Labor Unions in Sweden

Lucas Owen

2024-01-31

Prerequisite

```
library(questionr)
library(tidyverse)

# load data
unions <- read_csv(file = "data/union_sweden.csv")
```

We will be using data from the European Social Survey, specifically from wave 5 in 2010. The dataset, unions_sweden.csv, includes information on socioeconomic characteristics for over 1,289 respondents from Sweden. Our focus is on labor unions, and we will be utilizing the following variables:

Name	Description
`Country`	Name of the respondent's country of residence.
`union`	Whether the respondent is an active member of a labor union (`0`= No, `1`= Yes)
`gndr`	Respondent's gender (`0` = male, `1`= female)
`income`	Respondent's disposable income.
`sector`	Sector in which the respondent is employed. It follows the Industry, NACE classification.
`lrscale`	Placement on left right scale (where `0` = left, `1`= right)
`gincdif`	Respondent's answer to question "Government should reduce differences in income levels" (where `5` = Agree strongly, `4` = Agree, `3` = Neither agree nor disagree, `2` = Disagree, `1` = Disagree strongly)

Union's treatment effects

Estimate the differences between the control and treatment groups based on the "union" variable for the outcomes `income` and `gincdif`. Report these estimates. Do these estimates **identify** the causal effects of the "union" variable? Check the balance between the control and treatment groups.

```
## check balance between treatment and control

table(unions$union_f)
```

```
##
##   control treatment
##       603       686
```

```r
## estimate union effect on income

mean(unions$income[unions$union_f=="treatment"]) - mean(unions$income[unions$union_f=="control"])
```

```
## [1] 2869.76
```

```r
## estimate union effect on gincdif.

mean(unions$gincdif[unions$union_f=="treatment"]) - mean(unions$gincdif[unions$union_f=="control"])
```

```
## [1] 0.1153707
```

Confounding

Since this sample has been adjusted to achieve a nearly equal number of non-union (**control**) and union membership (**treatment**) respondents, we would anticipate an approximate 50% proportion of union members across various observable factors if, and only if, the **assumption of covariate balance** holds. This balance serves as a tentative assumption of the causal effects of union membership, conditioned on other factors.

To assess this assumption, let's examine the proportion of union members at each level of the left-right ideological scale (`lrscale`). Does the data corroborate this assumption?

```r
## Look at the proportion of union memberhsip with tapply

tapply(unions$union, unions$lrscale, mean)
```

```
##         0         1         2         3         4         5         6         7
## 0.7352941 0.6190476 0.5256410 0.6422764 0.5819672 0.5665399 0.5547945 0.4908257
##         8         9        10
## 0.4378378 0.4074074 0.3777778
```

```r
## new function: aggregate

aggregate(union ~ lrscale, data = unions, FUN= mean)
```

```
##    lrscale     union
## 1        0 0.7352941
## 2        1 0.6190476
## 3        2 0.5256410
## 4        3 0.6422764
## 5        4 0.5819672
## 6        5 0.5665399
## 7        6 0.5547945
## 8        7 0.4908257
## 9        8 0.4378378
## 10       9 0.4074074
## 11      10 0.3777778
```

Additionally, consider identifying another variable that may confound union membership.

```r
## check balance of union membersip based on gender

mean(unions$gndr[unions$union_f=="treatment"]) - mean(unions$gndr[unions$union_f=="control"])
```

```
## [1] 0.05445803
```

```r
## check proportion of union membership conditional on sector

aggregate(union ~ sector, data = unions, FUN= mean)
```

```
##                                                                  sector
## 1                                  A: Agriculture, forestry and fishing
## 2                                              B: Mining and quarrying
## 3                                                     C: Manufacturing
## 4                           D: Electricity, gas, steam and air conditioning
## 5                  E: Water supply, sewerage, waste management and remediation
## 6                                                      F: Construction
## 7  G: Wholesale and retail trade, repair of motor vehicles and motorcycles
## 8                                         H: Transportation and storage
## 9                                    I: Accommodation and food services
## 10                                  J: Information and communication
## 11                               K: Financial and insurance activities
## 12                                           L: Real estate activities
## 13                    M: Professional, scientific, and technical activites
## 14                        N: Adminsitrative and support service activities
## 15         O: Public adminsitration and defence, compulsory social security
## 16                                                       P: Education
## 17                                     Q: Human health and social work
## 18                               R: Arts, entertainment and recreation
## 19                                         S: Other service activities
## 20                              T: Activities of households as employers
##        union
## 1  0.4838710
## 2  1.0000000
## 3  0.5438596
## 4  0.8333333
## 5  1.0000000
## 6  0.5384615
## 7  0.4088050
## 8  0.5535714
## 9  0.2500000
## 10 0.5208333
## 11 0.4375000
## 12 0.5263158
## 13 0.4268293
## 14 0.6842105
## 15 0.7078652
## 16 0.5985915
## 17 0.5618557
## 18 0.4102564
## 19 0.6250000
## 20 0.0000000
```

Adjusting confounding

Use the `lm()` function to regress the outcomes `income` and `gincdif` on the treatment variable union using a linear model. Save the output in an object and use the `summary()` function to examine the estimated coefficients. Compare these coefficients with the previously calculated differences for the union variable.

```r
m1 <- lm(income ~ union, data = unions)
m2 <- lm(gincdif ~ union, data = unions)
summary(m1)
```

Call: lm(formula = income ~ union, data = unions)

Residuals: Min 1Q Median 3Q Max -16469.4 -5635.6 -797.3 5241.3 21968.4

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 17624.1 357.4 49.316 < 2e-16 ***union 2869.8 489.9 5.858 5.94e-09*** — Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8776 on 1287 degrees of freedom Multiple R-squared: 0.02597, Adjusted R-squared: 0.02522 F-statistic: 34.32 on 1 and 1287 DF, p-value: 5.935e-09

```r
library(stargazer)

table_1 <- stargazer(m1,m2,
          type="latex")     # change to type="text" to display the results in the console
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Feb 11, 2025 - 08:38:00

Table 2:

	Dependent variable:	
	income	gincdif
	(1)	(2)
union	2,869.760***	0.115**
	(489.871)	(0.053)
Constant	17,624.080***	3.637***
	(357.369)	(0.038)
Observations	1,289	1,289
R^2	0.026	0.004
Adjusted R^2	0.025	0.003
Residual Std. Error (df = 1287)	8,775.575	0.943
F Statistic (df = 1; 1287)	34.319***	4.800**
Note:	*p<0.1; **p<0.05; ***p<0.01	

Repeat the same exercise, but this time include the variables `lrscale` and `sector`. How much have change your estimations after **adjusting for confounding**? Can we say that we have identified a causal estimate?

```r
m3 <- lm(income ~ sector + lrscale + union, data = unions)
m4 <- lm(gincdif ~ sector + lrscale + union, data = unions)


stargazer(m3,m4,
          omit = "sector", # for space concern, omit reporting sector coefficients.
          type="latex")    # change to type="text" to display the results in the console
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Feb 11, 2025 - 08:38:00

Table 3:

	Dependent variable:	
	income	gincdif
	(1)	(2)
lrscale	815.810***	−0.159***
	(104.391)	(0.011)
union	3,329.746***	0.003
	(480.351)	(0.050)
Constant	4,383.720**	4.764***
	(1,705.640)	(0.177)
Observations	1,289	1,289
R^2	0.135	0.176
Adjusted R^2	0.121	0.162
Residual Std. Error (df = 1267)	8,334.993	0.865
F Statistic (df = 21; 1267)	9.414***	12.887***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Are there differences in redistributive attitudes between male and female union members? Compare the `gincdif` variable between the two groups.

```r
# group_by with dplyr

unions %>%
  group_by(union_f,gndr_f) %>%
  summarize(gincdif=mean(gincdif,na.rm=T)) %>%
  pivot_wider(names_from = c("union_f","gndr_f"),
              values_from = gincdif) %>%
  mutate(gndr_diff=treatment_Female-treatment_Male)
```

```
## # A tibble: 1 x 5
##   control_Female control_Male treatment_Female treatment_Male gndr_diff
##            <dbl>        <dbl>            <dbl>          <dbl>     <dbl>
## 1           3.72         3.56             3.84           3.65     0.196
```

```r
# with base R

women_union <- mean(unions$gincdif[unions$union_gndr=="trt_female"])
men_union <- mean(unions$gincdif[unions$union_gndr=="trt_male"])
women_union - men_union
```

```
## [1] 0.1958605
```

Also, examine union-gender differences by sector and save your visualizations to your local folder. Are these differences considered **causal effects**?

```r
unions %>%
  # group by categories
  group_by(sector,union_f,gndr_f) %>%
  # estimate the summaries of interest
  summarize(ginunions=mean(gincdif,na.rm=T)) %>%
  # pivot from long to wide
  pivot_wider(id_cols = "sector",
              names_from = c("union_f","gndr_f"), # note the two groups
              values_from = "ginunions") %>%
  mutate(diff=treatment_Female-treatment_Male,
         diff = round(diff,2))
```

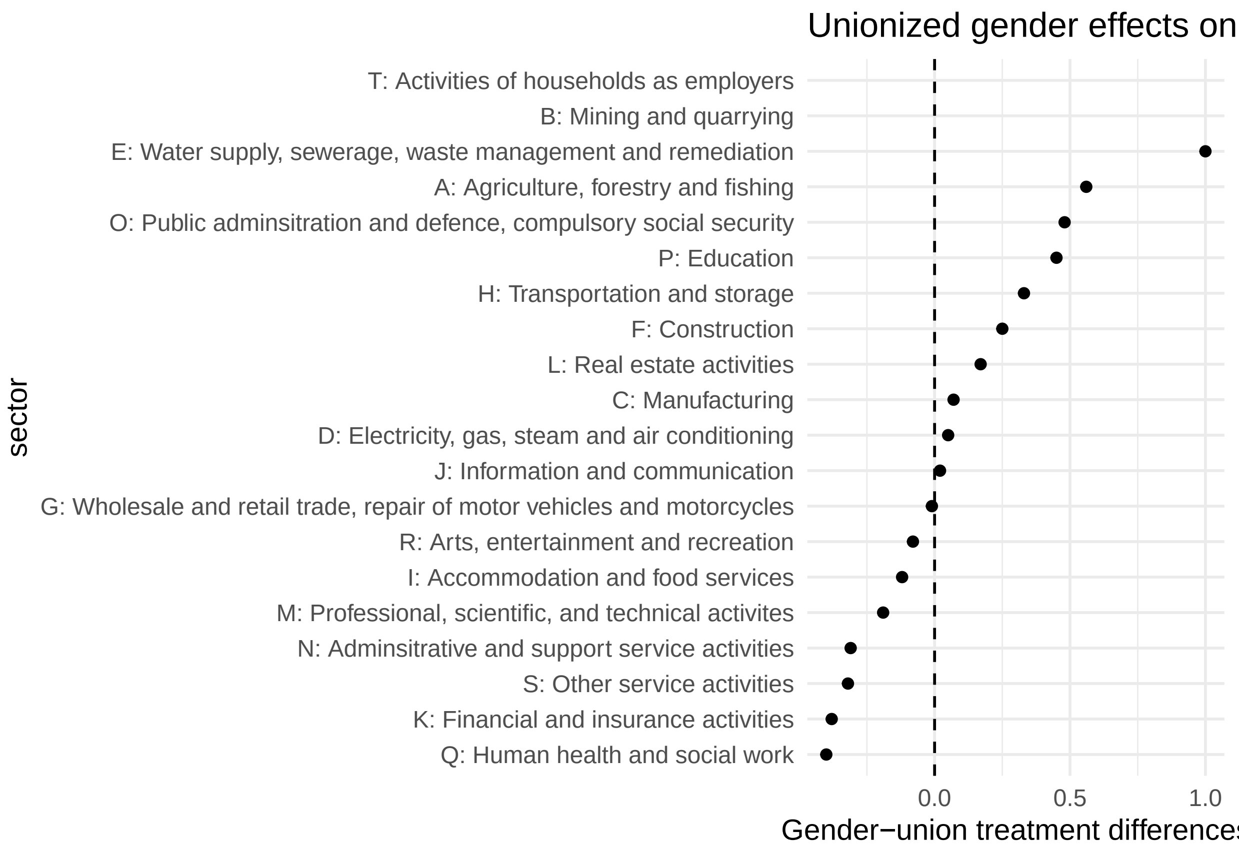
```
## # A tibble: 20 x 6
## # Groups:   sector [20]
##    sector      control_Female control_Male treatment_Female treatment_Male  diff
##    <chr>                <dbl>        <dbl>            <dbl>          <dbl> <dbl>
##  1 A: Agricul~           4            3.5              4              3.44  0.56
##  2 B: Mining ~          NA           NA               NA              4.5  NA
##  3 C: Manufac~           3.78         3.68             3.5            3.43  0.07
##  4 D: Electri~          NA            3.5              3.33           3.29  0.05
##  5 E: Water s~          NA           NA                4              3     1
##  6 F: Constru~           3.83         3.61             4              3.75  0.25
##  7 G: Wholesa~           3.79         3.47             3.68           3.68 -0.01
##  8 H: Transpo~           3.67         3.73             4.29           3.96  0.33
##  9 I: Accommo~           3.52         3.86             3.88           4    -0.12
## 10 J: Informa~           4            3.33             3.38           3.35  0.02
## 11 K: Financi~           3.55         3                3.62           4    -0.38
## 12 L: Real es~           4            3.71             3.5            3.33  0.17
## 13 M: Profess~           3.4          3.30             3.38           3.57 -0.19
## 14 N: Adminsi~           3.25         3.38             3.46           3.77 -0.31
## 15 O: Public ~           3.71         3.75             4.03           3.55  0.48
## 16 P: Educati~           3.76         3.62             3.92           3.47  0.45
## 17 Q: Human h~           3.89         3.36             3.94           4.33 -0.4
## 18 R: Arts, e~           3.27         3.83             4.17           4.25 -0.08
## 19 S: Other s~           3.43         3.88             4.18           4.5  -0.32
## 20 T: Activit~          NA            3               NA             NA    NA
```

```r
# save the data in an object for visualization

vis <-
  unions %>%
```

```r
  group_by(sector,union_f,gndr_f) %>%
  summarize(ginunions=mean(gincdif,na.rm=T)) %>%
  pivot_wider(id_cols = "sector",
              names_from = c("union_f","gndr_f"),
              values_from = "ginunions") %>%
  mutate(diff=treatment_Female-treatment_Male,
         diff = round(diff,2))

vis %>%
  ggplot(aes(x=diff,
             y=(sector=reorder(sector, diff)))) +
  geom_point() +
  geom_vline(xintercept = 0,linetype="dashed") +
  theme_minimal() +
  labs(y="sector",
       x="Gender-union treatment differences",
       title="Unionized gender effects on solidarity by sector")
```




```r
# width = 6
# ggsave("output/gndr_diff.punions", width = width, height = width/1.618)
```