Regression Analysis: Properties of Least Squares Estimators

EC163 Handout, Chapter 18

Abstract

We consider the regression model: $Y_i = X_i'\beta + u_i$, $i = 1, \ldots, n$. This note summarizes the results for asymptotic analysis of the least squares estimator $\hat{\beta}$, that makes it possible to test: (i) hypotheses about the individual coefficients of $\hat{\beta}$; and (ii) hypotheses that involve several coefficients, such as $R\beta = r$, where R is a known matrix and r is a known vector.

1 Regression Analysis and Least Squares Estimators

Consider the regression model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i = X_i'\beta + u_i, \qquad t = 1, \ldots, n, \qquad [\text{Matrix: } \mathbf{Y} = \mathbf{X}\beta + \mathbf{U}.] \tag{1}$$

with $\beta = (\beta_0\ \beta_1 \ldots \beta_k)'$, $X_i = (1\ x_{1i} \ldots x_{ki})'$, $\mathbf{Y} = (Y_1 \ldots Y_n)'$, $\mathbf{X} = (X_1 \ldots X_n)'$. The *least squares estimator* (LS) is given as the solution to

$$\min_\beta S(\beta) = \min_\beta \sum_{i=1}^n u_i^2 = \min_\beta \sum_{i=1}^n \left(Y_i - \beta' X_i\right)^2, \qquad [\text{Matrix: } \min_\beta (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta).]$$

The least squares estimator results by solving the first order condition for a minimum:

$$\left.\begin{array}{c} \frac{\partial}{\partial \beta_0} S(\hat{\beta}) = 0 \\ \frac{\partial}{\partial \beta_1} S(\hat{\beta}) = 0 \\ \vdots \\ \frac{\partial}{\partial \beta_k} S(\hat{\beta}) = 0 \end{array}\right\} \Leftrightarrow \left(\begin{array}{c} -2\sum_{i=1}^n \left(Y_i - X_i'\hat{\beta}\right) = 0 \\ -2\sum_{i=1}^n x_{1i}\left(Y_i - X_i'\hat{\beta}\right) = 0 \\ \vdots \\ -2\sum_{i=1}^n x_{ki}\left(Y_i - X_i'\hat{\beta}\right) = 0 \end{array}\right) \Leftrightarrow$$
$$\left(\begin{array}{c} \sum_{i=1}^n Y_i = \sum_{i=1}^n X_i'\hat{\beta} \\ \sum_{i=1}^n x_{1i} Y_i = \sum_{i=1}^n x_{1i} X_i'\hat{\beta} \\ \vdots \\ \sum_{i=1}^n x_{ki} Y_i = \sum_{i=1}^n x_{ki} X_i'\hat{\beta} \end{array}\right) \Leftrightarrow \sum_{i=1}^n X_i Y_i = \sum_{i=1}^n X_i X_i'\hat{\beta} \tag{2}$$

and the least squares estimator reads

$$\hat{\beta} = \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i Y_i. \qquad [\text{Matrix: } \hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}], \tag{3}$$

by substituting (1) into (3) we find that

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{U},$$

such that

$$\hat{\beta}-\beta=\left(\sum_{i=1}^{n}X_iX_i'\right)^{-1}\sum_{i=1}^{n}X_iu_i,\qquad [\text{Matrix: } \hat{\beta}-\beta=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{U}]. \tag{4}$$

A1 $E(U|X)=0$.

The expectation of $\hat{\beta}$:

$$E(\hat{\beta})=\beta+\left(\sum_{i=1}^{n}X_iX_i'\right)^{-1}\sum_{i=1}^{n}E(X_iu_i)=\beta,\qquad [\text{Matrix: } E(\hat{\beta})=\beta+\left(\mathbf{X}'\mathbf{X}\right)^{-1}E(\mathbf{X}'\mathbf{U})=\beta]. \tag{5}$$

We want to use the true finite sample distribution of $\hat{\beta}$. Since we can not always obtain this, we resort to the asymptotic distribution, which is likely to be a good approximation of the unknown finite sample distribution, if n is large. We consider the distribution of $\hat{\beta}$ as $n\to\infty$, where $\hat{\beta}$ is the LS estimator (note that $\hat{\beta}$ depends on n, although it is not clear from the definition).

When applying asymptotic results one should have the following in mind. The estimator, $\hat{\beta}_n$, has some (unknown) finite-sample distribution, which is approximated by its asymptotic distribution. The finite-sample distribution and the asymptotic distribution are (most likely) different, so we are making an error when we use the asymptotic distribution. However, when n is large, the difference between the two distributions is likely to be small, and the asymptotic distribution is then a good approximation.

2 Asymptotic Analysis of $\hat{\beta}$

In addition to Assumption A1 we assume:

A2 $(Y_i, X_{1i},\ldots,X_{ki})$, $i=1,\ldots,n$ are iid.

A3 $(u_i, X_{1i},\ldots,X_{ki})$ have finite fourth moments.

A4 $\mathbf{X}'\mathbf{X}$ has full rank. [Equivalent to: No perfect multicollinearity, or that $\det(\mathbf{X}'\mathbf{X})\neq 0$.]

We have already used A4 when we implicitly assumed that the inverse of $\mathbf{X}'\mathbf{X}$ was well defined. The two other assumptions will be used in the asymptotic analysis.

You are familiar with the law of large numbers and the central limit theorem for scalar variables. The results that concerns vectors and matrices are straightforward extensions.

Theorem 1 (Multivariate LLN) *Let $\{M_i\}$, $i=1,2,\ldots$ be a sequence of matrices, whose elements are iid random variables. Then it holds that $\frac{1}{n}\sum_{i=1}^{n}M_i\xrightarrow{p}E(M_i)$.*

So the univariate LLN is just a special case of the multivariate version. The same is true for the multivariate CLT.

Theorem 2 (Multivariate CLT) *Let $\{V_i\}$ be a sequence of m-dimensional random vectors that are iid, with mean $\mu_V=E(V_i)$ and covariance matrix $\Omega_V=\operatorname{var}(V_i)=E[(V_i-\mu_V)(V_i-\mu_V)']$. Then it holds that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(V_i-\mu_V)\xrightarrow{d}N_m(0,\Omega_V)$.*

Since $\{X_i\}$ is a sequence of iid random variables (vectors), it follows that $\{X_i X_i'\}$ is a sequence of iid random variables (matrices). So by the multivariate LLN it holds that $\frac{1}{n}\sum_{i=1}^n X_i X_i' \xrightarrow{p} Q_X \equiv E(X_i X_i')$. Similarly, $\{X_i u_i\}$ is a sequence of iid random variables (vectors), with expected value $E[X_i u_i] = E[E(X_i u_i|X_i)] = E[X_i E(u_i|X_i)] = E[X_i 0] = 0$. So by the multivariate central limit theorem we have that $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i u_i \xrightarrow{d} N_{k+1}(0, \Sigma_v)$, where $\Sigma_v \equiv \text{var}(X_i u_i) = E(X_i X_i' u_i^2)$. Here we are implicitly using Assumption A3 that guarantees that the expected values $E(X_i X_i')$ and $E(X_i X_i' u_i^2)$ are well defined (finite).

Theorem 3 (Linear Transformation of Gaussian Variables) *Let $Z \sim N_m(\mu, \Omega)$, for some vector, μ, $(m \times 1)$, and some matrix Ω, $(m \times m)$. Let A be a $l \times m$ matrix and b be a $l \times 1$ vector. Define the l-dimensional random variable $\tilde{Z} = AZ + b$. Then it holds that $\tilde{Z} \sim N_l(A\mu + b, A\Omega A')$.*

Theorem 4 (Asymptotic Linear Transformation of Gaussian Variables) *Let $Z_n \xrightarrow{d} N_m(\mu, \Omega)$, for some vector, μ, $(m\times 1)$, and some matrix Ω, $(m\times m)$. Let $A_n \xrightarrow{p} A$ and $b_n \xrightarrow{p} b$ for some constant $l\times m$ matrix, A, and some constant $l \times 1$ vector b. Define the l-dimensional random variable $\tilde{Z}_n = A_n Z_n + b_n$. Then it holds that $\tilde{Z}_n \xrightarrow{d} N_l(A\mu + b, A\Omega A')$.*

In the present context, we can set $A_n = \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1}$, $b_n = 0$, and $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i u_i$, and from the theorem it follows that

$$\sqrt{n}(\hat{\beta} - \beta) = \underbrace{\left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1}}_{\xrightarrow{p} Q_X^{-1}} \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i u_i}_{\xrightarrow{d} N_{k+1}(0,\Sigma_v)} \xrightarrow{d} N_{k+1}(0, Q_X^{-1}\Sigma_v Q_X^{-1}).$$

The covariance matrix $\Sigma_{\sqrt{n}(\hat{\beta}-\beta)} \equiv Q_X^{-1}\Sigma_v Q_X^{-1}$, can be estimated by $\hat{\Sigma}_{\sqrt{n}(\hat{\beta}_n-\beta)} \equiv \hat{Q}_X^{-1}\hat{\Sigma}_v \hat{Q}_X^{-1}$, where

$$\hat{Q}_X^{-1} \equiv \frac{1}{n}\sum_{i=1}^n X_i X_i' \qquad \text{and} \qquad \hat{\Sigma}_v \equiv \frac{1}{n-k-1}\sum_{i=1}^n X_i X_i' \hat{u}_i^2.$$

How can we be sure that $\hat{\Sigma}_{\sqrt{n}(\hat{\beta}_n-\beta)}$ is consistent for $\Sigma_{\sqrt{n}(\hat{\beta}_n-\beta)}$? We have already established that $\hat{Q}_X \xrightarrow{p} Q_X$ and since $\{X_i, Y_i\}$ is iid (Assumption A2), also $\{X_i X_i'(Y_i - X_i'\beta)^2\} = \{X_i X_i' u_i^2\}$ is iid, such that a LLN gives us $\frac{1}{n}\sum_{i=1}^n X_i X_i' u_i^2 \xrightarrow{p} E(X_i X_i' u_i^2)$. To establish that $\hat{\Sigma}_v \xrightarrow{p} \Sigma_v$, we first note that

$$\frac{1}{n}\sum_{i=1}^n X_i X_i' \hat{u}_i^2 = \frac{1}{n}\sum_{i=1}^n X_i X_i' u_i^2 + \frac{1}{n}\sum_{i=1}^n X_i X_i'(\hat{u}_i^2 - u_i^2),$$

and it can be shown that $\frac{1}{n}\sum_{i=1}^n X_i X_i'(\hat{u}_i^2 - u_i^2) \xrightarrow{p} 0$, (beyond the scope of EC163), such that $\hat{\Sigma}_v = \frac{n}{n-k-1}\frac{1}{n}\sum_{i=1}^n X_i X_i' \hat{u}_i^2 \xrightarrow{p} E(X_i X_i' u_i^2) = \Sigma_v$, using $\frac{n}{n-k-1} \to 1$ as $n \to \infty$. Since the mapping from $\{Q_x, \Sigma_v\} \mapsto Q_x^{-1}\Sigma_v Q_x^{-1}$ is continuous, we know that $\hat{Q}_X \xrightarrow{p} Q_X$ and $\hat{\Sigma}_v \xrightarrow{p} \Sigma_v$ implies that $\hat{\Sigma}_{\sqrt{n}(\hat{\beta}_n-\beta)} = \hat{Q}_x^{-1}\hat{\Sigma}_v \hat{Q}_x^{-1} \xrightarrow{p} Q_x^{-1}\Sigma_v Q_x^{-1} = \Sigma_{\sqrt{n}(\hat{\beta}_n-\beta)}$, as we wanted to show.

Multiplying $\sqrt{n}(\hat{\beta} - \beta)$ by $\frac{1}{\sqrt{n}}$ and adding β, shows that asymptotically $\hat{\beta}$ is normally distributed about β, with a covariance matrix that is given by $\Sigma_{\hat{\beta}} \equiv \frac{1}{n}\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}$. In practice we will use the estimate, $\hat{\Sigma}_{\hat{\beta}} \equiv \frac{1}{n}\hat{\Sigma}_{\sqrt{n}(\hat{\beta}_n-\beta)}$.

2.1 Test About a Single Regression Coefficient

Consider the vector of regression coefficients, $\beta = (\beta_0, \ldots, \beta_k)'$, and suppose that we are interested in the jth coefficient, β_j. We can let $d = (0, \ldots, 0, 1, 0, \ldots, 0)'$ denote the jth unit-vector (the vector which has 1 as its jth element and zero otherwise). Then we note that

$$d'\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}(d'\hat{\beta} - d'\beta) = \sqrt{n}(\hat{\beta}_j - \beta_j),$$

and by Theorem 4 it follows that $\sqrt{n}(\hat{\beta}_j - \beta_j) = d'\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N_1(0, d'\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}d)$. So for large n it holds that

$$\hat{\beta}_j - \beta_j \overset{A}{\sim} N_1(0, d'\hat{\Sigma}_{\hat{\beta}}d),$$

which allows us to construct the t-statistic of the hypothesis $H_0 : \beta_j = c$. It is given by

$$t_{\beta_j = c} = \frac{\hat{\beta}_j - c}{\sqrt{d'\hat{\Sigma}_{\hat{\beta}}d}},$$

which for large n, is approximately distributed as a standard normal, $N(0, 1)$. (For moderate values of n, it is typically better to use the t-distribution with $n - k - 1$ degrees of freedom.)

2.2 Test About a Multiple Regression Coefficients

To test hypotheses that involve multiple coefficients we need the following result.

Theorem 5 *Let* $Z \sim N_m(\mu, \Omega)$, *for some vector,* μ, $(m \times 1)$, *and some (full rank) matrix* Ω, $(m \times m)$. *Then it holds that*

$$(Z - \mu)'\Omega^{-1}(Z - \mu) \sim \chi^2_m.$$

Here we use χ^2_m to denote the chi-squared distribution with m degrees of freedom. In our asymptotic analysis the result we need is the following.

Theorem 6 *Let* $Z_n \xrightarrow{d} N_m(\mu, \Omega)$ *for some vector,* μ, $(m \times 1)$, *and some (full rank) matrix* Ω, $(m \times m)$. *Suppose that* $\hat{\mu} \xrightarrow{p} \mu$ *and that* $\hat{\Omega} \xrightarrow{p} \Omega$. *Then it holds that* $(Z_n - \hat{\mu})'\hat{\Omega}^{-1}(Z_n - \hat{\mu}) \xrightarrow{d} \chi^2_m$.

In our setting we have established that $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N_{k+1}(0, \Sigma_{\sqrt{n}(\hat{\beta}-\beta)})$ and $\hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)} \xrightarrow{p} \Sigma_{\sqrt{n}(\hat{\beta}-\beta)}$. Thus the theorem tells us that

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta)' \left[\hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)}\right]^{-1} \sqrt{n}(\hat{\beta} - \beta) &= (\hat{\beta} - \beta)' \left[\frac{1}{n}\hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)}\right]^{-1} (\hat{\beta} - \beta) \\
&= (\hat{\beta} - \beta)' \left[\hat{\Sigma}_{\hat{\beta}}\right]^{-1} (\hat{\beta} - \beta) \xrightarrow{d} \chi^2_{k+1}.
\end{aligned}$$

This enables us to test the hypothesis that the vector of regression parameters equals a particular vector, e.g., $H_0 : \beta = \beta^o$. All we need to do is to compute $(\hat{\beta} - \beta^o)' \left[\hat{\Sigma}_{\hat{\beta}}\right]^{-1} (\hat{\beta} - \beta^o)$ and compare this (scalar) number to the quantile (e.g. the 95%-quantile) of the χ^2_{k+1}-distribution.

An important distribution that is closely related to the χ^2-distribution is the $F_{q,\infty}$-distribution. It is defined as follows. Suppose that $Z \sim \chi^2_q$, then $U = Z/q \sim F_{q,\infty}$. So an $F_{q,\infty}$ is simply a χ^2_q that has been divided by its degrees of freedom. So, should we prefer to use an F-test to test $H_0 : \beta = \beta^o$, we would simply use that

$$F_{\beta=\beta^o} = \frac{(\hat{\beta} - \beta^o)' \left[\hat{\Sigma}_{\hat{\beta}}\right]^{-1} (\hat{\beta} - \beta^o)}{k+1} \xrightarrow{d} F_{k+1,\infty},$$

where $F_{\beta=\beta^o}$ denotes the test-statistic and $F_{k+1,\infty}$ represents the F-distribution with $(k+1, \infty)$ degrees of freedom. (An F-distribution has two degrees of freedom).

Typically we are interested in more complicated hypotheses than $\beta_j = c$ or $\beta = \beta^o$. A general class of hypotheses can be formulated as $H_0 : R\beta = r$, for some $q \times k+1$ matrix, R, and some $q \times 1$ vector, r.

How can we test hypotheses of this kind? First we note that Theorem 4 gives us that,

$$R\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N_q(R \cdot 0, R\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}R') = N_q(0, R\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}R').$$

The left hand side can be rewritten as $R\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}(R\hat{\beta} - R\beta) = \sqrt{n}[(R\hat{\beta} - r) - (R\beta - r)]$ which equals $\sqrt{n}(R\hat{\beta} - r)$ if the null hypothesis is true. So if we divide by $\sqrt{n}$ we get that $(R\hat{\beta} - r) \overset{A}{\sim} N_q(0, \frac{1}{n}R\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}R') = N_q(0, R\Sigma_{\hat{\beta}}R')$. Thus by using Theorem 6 we can construct a χ^2-test of the hypothesis $H_0 : R\beta = r$, using the test statistic: $(R\hat{\beta} - r)' \left[R\hat{\Sigma}_{\hat{\beta}}R'\right]^{-1} (R\hat{\beta} - r) \xrightarrow{d} \chi^2_q$, or the equivalent F-test, which is based on the statistics

$$F_{R\beta=r} = \frac{(R\hat{\beta} - r)' \left[R\hat{\Sigma}_{\hat{\beta}}R'\right]^{-1} (R\hat{\beta} - r)}{q} \xrightarrow{d} F_{q,\infty}.$$

Tables with critical values for the χ^2_m-distribution can be found in S&W on page 645. The F_{m_1,m_2}-distribution is tabulated on pages 647-649, and the $F_{m,\infty}$-distribution (which you will use most frequently) is tabulated on page 646, and (conveniently) on the very last page of the book.

2.3 A Simple Example

Suppose that we have estimated

$$\hat{\beta} = \begin{pmatrix} 4.0 \\ 2.5 \\ -1.5 \end{pmatrix} \qquad \text{and} \qquad \hat{\Sigma}_{\hat{\beta}} = \begin{pmatrix} \frac{801}{40} & \frac{27}{8} & -6 \\ \frac{27}{8} & \frac{3}{4} & -\frac{9}{8} \\ -6 & -\frac{9}{8} & \frac{15}{8} \end{pmatrix},$$

and wanted to test the following hypotheses:

1. $H_1 : \beta_3 = 0$. We note that H_1 is equivalent

   $$R_1\beta = r_1, \qquad \text{where } R_1 = (0, 0, 1) \text{ and } r_1 = 0.$$

   We find that

   $$R_1\hat{\beta} = -1.5 \qquad \text{and} \qquad R_1\hat{\Sigma}_{\hat{\beta}}R_1' = \frac{15}{8},$$

   such that the F-statistic of H_1 is given by

   $$\begin{aligned} F_{\beta_3=0} &= (R_1\hat{\beta} - r_1)' \left[R_1\hat{\Sigma}_{\hat{\beta}}R_1'\right]^{-1} (R_1\hat{\beta} - r_1) \\ &= (-1.5) \left[\frac{15}{8}\right]^{-1} (-1.5) = 1.2 \overset{A}{\sim} F_{1,\infty}. \end{aligned}$$

Note that when we test a single restriction, the F-statistic is the square of the t-statistic

$$F_{\beta_3=0} = \left(t_{\beta_3=0}\right)^2 = \left(\frac{\hat{\beta}_3 - 0}{\sqrt{\widehat{\mathrm{var}}(\hat{\beta}_3)}}\right)^2 .$$

2. $H_2 : \beta_2 = \beta_3$. This hypothesis is equivalent to $\beta_2 - \beta_3 = 0$, so we set

$$R_2 = (0, 1, -1) \text{ and } r_2 = 0,$$

and find

$$R_2\hat{\beta} = 2.5 - (-1.5) = 4, \qquad \text{and} \qquad R_2\hat{\Sigma}_{\hat{\beta}}R_2' = \frac{39}{8}.$$

So the F-test is given by

$$F_{\beta_2=\beta_3} = R_2\hat{\beta}\left[R_2\hat{\Sigma}_{\hat{\beta}}R_2'\right]^{-1} R_2\hat{\beta} = 4\left[\frac{39}{8}\right]^{-1} 4 = 3.2821 \overset{A}{\sim} F_{1,\infty}.$$

3. $H_3 : \beta_2 + \beta_3 = 0$? Here we set

$$R_3 = (0, 1, 1) \text{ and } r_3 = 0,$$

We find

$$R_3\hat{\beta} = 2.5 - 1.5 = 1, \qquad \text{and} \qquad R_3\hat{\Sigma}_{\hat{\beta}}R_3' = \frac{3}{8},$$

and the F-test is given by

$$F_{\beta_2+\beta_3=0} = 1\left[\frac{3}{8}\right]^{-1} 1 = 2.6666 \overset{A}{\sim} F_{1,\infty}.$$

4. $H_4 : \beta_2 = \beta_3 = 0$. Now we set

$$R_4 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; r_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Then

$$R_4\hat{\beta} = \begin{pmatrix} 2.5 \\ -1.5 \end{pmatrix} \qquad \text{and} \qquad R_4\hat{\Sigma}_{\hat{\beta}}R_4' = \begin{pmatrix} \frac{3}{4} & -\frac{9}{8} \\ -\frac{9}{8} & \frac{15}{8} \end{pmatrix},$$

such that

$$F_{\beta_2=\beta_3=0} = \frac{\begin{pmatrix} 2.5 & -1.5 \end{pmatrix}\begin{pmatrix} \frac{3}{4} & -\frac{9}{8} \\ -\frac{9}{8} & \frac{15}{8} \end{pmatrix}^{-1}\begin{pmatrix} 2.5 \\ -1.5 \end{pmatrix}}{2} = 17.666 \overset{A}{\sim} F_{2,\infty}.$$

3 Calculus with Conditional Expectations

When X_i is random, the concept of *conditional expectation* is an important tool in the analysis.

A conditional expected value, is written $E(Y|X = x)$, and denotes the expected value Y, when we know that $X = x$. (So $E(Y)$ is the expected value of Y when we know nothing about other random variables). When we write $E(Y|X)$, we think of it as a function of X. Since X is random then $E(Y|X)$ is also random (in the general case).

Some properties of conditional expected values are listed below.

1. If X and Y are independent, then $E(Y|X) = E(Y)$.
2. If $\mathrm{cov}(X,Y) \neq 0$, then $E(Y|X)$ depends on the X, in which case $E(Y|X)$ is a random variable (as X is random). So it is meaningful to talk about the expected value of a conditional expected value.
3. $E(Yf(X)|X) = E(Y|X)f(X)$, for any function f. So (functions of) the variable we condition on, can be taken outside the conditional expectation.
4. $E(Y) = E[E(Y|X)]$.

The properties of conditional expectations are very useful for our analysis. We shall make use of arguments such as the following example.

Example 7 *Suppose that* $E(u|X) = 0$. *Then we have that*

$$E(uX) \overset{(4)}{=} E[E(uX|X)] \overset{(3)}{=} E[E(u|X)X] = E[0X] = 0,$$

where the numbers refer to the properties listed above.