

# LSEM Individual

Antonio R. Linero

2022-10-06

Load the data:

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.6       v dplyr 1.0.8
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

meps <- read_csv("../Data/meps2011.csv") %>%
  filter(totexp > 0 & bmi > 0) %>%
  mutate(logY = log(totexp)) %>%
  mutate(smoke = ifelse(smoke == "No", 0, 1)) %>%
  select(-totexp) %>% select(logY, everything())

##
## -- Column specification -----
## cols(
##   age = col_double(),
##   bmi = col_double(),
##   edu = col_double(),
##   income = col_double(),
##   povlev = col_double(),
##   region = col_character(),
##   sex = col_character(),
##   marital = col_character(),
##   race = col_character(),
##   seatbelt = col_character(),
##   smoke = col_character(),
##   phealth = col_character(),
##   totexp = col_double()
## )

phealth <- meps$phealth
phealth <- case_when(phealth == "Poor" ~ 1, phealth == "Fair" ~ 2,
  phealth == "Good" ~ 3, phealth == "Very Good" ~ 4,
  phealth == "Excellent" ~ 5)
meps$phealth <- phealth
rm(phealth)
```

Then fit (say) the model for phealth

```
fit_m <- lm(phealth ~ smoke * (bmi + edu + log(income + 1000) + region + sex +
                                marital + race + seatbelt),
            data = meps)
```

This model is of the form

$$M_i = \beta_0 + X_i^\top \beta + A_i(\gamma_0 + X_i^\top \gamma) + \text{error}_i$$

So  $\tau(x) = \gamma_0 + X_i^\top \gamma$ . We can extract these coefficients as the interaction terms of the model:

```
gamma <- coef(fit_m)[str_detect(names(coef(fit_m)), "smoke")]
print(gamma)
```

```
##              smoke              smoke:bmi              smoke:edu
##          -0.564480622              0.006790751              0.012726249
## smoke:log(income + 1000) smoke:regionNortheast smoke:regionSouth
##              0.023094509              -0.077035078              0.055130227
## smoke:regionWest smoke:sexMale smoke:maritalMarried
##          -0.079752349              0.019620879              -0.078836787
## smoke:maritalSeparated smoke:maritalWidowed smoke:raceBlack
##          -0.042419273              0.126716746              -0.266407299
## smoke:raceIndig smoke:raceMulti smoke:racePacificIslander
##          -0.360579741              -0.124910254              0.654177813
## smoke:raceWhite smoke:seatbeltAlways smoke:seatbeltNever
##          -0.321200733              -0.010987533              -0.192360418
## smoke:seatbeltNoCar smoke:seatbeltSeldom smoke:seatbeltSometimes
##              0.066108809              0.166416727              0.120287792
```

We can then extract the design matrix using `model.matrix()`:

```
X_tau <- model.matrix(phealth ~ bmi + edu + log(income + 1000) + region + sex +
                      marital + race + seatbelt, data = meps)
head(X_tau)
```

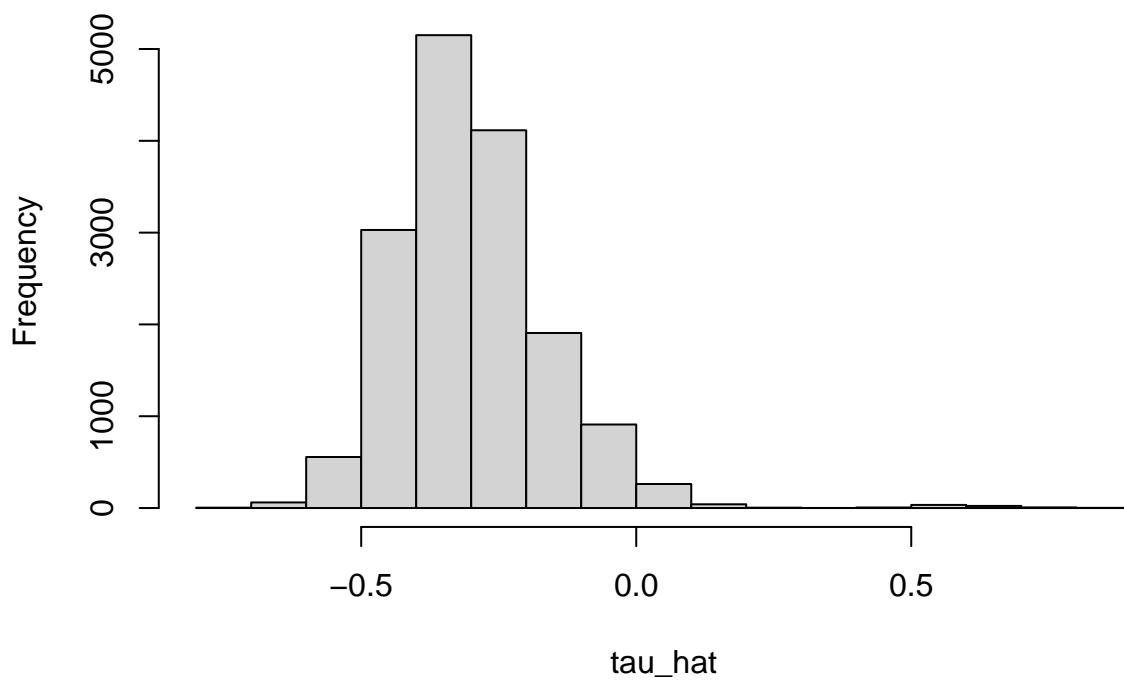
```
## (Intercept)  bmi edu log(income + 1000) regionNortheast regionSouth
## 1           1 39.1 14           11.282254           1           0
## 2           1 20.2 17           12.111388           0           0
## 3           1 21.0 14           10.275017           0           0
## 4           1 25.7 12           10.275017           0           0
## 5           1 23.0 12           9.667765           0           1
## 6           1 23.4 9           10.741168           0           0
## regionWest sexMale maritalMarried maritalSeparated maritalWidowed raceBlack
## 1           0           1           1           0           0           0
## 2           1           1           1           0           0           0
## 3           1           1           1           0           0           0
## 4           1           0           1           0           0           0
## 5           0           0           0           0           0           0
## 6           1           0           0           1           0           0
## raceIndig raceMulti racePacificIslander raceWhite seatbeltAlways
## 1           0           0           0           1           1
## 2           0           1           0           0           1
## 3           0           0           0           1           1
## 4           0           0           0           1           1
## 5           0           0           0           1           1
```

## 6	0	0	1	0	1
##	seatbeltNever	seatbeltNoCar	seatbeltSeldom	seatbeltSometimes	
## 1	0	0	0	0	
## 2	0	0	0	0	
## 3	0	0	0	0	
## 4	0	0	0	0	
## 5	0	0	0	0	
## 6	0	0	0	0	

So we can get  $\tau(X_i)$  estimates for each individual in the sample as:

```
tau_hat <- as.numeric(X_tau %*% gamma)
hist(tau_hat)
```

**Histogram of tau\_hat**



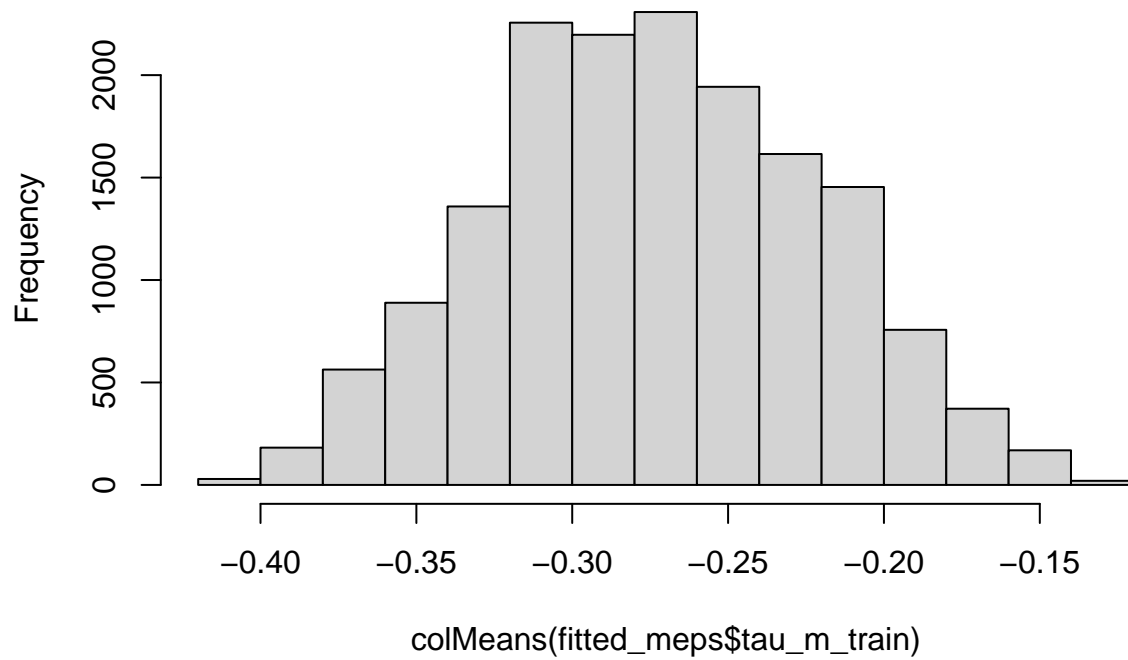
Comparing this with the estimates from BCMF:

```
fitted_meps <- readRDS("../Data/meps_fit.rds")
```

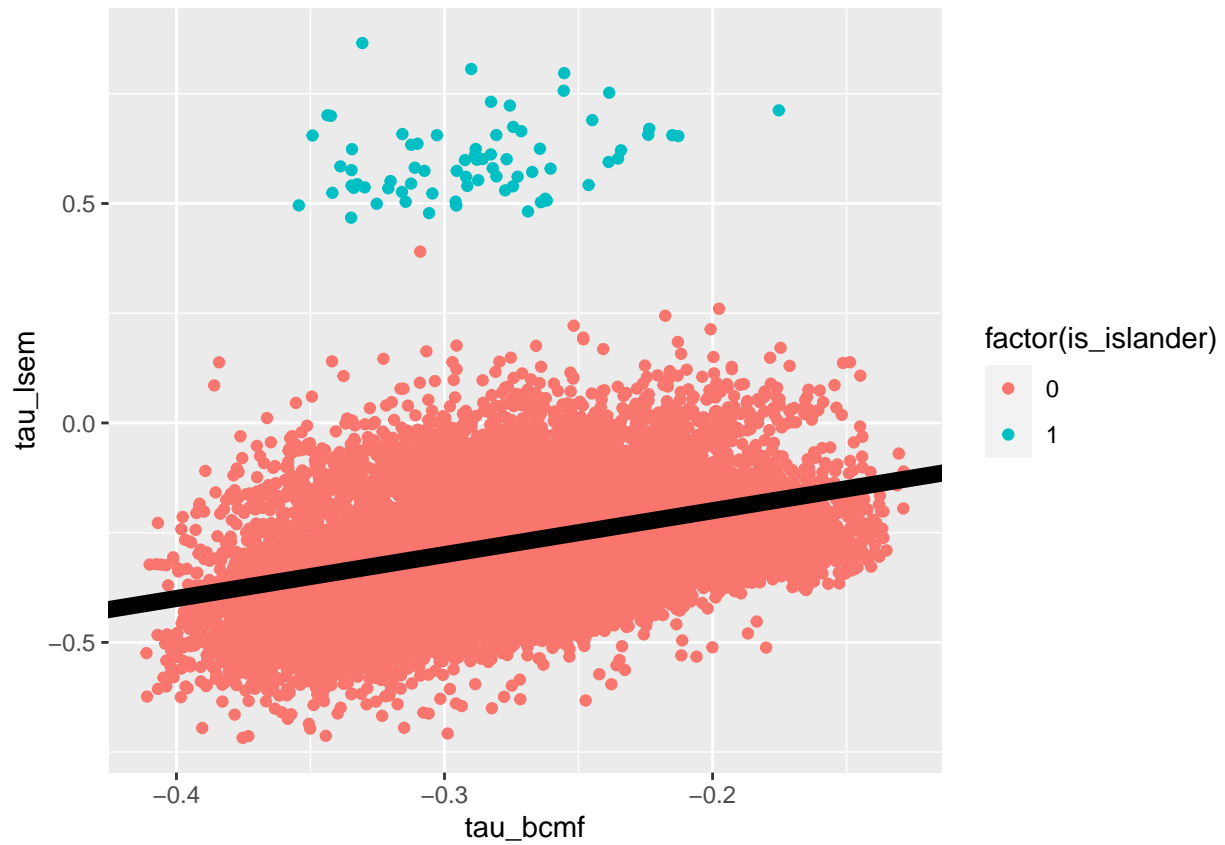
It turns out that the linear model behaves strangely for pacific islanders, probably because there aren't many pacific islanders in the sample:

```
hist(colMeans(fitted_meps$tau_m_train))
```

## Histogram of colMeans(fitted\_meps\$tau\_m\_train)



```
data_for_plot <- data.frame(tau_lsem = tau_hat,  
                             tau_bcmf = colMeans(fitted_meps$tau_m_train),  
                             is_islander = X_tau[, "racePacificIslander"])  
  
ggplot(data_for_plot, aes(x = tau_bcmf, y = tau_lsem,  
                           color = factor(is_islander))) +  
  geom_point() +  
  geom_abline(slope = 1, intercept = 0, size = 3)
```



Anyway, you can do all this to get LSEM-specific estimates of  $d(x)$  and  $\zeta(x)$  as well from the outcome model. You just extract a different set of coefficients and use the same design matrix that I got above.