EC114 Introduction to Quantitative Economics

14. Properties of Estimators

Marcus Chambers

Department of Economics
University of Essex

07/09 February 2012

Outline

1. Introduction
2. Unbiased Estimators
3. Efficient Estimators
4. Linear Estimators and Mean Square Error
5. An Example

Reference: R. L. Thomas, *Using Statistics in Economics*, McGraw-Hill, 2005, sections 11.1 and 11.2.

- To provide some background, consider a population of values for a random variable X.
- The variable X will have a probability distribution, which may be known or unkown (typically the latter).
- Suppose this distribution can be characterised by an unknown parameter θ.
- The parameter θ could represent the mean (μ) or variance (σ^2) of the distribution, or it could represent the regression slope parameter (β).
- Whatever it represents, the parameter θ needs to be estimated using sample information.

- Suppose we have a random sample of n observations taken from the population of X:

$$X_1, X_2, \ldots, X_n,$$

where X_i denotes the i'th observation.

- Let the estimator of θ be denoted Q, which will be some function of the observations i.e.

$$Q = Q(X_1, X_2, \ldots, X_n).$$

- For example, if θ were the population mean μ, then our estimator would be the sample mean

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

- Whatever the estimator Q, it will have a *sampling distribution*.
- This is because the value of Q changes with each different possible sample, and the sampling distribution represents the distribution of Q across all possible samples.
- We can therefore talk about quantities such as the mean and variance of the estimator Q i.e.

$$E(Q) \text{ and } E[Q - E(Q)]^2.$$

- In the regression model we have two estimators, a and b, of α and β, the intercept and slope parameters.
- Both a and b will have their own sampling distributions, as well as a joint sampling distribution.

- The *small sample properties of estimators* are determined by the sampling distribution for estimators obtained using a given sample size n.
- Such properties hold even when n may be small.
- This contrasts with *large sample*, or *asymptotic*, properties which are obtained as the sample size n gets larger and larger i.e. as $n \to \infty$.
- You have already seen the Central Limit Theorem, which is an example of an asymptotic property for the sample mean.
- It is often written in the form

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

- We shall not be concerned with asymptotic properties here, only small sample properties.

- A desirable small sample property for an estimator to possess is *unbiasedness*:

Definition

An estimator Q is said to be an *unbiased estimator* of θ if, and only if, $E(Q) = \theta$.

- Put another way, the mean of the sampling distribution of the estimator Q is equal to the true population parameter θ.
- Or, if we were to take many samples, the average of all Q's obtained would be equal to θ.
- This is illustrated on the next slide:

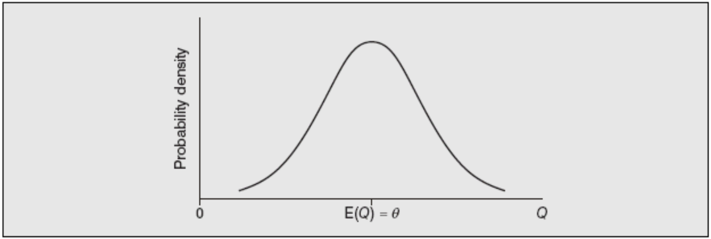

Figure 11.1 An unbiased estimator

- The distribution of Q is centred at θ, meaning that there is no systematic tendency to either over-estimate or under-estimate θ.
- This does not mean that in any given sample the estimator will equal θ (it almost certainly won't)!

- We have already seen that the sample mean is an unbiased estimator of the population mean: $E(\bar{X}) = \mu$.
- Note that the property of unbiasedness does not depend on the sample size; it is therefore a small sample property.
- If an estimator does not satisfy the unbiasedness property it is said to be *biased* and so there *is* a systematic tendency to error in estimating θ.
- The bias of an estimator Q is defined as

$$\text{bias}(Q) = E(Q) - \theta.$$

- If Q tends to over-estimate θ then $E(Q) > \theta$ and $\text{bias}(Q) > 0$.
- Alternatively, if Q tends to under-estimate θ then $E(Q) < \theta$ and $\text{bias}(Q) < 0$.

- An example of a biased estimator is

$$v^2 = \frac{\sum(X_i - \bar{X})^2}{n},$$

which is an estimator of σ^2, the population variance.

- It can be shown that

$$E(v^2) = \left(\frac{n-1}{n}\right)\sigma^2 < \sigma^2,$$

and hence bias$(v^2) < 0$.

- This motivates the unbiased estimator, s^2, of σ^2, in which n in the denominator is replaced by $n - 1$:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1};$$

this results in $E(s^2) = \sigma^2$.

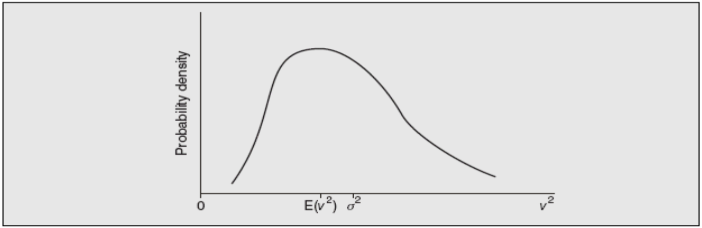

Figure 11.2 The downward bias of v^2

- The negative bias of v^2 is depicted above.
- But is unbiasedness all that we require of an estimator?

- In practice we are faced with using a single sample to determine our estimator Q.
- Although unbiasedness is a good property for Q to possess, we need to consider other aspects of the sampling distribution as well, such as the variance.
- Consider two unbiased estimators, Q_1 and Q_2, so that $E(Q_1) = \theta$ and $E(Q_2) = \theta$.
- Suppose, however, that the variance of Q_1 is larger than the variance of Q_2, so that $V(Q_1) > V(Q_2)$.
- Which estimator would we prefer?
- The diagram on the next slide will help answer this question. . .

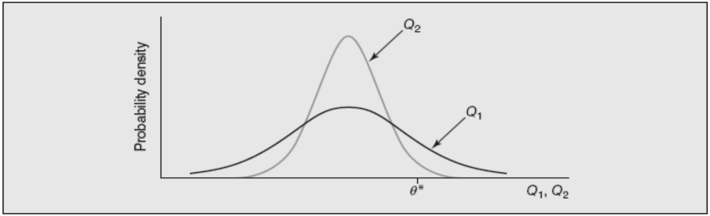

Figure 11.3 Unbiased estimation with $V(Q_1) > V(Q_2)$

- Although both estimators are unbiased, there is a higher probability of being far away from θ using the estimator Q_1 than with Q_2 i.e.

$$\Pr(Q_1 > \theta^*) > \Pr(Q_2 > \theta^*).$$

- We would therefore prefer to use Q_2 whose distribution is more condensed around θ than the distribution of Q_1.
- Another desirable property for an estimator to possess is efficiency:

Definition

An estimator Q is said to be an *efficient estimator* of θ if, and only if:

(i) it is unbiased, so that $E(Q) = \theta$; and

(ii) no other unbiased estimator of θ has a smaller variance.

- The two important properties for efficiency, therefore, are unbiasedness and smallest (or minimum) variance.
- Efficient estimators are sometimes also called *best unbiased estimators*.

- It can be difficult, in practice, to show that an unbiased estimator has the smallest variance among *all* estimators.
- It is often easier to restrict attention to *linear* estimators, that is, ones which are a linear combination of the sample observations.
- A linear estimator is therefore of the form

$$Q = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n,$$

  where $a_1, \ldots, a_n$ are a set of constants (or weights).
- An example of a linear estimator is the sample mean:

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n} = \frac{1}{n} X_1 + \ldots + \frac{1}{n} X_n,$$

  so the weights are $a_1 = 1/n, \ldots, a_n = 1/n$.

Definition

An estimator, Q, is said to be a *best linear unbiased estimator* (BLUE) of θ if, and only if:

(i) it is a linear estimator, that is, $Q = \sum_i a_i X_i$, where the a_i are constants;

(ii) it is unbiased, so that $E(Q) = \theta$; and

(iii) no other linear unbiased estimator has a smaller variance.

- Note that a BLUE estimator may not be the best possible, because there may exist an efficient nonlinear estimator with a smaller variance.
- In terms of estimating the population mean μ, it turns out that the sample mean $\bar{X}$ is both BLUE and efficient.

- The concept of efficiency is useful when comparing unbiased estimators, because we always prefer the one with the smaller variance.
- But suppose we are faced with the following problem.
- There are two estimators, Q_1 and Q_2, of a parameter θ.
- Q_2 is unbiased and efficient while Q_1 has a small (positive) bias but also a smaller variance than Q_2.
- We have $E(Q_1) > \theta$, $E(Q_2) = \theta$, $V(Q_1) < V(Q_2)$.
- Can we compare these estimators in a meaningful way?
- Consider the diagram on the next slide:

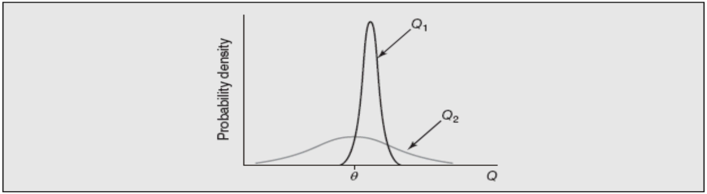

Figure 11.5 Unbiasness versus efficiency

- The distribution of Q_2 is centred at θ but has large dispersion while the distribution of Q_1 lies to the right of θ but with a smaller variance.
- There is a trade-off here between bias and variance.
- Can we combine both in a single measure?

- We shall consider the *error* of the estimator, given by $Q - \theta$.
- The *mean square error* (MSE) of Q is given by

$$\text{MSE}(Q) = E(Q - \theta)^2.$$

- It can be shown that

$$\text{MSE}(Q) = V(Q) + [\text{bias}(Q)]^2,$$

  i.e. the MSE of Q is equal to the variance plus the square of the bias.
- One way of choosing the preferred estimator would be to choose the one that has the smallest MSE, which places equal weight on variance and squared bias.

- A more general method uses weights to reflect the relative importance placed on variance and squared bias.
- We could then choose the estimator as the one which minimises

$$M(Q) = \phi V(Q) + (1 - \phi)[\text{bias}(Q)]^2,$$

where $0 < \phi < 1$ denotes the weight.
- The larger is ϕ, the more importance is placed on variance, while the smaller is ϕ, the more importance is placed on (squared) bias.

- It is usually a good idea, when estimating a parameter, to use all the sample information that is available.
- In general, the more information used, the more efficient the estimator will be.
- For example, it wouldn't make sense to ignore 50% of a sample when estimating the population mean – the more information (observations) the better the estimator.
- An estimator that uses all the sample information is said to be *sufficient*.
- However, if an estimator uses all the observations inappropriately, it will not be efficient despite being sufficient.
- The point is that an estimator cannot be efficient unless it is sufficient.

- **Example** (Thomas, p.327). A variable X is normally distributed with mean μ and variance σ^2. Three estimators of μ are proposed:

$$\hat{m} = \bar{X} - 10, \quad \tilde{m} = \bar{X} + \frac{5}{n}, \quad m^* = \left(\frac{n-1}{n-2}\right)\bar{X},$$

where $\bar{X}$ is the sample mean and n the sample size.

(a) Explain why all three estimators will have sampling distributions that are normal in shape.

(b) Recalling that $E(\bar{X}) = \mu$, use Theorem 1.1 to show that all the proposed estimators are biased and hence determine the bias in each case. If $n = 10$ and $\mu = 8$, which estimator has the smallest absolute bias?

(c) Recalling that $V(\bar{X}) = \sigma^2/n$, use Theorem 1.1 to find the variance of the sampling distribution for each estimator. Hence determine which estimator has the largest variance.

- **Solution (a).** First, note that $\hat{m}$, $\tilde{m}$ and m^* are all linear functions of $\bar{X}$.
- From the central limit theorem, we know that $\bar{X}$ has a normal distribution.
- As all linear functions of normally distributed variables are themselves normally distributed, it follows that the sampling distributions of $\hat{m}$, $\tilde{m}$ and m^* must all also be normal.

- **Solution (b).** Recall, from Theorem 1.1, that

$$E(a + bX) = a + bE(X),$$

  where a and b are constants.

- We begin by finding the expected values of each estimator:

$$\begin{aligned}
E(\hat{m}) &= E(\bar{X} - 10) = E(\bar{X}) - 10 = \mu - 10; \\
E(\tilde{m}) &= E\left(\bar{X} + \frac{5}{n}\right) = E(\bar{X}) + \frac{5}{n} = \mu + \frac{5}{n}; \\
E(m^*) &= E\left[\left(\frac{n-1}{n-2}\right)\bar{X}\right] = \left(\frac{n-1}{n-2}\right)E(\bar{X}) = \left(\frac{n-1}{n-2}\right)\mu.
\end{aligned}$$

- Thus all the estimators are biased.

- The biases are

$$
\begin{aligned}
\text{bias}(\hat{m}) &= E(\hat{m}) - \mu = \mu - 10 - \mu = -10; \\
\text{bias}(\tilde{m}) &= E(\tilde{m}) - \mu = \mu + \frac{5}{n} - \mu = \frac{5}{n}; \\
\text{bias}(m^*) &= E(m^*) - \mu = \left(\frac{n-1}{n-2}\right)\mu - \mu = \left(\frac{1}{n-2}\right)\mu.
\end{aligned}
$$

- When $n = 10$ and $\mu = 8$,

$$
\begin{aligned}
\text{bias}(\hat{m}) &= -10; \\
\text{bias}(\tilde{m}) &= \frac{5}{10} = 0.5; \\
\text{bias}(m^*) &= \left(\frac{1}{10-2}\right) 8 = 1.
\end{aligned}
$$

- Hence $\tilde{m}$ has the smallest bias in absolute terms.

- **Solution (c).** Recall, from Theorem 1.1, that

$$V(a+bX) = b^2 V(X),$$

where a and b are constants.

- The variances are

$$\begin{aligned} V(\hat{m}) &= V(\bar{X}) = \frac{\sigma^2}{n}; \\ V(\tilde{m}) &= V(\bar{X}) = \frac{\sigma^2}{n}; \\ V(m^*) &= \left(\frac{n-1}{n-2}\right)^2 V(\bar{X}) = \left(\frac{n-1}{n-2}\right)^2 \frac{\sigma^2}{n}. \end{aligned}$$

- Thus, since $(n-1)/(n-2) > 1$, it follows that m^* has the largest variance.

Summary

- Small sample properties of estimators.
- Next week:
    - the Classical two-variable regression model.