Intro to Causal Inference in Econometrics

Lecture 2

Edward Vytlacil

Yale University, Department of Economics

December 18, 2024

Overview for this lecture:

- **Agenda**
  1. Simple Linear Regression
  2. Omitted Variable Bias
  3. Multivariate Linear Regression.
- **Application:**
  - Wage Regression, The Mincer Model

Part I: Simple Linear Regression Models

Simple Linear Regression

Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 X + e. \tag{1}$$

where

- Y is an observed random variable, called the *outcome variable*, the *left-hand side variable*, the *dependent variable*, or the *regressand*,
- X is an observed random variable, called the *covariate* , the *right-hand side variable*, the *independent variable*, or the *regressor*,
- e is an unobserved random variable, called the *error term* ,
- β_0 is an unknown constant, called the regression *intercept* ,
- β_1 is an unknown constant, called the regression *slope* .

Simple Linear Regression

Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 X + e. \tag{1}$$

What is the meaning of (1)? How to interpret?

Simple Linear Regression

Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 X + e. \tag{1}$$

What is the meaning of (1)? How to interpret?

Two interpretations of (1):

1. Best Linear Predictor: BLP of Y given X is $\beta_0 + \beta_1 X$.
2. Causal or structural model.

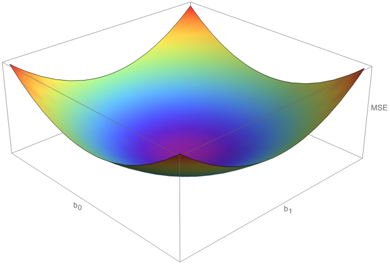
Interpretation 1: Linear Regression Model as BLP

For any (b_0, b_1), define

$$\text{MSE}(b_0, b_1) \equiv E[(Y - b_0 - b_1 X)^2],$$

and define

$$(\beta_0, \beta_1) \equiv \underset{(b_0, b_1)}{\text{argmin}}\ \text{MSE}(b_0, b_1). \qquad (2)$$



- $\text{MSE}(b_0, b_1)$ is the *mean squared prediction error* if using $b_0 + b_1 X$ to predict Y.
- $\beta_0 + \beta_1 X$ is called the *Best Linear Predictor* of Y given X, also called the *Linear Projection* of Y on X.
- β_0 and β_1 are called the *Linear Projection Coefficients*

Interpretation 1: Linear Regression Model as BLP

Define

$$e \equiv Y - \beta_0 - \beta_1 X, \tag{3}$$

$$\Rightarrow Y = \beta_0 - \beta_1 X + e, \tag{4}$$

with (β_0, β_1) defined by (2).

- e is defined as a prediction error, it has no economic or causal interpretation per se.
- Equation (4) is defined through BLP, no economic or causal interpretation per se.

Interpretation 1: Solving for β_0, β_1

$$(\beta_0, \beta_1) \equiv \underset{(b_0, b_1)}{\operatorname{argmin}} E\left[\left(Y - b_0 - b_1 X\right)^2\right], \tag{2}$$

$$e \equiv Y - \beta_0 - \beta_1 X. \tag{3}$$

Minimization problem (2) has FOC:

$$\mathbb{E}[Y - \beta_0 - \beta_1 X] = 0, \tag{5}$$

$$\mathbb{E}[(Y - \beta_0 - \beta_1 X])X] = 0, \tag{6}$$

which can be solved for β_0, β_1:

$$\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X], \tag{7}$$

$$\beta_1 = \frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(X)}. \tag{8}$$

Interpretation 1: FOC imply $\mathbb{E}[e] = \mathbb{E}[e\,X] = 0$.

$$(\beta_0, \beta_1) \equiv \underset{(b_0,b_1)}{\text{argmin}}\, E\big[(Y - b_0 - b_1X)^2\big], \tag{2}$$

$$e \equiv Y - \beta_0 - \beta_1X. \tag{3}$$

Minimization problem (2) has FOC:

$$\mathbb{E}[Y - \beta_0 - \beta_1X] = 0, \tag{5}$$

$$\mathbb{E}[(Y - \beta_0 - \beta_1X])\,X] = 0. \tag{6}$$

We can use (3) to rewrite FOC as

$$\mathbb{E}[e] = \mathbb{E}[e\,X] = 0.$$

Equivalent Ways to State Interpretation 1

$$\left(\beta_0, \beta_1\right) \equiv \underset{(b_0, b_1)}{\text{argmin}}\, E\left[\left(Y - b_0 - b_1 X\right)^2\right], \tag{2}$$

$$e \equiv Y - \beta_0 - \beta_1 X. \tag{3}$$

Equations (2)-(3) imply

$$Y = \beta_0 + \beta_1 X + e, \tag{9}$$

$$\mathbb{E}[e] = \mathbb{E}[eX] = 0, \tag{10}$$

and equations (9)-(10) imply (2)-(3) .

Interpretation 2: Linear Regression Model as Causal/Structural Model

For example, consider potential outcome notation with binary treatment:

$$Y = Y_0 + X(Y_1 - Y_0) = \begin{cases} Y_1 & \text{if } X = 1, \\ Y_0 & \text{if } X = 0. \end{cases} \tag{11}$$

where Y is observed outcome, X is binary treatment, and (Y_0, Y_1) are potential outcomes.

We can rewrite (11) as

$$Y = \beta_0 + \beta_1 X + e$$

where

$$\begin{aligned} \beta_0 &= \mathbb{E}[Y_0], \\ \beta_1 &= \mathbb{E}[Y_1 - Y_0 \mid X = 1], \\ e &= (Y_0 - \mathbb{E}[Y_0]) + ((Y_1 - Y_0) - \mathbb{E}[Y_1 - Y_0 \mid X = 1])\, X. \end{aligned}$$

Interpretation 2: Linear Regression Model as Causal/Structural Model

$$Y = \beta_0 + \beta_1 X + e$$

where

$$\begin{aligned}
\beta_0 &= \mathbb{E}[Y_0], \\
\beta_1 &= \mathbb{E}[Y_1 - Y_0 \mid X = 1], \\
e &= (Y_0 - \mathbb{E}[Y_0]) + ((Y_1 - Y_0) - \mathbb{E}[Y_1 - Y_0 \mid X = 1])\, X.
\end{aligned}$$

Here:

- Model is not defined as a predictive model.
- Interpretation of β_1 is as a causal parameter.
- e is not defined as prediction error, has interpretation from underlying causal model.
- $\mathbb{E}[e] = \mathbb{E}[Xe] = 0$ if $\mathbb{E}[Y_0 \mid X = 1] = \mathbb{E}[Y_0]$. In other words, if no selection bias.

Part II: Ordinary Least Squares for SLR

OLS Estimator for Simple Linear Regression

- The OLS estimators of β_0 and β_1 are defined as the values for b_0 and b_1 that minimize the sum of squared prediction errors over all observations

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(b_0, b_1)}{\operatorname{argmin}} \sum_{i=1}^{N} (Y_i - b_0 - b_1 X_i)^2.$$

- If we then *define* $\hat{e}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$, then the FOC for the above minimization problem can be expressed as:

$$\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} \hat{e}_i X_i = 0.$$

OLS Estimator for Simple Linear Regression

- The FOC for the OLS minimization problem can be expressed as:

$$\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} \hat{e}_i X_i = 0.$$

where $\hat{e}_i \equiv Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$

- Let s_X^2 denote sample variance of X_i, and s_{XY} denote sample covariance of (X_i, Y_i). Then, assuming $s_X^2 > 0$, can solve above FOC for $\left(\hat{\beta}_0, \hat{\beta}_1\right)$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2},$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}.$$

Application: Returns to Education – Estimating the Univariate Model

Consider "log-linear" regression

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{educ}_i + e_i,$$

where educ_i is years of schooling.

- Linear model with $Y_i = \ln \text{Wage}_i$, but non-linear model for wages.
- Interpretation of β_1?
  - Observations with 1 more year of school have β_1 higher log wage.
  - Approximately:
    Observations with 1 more year of school have $\beta_1 \cdot 100$ percent higher wage.

Application: Returns to Education – Estimating the Univariate Model

Consider “log-linear“ regression

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{educ}_i + e_i,$$

where educ_i is years of schooling.

```
> lm(logwage ~ education  , data = NLSYM)

Call:
lm(formula = logwage ~ education, data = NLSYM)

Coefficients:
(Intercept)    education
    0.96571      0.05209
>
> ggplot(data=NLSYM, aes(x= education, y=logwage)) +
+   geom_point(shape=1) +ggtitle("Log Wage Regression,NLSYM")+
+   theme_bw() + xlab("Education") + ylab("Log Wage") +
+   geom_smooth(method = lm, se = FALSE)
```

Returns to Education – Estimating the Univariate Model

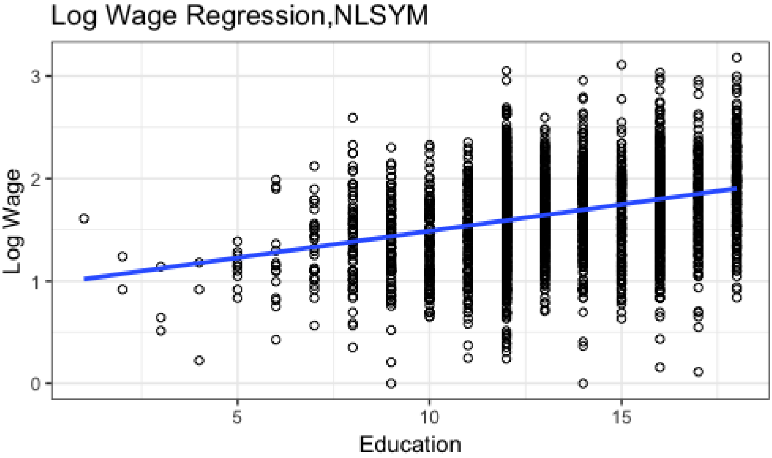


Figure 3.1: Linear regression of log wage on years of schooling.
Data Source: NLSYM

Simple Linear Regression

In general:

- $Y = \beta_0 + \beta_1 X + e$, with $\mathbb{E}[e] = \mathbb{E}[e\,X] = 0$ implies

$$\beta_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

$$\beta_0 = [Y] - \beta_1 \mathbb{E}[X]$$

- OLS regression:

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

Simple Linear Regression with Dummy Covariate

If X is binary:

- $Y = \beta_0 + \beta_1 X + e$, with $\mathbb{E}[e] = \mathbb{E}[e\, X] = 0$ implies

$$\begin{aligned}
\beta_1 &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \\
&= \mathbb{E}[Y_i \mid X_i = 1] - \mathbb{E}[Y_i \mid X_i = 0] \\
\beta_0 &= [Y] - \beta_1 \mathbb{E}[X] \\
&= \mathbb{E}[Y_i \mid X_i = 0],
\end{aligned}$$

- OLS regression:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{s_{XY}}{s_X^2} \\
&= \overline{Y}_1 - \overline{Y}_0 \\
\hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1 \overline{X} \\
&= \overline{Y}_0 ,
\end{aligned}$$

where $\overline{Y}_x$ denotes the sample mean of Y_i among those with $X_i = x$.

Application: Effect of PROGRESA – Estimating the Univariate Model

Consider

$$\text{SchoolEnroll}_i = \beta_0 + \beta_1 \text{Treat}_i + e_i,$$

where Treat_i is a dummy variable for being treated.

```
> mean.enroll <- with(dfPost,tapply(school,treat,mean))
> mean.enroll
        0         1
0.7456201 0.7859343
> mean.enroll[2]-mean.enroll[1]
         1
0.04031423
> lm(school ~ treat, data =dfPost)

Call:
lm(formula = school ~ treat, data = dfPost)

Coefficients:
(Intercept)        treat
    0.74562      0.04031
```

Part III: Omitted Variable Bias

Linear Regression and Structural Model: Wage Equation

Suppose a labor economist has a hedonic model of wages with

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i, \qquad E[e_i] = E[e_i X_{1i}] = E[e_i X_{2i}] = 0,$$

where:

- Y_i denotes wages,
- X_{1i} is a dummy variable for being African-American,
- X_{2i} years of education.

The parameter of interest is β_1, measuring taste-based discrimination.

What if we regress Y_i on X_{1i} alone by simple linear regression, omitting X_{2i}?

OLS with Omitted Variables

$$
\begin{aligned}
\hat{\beta}_1 = \frac{s_{X_1 Y}}{s^2_{X_1}} \xrightarrow{P} \frac{Cov(Y_i, X_{1i})}{Var(X_{1i})} \\
&= \frac{Cov(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i, X_{1i})}{Var(X_{1i})} \\
&= \frac{Cov(\beta_0, X_{1i})}{Var(X_{1i})} + \frac{Cov(\beta_1 X_{1i}, X_{1i})}{Var(X_{1i})} + \frac{Cov(\beta_2 X_{2i}, X_i)}{Var(X_{1i})} + \frac{Cov(e_i, X_i)}{Var(X_{1i})} \\
&= \beta_1 + \beta_2 \frac{Cov(X_{2i}, X_{1i})}{Var(X_{1i})} + \frac{Cov(e_i, X_i)}{Var(X_{1i})} \\
&= \beta_1 + \underbrace{\beta_2 \frac{Cov(X_{2i}, X_{1i})}{Var(X_{1i})}}_{\text{Omitted Variable Bias}}.
\end{aligned}
$$

Signing Omitted Variable Bias

$$\hat{\beta}_1 \xrightarrow{P} \beta_1 + \underbrace{\beta_2 \frac{Cov(X_{2i}, X_{1i})}{Var(X_{1i})}}_{\text{Omitted Variable Bias}}.$$

- $\frac{Cov(X_{2i}, X_{1i})}{Var(X_{1i})}$ is the coefficient of a projection of X_2 on X_1.
- In special case of binary X_{1i},
  - $\frac{Cov(X_{2i}, X_{1i})}{Var(X_{1i})} = E[X_{2i} \mid X_{1i} = 1] - E[X_{2i} \mid X_{1i} = 0]$,
  - Sign $\{Cov(X_{2i}, X_{1i})\} =$ Sign $\{E[X_{2i} \mid X_{1i} = 1] - E[X_{2i} \mid X_{1i} = 0]\}$.
  - Connection to selection bias for treatment effects?

Signing Omitted Variable Bias

$$\hat{\beta}_1 \xrightarrow{P} \beta_1 + \underbrace{\beta_2 \frac{Cov(X_{2i}, X_{1i})}{Var(X_{1i})}}_{\text{Omitted Variable Bias}}.$$

- Omitted variable bias is zero if:
    - $Cov(X_{2i}, X_{1i}) = 0$ (X_1 and X_2 are uncorrelated), or
    - $\beta_2 = 0$.
- Otherwise:
    - If $\text{Sign}\{\beta_2\} = \text{Sign}\{\text{Cov}(X_2, X_1)\}$, the bias is positive.
    - If $\text{Sign}\{\beta_2\} \neq \text{Sign}\{\text{Cov}(X_2, X_1)\}$, the bias is negative.

Part IV: Multivariate Linear Regression Models

Include more Covariates?

- For prediction, adding more covariates can improve precision of predictions.
- In causal or structural models, may believe that need other regressors in model for moment conditions to hold.
  - e.g., Include “measured confounders”, may believe that treatment receipt is “as-if” by randomized experiment conditional on other covariates.
  - recall omitted variable bias.

Multivariate Linear Regression Models

- Multivariate Linear Regression Models are of the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + e_i, \tag{12}$$

  with

$$E[e_i] = E[e_i X_{1i}] = \ldots = E[e_i X_{Ki}] = 0. \tag{13}$$

- Equations 12 and 13 equivalent to

$$(\beta_0, \beta_1, \ldots, \beta_K) = \underset{(b_0, b_1, \ldots, b_K)}{\operatorname{argmin}} E\left[\left(Y - b_0 - b_1 X_1 - \ldots - b_K X_K\right)^2\right]. \tag{14}$$

- If define model by linear projection (as solution to 14), then moment conditions (13) holds by FOC for equation 14.

Multivariate Linear Regression Models

- Multivariate Linear Regression Models are of the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + e_i, \tag{12}$$

with

$$E[e_i] = E[e_i X_{1i}] = \ldots = E[e_i X_{Ki}] = 0. \tag{13}$$

- Equations 12 and 13 equivalent to

$$(\beta_0, \beta_1, \ldots, \beta_K) = \underset{(b_0, b_1, \ldots, b_K)}{\operatorname{argmin}} E\left[\left(Y - b_0 - b_1 X_1 - \ldots - b_K X_K\right)^2\right]. \tag{14}$$

- If we are defining the regression model as a causal/structural model, then (13) need not hold, and $(\beta_0, \beta_1, \ldots, \beta_K)$ need not be solution to equation 14.

Marginal Effects

- We call the change in predicted Y induced by a *ceterus paribus* marginal increase in X_j the **marginal effect** of changing X_j on Y.
  - May or may not be causal
- For binary X_j this is understood to mean changing X_j from 0 to 1.
- What if there is a functional relationship between regressors, e.g., $X_2 = X_1^2$?
  - $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + e \Rightarrow$ marginal effect of X_1 equals $\beta_1 + 2\beta_2 X_1$.
  - Are marginal effects always constant in linear models? No!

Part V: Ordinary Least Squares for MLR

Multivariate OLS

- The OLS estimator of $(\beta_0, \ldots, \beta_K)$ are defined as the values for $(b_0, \ldots, b_K)$ that minimize the sum of squared prediction errors over all observations:

$$(\hat{\beta}_0, \hat{\beta}_1, \ldots., \hat{\beta}_K) = \underset{b_0,\ldots,b_K}{\arg\min} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i} - \ldots - b_K X_{Ki})^2 .$$

- If we then *define*

$$\hat{e}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_K X_{Ki},$$

then the FOC for the above minimization problem can be expressed as:

$$\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} \hat{e}_i X_{1i} = \sum_{i=1}^{n} \hat{e}_i X_{2i} = \ldots = \sum_{i=1}^{n} \hat{e}_i X_{Ki} = 0.$$

Multivariate OLS

- $(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_K)$ solve

$$\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} \hat{e}_i X_{1i} = \sum_{i=1}^{n} \hat{e}_i X_{2i} = \ldots = \sum_{i=1}^{n} \hat{e}_i X_{Ki} = 0,$$

  where $\hat{e}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_K X_{Ki}$.

- In order to have a unique solution to the OLS minimization problem, we need one additional assumption: No *perfect multicollinearity*.

- OLS estimator has desirable properties for estimation of $(\beta_0, \beta_1, ..., \beta_K)$ as long as no perfect multicollinearity and moment equation (13) holds.

Multicollinearity

- Two or more regressors are said to exhibit **perfect multicollinearity**, if one of the regressors is a perfect linear function of the others.
- Can not estimate OLS on a set of $X's$ that include collinear variables.
- Intuitively, multicollinearity is a problem because it is not possible to disentangle the effects of two variables that always move together.
- Multicollinearity most often can be avoided by choosing the appropriate set of covariates. Will come back to this later in the context of the "Dummy-Variable Trap".

Terminology

- We call $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + e_i$, the ***population regression function***.

- We call the function $\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_K X_{Ki}$ the ***fitted regression function***.

- We call $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \ldots + \hat{\beta}_K X_{Ki}$ the ***fitted values***.

- We call $\hat{e}_i = Y_i - \hat{Y}_i$ the ***residuals***.

Inference (optional)

Under appropriate regularity conditions,

$$\sqrt{N}(\hat{\beta}_j - \beta_j) \xrightarrow{d} N(0, V_j)$$

Suppose have consistent estimator of V_j,

$$\hat{V}_j \xrightarrow{p} V_j.$$

- Consistent standard errors:

$$\text{s.e.}(\hat{\beta}_j) = \sqrt{\frac{\hat{V}_j}{N}}$$

- 95% Asymptotic Confidence Interval:

$$\left[\hat{\beta}_j - 1.96 \cdot \text{s.e.}(\hat{\beta}_j),\ \hat{\beta}_j + 1.96 \cdot \text{s.e.}(\hat{\beta}_j)\right].$$

Inference (optional)

Under appropriate regularity conditions,

$$\sqrt{N}\left(\hat{\beta}_j - \beta_j\right) \xrightarrow{d} N\left(0, V_j\right)$$

Suppose have consistent estimator of V_j,

$$\hat{V}_j \xrightarrow{p} V_j.$$

- To test null $H_0 : \beta_j = \beta_{j0}$, can define test statistic

$$T = \frac{\hat{\beta}_j - \beta_{j0}}{\text{s.e.}\left(\hat{\beta}_j\right)}$$

$$\Rightarrow T \xrightarrow{d} N\left(0, 1\right) \qquad \text{under } H_0$$

Thus reject null at at 5% level if

$$\left|T\right| \geq 1.96.$$

Inference (optional)

Under appropriate regularity conditions,

$$\sqrt{N}\left(\hat{\beta}_j - \beta_j\right) \xrightarrow{d} N\left(0, V_j\right)$$

Suppose have consistent estimator of V_j,

$$\hat{V}_j \xrightarrow{p} V_j.$$

- Default s.e. in **R** and most packages are only consistent under homoscedasticity, if $\mathbb{E}\left[e_i^2 \mid X_{1i}, \ldots, X_{Ki}\right]$ is a constant. Typically implausible in economics.
- Typically heteroscedastic-robust s.e. in economics.
- Often believe data is “clustered”, use s.e. that are both heterscedastic-robust and robust to dependence within a cluster.

Part VI : Application – Estimating the Returns to Education

Estimating the Returns to Education

- Economists have studied and tried to estimate the returns to education for more than 60 years
- Mincer (JPE, 1958): "Investment in Human Capital and Personal Income Distribution"
- Implies $\texttt{ln(wage)} = \beta_0 + \beta_1 \texttt{educ} + \beta_2 \texttt{exp} + \beta_3 \texttt{exp}^2 + e$
- Wages increase at a constant rate (%) in education
- Regressions of $\texttt{ln(wage)}$ on $\texttt{educ}$, $\texttt{exp}$ and $\texttt{exp}^2$ (and other covariates) are called *Mincer regressions*.
- Original Mincer regressions did not include ethnicity, but often additionally include dummy variable for being African American.

NLSYM

- Following regressions use data from National Longitudinal Survey of Young Men (NLSYM) that was used by Card (1995).
- Data is from 1976, when respondents were between 24 and 34 years old.
- True years of work experience are not recorded in data set, and thus define "potential experience":

$$\text{experience} = \text{Age} - \text{Education} - 6.$$

- See card.dta documentation for more details.

Returns to Education – Estimating the Univariate Model

- Let's run some regressions

```
> reg.1<-lm(logwage ~ education  , data = NLSYM)
> reg.2<- lm(logwage ~ experience  , data = NLSYM)
> reg.1

Call:
lm(formula = logwage ~ education, data = NLSYM)

Coefficients:
(Intercept)    education
     0.9657       0.0521

> reg.2

Call:
lm(formula = logwage ~ experience, data = NLSYM)

Coefficients:
(Intercept)   experience
    1.64481      0.00134
```

Returns to Education – Estimating Bivariate Model

Lets estimate the bivariate model:

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{experience}_i + e_i$$

```
> reg.3<- lm(logwage ~ education+experience   , data = NLSYM)
> reg.3

Call:
lm(formula = logwage ~ education + experience, data = NLSYM)

Coefficients:
(Intercept)    education   experience
     0.0609       0.0932       0.0407
>
> with(NLSYM,cor(education,experience))
[1] -0.653
```

- Why, in this sample, are education and experience so strongly negatively correlated?

Fitted Regression Function: $\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \hat{\beta}_2 \text{experience}_i$

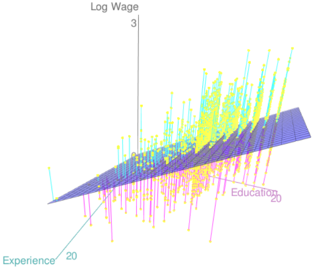


Figure 3.2: Linear regression of log wage on years of schooling and years of experience, using NLSYM.

Table of Results

- Can use `stargazer` to produce table of results.
- Problem: default s.e., p-values based on default s.e., require homoscedasticity for validity.
- Alternatively, can use `coeftest` with `vcov=vcovHC` for heteroscedastic-robust s.e.:

```
> library(sandwich)
> library(lmtest)
> reg.test.1<-coeftest(reg.1,vcov = vcovHC)
> reg.test.2<-coeftest(reg.2,vcov = vcovHC)
> reg.test.3<-coeftest(reg.3,vcov = vcovHC)
> stargazer(reg.1,reg.2,reg.3,
+           se= list(reg.test.1[,2], reg.test.2[,2],reg.test.3[,2]),
+           p=list(reg.test.1[,4], reg.test.2[,4],reg.test.3[,4]),
+           dep.var.labels="Enrollment", intercept.bottom = FALSE,
+           keep.stat=c("n","rsq"),
+           notes.append = FALSE, notes.align = "l", notes = "Reporting
    heteroscedastic-robust standard errors in parenthesis.")
```

Returns to Education – Estimating Bivariate Model

Table 3.1: Log Wage Regressions, Using NLSYM Data

	Dependent variable:		
	Log Wages		
Constant	0.966***	1.645***	0.061
	(0.039)	(0.019)	(0.065)
education	0.052***		0.093***
	(0.003)		(0.004)
experience		0.001	0.041***
		(0.002)	(0.002)
Observations	3,010	3,010	3,010
R^2	0.099	0.0002	0.181
Note:	Reporting heteroscedastic-robust standard errors in parenthesis.		

- How do estimated coefficients compare?
- Intuition?

Returns to Education – What if add age?

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{experience}_i + \beta_3 \text{age}_i + e_i$$

```
>  lm(logwage ~ education+experience +age  , data = NLSYM)

Call:
lm(formula = logwage ~ education + experience + age, data =
    NLSYM)

Coefficients:
(Intercept)    education   experience          age
     0.0609       0.0932       0.0407           NA
```

- Why the NA when we add age?
  - Perfect multicolinearity!
  - Recall experience is potential experience,

$$\text{experience} = \text{Age} - \text{Education} - 6.$$

Returns to Education – Estimating Multivariate Model

Lets estimate the model with quadratic in experience:

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{experience}_i + \beta_3 \text{experience}_i^2 + e_i,$$

implying marginal effect of experience:

$$\beta_2 + 2\beta_3 \text{experience}_i.$$

```
> reg.4<- lm(logwage ~ education+experience + I(experience^2) ,
      data = NLSYM)
```

Returns to Education – Estimating Multivariate Model

Table 3.2: Log Wage Regressions, Using NSLYM Data

	Dependent variable:			
	Log Wages			
	(1)	(2)	(3)	(4)
Constant	0.966^{***}	1.645^{***}	0.061	-0.137^{*}
	(0.039)	(0.019)	(0.065)	(0.070)
education	0.052^{***}		0.093^{***}	0.093^{***}
	(0.003)		(0.004)	(0.004)
experience		0.001	0.041^{***}	0.090^{***}
		(0.002)	(0.002)	(0.007)
I(experience^2)				-0.002^{***}
				(0.0003)
Observations	3,010	3,010	3,010	3,010
R^2	0.099	0.0002	0.181	0.196

In model with quadratic in experience,
Estimated Marginal Effect of Experience: $\hat{\beta}_2 + 2\beta_3$experience.

Returns to Education – Estimating Multivariate Model

Table 3.2: Log Wage Regressions, Using NSLYM Data

	Dependent variable:			
	Log Wages			
	(1)	(2)	(3)	(4)
Constant	0.966^{***}	1.645^{***}	0.061	-0.137^{*}
	(0.039)	(0.019)	(0.065)	(0.070)
education	0.052^{***}		0.093^{***}	0.093^{***}
	(0.003)		(0.004)	(0.004)
experience		0.001	0.041^{***}	0.090^{***}
		(0.002)	(0.002)	(0.007)
I(experience^2)				-0.002^{***}
				(0.0003)
Observations	3,010	3,010	3,010	3,010
R^2	0.099	0.0002	0.181	0.196

In model with quadratic in experience,
Estimated Marginal Effect of Experience: $\hat{\hat{\beta}}_2 + 2\beta_3$experience.

If experience$=0$, estimated marginal effect: 0.09.

Returns to Education – Estimating Multivariate Model

Table 3.2: Log Wage Regressions, Using NSLYM Data

	Dependent variable:			
	Log Wages			
	(1)	(2)	(3)	(4)
Constant	0.966^{***}	1.645^{***}	0.061	-0.137^{*}
	(0.039)	(0.019)	(0.065)	(0.070)
education	0.052^{***}		0.093^{***}	0.093^{***}
	(0.003)		(0.004)	(0.004)
experience		0.001	0.041^{***}	0.090^{***}
		(0.002)	(0.002)	(0.007)
I(experience^2)				-0.002^{***}
				(0.0003)
Observations	3,010	3,010	3,010	3,010
R^2	0.099	0.0002	0.181	0.196

In model with quadratic in experience,
Estimated Marginal Effect of Experience: $\hat{\beta}_2 + 2\beta_3$experience.

If experience$=$ 10, estimated marginal effect: $0.09 + 2 \cdot (-0.002) \cdot 10 = 0.05$.

Returns to Education – Estimating Multivariate Model

Table 3.2: Log Wage Regressions, Using NSLYM Data

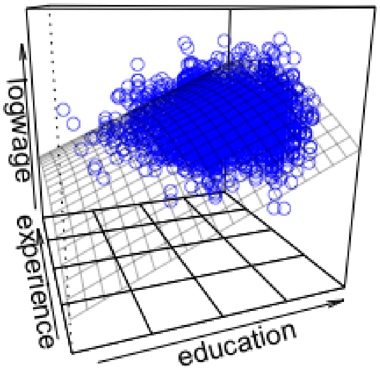
	Dependent variable:			
	Log Wages			
	(1)	(2)	(3)	(4)
Constant	0.966^{***}	1.645^{***}	0.061	-0.137^{*}
	(0.039)	(0.019)	(0.065)	(0.070)
education	0.052^{***}		0.093^{***}	0.093^{***}
	(0.003)		(0.004)	(0.004)
experience		0.001	0.041^{***}	0.090^{***}
		(0.002)	(0.002)	(0.007)
I(experience^2)				-0.002^{***}
				(0.0003)
Observations	3,010	3,010	3,010	3,010
R^2	0.099	0.0002	0.181	0.196

In model with quadratic in experience,
Estimated Marginal Effect of Experience: $\hat{\beta}_2 + 2\beta_3$experience.

If experience$=$ 20, estimated marginal effect: $0.09 + 2 \cdot (-0.002) \cdot 20 = 0.01$.

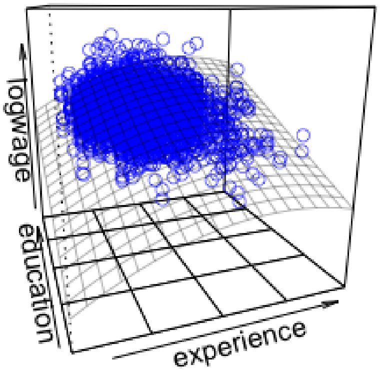
Fitted Regression Function:

$$\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \hat{\beta}_2 \text{experience}_i + \hat{\beta}_3 \text{experience}_i^2$$



Fitted Regression Function:

$$\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \hat{\beta}_2 \text{experience}_i + \hat{\beta}_3 \text{experience}_i^2$$



Returns to Education – Education and Ethnicity

- Linear Regression Models on Education and Ethnicity:

```

> reg.1<-lm(logwage ~ education  , data = NLSYM)
> reg.5<-lm(logwage ~ black  , data = NLSYM)
> reg.6<- lm(logwage ~ education + black  , data = NLSYM)
> with(NLSYM,cor(education,black))
[1] -0.2694
> with(NLSYM,tapply(education,black,mean))
    0     1
13.66 11.96
```

- Why not also add dummy variable for not being black?
  Dummy Variable Trap, would result in perfect multicolinearity.

Returns to Education – Estimating Bivariate Model

Table 3.3: Log Wage Regressions, Using NLSYM Data

	Dependent variable:		
	Log Wages		
	(1)	(2)	(3)
Constant	0.966***	1.731***	1.163***
	(0.039)	(0.009)	(0.040)
education	0.052***		0.042***
	(0.003)		(0.003)
black		−0.318***	−0.247***
		(0.018)	(0.018)
Observations	3,010	3,010	3,010
R^2	0.099	0.092	0.150
Note:	Reporting heteroscedastic-robust standard errors in parenthesis.		

- How does estimated coefficient compare?
- Intuition?

Returns to Education – Estimating Bivariate Model

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{Black}_i + e_i,$$

```
> reg.6<- lm(logwage ~ education + black  , data = NLSYM)
>
> equation.6w=function(x){coef(reg.6)[2]*x+coef(reg.6)[1]}
> equation.6b=function(x){coef(reg.6)[2]*x+coef(reg.6)[1]+coef(reg.6)[3]}
>
> ggplot(data=NLSYM, aes(x= education, y=logwage,color=as.factor(black)))+
+   geom_point(shape=1) +ggtitle("Log Wage Regression, NLSYM 1988")+ theme_bw() +
+   xlab("Education") + ylab("Log Wage")+
+   stat_function(fun=equation.6w,geom="line",color=scales::hue_pal()(2)[1])+
+   stat_function(fun=equation.6b,geom="line",color=scales::hue_pal()(2)[2])+
+   scale_color_hue(labels = c("Caucasian", "African Americans"))+
+   labs(colour="Ethnicity")
```

Fitted Regression Function: $\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \hat{\beta}_2 \text{Black}_i$

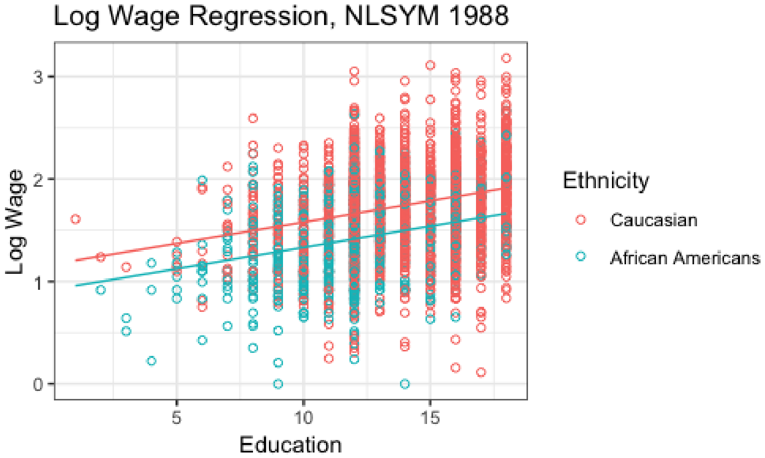


Figure 3.4: Linear regression of wage on years of schooling and dummy variable for being African-American.

Returns to Education – Estimating Model with Interactions, Education and Ethnicity

- Additive model

$$
\begin{aligned}
\ln \text{Wage}_i &= \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{Black}_i + e_i \\
&= \begin{cases} \beta_0 + \beta_1 \text{education}_i + e_i & \text{if not Black} \\ (\beta_0 + \beta_2) + \beta_1 \text{education}_i + e_i & \text{if Black} \end{cases}
\end{aligned}
$$

- Model with interactions:

$$
\begin{aligned}
\ln \text{Wage}_i &= \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{Black}_i + \beta_3 \text{Black}_i \cdot \text{education}_i + e_i \\
&= \begin{cases} \beta_0 + \beta_1 \text{education}_i + e_i & \text{if not Black} \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \text{education}_i + e_i & \text{if Black} \end{cases}
\end{aligned}
$$

- OLS regression on model with interactions equivalent to running regressions $\ln \text{Wage}_i = \beta_0 + \beta_1 \text{education}_i + e_i$ separately on black and non-black samples.

Returns to Education – Estimating Model with Interactions

```r
reg.7w <- lm(logwage ~ education, data = NLSYM[NLSYM$black==0,])
reg.7b <- lm(logwage ~ education, data = NLSYM[NLSYM$black==1,])
reg.7<- lm(logwage ~ education*black  , data = NLSYM)
```

Returns to Education – Estimating Model with Interactions

Table 3.4: Log Wage Regressions, Using NLSYM Data

	Dependent variable:					
	log wage					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.966***	1.731***	1.163***	1.245***	0.690***	1.245***
	(0.039)	(0.009)	(0.040)	(0.046)	(0.068)	(0.046)
education	0.052***		0.042***	0.036***	0.060***	0.036***
	(0.003)		(0.003)	(0.003)	(0.006)	(0.003)
black		−0.318***	−0.247***			−0.555***
		(0.018)	(0.018)			(0.082)
education:black						0.025***
						(0.006)
Sample:	Full	Full	Full	Non-Black	Black	Full
Observations	3,010	3,010	3,010	2,307	703	3,010
R^2	0.099	0.092	0.150	0.046	0.142	0.154

Note: Reporting heteroscedastic-robust standard errors in parenthesis.

Fitted Regression: $\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \hat{\beta}_2 \text{Black}_i + \hat{\beta}_3 \text{Black}_i \cdot \text{education}_i$

```
> reg.7

Call:
lm(formula = logwage ~ education * black, data = NLSYM)

Coefficients:
(Intercept)    education       black   education:black
     1.2453       0.0355     -0.5550            0.0249

 #fitted regression line for non-blacks:
> equation.7w=function(x){coef(reg.7)[2]*x+coef(reg.7)[1]}
 #fitted regression line for blacks:
> equation.7b=function(x){(coef(reg.7)[2]+coef(reg.7)[4])*x
     +coef(reg.7)[1]+coef(reg.7)[3]}
```

Fitted Regression: $\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \hat{\beta}_2 \text{Black}_i + \hat{\beta}_3 \text{Black}_i \cdot \text{education}_i$

```
> ggplot(data=NLSYM, aes(x= education, y=logwage,color=as.factor(black)))+
+   geom_point(shape=1) +
+   ggtitle("Log Wage Regression, With   Interaction, NLSYM Data")+
+   theme_bw() + xlab("Education") + ylab("Log Wage")+ylim(0,5) +
+   stat_function(fun=equation.7w,geom="line",color=scales::hue_pal()(2)[1])+
+   stat_function(fun=equation.7b,geom="line",color=scales::hue_pal()(2)[2])+
+   scale_color_hue(labels = c("Caucasian", "African Americans"))+
+   labs(colour="Ethnicity")
```

Returns to Education – Estimating the Model with Interactions

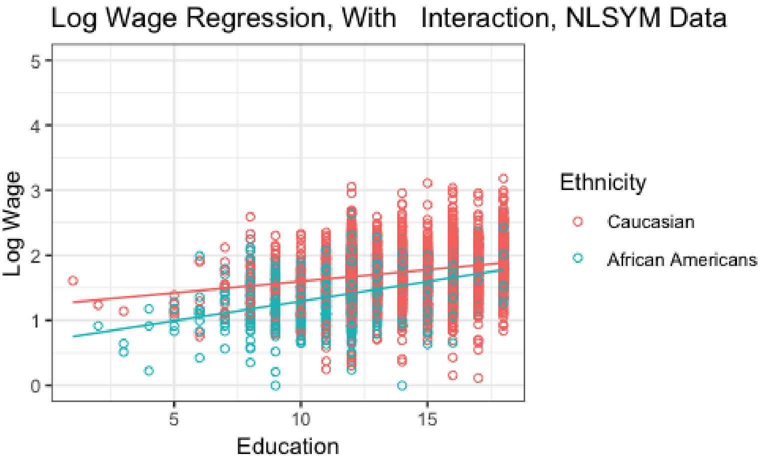


Figure 3.5: Linear regression of wage on years of schooling, dummy variable for being African-American, and interaction.

```r
> reg.8<- lm(logwage ~ education*black+experience
+               + I(experience^2)  , data = NLSYM)
> reg.9w <- lm(logwage ~ education+experience
+               + I(experience^2), data = NLSYM[NLSYM$black==0,])
> reg.9b <- lm(logwage ~ education+experience
+                 + I(experience^2), data = NLSYM[NLSYM$black==1,])
> reg.9<- lm(logwage ~ education*black+experience*black
+               + I(experience^2)*black  , data = NLSYM)
```

Table 3.5: Log Wage Regressions, Using NLSYM Data

	Dependent variable:				
	Log Wages				
	(1)	(2)	(3)	(4)	(5)
Constant	1.245***	0.369***	0.249***	0.208	0.249***
	(0.046)	(0.069)	(0.074)	(0.140)	(0.074)
education	0.036***	0.075***	0.080***	0.084***	0.080***
	(0.003)	(0.004)	(0.004)	(0.008)	(0.004)
black	−0.555***	−0.615***			−0.041
	(0.082)	(0.079)			(0.158)
experience		0.040***	0.045***	0.020***	0.045***
		(0.002)	(0.003)	(0.005)	(0.003)
education:black	0.025***	0.031***			0.004
	(0.006)	(0.006)			(0.009)
black:experience					−0.025***
					(0.006)
Sample:	*Full*	*Full*	*Non — Black*	*Black*	*Full*
Observations	3,010	3,010	2,307	703	3,010
R^2	0.154	0.233	0.159	0.164	0.238

Note: Reporting heteroscedastic-robust standard errors in parenthesis.

Fitted Regression Function:

$$\hat{\beta}_0 + \hat{\beta}_1 \text{education}_i + \beta_2 \text{experience}_i + \hat{\beta}_3 \text{Black}_i$$

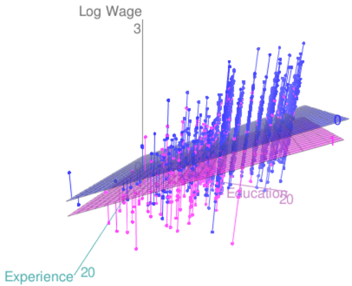


Figure 3.6: Linear regression of wage on years of schooling, years of experience, and dummy variable for being African-American.

Fitted Regression Function: Including Interactions with Being African-American

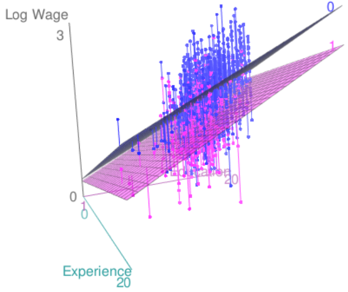


Figure 3.7: Linear regression of wage on years of schooling, years of experience, dummy variable for being African-American, and interactions with being African-American.

Additional Resources, Next Week

Additional resources for this lecture:

- code, data.
- Handouts:
  - Handout: Implementing OLS in **R**.
  - Handout: `stargazer`.