wage_salary_quantile_regression

Todd Nobles

2024-12-04

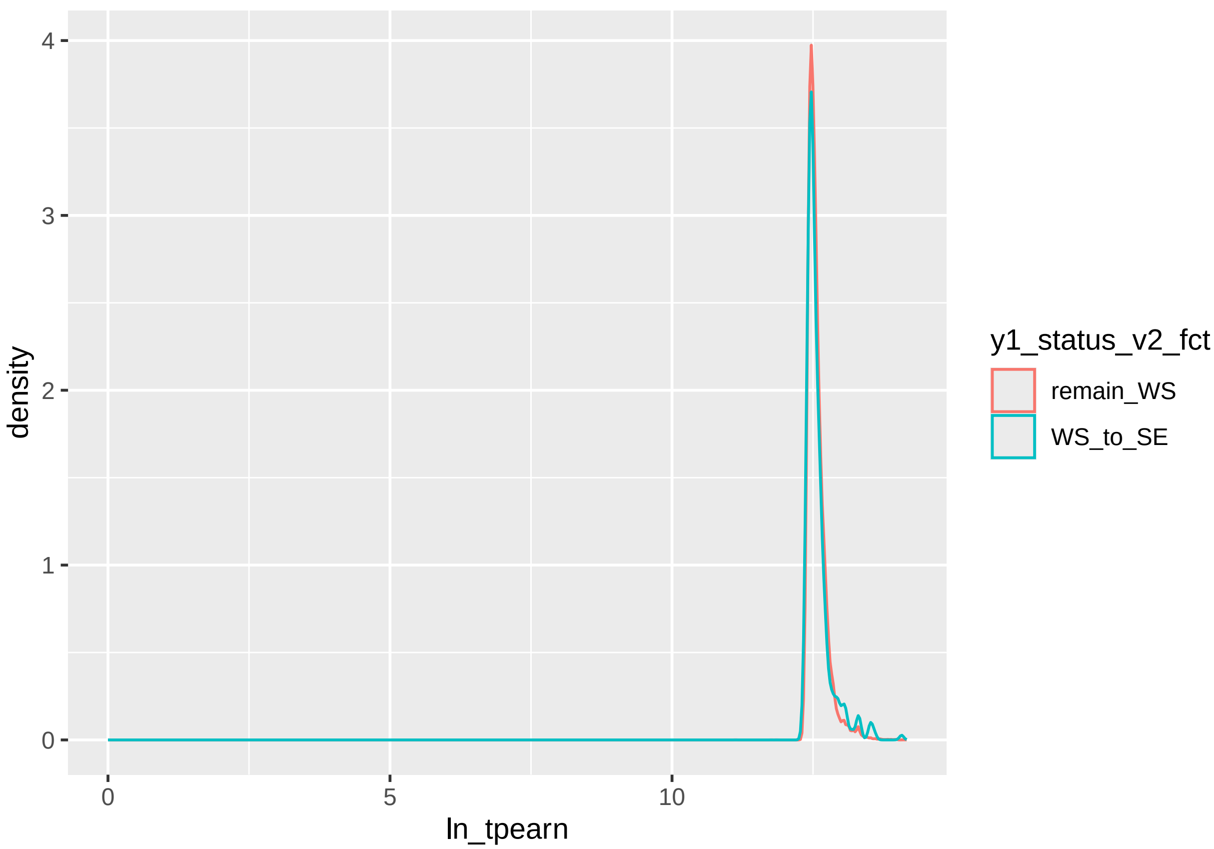
Reading in the data from Stata. At this point we've collapsed down to annual earnings, filtered out the first year for everyone and then

```
df <- read_dta("/Users/toddnobles/Documents/sipp_analyses/earnings_models.dta")
```

```
earnings <- df %>%
  mutate(across(c(sex, immigrant, educ_collapsed, race_collapsed,
                 y1_status_v2),
                as_factor,
                .names = "{col}_fct"),
         ssuid_spanel_pnum_id_fct = factor(ssuid_spanel_pnum_id),
         y1_status_v2_fct = fct_drop(y1_status_v2_fct),
         y1_status_v2_fct = fct_recode(y1_status_v2_fct, "remain_WS" = "W&S", "WS_to_SE" = "SE"))
```

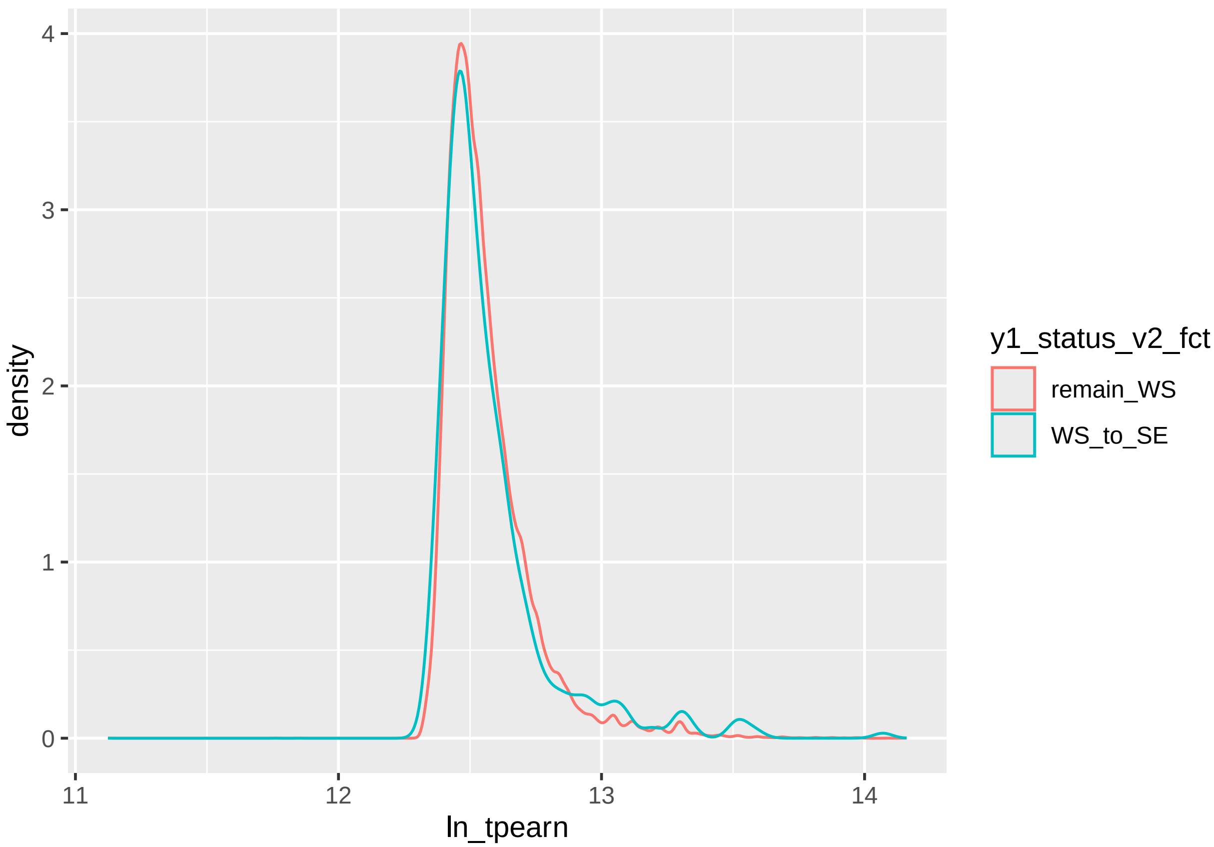
Below is a plot of the ln_tpearn distributions for our two groups.

```
earnings %>% ggplot() + geom_density(aes(x = ln_tpearn, color = y1_status_v2_fct))
```



Same plot but restricting the distribution a bit.

```
earnings %>%
  filter(ln_tpearn > 1) %>%
  ggplot() + geom_density(aes(x = ln_tpearn, color = y1_status_v2_fct))
```



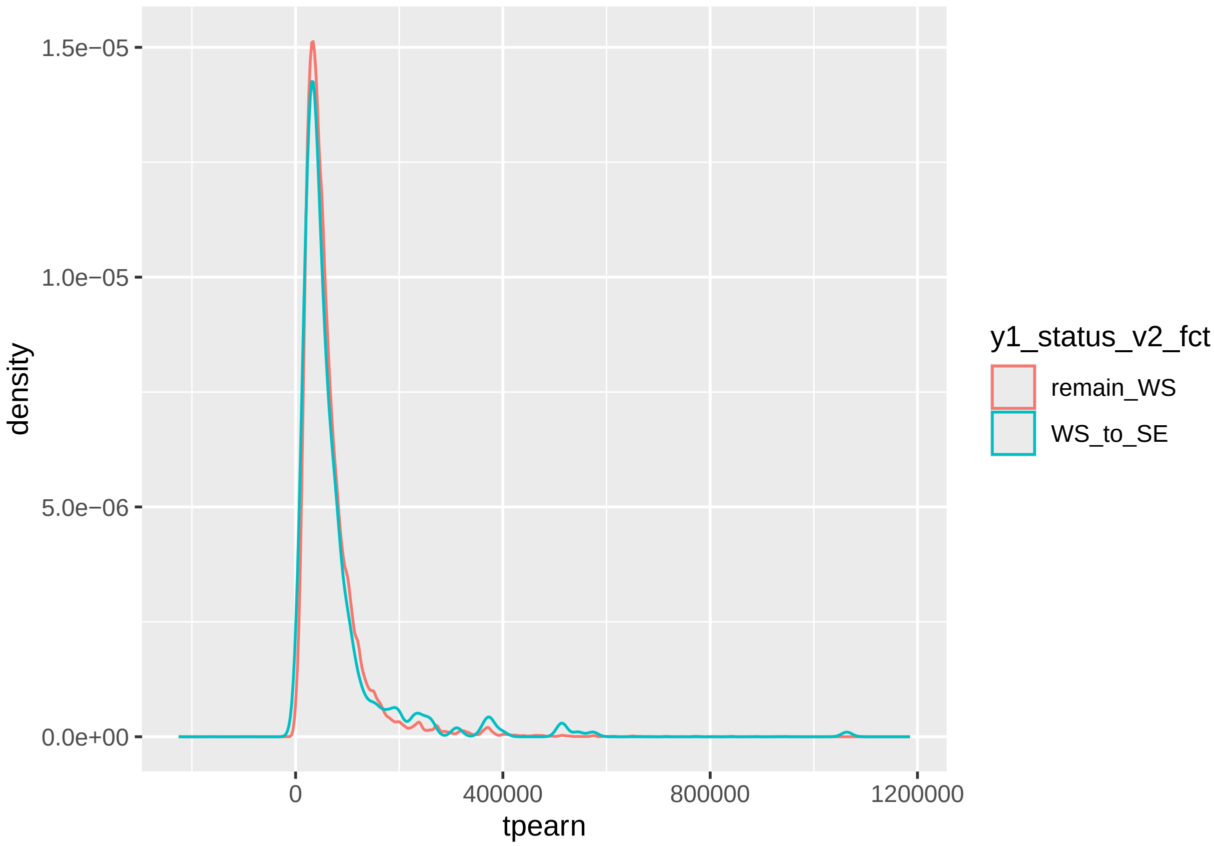
Descriptive table of means/medians

```
earnings %>%
  group_by(y1_status_v2_fct) %>%
  summarize(person_years = n(), people = n_distinct(ssuid_spanel_pnum_id), mean = mean(ln_tpearn), sd =
```

```
## # A tibble: 2 x 8
##   y1_status_v2_fct person_years people  mean    sd   p25   p50   p75
##   <fct>                   <int>  <int> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 remain_WS               50323  26592  12.6 0.178  12.5  12.5  12.6
## 2 WS_to_SE                  388    261  12.6 0.226  12.5  12.5  12.6
```

What about in our raw values?

```
earnings %>% ggplot() + geom_density(aes(x = tpearn, color = y1_status_v2_fct))
```




```
earnings %>%
  group_by(y1_status_v2_fct) %>%
  summarize(person_years = n(), people = n_distinct(ssuid_spanel_pnum_id), mean = mean(tpearn), sd = sd
```

```
## # A tibble: 2 x 8
##   y1_status_v2_fct person_years people   mean     sd    p25    p50    p75
##   <fct>                   <int>  <int>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 remain_WS               50323  26592 65208. 62650. 31180  48715  78008
## 2 WS_to_SE                  388    261 74157. 97330. 29764. 45810. 78664.
```

Bivariate Model

```
m1 <- lqmm(fixed = ln_tpearn ~ y1_status_v2_fct,
                     random = ~1,
                             group = + ssuid_spanel_pnum_id,
                             tau = c( 0.25, 0.5, .75),
                             data = earnings,
                             control = lqmmControl(method = "df"))
summary(m1)
```

```
## Call: lqmm(fixed = ln_tpearn ~ y1_status_v2_fct, random = ~1, group = +ssuid_spanel_pnum_id,
##     tau = c(0.25, 0.5, 0.75), data = earnings, control = lqmmControl(method = "df"))
##
## tau = 0.25
##
## Fixed effects:
##                                  Value Std. Error lower bound upper bound
## (Intercept)                1.2407e+01 8.4951e-04  1.2405e+01     12.4085
```

```
## y1_status_v2_fctWS_to_SE 1.4132e+00 1.3454e-02  1.3861e+00      1.4402
##                            Pr(>|t|)
## (Intercept)               < 2.2e-16 ***
## y1_status_v2_fctWS_to_SE < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## tau = 0.5
##
## Fixed effects:
##                                Value  Std. Error lower bound upper bound
## (Intercept)              12.56379755  0.00086026 12.56206879     12.5655
## y1_status_v2_fctWS_to_SE  0.01545359  0.01343593 -0.01154692      0.0425
##                          Pr(>|t|)
## (Intercept)                <2e-16 ***
## y1_status_v2_fctWS_to_SE   0.2557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## tau = 0.75
##
## Fixed effects:
##                               Value Std. Error lower bound upper bound
## (Intercept)              1.2642e+01 8.6564e-04  1.2641e+01     12.6441
## y1_status_v2_fctWS_to_SE 8.2687e-01 1.3446e-02  7.9984e-01      0.8539
##                           Pr(>|t|)
## (Intercept)              < 2.2e-16 ***
## y1_status_v2_fctWS_to_SE < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## AIC:
## [1]  -9729 (df = 4)  -5993 (df = 4) -35824 (df = 4)
```

Just out of curiosity, what if we add in age as a control.

```
m2 <- lqmm(fixed = ln_tpearn ~ y1_status_v2_fct + age + age2,
                       random = ~1,
                               group = + ssuid_spanel_pnum_id,
                               tau = c( 0.25, 0.5, .75),
                               data = earnings,
                               control = lqmmControl(method = "df"))

summary(m2)
```

```
## Call: lqmm(fixed = ln_tpearn ~ y1_status_v2_fct + age + age2, random = ~1,
##     group = +ssuid_spanel_pnum_id, tau = c(0.25, 0.5, 0.75),
##     data = earnings, control = lqmmControl(method = "df"))
##
## tau = 0.25
##
## Fixed effects:
##                                Value  Std. Error lower bound upper bound
## (Intercept)               1.2110e+01  1.8363e-02  1.2074e+01     12.1474
## y1_status_v2_fctWS_to_SE -3.0220e-01  7.0774e-01 -1.7245e+00      1.1200
```

```
## age                         1.8618e-02  6.4597e-04  1.7320e-02      0.0199
## age2                       -2.3057e-04  1.5104e-05 -2.6093e-04     -0.0002
##                            Pr(>|t|)
## (Intercept)                  <2e-16 ***
## y1_status_v2_fctWS_to_SE   0.6713
## age                          <2e-16 ***
## age2                         <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## tau = 0.5
##
## Fixed effects:
##                                  Value  Std. Error lower bound upper bound
## (Intercept)                  1.2134e+01  1.2442e-02  1.2109e+01     12.1592
## y1_status_v2_fctWS_to_SE  1.3367e-02  1.4852e-02 -1.6480e-02      0.0432
## age                          1.8872e-02  6.4231e-04  1.7581e-02      0.0202
## age2                        -1.9391e-04  7.6373e-06 -2.0926e-04     -0.0002
##                            Pr(>|t|)
## (Intercept)                  <2e-16 ***
## y1_status_v2_fctWS_to_SE   0.3725
## age                          <2e-16 ***
## age2                         <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## tau = 0.75
##
## Fixed effects:
##                                  Value  Std. Error lower bound upper bound
## (Intercept)                  1.2321e+01  1.9853e-02  1.2281e+01     12.3605
## y1_status_v2_fctWS_to_SE  6.7883e-01  1.3354e-01  4.1047e-01      0.9472
## age                          2.0312e-02  7.1644e-04  1.8873e-02      0.0218
## age2                        -2.4702e-04  9.5260e-06 -2.6616e-04     -0.0002
##                             Pr(>|t|)
## (Intercept)                < 2.2e-16 ***
## y1_status_v2_fctWS_to_SE 5.807e-06 ***
## age                        < 2.2e-16 ***
## age2                       < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## AIC:
## [1] -16988 (df = 6)  -8954 (df = 6) -29745 (df = 6)
```

Some non-parametric tests of differences between medians for the mood.medtest and distribution (sort of median) for the wilcox.test

```r
RVAideMemoire::mood.medtest(ln_tpearn ~ y1_status_v2_fct, data = earnings)
```

```
##
##  Mood's median test
##
## data:  ln_tpearn by y1_status_v2_fct
## X-squared = 1.6223, df = 1, p-value = 0.2028
```

```r
wilcox.test(ln_tpearn ~ y1_status_v2_fct, data = earnings)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  ln_tpearn by y1_status_v2_fct
## W = 10066554, p-value = 0.2901
## alternative hypothesis: true location shift is not equal to 0
```