities until another employment spell ends at 225 days when two of three subjects remain employed, the conditional probability falls to .67 and the cumulative probability falls to .67*.48 = .32. Two subjects remain employed to a duration of 250 days at which point another employment spell ends for a .5 conditional probability of survival, thus lowering the cumulative probability to .16 at that point. The subject with the longest employment duration (at least 315 days) remained employed at the end of the one year observation period so no additional attrition took place. Consequently, the cumulative employment probability remains unchanged at .16 beyond that point.

[11] The procedure can accommodate additional categorical data as well. For example, if the sex or race of each subject were known, a categorical variable could be added allowing for a comparison of employment survival probabilities and durations between males and females or whites and minorities.

[12] This example is intended to provide the reader with a basic understanding of the K-M procedure. Greater detail can be found in Appendix 2.

Figure 5 represents the cumulative employment survival probabilities obtained from the procedure described above. Up to 50 days, the cumulative probability of employment is 100% since no employment spell terminated prior to that point. The probability of continued employment falls with additional terminations until it reaches a low of .16 at 250 days. Note that, as was explained above, the existence of censored employment durations at 165, 215, and 315 days (each denoted with + in Figure 4) do not impact the employment survival probabilities. If the researcher is interested in expected durations of employment they can be obtained using either the mean or median duration of employment.

IV. Results: A KM Estimate of Firm-Specific Trucking Employment Survival Probabilities and Expected Employment Durations

The methodology described in the previous section was applied to the employment duration records for 543 truck drivers who began employment for New Prime Inc. in 2009. The employment records extended to August 31, 2013 and most of the drivers' employment spells ended prior to that date. There were 59 drivers whose employment continued beyond the duration of the end of the recording period; consequently those cases were censored leaving 486 employment termination events to be analyzed.

An SPSS file was created containing the duration of each driver's employment spell (EMPDURDAYS) and a status variable (TERM) equal to one if the employment terminated before August 31, 2013. The 59 censored cases were identified with TERM set to zero. The KM procedure available in SPSS was used to obtain the estimates of employment survival probabilities shown in Figure 6.[13] The expected employment durations as estimated by the mean and median employment durations are found in Table 2.

The cumulative survival function in Figure 6 illustrates how critically important it is for the forensic economist to consider the impact of employee attrition on the likelihood of continued employment with the firm in a wrongful termination or failure to hire case. The KM estimates show that there is less than a 50% chance that a new driver would still be at the firm one year hence (at EMDURDAYS of 365 the cumulative survival probability is .45), and the likelihood of continued employment falls to .20 at two years. The mean and median durations of employment can be used as estimates of the expected duration of employment. As shown in Table 2, Prime drivers hired in 2009 had mean and median durations of employment of 483 and 311 days, respectively.

The KM procedure also allows for the comparison of employment survival probabilities between different employee characteristics if such data are available. A categorical variable (LEADSEAT) set to one for a lead seat driver and for second seats was created and the KM procedure was run again with that categorical variable selected as a factor. The results are summarized by the two survival functions in Figure 7. Drivers having been promoted to lead seat had much higher employment survival probabilities that their counterparts who had not. SPSS also provides three alternative Chi-Square tests of the equality of the survival functions. Table 3 contains the results of these tests. The hypothesis that no statistically significant difference exists between the two survival distributions can be rejected with very high confidence.

As a practical matter, wrongful termination or failure to hire employment cases typically assign uniform time intervals for the purpose of estimating forgone earnings. A variation of the KM procedure described above can easily be tailored to accommodate the time intervals that are most appropriate for a specific situation. 'Lifetable analysis' is a variant of the KM method which allows the forensic economist to make such a choice. If, for example, foregone earnings need to be estimated and reported to the court on a quarterly basis, 91 or 91.25 day time intervals[14] can be selected so that employment survival probabilities can be calculated for each quarter of the study period. Data from the Prime database was used to calculate cumulative employment survival probabilities using time intervals of one-quarter in duration, and the results are found in Figure 8.[15]

[13] The 'Analyze/Survival/Kaplan-Meier' options are used in IBM SPSS Statistics Version 24 to access the KM procedure. KM is also available in other statistical software packages such as SAS, R, and Stata.

[14] The average quarter is 365/4 = 91.25 days in length unless it is a leap year in which case the average quarter is 91 days long.

[15] SPSS allows the selection of customized discrete time intervals with the 'Analyze/Survival/Life Tables' option. If the time interval is shortened to the point where there is only one termination per interval, the life table procedure will produce a survival function that is identical to the KM procedure.

V. Conclusion

The forensic economics literature makes it clear that employment attrition and job tenure need to be considered in wrongful termination or failure to hire cases where evidence suggests that the assumption of continued employment with the defendant firm, were if not for the employment law violation, is unreasonable. As discussed in the literature review, aggregate national and industry data can and have been used to estimate employment survival probabilities, but firm-specific data is preferred if available because the estimation of plaintiff forgone earnings should, as closely as possible, compare plaintiffs with “similarly situated” employees. (Baum, 2013, p. 43; White, Tranfa-Abboud, & Holt, 2003, p. 209)

The author was fortunate to have access to firm-specific data in a recent multiple-claimant sex discrimination case. That data was first used with the Binary Logistic (or Logit) model to estimate an employment survival function for our sample of truck drivers. The parametric constraints imposed by that model led to implausible results. A second procedure, the Kaplan – Meier right-censored survival model, was then used. This KM model is particularly well suited for the analysis of this type of employment data because the lengths of some employment spells are unobserved since they continued beyond the period of observation. By right-censoring these observations the KM procedure more accurately estimates the survival probabilities and expected employment durations needed by the forensic economist to accurately assess the economic damages associated with forgone earnings at the defendant firm.

References

Baum, C. L. (2013). Employee Tenure and Economic Losses in Wrongful Termination Cases. *Journal of Forensic Economics*, *24*(1), 41-66.

Ciecka, G. R. (2014). Worklife Expectancy of Railroad Workers Based on the Twenty-Fifth Actuarial Valuation Using Both Competing Risks/Multiple Decrement Theory and the Markov Railraod Model. *Journal of Forensic Economics*, *25*(2), 109-127.

Equal Employment Opportunity Commission v. New Prime, INC, 6:11-CV-03367-MDH (U.S. District Court for the Western District of Missouri, Southern Division August 31, 2015).

Kaplan, E., & Meier, P. (1958). Nonparamtric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282), 457-481.

Laerd Statistics. (n.d.). *Kaplan-Meier using SPSS Statistics*. Retrieved November 26, 2016, from https://statistics.laerd.com/spss-tutorials/kaplan-meier-using-spss-statistics.php

Report of the Special Master: Equal Employment Opportunity Commission v. New Prime Inc., 6:11-CV-03367-MDH (U.S. District Court for the Western District of Missouri, Southern Division August, 2015).

Revised Calculation of Economic Damages in the case of EEOC v. New Prime Inc., 6:11-CV-03367- MDH (Western District of Missouri, Southern Division May 14, 2015).

Trout, R. (1995). Duration of Employment In Wrongful Termintion Cases. *Journal of Forensic Economics*, *8*(2), 167-177.

White, P. F., Tranfa-Abboud, J. V., & Holt, F. M. (2003). The Use of Attrition Rates for Economic Loss Calculations in Employment Discrimination Cases: A Hypothetical Case Study. *Journal of Forensic Economics*, *16*(2), 209-223.

Figures

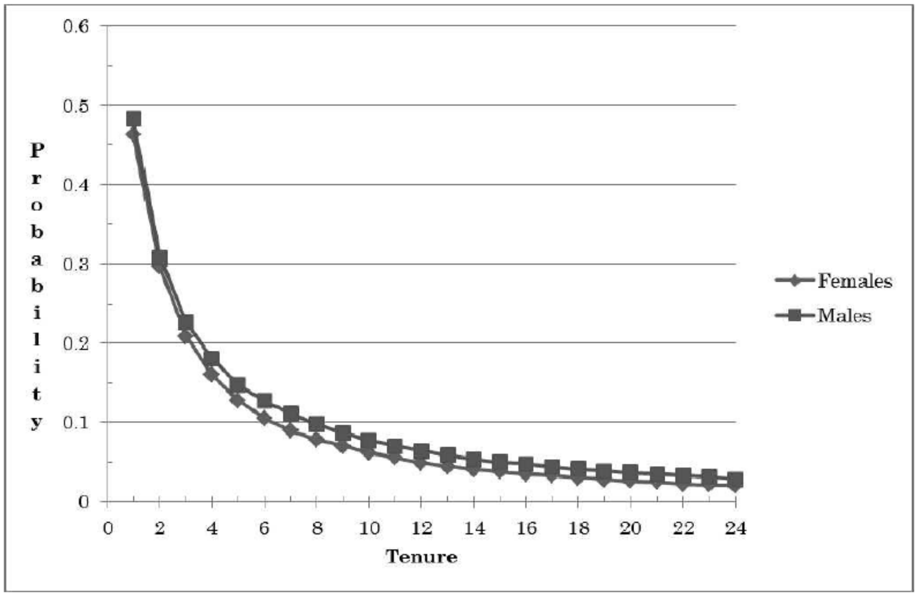


Figure 1 Survival Functions: The Cumulative Probabilities of Additional Years with an Employer by Gender (Baum, 2013, p. 49)

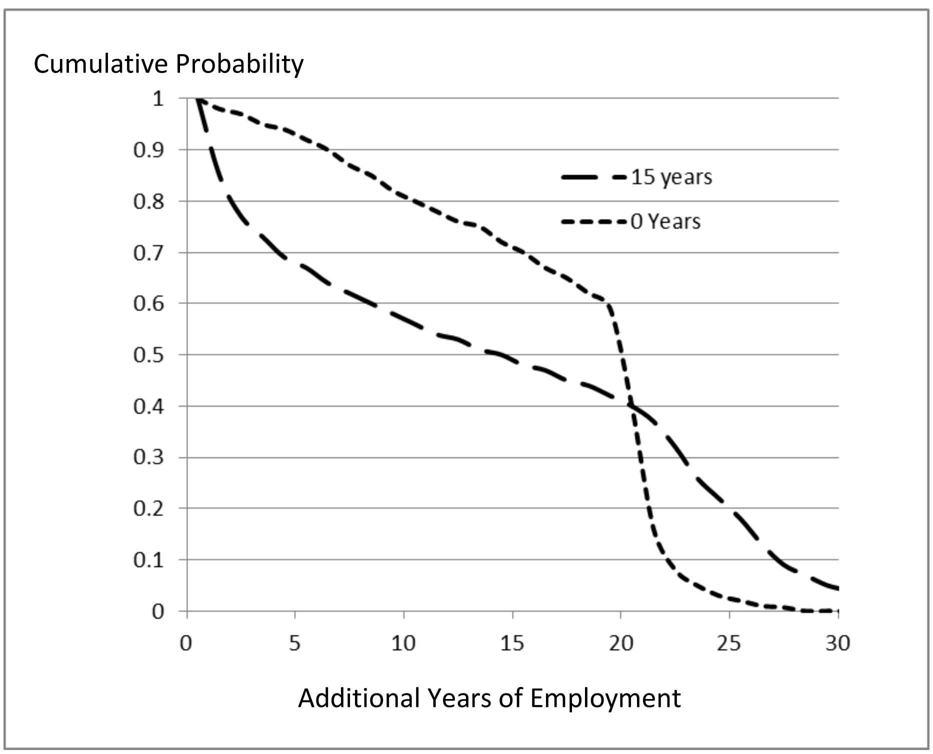


Figure 2 Cumulative Probabilities of Additional Years of Employment by Years of Prior Service[16] (Ciecka and Skoog, 2014)

[16] The author created Figure 2 using results from page 119, Table 3 of Ciecka and Skoog.

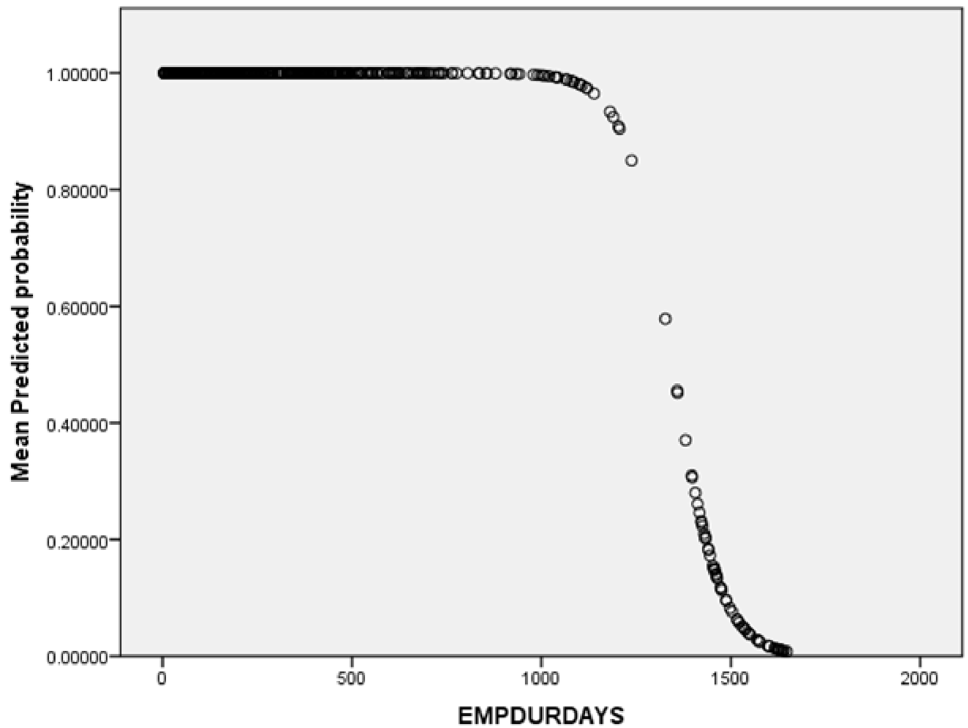


Figure 3 Cumulative Employment Survival Probabilities
Binary Logistic Model

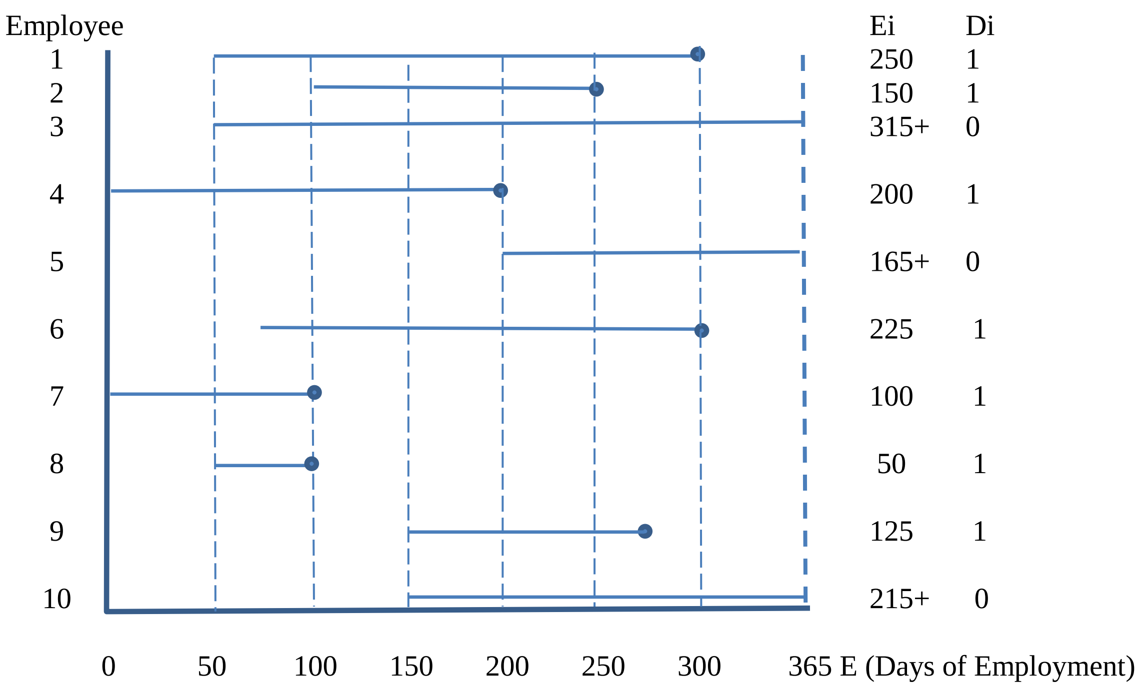


Figure 4 Hypothetical Employment Data for Kaplan – Meier
Right Censored Survival Analysis

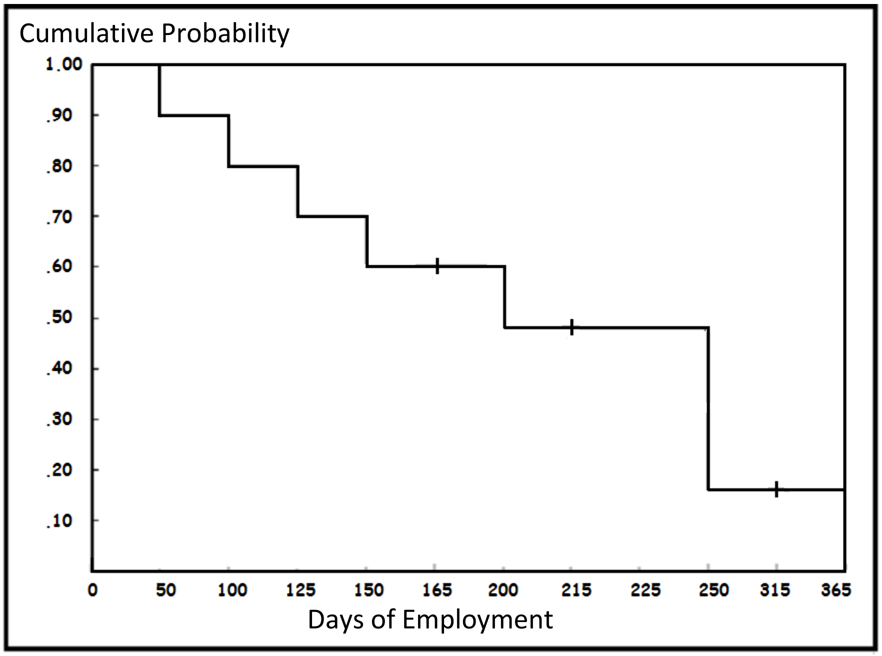


Figure 5 Kaplan – Meier Procedure Cumulative Employment Probabilities Obtained from Hypothetical Data in Table 1 (+ denotes Right Censored Employment Durations)

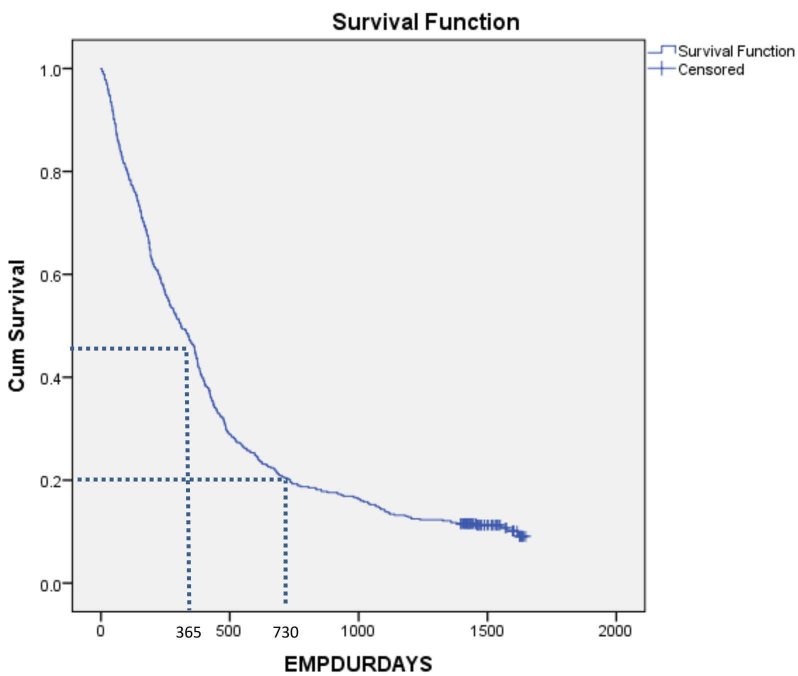


Figure 6 Cumulative Employment Survival Probabilities for Truck Drivers Obtained from KM Estimates of New Prime Inc. 2009-2013 Employment Records

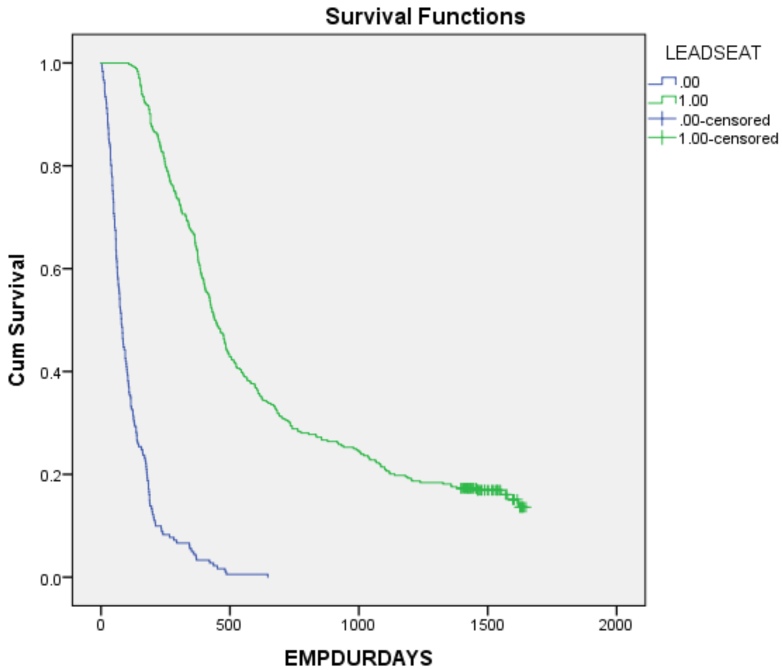


Figure 7 Cumulative Employment Survival Probabilities for Lead Seat and Second Seat Drivers

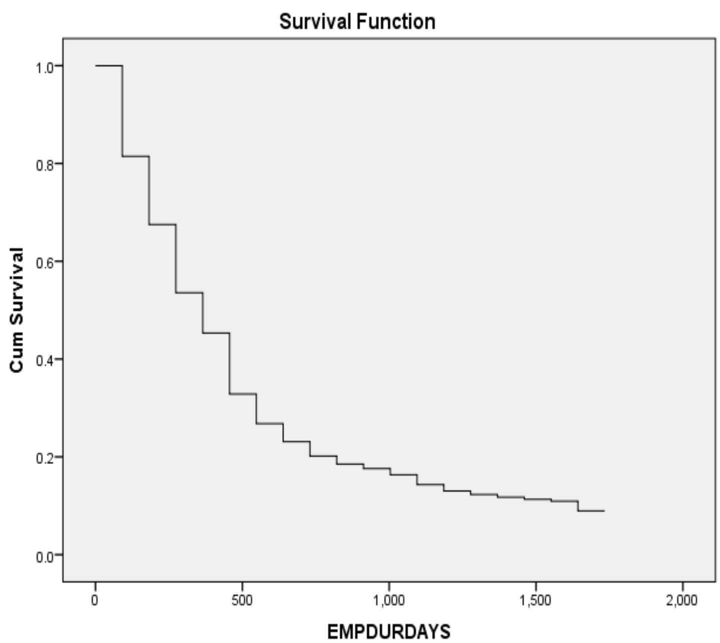


Figure 8 Cumulative Employment Survival Probabilities
With Quarterly Time Intervals

Tables

Table 1
Conditional and Cumulative Employment Survival Probabilities
Kaplan – Meier Procedure applied to Hypothetical Data

Time Interval	Elapsed Time in days	Emploment Status or Si	Cases Remaining	Conditional Probability of employment	Cumulative Probabilty of employment
1	50	1	9	0.90	0.90
2	100	1	8	0.89	0.80
3	125	1	7	0.88	0.70
4	150	1	6	0.86	0.60
5	165	0	5		
6	200	1	4	0.80	0.48
7	215	0	3		
8	225	1	2	0.67	0.32
9	250	1	1	0.50	0.16
10	315	0	0		

Table 2
Mean and Median Truck Driver Employment Duration
Obtained from KM Procedure applied to New Prime Inc.
2009-2013 Employment Records

Mean[a]				Median			
Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound			Lower Bound	Upper Bound
483.354	21.242	441.721	524.988	311.000	22.072	267.739	354.261

a. Estimation is limited to the largest survival time if it is censored.

Table 3

Chi-Square Test Results of Equality of Survival Distributions
Lead Seat verses Second Seat Drivers

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	505.339	1	.000
Breslow (Generalized Wilcoxon)	500.329	1	.000
Tarone-Ware	508.991	1	.000

Test of equality of survival distributions for the different levels of LEADSEAT.

Appendix 1

The Logit Model

The Logit Model is a qualitative choice model where the dependent variable Yi takes on the value of 0 or 1. As applied to the issue of continued employment, 0 and 1 represent the state of continued employment or not as a function of explanatory variables. The model assumes that the conditional employment survival probability distribution is consistent with the logistic distribution:

$$Pi = E(Yi = 1|Xi) = [1 + e^{-Zi}]^{-1} \qquad (1)$$

Where Pi = probability of an event occurring (in this case continued employment) conditional on Xi having occurred (conditional probability of Y given X), e is the base of the natural logarithm, and $Zi = \beta_0 + \beta_1 Xi$. Xi is a vector of explanatory variables and β_0 and β_1 are parameters to be estimated.
From (1), the *odds ratio* can be derived as:

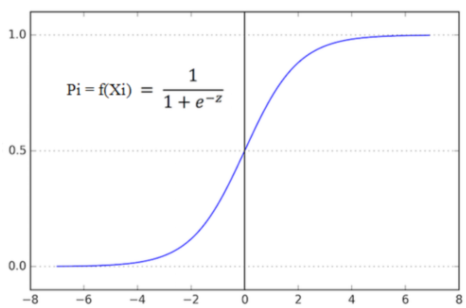
$$Pi/(1- Pi) = e^{Zi} \qquad (2)$$

The logit (Li) is obtained by taking the natural log of the odds ratio (2)

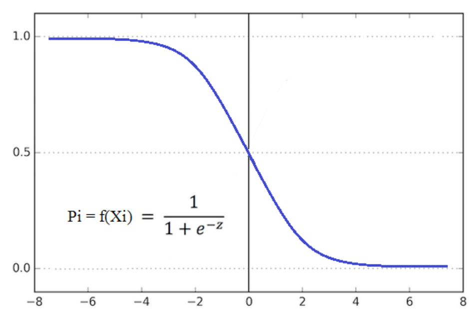
$$Li = \ln [Pi/(1- Pi)] = Zi = \beta_0 + \beta_1 Xi \qquad (3)$$

Li must be estimated using maximum likelihood estimation (MLE) because ordinary least squares regression cannot be used.[17] Depending upon Zi < 0 or Zi > 0, the logistic probability distribution can take on the following general relationship between Pi and Xi.

For Zi > 0



For Zi < 0



Appendix 2

The Kaplan–Meier estimator

The basis for this estimate is the well-known multiplication rule for probability:

$$P(A \; and \; B) = P(A) \cdot P(B \mid A), \qquad (1)$$

Where $P(A)$ – probability of event A occurring, $P(B \mid A)$ – probability of event B occurring if event A has already occurred (conditional probability of B given A). $P(A \; and \; B)$, the probability that both events A and B occur.

We let T measure the duration of a study period that encompasses a maximum length of employment. (It could also be length of life after some surgery, time of trouble-free operation of some device from the moment of the beginning of operation, etc.)

Let t_1, t_2, t_3, … denote actual durations of employment spells for all employees in increasing order. Let d_1, d_2, d_3, … denote the number of dismissals that happen at each of these times, and let n_1, n_2, n_3, … be the number of workers remaining employed during corresponding times but not censored. Censored employees are those who do not have employment terminated within study period. Note that $n_{k+1} = n_k - d_k - c_k$, where c_k is number of workers who has been working not more than t_k but continue to be employed.

[17] Since the dependent variable Pi takes on the value of 0 or 1, the logit is either ln(0/1) or ln(1/0), neither of which are defined.

The survival function $S(t) = P(T > t)$ represents probability that employee works longer than time t. $S(t) = 1$ for any $t < t_1$ (everyone works at least time t_1).

Then $S(t_k) = P(T > t_k)$ is the probability that employee works longer than time t_k (probability of surviving beyond time t_k). It is clear that $S(t_{k+1}) = P(T > t_{k+1})$ = "Probability of surviving beyond time t_{k+1}" depends conditionally on

$S(t_k) = P(T > t_k)$ = "Probability of surviving beyond time tk". By using formula (1), we can iteratively find a point estimate $\hat{S}(t)$ of the true survival function $S(t)$.

For any $t \in [t_1, t_2)$ (i. e, $t_1 \leq t < t_2$)

$$S(t) = P(T > t) = P(\text{works in } [0, t_1)) \cdot P(\text{works in } [t_1, t) \mid \text{worked in } [0, t_1))$$

$\Rightarrow$

$$\hat{S}(t) = 1 \cdot \frac{n_1 - d_1}{n_1} = \left(1 - \frac{d_1}{n_1}\right)$$

For any $t \in [t_2, t_3)$ (i. e, $t_2 \leq t < t_3$)

$$S(t) = P(T > t) = P(\text{works in } [t_1, t_2)) \cdot P(\text{works in } [t_2, t) \mid \text{worked in } [t_1, t_2))$$

$\Rightarrow$

$$\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right) \cdot \frac{n_1 - d_1}{n_1} = \left(1 - \frac{d_1}{n_1}\right) \cdot \left(1 - \frac{d_2}{n_2}\right)$$

By using this approach we can get the general formula for a point estimate of the survival function $S(t)$ for any time t, which is named the Kaplan–Meier estimator:

For any $t \in [t_k, t_{k+1})$ (i. e, $t_k \leq t < t_{k+1}$)

$$\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right) \cdot \left(1 - \frac{d_2}{n_2}\right) \cdot \ldots \cdot \left(1 - \frac{d_k}{n_k}\right) = \prod_{i=1}^{k} \left(1 - \frac{d_i}{n_i}\right). \qquad (2)$$

This is actually step function, but for big data sets its graph looks continuous as seen in Figures 5 and 6 showing our application of the Kaplan–Meier estimator to employment in the trucking industry.