- *Multiple Regression Analysis: Asymptotic Properties*. Wooldridge (2013), Chapter 5
    - Consistency - Introduction
    - Consistency of the OLS estimator
    - Inconsistency when we ignore one regressor.
    - Asymptotic Normality-Introduction
    - Asymptotic Normality of the OLS estimator
    - Lagrange Multiplier statistic

Consistency-Introduction

- Consistency is a minimal requirement for an estimator:

  "If you can't get it right as n goes to infinity, you shouldn't be in this business."

  C.W.Granger, 2003, Nobel Prize Winner in Economics

- In some cases an estimator does not have the desirable properties when the sample size is small, but when the sample is large some of the desirable properties might hold. As we let the sample size go to infinite we call these properties *asymptotic properties*.

Definition

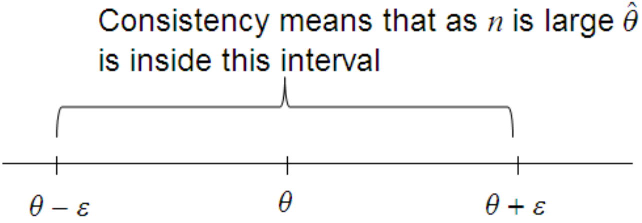
An estimator $\hat{\theta}$ is said to be a *consistent* estimator of θ if

$$\lim_{n\to\infty} \mathcal{P}(\theta - \varepsilon < \hat{\theta} < \theta + \varepsilon) = 1,$$

for all $\varepsilon > 0$. This property is often expressed as plim $\hat{\theta} = \theta$. If $\hat{\theta}$ is not consistent for θ, we say that it is *inconsistent*.

Intuition behind the technical definition of consistency

- Take ε equal to a very small number say $\varepsilon = 0.00001$.
- the sample size n is equal to a very large number say $n = 1000000$.



Thus if the sample size n is large $\hat{\theta}$ is close to θ.

- Notice that there are estimators that are unbiased but are not consistent.
- **Example:** if we have a random sample $(X_1, ..., X_n)$ and we would like to estimate the population mean $\mu_X = E(X_i), i = 1, ..., n$.
  - We consider the first observation X_1 as an estimator for the population mean.
  - Then, $E(X_1) = \mu_X$ (and therefore it is an unbiased estimator).
  - However, it can be shown that X_1 is not a consistent estimator of μ_X.

Proposition

If

$$\lim_{n\to\infty} E\left(\hat{\theta}\right) = \theta$$

and

$$\lim_{n\to\infty} Var\left(\hat{\theta}\right) = 0,$$

then plim $\hat{\theta} = \theta$.

Consider a random sample $(X_1, ..., X_n)$ such that $E(X_i) = \mu_X$ and $V(X_i) = \sigma_X^2$ (finite) where $i = 1, ..., n$ and let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

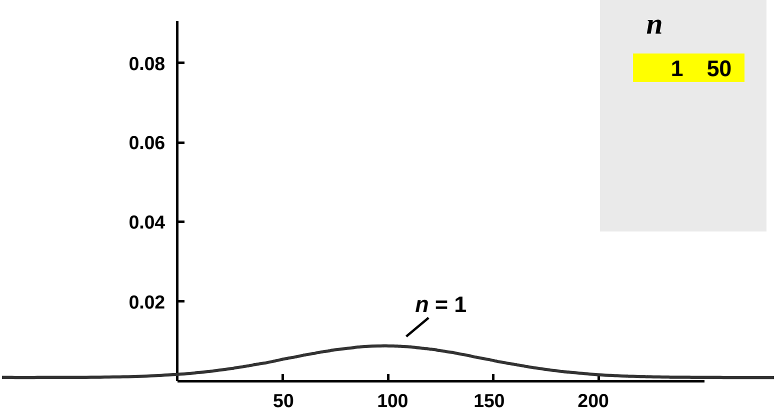
Properties of $\overline{X}$:

1. $E\left(\overline{X}\right) = \mu_X$.
2. $Var\left(\overline{X}\right) = \sigma_X^2/n$.
3. plim $\overline{X} = \mu_X$. (*Law of Large Numbers -LLN*)

Remark: An estimator can be biased and consistent.

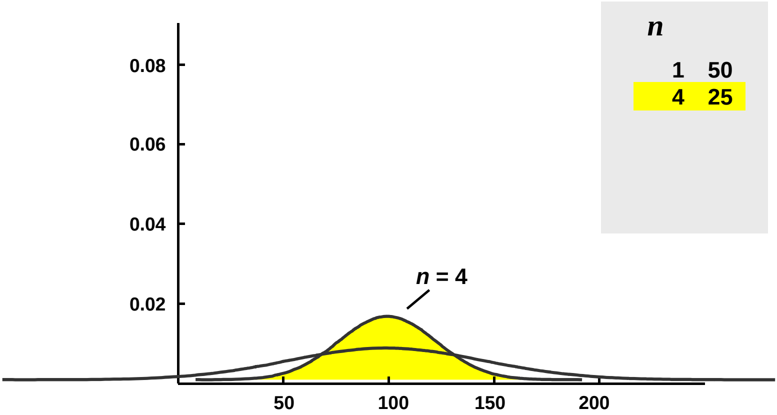
Example: $\widetilde{X} = \frac{1}{n+1}\sum_{i=1}^{n} X_i$.

Consistency of $\bar{X}$ for μ_X (intuition)



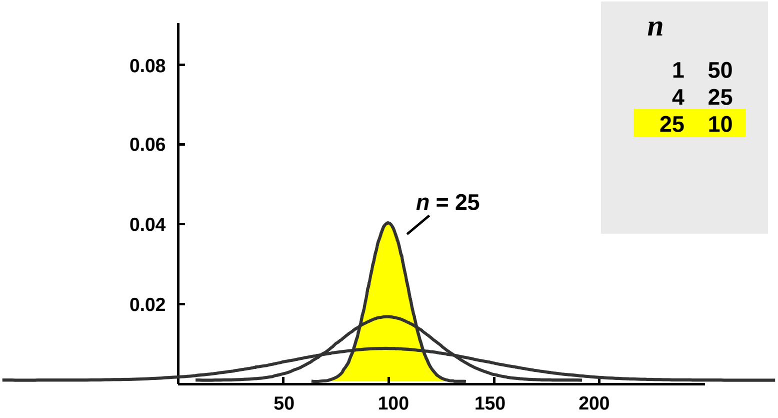
Density function of $\bar{X}$ ($\mu_X = 100$, $\sigma_X = 50$)

Consistency of $\bar{X}$ for μ_X (intuition)



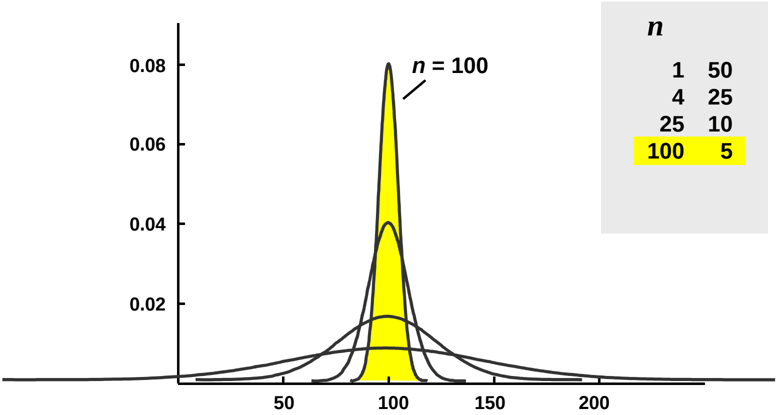
Density function of $\bar{X}$ ($\mu_X = 100$, $\sigma_X = 50$)

Consistency of $\bar{X}$ for μ_X (intuition)



Density function of $\bar{X}$ ($\mu_X = 100, \sigma_X = 50$)

Consistency of $\bar{X}$ for μ_X (intuition)



Density function of $\bar{X}$ ($\mu_X = 100$, $\sigma_X = 50$)

Remarks on Consistency:
In practice we deal with finite samples, not infinite ones. So why should we be interested in whether an estimator is consistent?

1. Sometimes it is impossible to find an estimator that is unbiased for small samples. If you can find one that is at least consistent, that may be better than having no estimate at all.
2. Often we are unable to say anything at all about the expectation of an estimator. The expected value can be applied in relatively simple contexts.

Consistency-Introduction

There are some theoretical results that help us to prove the consistency of estimators.

Property PLIM.1: Let $g(.)$ be a continuous function and suppose that

$$\text{plim}\ \hat{\theta} = \theta,$$

then

$$\text{plim}\ g(\hat{\theta}) = g(\theta).$$

Example: If $\hat{\theta} = \bar{X}$, $\text{plim}\ \bar{X} = \mu_X$. Consider the function $g(x) = x^2$. This function is continuous, therefore

$$\text{plim}\ \bar{X}^2 = \mu_X^2.$$

Example: Exercise C.5. of Wooldridge (2013, page 792)
Let Y denote a *Bernoulli*(θ) random variable, that is

$$Y = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{with probability } 1-\theta \end{cases}$$

where $0 < \theta < 1$. Given a random sample $(Y_1, ..., Y_n)$ we know that a consistent estimator for θ is $\bar{Y} = \sum_{i=1}^{n} Y_i/n$. We are interested in estimating the odds ratio

$$\gamma = \frac{\theta}{1-\theta},$$

which is the probability of success over the probability of failure. Show that

$$G_n = \frac{\bar{Y}}{1-\bar{Y}}$$

is a consistent estimator of γ.

Property PLIM.2: If plim $T_n = \alpha$ and plim $U_n = \beta$ then

(i)

$$\text{plim}\,(T_n + U_n) = \alpha + \beta.$$

(ii)

$$\text{plim}\,(T_n U_n) = \alpha\beta.$$

(iii)

$$\text{plim}\,\frac{T_n}{U_n} = \frac{\alpha}{\beta}.$$

Consistency of the OLS estimator

- Consider the multiple regression model

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

- Under some assumptions the OLS estimators $\hat{\beta}_j, j = 0, ..., k$ are unbiased. That is,

$$E(\hat{\beta}_j) = \beta_j,$$

  $j = 0, ..., k.$

- The required assumptions are:
  1. Linearity.
  2. Random sampling.
  3. No perfect multicollinearity.
  4. Zero Conditional Mean.
- But are these estimators consistent?
- That is, as $n \to \infty$, is $\hat{\beta}_j$ close to β_j?

Consistency of the OLS estimator

- Under the assumptions required for unbiasedness, the OLS estimator is consistent.
- To show this let us focus on the simple linear regression model for simplicity
$$y = \beta_0 + \beta_1 x + u.$$
- Consistency can be proved for the simple regression case in a manner similar to the proof of unbiasedness.
- Will need to take probability limit (plim) to establish consistency.

Consistency of the OLS estimator

Sketch of the prove of consistency of the OLS estimator $\hat{\beta}_1$ for the slope β_1 in the simple linear model:
The objective is to prove that

$$\text{plim } \hat{\beta}_1 = \beta_1.$$

That is, if the size of the sample is large $\hat{\beta}_1$ is close to β_1.
Sketch of the prove:

- Recall that the Ordinary Least Squares estimator for the slope β_1 is given by
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$
- Plug in $y_i = \beta_0 + \beta_1 x_i + u_i$ and we obtain
$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Consistency of the OLS estimator

- Notice that
$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$
- Using the properties above in can be shown that
$$\text{plim}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,u_i = E[(x - E(x))u]$$
and by definition $E\left[(x - E(x))u\right] = Cov(x, u)$.
- In addition
$$\text{plim}\ \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = E[(x - E(x))^2],$$
and by definition $E[(x - E(x))^2] = Var(x)$.
- Hence we can write
$$\text{plim}\ \hat{\beta}_1 = \beta_1 + \frac{Cov(x, u)}{Var(x)}$$
- If $Cov(x, u) = 0$ and $Var(x) > 0$, it follows that plim $\hat{\beta}_1 = \beta_1$

Consistency of the OLS estimator

- Similarly it can be shown that the OLS estimator is consistent in the Multiple Regression Model.

Remark: We require weaker Assumptions to prove consistency than to prove unbiasedeness.
Consider the multiple regression model

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

- To prove unbiasedness of the OLS estimator we assumed a zero conditional mean – $E(u|x_1, x_2, \ldots, x_k) = 0$.
- To prove consistency of the OLS estimator we can have the weaker assumption of zero mean and zero correlation – $E(u) = 0$ and $Cov(x_j, u) = 0$, for $j = 1, 2, \ldots, k$.
- Without these assumptions, OLS will be biased and inconsistent.

Inconsistency when we ignore regressors

- Assume that the true model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v,$$

where the error term v satisfies $E[v|x_1, x_2] = 0$.

- Suppose you think incorrectly that the model is

$$y = \beta_0 + \beta_1 x_1 + u.$$

- That is x_2 was omitted from the model. Hence, the error is

$$u = \beta_2 x_2 + v.$$

- Suppose we run a regression of y on x_1. The OLS estimator for β_1 is

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1) y_i}{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2}.$$

Inconsistency when we ignore regressors

- Then, it can be shown that

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \beta_2 \delta,$$

  where

$$\delta = \frac{Cov(x_1, x_2)}{Var(x_1)}$$

- The difference between plim $\tilde{\beta}_1$ and β_1 is called *asymptotic bias* and in this case is equal to $\beta_2 \delta$, where $\delta = \frac{Cov(x_1, x_2)}{Var(x_1)}$.
- So, thinking about the direction of the *asymptotic bias* is just like thinking about the direction of *bias* for an omitted variable.
- Main difference is that asymptotic bias uses the *population variance and covariance*, while bias uses the *sample variance and covariance*.
- Remember, inconsistency is a large sample problem – it doesn't go away as add data.

Asymptotic Normality-Introduction

- Consider a random variable $(X_1, ..., X_n)$ and assume that $X_i \sim N(\mu_X, \sigma_X^2), i = 1, ..., n$, then

$$\sum_{i=1}^{n} X_i \sim N(n\mu_X, n\sigma_X^2).$$

- Thus

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu_X, \sigma_X^2/n).$$

- If we standardize we have the statistic

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X} \sim N(0, 1)$$

- This result is *the main building block* to develop test statistics to test hypothesis and construct confidence intervals for μ_X.

Asymptotic Normality-Introduction

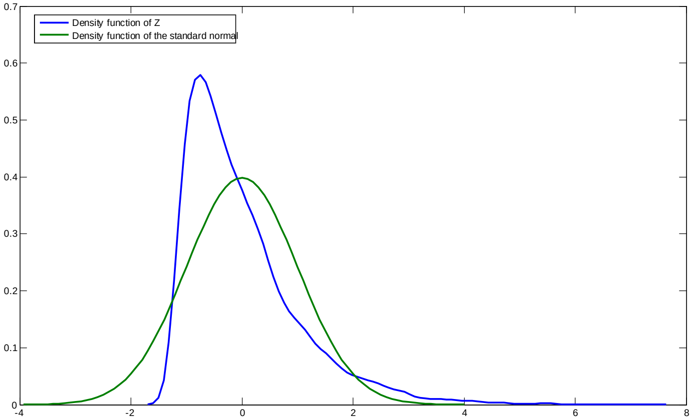
- However, what happens if the random variables X_i are not normally distributed?
- In that case we resort to the following theorem
- **Central Limit Theorem (CLT):** If the X_i in the sample are all drawn independently from the same distribution (the distribution of X), and provided that this distribution has finite variance, the distribution of

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X}$$

will converge to a standard normal distribution as n tends to infinity. We write $Z \overset{a}{\sim} N(0,1)$ where the symbol $\overset{a}{\sim}$ reads “distributed asymptotically” (it means that if the sample size is large the distribution of Z is close to the standard normal).

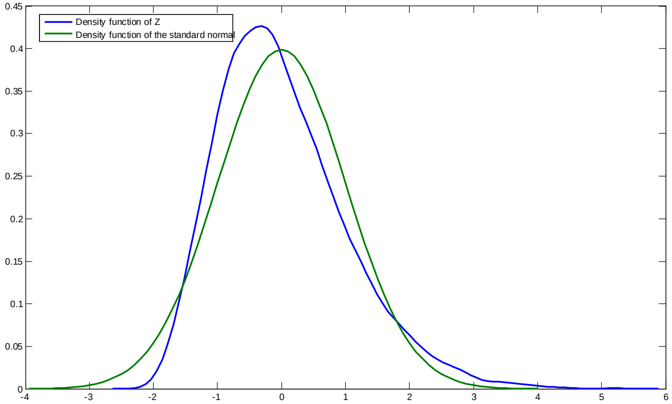
Asymptotic Normality-Introduction

Example: Sample Distribution of Z when $X_i \sim \chi^2(1), i = 1, ..., n$.
Sample size $n = 3$.



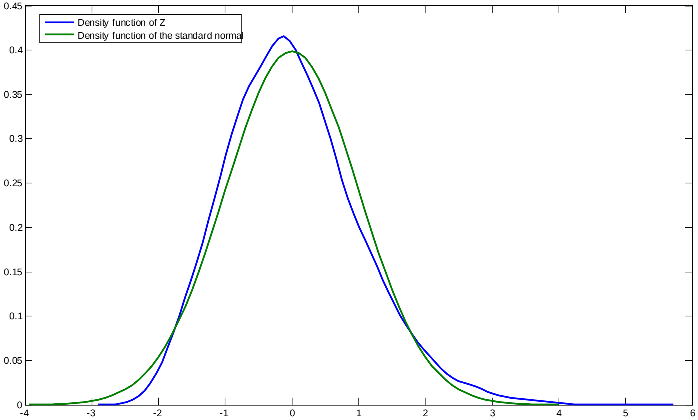
Asymptotic Normality-Introduction

Example: Sample Distribution of Z when $X_i \sim \chi^2(1), i = 1, ..., n$.
Sample size $n = 10$.



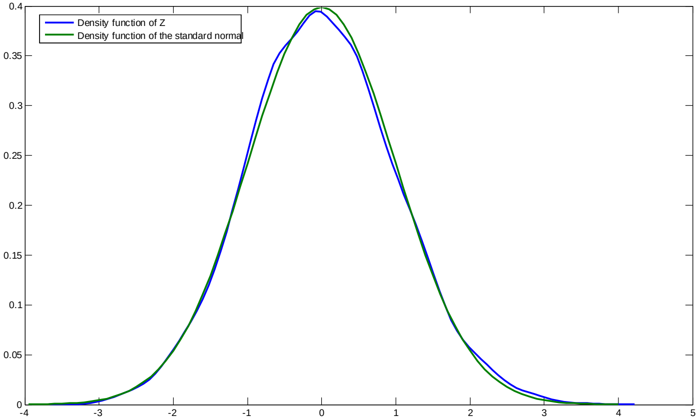
Asymptotic Normality-Introduction

Example: Sample Distribution of Z when $X_i \sim \chi^2(1), i = 1, ..., n$.
Sample size $n = 25$.



Asymptotic Normality-Introduction

Example: Sample Distribution of Z when $X_i \sim \chi^2(1), i = 1, ..., n$..
Sample size $n = 500$.



Asymptotic Normality of the OLS estimator

Consider the multiple regression model

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

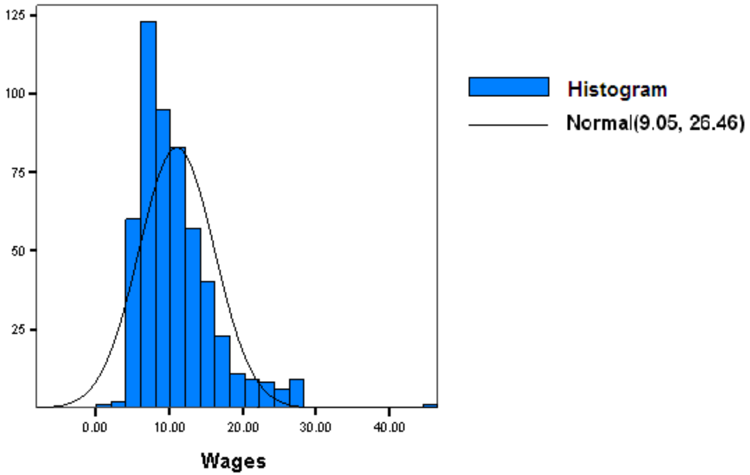
- Recall that under the *Classical Linear Model* assumptions, the sampling distributions are normal, so we could derive t and F distributions for testing.
- *Classical Linear Model* assumptions:
    1. Linearity.
    2. Random sampling.
    3. No perfect multicollinearity.
    4. Zero Conditional Mean.
    5. The regressors are independent of the error term and $u \sim N(0, \sigma^2)$.
- This exact normality was due to assuming the population error distribution was normal.
- This assumption of normal errors implied that the distribution of y, given the x's, was normal as well.

Asymptotic Normality of the OLS estimator

- Easy to come up with examples for which this exact normality assumption will fail.
- Variables, like wages, arrests, savings, etc. don't have symmetric distribution functions, hence can't be normal.

Asymptotic Normality of the OLS estimator

Histogram of Wages



Asymptotic Normality of the OLS estimator

- Normality assumption not needed to conclude OLS is *BLUE*, only for inference.
- Based on the *central limit theorem*, we can show that OLS estimator is *asymptotically normal*.
- *Asymptotic Normality* means that the distribution of the estimator (after a standardization) converges to the standard normal distribution function as $n \rightarrow \infty$. That is, the distribution of the standardized estimator is close to the the distribution of a standard normal random variable if the sample size is large.

Consider the multiple regression model

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

and denote the OLS estimators for the parameters $\beta_0, \beta_1, ..., \beta_k$ as $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$.
Under the Gauss-Markov Assumptions it is possible to use the *CLT* and the *LLN* to prove the following results:
1- Denote $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - ... - \hat{\beta}_k x_{ki}$, $i = 1, ..., n$, the residuals. It can be shown that

$$\text{plim } \hat{\sigma}^2 = \sigma^2,$$

where

$$\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{u}_i^2 \text{ and } \sigma^2 = Var(u).$$

That is, $\hat{\sigma}^2$ is a *consistent* estimator of σ^2.

Asymptotic Normality of the OLS estimator

2- For $j = 0, ..., k$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \overset{a}{\sim} N(0, 1),$$

where $se(\hat{\beta}_j)$ is the usual OLS *standard error*:

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j \left(1 - R_j^2\right)}},$$

the $SST_j = \sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from the regressing x_j on all other $x's$.

Asymptotic Normality of the OLS estimator

- Result **2** can be used to test hypothesis on the parameters or to construct confidence intervals.
- **Remark:** Note that the CLT *does not imply* that the error term is standard normal in large samples.
- Note that while we no longer need to assume normality with a large sample, we do still need *homoskedasticity*.
- Notice that since the $t(n-k-1)$ distribution is close to the $N(0,1)$ distribution when n is large, there is no problem in using the table of the t distribution to obtain the critical values.
- Asymptotic Normality of the OLS estimators also implies that the F statistic have approximate F distribution in large samples. So nothing changes from what we have done before.
- If u is not normally distributed, we sometimes will refer to the *standard error* as an *asymptotic standard error*.

Lagrange Multiplier statistic

- Once we are using large samples and relying on asymptotic normality for inference, we can use more than the t and F statistics.
- The *Lagrange Multiplier* or *LM* statistic is an alternative for testing multiple exclusion restrictions.
- Because the *LM* statistic uses an auxiliary regression it's sometimes called an nR^2 statistic.
- This test statistic does not require the assumption of normality, but we do still need *homoskedasticity*.

Lagrange Multiplier statistic

Suppose we have a standard model,
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_q x_q + \beta_{q+1} x_{q+1} + ... + \beta_k x_k + u$ and our null hypothesis is $H_0 : \beta_1 = 0, ..., \beta_q = 0$ vs $H_1 : H_0$ not true.

Steps to construct the LM test statistic

1. First, we run the restricted model, that is, we run the regression of y on an intercept and $x_{q+1}, x_{q+2}, ..., x_k$.
2. Now take the residuals of this regression $\tilde{u}$ and regress $\tilde{u}$ on an intercept and $x_1, x_2, ..., x_k$ (i.e. all variables). Denote R_u^2 the R^2 of this regression.
3. The LM statistic is given by $LM = nR_u^2$.

Lagrange Multiplier statistic

It can be shown that

$$LM \overset{a}{\sim} \chi^2(q),$$

that is the distribution of LM is asymptotically chi-square with q degrees of freedom.

- Suppose we would like to test H_0 at the α level.
- Denote c^{act} the actual value of the LM statistic in a given sample.
- The critical value is denoted as c_α and corresponds to the percentile $(1-\alpha)^{th}$ of the $\chi^2(q)$ distribution, that this the constant that satisfies

$$\mathcal{P}(X > c_\alpha) = \alpha,$$

  where X is a random variable with the $\chi^2(q)$ distribution.
- *Rejection rule:* Reject H_0 if $c^{act} > c_\alpha$.
- Alternatively, compute the p-value as

$$p - value = \mathcal{P}(X > c^{act}),$$

where X is a random variable with the $\chi^2(q)$ distribution.
Rejection rule: Reject H_0 if $p - value < \alpha$.

Lagrange Multiplier statistic

Example: We are studying the returns to education at junior colleges and four year colleges (universities) and we have the model

$$\log(wages) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u,$$

We would like to test

$$H_0 : \beta_1 = 0, \beta_2 = 0$$

The alternative hypothesis is

$$H_1 : \beta_1 \neq 0 \text{ and/or } \beta_2 \neq 0.$$

Lagrange Multiplier statistic

Running the regression of $\log(wages)$ on *exper* we obtained

$$\widehat{\log(wages)} = 1.70919 + 0.0044\, exper,$$
$$n = 6763, R^2 = 0.0911$$

Denote *Res1* the residuals of the regression above. Running the regression of *Res1* on *jc*, *univ* and *exper* we obtain

$$\widehat{Res1} = -0.23686 + 0.00054 jc + 0.0667 univ + 0.07688 exper,$$
$$n = 6763,\ R^2 = 0.1445$$

Test H_0 at 5% level:

Lagrange Multiplier statistic

- With a large sample, the result from an F test and from an LM test should be similar, though in finite samples the results of both tests will not be identical.