ECO 307: Celebration of Econometrics Knowledge Rehearsal

Fall 2015

Workers' compensation provides income to workers who are unable to continue working because of an injury sustained while on the job. The compensation is received for a period of time until the worker recovers from his or her injuries. The amount of compensation an individual worker receives is directly related to the worker's income. There is a limit, though, on the maximum weekly earnings that can be dispersed in the program. On July 15, 1980, the state of Kentucky raised this limit. This had no effect on low-wage earners, but it did affect workers with higher earnings.

Employers and researchers are interested in determining whether extending the limit on compensation causes workers to stay on workers' compensation for a longer duration, intentionally delaying their recovery. The change in policy only affects higher-income workers, because neither the old limit nor new limit was binding for lower-income workers.

They collect the following variables on over 7,000 workers who had taken workers compensation just before the new policy went into effect and just afterwards:

- `durat`: Duration in days collecting wokers' compensation benefits
- `afchnge`: Dummy variable equal to 1 for workers who became injured *after* the compensation limit was extended.
- `highearn`: Dummy variable equal to 1 for workers who have earnings high enough to be affected by the extended compensation limit.
- `male`: Dummy variable equal to 1 for people who are male.
- `married`: Dummy variable equal to 1 for people who are married
- `age`: Age of worker, in years
- `manuf`: Dummy variable equal to 1 for workers in the manufacturing industry
- `construc`: Dummy variable equal to 1 for workers in the construction industry

Consider the calls in R and output on the following pages to answer the questions on the last page.

```r
lm1 <- lm(log(durat) ~ afchnge + highearn + afchnge:highearn + male + married
          + age + I(age^2) + manuf + construc + male:manuf + male:construc, data=data)
summary(lm1)
```

```
##
## Call:
## lm(formula = log(durat) ~ afchnge + highearn + afchnge:highearn +
##     male + married + age + I(age^2) + manuf + construc + male:manuf +
##     male:construc, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3341 -0.7694  0.0548  0.7706  4.2824
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.4788940  0.1442806   3.319 0.000908 ***
## afchnge           0.0189871  0.0399107   0.476 0.634275
## highearn          0.0895407  0.0476875   1.878 0.060471 .
## male              0.0158846  0.0518644   0.306 0.759408
## married           0.0315616  0.0372522   0.847 0.396889
## age               0.0323542  0.0077224   4.190 2.83e-05 ***
## I(age^2)         -0.0002848  0.0000954  -2.986 0.002841 **
## manuf             0.0266023  0.0687226   0.387 0.698696
## construc          0.2244040  0.2549658   0.880 0.378818
## afchnge:highearn  0.2247161  0.0636398   3.531 0.000417 ***
## male:manuf       -0.1927440  0.0806691  -2.389 0.016907 *
## male:construc    -0.0329263  0.2592697  -0.127 0.898947
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.281 on 6810 degrees of freedom
## Multiple R-squared:  0.03402,    Adjusted R-squared:  0.03246
## F-statistic:  21.8 on 11 and 6810 DF,  p-value: < 2.2e-16
```

```r
lm2 <- lm( I(lm1$residuals^2) ~ lm1$fitted.values + I(lm1$fitted.values^2))
summary(lm2)
```

```
##
## Call:
## lm(formula = I(lm1$residuals^2) ~ lm1$fitted.values + I(lm1$fitted.values^2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7663 -1.5215 -1.0572  0.1911 16.8490
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.8977     0.7917   1.134    0.257
## lm1$fitted.values          0.8203     1.1822   0.694    0.488
## I(lm1$fitted.values^2)   -0.1924     0.4331  -0.444    0.657
##
## Residual standard error: 2.602 on 6819 degrees of freedom
## Multiple R-squared:  0.0007871,  Adjusted R-squared:  0.0004941
## F-statistic: 2.686 on 2 and 6819 DF,  p-value: 0.06824
```

```r
vv <- vcovHC(lm1, type = "HC1")
coeftest(lm1, vcov=vv)
```

```
##
## t test of coefficients:
##
##                      Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)        0.47889395  0.14216604  3.3686 0.0007598 ***
## afchnge            0.01898710  0.03957691  0.4798 0.6314192
## highearn           0.08954070  0.04714988  1.8991 0.0575981 .
## male               0.01588455  0.05078715  0.3128 0.7544672
## married            0.03156165  0.03654967  0.8635 0.3878778
## age                0.03235418  0.00764403  4.2326  2.34e-05 ***
## I(age^2)          -0.00028481  0.00009472 -3.0069 0.0026490 **
## manuf              0.02660234  0.06914214  0.3847 0.7004358
## construc           0.22440398  0.28270990  0.7938 0.4273625
## afchnge:highearn  0.22471608  0.06401542  3.5103 0.0004504 ***
## male:manuf        -0.19274400  0.08077152 -2.3863 0.0170466 *
## male:construc     -0.03292629  0.28642140 -0.1150 0.9084822
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lm3 <- lm(log(durat) ~ afchnge + highearn + afchnge:highearn + married
          + age + I(age^2) + manuf + construc, data=data)
summary(lm3)
```

```
##
## Call:
## lm(formula = log(durat) ~ afchnge + highearn + afchnge:highearn +
##     married + age + I(age^2) + manuf + construc, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3342 -0.7678  0.0571  0.7700  4.2310
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.587e-01  1.346e-01   3.409 0.000656 ***
## afchnge           2.165e-02  3.991e-02   0.543 0.587420
## highearn          7.886e-02  4.602e-02   1.713 0.086668 .
## married           3.060e-02  3.720e-02   0.823 0.410801
## age               3.424e-02  7.649e-03   4.476 7.72e-06 ***
## I(age^2)         -3.060e-04  9.482e-05  -3.227 0.001257 **
## manuf            -1.126e-01  3.591e-02  -3.137 0.001713 **
## construc          1.975e-01  4.560e-02   4.330 1.51e-05 ***
## afchnge:highearn  2.220e-01  6.365e-02   3.487 0.000491 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.281 on 6813 degrees of freedom
## Multiple R-squared:  0.03294,    Adjusted R-squared:  0.0318
## F-statistic: 29.01 on 8 and 6813 DF,  p-value: < 2.2e-16
```

```r
anova(lm3,lm1)
```

```
## Analysis of Variance Table
##
## Model 1: log(durat) ~ afchnge + highearn + afchnge:highearn + married +
##     age + I(age^2) + manuf + construc
## Model 2: log(durat) ~ afchnge + highearn + afchnge:highearn + male + married +
##     age + I(age^2) + manuf + construc + male:manuf + male:construc
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1   6813 11186
## 2   6810 11173  3     12.46 2.5314 0.05528 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
waldtest(lm3,lm1,vcov = vv)
```

```
## Wald test
##
## Model 1: log(durat) ~ afchnge + highearn + afchnge:highearn + married +
##     age + I(age^2) + manuf + construc
## Model 2: log(durat) ~ afchnge + highearn + afchnge:highearn + male + married +
##     age + I(age^2) + manuf + construc + male:manuf + male:construc
##   Res.Df Df      F  Pr(>F)
## 1   6813
## 2   6810  3 2.4601 0.06081 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lm4 <- lm(log(durat) ~ afchnge + highearn + afchnge:highearn + male + married
          + age + I(age^2), data=data)
anova(lm4,lm1)
```

```
## Analysis of Variance Table
##
## Model 1: log(durat) ~ afchnge + highearn + afchnge:highearn + male + married +
##     age + I(age^2)
## Model 2: log(durat) ~ afchnge + highearn + afchnge:highearn + male + married +
##     age + I(age^2) + manuf + construc + male:manuf + male:construc
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1   6814 11247
## 2   6810 11173  4    73.601 11.215 4.547e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
waldtest(lm4,lm1,vcov = vv)
```

```
## Wald test
##
## Model 1: log(durat) ~ afchnge + highearn + afchnge:highearn + male + married +
##     age + I(age^2)
## Model 2: log(durat) ~ afchnge + highearn + afchnge:highearn + male + married +
##     age + I(age^2) + manuf + construc + male:manuf + male:construc
##   Res.Df Df      F    Pr(>F)
## 1   6814
## 2   6810  4 11.773 1.573e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Whenever performing hypothesis tests to answer the questions below, please use a 10% significance level.

1. Is there evidence for heteroskedasticity in the regression model that includes all explanatory variables listed above? Test the appropriate hypothesis.
2. Taking into account all the explanatory variables described above, is there statistical evidence that extending the compensation limit led to an increase in the average duration that injured workers collect workers compensation? Test the appropriate hypothesis.
3. Interpret the estimated magnitude for the impact that extending the compensation limit has on workers' compensation duration? Is the effect positive or negative?
4. Interpret the meaning and magnitude for the coefficient on `highearn`.
5. Discuss at least two modeling strategies taken to reduce the potential for omitted variable bias.
6. Accounting for all the other explanatory variables in the model, is there evidence that men have a different average duration collecting workers compensation than women? Test the appropriate hypothesis.
7. Accounting for all the other explanatory variables in the model, is there evidence that men in the manufacturing industry have a different average duration collecting workers compensation than women in manufacturing? Test the appropriate hypothesis.
8. Accounting for all the other explanatory variables in the model, is there evidence that men in the manufacturing industry have a different average duration collecting workers compensation than women in manufacturing? Test the appropriate hypothesis.
9. Comment on how well all the explanatory variables explain duration on workers' compensation. Conduct a hypothesis test for the relevance of the group of explanatory variables used to explain workers compensation. What do you conclude from this analysis? Is the regression model appropriate?
10. Comment on how age affects the length of time that injured workers collect workers' compensation. Does duration increase with age or decrease with age? Does it depend on age? If so, how so?
11. Is the average duration on workers' compensation different depending on industry? Test the appropriate hypothesis.