Econometrics I - Case Study 3

Group 3, Instructor: Univ.Prof. David Preinerstorfer, Ph.D

Julian Hofmaninger, Tsz Lam Hung and Daniel Diederichs

December 8, 2025

Loading data

```
library(readxl)
data <- read.csv("change.csv")
data$gender <- as.factor(data$gender)
data$occupation <- as.factor(data$occupation)
data$medianwage <- as.factor(data$medianwage)
```

Descriptive Statistics

```
summary(data[, c(1,6,4)])
```

```
##       age          periodsincome       nchange
##  Min.   :22.00   Min.   : 1.00   Min.   :0.000
##  1st Qu.:26.00   1st Qu.: 8.00   1st Qu.:0.000
##  Median :31.00   Median :12.00   Median :1.000
##  Mean   :31.26   Mean   :10.03   Mean   :1.297
##  3rd Qu.:37.00   3rd Qu.:13.00   3rd Qu.:2.000
##  Max.   :42.00   Max.   :13.00   Max.   :9.000
```

To get a rough understanding of the dataset we choose to calculate the summary statistics for the numerical variables. From that we can see that our dataset mainly contains younger to middle-aged people ranging from 22 years to 42 years with the mean being 31.26 years. For `periodsincome` we can observe that there are people with positive incomes in just one year to a maximum of 13 years with the mean being 10.03 years The number of changes of employer between 1986 and 1998 ranges from no change at all to 9 changes with the median being 1.297 changes.

```
get_mode <- function(x) {
  x <- x[!is.na(x)]
  if (length(x) == 0) return(list(mode = NA, frequency = NA))
  tab <- table(x)
  max_count <- max(tab)
  modes <- names(tab)[tab == max_count]
  rel_freq <- max_count / length(x)
  return(list(mode = modes, relative_frequency = rel_freq))
```

```r
}
paste("Mode for Gender: ",
      ifelse(get_mode(data$gender)$mode == 0, "Men", "Women"),
      " they make up for ", round(get_mode(data$gender)$relative_frequency, 4) *
        100, "% ", "of the data", sep="")
```

[1] "Mode for Gender: Men they make up for 57.88% of the data"

```r
paste("Mode for occupation: ",
      ifelse(get_mode(data$occupation)$mode == 0, "Blue Collar","White Collar"),
      " it makes up for ",
      round(get_mode(data$occupation)$relative_frequency, 4) * 100, "% ",
      "of the data", sep="")
```

[1] "Mode for occupation: Blue Collar it makes up for 54.19% of the data"

```r
paste("Mode for wage category: ",
      ifelse(get_mode(data$medianwage)$mode == 1, "Lowest Wage Category",
             ifelse(get_mode(data$medianwage)$mode == 2,"2nd lowest wage category",
                    ifelse(get_mode(data$medianwage)$mode == 3, "Middlemost wage category",
                           ifelse(get_modedata$medianwage)$mode == 1,
                           "2nd highest wage category", "Highest wage category")))),
      " it makes up for ", round(get_mode(data$medianwage)$relative_frequency, 4) *
        100, "% ", "of the data", sep="")
```

[1] "Mode for wage category: Lowest Wage Category it makes up for 21.78% of the data"
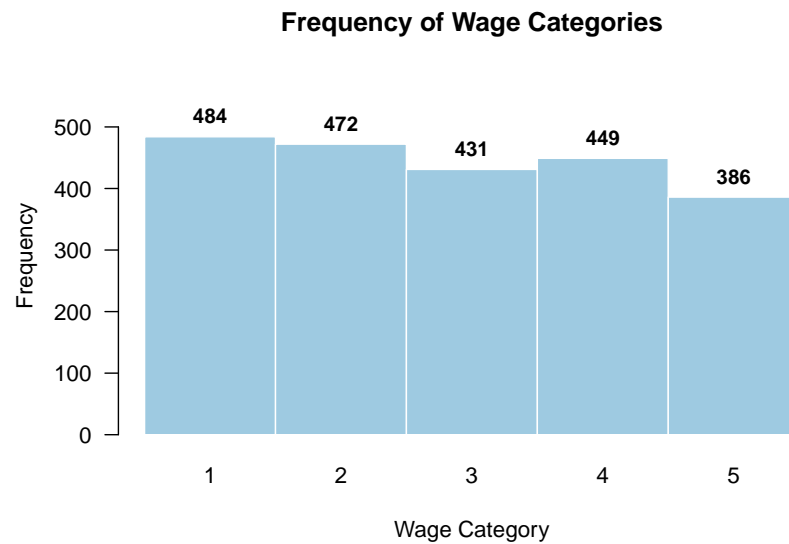
For all the categorical variables we calculated the mode. From that we can see that the majority of the
data is for men, who make up for roughly 58% of the observations. Furthermore, in terms of `occupation`
the majority of the observations is made for Blue Collar Workers with roughly 54%. Last but not least, the
lowest wage category is the most frequently observed category across all observation with a share of roughly
22%.

```r
medianwage_freq <- table(data$medianwage)
breaks <- as.numeric(names(medianwage_freq))
freqs <- as.numeric(medianwage_freq)

bp_medwage <- barplot(freqs,
             names.arg = breaks,
             col = "#9ECAE1",
             border = "white",
             space = 0,
             ylim = c(0, max(freqs) * 1.2),
             main = "Frequency of Wage Categories",
             xlab = "Wage Category", ylab = "Frequency",
             las = 1)
text(x = bp_medwage, y = freqs, labels = freqs, pos = 3, cex = 0.9, font = 2)
```
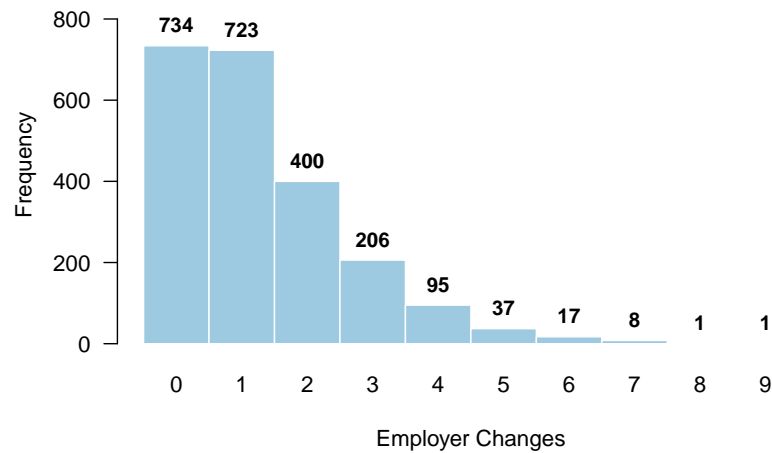
Frequency of Wage Categories

To get a better understanding of the distribution of the wage categories we decided to use a historgram. This shows us that the observations are quite uniformly distributed in terms of `medianwage`.

```r
nchange_freq <- table(data$nchange)
breaks <- as.numeric(names(nchange_freq))
freqs <- as.numeric(nchange_freq)

bp_nchange <- barplot(freqs,
            names.arg = breaks,
            col = "#9ECAE1",
            border = "white",
            space = 0,
            ylim = c(0, max(freqs) * 1.2),
            main = "Frequency of Employer Changes",
            xlab = "Employer Changes", ylab = "Frequency",
            las = 1)
text(x = bp_nchange, y = freqs, labels = freqs, pos = 3, cex = 0.9, font = 2)
```
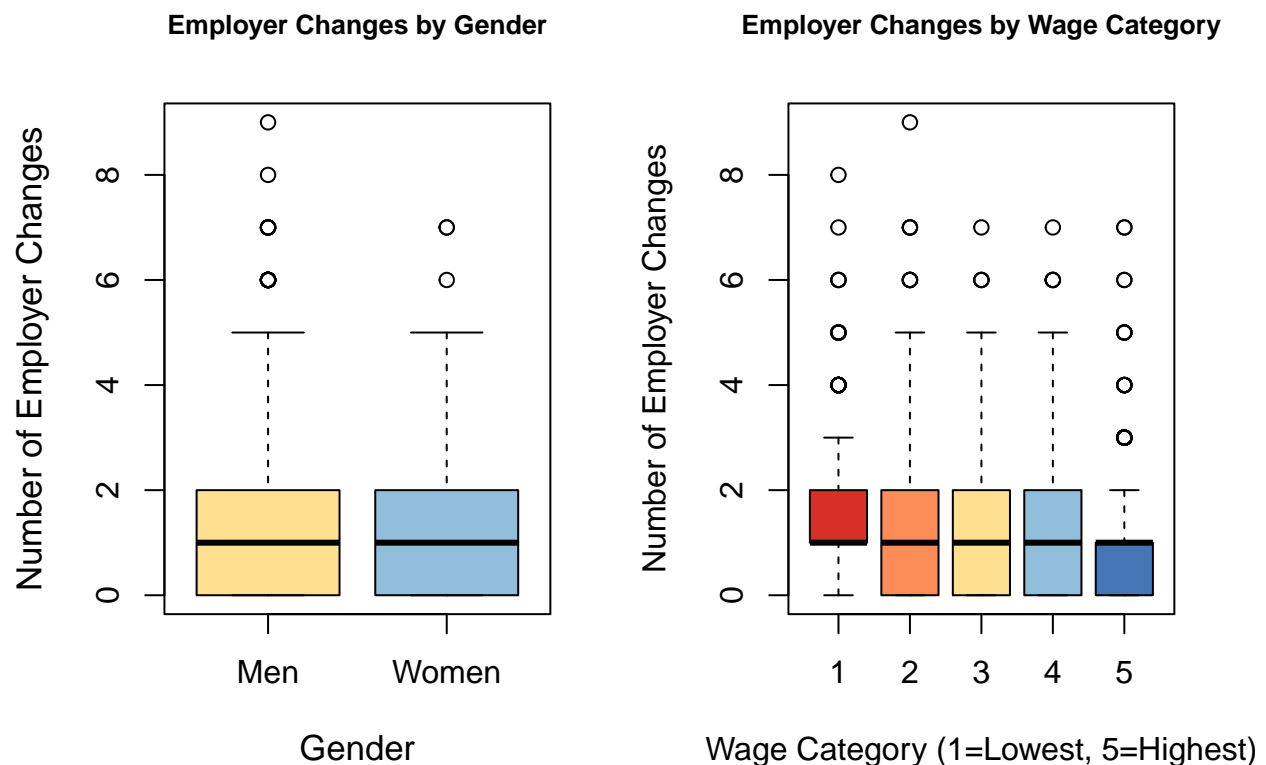
Frequency of Employer Changes

Moreover, we are interested in understanding the distribution of the number of employer changes in the data. Therefore, we used another histogram and from that we can see a strongly right-skewed distribution. The majority of the people observed rarely changed their employer in the relevant period.

```r
par(mfrow = c(1,2))
boxplot(nchange ~ gender, data = data,
        main = "Employer Changes by Gender",
        xlab = "Gender",
        col = c("#FEE090", "#91BFDB"),
        ylab = "Number of Employer Changes",
        names = c("Men", "Women"), cex.main = 0.8,
        cex.lab = 1.1, outline = TRUE)

boxplot(nchange ~ medianwage, data = data,
        main = "Employer Changes by Wage Category",
        xlab = "Wage Category (1=Lowest, 5=Highest)",
        ylab = "Number of Employer Changes", cex.main = 0.8,
        col = c("#D73027", "#FC8D59", "#FEE090", "#91BFDB", "#4575B4"),
        names = c("1", "2", "3",
                  "4", "5"),
        outline = TRUE)
```
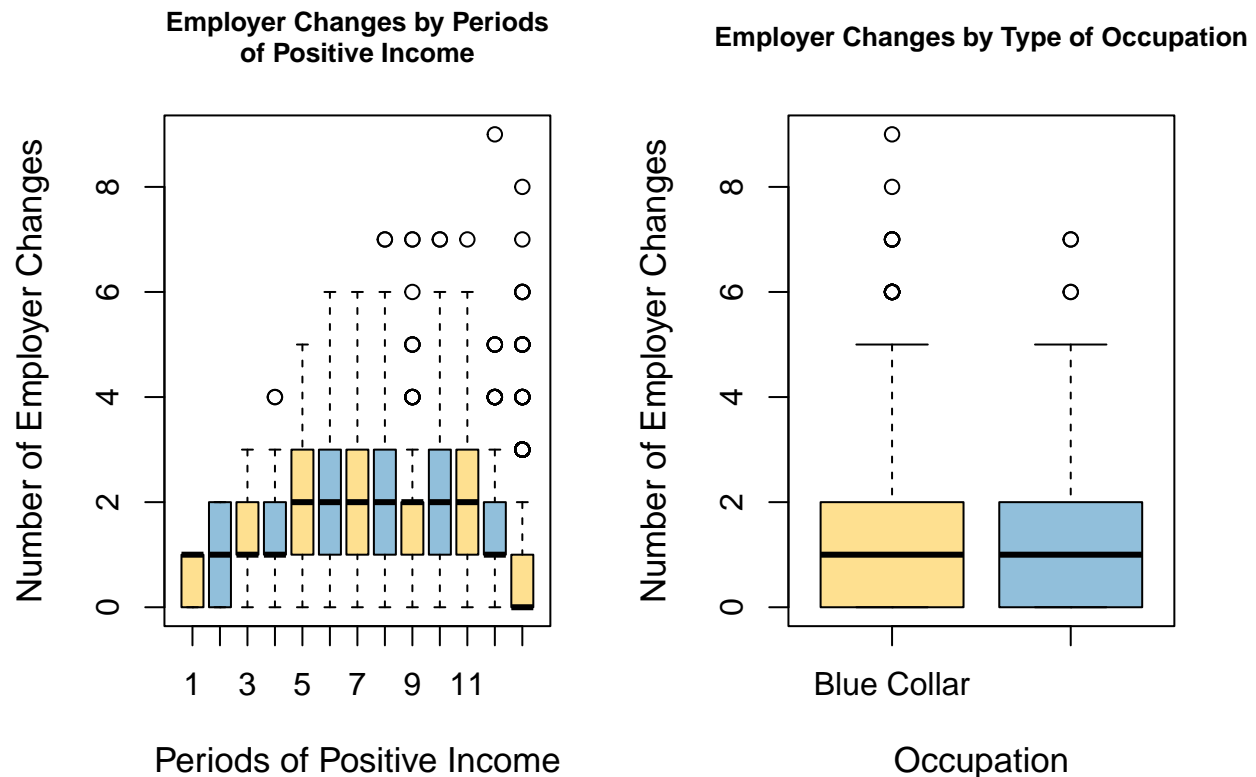
Employer Changes by Gender   **Employer Changes by Wage Category**

Next up is the investigation of potential relationships between observed properties of a person and how often that person changed its employer. Therefore we created a boxplot for number of employer changes depending on gender and another one where it depends on the wage category. From these we can see that interestingly the median is not different depending on the gender nor the wage category. However what is visible is that the more extreme outliers in terms of more employer changes are withing the male sample. Concerning the wage category we can that apart from the median the interquartile range is quite similar across the middle wage categories. However, for the lowest and the highest wage category it is a smaller. Furthermore, the upper whisker of the highest wage category is shorter indicating a slight trend towards less employer changes in that category

```r
par(mfrow = c(1,2))
boxplot(nchange ~ periodsincome, data = data,
        main = "Employer Changes by Periods\nof Positive Income",
        xlab = "Periods of Positive Income",
        col = c("#FEE090", "#91BFDB"),
        ylab = "Number of Employer Changes",
        cex.main = 0.8, cex.lab = 1.1,
        outline = TRUE)

boxplot(nchange ~ occupation, data = data,
        main = "Employer Changes by Type of Occupation",
        xlab = "Occupation",
        col = c("#FEE090", "#91BFDB"),
        ylab = "Number of Employer Changes",
        names = c("Blue Collar", "White Collar"),
        cex.main = 0.8, cex.lab = 1.1, outline = TRUE)
```

Employer Changes by Periods of Positive Income

Employer Changes by Type of Occupation

Additionally we created a boxplot for the number of employer changes depending on the periods of positive income in the relevant period and one for the number of employer changes depending on the type of occupation. For the first boxplot we can see an interesting relationship as for people with rather less periods of positive income we can see that the median is quite low, the range is compressed and there are no outliers. A similar pattern can be observed for individuals with up to four positive periods of income. From then on we see that the median increases slightly and the range of observations regarding the number of employer changes increases quite strongly. It is only for the people with very stable incomes (approx. 12-13 periods) that the median comes down again and the upper whiskers become shorter again as well. However, even for those we can see strong outliers in terms of more employer changes. Regarding the boxplot that visualises employer changes dependent on the type of occupation we can see a similar pattern as for the gender with the same median for blue and white collar people. However, the more extreme values are observed for blue collar workers hinting at a slightly higher job mobility for blue collar workers rather than for white collar employees.

1. Linear Regression Model

```
linear_model1 <- lm(nchange ~ gender + occupation + age + periodsincome +
                    medianwage, data=data)
summary(linear_model1)
```

```
##
## Call:
## lm(formula = nchange ~ gender + occupation + age + periodsincome +
##     medianwage, data = data)
```

```
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -2.2314 -0.9518 -0.3055  0.6277  7.1109
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.833747   0.165442  17.128  < 2e-16 ***
## gender1       -0.241707   0.067955  -3.557 0.000383 ***
## occupation1   -0.210958   0.062728  -3.363 0.000784 ***
## age           -0.027646   0.004587  -6.027 1.95e-09 ***
## periodsincome -0.031467   0.008411  -3.741 0.000188 ***
## medianwage2    0.096469   0.089360   1.080 0.280457
## medianwage3   -0.214844   0.096899  -2.217 0.026712 *
## medianwage4   -0.369697   0.100990  -3.661 0.000257 ***
## medianwage5   -0.362804   0.113603  -3.194 0.001425 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.302 on 2213 degrees of freedom
## Multiple R-squared:  0.0706, Adjusted R-squared:  0.06724
## F-statistic: 21.01 on 8 and 2213 DF,  p-value: < 2.2e-16
```

- **gender**: Since the coefficient for `gender` is negative the expected changes decrease with the gender going from male to female. Therefore, the model suggests, on a statistically significant level, that females change their employer less often than males. To be more precise, the coefficient tells us that women (ceteris paribus) have 0.242 less changes of employer than men have.

- **occupation**: From the negative coefficient for `occupation` we can conclude that the model suggests, on a statistically significant level, that blue collar workers change their job more often that white collar workers do. More precisely, the coefficient tells us that white collar workers (ceteris paribus) do have 0.211 less changes of employer than blue collar workers have.

- **age**: The coefficient for `age` is negative but absolutely speaking very small. Therefore, the model suggests, on a statistically significant level, that older people tend to change their employer less often than younger people. More concretely speaking the coefficient tells us that if a person is one year older (ceteris paribus) there are 0.028 less changes of employer for that person.

- **periodsincome**: The coefficient for `periodsincome` is negative but absolutely speaking very small. Therefore, the model suggests, on a statistically signficant level, that people who had more positive incomes in the period of observation tend to change their employer less often than people with less positive incomes. More concretely speaking the coefficient tells us that if a person has one more year of positive income (ceteris paribus) there are 0.031 less changes of employer for that person.

- **medianwage**: The lowest income category (1) is the baseline of the model. Therefore we can see from the coefficient `medianwage2` that the effect on the changes of employer for a person by going from the lowest income category to the second lowest is positive. That implies that when a person moves to the second lowest category of income (ceteris paribus) the same person is expected to change the employer more often. This effect however, is not statistically significant on any reasonable significance level. Additionally, moving to a higher category of income from the second lowest category (ceteris paribus) the person is always expected to have less changes of employer. This effect of less employer changes by moving to a higher income category from the second lowest category on is furthermore statistically signficant on a level of at least 5%.

Linear Hypotheses

$$H_0 : \beta_8 = \beta_7$$

$$H_1 : \beta_8 \neq \beta_7$$

```
library(car)
linearHypothesis(linear_model1, c("medianwage4=medianwage5"))
```

```
##
## Linear hypothesis test:
## medianwage4 - medianwage5 = 0
##
## Model 1: restricted model
## Model 2: nchange ~ gender + occupation + age + periodsincome + medianwage
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   2214 3750.7
## 2   2213 3750.7  1 0.0092695 0.0055 0.9411
```

From the test we get that $p\text{-}value = 0.9411$ which is greater than the significance level of $\alpha = 0.05$. Therefore, we retain the null-hypothesis, which suggests that the effect of income is the same for the two highest wage categories.

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

```
linearHypothesis(linear_model1, c("medianwage2=0"))
```

```
##
## Linear hypothesis test:
## medianwage2 = 0
##
## Model 1: restricted model
## Model 2: nchange ~ gender + occupation + age + periodsincome + medianwage
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   2214 3752.6
## 2   2213 3750.7  1    1.9752 1.1654 0.2805
```

From the test we get that $p\text{-}value = 0.2805$ which is greater than the significance level of $\alpha = 0.05$. Therefore, we retain the null-hypothesis which suggests that the effect of income is the same for the two lowest wage categories.

2. Linear Regression Model

```
linear_model2 <- lm(nchange ~ gender + occupation + age + periodsincome +
                    I(periodsincome^2) + medianwage, data=data)
summary(linear_model2)
```

```
##
## Call:
## lm(formula = nchange ~ gender + occupation + age + periodsincome +
##     I(periodsincome^2) + medianwage, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4566 -0.8680 -0.1929  0.6030  7.1634
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.718844   0.197325   3.643 0.000276 ***
## gender1            -0.287296   0.063818  -4.502 7.09e-06 ***
## occupation1        -0.203588   0.058861  -3.459 0.000553 ***
## age                -0.019319   0.004331  -4.461 8.56e-06 ***
## periodsincome       0.601521   0.037302  16.126  < 2e-16 ***
## I(periodsincome^2) -0.039405   0.002270 -17.362  < 2e-16 ***
## medianwage2         0.037506   0.083918   0.447 0.654964
## medianwage3        -0.231506   0.090929  -2.546 0.010963 *
## medianwage4        -0.277419   0.094911  -2.923 0.003503 **
## medianwage5        -0.277019   0.106712  -2.596 0.009495 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.222 on 2212 degrees of freedom
## Multiple R-squared:  0.1821, Adjusted R-squared:  0.1787
## F-statistic: 54.71 on 9 and 2212 DF,  p-value: < 2.2e-16
```

The *p-value* $< 2e\text{-}16$ for the coefficient of `I(periodsincome^2)` is smaller than any relevant significance level. Therefore, we conclude that the quadratic effect of `periodsincome` is significant.

```r
x_vertex <- -linear_model2$coefficients["periodsincome"]/
  (2*linear_model2$coefficients["I(periodsincome^2)"])
min_periodsincome <- min(data$periodsincome)
max_periodsincome <- max(data$periodsincome)


if(x_vertex >= min_periodsincome & x_vertex <= max_periodsincome){
  paste("The vertex (", round(x_vertex,3) ,") lies within the range (",
  round(min_periodsincome, 3), ";", round(max_periodsincome,3),
  ") of periodsincome: Non monotonic effect", sep="")
}else{
  print("The vertex (", round(x_vertex,3) ,") does not lie within the range (",
        round(min_periodsincome, 3), ";", round(max_periodsincome,3), ") of
        periodsincome: Monotone effect", sep="")
}
```

```
## [1] "The vertex (7.633) lies within the range (1;13) of periodsincome: Non monotonic effect"
```

To understand if the impact of `periods_income` is monotone we need to calculate the vertex of the parabola. After doing so we get that it is 7.633. Comparing this with the range of `periods_income` in our sample which is 1 to 13 we can conclude that the effect of `periods_income` is not monotone as it changes direction within the range of the values.

Effect of two additional years of `periodsincome`

The effect of two additional years of `periodsincome` (ceteris paribus) is given by

$$
\Delta\text{nchange} =
$$
$$
y(periodsincome + 2) - y(periodsincome) =
$$
$$
\beta_4 \cdot (periodsincome + 2) + \beta_5 \cdot (periodsincome + 2)^2 - [\beta_4 \cdot periodsincome + \beta_5 \cdot periodsincome^2]
$$

which further simplifies to:

$$
\Delta\text{nchange} = 2\beta_4 + 4\beta_5 \cdot periodsincome + 4\beta_5
$$

Alternatively, we could look at the marginal effects which can be used as an approximation. The marginal effect results from the derivation of the model equation with respect to `periodsincome`:

$$
\frac{\delta\text{y}}{\delta\text{periodsincome}} = \beta_4 + 2\beta_5 \cdot periodsincome
$$

From the exact version and the approximation we can see that the effect of `periodsincome` depends on the value of `periodsincome`. Therefore, to calculate the effect we need to set a concrete value. We decided to use the sample mean to calculate the effect.

```r
periodsincome_mean <- mean(data$periodsincome)

exact_effect <- 2*linear_model2$coefficients["periodsincome"] +
  4*linear_model2$coefficients["I(periodsincome^2)"]*periodsincome_mean +
  4*linear_model2$coefficients["I(periodsincome^2)"]

approx_effect <- (linear_model2$coefficients["periodsincome"] +
                  2*linear_model2$coefficients["I(periodsincome^2)"] *
                  periodsincome_mean)*2

paste("Exact effect of two additional years", round(exact_effect, 3))
```

```
## [1] "Exact effect of two additional years -0.536"
```

```r
paste("Approximated effect of two additional years", round(approx_effect, 3))
```

```
## [1] "Approximated effect of two additional years -0.378"
```

3. Linear Regression Model

```r
linear_model3 <- lm(nchange ~ gender + occupation + occupation:gender + age +
                    periodsincome + I(periodsincome^2) + medianwage, data=data)
summary(linear_model3)
```

```
##
## Call:
## lm(formula = nchange ~ gender + occupation + occupation:gender +
```

```
##     age + periodsincome + I(periodsincome^2) + medianwage, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4925 -0.8392 -0.1995  0.5760  7.1278
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.781988   0.199259   3.924 8.96e-05 ***
## gender1            -0.411302   0.085326  -4.820 1.53e-06 ***
## occupation1        -0.316108   0.078138  -4.046 5.40e-05 ***
## age                -0.019321   0.004327  -4.465 8.40e-06 ***
## periodsincome       0.601590   0.037271  16.141  < 2e-16 ***
## I(periodsincome^2) -0.039412   0.002268 -17.380  < 2e-16 ***
## medianwage2         0.010225   0.084769   0.121  0.90401
## medianwage3        -0.270297   0.092566  -2.920  0.00354 **
## medianwage4        -0.312367   0.096167  -3.248  0.00118 **
## medianwage5        -0.268944   0.106684  -2.521  0.01177 *
## gender1:occupation1 0.246230   0.112582   2.187  0.02884 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.221 on 2211 degrees of freedom
## Multiple R-squared:  0.1838, Adjusted R-squared:  0.1801
## F-statistic:  49.8 on 10 and 2211 DF,  p-value: < 2.2e-16
```

The $p\text{-}value = 0.029$ for the coefficient of `gender * occupation` is smaller than the significance level of $\alpha = 0.05$. Therefore, we conclude that the interaction effect of `gender` and `occupation` is significant.

```r
beta_gender <- function(occupation){
  return(linear_model3$coefficients["gender1"] +
           linear_model3$coefficients["gender1:occupation1"] * occupation)
}


beta_occupation <- function(gender){
  return(linear_model3$coefficients["occupation1"] +
           linear_model3$coefficients["gender1:occupation1"] * gender)
}


cat("Effect of being woman (vs man):\n",
    "  Blue collar:", round(beta_gender(0), 3), "\n",
    "  White collar:", round(beta_gender(1), 3), "\n\n")
```

```
## Effect of being woman (vs man):
##    Blue collar: -0.411
##    White collar: -0.165
```

```r
cat("Effect of being white collar (vs blue collar):\n",
    "  Men:", round(beta_occupation(0), 3), "\n",
    "  Women:", round(beta_occupation(1), 3), "\n")
```

```
## Effect of being white collar (vs blue collar):
##    Men: -0.316
##    Women: -0.07
```

From the calculations above we can see the marginal effects of `gender` and `occupation` on `nchange`. Since there is an interaction effect considered in our model. The marginal effect is now somehow dependent on the other covariate. However, by looking at the marginal effect of being a woman vs being a man we can see that the direction stays the same for blue collar and white collar people. For both types of `occupation` women are predicted to make less changes of employer. Even though the direction stays the same (women making less changes), we can observe that for white collar workers the difference between women and men is signficantly smaller. Concerning the effect of being a white collar employee compared to a blue collar worker we can also see that taking the gender into account does not change the direction. White collar people are expected to change ther employer less often then blue collar people do regardless of being a male or a female.

Point Prediction

For the given data, the model predicts the following number of employer changes:

```r
new_data <- data.frame(age=35,
                       gender=factor(1,levels=c(0,1)),
                       occupation=factor(0,levels=c(0,1)),
                       periodsincome=11,
                       medianwage=factor(2, levels = 1:5))
predict(linear_model3, newdata=new_data)
```

```
##        1
## 1.553282
```

Model Comparison

```r
AICs <- c(AIC(linear_model1),AIC(linear_model2),AIC(linear_model3))
BICs <- c(BIC(linear_model1),BIC(linear_model2),BIC(linear_model3))
AIC_BIC_table <- cbind(AICs, BICs)
rownames(AIC_BIC_table) <- c("Model 1", "Model 2", "Model 3")
AIC_BIC_table
```

```
##             AICs     BICs
## Model 1 7489.036 7546.097
## Model 2 7207.162 7269.930
## Model 3 7204.360 7272.834
```

```r
best_AIC_model <- rownames(AIC_BIC_table)[which.min(AIC_BIC_table[,1])]
best_AIC_value <- round(min(AIC_BIC_table[,1]), 2)

best_BIC_model <- rownames(AIC_BIC_table)[which.min(AIC_BIC_table[,2])]
best_BIC_value <- round(min(AIC_BIC_table[,2]), 2)

cat("According to AIC ", best_AIC_model, " should be selected with a value of",
    best_AIC_value, "\n", "According to BIC ", best_BIC_model,
    " should be selected with a value of", best_BIC_value, sep="")
```

```
## According to AIC Model 3 should be selected with a value of7204.36
## According to BIC Model 2 should be selected with a value of7269.93
```

Since there are two different models that perform best on `AIC` and `BIC` the decision which model to choose is not quite clear. Therefore, we decided to take a look at a third parameter that can give information about the quality of the model.

```
cat("Adjusted R-squared Model 2: ", summary(linear_model2)$adj.r.squared, "\n",
    "Adjusted R-squared Model 3: ", summary(linear_model3)$adj.r.squared, sep="")
```

```
## Adjusted R-squared Model 2: 0.1787324
## Adjusted R-squared Model 3: 0.1801348
```

We choose to look at the adjusted R^2. From that we can see that `model3` is marginally better. Hence, we will continue our analysis for `model3`. However, this decision is rather vague and choosing another approach might lead to different outcomes.
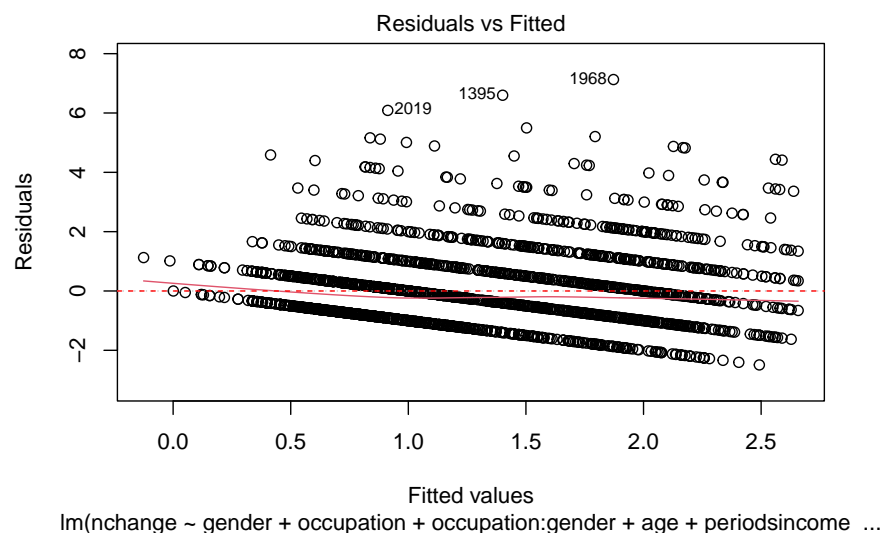
Residual Diagnosis

Standard Model Assumptions

$$\mathbb{E}(u|X_1, \ldots, X_K) = \mathbb{E}(u) = 0 \tag{1}$$

$$\mathbb{E}(Y|X_1, \ldots, X_K) = \beta_0 + \beta_1 \cdot X_1 + \ldots + \beta_K \cdot X \tag{2}$$

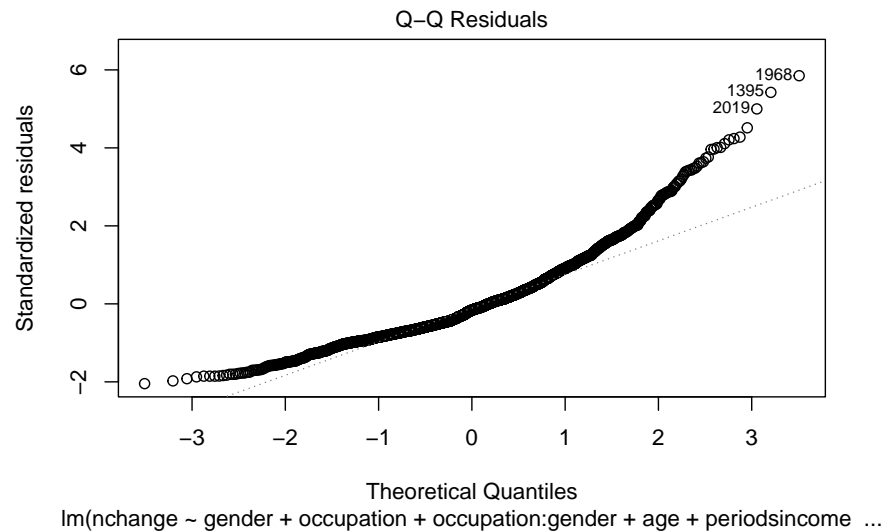$$\mathbb{V}(u|X_1, \ldots, X_K) = \sigma^2 = \mathbb{V}(u) = \sigma^2 \tag{3}$$

```
resids <- residuals(linear_model3)
plot(linear_model3, which=1)
abline(h=0, lty=2, col="red")
```


From the Residuals vs Fitted plot, the LOESS curve shows a slight decreasing, non-linear trend rather than staying flat at zero. This suggests that the model does not fully capture the functional (possibly non-linear) dependence of the response on the regressors. The LOESS curve is slightly above the zero line for smaller fitted values and slightly below for larger fitted values, so the Zero Conditional Mean Error assumption appears reasonably, but not perfectly, satisfied. The homoskedasticity assumption is more clearly violated, as the vertical spread of residuals around the zero line increases with the fitted values, indicating heteroskedasticity.

Normaility of errors

```r
plot(linear_model3, which=2)
```

Q–Q Residuals

lm(nchange ~ gender + occupation + occupation:gender + age + periodsincome ...

```r
tseries::jarque.bera.test(resids)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo


##
##   Jarque Bera Test
##
## data:  resids
## X-squared = 1291.9, df = 2, p-value < 2.2e-16
```

From the Q-Q-Plot we can already see that the residuals look skewed and we would suggest that there is no normal distribution of the residuals given. The `Jarque-Bera-Test` confirms this suggestion as the $p\text{-value} < 2.2e\text{-}16$ leads us to rejecting the null hypothesis that the errors follow a normal distribution.

Conclusions

Concluding from the diagnostic plots, the model likely does not fully capture non-linear relationships in the data. Because the Zero Conditional Mean Error assumption appears to be at least reasonably satisfied, the OLS point estimates of the coefficients should remain approximately unbiased and thus useful for describing the direction and rough magnitude of relationships. However, the residuals show clear heteroskedasticity and deviations from normality, so the usual OLS standard errors and the resulting t- and F-tests are no longer reliable; their p-values may be distorted. Therefore, statements about parameter significance should be made with caution.