Survival analysis of the unemployed in the 1986-1992 United States

Aditya Bisht, Irene Pham, Mohamed Dhmine, Mónica Rojas

2023-06-13

Abstract

This report describes a survival analysis study on unemployment data observed in the United States between the late 1980s and early 1990s. We aim to test how demographics and previous employment influence the ability of an unemployed subject to (re)gain full-time employment. For this analysis, we first create Kaplan-Meier estimators on one and more groups. Then we test the individual impact of the factors using stratified Logrank test. We finish the analysis by building Cox proportional hazard models: first by comparing results with select combinations of factors, then by performing model selection using Akaike Information Criterion. We hypothesized that certain factors will either have a positive (e.g. age and whether one receives insurance payout), or a negative effect (e.g. tenure in the last job) to time-to-event. However, we find that age and being insurance recipient are the strongest contributors to finding full-time employment among the subjects. All of the other factors except tenure in previous job are also relatively strong factors to consider.

1 Introduction

Time-To-Event (TTE) data is often analysed as a special type of time series problem, where an event is expected to occur after a variable time duration passes. Such analysis, also known as survival analysis, often frames the observed phenomenon around a group of subjects who may survive for different amounts of time depending on other measured factors.

While many such studies involve a medical scenario, in this report we consider a different problem: how much time an unemployed person takes to (re)gain employment. In macroeconomics, unemployment rate is used as a metric to measure the competence of an economy or its policymakers. While studying advanced economies, it is often interesting to study unemployment insurance and its role in the labour market movements.

Unemployment insurance, often issued by governments, takes form of payouts for those who recently lost a job. It aims to financially assist the unemployed while they search for another job. For instance, 2.5 million people received such benefits in France in 2021 ("People Receiving Unemployment Benefits in France 2021" n.d.). The amount of payment a recipient gets varies depending on several factors. Two of these of interest are disregard and replacement. The former is defined as additional payment for recipients accepting part-time work (McCall 1996). While the latter represents the base payment designated to the recipient.

To evaluate this, we consider the "Unemployment Duration" dataset (McCall 1996), which is available through the "Edcat" package in R ("R: Data Sets for Econometrics" n.d.). The observations are from data extracted from surveys taken between 1986 and 1992 (McCall 1996). Notably, the wages are measured in 1985 US Dollars ("R: Unemployment Duration" n.d.). While the study the data originates from aims to determine if *"the level of the disregard influences a… recipient's job search behavior"* (McCall 1996), we aim to use the dataset to study the effect of other factors on (re)gaining employment. Specifically, want to test if demographic factors and aspects of the previous employment influence the time taken to find another full-time job.

In our analysis, we will employ non-parametric models using one or more groups, (stratified) logrank test and semi-parametric models using Cox proportional hazard modelling. Before this, let us look at the dataset in more detail.

2 Data Description

The raw dataset contains 3343 observations and 11 features. In the independent variables, there is one categorical variable (ui) and five quantitative variables (age, tenure, reprate, disrate and logwage). The time variable (spell) is measured in fortnights (two weeks).

As a refined dataset, a minimal amount of processing was done before starting the study. Out of the features whose name starts with "censor", we kept only the first one (censor1), which is equal to 1 if full-time employment is (re)gained. The median age of the subjects is around 35 years; however, the censor is equal to 1 proportionally between the observations on either side. The "tenure" variable is heavily positively skewed, with at least a quarter of observations having the value 0. Its median is 2 years, but extreme values go as high as 40 years. For the analysis, the above two variables are binned in a specified way – "age" as new variable **"AgeCat"** by the decade (e.g. 20-30, 31-40, etc.) and "tenure" by one of the ways seniority is established by years of work experience. For the latter as **"JobRank"**, up to 2 years qualify as "Entry", 2-5 as "Mid", 5-10 as "Senior" and over 10 years as "Executive" ("Data-Processing. Rmd" 2023).

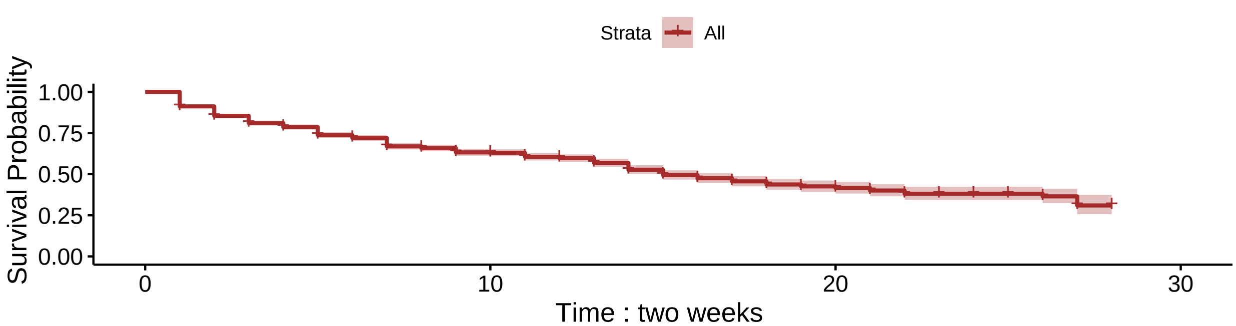
There is no instance of left censor in the dataset, while under 4% observations have right censor. The censor itself is well-distributed, with 32.1% with event equal to 1. The "spell" variable is heavily positively skewed, with an average value of 6.25 fortnights. "disrate" refers to the (weekly) disregard received proportional to their (weekly) salary in the previous job, and it only applies for those who found a part-time job. It is very positively skewed, with mean and median around 10-11% but the highest recorded value of 102%. "reprate", a similar metric but with the (weekly) insurance payout, is on average 45%, with the half of the sample having a value between 39.8% and 52%. Its distribution is positively skewed. The "ui" variable, which measures if the subject is receiving unemployment insurance, has a ratio 5:6 between its values ("no" and "yes" respectively). The "logwage" feature is described in natural log, with an average of 5.67 (around 290). Its distribution is slightly negatively skewed.

Out of the above features, we expect "age", "reprate" and "ui" to help improve the survival of the subjects. Conversely, we expect "tenure" to negatively impact the survival. With these observations on the dataset, we will now describe the analysis done on it and our findings.

3 Analysis

3.1 Nonparametric estimation of survival for one group

To better understand the general trend of the sample population, we plot the Kaplan-Meier estimation.



As per the dataset, the plot accounts for the entire range of time observations and the roughly one-third of the observations with censor = 0. The large amount of steps suggest the wide range of times the subjects were studied. This also implies that for censored observations, it is possible to have a lot of estimations influence the graph's shape. Given that the observation time for the subjects without the event follow a similar distribution as the entire dataset, the above is a major consideration.

From the plot we can concur that the event tends to be consistently more likely as the time passes. However, for further understanding on how different groups behave, we need to perform further tests, while will be discussed in the following sections.

3.2 Nonparametric estimation of survival for two groups or more

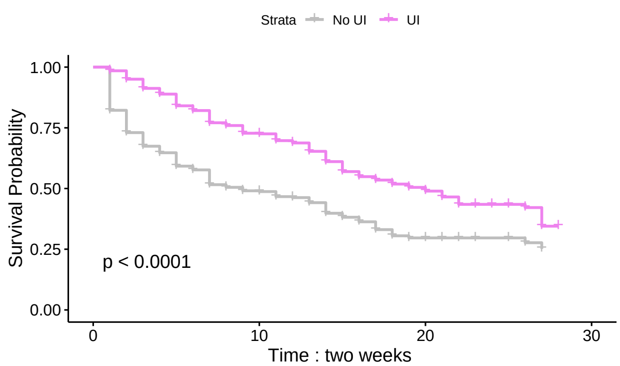
3.2.1 Logrank test

After exploring Kaplan-Meier estimation for a single group, our next step is to apply this analysis to multiple groups. In this part of our study, we'll make use of the log-rank test to ascertain whether there are significant differences between the groups under investigation.

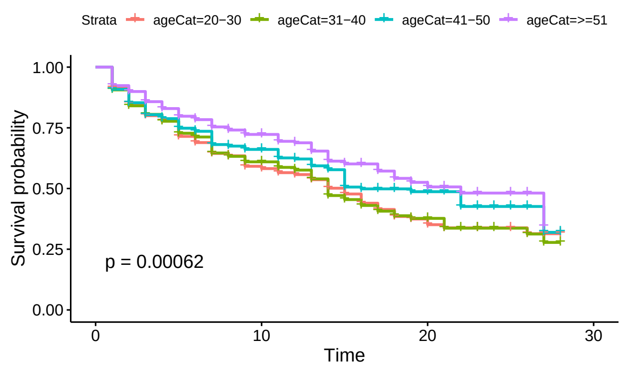
Initially, we'll focus our comparison on two key groups: individuals who filed Unemployment Insurance (UI) claims and those who did not. The goal here is to understand the effects of UI claims on the duration of unemployment.

Subsequently, we'll broaden the scope of our investigation to include an analysis across different age categories and job ranks. This is an important step towards identifying a more diverse range of factors that can impact the time it takes an individual to find full-time employment.

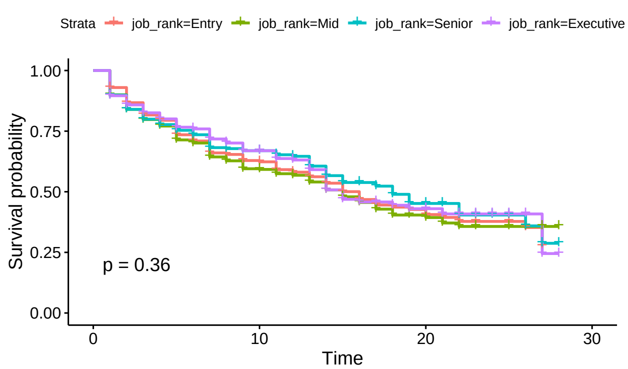
Through this analysis, we aim to provide an understanding of how various factors, from UI claims to age and job rank, impact unemployment durations.



(1)



(2)



(3)

Interpretation of Plots

Plot (1): Unemployment Duration and UI Claim Status The first plot examines the survival distributions of two distinct groups: those who filed an Unemployment Insurance (UI) claim and those who did not. The survival distribution in this context represents the time it takes for an unemployed individual to secure a full-time job.

Observing the plot, we notice that individuals who filed a UI claim typically take longer on average to find full-time employment compared to those who didn't file a UI claim.

To statistically validate this observation, a logrank test was conducted. The null hypothesis (H0) of the logrank test posits that the survival functions of the two groups are identical, implying that the UI claim status does not influence the time taken to secure full-time employment.

However, the obtained p-value from the logrank test is less than 0.005. Consequently, we reject the null hypothesis and conclude that there's a statistically significant difference in the survival distributions between the UI claim and non-UI claim groups.

Plot (2): Unemployment Duration Across Different Age Groups The second plot provides a comparison of the survival distributions across different age groups, in particular, "20-30", "31-40", "41-50", and ">=51".

The corresponding logrank test yields a p-value of 6e-04, which is statistically significant at the usual 0.05 level. This implies that at least one age group's survival function differs significantly from the others, meaning age seems to impact the time it takes to find full-time employment.

However, examining the Kaplan-Meier plot reveals that these differences are not constant over time. In the initial period, all age groups exhibit similar times to find full-time employment. Yet, as time progresses, a noticeable divergence occurs, specifically between the "20-30" and ">=51" groups. This suggests that older job seekers (51 and above) tend to experience longer periods of unemployment before securing full-time work compared to their younger counterparts (20-30). Possible explanations could include factors such as age discrimination or differing job preferences among age groups.

Plot (3): Unemployment Duration and Job Ranking Groups The third plot investigates the relationship between the time taken to find full-time employment and different job ranking groups based on tenure in the last job: 'entry', 'mid', 'senior', and 'executive'.

However, the p-value of the logrank test is 0.36, indicating no statistically significant difference in the time taken to secure full-time employment across the job ranking groups. This finding suggests that the tenure-based job rank in the previous job does not significantly affect the duration of unemployment before finding a new full-time job.

3.2.2 Logrank Stratified test

Having conducted our initial analyses, we now advance to a more nuanced level of investigation using stratified tests. Stratification allows us to control for one or more variables while testing the effects of others. This method will enable us to separate the influences of different factors, providing a clearer understanding of their individual impacts.

In our upcoming analyses, we will be stratifying our data by age categories and job ranks. This approach will help illuminate how unemployment duration is affected across these different strata, beyond the effects of filing a UI claim.

1. **Test 1: Unemployment Insurance and Age Category**

   ```
   ## Call:
   ## survdiff(formula = Surv(time = spell, event = censor1) ~ ui +
   ##     strata(ageCat), data = UnempDur_df)
   ##
   ##           N Observed Expected (O-E)^2/E (O-E)^2/V
   ## ui=no  1495      576      354     139.0       238
   ## ui=yes 1848      497      719      68.5       238
   ##
   ##  Chisq= 238  on 1 degrees of freedom, p= <2e-16
   ```

2. **Test 2: Unemployment Insurance and Job Rank**

   ```
   ## Call:
   ## survdiff(formula = Surv(time = spell, event = censor1) ~ ui +
   ##     strata(job_rank), data = UnempDur_df)
   ##
   ##           N Observed Expected (O-E)^2/E (O-E)^2/V
   ## ui=no  1495      576      347       150       264
   ## ui=yes 1848      497      726        72       264
   ##
   ##  Chisq= 264  on 1 degrees of freedom, p= <2e-16
   ```

Interpretation of tests

Test 1: Unemployment Insurance and Age Category :

The p-value is less than 2e-16, indicating that there is compelling statistical evidence to imply a significant difference in the duration to secure full-time employment between individuals who filed a UI claim and those who did not, within every age category.

In the group that did not file a UI claim, composed of 1495 individuals, 576 were able to secure full-time employment. This figure is significantly larger than the expected count of 354. Conversely, in the group that did file a UI claim, which includes 1848 individuals, only 497 managed to find full-time employment, a number considerably lower than the expected 719.

The results insinuate that filing a UI claim seems to extend the duration it takes an individual to secure a full-time job, irrespective of their age group. The procedure of finding full-time employment appears to be profoundly influenced by whether or not an individual files a UI claim, and this impact is uniformly observed across all age brackets.

Test 2: Unemployment Insurance and Job Rank :

This output represents the result of a stratified logrank test where we are comparing the survival distributions between those who filed a UI claim (ui = yes) and those who didn't file a UI claim (ui = no), while accounting for different job ranking groups.

Given the obtained p-value *p= <2e-16* , which is far below 0.05 conventional level of significance, we reject the null hypothesis. This means there is a statistically significant difference in the time to find a full-time job between those who filed a UI claim and those who didn't, even after stratifying by job rank. Furthermore, observing the Observed and Expected numbers, we see that among those who didn't file a UI claim, more individuals found a job than expected (576 observed vs. 347 expected). Conversely, fewer individuals who filed a UI claim found a job than expected (497 observed vs. 726 expected). This suggests that those who filed a UI claim generally take longer to find a full-time job than those who didn't, even when accounting for job rank.

3.3 Semi-parametric model

3.3.1 Cox Proportional Harzard

In this section, we apply Cox proportional hazard modeling on 4 sets of continuous and categorical covariates to exam their influences on time to find full-time job.

- **Cox.1**: Age, UI, Logwage, Tenure, Reprate and Disrate
- **Cox.2**: Age, UI, Logwage, Reprate and Disrate
- **Cox.3**: Categorical Age (<30, 31-40, 41-50, >51), UI, Logwage, Reprate and Disrate
- **Cox.4**: Categorical Age (<51, >=51), UI, Logwage, Reprate and Disrate

In **Cox.1** model, the effects of all the covariates, except for Tenure, are statistically significant. Given other covariates remain the same:

- Age: being 1 year older results in a 1 % decrease in the rate of finding a full time job. It's worth pointing out that the exp(coef) of age ranging between 0.9818 to 0.9947, which is a narrow bound. This means the exp(coef) for Age is quite consistent.
- UI: the exp(coef) of 0.36 indicates that claiming unemployment insurance introduces a 64% decrease in the chance of finding a full-time job. The variable exp(coef) also has a smaller range, from 0.31 for lower bound to 0.4 for upper bound at 95% confidence interval.
- Logwage: the exp(coef) shows that people having higher wage in the previous job have shorter unemployment duration. 1 unit increase causes a 60% increase in the hazard ratio.
- Tenure: there is almost no effect (exp(coef) is 1.0061) introduced for every unit increase in Tenure. The p-value of 0.3 also shows that this covariate is not statistically significant.
- Reprate: this covariate has a large range of exp(coef) between 1.96 to 9.07 at 95% confidence interval. It shows that for a unit increase in Reprate, there is a 3.8 difference in unit for hazard ratio.
- Disrate: On the other hand, having a lower Disrate results in a higher chance of finding a full-time job. Changing 1 unit in Disrate suggests a 0.165 change in the hazard ratio. Similar with Reprate, its exp(coef) also has a large range from 0.06 to 0.44.

The concordance index which is the equivalent of area under the ROC curve in Logistic regression is at 0.698 for this model, which shows the model prediction is better than random prediction.

Likelihood ratio test, Wald test and Score (logrank) test all examine whether any of the coeffients is equal to 0. The p-value of 3 tests under 2e-16 shows this model significant.

```
## Call:
## coxph(formula = Surv(spell, censor1) ~ age + ui + logwage + tenure +
##     reprate + disrate, data = UnempDur)
##
##   n= 3343, number of events= 1073
##
##              coef exp(coef)  se(coef)       z Pr(>|z|)
## age     -0.011829  0.988241  0.003336  -3.546 0.000391 ***
## uiyes   -1.034833  0.355286  0.064640 -16.009  < 2e-16 ***
## logwage  0.609357  1.839249  0.093579   6.512 7.43e-11 ***
## tenure   0.006059  1.006078  0.005867   1.033 0.301699
## reprate  1.350958  3.861124  0.436287   3.096 0.001958 **
## disrate -1.799896  0.165316  0.501554  -3.589 0.000332 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## age        0.9882     1.0119   0.98180    0.9947
## uiyes      0.3553     2.8146   0.31301    0.4033
## logwage    1.8392     0.5437   1.53104    2.2095
## tenure     1.0061     0.9940   0.99458    1.0177
## reprate    3.8611     0.2590   1.64190    9.0799
## disrate    0.1653     6.0490   0.06186    0.4418
##
```

```
## Concordance= 0.698  (se = 0.009 )
## Likelihood ratio test= 323.7  on 6 df,   p=<2e-16
## Wald test            = 328.2  on 6 df,   p=<2e-16
## Score (logrank) test = 344.6  on 6 df,   p=<2e-16
```

Nested model testing

Since Tenure covariate is not statistically significant according to **Cox.1** model, it is dropped in **Cox.2** model. We use Anova test to check if **Cox.1** model is significantly better. We fail to reject the null hypothesis due to the p-value of 0.306. This means including Tenure variable or not does not affect the predictive power of the model.

```
##                coef exp(coef)    se(coef)          z     Pr(>|z|)
## age     -0.01067116 0.9893856 0.003126009  -3.413670 6.409410e-04
## uiyes   -1.02793750 0.3577440 0.064263537 -15.995657 1.370051e-57
## logwage  0.61777222 1.8547914 0.093237146   6.625816 3.453341e-11
## reprate  1.32501477 3.7622409 0.435958787   3.039312 2.371192e-03
## disrate -1.77473147 0.1695290 0.500939811  -3.542804 3.958971e-04
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(spell, censor1)
##  Model 1: ~ age + ui + logwage + tenure + reprate + disrate
##  Model 2: ~ age + ui + logwage + reprate + disrate
##    loglik Chisq Df Pr(>|Chi|)
## 1 -7786.0
## 2 -7786.5 1.048  1      0.306
```

Test the effect of categorized Age covariate

In the section **3.2**, Age is divided into 4 bins. In **Cox.3** model, we'll use this categorical Age covariate instead of continuous Age covariate to check its effect to the Cox model. According to the summary, ageCat>=51 is the only significant Age covariate. It shows that people from 51 year-old has a chance of finding a full-time job that is 0.7 times that of those under 51, given all other covariates remain unchanged.

```
##                    coef exp(coef)   se(coef)           z     Pr(>|z|)
## ageCat31-40  0.02874503 1.0291622 0.07641637   0.3761633 7.067955e-01
## ageCat41-50 -0.15160226 0.8593300 0.08911514  -1.7011953 8.890633e-02
## ageCat>=51  -0.33739180 0.7136292 0.10804682  -3.1226444 1.792342e-03
## uiyes       -1.04070794 0.3532045 0.06426526 -16.1939424 5.564624e-59
## logwage      0.59990339 1.8219428 0.09295605   6.4536238 1.092067e-10
## reprate      1.31593485 3.7282347 0.43622102   3.0166699 2.555679e-03
## disrate     -1.76443191 0.1712841 0.50029904  -3.5267546 4.206865e-04
```

By replacing the continuous Age covariate with the categorical age_51 (whether or not the person is above 51 year-old), in **Cox.4** model, there is no significant change in term of coefficient value of other covariate. However, we see a slight increase is concordance index, from 0.698 to 0.699.

Since the concordance index only slightly higher in **Cox.4** model compared to **Cox.2** model, we use **Cox.2** model with most continuous covariates to examine further the assumption of Cox Proportional Harzard modeling.

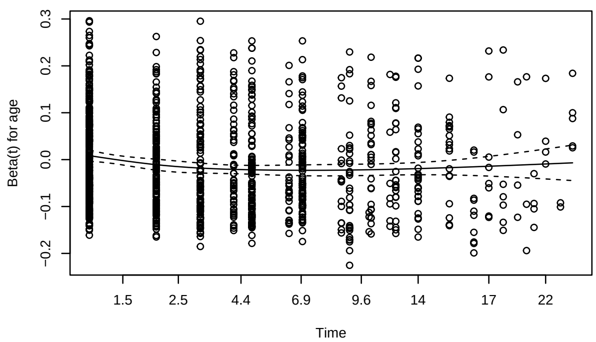
```
##                  coef exp(coef)   se(coef)          z     Pr(>|z|)
## age_51>=51 -0.3699201 0.6907896 0.10577118  -3.497362 4.698844e-04
## uiyes      -1.0426313 0.3525259 0.06384715 -16.330114 6.027188e-60
## logwage     0.5862581 1.7972507 0.09192779   6.377376 1.801477e-10
## reprate     1.3599071 3.8958316 0.43540020   3.123350 1.788051e-03
## disrate    -1.7980038 0.1656292 0.49999150  -3.596069 3.230624e-04
```

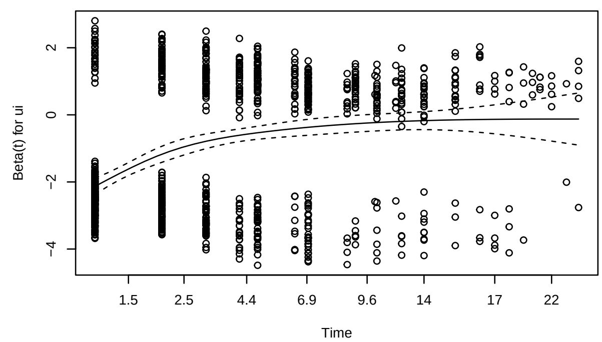
```
##           C       se(C)
## 0.698563255 0.008591571
```

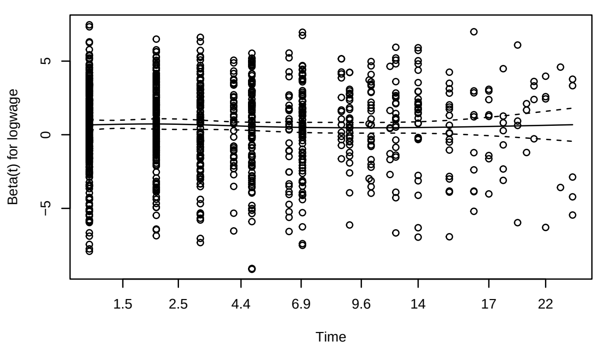
Test proportional Hazards assumption of Cox.2 model

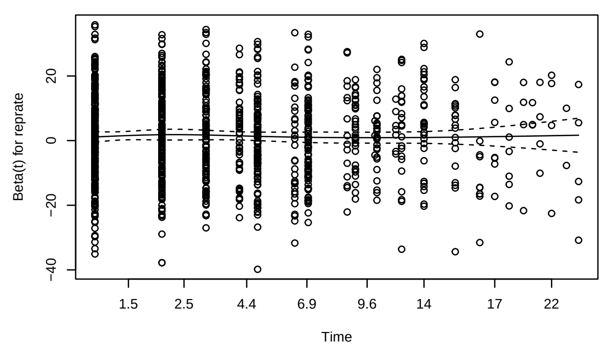
P-values of the Schoenfeld residuals test below are not statistically significant for all variables, except for UI covariate which has a p-value <2e-16 accompanied by a non-flat line residual. The global test is also significant. For this reason, we might conclude that the risk associated with UI covariate is not proportional.

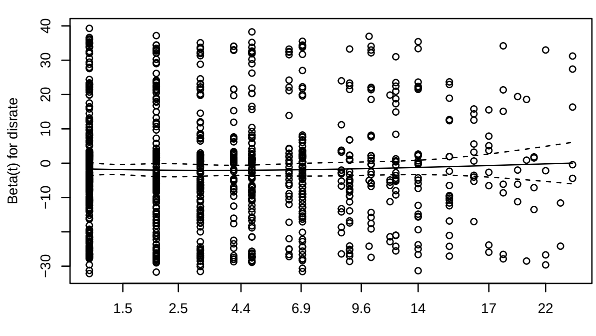
```
##            chisq df       p
## age        2.449  1    0.12
## ui       100.534  1 <2e-16
## logwage    0.139  1    0.71
## reprate    0.946  1    0.33
## disrate    1.327  1    0.25
## GLOBAL   113.710  5 <2e-16
```



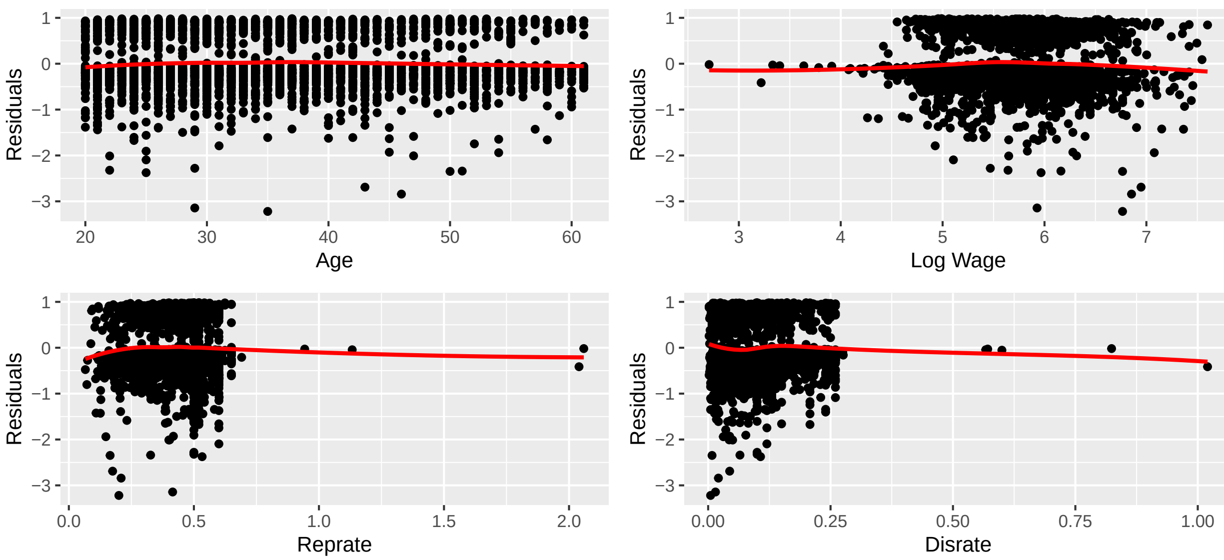








Test non linearity of Cox.2 model



For all continuous variable Age, Logwage, Reprate, Disrate martingale residual plots, the residuals fluctuate randomly around zero which does not suggest violations of the proportional hazards assumption.

3.3.2 Model selection based on AIC

3.3.2.1 Comparing non-nested models: AIC

Next step in our analysis is to compare Cox proportional Hazard Models. For this, we prepare the following Cox models:

- **M0**: null model
- **M_UI**: model with UI covariate
- **M_Job**: model with Job_rank covariate
- **M_AgeCat**: model with ageCat covariate.

Akaike Information Criterion is a general criterion of selecting survival models and attempts to determine how well the data fit the model. For now the best fit model is M_UI (AIC= 15664.39), since M0, M_Job and M_AgeCat show higher AIC scores, 15895.63, 15898.47 and 15883.57 respectively.

```
##       M0     M_UI    M_Job M_AgeCat
## 15895.63 15664.39 15898.47 15883.57
```

3.3.2.2 Step-wise model selection based on AIC

In the automatic model selection process, the covariates Age, UI, Logwage, Reprate, Disrate and Tenure are used to measure their significance to find a full-time job.

```
## Start:  AIC=15583.96
## Surv(ttr, event) ~ age + ui + logwage + tenure + reprate + disrate
##
##           Df   AIC
## - tenure   1 15583
## <none>       15584
## - reprate  1 15591
## - age      1 15595
## - disrate  1 15595
## - logwage  1 15625
## - ui       1 15835
##
```

```
##
## Step:  AIC=15583.01
## Surv(ttr, event) ~ age + ui + logwage + reprate + disrate
##
##           Df   AIC
## <none>       15583
## - reprate  1 15590
## - age      1 15593
## - disrate  1 15594
## - logwage  1 15625
## - ui       1 15833
```

We start with an AIC score of 15583.96 followed by an slight improvement in the step, where we have a score of 15583.01. Therefore, as we saw in class a smaller score “wins”, so this makes the second result a “better” model fit, as shown in the **Cox.2** model summary. With a concordance index of 0.698, last step includes Age, UI (UI=yes), Logwage, Reprate, Disrate as most significant variables. In this case, as previously shown, Tenure is also excluded by the automatic selection.

3.3.2.3 Transformed variables: AgeCat and JobRank

Using the transformed variables AgeCat and JobRank, we perform again the AIC test and analyse the results.

```
## Call:
## coxph(formula = Surv(ttr, event) ~ job_rank + ageCat, data = UnempDur_df)
##
##   n= 3343, number of events= 1073
##
##                        coef exp(coef)  se(coef)      z Pr(>|z|)
## job_rankMid        0.097743  1.102679  0.073975  1.321 0.186405
## job_rankSenior    -0.005383  0.994631  0.093863 -0.057 0.954267
## job_rankExecutive  0.078967  1.082169  0.101074  0.781 0.434637
## ageCat31-40       -0.010480  0.989575  0.075926 -0.138 0.890220
## ageCat41-50       -0.164157  0.848609  0.088601 -1.853 0.063916 .
## ageCat>=51        -0.397447  0.672033  0.110247 -3.605 0.000312 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                   exp(coef) exp(-coef) lower .95 upper .95
## job_rankMid          1.1027     0.9069    0.9539    1.2747
## job_rankSenior       0.9946     1.0054    0.8275    1.1955
## job_rankExecutive    1.0822     0.9241    0.8877    1.3193
## ageCat31-40          0.9896     1.0105    0.8527    1.1484
## ageCat41-50          0.8486     1.1784    0.7133    1.0095
## ageCat>=51           0.6720     1.4880    0.5414    0.8341
##
## Concordance= 0.535  (se = 0.01 )
## Likelihood ratio test= 20.38  on 6 df,   p=0.002
## Wald test            = 19.28  on 6 df,   p=0.004
## Score (logrank) test = 19.44  on 6 df,   p=0.003
```

From this result, with in AIC of 15887.25 the transformed variables are not performing better than the previous model and after comparing it with the null model (AIC = 15895.63) the model fits worse, nevertheless, and according to the summary AgeCat>=51 is the one and only significant Age variable as we saw also in **Cox.4** model.

4 Conclusion

After conducting non-parametric and semi-parametric tests on different covariates that might affect time to find a full-time job, we see a consistency in the result. As for UI covariate, in both studies, it shows that people claiming insurance take longer time to secure a full-time job. For Age, it is observed that younger groups have a shorter survival rate. Using semi-parametric test, the p-value of 51-plus group is significant, however, we cannot say the same to other age groups. On the other side, using log rank test, the **plot 2** shows there is considerable between the survival rate of 51-plus group and 20-30 group. Contrary to our initial assumption, surprisingly Tenure covariate has no effect on survival rate. Finally, Logwage, Reprate and Disrate are also good explanatory covariates on time to event. Using these covariates alongside with Age and UI give us the best model with the lowest AIC and a high concordance index.

References

“Data-Processing. Rmd.” 2023. https://github.com/adbcode/unemployment-survival-analysis/blob/master/resource/data-processing.rmd.

McCall, Brian P. 1996. “Unemployment Insurance Rules, Joblessness, and Part-Time Work.” *Econometrica* 64 (3): 647. https://doi.org/10.2307/2171865.

“People Receiving Unemployment Benefits in France 2021.” n.d. *Statista*. Accessed June 22, 2023. https://www.statista.com/statistics/757636/people-indemnified-unemployment-insurance/.

“R: Data Sets for Econometrics.” n.d. Accessed June 22, 2023. https://search.r-project.org/CRAN/refmans/Ecdat/html/00Index.html.

“R: Unemployment Duration.” n.d. Accessed June 22, 2023. https://search.r-project.org/CRAN/refmans/Ecdat/html/UnempDur.html.