Review of OLS and the Simple Linear Regression Model (SLRM)

\[
Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, 2, ..., N
\]

Using OLS to estimate and test hypotheses regarding $\beta_0$ and $\beta_1$

Properties of OLS Estimators

→ Algebraic vs Statistical

→ Small Sample vs Large Sample (for this review, mostly small sample)

Application

→ Predictive vs Causal

Extension

→ OLS under heteroskedasticity, Weighted Least Squares
(For now) Let $\hat{\beta}_0$ and $\hat{\beta}_1$ represent estimators (not necessarily OLS) for $\beta_0$ and $\beta_1$

Fitted Values: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

Residuals: $\hat{e}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$, $i = 1, 2, ..., N.$

OLS: Choose $\hat{\beta}_0, \hat{\beta}_1$ to minimize $SSR = \sum_{i=1}^{N} \hat{e}_i^2 = \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$
Deriving the formulas:

OLS: Choose $\hat{\beta}_0, \hat{\beta}_1$ to minimize $SSR = \sum_{i=1}^{N} \hat{e}_i^2 = \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$

\[
\frac{\partial SSR}{\partial \hat{\beta}_0} \bigg|_{\hat{\beta}_0^{ols}, \hat{\beta}_1^{ols}} = -2 \sum_{i=1}^{N} (Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols} X_i) = 0
\]

\[
\frac{\partial SSR}{\partial \hat{\beta}_1} \bigg|_{\hat{\beta}_0^{ols}, \hat{\beta}_1^{ols}} = -2 \sum_{i=1}^{N} (Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols} X_i)X_i = 0
\]

1st eq $\div N \Rightarrow \hat{\beta}_0^{ols} = \bar{Y} - \hat{\beta}_1^{ols} \bar{X}$

Sub into 2nd eq $\Rightarrow \sum_{i=1}^{N} ((Y_i - \bar{Y}) - \hat{\beta}_1^{ols} (X_i - \bar{X}))X_i = 0$

$\Rightarrow \hat{\beta}_1^{ols} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})X_i}{\sum_{i=1}^{N} (X_i - \bar{X})X_i} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \frac{s.cov(Y_i, X_i)}{s.var(X_i)}$
Remarks:

→ the S.O.C. for minimization hold (see notes)

→ solution requires $\sum_{i=1}^{N} (X_i - \bar{X})^2 \neq 0$

→ $\hat{\beta}_0^{ols} = \bar{Y} - \hat{\beta}_1^{ols} \bar{X} \Rightarrow \bar{Y} = \hat{\beta}_0^{ols} + \hat{\beta}_1^{ols} \bar{X}$ Est. reg. line passes through $(\bar{Y}, \bar{X})$

FOC can be written

$$\sum_{i=1}^{N} \hat{\varepsilon}_i^{ols} = 0 \text{ (implication: sample mean of } \hat{\varepsilon}_i^{ols} \text{ is zero)}$$

$$\sum_{i=1}^{N} \hat{\varepsilon}_i^{ols} X_i = 0 \text{ (because } \overline{\hat{\varepsilon}_i^{ols}} \text{, this says that } \text{s.cov}(\hat{\varepsilon}_i^{ols}, X_i) = 0)$$

where $\hat{\varepsilon}_i^{ols} = Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols} X_i$
For rest of the notes, drop ‘ols’ superscript

→ $\hat{\beta}_0, \hat{\beta}_1$ will refer to OLS estimators

→ $\hat{Y}_i$ will refer to OLS fitted values

→ $\hat{\varepsilon}_i$ will refer to OLS residuals

Beware of context:

- We can have other types of estimators, fitted values, residuals
- E.g. later on we have Weighted Least Squares (WLS) estimators, fitted values, residuals
Important Decomposition (for OLS on SLRM with intercept)

\[
\sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \sum_{t=1}^{N} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{N} \hat{\epsilon}_i^2 
\]

\[
SST = SSE + SSR
\]

\[
\frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \frac{1}{N-1} \sum_{t=1}^{N} (\hat{Y}_i - \bar{Y})^2 + \frac{1}{N-1} \sum_{i=1}^{N} \hat{\epsilon}_i^2
\]

This is a (sample) variance decomposition

Note that \( \bar{Y} = \hat{Y} \) (because \( Y_i = \hat{Y}_i + \hat{\epsilon}_i \) and \( \bar{\epsilon} = 0 \))
Proof:

\[ Y_i = \hat{Y}_i + \hat{\epsilon}_i \]

\[ Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + \hat{\epsilon}_i \]

\[ Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + \hat{\epsilon}_i \]

\[ \sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \sum_{i=1}^{N} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{N} \hat{\epsilon}_i^2 + 2 \sum_{i=1}^{N} (\hat{Y}_i - \bar{Y})\hat{\epsilon}_i \]

\[ \rightarrow \sum_{i=1}^{N} (\hat{Y}_i - \bar{Y})\hat{\epsilon}_i = 0 \text{ because } \sum_{i=1}^{N} (\hat{Y}_i - \bar{Y})\hat{\epsilon}_i = \hat{\beta}_1 \sum_{i=1}^{N} (X_i - \bar{X})\hat{\epsilon}_i = 0 \]

\[ \rightarrow \text{ From } SST = SSE + SSR, \text{ get } \]

\[ R^2 = 1 - \frac{SSR}{SST} \text{ Measure of “Goodness of Fit”} \]

When is \( R^2 = 1 \)? When is \( R^2 = 0 \)?
Finite Sample Statistical Properties of OLS estimators

→ Focus of $\hat{\beta}_1$ (similar remarks for $\hat{\beta}_0$)

→ We include intercept term throughout

More remarks:

1. $\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} (X_i - \bar{X})Y_i}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \sum_{i=1}^{N} \left[ \frac{(X_i - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right] Y_i = \sum_{i=1}^{N} w_i Y_i$

$\hat{\beta}_1$ is “linear” (specifically meaning that it is a weighted average of the $Y_i$’s

2. Properties of the weights

a. $\sum_{i=1}^{N} w_i = \sum_{i=1}^{N} \left[ \frac{(X_i - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right] = \frac{\sum_{i=1}^{N} (X_i - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = 0$
b. \[ \sum_{i=1}^{N} w_i^2 = \sum_{i=1}^{N} \left[ \frac{(X_i - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right] = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{\left( \sum_{i=1}^{N} (X_i - \bar{X})^2 \right)^2} = \frac{1}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \]

c. \[ \sum_{i=1}^{N} w_i X_i = \sum_{i=1}^{N} \left[ \frac{(X_i - \bar{X}) X_i}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right] = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = 1 \]

3. \[ \hat{\beta}_1 = \sum_{i=1}^{N} w_i Y_i = \sum_{i=1}^{N} w_i (\hat{\beta}_0 + \beta_1 X_i + \epsilon_i) = \beta_1 + \sum_{i=1}^{N} w_i \epsilon_i \]

useful for comparing \( \hat{\beta}_1 \) vs \( \beta_1 \)

4. Statistical Properties depend on properties of variables, error terms, etc.
Best case assumptions

A1  There exists (unknown) constants $\beta_0$ and $\beta_1$ such that the data $\{Y_i, X_i\}_{i=1}^N$ can be said to be related according to $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, where the noise term $\varepsilon_i$ satisfies:

A2  $E[\varepsilon_i | X_1, ..., X_N] = 0$ for all $i = 1,...,N$

A3  $E[\varepsilon_i^2 | X_1, ..., X_N] = \sigma^2$ for all $i = 1,...,N$

A4  $E[\varepsilon_i \varepsilon_j | X_1, ..., X_N] = 0$ for all $i, j = 1,...,N, i \neq j$.

We also assume:

A5  $\sum_{i=1}^N (X_i - \bar{X})^2 \neq 0$, i.e., $X_i$ is not a degenerate random variable (i.e., not a constant)
Example

\( \{Y_i, X_i\}_{i=1}^N \) is a random draw from two variables \( X \) and \( Y \) satisfying

\[
E[Y \mid X] = \beta_0 + \beta_1 X \text{ for some constants } \beta_0 \text{ and } \beta_1
\]

\[
\text{var}[Y \mid X] = \sigma^2
\]

Note: ‘random sampling’ means:

→ Sampling process does not bias probability of certain observations being picked
→ Data is representative of the population
→ Draw are independent
Define $\varepsilon_i = Y - \beta_0 - \beta_1 X$

$\rightarrow E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$ implies $E[\varepsilon_i \mid X_i] = 0$

$\rightarrow \{Y_i, X_i\}_{i=1}^N$ independent draws $\Rightarrow \{\varepsilon_i, X_i\}_{i=1}^N$ are independent draws

$\Rightarrow E[\varepsilon_i \mid X_1, X_2, \ldots, X_N] = 0$

Also $E[\varepsilon_i \varepsilon_j \mid X_1, \ldots, X_N] = 0$

$\rightarrow \text{var}[Y_i \mid X_i] = \sigma^2$ and independent draws $\Rightarrow \text{var}[\varepsilon_i \mid X_1, \ldots, X_N] = \sigma^2$

$\rightarrow$ If $X$ is a non-degenerate random variable, random sampling implies A5
Example \( Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \) where A1-A5 holds

but observations with \( Y_i \geq c \) are (systematically) missing (truncated data)

Your data: \( \{sY_i, sX_i\}_{i=1}^N \) where \( s = 1 \) if \( \beta_0 + \beta_1 X_i + \varepsilon_i \leq Y_i \)

Your regression is \( sY_i = s\beta_0 + \beta_1 sX_i + s\varepsilon_i \)

Assumption A2 for your regression is \( E[s\varepsilon_i | sX_1, sX_2, ..., sX_N] = 0 \)

Consider \( E[s\varepsilon_i | sX_i] \)

\( s \) depends on \( \varepsilon_i \) (they are correlated)

\( \Rightarrow \) correlation changes with \( X_i \)

\( \Rightarrow E[s\varepsilon_i | sX_i] \neq 0 \), so A2 for your regression does not hold.
Unbiasedness

\[ \hat{\beta}_1 = \sum_{i=1}^{N} w_i Y_i = \sum_{i=1}^{N} w_i (\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_1 + \sum_{i=1}^{N} w_i \varepsilon_i \]

\[ \Rightarrow E[\hat{\beta}_1 | X_1, \ldots, X_N] = \beta_1 + \sum_{i=1}^{N} w_i E[\varepsilon_i | X_1, \ldots, X_N] = \beta_1 \]

\[ \Rightarrow E[\hat{\beta}_1] = \beta_1 \]
Example Simulated data (blue) with A1-A5 holding

Only data with $Y_i \leq 1500$ are observed (red)

Regression on untruncated data in blue
- Unbiased (we know this theoretically)

Regression on truncated data in red
- Systematically biased downwards
Example: Measurement Error

\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \]

but \( X_i \) only observed with error, i.e., only \( X_i^* = X_i + u_i \) are observed

(assume \( \varepsilon_i \) and \( u_i \) are independent error terms)

You run \( Y_i = \beta_0 + \beta_1 X_i^* + \nu_i \)

We have

\[ Y_i = \beta_0 + \beta_1 (X_i^* - u_i) + \varepsilon_i = \beta_0 + \beta_1 X_i^* + \varepsilon_i - \beta_1 u_i = \beta_0 + \beta_1 X_i^* + \nu_i \]

\( \rightarrow \) \( X_i^* \) and \( u_i \) are positively correlated

\( \rightarrow \) \( X_i^* \) and \( \nu_i \) are negatively correlated if \( \beta_1 > 0 \), positively correlated if \( \beta_1 < 0 \)

\( \rightarrow \) \( \hat{\beta}_1 \) is biased towards zero
Variance

\[
\text{var}[\hat{\beta}_1 \mid X_1, \ldots, X_N] = \sum_{i=1}^{N} w_i^2 \text{var}[\varepsilon_i \mid X_1, \ldots, X_N] = \sigma^2 \left( \frac{\sigma^2}{\sigma^2} \right) = \frac{\sigma^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2}
\]

Variance of \( \hat{\beta}_1 \)

→ increases with \( \sigma^2 \)  → decreases with \( N \)  → increases with variation in \( X_i \)

But \( \sigma^2 \) unknown. Use \( \hat{\sigma}^2 = \frac{1}{N - 2} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 \), unbiased for \( \sigma^2 \)

\[
\widehat{\text{var}}[\hat{\beta}_1 \mid X_1, \ldots, X_N] = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2}
\]
OLS has minimum variance among all linear unbiased estimators

→ Let \( \hat{\beta}_1 = \sum_{i=1}^{N} (w_i + v_i)Y_i \)

→ Unbiasedness of \( \hat{\beta}_1 \) requires \( \sum_{i=1}^{N} v_iX_i = 0 \) and \( \sum_{i=1}^{N} v_i = 0 \)

\[ \hat{\beta}_1 = \beta_1 + \sum_{i=1}^{N} w_i\epsilon_i + \beta_0 \sum_{i=1}^{N} v_i + \beta_1 \sum_{i=1}^{N} v_iX_i + \sum_{i=1}^{N} v_i\epsilon_i \]

→ \( \tilde{\beta}_1 = \sum_{i=1}^{N} (w_i + v_i)Y_i = \sum_{i=1}^{N} (w_i + v_i)\epsilon_i \)

\[
\text{var}[\tilde{\beta}_1 \mid X_1, \ldots, X_N] = \sum_{i=1}^{N} (w_i + v_i)^2 \text{var}[\epsilon_i \mid X_1, \ldots, X_N] \\
= \sigma^2 \sum_{i=1}^{N} (w_i^2 + w_i v_i + v_i^2) \\
= \sigma^2 \sum_{i=1}^{N} w_i^2 + \sigma^2 \sum_{i=1}^{N} v_i w_i + \sigma^2 \sum_{i=1}^{n} v_i^2 \\
= \text{var}[\hat{\beta}_1 \mid X_1, \ldots, X_N] + \sigma^2 \sum_{i=1}^{N} v_i^2 \geq 0
\]
Hypothesis Testing \( H_0 : \beta_1 = \beta^* \) vs \( H_0 : \beta_1 \neq \beta^* \)

\[ A6 \quad \varepsilon_i \mid X_1, \ldots, X_N \sim N(0, \sigma^2), \]

Then under null hypothesis

\[
t = \frac{\hat{\beta}_1 - \beta^*_1}{\sqrt{\hat{\sigma}^2 / \sum_{i=1}^{N} (X_i - \bar{X})^2}} \mid X_1, \ldots, X_N \sim t_{n-2}
\]

(Proof omitted)

0.95 level of significance test: reject null hypothesis if \( t \) falls outside the \((-c, c)\) interval, where \( c \) is the 97.5th percentile of the \( t_{N-2} \) distribution.
A peek at large sample results

\[ \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N} (X_i - \bar{X}) \varepsilon_i}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \beta_1 + \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X}) \varepsilon_i \]

If data is such that

\[ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X}) \varepsilon_i \xrightarrow{\text{p}} \text{cov}(X_i, \varepsilon_i) \]

\[ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2 \xrightarrow{\text{p}} \text{var}(X_i) \]

Then

\[ \hat{\beta}_1 = \beta_1 + \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X}) \varepsilon_i \xrightarrow{\text{p}} \beta_1 \]

\[ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2 \xrightarrow{\text{p}} \text{var}(X_i) \neq 0 \]
### Application: Prediction

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.73</td>
<td>3.18</td>
</tr>
<tr>
<td>4.56</td>
<td>3.81</td>
</tr>
<tr>
<td>1.82</td>
<td>-0.61</td>
</tr>
<tr>
<td>2.2</td>
<td>-0.25</td>
</tr>
<tr>
<td>1.65</td>
<td>0.83</td>
</tr>
<tr>
<td>3.4</td>
<td>1.53</td>
</tr>
<tr>
<td>3.07</td>
<td>0.91</td>
</tr>
<tr>
<td>2.39</td>
<td>1.23</td>
</tr>
<tr>
<td>3.16</td>
<td>3.09</td>
</tr>
<tr>
<td>3.94</td>
<td>1.81</td>
</tr>
</tbody>
</table>

Want to predict $Y$ when $X = 4$

- Perhaps take average of $Y$ such that $X = 4$? No such data
- Only one observation with $X$ close to 4.
But if we know that \( Y = \beta_0 + \beta_1 X + \varepsilon, \ E[\varepsilon | X] = 0 \), then

\[ \to \text{use all observations to estimate } \beta_0, \beta_1 \]

\[ \to \text{use these estimates to estimate } E[Y | X = 4] = \beta_0 + \beta_1 4 \]

\[ \to \text{use } \hat{E}[Y | X = 4] = \hat{\beta}_0 + 4\hat{\beta}_1 \text{ as prediction} \]

For our data:

\[
\hat{Y} = -2.178 + 1.247 \times X, \\
\quad (0.964) \quad (0.308)
\]

Prediction is \( \hat{Y}_{X=4} = -2.178 + 1.247(4) = 2.81 \).
Always calculate prediction error variance!

→ Measure of uncertainty regarding prediction

→ Can show \( \text{var}[\hat{\epsilon}_{X=X^*}] = \sigma^2 \left[ 1 + \frac{1}{N} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right] \)

use \( \hat{\epsilon}_{X=X^*} = Y_{X=X^*} - \hat{Y}_{X=X^*} = \beta_0 + \beta_1 X^* + \epsilon_{X=X^*} - \hat{\beta}_0 - \hat{\beta}_1 X^* = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) X^* + \epsilon_{X=X^*} \),

\[
\text{var}[\hat{\beta}_0] = \frac{\sigma^2 \sum_{i=1}^{N} X_i^2}{N \sum_{i=1}^{N} (X_i - \bar{X})^2}, \quad \text{var}[\hat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2}, \quad \text{cov}[\hat{\beta}_0, \hat{\beta}_1] = \frac{-\sigma^2 \bar{X}}{\sum_{i=1}^{N} (X_i - \bar{X})^2}
\]

For our data, \( s.e.[\hat{\epsilon}_{X=4}] = 0.98 \)
Causality vs Prediction

Think carefully before placing “causal” interpretation to regression estimates

\( \hat{\beta}_1 \) measures a correlation, correlation ≠ causality

Example

\[
Y = \alpha_0 + \alpha_1 Z + u
\]
\[
X = Z + \nu
\]

where \( u \sim N(0, \sigma_u^2) \), \( \nu \sim N(0, \sigma_v^2) \), independent noise terms

assume \( Z \sim N(0, \sigma_Z^2) \) exogenous
Here $Z$ causes both $X$ and $Y$

$\rightarrow$ If $Z$ changes, both $X$ and $Y$ will change

$\rightarrow$ $X$ and $Y$ will appear correlated

$\rightarrow$ changes in $X$ not because of $Z$ does not affect $Y$

Suppose you have $\{Y_i, X_i\}_{i=1}^N$ random draw, and regress

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

what does $\hat{\beta}_1$ measure?
Can show \( E[Z \mid X] = \frac{\sigma^2_Z}{\sigma^2_Z + \sigma^2_v} X \)

\[ \text{Hint: } \begin{bmatrix} X \\ Z \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_Z + \sigma^2_v & \sigma^2_Z + \sigma^2_v \\ \sigma^2_Z + \sigma^2_v & \sigma^2_Z \end{bmatrix} \right) \]

Therefore

\[ E[Y \mid X] = \alpha_0 + \alpha_1 E[Z \mid X] + E[u \mid X] = \alpha_0 + \frac{\alpha_1 \sigma^2_Z}{\sigma^2_Z + \sigma^2_v} X \]

→ if we define \( \varepsilon = Y - \alpha_0 - \frac{\alpha_1 \sigma^2_Z}{\sigma^2_Z + \sigma^2_v} X \), then
\[ Y = \alpha_0 + \frac{\alpha_1 \sigma^2_Z}{\sigma^2_Z + \sigma^2_v} X + \varepsilon, \quad E[\varepsilon \mid X] = 0 \]

→ if we define \( \varepsilon = Y - \alpha_0 - 0X \), then
\[ Y = \alpha_0 + 0X + \underbrace{\alpha_1 Z + u}_{\varepsilon}, \quad E[\varepsilon \mid X] \neq 0 \]

We will have \( E[\hat{\beta}_1] = \frac{\alpha_1 \sigma^2_Z}{\sigma^2_Z + \sigma^2_v} \neq 0 \).
In large samples:

\[
\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})Y_i}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(\alpha_0 + \alpha_1Z_i + u_i)}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \alpha_1 \frac{\sum_{i=1}^{N} (X_i - \bar{X})Z_i}{\sum_{i=1}^{N} (X_i - \bar{X})^2} + \frac{\sum_{i=1}^{N} (X_i - \bar{X})u_i}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \alpha_1 \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(Z_i - \bar{Z}) + \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})u_i \to_p \alpha_1 \frac{\text{cov}(X, Z)}{\text{var}(X)} + 0 = \frac{\alpha_1 \sigma^2_Z}{\sigma^2_Z + \sigma_v^2}
\]
This is an omitted variable problem:

In the (multiple) regression \( Y = \beta_0 + \beta_1 X + \beta_2 Z + u \), \( E[u \mid X, Z] = 0 \),

We will have \( E[\hat{\beta}_1] = 0 \)

So far we have seen bias because of

- Measurement error
- Truncation
- Omitted variables
Simultaneity Bias

$q^d = \delta_0 + \delta_1 p + \varepsilon^d$  (demand, $\delta_1 < 0$)
$q^s = \alpha_0 + \alpha_1 p + \varepsilon^s$  (supply, $\alpha_1 > 0$)
$q^d = q^s$  (market clearing)

$\delta_0 + \delta_1 p + \varepsilon^d = \alpha_0 + \alpha_1 p + \varepsilon^s$

$p_t = \frac{\alpha_0 - \delta_0}{\delta_1 - \alpha_1} + \frac{\varepsilon^s - \varepsilon^d}{\delta_1 - \alpha_1}$ and
$q_t = \left( \delta_0 + \delta_1 \frac{\alpha_0 - \delta_0}{\delta_1 - \alpha_1} \right) + \frac{\delta_1 \varepsilon^s - \alpha_1 \varepsilon^d}{\delta_1 - \alpha_1}$

$\text{var}[P_t] = \frac{\sigma_s^2 + \sigma_d^2}{(\delta_1 - \alpha_1)^2}$ and $\text{cov}[Q_t, P_t] = \frac{\delta_1 \sigma_s^2 + \alpha_1 \sigma_d^2}{(\delta_1 - \alpha_1)^2}$

$\hat{\beta}_1^{OLS} \rightarrow_p \frac{\text{cov}[Q_t, P_t]}{\text{var}[P_t]} = \frac{\delta_1 \sigma_s^2 + \alpha_1 \sigma_d^2}{\sigma_s^2 + \sigma_d^2}$
**Heteroskedasticity**

A3’  \[ E[\varepsilon_i^2 | X_1, ..., X_N] = \sigma_i^2 = h(X_i) \] for all \( i = 1, ..., N \)

E.g. → OLS still unbiased/consistent  
(proof did not require A3)  
→ But usual variance formula is wrong

\[
\text{var}[\hat{\beta}_1^{OLS} | X_1, ..., X_N] = \sum_{i=1}^{N} w_i^2 \text{var}[\varepsilon_i | X_1, ..., X_N] = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2 \sigma_i^2}{\left[ \sum_{i=1}^{N} (X_i - \bar{X})^2 \right]^2}
\]
Problem: we do not have $\sigma_i^2$

Solution: can show that

$$\hat{\text{var}}[\hat{\beta}_1^{OLS}] = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2 \hat{\varepsilon}_i^2}{\left[ \sum_{i=1}^{N} (X_i - \bar{X})^2 \right]^2} \rightarrow_p \text{ var}[\hat{\beta}_1]$$

(Proof given in later lesson)

Furthermore,

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\text{var}}[\hat{\beta}_1]}} \overset{a}{\sim} N(0,1)$$
Downside: OLS not the most efficient estimator

(Indirect proof: demonstrate existence of more efficient estimator)

(Note the stronger assumptions)

A1 The sample \( \{Y_i, X_i\}_{i=1}^N \) is related according to \( Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \)

A2 \( E[\varepsilon_i \mid X_1, \ldots, X_N] = 0 \) for all \( i = 1, \ldots, N \)

A3 \( E[\varepsilon_i^2 \mid X_1, \ldots, X_N] = \sigma^2 h(X_i) \) for all \( i = 1, \ldots, N \), where \( h(X_i) \) is known

A4 \( E[\varepsilon_i \varepsilon_j \mid X_1, \ldots, X_N] = 0 \) for all \( i, j = 1, \ldots, N, i \neq j \)

A5 \( \sum_{i=1}^N (X_i - \bar{X})^2 \neq 0. \)
Apply weight $1/\sqrt{h(X_i)}$ to each observation:

$$\frac{Y_i}{\sqrt{h(X_i)}} = \beta_0 \left( \frac{1}{\sqrt{h(X_i)}} \right) + \beta_1 \frac{X_i}{\sqrt{h(X_i)}} + \frac{\varepsilon_i}{\sqrt{h(X_i)}}$$

or

$$Y_i^* = \beta_0 X_{0i}^* + \beta_1 X_{1i}^* + \varepsilon_i^*$$

(Note this is potentially a multiple regression model)

$$\text{var} \left[ \varepsilon_i^* \mid X_1, \ldots, X_N \right] = \text{var} \left[ \frac{\varepsilon_i}{\sqrt{h(X_i)}} \mid X_1, \ldots, X_N \right] = \sigma^2$$

OLS on weighted regression will give BLU estimators

“Weighted Least Squares” (special case of “Generalized Least Squares”)

ECON207 Session 2 Slide 33
Example

Suppose

\[ Y_i = \beta_1 X_i + \varepsilon_i \] (note no intercept term), and

\[ E[\varepsilon_i^2 \mid X_1, \ldots, X_N] = \sigma^2 X_i^2 \]

\[ \hat{\beta}_1^{OLS} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i (\beta_1 X_i + \varepsilon_i)}{\sum_{i=1}^{n} X_i^2} = \beta_1 + \frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2} \]

\[ \text{var}[\hat{\beta}_1^{OLS}] = \frac{\sum_{i=1}^{n} X_i^2 \text{var}[\varepsilon_i]}{\left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{\sigma^2 \sum_{i=1}^{n} X_i^4}{\left(\sum_{i=1}^{n} X_i^2\right)^2} \]
Weighted Least Squares

\[
\left( \frac{Y_i}{X_i} \right) = \beta_1 + \frac{\varepsilon_i}{X_i}, \text{ i.e., } Y_i^* = \beta_1 + \varepsilon_i^*, \text{ where } Y_i^* = \frac{Y_i}{X_i} \text{ and } \text{var}[\varepsilon_i^*] = \sigma^2
\]

\[
\rightarrow \hat{\beta}_1^{WLS} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i}{X_i}
\]

\[
\rightarrow \text{var}\left[\hat{\beta}_1^{WLS}\right] = \frac{\sigma^2}{N}
\]

Can show

\[
\frac{\sum_{i=1}^{n} X_i^4}{\left(\sum_{i=1}^{n} X_i^2\right)^2} \geq \frac{1}{N} \quad \text{(exercise!)}
\]

Therefore

\[
\text{var}\left[\hat{\beta}_1^{WLS}\right] \leq \text{var}\left[\hat{\beta}_1^{OLS}\right]
\]
Example: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $E[\varepsilon_i^2 \mid X_1, X_2, \ldots, X_N] = \sigma^2 X_i$

WLS: $\frac{Y_i}{X_i} = \beta_0 \left( \frac{1}{X_i} \right) + \beta_1 \frac{X_i}{X_i} + \frac{\varepsilon_i}{X_i}$

$\Rightarrow \frac{\hat{Y}_i}{X_i} = 2.532 + 4.826 \frac{1}{X_i}$, $R^2 = 0.056$, $N = 200$

(0.102) (1.407)

$\Rightarrow \hat{Y} = 4.826 + 2.532 X$, $R^2 = 0.681$, $N = 200$

(1.407) (0.102)

OLS: $\hat{Y} = 6.286 + 2.431 X$, $R^2 = 0.683$, $N = 200$

(1.781) (0.118)

OLS with h.c.s.e.: $\hat{Y} = 6.286 + 2.431 X$, $R^2 = 0.683$, $N = 200$

(1.864) (0.136)
How to calculate $R^2$ for WLS?

In our example:

$$
\hat{\epsilon}_{i}^{WLS} = Y_i - 4.826 - 2.532X_i
$$

$$
R_{WLS}^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{\epsilon}_{i}^{WLS})^2}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}
$$
Remarks:

Although WLS is BLU, note

→ Requires knowledge of form of heteroskedasticity

→ In Multiple Linear Regression Model, form of heteroskedasticity usually has unknown parameters (which must be estimated)

→ Heteroskedasticity-consistent standard errors are valid in OLS even if there is no heteroskedasticity

→ If heteroskedasticity is mild, maybe OLS (with heteroskedasticity-consistent s.e.) is good enough