Finite-Sample Properties of OLS in the Classical Linear Model

Walter Sosa-Escudero

March 2015

January 14, 2015

Classical linear model:

1. Linearity: $Y = X\beta + u$. X, an $n \times K$ random matrix, u, a $n \times 1$ random vector.
2. Strict exogeneity: $E(u|X) = 0$
3. No Multicollinearity: $\rho(X) = K$.
4. No heteroskedasticity/ serial correlation: $V(u|X) = \sigma^2 I_n$.

OLS Estimator: $\hat{\beta} = (X'X)^{-1}X'Y$
Estimator of σ^2: $S^2 = \sum_{i=1}^{2} e_i^2/(n-K)$.
Elementary econometrics courses spend considerable time deriving these estimators

Finite sample properties: properties of $\hat{\beta}$ and S^2 that can be verified for any fixed sample size n.

The goal is to derive some basic properties that hold under the classical linear model.

Conditional Expectations

$$E(Y|X = x) = \int y \;\; f_{Y|X} dy$$

Idea: how the expected value of Y changes when X changes. Seen as a *function* of X, if X is a random variable, then $E(Y|X)$ is a random variable.

Some Properties

- $Y = a + bX + U$, then $E(Y|X) = a + bX + E(U|X)$.
- $E(g(X)|X) = g(X)$
- $E(Y|X) = E(Y)$ if Y and X are independent.
- $E(Y|X, g(X)) = E(Y|X)$
- $E(Y) = E\left[E(Y|X)\right]$ (Law of Iterated Expectations).

Finite Sample Properties of $\hat{\beta}$

Properties of $\hat{\beta}$

1. Unbiasedness: $E(\hat{\beta}) = \beta$.
2. Variance: $V(\hat{\beta}) = \sigma^2 E\left[(X'X)^{-1}\right]$.
3. Gauss/Markov Theorem: $\hat{\beta}$ is the 'best linear unbiased estimator'.

Unbiasedness: $E(\hat{\beta}) = \beta$

First note:

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

By LIE $E(\hat{\beta}) = E[E(\hat{\beta}|X)]$

$$\begin{aligned}E(\hat{\beta}|X) &= \beta + E\left[(X'X)^{-1}X'u|X\right] \\ &= \beta + (X'X)^{-1}X'E\left[u|X\right] \\ &= \beta \qquad \text{(Since } E(u|X) = 0\text{)}\end{aligned}$$

Then, replacing above

$$E(\hat{\beta}) = E[E(\hat{\beta}|X)] = E(\beta) = \beta$$

- How does heteroskedasticity affect unbiasedness?
- Normality?
- Which assumptions do we use and which ones we don't?

Variance: $V(\hat{\beta}) = \sigma^2 E\left[(X'X)^{-1}\right]$.

We need an extra result

Result: $V(\hat{\beta}) = E[V(\hat{\beta}|X)] + V\left[E(\hat{\beta}|X)\right]$

By unbiasedness, $E(\hat{\beta}|X) = \beta$, so $V(\beta) = 0$. Hence, we only need to get the first term.

$$
\begin{aligned}
V(\hat{\beta}|X) &= V(\hat{\beta}-\beta|X) \quad \text{(β is not-random)} \\
&= V\left[(X'X)^{-1}X'u|X\right] \quad \text{(from previous proof...)} \\
&= E\Big[(X'X)^{-1}X'uu'X(X'X)^{-1}\;|X\Big] \\
&= (X'X)^{-1}X'E(uu'|X)X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \quad \text{(by Assumption 4)} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

Now, going back to our previous result.

$$
V(\hat{\beta}) = E\left[\sigma^2(X'X)^{-1}\right] = \sigma^2 E\left[(X'X)^{-1}\right]
$$

Gauss/Markow Theorem: $\hat{\beta}$ is the best linear unbiased estimator.

Formally: *For the classical linear model, for any linear unbiased estimator* $\tilde{\beta}$*,*

$$V(\tilde{\beta}|X) - V(\hat{\beta}|X) \geq 0$$

that is, $V(\tilde{\beta}|X) - V(\hat{\beta}|X)$ *is a positive semidefinite matrix.*

Before attacking the proof: 'better' stands for 'smaller variance'. So the GMT says that among all linear and unbiased estimators of β for the classical linear model, $\hat{\beta}$ is the 'best'

It is a rather restrictive notion.

Proof:

$\tilde{\beta}$ linear: there is $A_{K\times n}$ that depends on X, with rank K, such that $\tilde{\beta} = AY$.
Under the classical linear model

$$E(\tilde{\beta}|X) = E(AY|X) = E\left(A(X\beta + u)|X\right) = AX\beta \tag{1}$$

$\tilde{\beta}$ unbiased:

$$E(\tilde{\beta}|X) = \beta \tag{2}$$

$\tilde{\beta}$ linear <u>and</u> unbiased: (1) and (2) hold simultaneouly. This requires $AX = I$.

Trivially, $\tilde{\beta} = \hat{\beta} + \tilde{\beta} - \hat{\beta} \equiv \hat{\beta} + \hat{\gamma}$, with $\hat{\gamma} \equiv \tilde{\beta} - \hat{\beta}$.

Note that: $V(\tilde{\beta}|X) = V(\hat{\beta}|X) + V(\hat{\gamma}|X)$ iff $Cov(\hat{\beta}, \hat{\gamma}|X) = 0$.

So if we prove $Cov(\hat{\beta}, \hat{\gamma}|X) = 0$, we have the result. Why?

First note that trivially $E(\hat{\gamma}|X) = 0$ (Why?)

Hence: $Cov(\hat{\beta}, \hat{\gamma} \mid X) = E[(\hat{\beta} - \beta)\hat{\gamma}' \mid X]$

Note that

$$\begin{aligned} \hat{\gamma} &= AY - (X'X)^{-1}X'Y \\ &= (A - (X'X)^{-1}X')Y \\ &= (A - (X'X)^{-1}X')(X\beta + u) \\ &= (A - (X'X)^{-1}X')u \quad \text{(since } AX = I\text{)} \end{aligned}$$

Now replace to get:

$$\begin{aligned} Cov(\hat{\beta}, \hat{\gamma} \mid X) &= E[(\hat{\beta} - \beta)\hat{\gamma}' \mid X] \\ &= E[(X'X)^{-1}X')uu'(A - (X'X)^{-1}X')' \mid X] \\ &= \sigma^2[(X'X)^{-1}X'(A' - X(X'X)^{-1})] \\ &= \sigma^2[(X'X)^{-1}X'A' - (X'X)^{-1}X'X(X'X)^{-1})] \\ &= 0 \end{aligned}$$

where we used $V(u|X) = E(uu'|X) = \sigma^2 I_n$, and, again, $AX = I$. So, by our previous argument we get:

$$V(\tilde{\beta} \mid X) - V(\hat{\beta} \mid X) = V(\hat{\gamma} \mid X)$$

which is by construction positive semidefinite.

Can we obtain an ‘unconditional’ version of the Gauss-Markow Theorem? Yes!

I’ll leave it as an exercise. See Problem 4b) (pp.32) in Hayashi’s text.

On exogeneity

Assumption 2: Strict Exogeneity

$$E(u_i|X) = 0, \qquad i = 1, 2, \ldots, n$$

In basic courses it is assumed that $E(u_i) = 0$. Which one is stronger?

Implications of strict exogeneity:

- $E(u_i) = 0, \qquad i = 1, \ldots, n.$
  Proof: By the law of iterated expectations and strict exogeneity:
  $$E(u) = E[E(u|X)] = E(0) = 0$$
  In words: on average, the model is exactly linear.

- $E(x_{jk}u_i) = 0, \qquad j, i = 1, \ldots, n; \;\; k = 1, \ldots, K$
  In words: explanatory variables are uncorrelated with the error terms of *all* observations.
  Proof: as exercise.

Finite Sample Properties of S^2

Remember that we proposed:

$$S^2 = \frac{\sum e_i^2}{n-K} = \frac{e'e}{n-K}$$

as an estimator for σ^2.

Result: S^2 is unbiased $(E(S^2|X) = \sigma^2)$

Trace: Let A be a square $m \times m$ matrix. Its *trace* of A is the sum of all its principal diagonal elements: $tr(A) = \sum_{i=1}^{m} A_{ii}$.

Simple properties:

- If A is a scalar, trivially $tr(A) = A$
- $tr(AB) = tr(BA)$
- $tr(AB) = tr(A) + tr(B)$

The M matrix: $M \equiv I_n - X(X'X)^{-1}X'$

Some properties (check as homework)

- $M = M'$ (symmetric), $M = MM$ (idempotent)
- $tr(M) = n - K$
- $e = Mu$.

$e = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = (I - X(X'X)^{-1}X')Y = MY = MX\beta + Mu$. Now note that $MX = (I - X(X'X)^{-1}X')X = X - X(X'X)^{-1}X)X = 0$.

Proof:

$$
\begin{aligned}
E(S^2 \mid X) = \frac{E(e'e \mid X)}{n-K} &= \frac{E(u'M'Mu \mid X)}{n-K} \\
&= \frac{E(u'Mu \mid X)}{n-K} \\
&= \frac{E(tr(u'Mu) \mid X)}{n-K} \\
&= \frac{E(tr(uu'M) \mid X)}{n-K} \\
&= \frac{tr(E(uu'M \mid X))}{n-K} \\
&= \frac{tr(\sigma^2 I_n M)}{n-K} \\
&= \frac{\sigma^2 tr(M)}{n-K} \\
&= \sigma^2
\end{aligned}
$$

Hypothesis Testing

Assumption 5: normality. $u|X$ is normally distributed (it is a vector, so this involves the multivariate normal).

Note that this together with the classical assumptions imply

$$u|X \sim N(0, \sigma^2 I_n)$$

Remember that $\hat{\beta} = \beta + (X'X)^{-1}X'u$.Then

$$\hat{\beta} \,|\, X \sim N\big(\beta, \sigma^2 (X'X)^{-1}\big)$$

Hypothesis about single coefficients

$H_0 : \beta_j = \beta_{j0}$ vs. $H_A : \beta_j \neq \beta_{j0}$

Let a_{is} denote de (i, s) element of $(X'X)^{-1}$.
Then, when H_0 is true $\hat{\beta}_j - \beta_{j0} \sim N(0, \sigma^2 a_{jj})$ so

$$z_j \equiv \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\sigma^2 a_{jj}}} \sim N(0, 1)$$

- Note that $\sigma^2 a_{jj} = V(\beta_j)$.
- Special case: $\beta_{j0} = 0$ 'significance hypothesis'.
- The distribution of z_k does not depend on X.
- If σ^2 is observed, reject if z_j lies outside an acceptance region.

The problem is that σ^2 is not observed. Define:

$$t_j \equiv \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{S^2 a_{jj}}}$$

which is z_k with σ^2 replaced by its unbiased estimator S^2.

Result: Under assumptions 1 to 5 and when H_0 holds, $t_j \sim t(n-K)$.

Hypothesis about linear combinations of β.

$H_0 : c'\beta - r = 0$ vs. $H_A : c'\beta - r \neq 0, \;\; c \in \Re^K, \; r \in \Re$.

WLOG, supose $K = 3$ so,

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad i = 1, \ldots, n$$

Consider the following hypotheses:

- a) $H_0 : \beta_2 = \beta_3$, or $H_0 : \beta_2 - \beta_3 = 0$. In this case $c = (0, 1, -1)$ and $r = 0$.
- b) $H_0 : \beta_2 + \beta_3 = 1$, so now $c = (0, 1, 1)$ and $r = 1$.

To derive an appropiate test statistic note:

$$c'\hat{\beta} - r \sim N(0, \sigma^2 c'(X'X)c) \sim N(0,1)$$

So

$$z = \frac{c'\hat{\beta} - r}{\sqrt{\sigma^2 c'(X'X)c)}} \sim N(0,1)$$

And again, by the same argument as before, a 'feasible' version is

$$t = \frac{c'\hat{\beta} - r}{\sqrt{S^2 c'(X'X)^{-1}c)}} \sim t(n-K)$$

As a simple exercise, the appropriate statistics for the cases considered before are

- a) $c'\hat{\beta} - r = \hat{\beta}_2 - \hat{\beta}_3$ and $\sigma^2 c'(X'X)^{-1}c = \hat{V}(\beta_2) + \hat{V}(\beta_2) - 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$, so

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\hat{V}(\beta_2) + \hat{V}(\beta_2) - 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$$

- b) $c'\hat{\beta} - r = \hat{\beta}_2 + \hat{\beta}_3 - 1$, and $\sigma^2 c'(X'X)^{-1}c = \hat{V}(\beta_2) + \hat{V}(\beta_2) + 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$, so

$$t = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{\hat{V}(\beta_2) + \hat{V}(\beta_2) + 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$$

Multiple Linear Hypothesis

$H_0 : R\beta - r = 0$, R is a $q \times K$ matrix with $\rho(R) = q$, and $r \in \Re^q$

Example. In our previous case consider the *multiple* hypothesis

$$H_0 : \beta_2 = 0 : \beta_3 = 0$$

These are actually two joint hypothesis about the coefficient vector β. In this case

$$R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with $q = 2$. r is the number of restrictions.

- *What is the 'full row rank'requirement, $\rho(R) = q$, asking for?*

Consider the following test statistic:

$$F = \frac{(R\hat{\beta} - r)' \left[R(X'X)^{-1}R' \right]^{-1} (R\hat{\beta} - r)' \;/\; q}{S^2}$$

Result: under all assumptions and when H_0 is true, $F \sim F(q, n-K)$.

Intuition: Note that $V(R\hat{\beta}) = \sigma^2 R(X'X)^{-1}R'$. Then, $\hat{V}(R\hat{\beta}) = S^2 R(X'X)^{-1}R'$, so

$$F = (R\hat{\beta} - r)'\hat{V}(R\hat{\beta}|X)^{-1}(R\hat{\beta} - r)' \;/\; q$$

F is actually checking how large $R\hat{\beta} - r$ is.