Desirable Statistical Properties of Estimators

1. Two Categories of Statistical Properties

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

(1) *Small-sample*, or *finite-sample*, properties of estimators

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

S1. Unbiasedness

S2. Minimum Variance

S3. Efficiency

(2) *Large-sample*, or *asymptotic*, properties of estimators

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 $\hat{\beta}_j$ for different sample sizes**.

1.1 Small-Sample (Finite-Sample) Properties

- The ***small-sample, or finite-sample, properties*** **of the estimator** $\hat{\beta}_j$ refer to **the properties of the sampling distribution of** $\hat{\beta}_j$ ***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* <** ∞.

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

- The **sampling distribution of $\hat{\beta}_j$** 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 $\hat{\beta}_j$.

  - **For each of these samples of N observations**, the **formula for $\hat{\beta}_j$** is used to **compute a *numerical estimate*** of the population parameter β_j.

    - Each sample yields a ***different* numerical *estimate* of the unknown parameter β_j**.
      Why? Because each sample typically contains different sample values of the observable random variables that enter the formula for the estimator $\hat{\beta}_j$.

  - If we tabulate or plot these different sample estimates of the parameter β_j for a very large number of samples of size N, we obtain the ***small-sample, or finite-sample, distribution* of the estimator** $\hat{\beta}_j$.

1.2 Large-Sample (Asymptotic) Properties

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

  Definition: The probability distribution to which the sampling distribution of $\hat{\beta}_j$ converges ***as sample size N becomes indefinitely large*** (i.e., as $N \to \infty$) is called the ***ultimate distribution*** **of the estimator** $\hat{\beta}_j$.

2. Small-Sample Estimator Properties

Nature of Small-Sample Properties

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

    1. a ***mean***, or expectation, denoted as $E(\hat{\beta}_j)$, and

    2. a ***variance*** denoted as $Var(\hat{\beta}_j)$.

- The ***small-sample*** **properties of the estimator** $\hat{\beta}_j$ are defined in terms of the ***mean*** $\mathbf{E}(\hat{\beta}_j)$ and the ***variance*** $\mathbf{Var}(\hat{\beta}_j)$ of the ***finite-sample*** **distribution of the estimator** $\hat{\beta}_j$ **for any finite sample size** $N < \infty$**.**

S1: Unbiasedness

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

$$\mathbf{E}(\hat{\beta}_j) = \beta_j \qquad \text{for any given finite sample size } N < \infty.$$

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

$$\mathbf{Bias}(\hat{\beta}_j) = \mathbf{E}(\hat{\beta}_j) - \beta_j$$ = the ***mean* of** $\hat{\beta}_j$ *minus* the ***true value* of β_j**.

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

$$\text{Bias}(\hat{\beta}_j) = E(\hat{\beta}_j) - \beta_j = 0 \quad \Leftrightarrow \quad E(\hat{\beta}_j) = \beta_j.$$

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

$$\text{Bias}(\hat{\beta}_j) = E(\hat{\beta}_j) - \beta_j \neq 0 \quad \Leftrightarrow \quad E(\hat{\beta}_j) \neq \beta_j.$$

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

$$\text{Bias}\left(\hat{\beta}_j\right) = E\left(\hat{\beta}_j\right) - \beta_j > 0 \quad \Leftrightarrow \quad E\left(\hat{\beta}_j\right) > \beta_j.$$

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

$$\text{Bias}\left(\hat{\beta}_j\right) = E\left(\hat{\beta}_j\right) - \beta_j < 0 \quad \Leftrightarrow \quad E\left(\hat{\beta}_j\right) < \beta_j.$$

Meaning of the Unbiasedness Property

- The estimator $\hat{\beta}_j$ is an unbiased estimator of β_j if ***on average*** **it *equals* the true parameter value β_j**.
  - This means that on average the estimator $\hat{\beta}_j$ is correct, even though any single estimate of β_j for a particular sample of data may not equal β_j.
  - More technically, it means that **the finite-sample distribution of the estimator $\hat{\beta}_j$ is *centered* on the value β_j**, 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 $\text{Bias}(\hat{\beta}_j)$, the more accurate on average is the estimator $\hat{\beta}_j$ in estimating the population parameter β_j.
  - Thus, **an *unbiased* estimator** for which $\text{Bias}(\hat{\beta}_j) = 0$ -- that is, for which $E(\hat{\beta}_j) = \beta_j$ -- **is *on average* a perfectly accurate estimator of β_j**.
- Given a choice between two estimators of the same population parameter β_j, 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 $\hat{\beta}_j$ is a *minimum-variance* estimator** of the population parameter β_j if **the *variance* of the finite-sample distribution of $\hat{\beta}_j$** is ***less than or equal to*** the ***variance* of the finite-sample distribution of $\tilde{\beta}_j$**, where $\tilde{\beta}_j$ **is *any other estimator* of the population parameter β_j**; i.e., if

$$\mathbf{Var}\left(\hat{\beta}_j\right) \le \mathbf{Var}\left(\tilde{\beta}_j\right) \text{ for all finite sample sizes N such that } 0 < N < \infty$$

where

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \mathrm{E}\left[\hat{\beta}_j - \mathrm{E}\left(\hat{\beta}_j\right)\right]^2 = \text{ the variance of the estimator } \hat{\beta}_j;$$

$$\mathrm{Var}\left(\tilde{\beta}_j\right) = \mathrm{E}\left[\tilde{\beta}_j - \mathrm{E}\left(\tilde{\beta}_j\right)\right]^2 = \text{ the variance of any other estimator } \tilde{\beta}_j.$$

Note: Either or both of the estimators $\hat{\beta}_j$ and $\tilde{\beta}_j$ 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** $\hat{\beta}_j$ is that $E(\hat{\beta}_j) = \beta_j$, i.e., $\hat{\beta}_j$ must be an ***unbiased* estimator of the population parameter** β_j.

Definition of Efficiency: Efficiency = Unbiasedness + Minimum Variance

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

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

$$\text{Var}(\hat{\beta}_j) \le \text{Var}(\tilde{\beta}_j) \textbf{ for all finite N} \text{ where } E(\hat{\beta}_j) = \beta_j \text{ and } E(\tilde{\beta}_j) = \beta_j.$$

Note: **Both the estimators** $\hat{\beta}_j$ **and** $\tilde{\beta}_j$ **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 $\tilde{\beta}_j$ is *any* other *unbiased* estimator** of the population parameter β_j, then **the estimator $\hat{\beta}_j$ is the *best unbiased***, or ***minimum-variance unbiased***, **estimator of β_j.**

3. Large Sample Estimator Properties

L1: <u>Consistency</u> – the minimal requirement of any useful estimator

<u>Definition of Consistency</u>

Verbal Definition: Let $\hat{\beta}_j(N)$ denote an estimator of the population parameter β_j based on a sample of size N observations. **The estimator $\hat{\beta}_j(N)$ is a *consistent* estimator** of the population parameter β_j **if its sampling distribution *collapses on*, or *converges to*, the value of the population parameter β_j as $N \rightarrow \infty$.**

Formal Definition: **The estimator** $\hat{\beta}_j(N)$ **is a *consistent* estimator of the population parameter β_j if the *probability limit of* $\hat{\beta}_j(N)$ is β_j**, i.e., if

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}_j(N) = \beta_j \quad \textit{or} \quad \lim_{N\to\infty} \Pr\left(\left|\hat{\beta}_j(N) - \beta_j\right| \le \varepsilon\right) = 1 \quad \textit{or} \quad \Pr\left(\left|\hat{\beta}_j - \beta_j\right| \le \varepsilon\right) \rightarrow 1 \text{ as } N \rightarrow \infty.$$

The estimator $\hat{\beta}_j(N)$ **is a *consistent* estimator of the population parameter β_j**

- if **the probability that** $\hat{\beta}_j(N)$ **is arbitrarily close to β_j approaches 1** as the sample size $N \rightarrow \infty$

or

- if **the estimator** $\hat{\beta}_j(N)$ ***converges in probability* to the population parameter β_j.**

Intuitive Meaning of the Consistency Property

- As sample size N becomes larger and larger, the sampling distribution of $\hat{\beta}_j(N)$ becomes more and more concentrated around $\boldsymbol{\beta_j}$.

- As sample size N becomes larger and larger, the value of $\hat{\beta}_j(N)$ is more and more likely to be very close to $\boldsymbol{\beta_j}$.

A Sufficient Condition for Consistency

One way of determining if the estimator $\hat{\beta}_j(N)$ is consistent is to trace the behavior of the sampling distribution of $\hat{\beta}_j(N)$ as sample size N becomes larger and larger.

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

- Recall that the bias of $\hat{\beta}_j(N)$ is defined as $\text{Bias}\left(\hat{\beta}_j(N)\right) = \text{E}\left(\hat{\beta}_j(N)\right) - \beta_j$.

  Thus, the bias of $\hat{\beta}_j(N)$ approaches zero as $N \to \infty$ **if and only if the *mean* or *expectation* of the sampling distribution of $\hat{\beta}_j(N)$ approaches β_j** as $N \to \infty$:

$$\lim_{N\to\infty} \text{Bias}\left(\hat{\beta}_j(N)\right) = \lim_{N\to\infty} \text{E}\left(\hat{\beta}_j(N)\right) - \beta_j = 0 \quad \Leftrightarrow \quad \lim_{N\to\infty} \text{E}\left(\hat{\beta}_j(N)\right) = \beta_j.$$

❑ **Result: A *sufficient condition for consistency* of the estimator $\hat{\beta}_j(N)$** is that

$$\lim_{N\to\infty} \text{Bias}\left(\hat{\beta}_j(N)\right) = 0 \quad \textit{or} \quad \lim_{N\to\infty} \text{E}\left(\hat{\beta}_j(N)\right) = \beta_j \quad \textbf{\textit{and}} \quad \lim_{N\to\infty} \text{Var}\left(\hat{\beta}_j(N)\right) = 0.$$

This condition states that **if both the *bias* and *variance* of the estimator** $\hat{\beta}_j(N)$ ***approach zero* as sample size $N \to \infty$ then** $\hat{\beta}_j(N)$ **is a *consistent* estimator of β_j.**