ECON 351* -- NOTE 3

Desirable Statistical Properties of Estimators

1. Introduction

We derived in *Note 2* the OLS (Ordinary Least Squares) estimators $\hat{\beta}_j$ (j = 1, 2) of the regression coefficients β_j (j = 1, 2) in the simple linear regression model given by the **population regression equation**, or PRE

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{1}$$

where u_i is an iid random error term.

The **OLS estimators of the regression coefficients β_1 and β_2** are:

$$\hat{\beta}_2 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X - \bar{X})^2} = \frac{\sum_i X_i Y_i - N\bar{X}\bar{Y}}{\sum_i X_i^2 - N\bar{X}^2}. \tag{2.2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}. \tag{2.1}$$

where

$$y_i \equiv Y_i - \bar{Y} \qquad (i = 1, ..., N)$$
$$x_i \equiv X_i - \bar{X} \qquad (i = 1, ..., N)$$
$$\bar{Y} = \Sigma_i Y_i / N = \frac{\Sigma_i Y_i}{N} = \text{ the sample mean of the } Y_i \text{ values.}$$
$$\bar{X} = \Sigma_i X_i / N = \frac{\Sigma_i X_i}{N} = \text{ the sample mean of the } X_i \text{ values.}$$

The reason we use these **OLS coefficient estimators** is that, **under assumptions A1-A8 of the classical linear regression model**, they **have a number of desirable statistical properties**.

Note 4 outlines these desirable statistical properties of the OLS coefficient estimators.

Why are statistical properties of estimators important? These statistical properties are extremely important because they provide **criteria for choosing among alternative estimators**. Knowledge of these properties is therefore essential to understanding why we use the estimation procedures we do in econometric analysis.

2. Distinction Between an "Estimator" and an "Estimate"

Recall the distinction between an **estimator** and an **estimate**. Let some unknown population parameter be denoted in general as θ, and let an ***estimator* of θ** be denoted as $\hat{\theta}$.

3. Two Categories of Statistical Properties

There are ***two categories of statistical properties.***

(1) ***Small-sample*****, or** ***finite-sample*****, properties**;

The most fundamental ***desirable small-sample properties*** of an estimator are:

S1. Unbiasedness;

S2. Minimum Variance;

S3. Efficiency.

(2) ***Large-sample*****, or** ***asymptotic*****, properties**.

The most important ***desirable large-sample property*** of an estimator is:

L1. Consistency.

Both sets of statistical properties refer to the **properties of the *sampling distribution*, or *probability distribution*, of the estimator $\hat{\theta}$ for different sample sizes**.

3.1 Small-Sample (Finite-Sample) Properties

- The ***small-sample, or finite-sample, properties*** **of the estimator** $\hat{\theta}$ refer to **the properties of the sampling distribution of** $\hat{\theta}$ ***for any sample of fixed size n***, where ***n*** **is a** ***finite*** **number** (i.e., a number less than infinity) denoting the number of observations in the sample.

  n = number of sample observations, **where** $n < \infty$.

  Definition: The **sampling distribution of** $\hat{\theta}$ ***for any finite sample size n*** $< \infty$ is called the ***small-sample, or finite-sample, distribution*** **of the estimator** $\hat{\theta}$. In fact, there is a family of finite-sample distributions for the estimator $\hat{\theta}$, one for each finite value of n.

- The **sampling distribution of** $\hat{\theta}$ is based on the **concept of** ***repeated sampling***.

  - Suppose a **large number of samples of size n** are randomly selected from some underlying population.
    - Each of these samples **contains** ***n*** **observations**.
    - Each of these samples in general **contains** ***different*** **sample values** of the observable random variables that enter the formula for the estimator $\hat{\theta}$.
  - **For each of these samples of n observations**, the **formula for** $\hat{\theta}$ is used to **compute a** ***numerical estimate*** of the population parameter θ.
    - Each sample yields a ***different*** **numerical** ***estimate*** **of the unknown parameter** θ.
      Why? Because each sample typically contains different sample values of the observable random variables that enter the formula for the estimator $\hat{\theta}$.
  - If we tabulate or plot these different sample estimates of the parameter θ for a very large number of samples of size n, we obtain the ***small-sample, or finite-sample, distribution*** **of the estimator** $\hat{\theta}$.

3.2 Large-Sample (Asymptotic) Properties

- The ***large-sample*, or *asymptotic*, properties of the estimator** $\hat{\theta}$ refer to **the properties of the sampling distribution of** $\hat{\theta}$ ***as the sample size n becomes indefinitely large***, i.e., as sample size n approaches infinity (**as** $\mathbf{n \to \infty}$).

  Definition: The probability distribution to which the sampling distribution of $\hat{\theta}$ converges ***as sample size n becomes indefinitely large*** (i.e., as $n \to \infty$) is called the ***asymptotic, or large-sample, distribution*** **of the estimator** $\hat{\theta}$.

- The **properties of the *asymptotic* distribution of** $\hat{\theta}$ are what we call the ***large-sample*, or *asymptotic*, *properties*** **of the estimator** $\hat{\theta}$.

- To define all the ***large-sample*** **properties** of an estimator, we need to distinguish between ***two*** **large-sample distributions** of an estimator:

  1. the ***asymptotic*** **distribution** of the estimator;

  2. the ***ultimate*, or *final*, distribution** of the estimator.

4. Small-Sample Properties

Nature of Small-Sample Properties

- The ***small-sample, or finite-sample, distribution*** of the estimator $\hat{\theta}$ for any finite sample size $n < \infty$ has

  1. a ***mean***, or expectation, denoted as $E(\hat{\theta})$, and
  2. a ***variance*** denoted as $Var(\hat{\theta})$.

- The ***small-sample*** **properties of the estimator $\hat{\theta}$** are defined in terms of the ***mean*** **$E(\hat{\theta})$** and the ***variance*** **$Var(\hat{\theta})$** of the ***finite-sample*** **distribution of the estimator $\hat{\theta}$**.

S1: Unbiasedness

Definition of Unbiasedness: The **estimator** $\hat{\theta}$ is an ***unbiased*** **estimator** of the population parameter θ if the mean or expectation of the finite-sample distribution of $\hat{\theta}$ is equal to the true θ. That is, $\hat{\theta}$ **is an *unbiased* estimator of θ** if

$$\mathbf{E(\hat{\theta}) = \theta} \qquad \text{for any given finite sample size } n < \infty.$$

Definition of the Bias of an Estimator: The ***bias*** **of the estimator** $\hat{\theta}$ is defined as

$$\mathbf{Bias(\hat{\theta}) = E(\hat{\theta}) - \theta}$$ = the ***mean*** **of** $\hat{\theta}$ *minus* the ***true value*** **of θ**.

- The estimator $\hat{\theta}$ **is an *unbiased* estimator** of the population parameter θ if the bias of $\hat{\theta}$ is equal to zero; i.e., if

$$\mathbf{Bias(\hat{\theta}) = E(\hat{\theta}) - \theta = 0} \quad \Leftrightarrow \quad \mathbf{E(\hat{\theta}) = \theta}.$$

- Alternatively, the estimator $\hat{\theta}$ **is a *biased* estimator** of the population parameter θ if the bias of $\hat{\theta}$ is non-zero; i.e., if

$$\mathbf{Bias(\hat{\theta}) = E(\hat{\theta}) - \theta \neq 0} \quad \Leftrightarrow \quad \mathbf{E(\hat{\theta}) \neq \theta}.$$

1. The estimator $\hat{\theta}$ **is an *upward biased*** (or **positively biased**) **estimator** the population parameter θ if the **bias of $\hat{\theta}$ is *greater than zero***; i.e., if

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta > 0 \quad \Leftrightarrow \quad E(\hat{\theta}) > \theta.$$

2. The estimator $\hat{\theta}$ is a ***downward biased*** (or **negatively biased**) **estimator** of the population parameter θ if the **bias of $\hat{\theta}$ is *less than zero***; i.e., if

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta < 0 \quad \Leftrightarrow \quad E(\hat{\theta}) < \theta.$$

Meaning of the Unbiasedness Property

- The estimator $\hat{\theta}$ is an unbiased estimator of θ if on average it equals the true parameter value θ.
  - This means that on average the estimator $\hat{\theta}$ is correct, even though any single estimate of θ for a particular sample of data may not equal θ.
  - More technically, it means that **the finite-sample distribution of the estimator $\hat{\theta}$ is *centered* on the value θ**, not on some other real value.

- The ***bias*** **of an estimator** is an ***inverse*** **measure** of its **average accuracy**.
  - The smaller in absolute value is $\text{Bias}(\hat{\theta})$, the more accurate on average is the estimator $\hat{\theta}$ in estimating the population parameter θ.
  - Thus, an unbiased estimator for which $\text{Bias}(\hat{\theta}) = 0$ -- that is, for which $E(\hat{\theta}) = \theta$ -- is *on average* a perfectly accurate estimator of θ.

- Given a choice between two estimators of the same population parameter θ, of which one is biased and the other is unbiased, we prefer the unbiased estimator because it is more accurate on average than the biased estimator.

S2: Minimum Variance

Definition of Minimum Variance: The **estimator $\hat{\theta}$ is a *minimum-variance* estimator** of the population parameter θ if the variance of the finite-sample distribution of $\hat{\theta}$ is ***less than or equal to*** the variance of the finite-sample distribution of $\tilde{\theta}$, where $\tilde{\theta}$ is ***any other estimator*** of the population parameter θ; i.e., if

$$\mathbf{Var}(\hat{\theta}) \le \mathbf{Var}(\tilde{\theta}) \text{ for } \textbf{all } \textbf{\textit{finite}} \textbf{ sample sizes n} \text{ such that } 0 < n < \infty$$

where

$$\mathbf{Var}(\hat{\theta}) = \mathbf{E}\left[\hat{\theta} - \mathbf{E}(\hat{\theta})\right]^2 = \text{ the } \textbf{variance of the estimator } \hat{\theta};$$

$$\mathbf{Var}(\tilde{\theta}) = \mathbf{E}\left[\tilde{\theta} - \mathbf{E}(\tilde{\theta})\right]^2 = \text{ the } \textbf{variance of } \textbf{\textit{any other estimator}} \ \tilde{\theta}.$$

Note: Either or both of the estimators $\hat{\theta}$ and $\tilde{\theta}$ may be biased. The minimum variance property implies nothing about whether the estimators are biased or unbiased.

Meaning of the Minimum Variance Property

- The ***variance*** **of an estimator** is an ***inverse*** **measure** of its **statistical precision**, i.e., of its dispersion or spread around its mean.

  The ***smaller*** **the variance** of an estimator, the ***more*** **statistically precise** it is.

- A ***minimum variance*** **estimator** is therefore the statistically ***most precise*** **estimator** of an unknown population parameter, although it may be biased or unbiased.

S3: Efficiency

A Necessary Condition for Efficiency -- Unbiasedness

The small-sample property of efficiency is defined only for ***unbiased*** **estimators**.

Therefore, a ***necessary condition for efficiency*** **of the estimator** $\hat{\theta}$ is that $\mathbf{E(\hat{\theta}) = \theta}$, i.e., $\hat{\theta}$ must be an ***unbiased*** **estimator of the population parameter** θ.

Definition of Efficiency: Efficiency = Unbiasedness + Minimum Variance

Verbal Definition: If $\hat{\theta}$ and $\tilde{\theta}$ are two unbiased estimators of the population parameter θ, then the estimator $\hat{\theta}$ is efficient relative to the estimator $\tilde{\theta}$ if the variance of $\hat{\theta}$ is smaller than the variance of $\tilde{\theta}$ for any finite sample size $n < \infty$.

Formal Definition: Let $\hat{\theta}$ **and** $\tilde{\theta}$ be two ***unbiased*** **estimators of the population parameter θ**, such that $E(\hat{\theta}) = \theta$ and $E(\tilde{\theta}) = \theta$. Then **the estimator** $\hat{\theta}$ **is *efficient relative to* the estimator** $\tilde{\theta}$ if the variance of the finite-sample distribution of $\hat{\theta}$ is less than or at most equal to the variance of the finite-sample distribution of $\tilde{\theta}$; i.e. if

$$\mathbf{Var(\hat{\theta}) \le Var(\tilde{\theta})} \textbf{ for all finite n} \text{ where } \mathbf{E(\hat{\theta}) = \theta} \text{ and } \mathbf{E(\tilde{\theta}) = \theta}.$$

> ***Note:*** **Both the estimators** $\hat{\theta}$ **and** $\tilde{\theta}$ **must be *unbiased***, since the efficiency property refers only to the variances of unbiased estimators.

Meaning of the Efficiency Property

- Efficiency is a desirable statistical property because of two unbiased estimators of the same population parameter, we prefer the one that has the smaller variance, i.e., the one that is statistically more precise.

- In the above definition of efficiency, **if** $\tilde{\theta}$ **is *any* other *unbiased* estimator** of the population parameter θ, then **the estimator** $\hat{\theta}$ **is the *best unbiased***, or ***minimum-variance unbiased***, **estimator of θ**.

5. Large-Sample Properties

Nature of Large-Sample Properties

- The large-sample properties of an estimator are the properties of the sampling distribution of that estimator **as sample size n becomes very large**, as n approaches infinity, **as** $\mathbf{n \to \infty}$.

- Recall that **the *sampling distribution* of an estimator differs for different sample sizes** -- i.e., **for different values of n**.

  The sampling distribution of a given estimator for one sample size is different from the sampling distribution of that same estimator for some other sample size.

  Consider the estimator $\hat{\theta}$ for **two different values of n**, $\mathbf{n_1}$ and $\mathbf{n_2}$.

  In general, **the sampling distributions of** $\hat{\theta}_{n_1}$ **and** $\hat{\theta}_{n_2}$ **are *different***: they can have

  - ***different*** **means**
  - ***different*** **variances**
  - ***different*** **mathematical forms**

Desirable Large-Sample Properties

L1. **Consistency**;

L2. **Asymptotic Unbiasedness**;

L3. **Asymptotic Efficiency**.

We need to distinguish between ***two*** **large-sample distributions** of an estimator:

1. the ***asymptotic*** **distribution** of the estimator;

2. the ***ultimate*, or final, distribution** of the estimator.

The Ultimate Distribution of an Estimator

- The asymptotic distribution of an estimator $\hat{\theta}$ is not necessarily the final or ultimate form that the sampling distribution of the estimator takes as sample size n approaches infinity (as $n \to \infty$).

- For many estimators, **the sampling distribution *collapses to a single point* as sample size n approaches infinity**.

  - More specifically, **as sample size n → ∞**, the sampling distribution of the estimator ***collapses* to a column of *unit probability mass*** (or unit probability density) **at a single point** on the real line.
  - Such an estimator is said to ***converge in probability*** to some value.
  - A distribution that is completely concentrated at one point (on the real line) is called a ***degenerate* distribution**. Graphically, a degenerate distribution is represented by **a perpendicular to the real line with height equal to one**.

Distinguishing Between an Estimator's Asymptotic and Ultimate Distributions

- The ***ultimate* distribution** of an estimator is the distribution to which the sampling distribution of the estimator finally converges as sample size n "reaches" infinity.

- The ***asymptotic* distribution** of an estimator is the distribution to which the sampling distribution of the estimator converges just before it collapses (if indeed it does) as sample size n "approaches" infinity.

  - **If the *ultimate* distribution of an estimator is *degenerate***, then its ***asymptotic* distribution** is **<u>not identical</u>** to its ***ultimate* (final) distribution**.

  - But **if the *ultimate* distribution of an estimator is *non-degenerate***, then its ***asymptotic* distribution** is **<u>identical</u>** to its ***ultimate* (final) distribution**.

L1: Consistency

A Necessary Condition for Consistency

Let $\hat{\theta}_n$ be **an estimator of the population parameter θ based on a sample of size n observations**.

A ***necessary condition for consistency*** **of the estimator** $\hat{\theta}_n$ is that **the *ultimate* distribution of** $\hat{\theta}_n$ **be a *degenerate* distribution** at some point on the real line, meaning that it converges to a single point on the real line.

The Probability Limit of an Estimator -- the **plim** of an estimator

Short Definition: The probability limit of an estimator $\hat{\theta}_n$ is the value -- or point on the real line -- to which that estimator's sampling distribution converges as sample size n increases without limit.

Long Definition: If the **estimator** $\hat{\theta}_n$ **has a *degenerate ultimate* distribution** -- i.e., if the sampling distribution of $\hat{\theta}_n$ collapses on a single point as sample size $n \to \infty$ -- then **that point on which the sampling distribution of** $\hat{\theta}_n$ **converges** is called the ***probability limit*** **of** $\hat{\theta}_n$, and is denoted as $\text{plim}\,\hat{\theta}_n$ or $\underset{n\to\infty}{\text{plim}}\,\hat{\theta}_n$ where "plim" means "probability limit".

Formal Definition of Probability Limit: The **point θ_0** on the real line is the ***probability limit* of the estimator $\hat{\theta}_n$** if the ***ultimate* sampling distribution of** $\hat{\theta}_n$ **is *degenerate* at the point θ_0**, meaning that the sampling distribution of $\hat{\theta}_n$ collapses to a column of unit density on the point θ_0 as sample size $n \rightarrow \infty$.

- This definition can be written concisely as

$$\text{plim}\ \hat{\theta} = \theta_0 \qquad \textit{or} \qquad \underset{n\rightarrow\infty}{\text{plim}}\ \hat{\theta}_n = \theta_0 .$$

- The three statements above are equivalent to the statement

$$\lim_{n\rightarrow\infty} \Pr\left(\theta_0 - \varepsilon \le \hat{\theta}_n \le \theta_0 + \varepsilon\right) = \lim_{n\rightarrow\infty} \Pr\left(-\varepsilon \le \hat{\theta}_n - \theta_0 \le +\varepsilon\right) = \lim_{n\rightarrow\infty} \Pr\left(\left|\hat{\theta}_n - \theta_0\right| \le \varepsilon\right) = 1$$

  where $\varepsilon > 0$ is an arbitrarily small positive number and $\left|\hat{\theta}_n - \theta_0\right|$ denotes the absolute value of the difference between the estimator $\hat{\theta}_n$ and the point θ_0.

  1. This statement means that as sample size $n \rightarrow \infty$, the probability that the estimator $\hat{\theta}_n$ is arbitrarily close to the point θ_0 approaches one.

  2. The more technical way of saying this is that **the estimator $\hat{\theta}_n$ *converges in probability* to the point θ_0**.

Definition of Consistency

Verbal Definition: **The estimator $\hat{\theta}_n$ is a *consistent* estimator** of the population parameter θ if its sampling distribution collapses on, or converges to, the value of the population parameter θ as $n \to \infty$.

Formal Definition: **The estimator $\hat{\theta}_n$ is a *consistent* estimator of the population parameter θ if the *probability limit of* $\hat{\theta}_n$ is θ**, i.e., if

$$\underset{n\to\infty}{\text{plim}}\, \hat{\theta}_n = \theta \quad or \quad \lim_{n\to\infty} \Pr\left(\left|\hat{\theta}_n - \theta\right| \le \varepsilon\right) = 1 \quad or \quad \Pr\left(\left|\hat{\theta}_n - \theta\right| \le \varepsilon\right) \to 1 \text{ as } n \to \infty.$$

The estimator $\hat{\theta}_n$ **is a *consistent* estimator of the population parameter θ**

- if **the probability that $\hat{\theta}_n$ is arbitrarily close to θ approaches 1** as the sample size $\mathbf{n \to \infty}$

or

- if **the estimator $\hat{\theta}_n$ *converges in probability* to the population parameter θ**.

Meaning of the Consistency Property

- As sample size n becomes larger and larger, the sampling distribution of $\hat{\theta}_n$ becomes more and more concentrated around θ.

- As sample size n becomes larger and larger, the value of $\hat{\theta}_n$ is more and more likely to be very close to θ.

A Sufficient Condition for Consistency

One way of determining if the estimator $\hat{\theta}_n$ is consistent is to trace the behavior of the sampling distribution of $\hat{\theta}_n$ as sample size n becomes larger and larger.

- If as $n \to \infty$ (sample size n approaches infinity) both the **bias** of $\hat{\theta}_n$ and the **variance** of $\hat{\theta}_n$ ***approach zero***, then $\hat{\theta}_n$ is a ***consistent* estimator of the parameter θ**.

- Recall that the bias of $\hat{\theta}_n$ is defined as

$$\text{Bias}\left(\hat{\theta}_n\right) = E\left(\hat{\theta}_n\right) - \theta.$$

Thus, the bias of $\hat{\theta}_n$ approaches zero as $n \to \infty$ if and only if the mean or expectation of the sampling distribution of $\hat{\theta}_n$ approaches θ as $n \to \infty$:

$$\lim_{n\to\infty} \text{Bias}\left(\hat{\theta}_n\right) = \lim_{n\to\infty} E\left(\hat{\theta}_n\right) - \theta = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta.$$

- ❑ **<u>Result</u>: A *sufficient condition for consistency* of the estimator $\hat{\theta}_n$** is that

$$\lim_{n\to\infty} \text{Bias}\left(\hat{\theta}_n\right) = 0 \quad \textit{or} \quad \lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta \qquad \textit{and} \qquad \lim_{n\to\infty} \text{Var}\left(\hat{\theta}_n\right) = 0.$$

This condition states that **if both the *bias* and *variance* of the estimator $\hat{\theta}_n$ *approach zero* as sample size $n \to \infty$, then $\hat{\theta}_n$ is a *consistent* estimator of θ.**