5 Further Topics In Linear Regression

5.1 Large Sample Properties of OLS Estimators

Consistency

Theorem 5.1: Under the classical assumptions 1–5 the OLS estimators are consistent. That is

(1) $$\hat{\beta}_j \xrightarrow{P} \beta_j \text{ as } n \to \infty,$$

where "$\xrightarrow{P}$" means that for increasing sample size n and for any $\epsilon > 0$:

(2) $$\lim_{n\to\infty} P(|\hat{\beta}_j - \beta_j| > \epsilon) \to 0$$

$\hat{\beta}_j \xrightarrow{P} \beta_j$ is denoted usually as

(3) $$\text{plim}\,\hat{\beta}_j = \beta_j$$

"plim" standing for "probability limit".

Consistency of and estimator is an important and desirable property.

Example:

The Law of Large Numbers (LLN):

Let $x_1, \ldots, x_n$ be independent random variables with $\mathbb{E}[x_i] = \mu$ and $\mathbb{V}\text{ar}[x_i] = \sigma^2 < \infty$, then

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{P} \mu, \tag{4}$$

where "$\xrightarrow{P}$" means that for any $\epsilon > 0$:

$$\lim_{n\to\infty} P(|\bar{x}_n - \mu| > \epsilon) \to 0.$$

Remark 5.1: Correlation of the error term u with any of the explanatory variables $x_1, \ldots, x_k$ in regression

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

implies that OLS estimators $\widehat{\beta}_i$ are both *biased* and *inconsistent*!

Asymptotic Normality and Large Sample Inference

If the error terms $u_1, \ldots, u_n$ in the regression are not normal, i.e., Assumption 6 does not hold, then the distribution of $\widehat{\beta}_j$ is not any more (exactly) normal. This implies that the normal theory inference is not any more exactly valid.

However, because the OLS estimators are weighted sums of random variables, the CLT applies and we have the important result that the OLS estimators are in any case asymptotically normal (if assumptions 1–5 are in effect).

Central Limit Theorem (CLT): Let $x_1, \ldots, x_n$ be independent and identically distributed (iid) random variables with
$\mathbb{E}[x_i] = \mu$ and $\mathbb{V}\text{ar}[x_i] = \sigma^2 < \infty$. Then

$$z_n = \frac{(\bar{x}_n - \mu)}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0, 1) \tag{5}$$

as $n \to \infty$, where "$\overset{a}{\sim}$" means that the distribution of z_n approaches to the standard normal distribution.

The notation $\overset{a}{\sim}$ stands for the phrase "the asymptotic distribution of z_n is $N(0, 1)$".

As a result

(6) $$\frac{\widehat{\beta}_j - \beta_j}{\sigma_{\widehat{\beta}_j}} \overset{a}{\sim} N(0,1)$$

and

(7) $$\frac{\widehat{\beta}_j - \beta_j}{s_{\widehat{\beta}_j}} \overset{a}{\sim} t_{n-k-1}.$$

In the same manner the F-test discussed in Chapter 4 are also asymptotic.

Test statistics relying on asymptotic distribution results are commonly called asymptotic or large sample test statistics.

Other Large Sample Tests: The Lagrange Multiplier Test

In statistics there are three kinds of all purpose large sample statistics: The Likelihood Ratio (LR), the Wald (W), and the Lagrange Multiplier (LM) test statistics. These can be used in testing complicated restrictions on parameters.

Under the null hypothesis the asymptotic distribution for each statistic is χ_q^2 with q degrees of freedom, where q equals the number of imposed restrictions.

Asymptotically all these test lead to the same result, but in finite samples they may differ.

Modern econometric packages have these statistics available.

5.2 Effects of Data scaling on OLS Statistics

We have earlier discussed scaling and adding a constant to y and x variable in the simple regression.

The same rules apply here too. Here we discuss only standardization of variables.

Beta Coefficients

Sometimes it is desirable to standardize all the variables, such that

(8) $$y_i^* = \frac{y_i - \bar{y}}{s_y}$$

and

(9) $$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j},$$

where s_y and s_j are the standard deviations of y and x_j, respectively, $j = 1, \ldots, k$.

Then

(10) $$\hat{y}_i^* = \hat{\beta}_1^* x_{i1}^* + \cdots + \hat{\beta}_k^* x_{ik}^*,$$

where

(11) $$\hat{\beta}_j^* = \frac{s_j}{s_y}\hat{\beta}_j.$$

β_j^*s are called standardized coefficients or <u>beta coefficients</u>.

Remark 5.2: t-statistics R-squares etc do not change in standardization.

The interpretation of the standardized coefficients is that β_j^* indicates the change in standardized value of y as x_j changes by one standard deviation, ceteris paribus.

5.3 Prediction

Confidence Interval on Estimated Mean

We shall now find a confidence band for the mean value of the response variable Y for a specific set of values $c_1, \ldots, c_k$ of the predictor variables, which have not necessarily been used in developing the regression equation. Let

$$\begin{aligned} \mu_{Y|\mathbf{x}} &:= E(Y|X_1 = c_1, X_2 = c_2, \ldots, X_k = c_k) \\ &= \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \cdots + \beta_k c_k \\ &= \mathbf{x}'_{1\times(k+1)}\beta_{(k+1)\times 1}, \quad \text{where} \end{aligned} \tag{12}$$

$$\mathbf{x}' = (1, c_1, c_2, \ldots, c_k), \quad \beta = (\beta_0, \beta_1, \ldots, \beta_k)'.$$

An unbiased estimator for $\mu_{Y|\mathbf{x}}$ is

$$\widehat{\mu}_{Y|\mathbf{x}} = b_0 + b_1 c_1 + b_2 c_2 + \cdots + b_k c_k = \mathbf{x}'\mathbf{b}. \tag{13}$$

The variance of $\widehat{\mu}_{Y|\mathbf{x}}$ is

$$\begin{aligned} \text{Var}\,(\widehat{\mu}_{Y|\mathbf{x}}) &= \text{Var}\,(\mathbf{x}'\mathbf{b}) = \mathbf{x}'\text{Var}\,(\mathbf{b})\mathbf{x} \\ &= \mathbf{x}'\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} = \sigma_u^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}, \end{aligned} \tag{14}$$

such that

$$\widehat{\mu}_{Y|\mathbf{x}} \sim N(\mathbf{x}'\beta, \sigma_u^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}). \tag{15}$$

Standardizing and replacing the unknown σ_u^2 by its estimator $\hat{\sigma}_u^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{\text{SSR}}{n-k-1}$ yields

$$t = \frac{\hat{\mu}_{Y|\mathbf{x}} - \mu_{Y|\mathbf{x}}}{\sqrt{\hat{\sigma}_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}} \sim t(n-k-1). \tag{16}$$

A $(1-\alpha)$ confidence interval on $\mu_{Y|\mathbf{x}}$ is thus

$$\left[\hat{\mu}_{Y|\mathbf{x}} \pm t_{\frac{\alpha}{2}}(n-k-1)\sqrt{\hat{\sigma}_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}\right]. \tag{17}$$

Recalling, that the student distribution approaches the normal distribution for $df \to \infty$, we may also write for $n \gg k$:

$$\left[\hat{\mu}_{Y|\mathbf{x}} \pm z_{\frac{\alpha}{2}}\sqrt{\hat{\sigma}_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}\right], \tag{18}$$

where $z_{\frac{\alpha}{2}}$ denotes the $\frac{\alpha}{2}$ fractile of the standard normal distribution.

Prediction Interval on Single Response

Consider next predicting a single response $Y|\mathbf{x} = \mu_{Y|\mathbf{x}} + u$. The scalar product $\mathbf{x}'\mathbf{b}$ is also an unbiased estimator of $Y|\mathbf{x}$ since

$$E(Y|\mathbf{x}) = E(\mu_{Y|\mathbf{x}}) + E(u) = \mathbf{x}'\beta. \tag{19}$$

But the variance of $\widehat{Y}|\mathbf{x}$ is larger than the variance of $\widehat{\mu}_{Y|\mathbf{x}}$ due to the additional variation in u. More specifically:

$$\begin{aligned} \text{Var}\,(\widehat{Y}|\mathbf{x}) &= \text{Var}\,(\widehat{\mu}_{Y|\mathbf{x}}) + \text{Var}\,(u) \\ &= \sigma_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} + \sigma_u^2. \end{aligned} \tag{20}$$

That is,

$$\widehat{Y}|\mathbf{x} \sim N(\mathbf{x}'\beta, \sigma_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} + \sigma_u^2). \tag{21}$$

A similiar argument as for the confidence interval on the estimated mean yields as a $(1-\alpha)$ prediction interval for an individual response:

(22)

$$\left[\widehat{Y}|\mathbf{x} \pm t_{\frac{\alpha}{2}}(n-k-1)\sqrt{\widehat{\sigma}_u^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}+\widehat{\sigma}_u^2}\right].$$

Again, since the student distribution approaches the normal distribution for $df \to \infty$, we may also write for $n \gg k$:

(23) $$\left[\widehat{\mu}_{Y|\mathbf{x}} \pm z_{\frac{\alpha}{2}}\sqrt{\widehat{\sigma}_u^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}+\widehat{\sigma}_u^2}\right],$$

where $z_{\frac{\alpha}{2}}$ denotes the $\frac{\alpha}{2}$ fractile of the standard normal distribution.

Note that the prediction interval for an individual response is wider than the confidence interval for the corresponding mean.

If you wish to avoid the matrix algebra involved in calculating the standard errors for the confidence bands, you may use the following trick. Write $\beta_0 = \mu_{Y|\mathbf{x}} - \beta_1 c_1 - \ldots - \beta_k c_k$ and plug this into the equation

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

to obtain

(24)

$$y = \mu_{Y|\mathbf{x}} + \beta_1(x_1 - c_1) + \ldots + \beta_k(x_k - c_k) + u.$$

In other words, subtract c_j from each observation x_j and run the gegression of

$$y_i \text{ on } (x_{i1} - c_1), \ldots, (x_{ik} - c_k),\ i = 1, \ldots, n.$$

The parameter estimate of the *intercept* will be $\hat{\mu}_{Y|\mathbf{x}}$ and, more important, its standard error

(25) $$SE(\hat{\mu}_{Y|\mathbf{x}}) = \sqrt{\hat{\sigma}_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

needed in the construction of the confidence band for $\mu_{Y|\mathbf{x}}$. Augment this with $\hat{\sigma}_u^2$ from the regression output to obtain

(26) $$SE(\hat{Y}|\mathbf{x}) = \sqrt{\hat{\sigma}_u^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} + \hat{\sigma}_u^2}$$

for prediction intervals of individual responses.

Predicting y from $\log(y)$ specification

Consider the model

(27) $\log(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$

Estimated model

(28) $\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k.$

A natural prediction for y would be

$$\hat{y} = e^{\widehat{\log(y)}}.$$

This, however, systematically underestimates the expected value of y.

This is because, if $u \sim N(0, \sigma_u^2)$, $\mathsf{E}[e^u] = e^{\frac{1}{2}\sigma_u^2}$.

Then given x-values this implies

(29) $\mathsf{E}[y|\mathbf{x}] = \mathsf{E}[e^{\log(y)}|\mathbf{x}] = e^{\mathbf{x}'\mathbf{b}}\mathsf{E}[e^u|\mathbf{x}] = e^{\frac{1}{2}\sigma_u^2}e^{\mathbf{x}'\mathbf{b}}$,

where $\mathbf{b} = (\beta_0, \beta_1, \ldots, \beta_k)'$ and $\mathbf{x} = (1, x_1, \ldots, x_k)'$.

Thus an appropriate predictor for y is

$$\hat{y} = e^{\hat{\sigma}_u^2/2} e^{\widehat{\log(y)}}. \tag{30}$$

If the normality of u does not hold then let $\mathrm{E}[e^u] = \alpha_0$ and (29) becomes

$$\mathrm{E}[y|\mathbf{x}] = \alpha_0 e^{\log(y)} = \alpha_0 e^{\mathbf{x}'\mathbf{b}}, \tag{31}$$

where α_0 is an unknown parameter.

It turns out that a consistent estimator of α_0 is found as follows:

(1) Obtain the fitted value of $\widehat{\log(y)}_i$
(2) For each observation i, create $\hat{m}_i = e^{\widehat{\log(y)}_i}$
(3) Regress y on $\hat{m}_i$ without an intercept, and use the estimated regression coefficient as an estimate of α_0.

Example 5.1: Consider the wage example. The model to be estimated is

$$\log(\text{wage}) = \beta_0 + \beta_1\text{educ} + \beta_2\text{exper} + \beta_3\text{tenure} + u.$$

Estimating the parameters, generating the $\widehat{m}_i$ series, and estimating regression

$$\text{wage} = \alpha_0 \widehat{m} + v,$$

where v is an error term, produces $\widehat{\alpha}_0 = 1.1227$. Note that $e^{\widehat{\sigma}_u^2/2} = e^{(0.440862)^2/2} \approx 1.10206$, which differs from the α_0-estimate, indicating that the residuals are not normally distributed.

Note. It can be shown that α_0 must be always larger than 1. If nonetheless you get an estimate which is smaller than one, then it is an indication that assumption 3 about independence of the error term and the explanatory variables does not hold.

5.4. Functional Form of Regression

In economic applications most nonlinear relationships between explained explanatory variables are are worked out by taking logarithms or containing quadratics of exlanatory variables in the model.

Using logarithmic transformations

Consider the model of an earlier example

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \text{rooms} + u. \quad (32)$$

Coefficient β_1 is the elasticity of *price* with respect to pollution *nox*), while $100 \times \beta_2$ is approximately the percentage change in *price* when the *rooms* increases by one.

Example 5.2: Using Wooldridge's data set *hprice2.xls* we get the follwoing estimates for (32)

Dependent Variable: LOG(PRICE)
Method: Least Squares
Date: 10/10/06 Time: 00:15
Sample: 1 506
Included observations: 506

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.233738	0.187741	49.18350	0.0000
LOG(NOX)	-0.717673	0.066340	-10.81816	0.0000
ROOMS	0.305918	0.019017	16.08626	0.0000

R-squared	0.513717	Mean dependent var	9.941057
Adjusted R-squared	0.511784	S.D. dependent var	0.409255
S.E. of regression	0.285956	Akaike info criterion	0.339957
Sum squared resid	41.13085	Schwarz criterion	0.365015
Log likelihood	-83.00902	F-statistic	265.6890
Durbin-Watson stat	0.603290	Prob(F-statistic)	0.000000

Thus the estimate model is

$$
\begin{aligned}
\widehat{\log(\text{price})} &= \underset{(0.188)}{9.234} - \underset{(0.066)}{0.718}\log(\text{nox}) + \underset{(0.019)}{0.306}\,\text{rooms} \\
R^2 &= 0.514 \\
n &= 506,
\end{aligned}
$$

where standard errors are in parnetheses

When *nox* increases by 1%, *price* falls by 0.718% (holding *rooms* fixed). When number of rooms increases by one, *price* increases by approximately $100 \times 0.306 = 30.6\%$.

Remark 5.3: Approximation $\%\Delta y \approx 100 \times \Delta \log y$ becomes inaccurate when the change in $\log y$ becomes large. Generally if we have an estimated model

$$\widehat{\log y} = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2, \tag{33}$$

fixing x_1, we have $\Delta\widehat{\log y} = \hat{\beta}_2 \Delta x_2$, from which we get the exact percentage change as

$$\widehat{\%\Delta y} = 100 \times \left(\exp(\hat{\beta}_2 \Delta x_2) - 1\right). \tag{34}$$

Example 5.3: (Continued) In the previous example we have $\Delta x_2 = 1$ and $\widehat{\beta}_2 = 0.305918$. Thus we get

$$\widehat{\%\Delta y} = 100 \times (\exp(0.305918) - 1) \approx 35.8\%,$$

which is notably larger than the approximate change 30.6%.

Advantages of using log transformations:

• interpretation (elasticity, perentage change)

• changing scale does not change slope coefficients

• if $y > 0$ log-transformation usually make variables closer to normality

Usually log-transformations are quite routinely taken for series that are positive monetary values (wages, salaries, firm sales, firm market values, stock indices, etc.) also logs are often taken from variables measuring population, total number of employees, etc that are usually large integers.

Variables measured in years (education, experience, tenure, age, etc) are usually used in their original form.

Remark 5.4: log-transformations cannot be used if a variable takes zero or negative values!

Quadratic Terms

Consider model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u, \tag{35}$$

where x^2 is the quadratic term.

The interpretation of the model changes. β_1 *does not* any more measure the change in y when x changes by one unit (i.e., β_1 is not any more the slope coefficient).

To see this, write the estimated model

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2, \tag{36}$$

then approximately

$$\Delta\hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x)\Delta x, \tag{37}$$

so that the slope is

$$\frac{\Delta\hat{y}}{\Delta x} \approx \hat{\beta}_1 + 2\hat{\beta}_2 x. \tag{38}$$

Example 5.5: Consider the wage example and estimate the model (data: wage1.xls)

$$\log(\text{wage}) = \beta_0 + \beta_1 \,\text{exper} + \beta_2 \,\text{exper}^2 + u.$$

Estimating with EViews, we obtain:

```
Dependent Variable: LOG(WAGE)
Method: Least Squares
Sample: 1 526
Included observations: 526
===============================================================
Variable          Coefficient   Std. Error      t-Stat     Prob.
---------------------------------------------------------------
C                    1.295291     0.049481    26.17746    0.0000
EXPER                0.045534     0.005859    7.771027    0.0000
EXPER^2             -0.000944     0.000129   -7.312008    0.0000
===============================================================
R-squared            0.104001  Mean dependent var       1.623268
Adjusted R-squared   0.100574  S.D. dependent var       0.531538
S.E. of regression   0.504101  Akaike info criterion    1.473605
Sum squared resid    132.9034  Schwarz criterion        1.497932
Log likelihood      -384.5581  F-statistic              30.35285
Durbin-Watson stat   1.795529  Prob(F-statistic)        0.000000
===============================================================
```

Thus, the estimated equation is

$$\begin{array}{rcl} \log(\widehat{\text{wage}}) & = & \underset{(0.049)}{1.30} + \underset{(0.0059)}{0.046}\,\text{exper} - \underset{(0.000129)}{0.000944}\,\text{exper}^2 \\ R^2 & = & 0.104 \\ n & = & 526 \end{array}$$

Experience (exper) has a diminishing effect on wage. The first year increases wage about by 4.6%, the sencond year [using (38)] by $.045534 - 2 \times (.000944) \times 1 = 0.043646 \approx 4.4\%$. Going from 10 to 11 years of experience, wage is predicted to change by $.045534 - 2 \times (.000944) \times 10 \approx 2.7\%$.

The predicted maximum wage is achieved at experience of

$$\text{exper} = -\frac{\hat{\beta}_1}{2\hat{\beta}_2} = -\frac{.045534}{2(-0.000944)} \approx 24\,\text{years}.$$

Note that we have omitted other important factors (education, etc.) from this example.

Interaction terms

Consider a model with two explanatory variables such that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + u. \quad (39)$$

The cross-product term $x_1 x_2$ is called the *interaction* effect of x_1 and x_2. Usually it comes in a natural way to the model.

For example, consider the simple consumption function

$$C = \beta_0 + \beta_1 Y + u, \quad (40)$$

where C denotes consumption and Y income.

Suppose, the marginal propensity to consume (mpc), β_1 depends on the level of wealth A such that

$$\beta_1 = \beta_y + \beta_{ay} A. \tag{41}$$

This implies the interaction term $A \cdot Y$ to the model (40), as

$$\begin{aligned} C &= \beta_0 + \beta_1 Y + u \\ &= \beta_0 + (\beta_y + \beta_{ay} A) Y + u \\ &= \beta_0 + \beta_y Y + \beta_{ay} AY + u. \end{aligned} \tag{42}$$