Intelligent Business

FKA Multivariate regression

Taking back control …

…of confounding factors

by Ralf Martin (r.martin@imperial.ac.uk)

Objectives for this lecture

- Learn how to include further (control) variables in a regression $\Rightarrow$ Multivariate regression
- Interpret multivariate regressions properly
- Understand that including more variables is not always better
- Learn how to perform hypothesis tests involving several parameters

Why Multivariate Regression?

= we have more than one explanatory variable

e.g. $$Wage = \beta_1 + \beta_2 EDU + \beta_3 FEMALE + u$$

Years of Schooling

$$= \begin{cases} 1 \text{ for women} \\ 0 \text{ for men} \end{cases}$$

Why?

1. Interest in several effects at once

2. **<u>Address confounding/endogeneity</u>**

3. Can help to reduce variance of estimate

Endogeneity and Multivariate Regression

Suppose we are only interested in schooling

$$Wage = \beta_1 + \beta_2 EDU + \epsilon$$

so that $\epsilon = \beta_3 FEMALE + u$

Am not trying (or succeeding) to be misogynistic here. It's simply what we find in the data as we shall see later

e.g. women tend to have less schooling

- Suppose we are only interested in the effect of schooling
- But there is a correlation between schooling and gender and gender has also a separate effect on wages

$\Longrightarrow$Estimate of β_2 is biased

Your turn: What bias do you expect? Upward, downward, none?

$$Wage = \beta_1 + \beta_2 EDU + \epsilon$$

$$\epsilon = \beta_3 FEMALE + u$$

Go to **www.menti.com** and use the code **76 14 60 0**

Mentimeter

Regressing Wages on Schooling - What bias do you expect because of gender on the schooling coefficient?

$$Wage = \beta_1 + \beta_2 EDU + \epsilon$$

Upward Bias	Downward Bias	No Bias
0	0	0

Endogeneity and Multivariate Regression

-

$$Wage = \beta_1 + \beta_2 EDU + \epsilon$$

+

$$\epsilon = \beta_3 FEMALE + u$$

$\beta_3 < 0$

- Negative effect of FEMALE on WAGE and EDU implies positive correlation between ϵ
- We get upward bias when attempting to estimate β_2

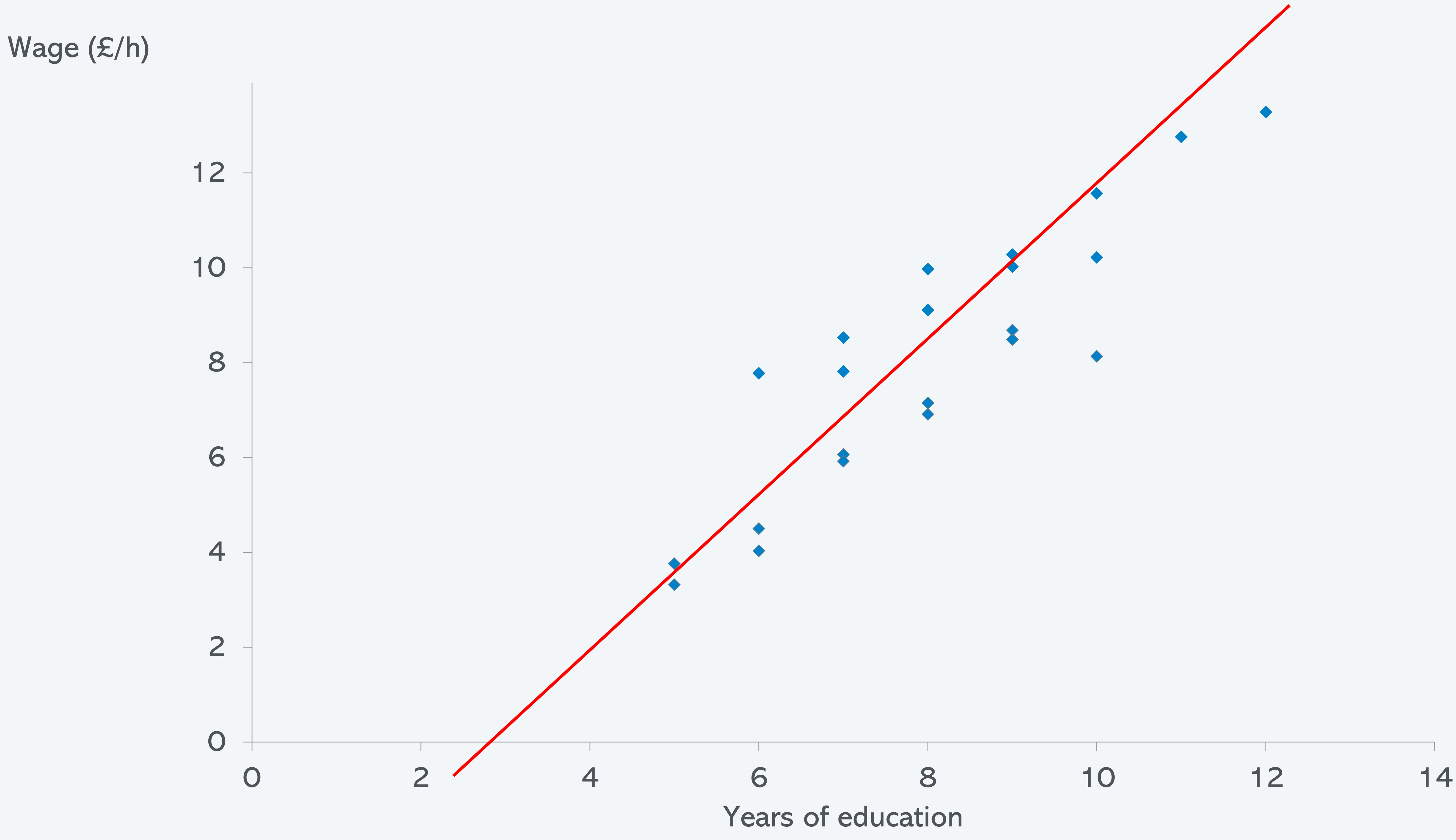
Endogeneity and Multivariate Regression

- To avoid the problem we can include $FEMALE$ as additional variable:

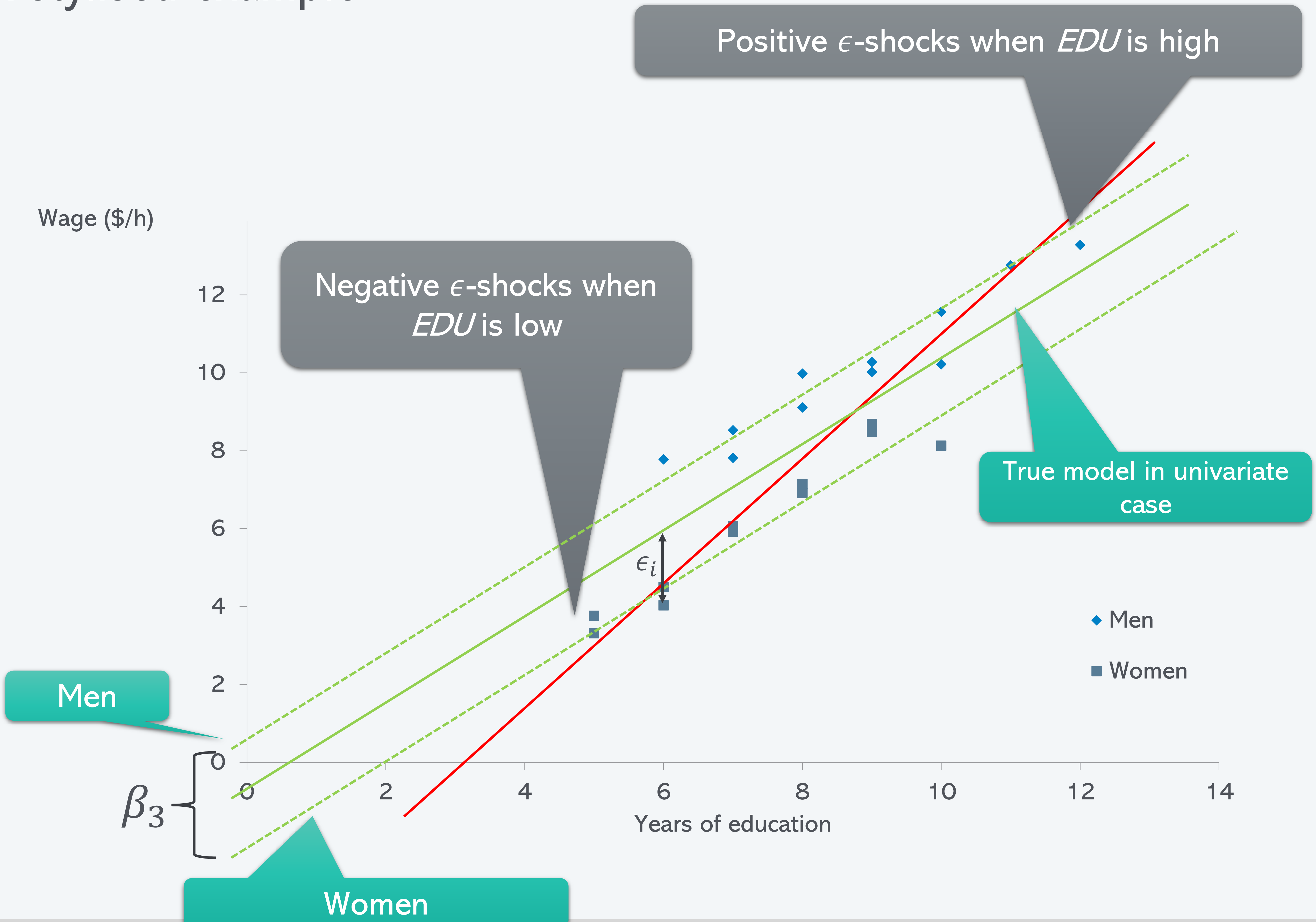
$$Wage = \beta_1 + \beta_2 EDU + \beta_3 FEMALE + u$$

- We get a different regression model with a different residual term u
- This will lead to an **unbiased** $\hat{\beta}_2$ if EDU is independent (i.e. uncorrelated) with u

A stylised example



A stylised example



Causality vs all else equal

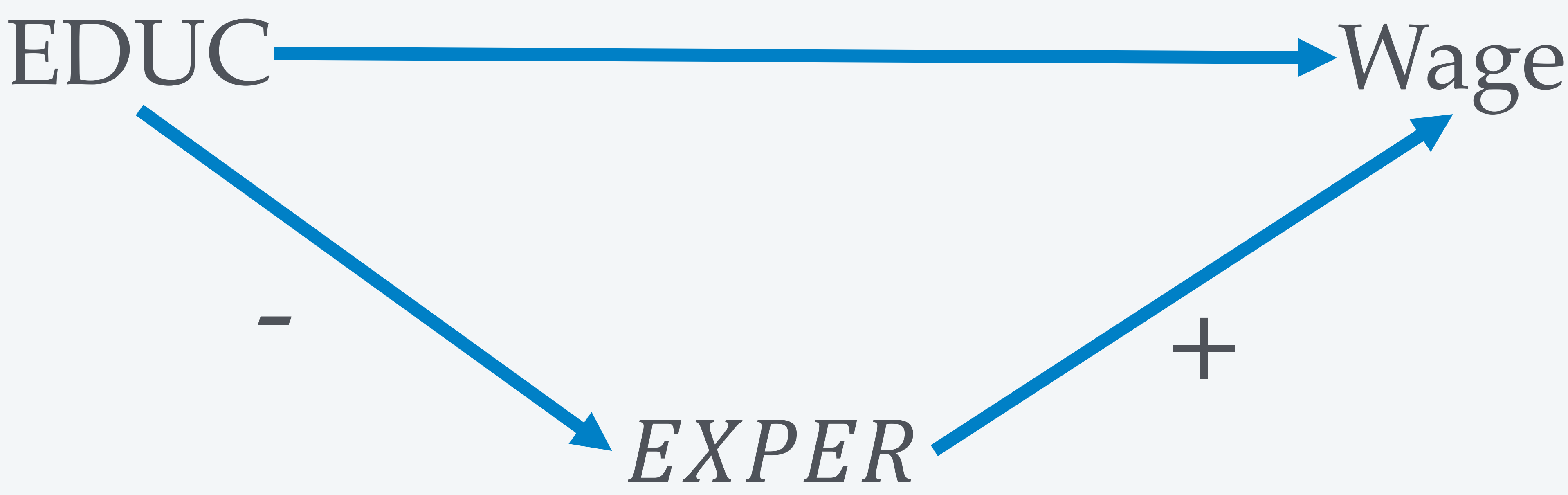
Another variable that is likely correlated with schooling and affecting wages is experience (EXPER).

$$Wage = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + u$$

If we run a regression of this equation (and EDUC and EXPER are independent of u) the estimate of β_2 gives us the change Wage for one year more of schooling keeping experience (and everything else constant)

However, it might not give us the causal effect of increasing schooling on wages.

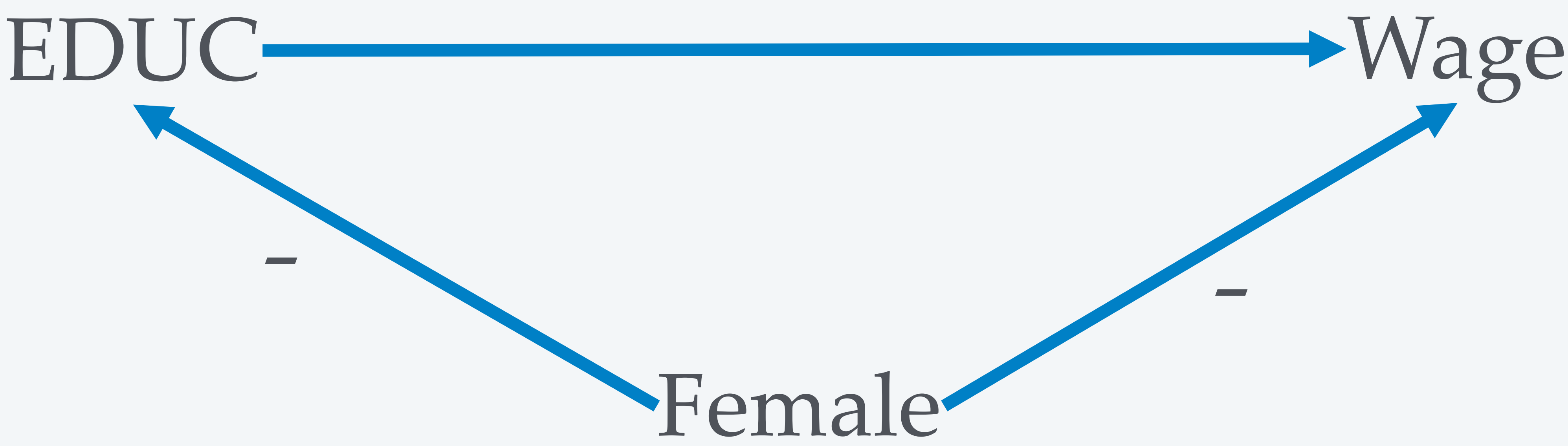
Directions of causality EDU☞EXP



The reason why EDUC and EXPER are correlated is likely because of a chain of causality from schooling to experience (i.e. if you go to school longer you don't have so much time to get job experience; also not that S is typically determined before EXPER which supports the suggested chain)

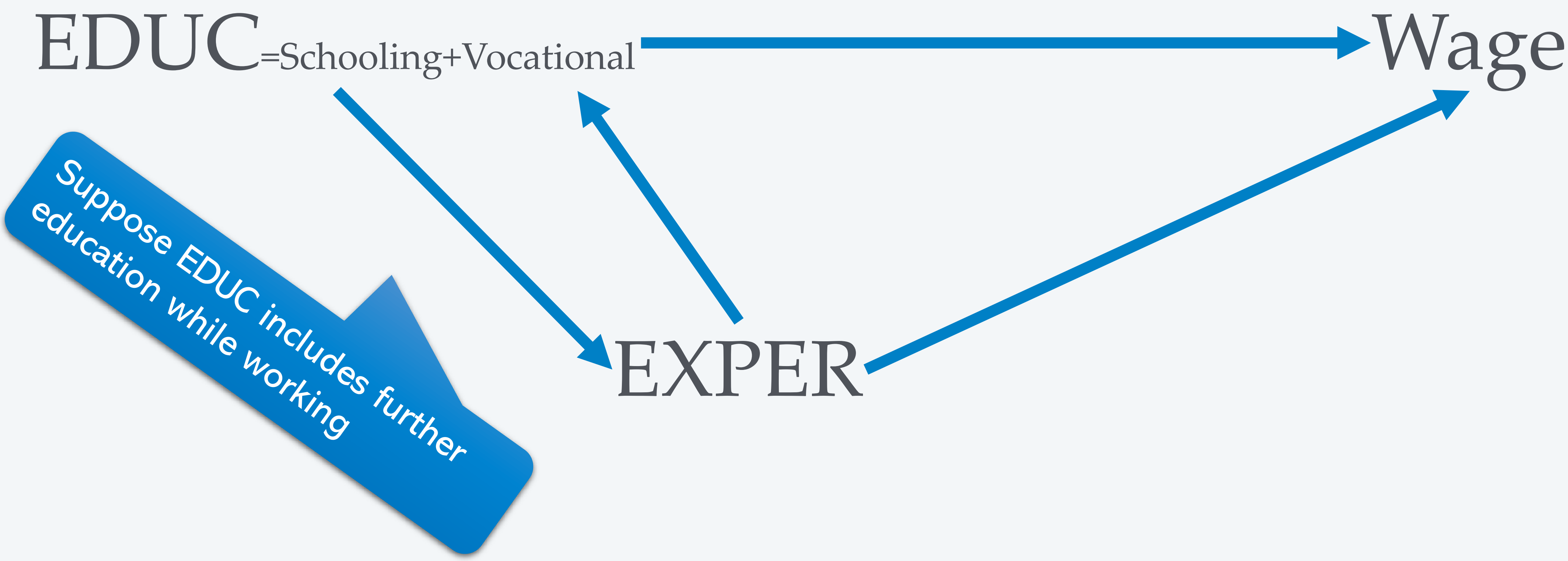
If you include EXPER as separate explanatory variable then your coefficient on EDUC will not reflect this causal channel. This is good if you really want the all else equal effect of EXPER. However, if you want the full causal effect of EDUC (e.g. you want to advise the government what an extra year of schooling does to wages) you get the wrong answer as you are pretending that you can have extra schooling without reducing people's experience. So it would be better to exclude EXPER.

Directions of causality EDUC☜FEMALE



- Gender is mostly (but not exclusively) determined before schooling
- Hence the reason why EDUC and Female variable are correlated because of a causality chain from Female to EDUC.
- In this case it is vital to include the Female variable to get the correct causal estimate of a change in EDUC

Directions of causality EDUC☞☜EXPER



If the causality between the two explanatory variables goes both ways we are in trouble as far as finding the causal effect of EDUC is concerned (we are cool for finding the ceteris paribus effect). Both including or dropping the gender variable will lead to a biased estimate. We have to use other methods some of which we shall discuss later in the module.

Key insight: You can be too controlling

- More control variables are not always better to identify a causal effect
- To include or not include → depends on direction of causation between control and x var of interest
- Sometimes there is no clear cut answer as causation goes both ways

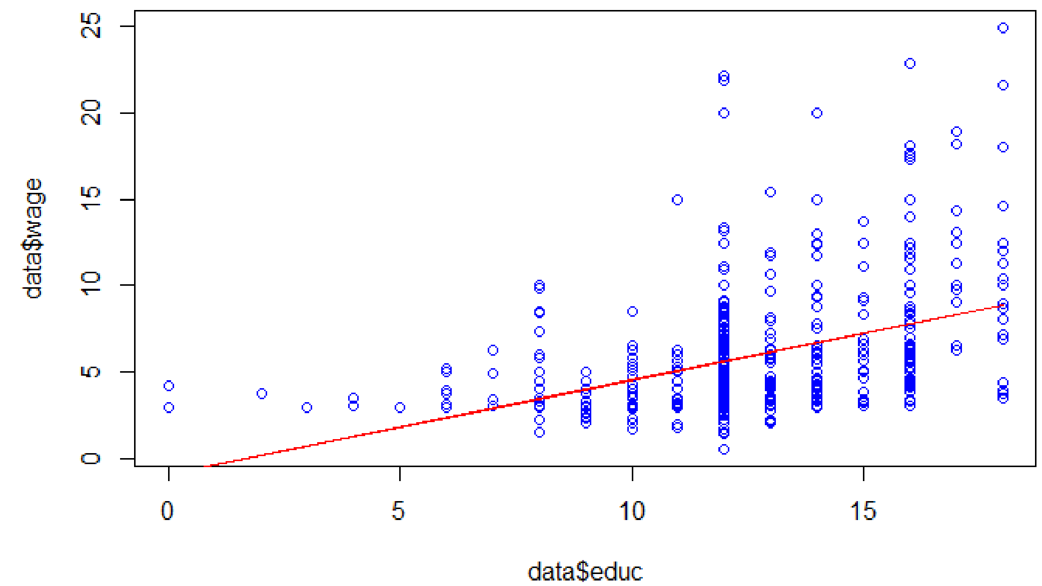
- Report regression with and without control and discuss limitations of your analysis
- More research with other data or better model (e.g. Instrumental Variables which we discuss later) might be needed.
- Might sometimes be beyond the scope of a study (e.g. in group coursework)

Multivariate OLS in practice – Let’s start univariate

```
data <- read.csv("https://www.dropbox.com/s/9agc2vmamfztlel/WAGE1.csv?dl=1")

mod1 <- lm(wage ~ educ, data)
summary(mod1)

##
## Call:
## lm(formula = wage ~ educ, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3396 -2.1501 -0.9674  1.1921 16.6085
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.90485    0.68497  -1.321    0.187
## educ         0.54        0.05325  10.167   <2e-16 ***
## ---
## Signif. codes:        1 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
##
##
```



$$\widehat{WAGE} = -0.9 + 0.54 \times EDUC$$

Indicates that earnings per hour increase by $0.54 for every extra year of schooling

EDUC vs EXPER

```
> summary(lm(exper ~ educ   , data))

Call:
lm(formula = exper ~ educ, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-20.652  -9.971  -2.971   9.125  30.625 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  35.4615     2.6279  13.494  < 2e-16 ***
educ         -1.4682     0.2043  -7.187  2.3e-12 ***
---
```

- One more year of education means 1.4 years less experience

Your turn: If we include *exper* as additional variable in a regression of WAGE what effect you expect that to have on the coefficient for education?
(a) EDUC coefficient goes up
(b) EDUC coefficient goes down
(c) EDUC coefficient remains unchanged

$$Wage = \beta_1 + \boldsymbol{\beta_2} EDU + \epsilon$$

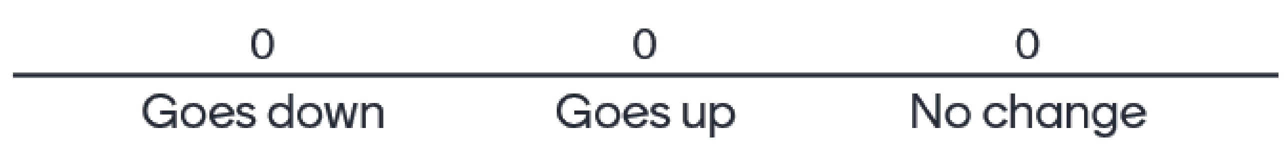
vs

$$Wage = \beta_1 + \boldsymbol{\beta_2} EDU + \beta_3 EXPER + \epsilon$$

Go to **www.menti.com** and use the code **76 14 60 0**

Mentimeter

What is the effect of including EXPER on the EDUC coefficient?



Multivariate OLS in practice

```
> summary(lm(wage ~ educ + exper, data))

Call:
lm(formula = wage ~ educ + exper, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.5532 -1.9801 -0.7071  1.2030 15.8370 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.39054    0.76657  -4.423 1.18e-05 ***
educ         0.64427    0.05381  11.974  < 2e-16 ***
exper        0.07010    0.01098   6.385 3.78e-10 ***
---
```

$$\widehat{WAGE} = -3.39 + 0.644 \times EDUC + 0.07 \times EXPER$$

- It indicates that earnings per hour increase by \$0.64 for every extra year of schooling and by \$0.07 for every extra year of work experience.
- EDUC coefficient went up (previously 0.54) because of negative correlation between EDUC and EXPER and because EXPER has positive influence on wage.
- EDUC coefficient represents now “all else equal” but no longer causal effect

Wage & Gender

```
> summary(lm(educ ~ female,data))

Call:
lm(formula = educ ~ female, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.3175  -0.7883  -0.3175   1.6825   5.6825 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  12.7883     0.1668  76.652   <2e-16 ***
female       -0.4709     0.2410  -1.953   0.0513 .  
---
```

- Women tend to be less educated than men

Wage & Gender

```
> summary(lm(wage ~ educ+female,data))

Call:
lm(formula = wage ~ educ + female, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.9890 -1.8702 -0.6651  1.0447 15.4998 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.62282    0.67253   0.926    0.355    
educ         0.50645    0.05039  10.051  < 2e-16 ***
female      -2.27336    0.27904  -8.147 2.76e-15 ***
---
```

- Including the Female variable makes the EDUC coefficient smaller
- This is because women are paid less than men irrespective of education
- Hence part of what we thought was the wage depressing effect of little education is actually the wage depressing effect of being female
- Compare with EXPER which had a positive effect on wages

More than 2 variables

```
summary(lm(wage  ~ educ+exper+female,data))
```

```
##
## Call:
## lm(formula = wage ~ educ + exper + female, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3856 -1.9652 -0.4931  1.1199 14.8217
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.73448    0.75362  -2.302   0.0218 *
## educ         0.60258    0.05112  11.788  < 2e-16 ***
## exper        0.06424    0.01040   6.177 1.32e-09 ***
## female      -2.15552    0.27031  -7.974 9.74e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.078 on 522 degrees of freedom
## Multiple R-squared:  0.3093, Adjusted R-squared:  0.3053
## F-statistic: 77.92 on 3 and 522 DF,  p-value: < 2.2e-16
```

Perfect Multi-collinearity

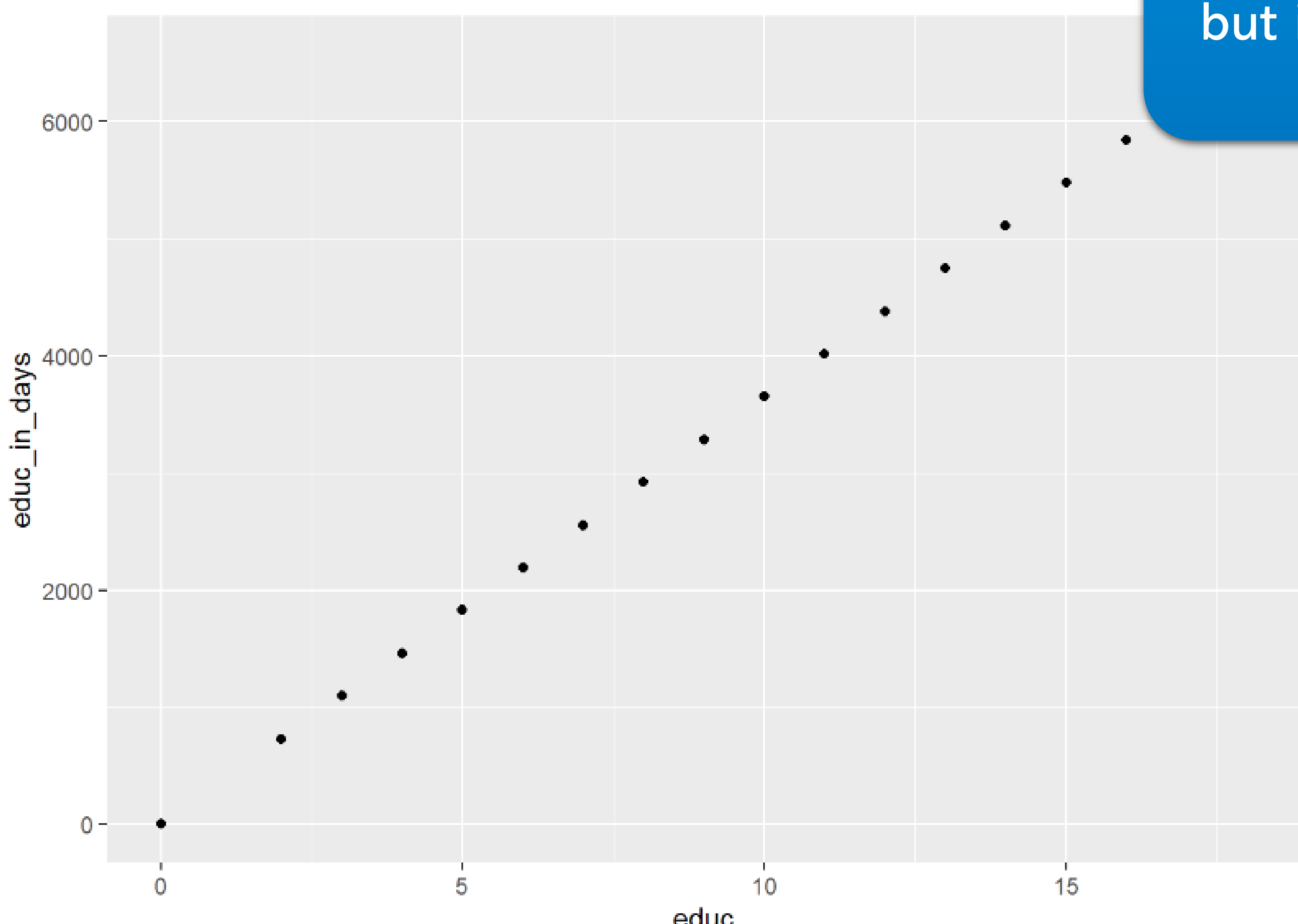
- A specific problem in the multivariate case
- Sample has not enough variation in the explanatory variables

```
library(ggplot2)
data=data %>% mutate(educ_in_days=educ*365)
cor(data %>% select(educ,educ_in_days))
```

```
##              educ educ_in_days
## educ            1            1
## educ_in_days    1            1
```

```
ggplot(data,aes(x=educ,y=educ_in_days))+geom_point()
```

Implies perfect multi-collinearity
educ in days will be different numbers but it will be perfectly correlated with *educ* in years



Perfect Multi-collinearity

```r
reg2=lm(wage~female+educ+educ_in_days,data)
reg2 %>% summary()
```

```
##
## Call:
## lm(formula = wage ~ female + educ + educ_in_days, data = data
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9890 -1.8702 -0.6651  1.0447 15.4998
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.62282    0.67253   0.926    0.355
## female       -2.27336    0.27904  -8.147 2.76e-15 ***
## educ          0.50645    0.05039  10.051  < 2e-16 **
## educ_in_days       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 3.186 on 523 degrees of freedom
## Multiple R-squared:  0.2588, Adjusted R-squared:  0.256
## F-statistic: 91.32 on 2 and 523 DF,  p-value: < 2.2e-16
```

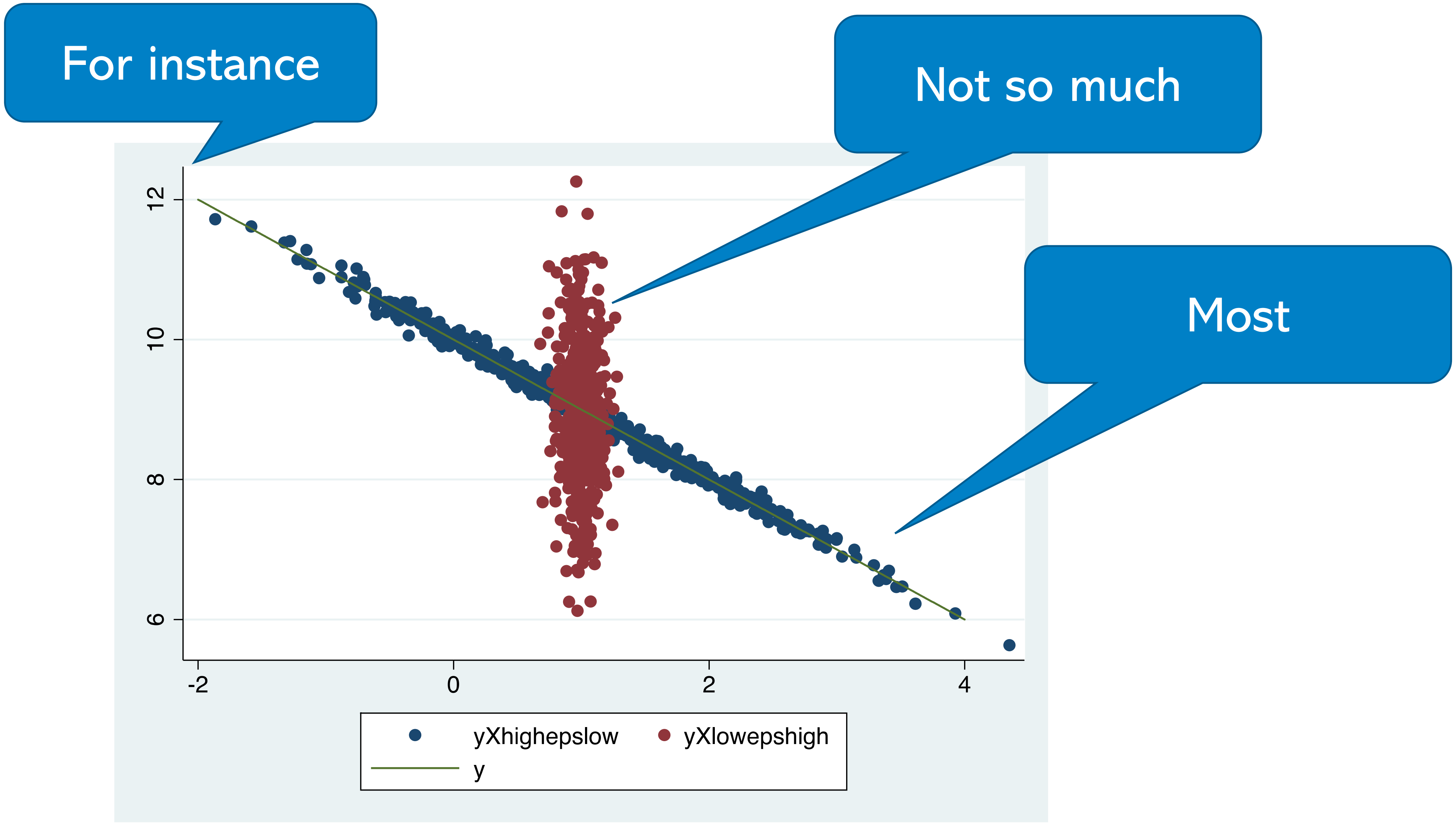
We cannot work the effect of "educ in days" holding "educ in years" constant because all observations with a given "educ in years" all have the same "educ in days"

- R will drop one of the variables to allow OLS
- Which one is dropped has no meaning

Extra Slides

Accounting for variation: R^2

- How much of the variation in Y is accounted for by $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon$?



$$R^2 = \frac{VAR(\hat{Y})}{VAR(Y)}$$

Finding R^2

```
summary(lm(wage  ~ educ+exper+female,data))
```

```
##
## Call:
## lm(formula = wage ~ educ + exper + female, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3856 -1.9652 -0.4931  1.1199 14.8217
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.73448    0.75362  -2.302   0.0218 *
## educ         0.60258    0.05112  11.788  < 2e-16 ***
## exper        0.06424    0.01040   6.177 1.32e-09 ***
## female      -2.15552    0.27031  -7.974 9.74e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.078 on 522 degrees of freedom
## Multiple R-squared:  0.3093, Adjusted R-squared:  0.3053
## F-statistic: 77.92 on 3 and 522 DF,  p-value: < 2.2e-16
```

- Accounting is not necessarily explaining
- R^2 is mechanically increasing as we add further variables
- If we have as many parameters as observations R^2 is always 100% (e.g. consider 2 observations)
- Hence Adjusted $\bar{R}^2 = 1 - \frac{(1-R^2)(n-1)}{n-(k+1)}$ where k=Number of variables
- i.e. the higher k the lower $\bar{R}$

Back to the criminal foreigners

```
## lm(formula = crimesPc ~ b_migr11 + pop11, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6243 -0.4052 -0.1253  0.2347 13.8304
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.124e+00  1.018e-01  11.034  < 2e-16 ***
## b_migr11     4.105e-02  5.335e-03   7.694 1.77e-13 ***
## pop11       -1.033e-06  5.078e-07  -2.034   0.0428 *
## ---
## S                                                 ' 0.1 ' ' 1
##
```

…those areas are less “crime intensive”

```
## lm(formula = b_migr11 ~ pop11, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.039  -5.187  -2.698   1.225  40.835
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.240e+00  9.530e-01   6.548 2.18e-10 ***
## pop11       3.088e-05  4.883e-06   6.326 8.02e-10 ***
```

Foreigners come to more populous areas but ….

Back to the criminal foreigners

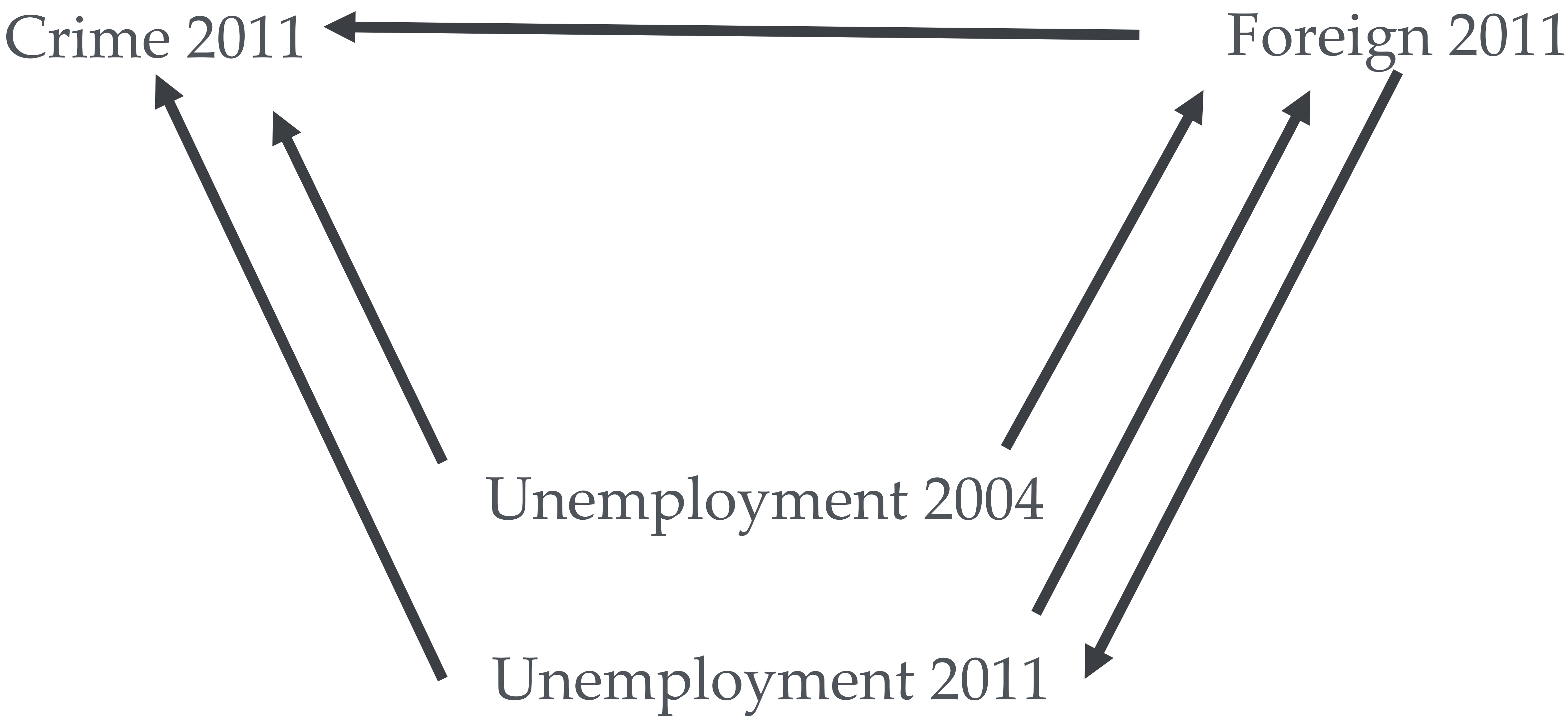
```
## lm(formula = crimesPc ~ b_migr11 + pop11 + urate2011 +
medianage,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8873 -0.2680 -0.0783  0.1434  3.1754
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.689e+00  4.855e-01   7.599 3.57e-13 ***
## b_migr11     5.446e-03  3.879e-03   1.404  0.16130
## pop11       -8.656e-07  2.793e-07  -3.099  0.00212 **
## urate2011    4.016e-02  9.320e-03   4.309 2.20e-05 ***
## medianage   -6.305e-02  1.027e-02  -6.138 2.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
##
## Residual standard error: 0.4774 on 310 degrees of
freedom
##   (105 observations deleted due to missingness)
## Multiple R-squared:  0.3468, Adjusted R-squared:
0.3383
```

Migration Effect goes away if we control for population unemployment rate & median age

Is it always a good idea to control for those variables?

Can u think of an alternative strategy?

An Alternative Strategy



An Alternative Strategy

An alternative strategy: Unemployment in 2004 can’t be affected by the surge in migration after 2004

```
summary(lm(crimesPc~b_migr11+pop11+medianage+urate2004,df %>% filter(crimesPc<15)))
```

```
##
## Call:
## lm(formula = crimesPc ~ b_migr11 + pop11 + medianage + urate2004,
##     data = df %>% filter(crimesPc < 15))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9334 -0.3021 -0.0885  0.1659  3.1744
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.037e+00  5.105e-01   7.907 4.44e-14 ***
## b_migr11     2.124e-03  4.075e-03   0.521  0.60257
## pop11       -8.958e-07  2.911e-07  -3.077  0.00227 **
## medianage   -6.787e-02  1.086e-02  -6.252 1.31e-09 ***
## urate2004    4.623e-02  1.825e-02   2.534  0.01178 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5011 on 316 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.3226, Adjusted R-squared:  0.314
## F-statistic: 37.62 on 4 and 316 DF,  p-value: < 2.2e-16
```

Migration coefficient:
- Still not significant
- Value has become slightly lower

Imperfect Multi-collinearity

Explanatory variables are closely but no perfectly correlated

Consequences:

- We can estimate all coefficients
- Variance of estimates might be high i.e. estimates could be quite far off from true value.
- However: estimates will be unbiased (if x not correlated with ϵ)

i.e. it's hard – but no impossible – for the OLS algorithm to distinguish between the separate effects for all the variables

So what's the problem?

- Possibly none
- Sometimes it won't be possible to reliably identify all desired effects
- You might think something doesn't matter when it does.

Imperfect Multi-collinearity: An example

```
evenmore=read.csv( "https://www.dropbox.com/s/pwotro2ghawkppg/foreign_evenmore.csv?dl=1")
df=df%>% inner_join(evenmore,by="area")



rr=lm(crimesPc~b_migr11+urate2011+
            pop11+
            shxage0t17+
            shxage18t29+shxage30t44+shxage45t64+meanage,df  %>% filter(crimesPc<150))
rr%>% summary()
```

```
##
## Call:
## lm(formula = crimesPc ~ b_migr11 + urate2011 + pop11 + shxage0t17 +
##     shxage18t29 + shxage30t44 + shxage45t64 + meanage, data = df %>%
##     filter(crimesPc < 150))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9592 -0.2153 -0.0735  0.1329  3.1625
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.089e+00  3.394e+01   0.238  0.81180
## b_migr11     1.646e-04  5.270e-03   0.031  0.97510
## urate2011    3.746e-02  9.443e-03   3.967 9.07e-05 ***
## pop11       -8.774e-07  2.736e-07  -3.206  0.00149 **
## shxage0t17  -6.445e-02  3.035e-01  -0.212  0.83198
## shxage18t29 -5.900e-03  2.483e-01  -0.024  0.98106
## shxage30t44 -2.058e-02  1.833e-01  -0.112  0.91064
## shxage45t64 -8.662e-02  1.187e-01  -0.730  0.46614
## meanage     -6.790e-02  4.269e-01  -0.159  0.87372
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

We include a wider set of controls for age bands; e.g. shxage017 reports the share of 0-17 year olds in percent.

None of the age variables is significant. Does this mean the age of the population is not important in explaining crime?

Age mattered before. The reason it doesn't matter now is because of collinearity

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.037e+00  5.105e-01   7.907 4.44e-14 ***
## b_migr11     2.124e-03  4.075e-03   0.521  0.60257
## pop11       -8.958e-07  2.911e-07  -3.077  0.00227 **
## medianage   -6.787e-02  1.086e-02  -6.252 1.31e-09 ***
## urate2004    4.623e-02  1.825e-02   2.534  0.01178 *
```

Imperfect Multi-collinearity: An example

```r
cor(df %>% select(shxage0t17,
                  shxage18t29,
                  shxage30t44,
                  shxage45t64,meanage),use="complete.obs")
```

```
##             shxage0t17 shxage18t29 shxage30t44 shxage45t64    meanage
## shxage0t17   1.00000000  0.01257122   0.3281674  -0.2878871 -0.5427118
## shxage18t29  0.01257122  1.00000000   0.5810169  -0.9006728 -0.8061182
## shxage30t44  0.32816735  0.58101695   1.0000000  -0.7408842 -0.8229079
## shxage45t64 -0.28788711 -0.90067279  -0.7408842   1.0000000  0.8938519
## meanage     -0.54271181 -0.80611820  -0.8229079   0.8938519  1.0000000
```

- We see that that some the age variables are highly correlated
- What matters is if a lot ofthe variation of an x variable is accounted for by a linear combination of all other x variables.
- We can examine this by looking at R2 in regressions of the following kind:

$$X_1 = \gamma_{11} + \gamma_{12} X_2 + \gamma_{13} X_3 + \cdots + u$$

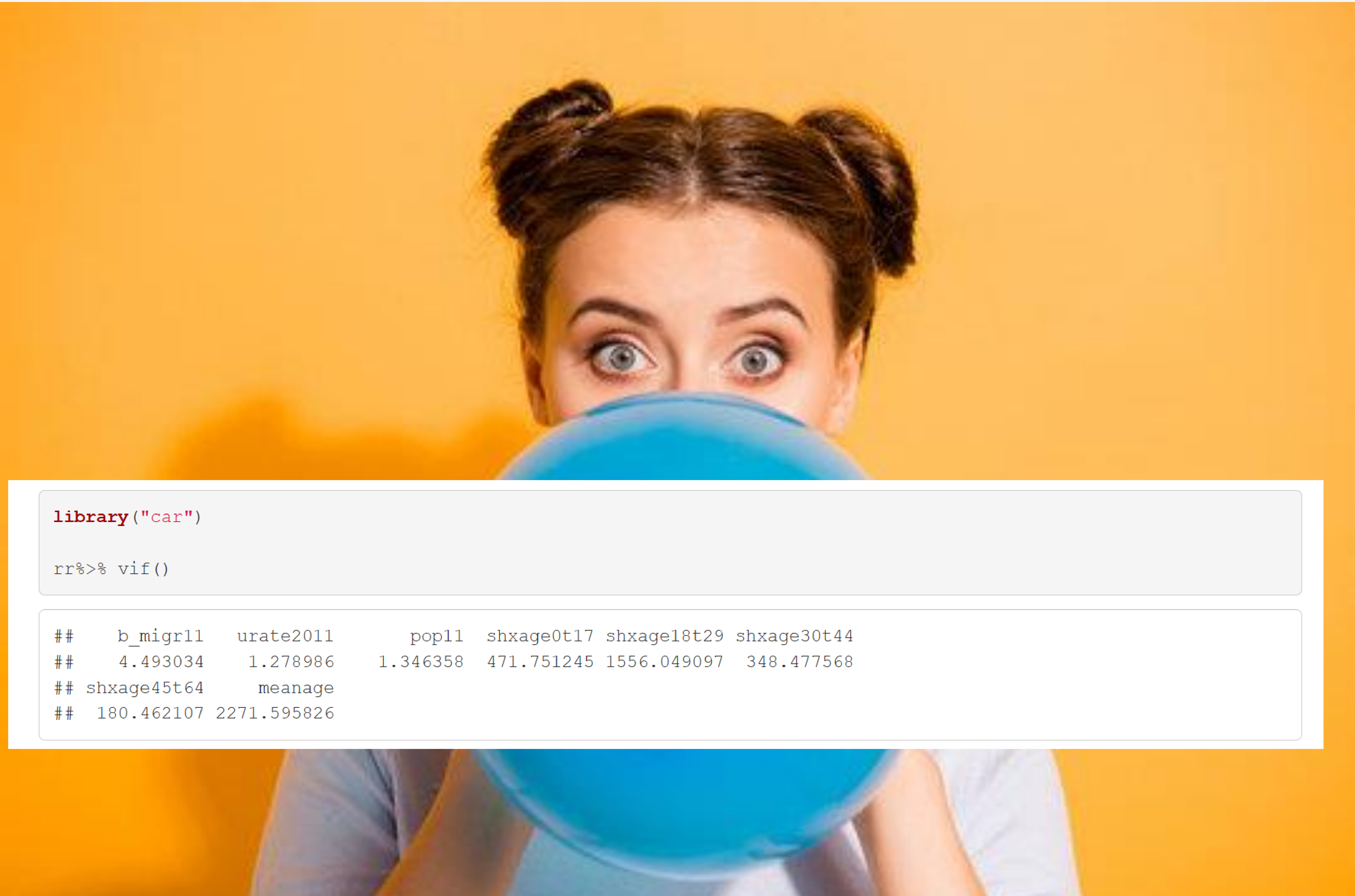
- i.e. we regress the explanatory variables on each other and compute R2 each time
- The Variance Inflation Factor (VIF) is an index that informs us about this by computing for every x-variable:

$$VIF = \frac{1}{1 - R^2}$$

VIF=1 $\rightarrow$ R2=0 $\rightarrow$ no problem

VIF>5 $\rightarrow$ R2>80% $\rightarrow$ maybe an issue

The Variance inflation factor in practice



```r
library("car")

rr%>% vif()
```

```
##    b_migr11   urate2011      pop11   shxage0t17 shxage18t29 shxage30t44
##    4.493034    1.278986    1.346358   471.751245 1556.049097  348.477568
## shxage45t64     meanage
##  180.462107 2271.595826
```

Joint Hypothesis Test

Testing multiple restrictions at once; e.g. does age really not matter in the regression above?

```
library("car")
linearHypothesis(rr,  c("shxage0t17 =0" ,
                        "shxage18t29=0",
                        "shxage30t44=0",
                        "shxage45t64=0",
                        "meanage=0"
                        ) )
```

F-test….but it's enough to look at the P value

```
## Linear hypothesis test
##
## Hypothesis:
## shxage0t17 = 0
## shxage18t29 = 0
## shxage30t44 = 0
## shxage45t64 = 0
## meanage = 0
##
## Model 1: restricted model
## Model 2: crimesPc ~ b_migr11 + urate2011 [illegible]11 + shxage0t17 + shxage18t29 +
##     shxage30t44 + shxage45t64 + meanage
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    311 79.228
## 2    306 66.011  5    13.217 12.254 7.689e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F-tests

Unrestricted Model	Restricted Model
$Y = \beta_{ux}X + \beta_{u1}AGE_1 + \beta_{u2}AGE_2 + \epsilon_u$	$Y = \beta_{rx}X + 0 \times AGE_1 + 0 \times AGE_2 + \epsilon_r$

2 parameters are restricted to be 0

We can compute an F-statistic as

$$F = \frac{\frac{RSS_r - RSS_u}{p_u - p_r}}{\frac{RSS_u}{n - p_u}}$$

- How much more error do we get when restricting the model.
- If it's a lot (F big) then the restriction should be rejected
- How do we know when F is big?
- Somebody worked out how F is distributed (turns out his name was Fisher)

where $RSS_r = \sum_i \hat{\epsilon}_{ri}^2$

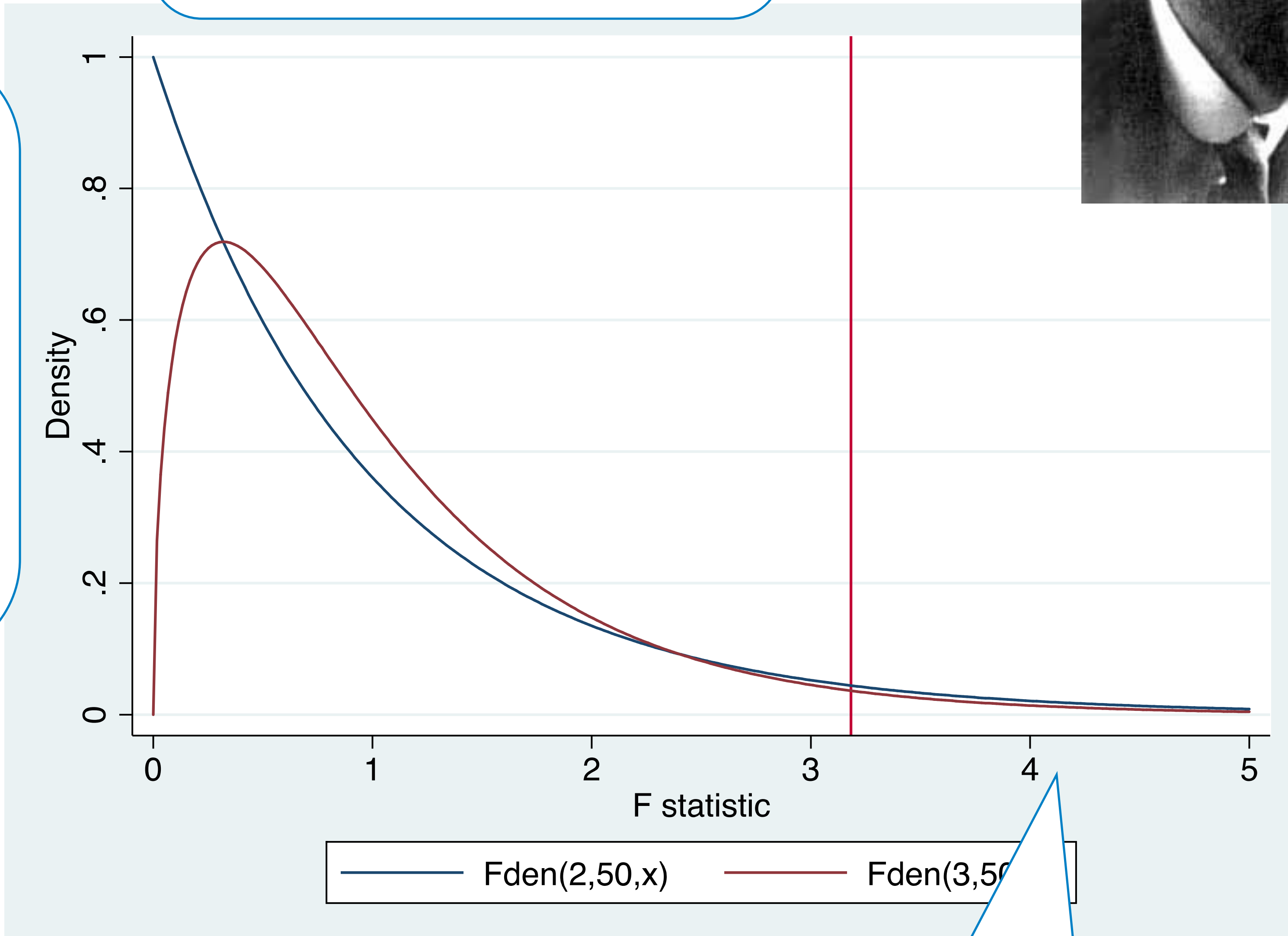
F distribution

```
> qf(0.95, 3, 50)
[1] 2.790008
```

F distribution has two arguments

1. Number of restrictions
2. Degrees of freedom unrestricted model

After R.A. Fisher (1890-1962)
(did not work for Guinness)



- If each the F statistic is large it means that some or all of the hypothesis jointly tested are probably not true

Find critical value by equating this area to significance level; 5%

Takeaways

- We can easily include further variables in a regression
- There are two reasons we might want to do that
  1. To deal with endogeneity
  2. We are interested in several variables at the same time
- Be careful about the causal relationships between explanatory variables
- There might be collinearity, which might imply that we cannot (precisely) distinguish between the effects of several explanatory variables.
- With several explanatory variables we might want to test several hypothesis combined.
- We can use an F-test for that.