Survival Unemployment

Final Project: PSTAT 175

Adam Trujillo, Maxine Wang, Venk Muriki

12/7/20

Part 1: Introduction

Scientific Question

Does the distinction between being white and non-white play a role in determining how long it takes an individual to get and keep a full-time job when unemployed?

Dataset Decription

This dataset is from the January Current Population Survey's Displaced Workers Supplements (DWS) for 1986, 1988, 1990, and 1992. The source was Econometrics Academy (https://sites.google.com/site/econometricsacademy/econometrics-models/survival-analysis), an online educational and non-profit organization. This dataset describes the United State's unemployed population, specifically their time until employment and other aspects of the individuals who are unemployed. There were 3343 subjects/data points in this dataset. The available covariates describe the subject who is/was unemployed.

The following is a list of the covariates available along with this description:

"spell": length of spell (unemployment) in number of two-week intervals
"event": (1) re-employed at a full-time job
"censor2": (1) re-employed at a part time
"censor3": (1) re-employed, but left job (part time or full time status is unknown)
"censor4": (1) still jobless
"ui": filled out unemployment insurance claim (1) yes (0) no
"reprate": eligible replacement rate
"logwage": log of weekly earnings from lost job
"tenure": years tenure in lost job (ranges from 0 to 40)
"disrate": eligible replacement rate
"slack": lost job due to slack work (1) yes (0) no
"abolpos": lost job due to abolished position (1) yes (0) no
"explose": expected to lose job (1) yes (0) no
"stateur": state unemployment rate ("houshead": only moneymaker of the house (1) yes (0) no
"married": is the subject married (1) yes (0) no
"female": (1) female (0) male
"child": (1) has children (0) does not have children
"ychild": (1) children under five (0) not children un-

der five
"nonwhite": (1) not white (0) white
"age": age in years (when the subject entered the study)
"schlt12": less than 12 years of schooling (1) yes (0) no
"schgt12": more than 12 years of schooling (1) yes (0) no
"smsa": resides in standard metropolitan statistical area (1) yes (0) no
"bluecoll": Blue collar as previous job (1) yes (0) no
"mining": Mining as previous job (1) yes (0) no
"constr": Construction as previous job (1) yes (0) no
"transp": Transportation as previous job (1) yes (0) no
"trade": Trade as previous job (1) yes (0) no
"fire": Fire as previous job (1) yes (0) no
"services": Services as previous job (1) yes (0) no
"pubadmin": Public Administration as previous job (1) yes (0) no
"year85": lost job in 1985 (1) yes (0) no
"year87": lost job in 1987 (1) yes (0) no
"year89": lost job in 1989 (1) yes (0) no
"midatl": region of residence is middle atlantic (1)

yes (0) no

"encen": region of residence is east north central (1)

yes (0) no

"wncen": region of residence is west north central (1)

yes (0) no

"southatl": region of residence is east south atlantic (1) yes (0) no

"escen": region of residence is east south central (1)

yes (0) no

"wscen": region of residence is west south central (1)

yes (0) no

"mountain": region of residence is mountain (1) yes (0) no

"pacific": region of residence is pacific (1) yes (0) no

Methods

The main question of this paper regards the covariate "non-white" and the censoring denoted as the covariate "event" which defines if a subject gets premanently re-employed at a full time job. After exploring this covariate, we will explore how being white relates to censor2 (a subject getting re-employed at and keeping a part time job), censor3 (a subject getting re-employed but then leaving the job again), and a few combinations between these two censors. Since censor4 is not independent of event, censor2, and censor3, this covariate is not giving us any new information, so we will not investigate this covariate, nor include it as part of our analyses from here on out.

For the main model regarding a subject getting re-employed at a full time job, we will first use forward selection to find a model with the lowest BIC. As part of model selection, we will not include spell (time covariate), the different type of censoring covariates (event, censor2, and censor3), and nonwhite (the covariate we are testing for) in the scope. The resulting model's covariates will be controlled for using stratification. Finally, we will add interaction terms of our interest to see if any are significant. The interaction terms that we are interested in include nonwhite's interaction with level of education, having children or not, marital status, residence, gender, and white collar or blue collar worker. If any interaction terms are significant, we will add them to the model and stratify on them.

After we obtain a satisfactory final model, we will perform diagnostic checks with Log-Log plots and the Goodness of Fit test to assert the Cox Proportional Hazards Assumption. Finally the analysis of the final model regarding non-white vs. white subjects' re-employment at a full time job will be conducted using Cox Regression Coefficients, Likelihood Ratio Tests, and Confidence Intervals of Hazard Ratios.

For the final portion of this paper, we will explore the different censors that this dataset provides using Competing Risk Models. We will be comparing 4 other different types of models, one for censor2 (getting re-employed at a and keeping part-time job), one for censor3 (getting re-employed and then leaving the job during the study), one for keepjob = event+censor2 (getting re-employed at a part-time or full-time job and remaining there throughout the entire study), and one for anyjob = event+censor2+censor3 (getting re-employed at a part-time or full-time job, disregarding what happens after the initial re-employment). For each, we will conduct forward selection and explore signficant interaction terms of our interst to find the build the best model with the lowest BIC (just as in the main part). Then we will look at the Hazard Proportion and Confidence Intervals for each of these new models in order to compare and further explore the relationship of being non-white/white in regards to unemployment more specifically with the type of job an individual can be re-employed with.
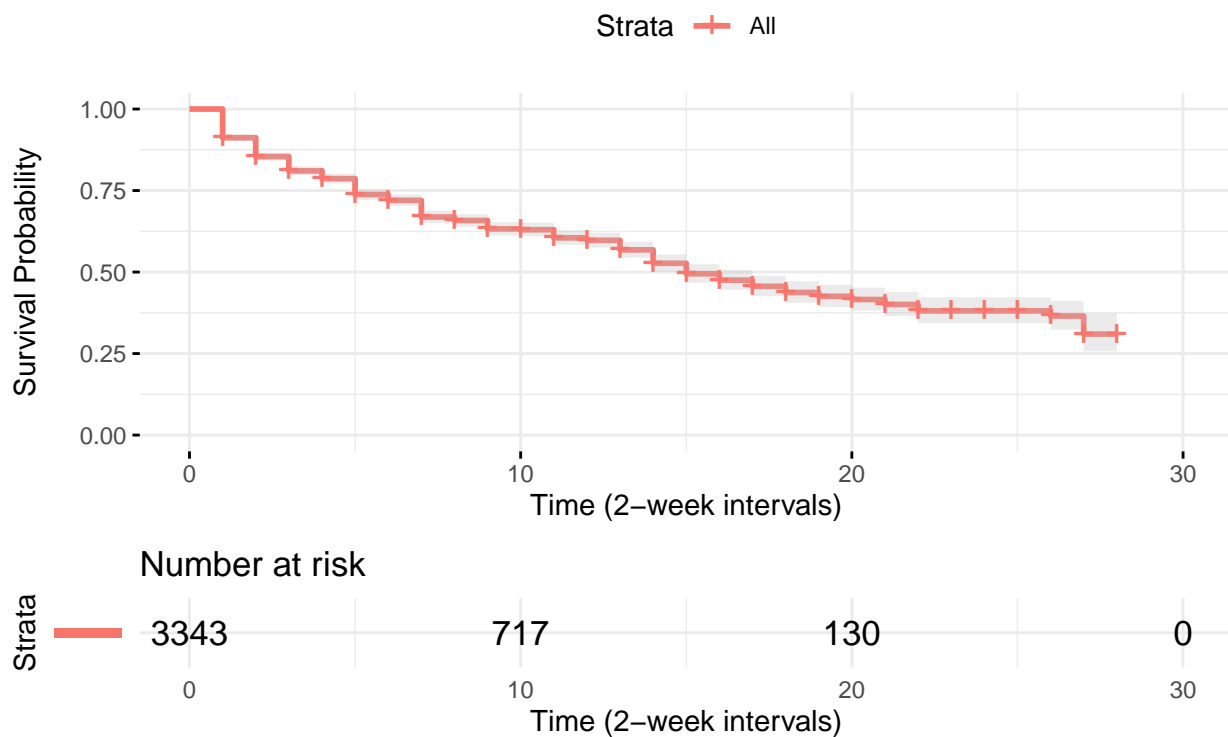
Part 2: Plots of the Data

Preliminary KM Estimate Plot of All Subjects (for the main censoring effect: event, getting re-employed at a full-time job)

```
unemp.fit <- surv_fit(Surv(spell, event) ~ 1, data = unemployment)
## Time is numbers of two-week intervals
```

```
ggsurvplot(unemp.fit,
           data = unemployment,
           risk.table = TRUE,
           conf.int = TRUE,
           xlab = "Time (2-week intervals)",
           ylab = "Survival Probability",
           title = "Kaplan-Meier Curve with 95% Confidence Interval
for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992",
           ggtheme = theme_minimal(),
           risk.table.y.text.col = T,
           risk.table.y.text = FALSE)
```


KM Estimate Plot to grouped into Non-white and White (for the main censoring effect: event, getting re-employed at a full-time job)

```
unemp.fit.nonwhite <- surv_fit(Surv(spell, event) ~ nonwhite, data = unemployment)
## Time is numbers of two-week intervals
ggsurvplot(unemp.fit.nonwhite,
           data = unemployment,
           risk.table = TRUE,
           conf.int = TRUE,
           xlab = "Time (2 week intervals)",
           ylab = "Survival Probability",
           title = "Kaplan-Meier Curve for White and Non-White Groups\n with 95% Confidence Interval for
```

```
        ggtheme = theme_minimal(),
        risk.table.y.text.col = T,
        risk.table.y.text = FALSE)
```

Kaplan–Meier Curve for White and Non–White Groups with 95% Confidence Interval for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992
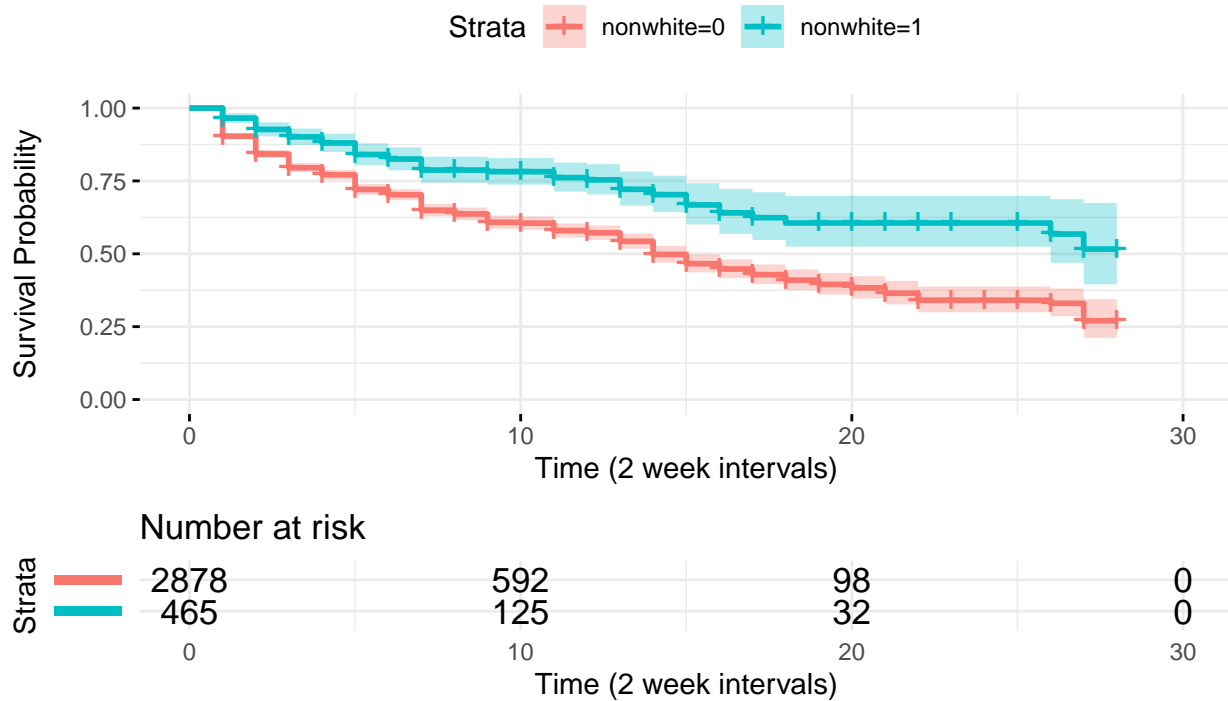
From this graph we can see that nonwhite subjects (1) have a higher survival probability (longer unemployment spell) than the white subjects (0). Therefore there seems to be a difference in the 2 groups, leaning towards to idea that nonwhite subjects have longer spells of unemployment till a full-time job. It is interesting to note that there are 2878 white subjects and 465 non-white subjects when looking at each censor.

Part 3: Model Selection

We are going to use forward selection to select the model with the lowest BIC since using BIC will greatly reduce the number of covariates in our model. The k = log(N) parameter in the step() function will produce a BIC where N is number of observations in the study. The scope of the search will not include spell, event, censor2, censor 3, censor 4, and nonwhite.

The output for the forward selection process using BIC is shown in Appendix A.

With BIC=15511.68, the best model is: $Surv(spell, event) \sim ui + bluecoll + logwage + slack + married + age + stateur + houshead + explose + wscen$

Control for forward selection model covariates with nonwhite covaraite (WITHOUT interaction)

In this part, we will use event (getting re-employed at and keeping a full-time job) as our status variable. In Part 6, where we look at competing risks models, we will use other status variables censor2 and censor3 as well.

Since age, logwage, and stateur are not categorical variables, we need to bin the data so we can stratify on them. We will bin the data based on the quantile output of the covariate.

```r
quantile(unemployment$age)
unemployment$age.cat <- cut(unemployment$age, breaks = c(20, 27, 34, 43, 61), right = TRUE,
              labels = c("20-27", "27-34", "34-43", "43-61"), include.lowest = TRUE)

quantile(unemployment$logwage)
unemployment$logwage.cat <- cut(unemployment$logwage,
                                breaks = c(2.70805, 5.29832, 5.67675, 6.05209, 7.60040),
                                right = TRUE,
                                labels = c("2.7-5.3", "5.3-5.7", "5.7-6.1", "6.1-7.6"),
                                include.lowest = TRUE)

quantile(unemployment$stateur)
unemployment$stateur.cat <- cut(unemployment$stateur,
                                breaks = c(2.5, 5.4, 6.6, 7.4, 13),
                                right = TRUE,
                                labels = c("2.5-5.4", "5.4-6.6", "6.6-7.4", "7.4-13"),
                                include.lowest = TRUE)
```

Here is the ANOVA Log Ratio test on the model in which all the supporting covariates are stratified on:

```r
race_control_no_int <- coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) +
                             strata(logwage.cat) + strata(slack) + strata(married) +
                             strata(age.cat) + strata(stateur.cat) + strata(houshead) +
                             strata(explose) + strata(wscen) + nonwhite, data = unemployment)

anova(race_control_no_int)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##           loglik  Chisq Df Pr(>|Chi|)
## NULL      -386.70
## nonwhite  -379.18 15.036  1  0.0001055 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After stratifying on the forward selection model covariates, the p-value to see if there is a difference in the rates of getting re-employed (hazard rates) among whites vs non-whites is significant at 0.0001055, considering a 0.05 significance level.

Adding Interaction Terms to the Model

Now we can control for interaction terms and see if the test comes out significant. The interaction terms that we sought to control for are that of nonwhite, with, education (schlt12 + schgt12), child (ychild + child),

married, area of residence (smsa), gender (female), and whether subject was a white or blue collar worker (bluecoll).

We will find the most significant interaction terms by removing the most insignificant (highest p-value) interaction term one by one until all interaction terms are significant at a level of 0.05. The p-value outputs and sequence of this selection is displayed in Appendix B, but the Anova LR test output of final coxph model with the most significant interaction term is below:

```
race_control_int <- coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) +
                          strata(logwage.cat) + strata(slack) + strata(married) +
                          strata(age.cat) + strata(stateur.cat) + strata(houshead) +
                          strata(explose) + strata(wscen) + nonwhite*(schgt12) +
                          nonwhite, data = unemployment)

anova(race_control_int)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                   loglik   Chisq Df Pr(>|Chi|)
## NULL             -386.70
## nonwhite         -379.18 15.0361  1  0.0001055 ***
## schgt12          -379.18  0.0007  1  0.9792082
## nonwhite:schgt12 -376.04  6.2775  1  0.0122279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term between nonwhite and schooling greater than 12 years turned out to be the only significant interaction term, at a significance value of 0.05, since the p-value is 0.0122279.

Inputting significant interaction terms into model

The main effect terms are automatically included when including interaction terms. We will also stratify on the interaction term so we control for it without testing for it.

```
race_control_cox_best_int <- coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) +
                                   strata(logwage.cat) + strata(slack) + strata(married) +
                                   strata(age.cat) + strata(stateur.cat) + strata(houshead) +
                                   strata(explose) + strata(wscen) + strata(nonwhite*schgt12) +
                                   nonwhite, data = unemployment)
anova(race_control_cox_best_int)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##           loglik  Chisq Df Pr(>|Chi|)
## NULL     -361.07
## nonwhite -350.71 20.716  1  5.326e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA LR test output has a p-value 5.326e-06 which is statistically significant at a significance level of 0.05. This means that there is a significant difference in the rates of getting re-employed at a full-time job between whites and non-whites after controlling for the stratified, significant covariates and interaction terms.

Therefore our final model (race_control_cox_best_int) for subjects getting and keeping a full-time job is: $Surv(spell, event) \sim strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + strata(married) + strata(age.cat) + strata(stateur.cat) + strata(houshead) + strata(explose) + strata(wscen) + strata(nonwhite * schgt12) + nonwhite$

Part 4: Performing Diagnostic Checks

We will use Log-Log Plots and the Goodness of Fit Test in order to check that the Proportional Hazards Assumption is reasonable for this model.

Log-Log Plots

```
cloglog <- function(x){log(-log(x))}
ggsurvplot(unemp.fit.nonwhite, fun=cloglog, xlab = "Time (in number of two-week intervals)",
           ylab = "Survival Probability",
           title = "Log-Log Plot for NonWhite Covariate",
           ggtheme = theme_minimal())
```
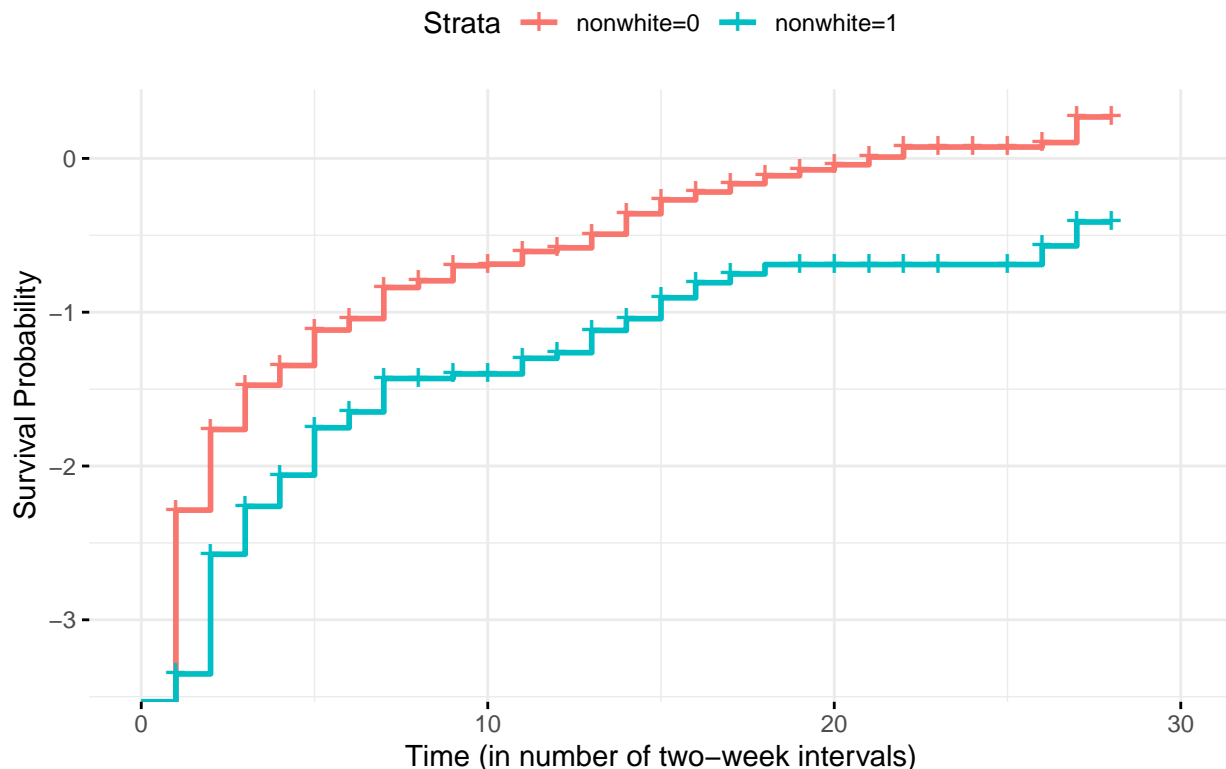

The white and nonwhite groups' log-log curves look generally parallel except for a section from the 20th to 25th two-week interval, which looks like the nonwhite group falls flat, but that should not be of much

concern because they are still generally parallel everywhere else and end in the same way as well. But maybe we should take a look at two-week intervals 20 to 30. There is also a straight section from about the 7th interval to the 10th, so that may also be another place we should look.

There is no obvious intersection, convergence, or divergence between the 2 curves therefore we can conclude that our Proportional Hazards Model is reasonable.

Goodness of Fit Test

```
cox.zph(race_control_cox_best_int, transform = "rank")
```

```
##              rho chisq     p
## nonwhite 0.0221 0.556 0.456
```

H_0: the PH Assumption is reasonable vs. H_1: the PH Assumption is violated.

Since Goodness of Fit Test results in a p-value of 0.456 for our covariate of interest (non-white), then accept the null hypothesis because it is greater than our significance value of 0.05. Therefore, we can conclude that the Proportional Hazards Assumption is reasonable.

Part 5: Analysis of Main Event (Getting and Keeping a Full-time job)

For our cox proportional hazards model, we will be testing that for the coefficient, $\beta_j = log$(Hazard Ratio); $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ for each of the covariate of interest, non-white. In plain English, we will be testing to see if there is a statistically significant difference in the survival probability of white and non-white subjects.

```
# using our final model: race_control_cox_best_int
anova(race_control_cox_best_int)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##           loglik  Chisq Df Pr(>|Chi|)
## NULL     -361.07
## nonwhite -350.71 20.716  1  5.326e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Likelihood Ratio Test

As mentioned earlier, the p-value for the Likelihood Ratio Test (which can be seen at the bottom of the summary of our Final Cox Proportional Hazard Model), is 5.326e-06 which is less than our significance value of 0.05, therefore we can reject the null hypothesis that $\beta_j = 0$ and conclude that $H_1 : \beta_j \neq 0$. This means that there is a statistically significant difference between the baseline hazard rate and the non-white hazard rate.

```
summary(race_control_cox_best_int)
```

```
## Call:
## coxph(formula = Surv(spell, event) ~ strata(ui) + strata(bluecoll) +
##     strata(logwage.cat) + strata(slack) + strata(married) + strata(age.cat) +
##     strata(stateur.cat) + strata(houshead) + strata(explose) +
##     strata(wscen) + strata(nonwhite * schgt12) + nonwhite, data = unemployment)
##
##   n= 3343, number of events= 1073
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## nonwhite -1.2684    0.2813   0.3112 -4.076 4.59e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## nonwhite    0.2813      3.555    0.1528    0.5177
##
## Concordance= 0.546  (se = 0.013 )
## Likelihood ratio test= 20.72  on 1 df,    p=5e-06
## Wald test            = 16.61  on 1 df,    p=5e-05
## Score (logrank) test = 18.67  on 1 df,    p=2e-05
```

Confidence Intervals

The 95% confidence interval for the non-white covariate hazard ratio is (0.1528, 0.5177). This excludes 1, which confirms our LR test results that there is a significant difference between the non-white covaraite hazard rate and the baseline hazard rate for white subjects. Since the confidence interval is wide, there is some uncertainty in estimating the magnitude of the hazard ratio.

Coefficient/Hazard Ratio

More specifically, the Cox Proportional Hazards Model's Coefficient ($beta_j$) is -1.2684 therefore, since it is negative we can conclude that the hazard rate for non-white subjects (h_1) is less than the baseline hazard rate (h_0), therefore the survival probability (unemployment spell) for non-white subjects is significantly higher than the baseline population (white subjects) for this dataset. For this paper, we defined the "event" which occurs as getting re-employed, therefore survival is defined as not getting re-employed. Therefore the negative coefficient of our final Cox PH Model tells us that non-white subjects remain unemployed for longer than white subjects. The hazard ratio is 0.2813, which implies that the rates at which non-white subjects can land a non-provisional full-time job is 28.13% of the rates at which white subjects can.

Part 6: Competing Risk Models

Since this dataset explores a situation which has many different outcomes, different censoring types were provided. For example, a subject could be re-employed at a full-time or part-time job, or they could be re-employed and then leave the job again during the study, or they could never get re-employed at all. For this section, we will explore the different types of censoring this dataset provides using Competing Risk Models.

First, we will explore censor2 which defines a subject getting re-employed at a part time job and keeping it. Second we will explore censor3 which defines a subject who got re-employed but then left the job again.

Then we will combine our original event (event: full-time) and censor2 to look non-white subjects getting re-employed and staying employed till the end of the study, at any type of job. Finally, we will combine our original censor (event: full-time), censor2, and censor3 to explore non-white subjects getting initially re-employed at any type of job (disregarding the type of job or if they do not keep the job).

We will use Forward Selection (BIC) to find the best model and the covariates we need to stratify on for each different type of censoring. Since the output for forward selection takes up too much space, is is not shown (reference Appendix A for the example of output if needed), but the code can be referenced in Appendix C. We will add the same interaction terms of interest as we did in the main model selection in Part 3 (reference Appendix D for this code if needed) to see if any are significant, if they are they will be added to the model for interpretation.

Getting a Part-Time Job and Keeping it (censor2)

```
unemp.fit.nonwhite.2 <- surv_fit(Surv(spell, censor2) ~ nonwhite, data = unemployment)
## Time is numbers of two-week intervals
ggsurvplot(unemp.fit.nonwhite.2,
           data = unemployment,
           risk.table = TRUE,
           conf.int = TRUE,
           xlab = "Time (2 week intervals)",
           ylab = "Survival Probability",
           title = "Kaplan-Meier Curve for Censor2 \nwith 95% Confidence Interval
           for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992",
           ggtheme = theme_minimal(),
           risk.table.y.text.col = T,
           risk.table.y.text = FALSE)
```

Kaplan–Meier Curve for Censor2
with 95% Confidence Interval
for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992
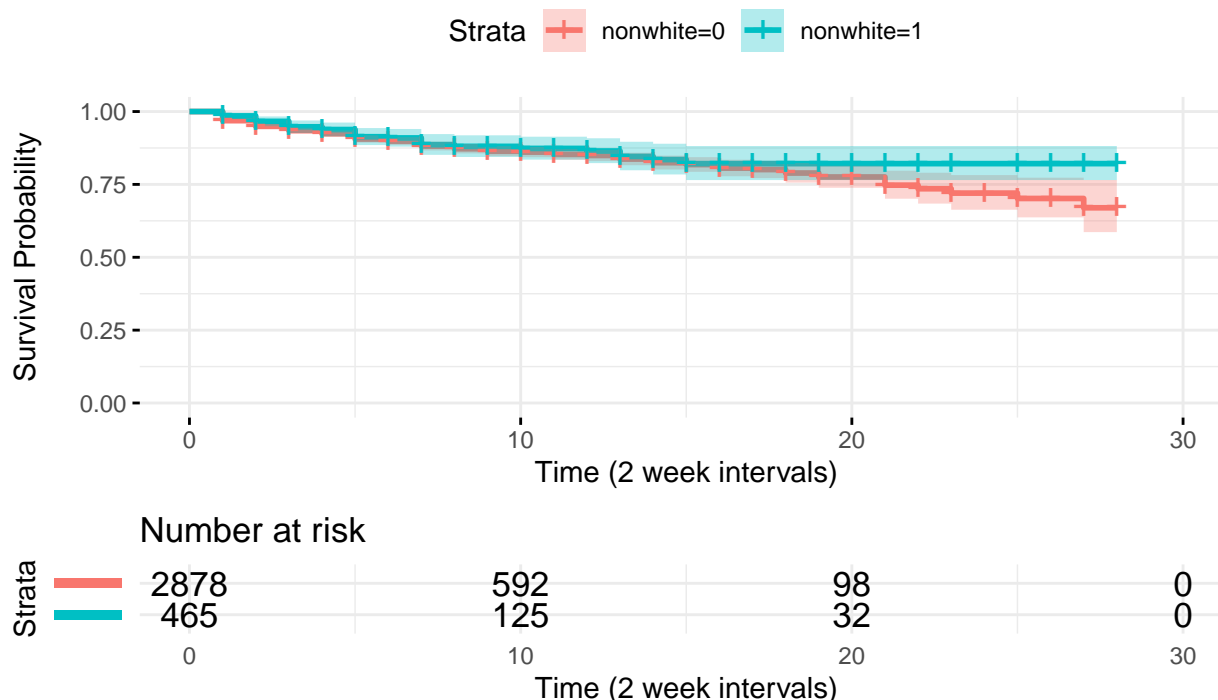
There is not an extreme difference in white and non-white KM plots. It is mainly the same and starts to separate after 20 two-week intervals Non-white (1) subjects have a higher survival probability (longer unemployment spell) after the 20th quarter of the study, when the 2 groups start to diverge.

In Appendix D, we tried to search for significant interaction terms to add to our model that we built with forward selection and BIC in Appendix C.

```
part_time <- coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) +
                   strata(stateur.cat) + nonwhite, data = unemployment)
# adding interaction terms of interest
part_time_int <- coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) +
                       strata(stateur.cat) + nonwhite*smsa + nonwhite, data = unemployment)
anova(part_time_int)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, censor2)
## Terms added sequentially (first to last)
##
##                loglik  Chisq Df Pr(>|Chi|)
## NULL          -1560.7
## nonwhite      -1557.8 5.9570  1    0.01466 *
## smsa          -1557.7 0.0415  1    0.83865
## nonwhite:smsa -1555.0 5.5042  1    0.01897 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As shown in Appendix D, this process resulted with nonwhite*smsa being the only significant interaction term with a p-value of 0.01897, so we stratified for this interaction term in our final model.

```
part_time_best_int <- coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) +
                                strata(stateur.cat) + strata(nonwhite*smsa) +
                                nonwhite, data = unemployment)
anova(part_time_best_int)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, censor2)
## Terms added sequentially (first to last)
##
##            loglik  Chisq Df Pr(>|Chi|)
## NULL      -1458.8
## nonwhite  -1458.6 0.3294  1      0.566
```

After controlling for the other significant covariates interaction terms, the final model is: $Surv(spell, censor2) \sim strata(ui) + strata(female) + strata(stateur.cat) + strata(nonwhite * smsa) + nonwhite$

When considering censor2, where the subject gets re-employed at and keeps a part-time job, the p-value of the Likelihood Ratio Test is 0.566 which is larger than the significance level 0.05.

```
summary(part_time_best_int)
```

```
## Call:
## coxph(formula = Surv(spell, censor2) ~ strata(ui) + strata(female) +
##     strata(stateur.cat) + strata(nonwhite * smsa) + nonwhite,
##     data = unemployment)
##
##   n= 3343, number of events= 339
##
##           coef exp(coef) se(coef)     z Pr(>|z|)
## nonwhite 0.153     1.165    0.261 0.586    0.558
##
##          exp(coef) exp(-coef) lower .95 upper .95
## nonwhite     1.165     0.8581    0.6988     1.943
##
## Concordance= 0.509  (se = 0.008 )
## Likelihood ratio test= 0.33  on 1 df,    p=0.6
## Wald test            = 0.34  on 1 df,    p=0.6
## Score (logrank) test = 0.34  on 1 df,    p=0.6
```

The 95% confidence interval for the hazard ratio is (0.6988, 1.943) which includes 1. Since the confidence interval is wide, there is uncertainty in estimating the magnitude of the hazard ratio.

The p-value and confidence interval confirm that there is no statistically significant evidence that there is a difference between non-white and white subjects for the time it takes them to get re-employed at and keep a part time job. Therefore there is no need for further investigation.

Getting a job and then losing it (censor3)

```
unemp.fit.nonwhite.3 <- surv_fit(Surv(spell, censor3) ~ nonwhite, data = unemployment)
## Time is numbers of two-week intervals
```

```
ggsurvplot(unemp.fit.nonwhite.3,
           data = unemployment,
           risk.table = TRUE,
           conf.int = TRUE,
           xlab = "Time (2 week intervals)",
           ylab = "Survival Probability",
           title = "Kaplan-Meier Curve for Censor3 \nwith 95% Confidence Interval
           for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992",
           ggtheme = theme_minimal(),
           risk.table.y.text.col = T,
           risk.table.y.text = FALSE)
```


Kaplan−Meier Curve for Censor3
with 95% Confidence Interval
for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992
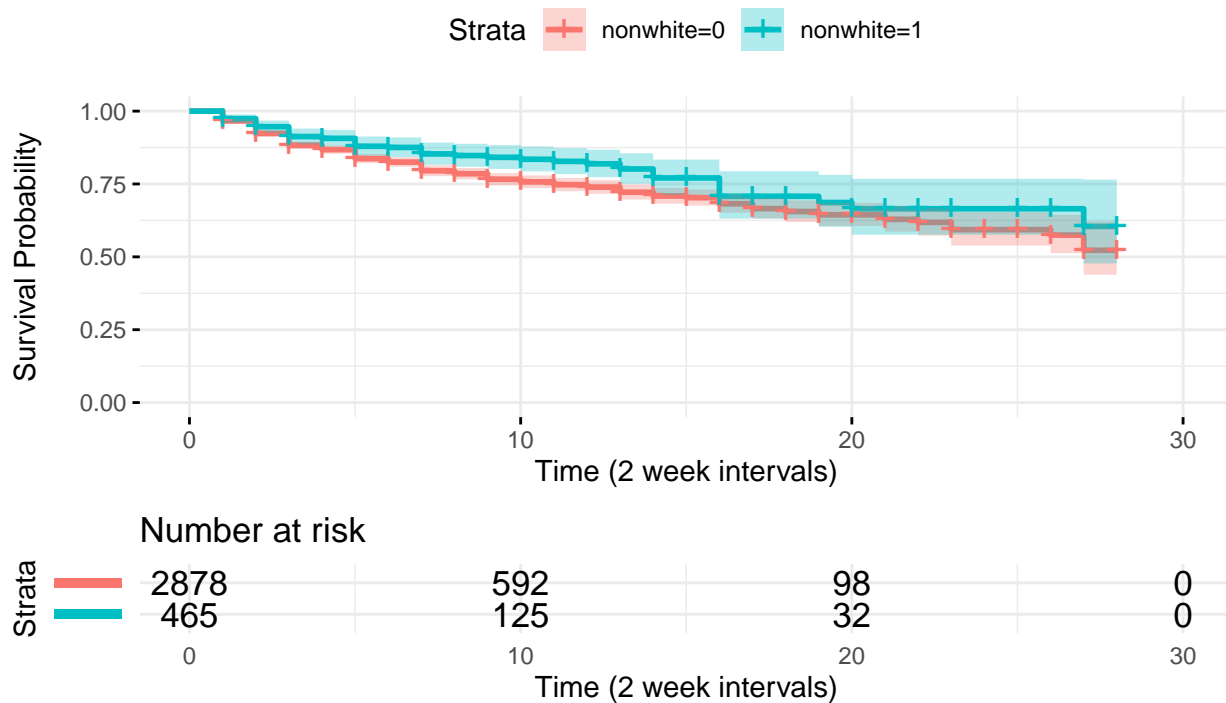
There is a slight difference in the KM plots for white and non-white throughout the entire time of the study, with the non-white (1) group having a higher survival probability (longer unemployment spell) than the white group (0).

We need to bin tenure to transform it into a categorical covariate.

```
quantile(unemployment$tenure, probs = c(0, 1/3, 2/3, 1))
unemployment$tenure.cat <- cut(unemployment$tenure,
                               breaks = c(0, 1, 4, 40),
                               right = TRUE,
                               labels = c("0-1", "1-4", "4-40"),
                               include.lowest = TRUE)
```

We will now control for the other covariates found from forward selection (see code in Appendix C). As seen in Appendix D, since no interaction terms were significant, so we don't need to include them into the model.

```r
cen3_cox <- coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) +
                    strata(slack) + strata(houshead) + strata(age.cat) +
                    strata(year87) + strata(escen) + strata(midatl) +
                    strata(abolpos) + nonwhite, data = unemployment)
anova(cen3_cox)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, censor3)
## Terms added sequentially (first to last)
##
##           loglik Chisq Df Pr(>|Chi|)
## NULL     -1159.1
## nonwhite -1156.3 5.599  1    0.01797 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After controlling for the other significant covariates, the final model is: $Surv(spell, censor3) \sim strata(ui) + strata(tenure.cat) + strata(slack) + strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) + strata(midatl) + strata(abolpos) + nonwhite$

When considering censor3, where the subject gets re-employed but then loses the job again, the p-value of the Likelihood Ratio Test is 0.01797 which is statistically significant at 0.05 significance level.

```r
summary(cen3_cox)
```

```
## Call:
## coxph(formula = Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) +
##     strata(slack) + strata(houshead) + strata(age.cat) + strata(year87) +
##     strata(escen) + strata(midatl) + strata(abolpos) + nonwhite,
##     data = unemployment)
##
##   n= 3343, number of events= 574
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## nonwhite -0.3333    0.7166   0.1454 -2.292   0.0219 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## nonwhite    0.7166      1.396    0.5389    0.9529
##
## Concordance= 0.526  (se = 0.011 )
## Likelihood ratio test= 5.6  on 1 df,   p=0.02
## Wald test            = 5.25  on 1 df,   p=0.02
## Score (logrank) test = 5.29  on 1 df,   p=0.02
```

The 95% confidence interval for the hazard ratio is (0.5389, 0.9529) which does not include 1. Since the confidence interval is pretty wide, there is some uncertainty in estimating the magnitude of the hazard ratio.

So the LR test p-value and confidence interval confirm that there is a statistically significant difference between the time it takes non-white and white subjects to get re-employed and then leave the job again during the study.

The Cox Proportional Hazards Model's Coefficient ($beta_j$) is -0.3333 therefore, since it is negative we can conclude that the hazard rate for non-white subjects (h_1) is less than the baseline hazard rate (h_0), therefore the survival probability for non-white subjects is significantly higher than the baseline population for this dataset. Regarding specifically censor3, the negative coefficient of this censor's Cox PH Model tells us that non-white subjects take longer to get re-employed and leave the job again during the study. The hazard ratio is 0.7166, which implies that the rates at which non-white subjects can get any job and lose it is 71.66% of the rates at which white subjects can.

Getting a Part-Time or Full-Time Job and Keeping that job throughout the study (keepjob = event+censor2)

```
unemp.fit.nonwhite.keep <- surv_fit(Surv(spell, keepjob) ~ nonwhite, data = unemployment)
## Time is numbers of two-week intervals
ggsurvplot(unemp.fit.nonwhite.keep,
           data = unemployment,
           risk.table = TRUE,
           conf.int = TRUE,
           xlab = "Time (2 week intervals)",
           ylab = "Survival Probability",
           title = "Kaplan-Meier Curve for KeepJob Censor \nwith 95% Confidence Interval
           for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992",
           ggtheme = theme_minimal(),
           risk.table.y.text.col = T,
           risk.table.y.text = FALSE)
```
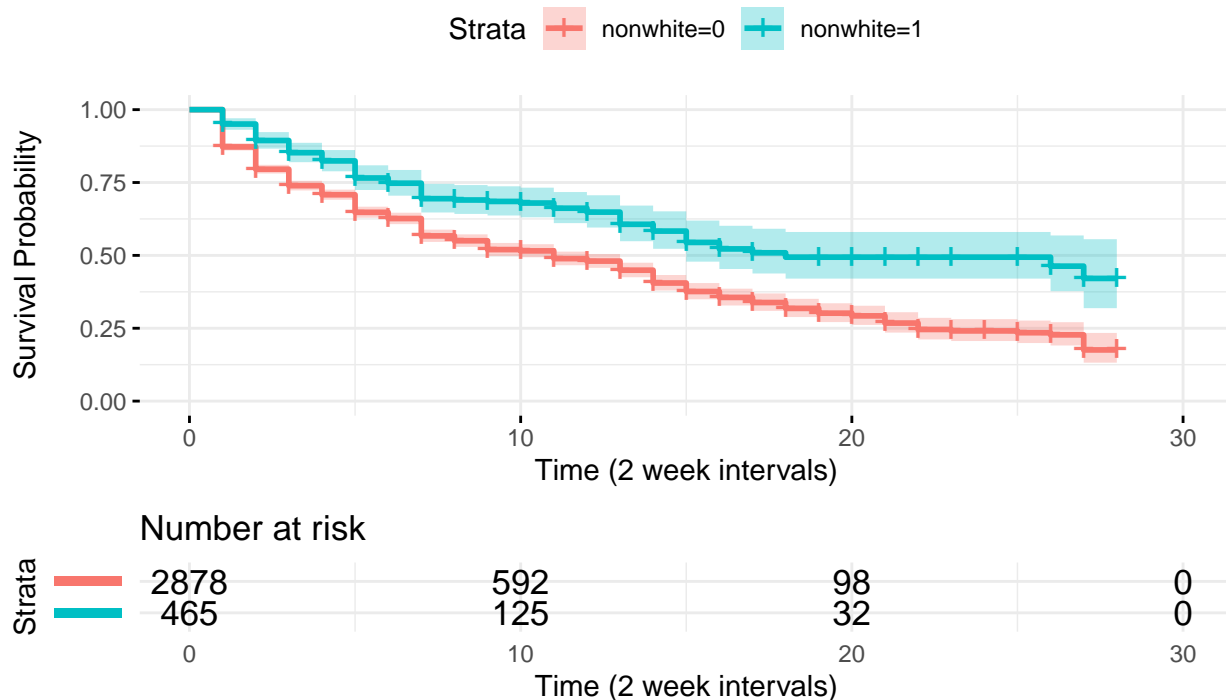

The difference in the survival probability for the 2 groups is bigger for this censor than the first 2 censors,

this probably means that we will discover more significant results when we dive deeper into this study. The non-white (1) group still has a higher survival probability (longer unemployment spell) than the white (0) group.

We need to transform disrate into a categorical variable.

```
quantile(unemployment$disrate)
unemployment$disrate.cat <- cut(unemployment$disrate,
                                breaks = c(0.0020, 0.0525, 0.1000, 0.1500, 1.0200),
                                right = TRUE,
                                labels = c("0-0.05", "0.05-0.1", "0.1-0.15", "0.15-1"),
                                include.lowest = TRUE)
```

We will now control for the other covariates found from forward selection (see code in Appendix C). As seen in Appendix D, since no interaction terms were significant, we don't need to include them into the model.

```
keepjob_cox <- coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) +
                     strata(stateur.cat) + strata(slack) + strata(married) +
                     strata(age.cat) + strata(houshead) + strata(explose) +
                     strata(wscen) + strata(disrate.cat) + nonwhite, data = unemployment)
anova(keepjob_cox)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, keepjob)
## Terms added sequentially (first to last)
##
##           loglik  Chisq Df Pr(>|Chi|)
## NULL     -471.60
## nonwhite -466.54 10.121  1   0.001465 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After controlling for the other significant covariates, the final model is: $Surv(spell, keepjob)\ strata(ui) + strata(bluecoll) + strata(stateur.cat) + strata(slack) + strata(married) + strata(age.cat) + strata(houshead) + strata(explose) + strata(wscen) + strata(disrate.cat) + nonwhite$

When considering event and censor2, where the subject gets a part-time or full-time job and keeps it throughout the entire study, the p-value of the Likelihood Ratio Test is 0.001465 which is statistically significant at 0.05 significance level.

```
summary(keepjob_cox)
```

```
## Call:
## coxph(formula = Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) +
##     strata(stateur.cat) + strata(slack) + strata(married) + strata(age.cat) +
##     strata(houshead) + strata(explose) + strata(wscen) + strata(disrate.cat) +
##     nonwhite, data = unemployment)
##
##   n= 3343, number of events= 1412
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## nonwhite -0.5715    0.5647   0.1852 -3.087  0.00202 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## nonwhite    0.5647      1.771    0.3928    0.8117
##
## Concordance= 0.543  (se = 0.012 )
## Likelihood ratio test= 10.12  on 1 df,   p=0.001
## Wald test            = 9.53  on 1 df,   p=0.002
## Score (logrank) test = 9.75  on 1 df,   p=0.002
```

The 95% confidence interval for the hazard ratio is (0.3928, 0.8117) which does not include 1. Since the confidence interval is pretty wide, there is some uncertainty in estimating the magnitude of the hazard ratio.

The LR test low p-value and confidence interval directs us to conclude that there is a statistically significant difference between the time it takes non-white and white subjects to get re-employed at any type of job and keeping it throughout the entire study.

The Cox Proportional Hazards Model's Coefficient ($beta_j$) is -0.5715 therefore, since it is negative we can conclude that the hazard rate for non-white subjects (h_1) is less than the baseline hazard rate (h_0), therefore the survival probability for non-white subjects is significantly higher than the baseline population for this dataset. Regarding specifically keepjob censor (event+censor2), the negative coefficient of this censor's Cox PH Model tells us that non-white subjects take longer to get re-employed and keep the job. The hazard ratio is 0.5647, which implies that the rates at which non-white subjects can get any job and keep it is 56.47% of the rates at which white subjects can.

Getting Re-Employed at all (anyjob = event+censor2+censor3)

```
unemp.fit.nonwhite.any <- surv_fit(Surv(spell, anyjob) ~ nonwhite, data = unemployment)
## Time is numbers of two-week intervals
ggsurvplot(unemp.fit.nonwhite.any,
          data = unemployment,
          risk.table = TRUE,
          conf.int = TRUE,
          xlab = "Time (2 week intervals)",
          ylab = "Survival Probability",
          title = "Kaplan-Meier Curve for AnyJob Censor \nwith 95% Confidence Interval
          for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992",
          ggtheme = theme_minimal(),
          risk.table.y.text.col = T,
          risk.table.y.text = FALSE)
```

Kaplan–Meier Curve for AnyJob Censor
with 95% Confidence Interval
for Unemployed U.S. individuals in 1986, 1988, 1990, and 1992
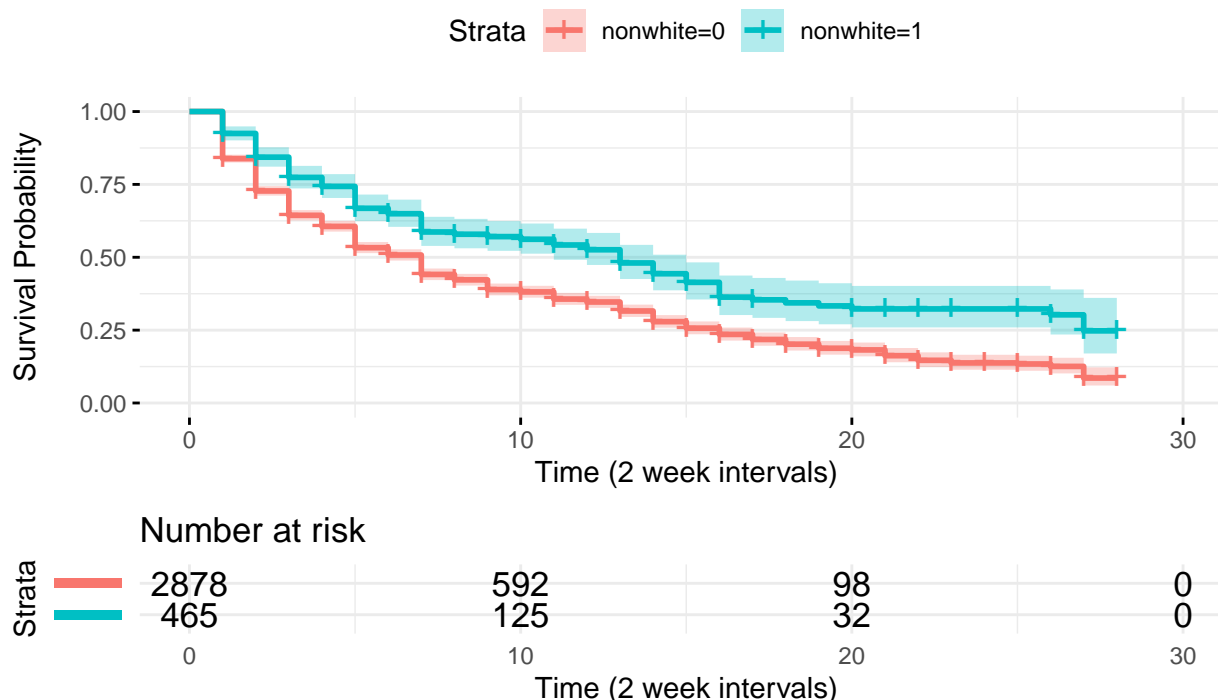
Finally, the difference between the 2 groups is quite clear for the censor that regards any initial re-employment. Once again the non-white (1) group has a higher survival probability (longer unemployment spell) than the white group throughout the entire time of the study just as in the rest of the different types of censors.

We will now control for the other covariates found from forward selection (see code in Appendix C). As seen in Appendix D, since no interaction terms were significant, we don't need to include them into the model.

```
any_cox <- coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
                strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
                strata(married) + strata(year87) + strata(midatl) + strata(encen) +
                strata(child) + nonwhite, data = unemployment)
anova(any_cox)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, anyjob)
## Terms added sequentially (first to last)
##
##           loglik  Chisq Df Pr(>|Chi|)
## NULL     -1060.4
## nonwhite -1048.0 24.809  1  6.329e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After controlling for the other significant covariates, the final model is: $Surv(spell, anyjob)$ $strata(ui) + strata(slack) + strata(stateur.cat) + strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) + strata(married) + strata(year87) + strata(midatl) + strata(encen) + strata(child) + nonwhite$

When considering event, censor2, and censor3, where the subject gets re-employed at all (disregarding what happens after initial re-employment), the p-value of the Likelihood Ratio Test is 6.329e-07 which is statistically significant at 0.05 significance level.

```
summary(any_cox)
```

```
## Call:
## coxph(formula = Surv(spell, anyjob) ~ strata(ui) + strata(slack) +
##     strata(stateur.cat) + strata(age.cat) + strata(houshead) +
##     strata(bluecoll) + strata(abolpos) + strata(married) + strata(year87) +
##     strata(midatl) + strata(encen) + strata(child) + nonwhite,
##     data = unemployment)
##
##   n= 3343, number of events= 1986
##
##            coef exp(coef) se(coef)      z Pr(>|z|)
## nonwhite -0.6218    0.5370   0.1299 -4.785 1.71e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## nonwhite     0.537      1.862    0.4163    0.6927
##
## Concordance= 0.538  (se = 0.009 )
## Likelihood ratio test= 24.81  on 1 df,    p=6e-07
## Wald test            = 22.9  on 1 df,    p=2e-06
## Score (logrank) test = 23.51  on 1 df,    p=1e-06
```

The 95% confidence interval for the hazard ratio is (0.4163, 0.6927) which does not include 1. Since the confidence interval is pretty wide, there is some uncertainty in estimating the magnitude of the hazard ratio.

Based on the low p-value and confidence interval, we can conclude that there is a difference between the time it takes non-white and white subjects to get initially re-employed at any type of job.

The Cox Proportional Hazards Model's Coefficient ($beta_j$) is -0.6218 therefore, since it is negative we can conclude that the hazard rate for non-white subjects (h_1) is less than the baseline hazard rate (h_0), therefore the survival probability for non-white subjects is significantly higher than the baseline population for this dataset. Regarding specifically censor3, the negative coefficient of this censor's Cox PH Model tells us that non-white subjects take longer to get re-employed at all. The hazard ratio is 0.5370, which implies that the rates at which non-white subjects can get initially re-employed is 53.7% of the rates at which white subjects can.

Conclusion

The objective of this study was to investigate if there is a distinction between white and non-white individuals regarding how long it takes for them to get re-employed at a non-provisional full-time job. For this study we defined the "event" to occur as "getting a job" and "survival" to be "unemployment". We came to the conclusion that there is a statistically significant difference in the time it takes for white and non-white subjects to get re-employed at a full-time job and remain in that job until the study ends. The coefficient -1.2684 from the final Cox Proportional Hazard Model shows that non-white subjects have higher survival rates than the baseline. Therefore, since non-white subjects have higher survival rates, they tend to remain unemployed for longer.

Beyond our Cox PH model with full-time re-employment as the status variable, we decided to explore the other types of censoring provided by this dataset using competing-risk models (CRM). The other censoring is as follows: censor2 which denotes the subject getting re-employed at a part time job and keeping that job, censor3 which denotes the subject getting re-employed but then losing the job again, keepjob(event+censor2) which defines a subject getting a full-time or part-time job and then keeping it throughout the entire study, and anyjob(event+censor2+censor3) which denotes a subject getting a re-employed in any type of job, but disregards what happens after initial re-employment.

The first altered model regarded censor2 (a subject getting re-employed at and staying in a part-time job). After finding the most suitable model for this new censor with forward selection and exploring our interaction terms, we found that the interaction between being non-white and residing in a standard metropolitan statistical area is significant, therefore we included it in the model for analysis. After running an ANOVA Likelihood Ratio test on this model the non-white covariate became insignificant. Therefore, being non-white makes no statistically significant difference in the time it takes for a subject to get permanently re-employed at a part-time job.

The second, third and fourth altered models regard censor3 (a subject getting re-employed at a job and then losing it again before the study ends), the keepjob censor (a subject getting a full-time or part-time job and then keeping it throughout the entire study), and the anyjob censor (a subject getting a re-employed in any type of job). We found the best models using forward selection and testing the significance of each interaction term just as we did with every other model. Then we ran an ANOVA Likelihood Ratio test for each of these additional models and they all resulted in statistically significant results. For each of these models the coefficient from the final Cox Proportional Hazard Model was negative, meaning that the survival rate of non-white subjects is higher than the baseline. Therefore non-white individuals are unemployed significantly longer than the baseline for temporary re-employment, non-temporary re-employment, and initial re-employment in any type of job.

Comparing these risk models, we see that censor3 has the highest hazard ratio (0.7166), therefore non-white individuals have a the highest chance of being re-employed more quickly for temporary re-employment. The second highest hazard ratio (0.5647) is for keepjob which implies that non-white individuals have a moderate chance of being re-employed in a timely manner for permanent re-employment at a part-time or full-time job. What is interesting is that the hazard ratio for initial re-employment is slightly lower than the ratio for non-temporary re-employment, comparing a hazard ratio of 0.5370 for initial re-employment to any job type to the hazard ratio of 0.5647 for non-temporary re-employment. Finally, the censor with the lowest hazard ratio is for permanent full-time re-employment. These conclusions answer our original question, that there are discrepancies on the time of re-employment at a non-temporary full-time job between white and non-white individuals. As mentioned in Part 3, the hazard ratio was 0.2183, which means that for every 1 non-temporary full-time re-employment per two week interval for a white subject there are 0.2183 non-temporary full-time re-employments per two week interval for a non-white subject.

In conclusion, for the main scientific question of this study, we see that the distinction between being white and non-white does in fact play a role in determining how long it takes an individual to get and keep a full-time job when unemployed. Specifically, it will take a non-white individual longer to get re-employed at a permanent full-time job than a white individual. For all of the competing-risk models, there were significant differences in re-employment rates between white and non-white subjects, with the exception of part-time re-employment. Although the main focus of this paper was to discover if discrimination based on race in the re-employment process is statistically significant, this topic could be further explored in different areas of discrimination including gender, religious beliefs, sexual orientation, etc. It could also be taken one step further by looking at different combinations of these social groups and the interaction between them.

Appendix

Appendix A: original model selection (forward selection)

This is the model selection for getting and keeping a full-time job. Since this output is very long, we are only showing it for this competing risk model. All the other forward selection outputs for the other competing risks will be done the same way as shown below, except with the status variable changed.

```
unemp.fit.scope <- coxph(Surv(spell, event) ~
                         ui + reprate + logwage + tenure + disrate +
                         slack + abolpos + explose + stateur + houshead +
                         married + child + ychild + female + age +
                         schlt12 + schgt12 + smsa + bluecoll + mining +
                         constr + transp + trade + fire + services +
                         pubadmin + year85 + year87 + year89 + midatl +
                         encen + wncen + southatl + escen + wscen +
                         mountain + pacific, data = unemployment)

fit0 <- coxph(Surv(spell, event) ~ 1, data = unemployment)  # starting model
fitf_bic <- step(fit0, scope = formula(unemp.fit.scope), direction = "forward",
                 k = log(nrow(unemployment)))
```

```
## Start:  AIC=15895.63
## Surv(spell, event) ~ 1
##
##              Df   AIC
## + ui          1 15670
## + bluecoll    1 15841
## + slack       1 15858
## + schgt12     1 15873
## + disrate     1 15874
## + stateur     1 15874
## + logwage     1 15882
## + year89      1 15885
## + age         1 15887
## + married     1 15887
## + fire        1 15888
## + smsa        1 15888
## + year87      1 15889
## + houshead    1 15890
## + reprate     1 15893
## + explose     1 15893
## + encen       1 15894
## + wscen       1 15894
## + services    1 15894
## <none>          15896
## + schlt12     1 15897
## + abolpos     1 15898
## + southatl    1 15899
## + midatl      1 15900
## + year85      1 15902
## + tenure      1 15902
## + wncen       1 15902
```

```
## + mining    1 15902
## + child     1 15902
## + trade     1 15902
## + ychild    1 15903
## + constr    1 15903
## + escen     1 15904
## + female    1 15904
## + mountain  1 15904
## + pubadmin  1 15904
## + transp    1 15904
## + pacific   1 15904
##
## Step:  AIC=15670.5
## Surv(spell, event) ~ ui
##
##             Df   AIC
## + bluecoll  1 15611
## + logwage   1 15619
## + slack     1 15632
## + schgt12   1 15643
## + married   1 15644
## + disrate   1 15645
## + stateur   1 15650
## + houshead  1 15654
## + reprate   1 15657
## + year89    1 15662
## + explose   1 15664
## + smsa      1 15664
## + fire      1 15666
## + encen     1 15668
## <none>        15670
## + schlt12   1 15670
## + year87    1 15671
## + wscen     1 15673
## + services  1 15673
## + abolpos   1 15673
## + year85    1 15675
## + tenure    1 15676
## + constr    1 15676
## + southatl  1 15676
## + wncen     1 15676
## + child     1 15677
## + age       1 15677
## + midatl    1 15677
## + mountain  1 15678
## + ychild    1 15678
## + escen     1 15678
## + transp    1 15678
## + pubadmin  1 15678
## + mining    1 15678
## + female    1 15678
## + trade     1 15678
## + pacific   1 15678
##
```

```
## Step:  AIC=15611.09
## Surv(spell, event) ~ ui + bluecoll
##
##             Df    AIC
## + logwage    1  15584
## + married    1  15586
## + houshead   1  15593
## + disrate    1  15594
## + slack      1  15594
## + stateur    1  15598
## + explose    1  15604
## + year89     1  15604
## + reprate    1  15609
## + encen      1  15609
## + schgt12    1  15610
## + year87     1  15611
## <none>          15611
## + smsa       1  15612
## + female     1  15612
## + wscen      1  15613
## + child      1  15616
## + age        1  15616
## + fire       1  15616
## + wncen      1  15617
## + ychild     1  15617
## + southatl   1  15617
## + year85     1  15617
## + trade      1  15618
## + mountain   1  15618
## + tenure     1  15618
## + midatl     1  15618
## + services   1  15618
## + schlt12    1  15619
## + transp     1  15619
## + abolpos    1  15619
## + constr     1  15619
## + escen      1  15619
## + mining     1  15619
## + pacific    1  15619
## + pubadmin   1  15619
##
## Step:  AIC=15583.84
## Surv(spell, event) ~ ui + bluecoll + logwage
##
##             Df    AIC
## + slack      1  15568
## + married    1  15570
## + stateur    1  15572
## + explose    1  15577
## + year89     1  15578
## + age        1  15578
## + houshead   1  15579
## + encen      1  15582
## + year87     1  15583
```

23

```
## + disrate   1 15584
## <none>         15584
## + wscen     1 15586
## + reprate   1 15587
## + smsa      1 15587
## + child     1 15589
## + southatl  1 15589
## + wncen     1 15589
## + services  1 15590
## + schgt12   1 15590
## + ychild    1 15590
## + fire      1 15590
## + transp    1 15590
## + mountain  1 15590
## + midatl    1 15590
## + year85    1 15591
## + mining    1 15591
## + tenure    1 15592
## + constr    1 15592
## + trade     1 15592
## + female    1 15592
## + escen     1 15592
## + abolpos   1 15592
## + pacific   1 15592
## + schlt12   1 15592
## + pubadmin  1 15592
##
## Step:  AIC=15568.18
## Surv(spell, event) ~ ui + bluecoll + logwage + slack
##
##             Df   AIC
## + married   1 15557
## + stateur   1 15559
## + age       1 15560
## + year89    1 15563
## + houshead  1 15564
## + explose   1 15565
## + encen     1 15567
## <none>        15568
## + year87    1 15569
## + disrate   1 15569
## + wscen     1 15570
## + reprate   1 15571
## + abolpos   1 15571
## + smsa      1 15571
## + services  1 15573
## + child     1 15574
## + southatl  1 15574
## + schgt12   1 15574
## + ychild    1 15574
## + tenure    1 15574
## + wncen     1 15574
## + transp    1 15574
## + mountain  1 15574
```

```
## + fire      1 15575
## + year85    1 15575
## + midatl    1 15575
## + trade     1 15575
## + female    1 15576
## + mining    1 15576
## + pacific   1 15576
## + constr    1 15576
## + pubadmin  1 15576
## + schlt12   1 15576
## + escen     1 15576
##
## Step:  AIC=15557.35
## Surv(spell, event) ~ ui + bluecoll + logwage + slack + married
##
##            Df   AIC
## + age       1 15541
## + stateur   1 15547
## + year89    1 15552
## + houshead  1 15554
## + explose   1 15554
## + encen     1 15556
## <none>        15557
## + year87    1 15559
## + disrate   1 15559
## + smsa      1 15559
## + reprate   1 15559
## + wscen     1 15561
## + abolpos   1 15561
## + services  1 15562
## + tenure    1 15562
## + schgt12   1 15563
## + southatl  1 15563
## + mountain  1 15563
## + transp    1 15563
## + wncen     1 15564
## + fire      1 15564
## + year85    1 15564
## + trade     1 15565
## + midatl    1 15565
## + mining    1 15565
## + female    1 15565
## + ychild    1 15565
## + pacific   1 15565
## + schlt12   1 15565
## + pubadmin  1 15565
## + constr    1 15566
## + escen     1 15566
## + child     1 15566
##
## Step:  AIC=15540.74
## Surv(spell, event) ~ ui + bluecoll + logwage + slack + married +
##      age
##
```

25

```
##              Df   AIC
## + stateur    1 15528
## + houshead   1 15530
## + year89     1 15535
## + explose    1 15538
## + encen      1 15540
## <none>         15541
## + disrate    1 15542
## + smsa       1 15542
## + year87     1 15542
## + reprate    1 15544
## + abolpos    1 15544
## + services   1 15545
## + wscen      1 15545
## + southatl   1 15545
## + schgt12    1 15546
## + mountain   1 15546
## + wncen      1 15547
## + transp     1 15547
## + ychild     1 15547
## + year85     1 15548
## + child      1 15548
## + fire       1 15548
## + mining     1 15548
## + trade      1 15548
## + midatl     1 15548
## + female     1 15549
## + pacific    1 15549
## + pubadmin   1 15549
## + schlt12    1 15549
## + tenure     1 15549
## + escen      1 15549
## + constr     1 15549
##
## Step:  AIC=15528.38
## Surv(spell, event) ~ ui + bluecoll + logwage + slack + married +
##     age + stateur
##
##              Df   AIC
## + houshead   1 15515
## + explose    1 15526
## + wscen      1 15527
## <none>         15528
## + year87     1 15531
## + disrate    1 15531
## + smsa       1 15531
## + encen      1 15532
## + year89     1 15532
## + services   1 15532
## + reprate    1 15533
## + abolpos    1 15533
## + schgt12    1 15534
## + mountain   1 15534
## + ychild     1 15534
```

26

```
## + transp    1 15535
## + southatl  1 15535
## + midatl    1 15535
## + child     1 15535
## + fire      1 15536
## + escen     1 15536
## + wncen     1 15536
## + trade     1 15536
## + pacific   1 15536
## + mining    1 15536
## + female    1 15536
## + schlt12   1 15536
## + pubadmin  1 15536
## + year85    1 15536
## + constr    1 15536
## + tenure    1 15536
##
## Step:  AIC=15515.18
## Surv(spell, event) ~ ui + bluecoll + logwage + slack + married +
##     age + stateur + houshead
##
##            Df   AIC
## + explose   1 15513
## + wscen     1 15514
## <none>        15515
## + year89    1 15517
## + smsa      1 15518
## + encen     1 15518
## + ychild    1 15519
## + year87    1 15519
## + disrate   1 15519
## + services  1 15519
## + reprate   1 15519
## + abolpos   1 15520
## + mountain  1 15520
## + child     1 15520
## + female    1 15521
## + transp    1 15521
## + schgt12   1 15521
## + southatl  1 15522
## + fire      1 15522
## + midatl    1 15523
## + trade     1 15523
## + escen     1 15523
## + pacific   1 15523
## + mining    1 15523
## + wncen     1 15523
## + constr    1 15523
## + pubadmin  1 15523
## + year85    1 15523
## + schlt12   1 15523
## + tenure    1 15523
##
## Step:  AIC=15513.21
```

```
## Surv(spell, event) ~ ui + bluecoll + logwage + slack + married +
##     age + stateur + houshead + explose
##
##             Df   AIC
## + wscen    1 15512
## <none>       15513
## + year89   1 15514
## + smsa     1 15516
## + disrate  1 15516
## + services 1 15517
## + ychild   1 15517
## + encen    1 15517
## + mountain 1 15517
## + year87   1 15517
## + reprate  1 15518
## + abolpos  1 15518
## + child    1 15518
## + schgt12  1 15519
## + female   1 15519
## + transp   1 15519
## + southatl 1 15519
## + fire     1 15520
## + midatl   1 15521
## + escen    1 15521
## + pacific  1 15521
## + constr   1 15521
## + mining   1 15521
## + pubadmin 1 15521
## + trade    1 15521
## + wncen    1 15521
## + year85   1 15521
## + schlt12  1 15521
## + tenure   1 15521
##
## Step:  AIC=15511.68
## Surv(spell, event) ~ ui + bluecoll + logwage + slack + married +
##     age + stateur + houshead + explose + wscen
##
##             Df   AIC
## <none>       15512
## + year89   1 15514
## + disrate  1 15514
## + ychild   1 15515
## + services 1 15515
## + smsa     1 15515
## + southatl 1 15516
## + year87   1 15516
## + child    1 15516
## + abolpos  1 15517
## + reprate  1 15517
## + mountain 1 15517
## + schgt12  1 15517
## + transp   1 15517
## + encen    1 15518
```

```
## + female    1 15518
## + escen     1 15518
## + fire      1 15519
## + mining    1 15519
## + pubadmin  1 15519
## + wncen     1 15519
## + constr    1 15519
## + midatl    1 15520
## + pacific   1 15520
## + trade     1 15520
## + schlt12   1 15520
## + year85    1 15520
## + tenure    1 15520
```

Appendix B: interaction term significance testing for getting and keeping a full-time job

Since this process output is very long, we will be showing the example here. For the other competing risk models analyzed in Appendix D, we won't be showing the outputs as they are very long, but we will show the conclusions.

This has all the interaction term covariates we want to test for:

```
anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                    loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schlt12           -378.56  1.2422  1  0.2650393
## schgt12           -378.46  0.1899  1  0.6629838
## ychild            -378.40  0.1201  1  0.7289583
## child             -377.84  1.1343  1  0.2868655
## married           -377.84  0.0000  0  1.0000000
## female            -377.76  0.1502  1  0.6983423
## smsa              -377.55  0.4272  1  0.5133607
## bluecoll          -377.55  0.0000  0  1.0000000
## nonwhite:schlt12  -377.55  0.0013  1  0.9711732
## nonwhite:schgt12  -373.63  7.8397  1  0.0051110 **
## nonwhite:ychild   -373.59  0.0806  1  0.7765037
## nonwhite:child    -373.32  0.5347  1  0.4646540
## nonwhite:married  -373.31  0.0309  1  0.8604295
## nonwhite:female   -372.59  1.4333  1  0.2312334
## nonwhite:smsa     -372.59  0.0029  1  0.9574141
## nonwhite:bluecoll -372.21  0.7576  1  0.3840843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:schlt12 is most insignificant with a 0.9711732 p-value, we will remove it.

```r
anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                    loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## ychild            -379.12  0.1257  1  0.7229654
## child             -378.53  1.1797  1  0.2774104
## married           -378.53  0.0000  0  1.0000000
## female            -378.47  0.1185  1  0.7306681
## smsa              -378.28  0.3737  1  0.5409737
## bluecoll          -378.28  0.0000  0  1.0000000
## nonwhite:schgt12  -375.23  6.0972  1  0.0135399 *
## nonwhite:ychild   -375.16  0.1537  1  0.6950029
## nonwhite:child    -374.82  0.6820  1  0.4088964
## nonwhite:married  -374.81  0.0068  1  0.9343483
## nonwhite:female   -374.21  1.2096  1  0.2714058
## nonwhite:smsa     -374.19  0.0300  1  0.8624779
## nonwhite:bluecoll -373.92  0.5428  1  0.4612740
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:married is most insignificant with a 0.9343483 p-value, we will remove it.

```r
anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                    loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## ychild            -379.12  0.1257  1  0.7229654
## child             -378.53  1.1797  1  0.2774104
## female            -378.47  0.1185  1  0.7306681
## smsa              -378.28  0.3737  1  0.5409737
## bluecoll          -378.28  0.0000  0  1.0000000
## nonwhite:schgt12  -375.23  6.0972  1  0.0135399 *
## nonwhite:ychild   -375.16  0.1537  1  0.6950029
## nonwhite:child    -374.82  0.6820  1  0.4088964
## nonwhite:female   -374.22  1.1905  1  0.2752323
## nonwhite:smsa     -374.20  0.0333  1  0.8551838
## nonwhite:bluecoll -373.93  0.5372  1  0.4635855
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:smsa is most insignificant with a 0.8551838 p-value, we will remove it.

```r
anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                    loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## ychild            -379.12  0.1257  1  0.7229654
## child             -378.53  1.1797  1  0.2774104
## female            -378.47  0.1185  1  0.7306681
## bluecoll          -378.47  0.0000  0  1.0000000
## nonwhite:schgt12  -375.41  6.1162  1  0.0133946 *
## nonwhite:ychild   -375.33  0.1643  1  0.6851970
## nonwhite:child    -375.01  0.6312  1  0.4269187
## nonwhite:female   -374.42  1.1807  1  0.2772232
## nonwhite:bluecoll -374.16  0.5219  1  0.4700214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:ychild is most insignificant with a 0.6851970 p-value, we will remove it.

```r
anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                    loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## child             -378.87  0.6280  1  0.4280935
## female            -378.79  0.1562  1  0.6927126
## bluecoll          -378.79  0.0000  0  1.0000000
## nonwhite:schgt12  -375.70  6.1692  1  0.0129995 *
## nonwhite:child    -375.32  0.7606  1  0.3831553
## nonwhite:female   -374.73  1.1880  1  0.2757374
## nonwhite:bluecoll -374.43  0.6014  1  0.4380574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:bluecoll is most insignificant with a 0.4380574 p-value, we will remove it.

```r
anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
```

```
##
##                     loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## child             -378.87  0.6280  1  0.4280935
## female            -378.79  0.1562  1  0.6927126
## nonwhite:schgt12  -375.70  6.1692  1  0.0129995 *
## nonwhite:child    -375.32  0.7606  1  0.3831553
## nonwhite:female   -374.73  1.1880  1  0.2757374
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:child is most insignificant with a 0.3831553 p-value, we will remove it.

`anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + `

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                     loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## female            -379.13  0.0950  1  0.7578971
## nonwhite:schgt12  -375.99  6.2824  1  0.0121943 *
## nonwhite:female   -375.66  0.6687  1  0.4135034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:female is most insignificant with a 0.4135034 p-value, we will remove it.

`anova(coxph(Surv(spell, event) ~ strata(ui) + strata(bluecoll) + strata(logwage.cat) + strata(slack) + `

```
## Analysis of Deviance Table
##  Cox model: response is Surv(spell, event)
## Terms added sequentially (first to last)
##
##                     loglik   Chisq Df Pr(>|Chi|)
## NULL              -386.70
## nonwhite          -379.18 15.0361  1  0.0001055 ***
## schgt12           -379.18  0.0007  1  0.9792082
## nonwhite:schgt12  -376.04  6.2775  1  0.0122279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since nonwhite:schgt12 is the only significant term left with a 0.0122279 p-value, we will stop here.

Appendix C: competing risk modeling (forward selection) without interaction terms

CENSOR2: Getting and keeping a Part-Time Job model selection (forward selection)

```
unemp.fit.scope2 <- coxph(Surv(spell, censor2) ~
                          ui + reprate + logwage + tenure + disrate +
                          slack + abolpos + explose + stateur + houshead +
                          married + child + ychild + age +
                          schlt12 + schgt12 + smsa + bluecoll + mining +
                          constr + transp + trade + fire + services +
                          pubadmin + year85 + year87 + year89 + midatl +
                          encen + wncen + southatl + escen + wscen +
                          mountain + pacific + female, data = unemployment)

fit0_2 <- coxph(Surv(spell, censor2) ~ 1, data = unemployment)  # starting model
fitf_bic_2 <- step(fit0_2, scope = formula(unemp.fit.scope2), direction = "forward",
                   k = log(nrow(unemployment)))
```

Final output: AIC=4956.33; $Surv(spell, censor2)\ ui + female + stateur$

CENSOR3: Getting a job and then losing it model selection (forward selection)

```
unemp.fit.scope3 <- coxph(Surv(spell, censor3) ~
                          ui + reprate + logwage + tenure + disrate +
                          slack + abolpos + explose + stateur + houshead +
                          married + child + ychild + age +
                          schlt12 + schgt12 + smsa + bluecoll + mining +
                          constr + transp + trade + fire + services +
                          pubadmin + year85 + year87 + year89 + midatl +
                          encen + wncen + southatl + escen + wscen +
                          mountain + pacific + female, data = unemployment)

fit0_3 <- coxph(Surv(spell, censor3) ~ 1, data = unemployment)  # starting model
fitf_bic_3 <- step(fit0_3, scope = formula(unemp.fit.scope3), direction = "forward",
                   k = log(nrow(unemployment)))
```

Final output: AIC=8303.19; $Surv(spell, censor3)\ ui + tenure + slack + houshead + age + year87 + escen + midatl + abolpos$

EVENT+CENSOR2 (keepjob): Getting a Part-Time or Full-Time Job and Staying there throughout the study, model selection (forward selection)

```
unemp.fit.scope5 <- coxph(Surv(spell, keepjob) ~
                          ui + reprate + logwage + tenure + disrate +
                          slack + abolpos + explose + stateur + houshead +
                          married + child + ychild + age +
                          schlt12 + schgt12 + smsa + bluecoll + mining +
                          constr + transp + trade + fire + services +
                          pubadmin + year85 + year87 + year89 + midatl +
                          encen + wncen + southatl + escen + wscen +
                          mountain + pacific + female, data = unemployment)

fit0_5 <- coxph(Surv(spell, keepjob) ~ 1, data = unemployment)  # starting model
fitf_bic_5 <- step(fit0_5, scope = formula(unemp.fit.scope5), direction = "forward",
                   k = log(nrow(unemployment)))
```

Final output: AIC=20455.75; $Surv(spell, keepjob)\ ui+bluecoll+stateur+slack+married+age+houshead+$
$explose + wscen + disrate$

EVENT+CENSOR2+CENSOR3 (anyjob): Getting Re-Employed at all model selection (forward selection)

```
unemp.fit.scope4 <- coxph(Surv(spell, anyjob) ~
                        ui + reprate + logwage + tenure + disrate +
                        slack + abolpos + explose + stateur + houshead +
                        married + child + ychild + age +
                        schlt12 + schgt12 + smsa + bluecoll + mining +
                        constr + transp + trade + fire + services +
                        pubadmin + year85 + year87 + year89 + midatl +
                        encen + wncen + southatl + escen + wscen +
                        mountain + pacific + female, data = unemployment)

fit0_4 <- coxph(Surv(spell, anyjob) ~ 1, data = unemployment)  # starting model
fitf_bic_4 <- step(fit0_4, scope = formula(unemp.fit.scope4), direction = "forward",
                k = log(nrow(unemployment)))
```

Final output: AIC=28654.19; $Surv(spell, anyjob)\ ui+slack+stateur+age+houshead+bluecoll+abolpos+$
$married + year87 + midatl + encen + child$

Appendix D: Finding Significant Interaction Terms in the Competing Risk Models in Part 6

CENSOR2: Getting and Keeping a Part-Time Job

This has all the interaction term covariates we want to test for:

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(schlt12 + schgt12 + ychild + child + married + female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite:schlt12 is most insignificant with a 0.76237 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(schgt12 + ychild + child + married + female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*married is most insignificant with a 0.41739 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(schgt12 + ychild + child + female + smsa + bluecoll) + nonwhite,
        data = unemployment))
```

Since nonwhite*schgt12 is most insignificant with a 0.33142 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(ychild + child + female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*bluecoll is most insignificant with a 0.38266 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(ychild + child + female + smsa) + nonwhite, data = unemployment))
```

Since nonwhite*female is most insignificant with a 0.30977 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(ychild + child + smsa) + nonwhite, data = unemployment))
```

Since nonwhite*child is most insignificant with a 0.26028 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(ychild + smsa) + nonwhite, data = unemployment))
```

Since nonwhite*ychild is most insignificant with a 0.12926 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor2) ~ strata(ui) + strata(female) + strata(stateur.cat) +
            nonwhite*(smsa) + nonwhite, data = unemployment))
```

Since nonwhite*smsa is the only significant term left with a 0.01897 p-value, we will stop here.

CENSOR3: Getting a job and then losing it

This has all the interaction term covariates we want to test for:

```
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) +
            nonwhite*(schlt12 + schgt12 + ychild + child + married + female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite:female is most insignificant with a 0.885639 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) +
            nonwhite*(schlt12 + schgt12 + ychild + child + married + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*married is most insignificant with a 0.778744 p-value, we will remove it.

```
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) +
            nonwhite*(schlt12 + schgt12 + ychild + child + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*ychild is most insignificant with a 0.765513 p-value, we will remove it.

```r
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) +
            nonwhite*(schlt12 + schgt12 + child + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*smsa is most insignificant with a 0.53245 p-value, we will remove it.

```r
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) +
            nonwhite*(schlt12 + schgt12 + child + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*bluecoll is most insignificant with a 0.532088 p-value, we will remove it.

```r
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) + nonwhite*(schlt12 + schgt12 + child) +
            nonwhite, data = unemployment))
```

Since nonwhite*child is most insignificant with a 0.32835 p-value, we will remove it.

```r
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) + strata(slack) +
            strata(houshead) + strata(age.cat) + strata(year87) + strata(escen) +
            strata(midatl) + strata(abolpos) + nonwhite*(schlt12 + schgt12) +
            nonwhite, data = unemployment))
```

Since nonwhite*schgt12 is most insignificant with a 0.22062 p-value, we will remove it.

```r
anova(coxph(Surv(spell, censor3) ~ strata(ui) + strata(tenure.cat) +
            strata(slack) + strata(houshead) + strata(age.cat) + strata(year87) +
            strata(escen) + strata(midatl) + strata(abolpos) + nonwhite*(schlt12) +
            nonwhite, data = unemployment))
```

Since nonwhite*schlt12 is insignificant with a 0.10488 p-value, we will remove it. We will stop here. There are no significant interaction terms to add to the model.

EVENT+CENSOR2 (keepjob): Getting a Part-Time or Full-Time Job and Staying there throughout the study

This has all the interaction term covariates we want to test for:

```r
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(schlt12 + schgt12 + ychild + child + married + female +
                      smsa + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite:ychild is most insignificant with a 0.962423 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(schlt12 + schgt12 + child + married + female + smsa +
                      bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*child is most insignificant with a 0.689204 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(schlt12 + schgt12 + married + female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*schlt12 is most insignificant with a 0.605692 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(schgt12 + married + female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*female is most insignificant with a 0.608522 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(schgt12 + married + smsa + bluecoll) + nonwhite,
        data = unemployment))
```

Since nonwhite*schgt12 is most insignificant with a 0.489218 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(married + smsa + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*smsa is most insignificant with a 0.418518 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) +
            nonwhite*(married + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*married is most insignificant with a 0.442988 p-value, we will remove it.

```
anova(coxph(Surv(spell, keepjob) ~ strata(ui) + strata(bluecoll) + strata(stateur.cat) +
            strata(slack) + strata(married) + strata(age.cat) + strata(houshead) +
            strata(explose) + strata(wscen) + strata(disrate.cat) + nonwhite*(bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*bluecoll is insignificant with a 0.313117 p-value, we will remove it. We will stop here. There
are no significant interaction terms to add to the model.

EVENT+CENSOR2+CENSOR3 (anyjob): Getting Re-Employed at all

This has all the interaction term covariates we want to test for:

```r
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) +
            strata(child) + nonwhite*(schlt12 + schgt12 + ychild + child + married +
                                      female + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite:schgt12 is most insignificant with a 0.97635 p-value, we will remove it.

```r
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) +
            strata(child) + nonwhite*(schlt12 + ychild + child + married + female +
                                      smsa + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*female is most insignificant with a 0.89015 p-value, we will remove it.

```r
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) +
            strata(child) + nonwhite*(schlt12 + ychild + child + married + smsa
                                      + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*ychild is most insignificant with a 0.79989 p-value, we will remove it.

```r
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) +
            strata(child) + nonwhite*(schlt12 + child + married + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*married is most insignificant with a 0.49618 p-value, we will remove it.

```r
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) +
            strata(child) + nonwhite*(schlt12 + child + smsa + bluecoll) +
            nonwhite, data = unemployment))
```

Since nonwhite*child is most insignificant with a 0.46494 p-value, we will remove it.

```r
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) +
            strata(child) + nonwhite*(schlt12 + smsa + bluecoll) + nonwhite,
          data = unemployment))
```

Since nonwhite*schlt12 is most insignificant with a 0.42408 p-value, we will remove it.

```
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) + strata(child) +
            nonwhite*(smsa + bluecoll) + nonwhite, data = unemployment))
```

Since nonwhite*bluecoll is most insignificant with a 0.15456 p-value, we will remove it.

```
anova(coxph(Surv(spell, anyjob) ~ strata(ui) + strata(slack) + strata(stateur.cat) +
            strata(age.cat) + strata(houshead) + strata(bluecoll) + strata(abolpos) +
            strata(married) + strata(year87) + strata(midatl) + strata(encen) + strata(child) +
            nonwhite*(smsa) + nonwhite, data = unemployment))
```

Since nonwhite*smsa is insignificant with a 0.13174 p-value, we will remove it. We will stop here. There are no significant interaction terms to add to the model.

Bibliography

Colin Cameron, Economics, U.C. Davis Personal Home Page, cameron.econ.ucdavis.edu/mmabook/mma17p4duration.txt.

Katchova, Ani L. (2013) "Survival Analysis - Econometrics Academy." Google Sites, sites.google.com/site/econometricsacadem models/survival-analysis.

Kleinbaum, David G., and Mitchel Klein. Survival Analysis: A Self Learning Text. Third Edition ed., Springer, 2012.

McCall, Brian P. "Unemployment Insurance Rules, Joblessness, and Part-Time Work." Econometrica, vol. 64, no. 3, 1996, pp. 647-682. JSTOR, www.jstor.org/stable/2171865. Accessed 4 Dec. 2020.