Simple Linear Regression

William Brasic

The University of Arizona

Simple Linear Regression

Definition 1: Simple Linear Regression

The population simple linear regression model for $i = 1, \ldots, n$ is given as

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

- y_i is often called a dependent variable, response variable, predicted variable, or outcome variable.
- x_i is often called an independent variable, explanatory variable, regressor, or covariate.
- u_i is often called the error term or idiosyncratic shock.
  - It represents all other factors than x_i that explain y_i.
- β_0 is the intercept/constant term.
- β_1 is the slope parameter.
  - It represents the effect of x on y.

How do we Estimate the Parameters?

Question 1: How do we Estimate the Parameters?

How would you estimate β_0 and β_1?

How do we Estimate the Parameters?

Question 1: How do we Estimate the Parameters?

How would you estimate β_0 and β_1?

Answer to Question 1

Ordinary Least Squares (OLS).

- We try to minimize the squared difference between our observed y_i and our predicted $\widehat{y_i}$ for each observation i.
  - $\widehat{y_i}$ is often called a fitted value for observation i.

Residuals

Definition 2: Residual

A residual $\widehat{u}_i$ is defined as

$$\widehat{u}_i = y_i - \widehat{y}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i.$$

- The residual is the estimated error for each i.
- OLS wants to minimize these residuals for each i.

Sum of Squared Residuals (SSR)

Definition 3: Sum of Squared Residuals

The sum of squared residuals (SSR) is defined as

$$
\begin{aligned}
SSR = \sum_{i=1}^{n} \widehat{u}_i^2 &= \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \\
&= \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2 .
\end{aligned}
$$

- The squaring of the residuals makes everything positive.
- We use SSR instead of the summing the absolute values because the latter is not differentiable at zero.

Ordinary Least Squares (OLS)

Definition 4: Ordinary Least Squares (OLS)

The ordinary least squares (OLS) algorithm minimizes the SSR to obtain $\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$:

$$\underset{b_0, b_1}{\arg\min} \sum_{i=1}^{n} \widehat{u}_i^2 = \left(\widehat{\beta}_0, \widehat{\beta}_1\right).$$

How to Minimize the SSR?

Question 2: How to Minimize the SSR?

How do we minimize the SSR?

How to Minimize the SSR?

Question 2: How to Minimize the SSR?

How do we minimize the SSR?

Answer to Question 2

Calculus! Take the derivative of the SSR with respect to each parameter. Then, set the derivatives equal to zero and solve the equations for $\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$.

The OLS Solution

Theorem 1: The OLS Solution

The OLS Solution is given by

$$\begin{aligned}\widehat{\beta}_0 &= \frac{1}{n}\sum_{i=1}^{n} y_i - \widehat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \overline{Y} - \widehat{\beta}_1 \overline{X} \\ \widehat{\beta}_1. &= \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \\ &= \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}.\end{aligned}$$

Proof of the OLS Solution

Proof 1: OLS Solution's Constant Term Part 1

The problem is to find the intercept term $\widehat{\beta}_0$ by minimizing the SSR:

$$\min_{\widehat{\beta}_0} \sum_{i=1}^{n} \widehat{u}_i^2 = \min_{\widehat{\beta}_0} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 .$$

To do so, we take the derivative of the SSR with respect to $\widehat{\beta}_0$ and set it equal to zero (first order condition).

Proof of the OLS Solution

Proof 1: OLS Solution's Constant Term Part 2

$$
\begin{aligned}
\frac{\partial SSR}{\partial \widehat{\beta}_0} &= 2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)(-1) \\
&= -2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)
\end{aligned}
$$

Proof of the OLS Solution

Proof 1: OLS Solution's Constant Term Part 2

$$\frac{\partial SSR}{\partial \widehat{\beta}_0} = 2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)(-1)$$
$$= -2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)$$

Setting this equal to zero we have

$$-2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0 \iff \sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0.$$

Proof of the OLS Solution

Proof 1: OLS Solution's Constant Term Part 3

Solving the equation for $\widehat{\beta}_0$ we get

$$
\begin{aligned}
\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0 &\iff \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \widehat{\beta}_0 - \sum_{i=1}^{n} \widehat{\beta}_1 x_i = 0 \\
&\iff \sum_{i=1}^{n} \widehat{\beta}_0 = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \widehat{\beta}_1 x_i
\end{aligned}
$$

Proof of the OLS Solution

Proof 1: OLS Solution's Constant Term Part 4

Lastly,

$$\begin{aligned}\sum_{i=1}^{n} \widehat{\beta}_0 = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \widehat{\beta}_1 x_i &\iff n\widehat{\beta}_0 = \sum_{i=1}^{n} y_i - \widehat{\beta}_1 \sum_{i=1}^{n} x_i \\ &\iff \widehat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \widehat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i \\ &\iff \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}. \quad \square\end{aligned}$$

Hooray!

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 1

The problem is to find the slope term $\widehat{\beta}_1$ by minimizing the SSR:

$$\min_{\widehat{\beta}_1} \sum_{i=1}^{n} \widehat{u}_i^2 = \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 1

The problem is to find the slope term $\widehat{\beta}_1$ by minimizing the SSR:

$$\begin{aligned}\min_{\widehat{\beta}_1} \sum_{i=1}^{n} \widehat{u}_i^2 &= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2 \\ &= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \left(\overline{Y} - \widehat{\beta}_1 \overline{X}\right) - \widehat{\beta}_1 x_i\right)^2\end{aligned}$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 1

The problem is to find the slope term $\widehat{\beta}_1$ by minimizing the SSR:

$$\begin{aligned}
\min_{\widehat{\beta}_1} \sum_{i=1}^{n} \widehat{u}_i^2 &= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2 \\
&= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \left(\overline{Y} - \widehat{\beta}_1 \overline{X}\right) - \widehat{\beta}_1 x_i\right)^2 \\
&= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \overline{Y} + \widehat{\beta}_1 \overline{X} - \widehat{\beta}_1 x_i\right)^2
\end{aligned}$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 1

The problem is to find the slope term $\widehat{\beta}_1$ by minimizing the SSR:

$$\begin{aligned}
\min_{\widehat{\beta}_1} \sum_{i=1}^{n} \widehat{u}_i^2 &= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 \\
&= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \left(\overline{Y} - \widehat{\beta}_1 \overline{X} \right) - \widehat{\beta}_1 x_i \right)^2 \\
&= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \overline{Y} + \widehat{\beta}_1 \overline{X} - \widehat{\beta}_1 x_i \right)^2 \\
&= \min_{\widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \overline{Y} - \widehat{\beta}_1 \left(x_i - \overline{X} \right) \right)^2 .
\end{aligned}$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 2

Now we take the derivative of the SSR with respect to $\widehat{\beta}_1$:

$$\frac{\partial SSR}{\partial \widehat{\beta}_1} = -2 \sum_{i=1}^{n} \left(y_i - \overline{Y} - \widehat{\beta}_1 \left(x_i - \overline{X} \right) \right) \left(x_i - \overline{X} \right).$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 3

Setting this derivative equal to zero gives

$$-2\sum_{i=1}^{n}\left(y_i-\overline{Y}-\widehat{\beta}_1\left(x_i-\overline{X}\right)\right)\left(x_i-\overline{X}\right)=0$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 3

Setting this derivative equal to zero gives

$$
\begin{aligned}
& -2\sum_{i=1}^{n}\left(y_i-\overline{Y}-\widehat{\beta}_1\left(x_i-\overline{X}\right)\right)\left(x_i-\overline{X}\right)=0 \\
\iff & \sum_{i=1}^{n}\left(y_i-\overline{Y}-\widehat{\beta}_1\left(x_i-\overline{X}\right)\right)\left(x_i-\overline{X}\right)=0
\end{aligned}
$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 3

Setting this derivative equal to zero gives

$$
\begin{aligned}
& -2\sum_{i=1}^{n}\left(y_i-\overline{Y}-\widehat{\beta}_1\left(x_i-\overline{X}\right)\right)\left(x_i-\overline{X}\right)=0 \\
\iff & \sum_{i=1}^{n}\left(y_i-\overline{Y}-\widehat{\beta}_1\left(x_i-\overline{X}\right)\right)\left(x_i-\overline{X}\right)=0 \\
\iff & \sum_{i=1}^{n}\left(\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right)-\widehat{\beta}_1\left(x_i-\overline{X}\right)\left(x_i-\overline{X}\right)\right)=0.
\end{aligned}
$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 4

Some algebra gives

$$\sum_{i=1}^{n}\Big(\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right)-\widehat{\beta}_1\left(x_i-\overline{X}\right)\left(x_i-\overline{X}\right)\Big)=0$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 4

Some algebra gives

$$\sum_{i=1}^{n}\Big(\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right)-\widehat{\beta}_1\left(x_i-\overline{X}\right)\left(x_i-\overline{X}\right)\Big)=0$$

$$\iff \sum_{i=1}^{n}\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right)-\sum_{i=1}^{n}\widehat{\beta}_1\left(x_i-\overline{X}\right)^2=0$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 4

Some algebra gives

$$\sum_{i=1}^{n}\left(\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right)-\widehat{\beta}_1\left(x_i-\overline{X}\right)\left(x_i-\overline{X}\right)\right)=0$$

$$\iff \sum_{i=1}^{n}\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right)-\sum_{i=1}^{n}\widehat{\beta}_1\left(x_i-\overline{X}\right)^2=0$$

$$\iff \widehat{\beta}_1\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2=\sum_{i=1}^{n}\left(y_i-\overline{Y}\right)\left(x_i-\overline{X}\right).$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 5

Solving for $\widehat{\beta}_1$ yields

$$\widehat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2 = \sum_{i=1}^{n} \left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)$$

$$\iff \widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}.$$

Proof of the OLS Solution

Proof 2: OLS Solution's Slope Term Part 6

Lastly, multiplying and dividing by $\frac{1}{n-1}$ doesn't change anything. So,

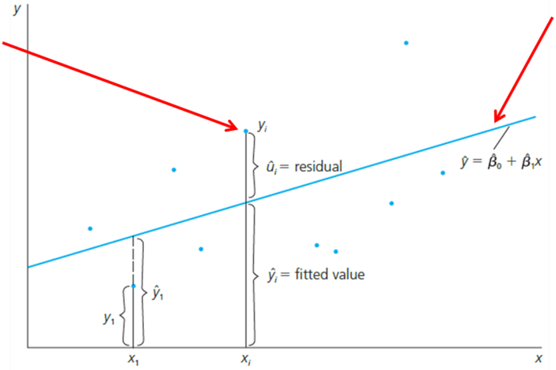
$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \\
&= \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \\
&= \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}. \quad \square
\end{aligned}$$

Horray!

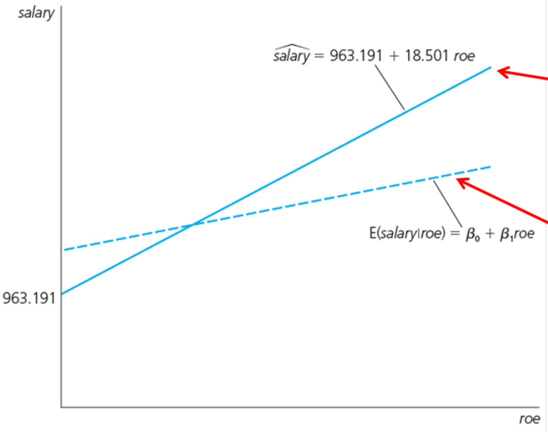
Visualization of What We Are Doing

For example, the i-th data point (x_i, y_i)

Fitted regression line



Visualization of What We Are Doing



Fitted regression line
(depends on sample)

Unknown population regression line

Properties of OLS

Property 1: OLS Residuals Sum to Zero

Recall from the F.O.C. for $\widehat{\beta}_0$ from Part 3 of the proof that

$$\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = \sum_{i=1}^{n}\widehat{u}_i = 0.$$

- This directly implies the mean of the residuals $\overline{U} = \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i = 0$.
- If we don't included an intercept, then the residuals don't necessary sum to zero.

Properties of OLS

Property 2: Zero Sample Covariance Between Residuals and Regressor

Without first substituting in $\widehat{\beta}_0$ for the F.O.C. for $\widehat{\beta}_1$, we could've written it as

$$\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) x_i = \sum_{i=1}^{n} \widehat{u}_i x_i = 0.$$

Properties of OLS

Property 3: Zero Sample Covariance Between Residuals and Regressor

Without first substituting in $\widehat{\beta}_0$ for the F.O.C. for $\widehat{\beta}_1$, we could've written it as

$$\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) x_i = \sum_{i=1}^{n} \widehat{u}_i x_i = 0.$$

Thus,

$$\begin{aligned}
\widehat{\sigma}_{x\widehat{u}} &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\widehat{u}_i - \overline{U}\right)\left(x_i - \overline{X}\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\widehat{u}_i x_i - \frac{\overline{X}}{n-1}\sum_{i=1}^{n}\widehat{u}_i - \frac{\overline{U}}{n-1}\sum_{i=1}^{n}x_i + \frac{n\overline{XU}}{n-1} \\
&= 0.
\end{aligned}$$

Properties of OLS

Property 4: Sample Means of Dependent Variable and Regressor Lie on Regression Line

Recall $\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$. So,

$$\overline{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{X}.$$

- If we try to predict y_i with $\overline{X}$, the prediction will be $\overline{Y}$.

Total Sum of Squares (SST)

Definition 5: Total Sum of Squares (SST)

The total sum of squares (SST) is defined as

$$SST = \sum_{i=1}^{n} \left(y_i - \overline{Y}\right)^2.$$

- Measures how much the observed outcome varies with respect to its mean.

Explained Sum of Squares (SSE)

Definition 6: Explained Sum of Squares (SSE)

The explained sum of squares (SSE) is defined as

$$SSE = \sum_{i=1}^{n} \left(\widehat{y}_i - \overline{Y}\right)^2.$$

- Measures how much the predicted outcome varies with respect to the mean of the observe outcome.

SST = SSE + SSR

Property 5: SST = SSE + SSR

We can write the SST in terms of the SSE and SSR as

$$SST = SSE + SSR.$$

R-squared

Definition 7: R-squared

The R-squared measures how well our regressor explains our outcome and is defined as

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}.$$

- R^2 lies between zero and one.
- Higher R^2 suggests better model fit.
- R^2 is overrated as it always increases as the number. covariates in model increases.

Adjusted R-squared

Definition 8: Adjusted R-squared

The adjusted R-squared is a modified version of the R-square defined as

$$\widetilde{R}^2 = 1 - \frac{SSR/(n-k)}{SST/(n-1)} = 1 - \frac{\left(1 - R^2\right)(n-1)}{n-k}$$

where k is the number of estimated parameters (including the intercept).

- Penalization for adding more regressors.
- Does not lie between zero and one.
  - Can be negative when a horizontal regression lie (such as at $\overline{Y}$) predicts y better.

Regression with Logs

Property 6: Level-Level Regression Model

A level-level simple linear regression model is given by

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

It follows that

$$\frac{dy}{dx} = \beta_1.$$

- When x changes by 1 unit, y changes by β_1 units.
- When x changes by 3 units, y changes by $3 * \beta_1$. units

Regression with Logs

Property 7: Log-Level Regression Model

A log-level simple linear regression model is given by

$$\ln(y_i) = \beta_0 + \beta_1 x_i + u_i.$$

It follows that

$$\frac{d\ln(y)}{dx} = \beta_1.$$

- This represents a semi-elasticity of y with respect to x.
- When x changes by 1 unit, y changes by approximately $100 * \beta_1$ percent.
- When x changes by 5 units, y changes by approximately $100 * 5 * \beta_1$ percent.

Regression with Logs

Proof 3: Property 7 Part 1

If $\ln(y) = \beta_0 + \beta_1 x + u$, then

$$\frac{d\ln(y)}{dx} = \beta_1.$$

Note that $\ln(y)$ is a function of x Let $u = \ln(y)$. By the chain rule

$$\frac{d\ln(y)}{dx} = \frac{du}{dx} = \frac{du}{dy}\frac{dy}{dx} = \frac{1}{y}\frac{dy}{dx}$$

Setting these equations equal to each other we see

$$\frac{1}{y}\frac{dy}{dx} = \beta_1 \iff \frac{dy}{y} = \beta_1 dx.$$

Regression with Logs

Proof 3: Property 7 Part 2

$\frac{dy}{y}$ represents the relative or proportional change in y telling us how much y changes relative to its current value. Think of dy as $y_1 - y$ then $\frac{dy}{y}$ is the exact formula for the proportional change in y. dx represents a small change in x. So, given a small change in x (dx), the proportional change in y ($\frac{dy}{y}$) is β_1. To get the percentage change from the proportional change, we just multiply both sides of the equation by 100. This gives

$$100\frac{dy}{y} = 100\beta_1 dx.$$

Thus, the percentage change in y ($100\frac{dy}{y}$) given a small change in x (dx) is given by $100\beta_1$.

Regression with Logs

Property 8: Log-Log Regression Model

A log-log simple linear regression model is given by

$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + u_i.$$

It follows that

$$\frac{d\ln(y)}{d\ln(x)} = \beta_1.$$

- This represents an elasticity of y with respect to x.
- When x changes by 1 percent, y changes by approximately β_1 percent.
- When x changes by 5 percent, y changes by approximately $5\beta_1$ percent.

The Meaning of “Linear” in Linear Regression

Property 9: The Meaning of “Linear” in Linear Regression

When we estimate a linear regression model, we mean linear in parameters, not necessarily linear in covariates.

- The model $y_i = \beta_0 + \beta_1 \sqrt{x_i} + u_i$ is linear in parameters so we are fine.
- The model $y_i = \beta_0 + \sqrt{\beta_1} x_i + u_i$ is not linear in parameters and we cannot cast it as a linear regression model.

The Meaning of “Linear” in Linear Regression

Question 3

Is the model $y_i = e^{\beta_0} x_i^{\beta_1} e^{u_i}$ linear in parameters? Could we cast it as a linear regression problem?

The Meaning of "Linear" in Linear Regression

Question 3

Is the model $y_i = e^{\beta_0} x_i^{\beta_1} e^{u_i}$ linear in parameters? Could we cast it as a linear regression problem?

Answer to Question 3

As it stands, it is not linear in parameters because y_i is a non-linear function of the parameters. However, we could transform it into a linear regression problem by taking the log of both sides of the equation giving

$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + u_i$$

which is a log-log model we can estimate via OLS.

OLS Does Not Necessarily Give us Causality

Property 10: OLS Does Not Necessarily Give us Causality

While OLS will generate the best possible line by minimizing the SSR, this does *not* mean $\widehat{\beta}_1$ measures a causal relationship. To identify the causal relationship between y and x, we need

$$\mathbb{E}\left[\widehat{\beta}_1\right] = \beta_1.$$

- This means $\widehat{\beta}_1$ is unbiased for β_1.
- We need the following four assumptions to be able to conclude $\widehat{\beta}_1$ is unbiased for β_1.

Linear in Parameters

SLR Assumption 1: Linear in Parameters

The population model is a linear function of the parameters.

- For instance, $y_i = \beta_0 + \beta_1 x_i + u_i$.

Random Sampling

SLR Assumption 2: Random Sampling

We have a random (i.i.d.) sample $\{(y_i, x_i)\}_{i=1}^{n}$ from the population of interest.

- This will ensure Assumption 4 holds for the entire sample and not just subsets.

Non-Zero Regressor Variance

SLR Assumption 3: Non-Zero Regressor Variance

The sample outcomes of our regressor, namely $\{x_i\}_{i=1}^n$, are not all the same value, i.e., $\widehat{\sigma}_x^2 \neq 0$.

- Easy to verify by loading data into software and calculating the sample variance of $\{x_i\}_{i=1}^n$.

Zero Conditional Mean

SLR Assumption 4: Zero Conditional Mean

The expectation of the error term conditioned on the regressor is zero, i.e., $\mathbb{E}[u_i \mid x_i] = 0$ for each $i = 1, \ldots, n$.

- Also called the exogeneity assumption
- Important implications include:
  1. $\mathbb{E}[u_i] = \mathbb{E}[\mathbb{E}[u_i \mid x_i]] = \mathbb{E}[0] = 0$.
  2. $\text{Cov}[u_i, x_i] = \mathbb{E}[u_i x_i] - \mathbb{E}[u_i]\mathbb{E}[x_i] = \mathbb{E}[u_i x_i] = \mathbb{E}[\mathbb{E}[u_i x_i \mid x_i]] = \mathbb{E}[x_i \mathbb{E}[u_i \mid x_i]] = 0$.
  3. $\mathbb{E}[y_i \mid x_i] = \mathbb{E}[\beta_0 + \beta_1 x_i + u_i \mid x_i] = \beta_0 + \beta_1 x_i$ so β_1 represents the average impact of x_i on y_i .
  - Sample analogs of 1. and 2. are justified by Properties 1 and 2.

Unbiasedness of OLS

Theorem 2: Unbiasedness of OLS

Under SLR Assumptions 1-4, the OLS estimator is unbiased, i.e.,

$$\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] = \beta_0.$$
$$\mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] = \beta_1.$$

- This means that on average if we take many samples and compute $\widehat{\beta}_0$ and $\widehat{\beta}_1$ we will get β_0 and β_1.
  - The sampling distribution of our estimated parameters is centered around their true values.
- By the law of total expectation, $\mathbb{E}\left[\widehat{\beta}_0\right] = \beta_0$ and $\mathbb{E}\left[\widehat{\beta}_1\right] = \beta_1$.

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 1

Using SLR Assumption 1 of $y_i = \beta_0 + \beta_1 x_i + u_i$ and Assumption 3 of $\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2 > 0$,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 1

Using SLR Assumption 1 of $y_i = \beta_0 + \beta_1 x_i + u_i$ and Assumption 3 of $\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2 > 0$,

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i + u_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}\end{aligned}$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 1

Using SLR Assumption 1 of $y_i = \beta_0 + \beta_1 x_i + u_i$ and Assumption 3 of $\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2 > 0$,

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i + u_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i + u_i - \beta_0 - \beta_1\overline{X}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}.
\end{aligned}$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 2

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(\beta_0 + \beta_1 x_i + u_i - \beta_0 - \beta_1 \overline{X}\right)\left(x_i - \overline{X}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 2

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(\beta_0+\beta_1 x_i+u_i-\beta_0-\beta_1\overline{X}\right)\left(x_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\\ &= \frac{\sum_{i=1}^{n}\left(\beta_1\left(x_i-\overline{X}\right)+u_i\right)\left(x_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\end{aligned}$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 2

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(\beta_0+\beta_1 x_i+u_i-\beta_0-\beta_1\overline{X}\right)\left(x_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(\beta_1\left(x_i-\overline{X}\right)+u_i\right)\left(x_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \frac{\beta_1\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}+\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}
\end{aligned}
$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 2

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(\beta_0+\beta_1 x_i+u_i-\beta_0-\beta_1\overline{X}\right)\left(x_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(\beta_1\left(x_i-\overline{X}\right)+u_i\right)\left(x_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \frac{\beta_1\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}+\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \beta_1+\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}.
\end{aligned}
$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 3

Now, using SLR Assumption 2 of random sampling and SLR Assumption 4 of exogeneity,

$$\mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] = \mathbb{E}[\beta_1] + \mathbb{E}\left[\left.\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}\right| x_i\right]$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 3

Now, using SLR Assumption 2 of random sampling and SLR Assumption 4 of exogeneity,

$$\begin{aligned}\mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] &= \mathbb{E}[\beta_1] + \mathbb{E}\left[\left.\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}\right| x_i\right] \\ &= \beta_1 + \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\mathbb{E}[u_i \mid x_i]}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}\end{aligned}$$

Proof of the Unbiasedness of OLS

Proof 4: Unbiasedness of the OLS Slope Term Part 3

Now, using SLR Assumption 2 of random sampling and SLR Assumption 4 of exogeneity,

$$\begin{aligned} \mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] &= \mathbb{E}[\beta_1] + \mathbb{E}\left[\left.\frac{\sum_{i=1}^n \left(x_i - \overline{X}\right) u_i}{\sum_{i=1}^n \left(x_i - \overline{X}\right)^2}\right| x_i\right] \\ &= \beta_1 + \frac{\sum_{i=1}^n \left(x_i - \overline{X}\right) \mathbb{E}[u_i \mid x_i]}{\sum_{i=1}^n \left(x_i - \overline{X}\right)^2} \\ &= \beta_1. \quad \square \end{aligned}$$

Hooray!

Proof of the Unbiasedness of OLS

Proof 5: Unbiasedness of the OLS Intercept Term

$$\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] = \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right]$$

Proof of the Unbiasedness of OLS

Proof 5: Unbiasedness of the OLS Intercept Term

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] &= \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}\left[\beta_0 + \beta_1 \overline{X} - \widehat{\beta}_1 \overline{X} \mid x_i\right]
\end{aligned}$$

Proof of the Unbiasedness of OLS

Proof 5: Unbiasedness of the OLS Intercept Term

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] &= \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}\left[\beta_0 + \beta_1 \overline{X} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}[\beta_0 \mid x_i] + \mathbb{E}\left[\beta_1 \overline{X} \mid x_i\right] - \mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] \overline{X}
\end{aligned}$$

Proof of the Unbiasedness of OLS

Proof 5: Unbiasedness of the OLS Intercept Term

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] &= \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}\left[\beta_0 + \beta_1 \overline{X} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}[\beta_0 \mid x_i] + \mathbb{E}\left[\beta_1 \overline{X} \mid x_i\right] - \mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] \overline{X} \\
&= \beta_0 + \beta_1 \overline{X} - \beta_1 \overline{X}
\end{aligned}
$$

Proof of the Unbiasedness of OLS

Proof 5: Unbiasedness of the OLS Intercept Term

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] &= \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}\left[\beta_0 + \beta_1 \overline{X} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\
&= \mathbb{E}[\beta_0 \mid x_i] + \mathbb{E}\left[\beta_1 \overline{X} \mid x_i\right] - \mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] \overline{X} \\
&= \beta_0 + \beta_1 \overline{X} - \beta_1 \overline{X} \\
&= \beta_0. \quad \square
\end{aligned}
$$

Horray!

Variance of the Parameter Estimates

Question 4

Under SLR Assumptions 1-4 we will, on average, correctly estimate β_0 and β_1. However, what is the variability of this estimate?

- By "on average", we mean the sampling distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ will be centered around β_0 and β_1, respectively.

Homoskedasticity

SLR Assumption 5: Homoskedastic Errors

Homoskedasticity states $\mathbb{V}[u_i \mid x_i] = \sigma^2$ for each $i = 1, \ldots, n$.

- Under SLR Assumptions 1-5, $\mathbb{V}[u_i] = \mathbb{E}[\mathbb{V}[u_i \mid x_i]] + \mathbb{V}[\mathbb{E}[u_i \mid x_i]] = \mathbb{E}[\sigma^2] = \sigma^2$.
- The variance of our error is constant across all observations when conditioning on our regressor.
- When this assumption fails, we say the errors are heteroskedastic.
- Really not that important since White's 1980 correction.

Variance of OLS Estimates

Theorem 3: Variance of OLS Estimates

Under SLR Assumptions 1-5, the variance of the OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are

$$\mathbb{V}\left[\widehat{\beta}_0 \mid x_i\right] = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n \left(x_i - \overline{X}\right)^2}.$$

$$\mathbb{V}\left[\widehat{\beta}_1 \mid x_i\right] = \frac{\sigma^2}{\sum_{i=1}^n \left(x_i - \overline{X}\right)^2}.$$

- As σ^2 increases, the variance of our estimators increase.
- As the SST of x (denominator) increases, the variance of our estimators decrease.
- As n increases the SST of x increases implying the variance of our estimates decreases.

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Estimate Part 1

Recall from Proof 4, Part 2 that

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}.$$

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Estimate Part 1

Recall from Proof 4, Part 2 that

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right) u_i}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}.$$

Taking the variance of both sides of this equation yields

$$\mathbb{V}\left[\widehat{\beta}_1 - \beta_1 \mid x_i\right] = \mathbb{V}\left[\left.\frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right) u_i}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}\right| x_i\right].$$

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Estimate Part 2

Inspecting the left hand side,

$$\mathbb{V}\left[\widehat{\beta}_1 - \beta_1 \mid x_i\right] = \mathbb{V}\left[\widehat{\beta}_1 \mid x_i\right] + \mathbb{V}[\beta_1 \mid x_i] - 2\text{Cov}\left[\widehat{\beta}_1, \beta_1 \mid x_i\right].$$

Since we treat β_1 as a constant,

$$\mathbb{V}\left[\widehat{\beta}_1 - \beta_1 \mid x_i\right] = \mathbb{V}\left[\widehat{\beta}_1 \mid x_i\right].$$

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Term Estimate Part 3

Inspecting the right hand side from Part 1,

$$\mathbb{V}\left[\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\middle|x_i\right]=\frac{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\right]^2\mathbb{V}[u_i\mid x_i]}{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right]^2}$$

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Term Estimate Part 3

Inspecting the right hand side from Part 1,

$$\mathbb{V}\left[\left.\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\right|x_i\right]=\frac{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\right]^2\mathbb{V}[u_i\mid x_i]}{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right]^2}$$

$$=\frac{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\right]^2\sigma^2}{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right]^2}$$

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Term Estimate Part 3

Inspecting the right hand side from Part 1,

$$\begin{aligned}
\mathbb{V}\left[\left.\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\right|x_i\right] &= \frac{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\right]^2\mathbb{V}[u_i \mid x_i]}{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right]^2} \\
&= \frac{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\right]^2\sigma^2}{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right]^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}.
\end{aligned}$$

Variance of OLS Estimates

Proof 6: Variance of OLS Slope Term Estimate Part 4

Putting the right and left hand sides together, we have

$$\mathbb{V}\left[\widehat{\beta}_1 \mid x_i\right] = \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}.$$

Hooray!

- By the law of total variance, $\mathbb{V}\left[\widehat{\beta}_1 \mid x_i\right] = \mathbb{V}\left[\widehat{\beta}_1\right]$.
- This is the true conditional variance of our estimator $\widehat{\beta}_1$.
  - How do we obtain the estimate of this conditional variance?

Conditional Variance of Error Term

Property 11: Conditional Variance of Error Term

Under SLR Assumptions 1-5, the conditional variance of u_i for each $i = 1, \ldots, n$ can be written as

$$\begin{aligned}\sigma^2 &= \mathbb{V}\left[u_i \mid x_i\right] \\ &= \mathbb{E}\left[u_i^2 \mid x_i\right] - \mathbb{E}\left[u_i \mid x_i\right]^2 \\ &= \mathbb{E}\left[u_i^2 \mid x_i\right].\end{aligned}$$

- If we had data on u_i, an unbiased estimate of σ^2 would then be $\frac{1}{n}\sum_{i=1}^{n} u_i^2$.
  - We only have data on $\widehat{u}_i$ though. What do we do?

Unbiased Estimator of the Variance of the Error Term

Theorem 4: Unbiased Estimator of the Variance of the Error Term

An unbiased estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{SSR}{n-k}.$$

where k is the number of estimated parameters (including intercept).

- In the simple linear regression model, $k = 2$.
- The $n-k$ term is called the *degree of freedom correction*.
- $\mathbb{E}[SSR] = (n-k)\sigma^2$.

Estimator of the Variance of the OLS Slope Term Estimator

Theorem 5: Estimator of the Variance of the OLS Slope Term Estimator

Under SLR Assumptions 1-5, an unbiased estimator of $\mathbb{V}\left[\widehat{\beta}_1 \mid x_i\right]$ is

$$\widehat{\mathbb{V}}\left[\widehat{\beta}_1 \mid x_i\right] = \frac{\widehat{\sigma}^2}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}.$$

Standard Errors

Definition 9: Standard Errors of OLS Estimates

The standard errors of our OLS estimators are

$$\text{se}\left[\widehat{\beta}_0 \mid x_i\right] = \sqrt{\widehat{\mathbb{V}}\left[\widehat{\beta}_0 \mid x_i\right]}$$
$$\text{se}\left[\widehat{\beta}_1 \mid x_i\right] = \sqrt{\widehat{\mathbb{V}}\left[\widehat{\beta}_1 \mid x_i\right]}.$$

- The standard error of an estimator is simply its estimated standard deviation.

Thank You!