# Simple Econometric Theory Review

## Kevin's Econometrics Resources

Kevin Li

# Model Specification

For independent variables $X_1, X_2, \dots, X_p$, and outcome variable $Y$ for units $i = 1, 2, \dots, n$:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

$\beta_0, \beta_1, \dots, \beta_p$ are parameters that describe the deterministic part of the relationship between $Y$ and $X_1, \dots, X_p$.

- Read: the part of $Y$ explained by $X_1, \dots, X_p$.

$\varepsilon_i$ error term is the non-deterministic relationship between $Y$ and $X_1, \dots, X_p$.

- Read: part of $Y$ **not** explained by $X_1, \dots, X_p$.
- $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$

Additional assumptions will be imposed on the model later.

# Matrix Form

Condensed form:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \quad \mathbf{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \end{pmatrix}$$

Even more condensed matrix form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

# Sum of Squared Errors

Naturally, we want to choose the values $b_0, \dots, b_p$ for the unkown $\beta_0, \dots, \beta_p$ that **minimise** the sum (squared) error of predicted $\hat{Y}_i$ in respect to the true population.

- Actual true Y values: $Y_i$, with unknown $\beta$
- Predicted $\hat{Y}$ values, with some choice of $\beta$ value of b.

Thus, the sum (squared) error is the sum of the differences between actual $Y_i$ and predicted $\hat{Y}_i$:

$$\text{SSE} = \sum (Y_i - \hat{Y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$$

Why squared?

1. gets rid of direction, only keeps magnitude
2. Easier for calculus as absolute value function is non-differentiable at vertex.
3. Nice properties (see later in the slides).

# Ordinary Least Squares (1)

We want to minimise sum of squared errors to find the values of $\beta$

$$\text{SSE} = \sum (Y_i - \hat{Y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$$

Remember predictions $\hat{\mathbf{y}}$ use the values of vector $\mathbf{b}$ for the unknown $\boldsymbol{\beta}$. That means $\hat{\mathbf{y}} = \mathbf{Xb}$

- (This is because we just plug $\mathbf{b} = \boldsymbol{\beta}$ into the linear regression, and $\boldsymbol{\varepsilon} = 0$ since $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$).

Knowing $\hat{\mathbf{y}} = \mathbf{Xb}$, we can plug it in to the SSE and expand.

$$\begin{aligned}
\text{SSE} &= (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\
&= (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Xb} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}
\end{aligned}$$

# Ordinary Least Squares (2)

We want to minimise the SSE, so take the derivative in respect to $\mathbf{b}$ and set equal to 0:

$$\frac{\partial \text{SSE}}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = 0$$

Re-arrange the equation to get

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \quad \implies \quad \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

This means this value of $\mathbf{b}$ is the value of $\beta$ that minimises the error of the predicted $\hat{\mathbf{y}}$. Thus, this value is the **ordinary least squares estimate** of the unkown $\beta$:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

# Estimator Properties

When we estimate $\beta$ (or any parameter), we typically use a sample of the population.

- What if we used a different sample to calculate the parameter? We would get a slightly different $\hat{\beta}$ estimate since the sample data is slightly different.

**Sampling distribution** is the distribution of all estimated $\hat{\beta}$ from different samples, taking an infinite number of samples.

- Imagine you take one sample, and estimate $\hat{\beta}$. Then, take another sample and estimate $\hat{\beta}$. Then again and again. Plot all of the $\hat{\beta}$ in a distribution to get the sampling distribution.

**Unbiasedness** is if the expected value of the sampling distribution equals the true population value of $\beta$. In other words: $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

**Standard Error** is the standard deviation of the sampling distribution.

# Unbiasedness of OLS (1)

Theorem: Part of the **Gauss-Markov Theorem** states that under 4 conditions, the OLS estimate of $\beta$ is **unbiased**: $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$

1. The population model can be expressed as a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
2. Random Sampling of sample from the population.
3. No perfect multicollinearity - i.e. no exact linear correlations between any two independent variables $X_1, X_2, \ldots, X_p$, or any linear combinations of the two. Basically, $\mathbf{X}$ must be full-rank.
4. **Strict Exogeneity**: Formally defined as $\mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$. This implies that $\mathrm{Cov}(\varepsilon, X_j) = 0$ for any explanatory variable $X_j = X_1, \ldots, X_p$.

Violations of strict exogeneity often caused by omitted confoudners (see causal frameworks).

# Unbiasedness of OLS (2)

Proof: start with our OLS solution.

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

We know that $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ (linear model). So plug in and simplify.

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon
\end{aligned}$$

Now we want to prove $\mathbb{E}[\hat{\beta}] = \beta$. So we want to take the expected value of $\hat{\beta}$:

$$\mathbb{E}[\hat{\beta}|\mathbf{X}] = \mathbb{E}[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon|\mathbf{X}]$$

# Unbiasedness of OLS (3)

1. $\beta$ (true population value) is a constant, and the expected value of a constant is itself.
2. $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is also a constant (sample data values). We know that the expected value of a constant times a variable, equals the constant times the expected value of the variable.

Thus we can rewrite the above to:

$$\mathbb{E}[\hat{\beta}|\mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\varepsilon|\mathbf{X}]$$

Recall Gauss-Markov condition (4), strict exogeneity: $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$. Thus:

$$\mathbb{E}[\hat{\beta}|X] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(0) = \beta$$

Finally, law of iterated expecations (LIE) gets us:

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[\mathbb{E}[\hat{\beta}]] = \beta$$

# Variance of OLS (1)

# Variance of OLS (2)

## Hypothesis Testing