The multiple regression model: Inference

Chapters 4: Introductory Econometrics 771

Prof Dieter von Fintel

Department of Economics
Stellenbosch University

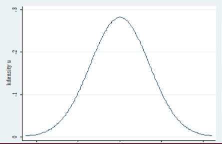
21 February 2024

Overview

Sampling Distributions of OLS Estimators

$\widehat{\beta}_j$ is estimated using **one** possible sample of **many** from the population: they are therefore **random variables**

- Statistical **theory**: estimates unbiased $E(\widehat{\beta}_j|X) = \beta_j$ under **MLR 1-4**
- Adding **MLR5** gives a formula for how "spread" out the distributions of the random variables are: $Var(\widehat{\beta}|X) = (X'X)^{-1}\sigma^2$
  - Only *one* of many possible samples observed (=no variation), need formula
- Still don't know about the **shape** of the distribution
  - All we know: distribution of $\widehat{\beta}_j$ depends on distribution of $\widehat{u}$
- Add **MLR**.6 – NORMALITY
  - Errors normally distributed $u \sim N(0; \sigma^2)$ & independent of X



Focus on distribution of u

- Recall

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{u}$$
$$= \boldsymbol{\beta} + \sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{x}_i')^{-1}\boldsymbol{x}_i u_i$$

  - ...or a function of a sum containing u_i
  - A linear combination of $u_i \sim N$ is normally distributed
  - Therefore $\widehat{\beta}$ is normally distributed if we assume $u_i \sim N$
- Put it all together:
  - From MLR1-4: $E(\widehat{\beta}_j|X) = \beta_j$
  - From MLR5: we get formula for $Var(\widehat{\beta}_j|X)$
  - Now add MLR6: $\widehat{\beta}_j \sim N(\beta_j; Var(\beta_j))$

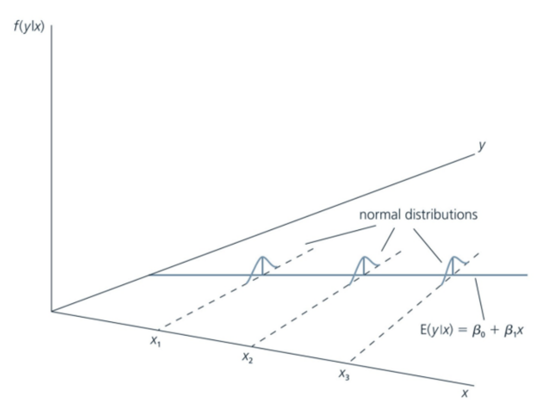
Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT forward together sonke siya phambili saam vorentoe

$$u \sim N(0;\sigma^2)$$

$$\Rightarrow y|\boldsymbol{x} \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k;\sigma^2) \text{ if MLR1-6 hold}$$

FIGURE 4.1

The homoskedastic normal distribution with a single explanatory variable.



Normality: valid or "convenient"?

- Generally valid by the **Central Limit Theorem**
  - Residuals represent unmodelled variables
  - Sum of many independently distributed variables $\rightarrow$ normal distribution when n is "large"
  - But we need to establish empirically whether this holds ($S \rightarrow 0; K \rightarrow 3$)
  - Sometimes transformations of variables can introduce normality
    - $log(Price), log(Income)$
- But also convenient
  - Now $\widehat{\beta}_j$ are also normally distributed by properties of normal distribution (linear combination of normal residuals are normally distributed)
  - Normal distribution defined completely by its mean and variance
  - Normality allows us to construct other test statistics (F, T and χ^2) to test a variety of hypotheses

T-test

- Slight adjustment: if $u \sim Normal \Rightarrow \widehat{\beta}_j$ has a t-distribution **when** we **estimate** $\widehat{\sigma}^2 = \frac{SSR}{n-k-1}$ which is contained in $se(\widehat{\beta}_j)$ - this is *usually* the case

$$T = \frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)}$$

- Normality is only a valid assumption if population σ^2 is known
  - t-distribution is flatter and wider to account for additional uncertainty that comes from estimating $\widehat{\sigma}^2$
  - As n grows large, a t distribution tends towards a normal distribution (it grows flatter and more symmetric)
    - Normal is a relevant and frequent special case of the t distribution

T Test

- Hypotheses about **population** parameters
  - Question: is it possible that a **sample** estimate could have been drawn from a population with hypothesised properties?

$$PRF : log(wage) = \beta_0 + \beta_1 educ + \beta_2 exp + \beta_3 tenure + u$$

  - Does tenure influence wages once we have controlled for education and experience? **OR** is $\beta_3 = 0$ in the population?

Hypothesis testing

- Hypothesise a population value of estimate (β^*)

$$H_0 : \beta_j = \beta^*$$
$$H_a : \beta_j \neq \beta^*$$

  - How "close" is estimated value to the hypothesised population value?
    - Too far – **reject** the possibility
    - Close enough – **do not reject**
  - How close is "close" – *depends on* ***distribution*** *of estimate*
- Particular case:
  - Does x have an influence on y once other variables are controlled for?

$$H_0 : \beta_j = 0$$
$$H_a : \beta_j \neq 0$$

Insert hypothesised value: $T = \dfrac{\widehat{\beta_j} - 0}{se(\widehat{\beta_j})} \sim t_{n-(k+1)}$

What are we saying?

- We assume (hypothesise) a specific population distribution
  - Then we "place" the estimate in the context of the hypothesised distribution
  - If the hypothesised value is "too far" from the mean of this distribution, then we reject this hypothesised population value
    - Gives us a sense of what probability that estimate could be drawn from specific population
  - For instance, a large T value will be far from 0 in absolute value
    - Because $\widehat{\beta}$ is large and $se(\widehat{\beta})$ is low
  - We then reject that the estimate is 0 (ie there **IS** a relationship)
  - Statistically SIGNIFICANT
- Note that we do not "accept" a hypothesis if it is "too close"
  - We simply "do not reject"

Testing one-sided alternatives

$$H_0 : \beta_j = 0$$
$$H_a : \beta_j > 0$$

- Choose *significance level* (α)
  - The probability of rejecting H_0 when it could be true
    - Usually 1%, 5%, 10%
  - Acknowledging that we could be making mistakes in the process
- Establish "sufficiently" large value of $\widehat{\beta}_j$ where we conclude that it is "sufficiently" likely to be larger than zero
  - CRITICAL value
  - 95th percentile of *theoretical* t-distribution if 5% level of significance
- Rejection rule
  - If *calculated* T is larger than *critical* t, reject H_0 in favour of H_a

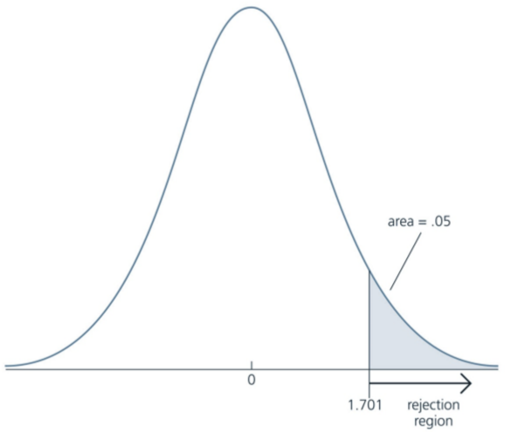
Critical t depends on...

- Degrees of freedom (see tables)
  - $n - k - 1$
- Level of significance (α) chosen
  - "Statistically (in)significant"
- Whether it is a one or two-tailed test
  - Always draw a picture and make sure how you split the distribution
  - Will see later...

Rejection rule

FIGURE 4.2

5% rejection rule for the alternative $H_1: \beta_j > 0$ with 28 *df*.

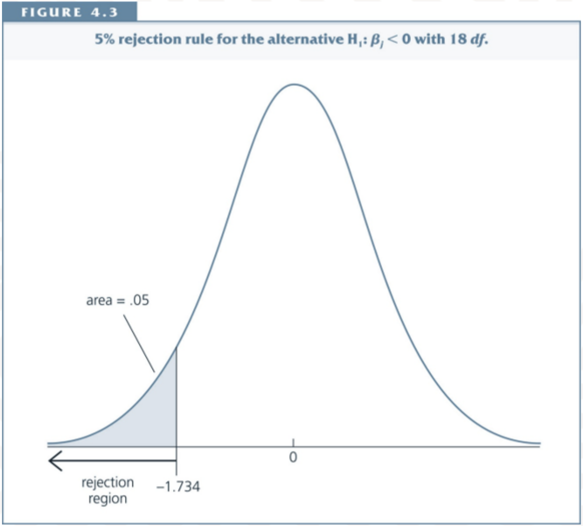


Testing one-sided alternatives

$$H_0 : \beta_j = 0$$
$$H_a : \beta_j < 0$$

- Very similar for testing whether estimate is negative
- Except now our rejection region is in bottom tail of distribution
  - And critical t is just the negative of the previous
  - Now if calculated value is *smaller* than the critical value, reject H_0 in favour of H_a
    - Same as taking the absolute value of the critical value and rejecting if calculated T is larger
    - Symmetry of T-distribution

Rejection rule



Illustration

- Test whether $H_0 : \beta_{educ} = 0$ or $H_a : \beta_{educ} > 0$ at a 5% level of significance

$$T = \frac{\widehat{\beta}_{educ} - 0}{se(\widehat{\beta}_{educ})} = \frac{0.1816909 - 0}{0.0019129} = 94.981$$

$$d.f. = n - (k+1) = 23436 - (3+1) = 23432$$

$$t_{23432;0.95} = 1.645$$

$T > t \Rightarrow$ Reject $H_0 \Rightarrow$ Statistically significant at a 100-95%=5% level

```
. reg lwage educ exp exp2

      Source |       SS           df       MS      Number of obs   =    23436
-------------+----------------------------------   F(  3, 23432)   =  3291.15
       Model |  8719.68073         3  2906.56024   Prob > F        =   0.0000
    Residual |  20693.8116    23432  .883143204   R-squared       =   0.2965
-------------+----------------------------------   Adj R-squared   =   0.2964
       Total |  29413.4923    23435  1.25510955   Root MSE        =   .93976

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1816909   .0019129    94.98   0.000     .1779414    .1854404
         exp |   .0365736   .0018376    19.90   0.000     .0329718    .0401754
        exp2 |  -.0002069   .0000351    -5.89   0.000    -.0002757   -.0001381
       _cons |  -.3953021   .0299979   -13.18   0.000    -.4540999   -.3365043
------------------------------------------------------------------------------
```

Illustration

- Test whether $H_0 : \beta_{exper^2} = 0$ or $H_a : \beta_{exper^2} < 0$ at a 1% level of significance

$$T = \frac{\widehat{\beta}_{educ} - 0}{se(\widehat{\beta}_{educ})} = \frac{-0.0002069 - 0}{0.0000351} = -5.8946 \Rightarrow |T| = 5.8646$$

$$d.f. = n - (k+1) = 23436 - (3+1) = 23432$$

$$t_{23432;0.99} = 2.326$$

$$|T| > t \Rightarrow \text{ Reject } H_0 \Rightarrow \text{ Statistically significant}$$

```
. reg lwage educ exp exp2

      Source |       SS       df       MS              Number of obs =   23436
-------------+------------------------------           F(  3, 23432) = 3291.15
       Model |  8719.68073     3  2906.56024           Prob > F      =  0.0000
    Residual |  20693.8116 23432  .883143204           R-squared     =  0.2965
-------------+------------------------------           Adj R-squared =  0.2964
       Total |  29413.4923 23435  1.25510955           Root MSE      =  .93976

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1816909   .0019129    94.98   0.000     .1779414    .1854404
         exp |   .0365736   .0018376    19.90   0.000     .0329718    .0401754
        exp2 |  -.0002069   .0000351    -5.89   0.000    -.0002757   -.0001381
       _cons |  -.3953021   .0299979   -13.18   0.000    -.4540999   -.3365043
------------------------------------------------------------------------------
```

Two-sided alternatives

$$H_0 : \beta_j = 0$$
$$H_a : \beta_j \neq 0$$

- This is the usual format used in hypothesis testing
  - Do not specify whether the effect is positive or negative a priori
  - We say it could be larger OR smaller than 0

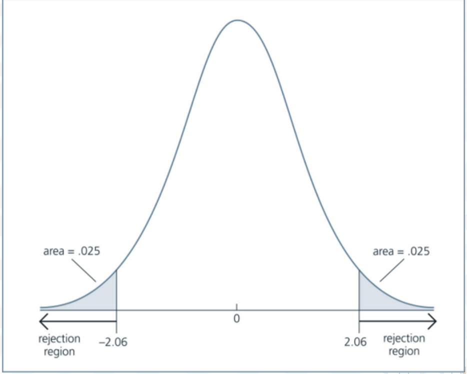
Rejection rules

"Split" the critical area: $0.025 + 0.025 = 0.05 \Rightarrow$ 5% level of significance

- **Reject** if T is *above* the *upper* critical value **or** *below lower* critical value
- Equivalent to checking whether $|T|$ is larger than upper critical value

FIGURE 4.4

5% rejection rule for the alternative $H_1: \beta_j \neq 0$ with 25 *df*.



Illustration

- Test whether $H_0 : \beta_{experience} = 0$ or $H_a : \beta_{experience} \neq 0$

$$T = \frac{\widehat{\beta}_{exp} - 0}{se(\widehat{\beta}_{exp})} = \frac{0.0365736 - 0}{0.0018376} = 19.90$$

$$d.f. = n - (k+1) = 23436 - (3+1) = 23432$$

$$t_{23432;\mathbf{0.995}} = 2.576 \text{ (where } 0.995 = 0.005 + 0.99)$$

$$|T| > t \Rightarrow \text{ Reject } H_0 \Rightarrow \text{ Statistically significant at } 1\% \text{ level}$$

```
. reg lwage educ exp exp2
```

Source	SS	df	MS
Model	8719.68073	3	2906.56024
Residual	20693.8116	23432	.883143204
Total	29413.4923	23435	1.25510955

Number of obs	= 23436
F(3, 23432)	= 3291.15
Prob > F	= 0.0000
R-squared	= 0.2965
Adj R-squared	= 0.2964
Root MSE	= .93976

lwage1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1816909	.0019129	94.98	0.000	.1779414	.1854404
exp	.0365736	.0018376	19.90	0.000	.0329718	.0401754
exp2	-.0002069	.0000351	-5.89	0.000	-.0002757	-.0001381
_cons	-.3953021	.0299979	-13.18	0.000	-.4540999	-.3365043

Smaller sample?

- Test for $\beta_{numeracy}$?

$$T = -0.08$$
$$t_{19;0.975} = 2.093$$
$$|T| < t \Rightarrow \text{Do } \textbf{not} \text{ reject } H_0 \Rightarrow \text{ Statistically } \textbf{in}\text{significant at } 5\% \text{ level}$$

```
. reg lwages educ numeracy if age==21
```

Source	SS	df	MS
Model	7.14221169	2	3.57110584
Residual	28.3220949	19	1.49063657
Total	35.4643066	21	1.6887765

Number of obs	=	22
F(2, 19)	=	2.40
Prob > F	=	0.1181
R-squared	=	0.2014
Adj R-squared	=	0.1173
Root MSE	=	1.2209

lwages	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.3901161	.1802502	2.16	0.043	.012848	.7673841
numeracy	-.0191996	.2461784	-0.08	0.939	-.534457	.4960577
_cons	-1.961189	1.946364	-1.01	0.326	-6.034976	2.112598

More generally

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT forward together sonke siya phambili saam vorentoe

- For instance, test that an elasticity (from a log-log model) equals $a_j = 1$

$$H_0 : \beta_j = a_j$$
$$T = \frac{\widehat{\beta}_j - a_j}{se(\widehat{\beta}_j)} \sim t_{n-(k+1)}$$
$$T = \frac{estimate - hypothesised}{se}$$

Hypothesis Testing – Some Tacit Knowledge

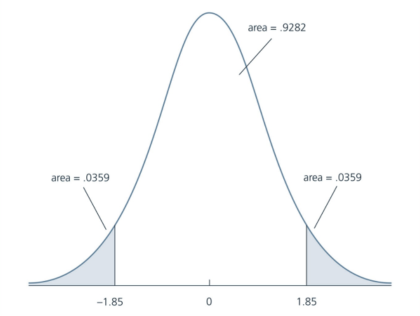
- *p*-value
  - Exact level of significance $P(|T| > |t|)$
  - The lowest level of significance at which one would reject H_0
  - Allows us to see what the conclusion would be at all levels of significance
  - Example: p-value = 0.04 (ie probability above calculated T-value is 0.04)
    - Draw a graph for yourself
    - At 1% that T value is not in rejection region
    - 3.9% do not reject
    - 4% reject
    - 4.1% reject
    - 5% reject
    - 10% reject
  - Small p-values lead to rejection of H_0

p-values for different alternatives

- Two-sided: Add area in both tails
- One-sided
  - Take only area in relevant tail according to H_a
  - Packages compute two-sided alternative
  - But for one-sided just divide by 2

FIGURE 4.6

Obtaining the p-value against a two-sided alternative, when $t = 1.85$ and $df = 40$.



Hypothesis Testing – Some Tacit Knowledge

- "Rejecting" and "Non-rejecting"
  - Never "accept" a hypothesis
  - As evident from confidence intervals (later), a range of possibilities could be the "true" population value
    - Which do we therefore "accept"?
  - Formulating null and alternative
    - Theory guides
    - Testing statistical properties (eg T-test of significance)

Choosing α

- Usually 1%, 5% and 10%
- "conventional" levels of statistical significance
  - But not grounded on any statistical basis

$$\alpha = P(\text{Type I Error}) = P(\text{Reject true hypothesis}) = \text{"false positive"}$$
$$P(\text{Type II Error}) = P(\text{Not Rejecting false hypothesis}) = \text{"false negative}$$

  - Trade-off!
  - Power of Test = 1 – P(Type II Error)
    - In practise: just choose α, but realise that you sacrifice power

Economic vs Statistical Significance

- McCloskey
- An estimate can "mean something" in economic magnitude even if it marginally fails a statistical test
- But, if two different values are statistically plausible, they may have varying practical implications
  - Eg small differences in MPC could have large impacts on consumption multiplier
- Often have statistically significant coefficient with no economically significant magnitude
  - Or it has the wrong sign
  - Must still go through some introspection of model
    - Ie don't blindly apply statistical criteria ONLY

Confidence Intervals

- Find a confidence interval for the population β

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$$

$$\vdots$$

$$P\left(\widehat{\beta}_j - t_{\alpha/2} se(\widehat{\beta}_j) \leq \beta_j \leq \widehat{\beta}_j - t_{\alpha/2} se(\widehat{\beta}_j)\right) = 1 - \alpha$$

$$(1-\alpha)\% \text{ confidence interval } = \widehat{\beta}_j \pm t_{\alpha/2} se(\widehat{\beta}_j)$$

- High Standard Error = Broad Confidence Interval
- Level of significence (α) = 5%
  - Confidence interval of 95%

Confidence Intervals - semantics

Stellenbosch University Iyunivesithi Universiteit — forward together sonke siya phambili saam vorentoe

- If α = 5%, is there a 95% chance that the true β lies within that specific interval? **NO!**
  - If 100 different samples were taken
  - ...would obtain 100 different estimates $\widehat{\beta}$
  - ...and 100 different confidence intervals
  - 95 out of the 100 times the true population β would fall inside the calculated confidence interval (subject to MLR1-6 holding!)
    - So there is still a 5% chance that a specific confidence interval does NOT contain the true β
    - Remember Type I error?
  - Can use to test hypothesis
    - $H_0 : \beta = 0$
    - If 0 lies within confidence interval, we do not reject at α%

- Calculate a 95% and a 99% confidence interval for $\beta_{education}$
  - $0.1816909 \pm 1.96 \times 0.0019129 = [0.17794162; 0.18544018]$
  - $0.1816909 \pm 2.576 \times 0.0019129 = [0.17676327; 0.18661853]$
  - Which is wider? Why?
- We get a “range” of possible population values
  - The reason why we never “accept” a hypothesis!

```
. reg lwage educ exp exp2
```

Source	SS	df	MS
Model	8719.68073	3	2906.56024
Residual	20693.8116	23432	.883143204
Total	29413.4923	23435	1.25510955

Number of obs	=	23436
F(3, 23432)	=	3291.15
Prob > F	=	0.0000
R-squared	=	0.2965
Adj R-squared	=	0.2964
Root MSE	=	.93976

lwage1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1816909	.0019129	94.98	0.000	.1779414	.1854404
exp	.0365736	.0018376	19.90	0.000	.0329718	.0401754
exp2	-.0002069	.0000351	-5.89	0.000	-.0002757	-.0001381
_cons	-.3953021	.0299979	-13.18	0.000	-.4540999	-.3365043

Confidence Intervals vs Test of Significance

- First gives range of *magnitudes*, while second only offers a *binary* conclusion
  - If test of significance is rejected, then what *is* a plausible value for the population value?
- However, in the second approach we can use p-values, while for confidence intervals we need to specify α upfront
  - Both approaches are therefore flexible in one sense or the other
- NB NOTE
  - ALL hypothesis tests depend on **MLR**1-6
    - If normality of errors is violated, $\widehat{\beta}$ are not t-distributed
    - If homoskedasticity is violated, standard errors are biased, and so are the calculated T-stats
    - If MLR 1-4 are violated, limited value for making conclusions about the population from biased estimates and inferences

Tests about more than one parameter

- Does an additional year at school have the same value as an additional year at university?

$$
\begin{aligned}
log(wage) &= \beta_0 + \beta_1 school + \beta_2 university + u \\
H_0 &: \beta_1 = \beta_2 \Leftrightarrow \beta_1 - \beta_2 = 0 \\
H_a &: \beta_1 < \beta_2 \Leftrightarrow \beta_1 - \beta_2 < 0 \\
T &= \frac{(\widehat{\beta}_1 - \widehat{\beta}_2) - 0}{se(\widehat{\beta}_1 - \widehat{\beta}_2)}
\end{aligned}
$$

- Similar approach to before, except *se* is not so easy to get from our normal regression output

Standard error?

$$Var(\widehat{\beta}_1 - \widehat{\beta}_2) = Var(\widehat{\beta}_1) + Var(\widehat{\beta}_2) - 2Cov(\widehat{\beta}_1; \widehat{\beta}_2)$$

$$se(\widehat{\beta}_1 - \widehat{\beta}_2) = \sqrt{Var(\widehat{\beta}_1 - \widehat{\beta}_2)}$$

- But we need to dig to find $Cov(\widehat{\beta}_1; \widehat{\beta}_2)$
- Another way is to estimate a similar but informative model, using resulting coefficient and standard errors with a T test

$$\begin{aligned} log(wage) &= \beta_0 + \beta_1 school + \beta_2 university + u \\ &\text{Let } \theta = \beta_1 - \beta_2 \text{, the parameter we are interested in testing} \\ &\Rightarrow \beta_1 = \theta + \beta_2 \\ \Rightarrow log(wage) &= \beta_0 + (\theta + \beta_2) school + \beta_2 university + u \\ &= \beta_0 + \theta school + \beta_2 (school + university) + u \end{aligned}$$

```
. gen school = (educ<=12)*educ
(290 missing values generated)

. gen university = (educ>12)*(educ-12)
(290 missing values generated)

. replace school=12 if university!=0
(4590 real changes made)
```

```
. reg lwage school university

      Source |       SS       df       MS              Number of obs =   23436
-------------+------------------------------           F(  2, 23433) = 4639.37
       Model |  8343.18513     2  4171.59256           Prob > F      =  0.0000
    Residual |  21070.3072 23433  .899172413           R-squared     =  0.2837
-------------+------------------------------           Adj R-squared =  0.2836
       Total |  29413.4923 23435  1.25510955           Root MSE      =  .94825

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1040755   .0017539    59.34   0.000     .1006377    .1075133
  university |   .4678748   .0082577    56.66   0.000     .4516892    .4840604
       _cons |   .9226882    .016209    56.92   0.000     .8909175     .954459
------------------------------------------------------------------------------
```

LFS

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together sonke siya phambili saam vorentoe

- Is the difference statistically significant?

```
. reg lwage school educ
```

Source	SS	df	MS
Model	8343.18513	2	4171.59256
Residual	21070.3072	23433	.899172413
Total	29413.4923	23435	1.25510955

Number of obs	=	23436
F(2, 23433)	=	4639.37
Prob > F	=	0.0000
R-squared	=	0.2837
Adj R-squared	=	0.2836
Root MSE	=	.94825

lwage1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
school	-.3637993	.0089007	-40.87	0.000	-.3812452	-.3463533
educ	.4678748	.0082577	56.66	0.000	.4516892	.4840604
_cons	.9226882	.016209	56.92	0.000	.8909175	.954459

OR… a lot easier

```
. reg lwage school university

      Source |       SS       df       MS              Number of obs =   23436
-------------+------------------------------           F(  2, 23433) = 4639.37
       Model |  8343.18513     2  4171.59256           Prob > F      =  0.0000
    Residual |  21070.3072 23433  .899172413           R-squared     =  0.2837
-------------+------------------------------           Adj R-squared =  0.2836
       Total |  29413.4923 23435  1.25510955           Root MSE      =  .94825

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1040755   .0017539    59.34   0.000     .1006377    .1075133
  university |   .4678748   .0082577    56.66   0.000     .4516892    .4840604
       _cons |   .9226882    .016209    56.92   0.000     .8909175     .954459
------------------------------------------------------------------------------

. test school=university

 ( 1)  school - university = 0

       F(  1, 23433) = 1670.61
            Prob > F =    0.0000
```

Testing Exclusion Restrictions

- Joint hypothesis: entire subsection of model is equal to zero

$$log(wage) = \beta_0 + \beta_1 school + \beta_2 university + \beta_3 experience + \beta_4 tenure + u$$

BEFORE JOB MARKET **AFTER JOB MARKET**

- Test whether learning acquired during the job market has no effect on wages
- **JOINT** hypothesis

$$H_0 : \beta_3 = \beta_4 = 0$$
$$H_a : H_0 \text{ not true}$$

- If one of these is different from zero, then the alternative will often hold

Exclusion Restrictions

UR: $log(wage) = \beta_0 + \beta_1 school + \beta_2 university + \beta_3 experience + \beta_4 tenure + u$

Impose hypothesised restriction by setting $\beta_3 = \beta_4 = 0$

R: $log(wage) = \beta_0 + \beta_1 school + \beta_2 university + u$

- $SSR \downarrow$ when $k \uparrow$ (or moving from restricted (R) to understricted (UR))
  - But is the *relative* drop in SSR “large enough” to suggest that the variables are relevant to the model?
- SSR of unrestricted model vs restricted model
  - More parameters vs fewer parameters
  - Lower SSR vs higher SSR
  - Higher R^2 vs lower R^2

Testing exclusion restrictions

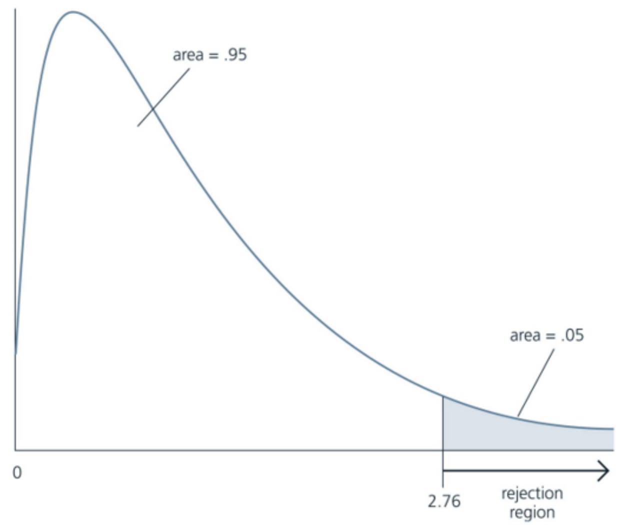
- F = the relative increase in SSR by excluding group of variables (scaled by d.f.)

$$F = \frac{\left(SSR_R - SSR_{UR}\right)/q}{SSR_{UR}/\left(n-k-1\right)} \sim F_{q;n-k-1}$$

  - F always > 0, therefore only one-sided test
  - $q = [n-(k+1-q)] - [n-k-1]$= number of restrictions = numerator d.f.
  - $n-k-1$ = denominator d.f.
- If F is "large": "substantial" enough increase in SSR to reject H_0
  - In other words F larger than a critical value from the F theoretical distribution

FIGURE 4.7

The 5% critical value and rejection region in an $F_{3,60}$ distribution.



```
. reg lwage school university exp tenure

      Source |       SS       df       MS              Number of obs =   19734
-------------+------------------------------           F(  4, 19729) = 3391.71
       Model |  8632.98807     4  2158.24702           Prob > F      =  0.0000
    Residual |  12554.1738 19729  .636330974           R-squared     =  0.4075
-------------+------------------------------           Adj R-squared =  0.4073
       Total |  21187.1619 19733  1.07369188           Root MSE      =   .7977

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1325051    .002059    64.36   0.000     .1284693    .1365408
  university |   .3974043    .007653    51.93   0.000     .3824039    .4124047
         exp |   .0120163   .0006381    18.83   0.000     .0107656     .013267
      tenure |   .0311069   .0008276    37.59   0.000     .0294847    .0327291
       _cons |   .2394071   .0278982     8.58   0.000     .1847243    .2940898
------------------------------------------------------------------------------

. reg lwage school university if e(sample)==1

      Source |       SS       df       MS              Number of obs =   19734
-------------+------------------------------           F(  2, 19731) = 4409.22
       Model |  6544.35606     2  3272.17803           Prob > F      =  0.0000
    Residual |  14642.8058 19731  .742121829           R-squared     =  0.3089
-------------+------------------------------           Adj R-squared =  0.3088
       Total |  21187.1619 19733  1.07369188           Root MSE      =  .86146

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .0991527   .0017681    56.08   0.000      .095687    .1026184
  university |   .4642348   .0081604    56.89   0.000     .4482398    .4802298
       _cons |   1.005904   .0164882    61.01   0.000     .9735857    1.038222
------------------------------------------------------------------------------
```

$$H_0 : \beta_{exp} = \beta_{tenure} = 0$$

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)} = \frac{(14642.8058 - 12554.1738)/2}{12554.1738/19729} = 1641.1522$$

$$F_{2;19729;0.99} = 4.61$$

$$F > F_{2;19729;0.99} \Rightarrow \text{reject } H_0 \text{ at a 1\% level of significance}$$

Relationship between F and t stats

$$t^2_{n-k-1} = F_{1;n-k-1}$$

- Using F-stat approach to exclude one variable at a time gives same conclusions as individual T-tests
- But T can test two-sided hypotheses
- F test is good at testing jointly
  - If joint test shows that whole group is insignificant, it may still be true that one variable does have some important explanatory power

R^2 form of F-stat

$$F = \frac{\left(R^2_{UR} - R^2_R\right)/q}{\left(1 - R^2_{UR}\right)/(n-k-1)}$$

- Note the order of the R^2's, to ensure that $F > 0$
- Interpretation: has relative R^2 increased significantly after adding variables to unrestricted model?

```
. reg lwage school university exp tenure

      Source |       SS           df       MS      Number of obs   =    19734
-------------+----------------------------------   F(  4, 19729)   =  3391.71
       Model |  8632.98807         4  2158.24702   Prob > F        =   0.0000
    Residual |  12554.1738    19729  .636330974   R-squared       =   0.4075
-------------+----------------------------------   Adj R-squared   =   0.4073
       Total |  21187.1619    19733  1.07369188   Root MSE        =    .7977

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1325051    .002059    64.36   0.000     .1284693    .1365408
  university |   .3974043    .007653    51.93   0.000     .3824039    .4124047
         exp |   .0120163   .0006381    18.83   0.000     .0107656     .013267
      tenure |   .0311069   .0008276    37.59   0.000     .0294847    .0327291
       _cons |   .2394071   .0278982     8.58   0.000     .1847243    .2940898
------------------------------------------------------------------------------

. reg lwage school university if e(sample)==1

      Source |       SS           df       MS      Number of obs   =    19734
-------------+----------------------------------   F(  2, 19731)   =  4409.22
       Model |  6544.35606         2  3272.17803   Prob > F        =   0.0000
    Residual |  14642.8058    19731  .742121829   R-squared       =   0.3089
-------------+----------------------------------   Adj R-squared   =   0.3088
       Total |  21187.1619    19733  1.07369188   Root MSE        =   .86146

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .0991527   .0017681    56.08   0.000      .095687    .1026184
  university |   .4642348   .0081604    56.89   0.000     .4482398    .4802298
       _cons |   1.005904   .0164882    61.01   0.000     .9735857    1.038222
------------------------------------------------------------------------------
```

$$H_0 : \beta_{exp} = \beta_{tenure} = 0$$

$$F = \frac{\left(R^2_{UR} - R^2_R\right)/q}{\left(1 - R^2_{UR}\right)/\left(n-k-1\right)} = \frac{\left(0.4075 - 0.3089\right)/2}{\left(1 - 0.4075\right)/19729} = 1641.15$$

$$F_{2;19729;0.99} = 4.61$$

$$F > F_{2;19729;0.99} \Rightarrow \text{reject } H_0 \text{ at a 1\% level of significance}$$

Or more simply in STATA


```
. reg lwage school university exp tenure

      Source |       SS       df       MS              Number of obs =   19734
-------------+------------------------------           F(  4, 19729) = 3391.71
       Model |  8632.98807     4  2158.24702           Prob > F      =  0.0000
    Residual |  12554.1738 19729  .636330974           R-squared     =  0.4075
-------------+------------------------------           Adj R-squared =  0.4073
       Total |  21187.1619 19733  1.07369188           Root MSE      =   .7977

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1325051    .002059    64.36   0.000     .1284693    .1365408
  university |   .3974043    .007653    51.93   0.000     .3824039    .4124047
         exp |   .0120163   .0006381    18.83   0.000     .0107656     .013267
      tenure |   .0311069   .0008276    37.59   0.000     .0294847    .0327291
       _cons |   .2394071   .0278982     8.58   0.000     .1847243    .2940898
------------------------------------------------------------------------------

. test exp=tenure=0

 ( 1)  exp - tenure = 0
 ( 2)  exp = 0

       F(  2, 19729) = 1641.15
            Prob > F =    0.0000
```

Computing p-values for F tests

- p-value = $P(F > F_{critical})$
- Better to use, because size of F-stat is dependent on d.f.
- Still same interpretation as other p-values
  - Small p-value is evidence against H_0

F statistic for overall significance of regression

- Are all the regressors jointly (in)significant?
  - Routinely reported by regression packages

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

  - Same approach as before, with k restrictions
  - The only difference is that the restricted model just has an intercept

Testing General Linear Restrictions

- Test whether coefficients are specific **non-zero** values
  - Rewrite models to do this
  - Cannot always use the R^2 version of the test statistic
    - Particularly if we have a different dependent variable after rewriting the model

$$H_0 : \beta_{experience} = 0.01 \text{ and } \beta_{university} = 3 \times \beta_{school}$$

Unrestricted (UR):

$$log(wage) = \beta_0 + \beta_1 school + \beta_2 university + \beta_3 experience + \beta_4 tenure + u$$

Restricted (R):

$$log(wage) = \beta_0 + \beta_1 school + 3\beta_1 university + 0.01 \times experience + \beta_4 tenure + u$$
$$\Rightarrow log(wage) - 0.01 \times exper = \beta_0 + \beta_1 (school + 3 \times university) + \beta_4 tenure + u$$

```
. reg lwage school university exp tenure

      Source |       SS       df       MS              Number of obs =   19734
-------------+------------------------------           F(  4, 19729) = 3391.71
       Model |  8632.98807     4  2158.24702           Prob > F      =  0.0000
    Residual |  12554.1738 19729  .636330974           R-squared     =  0.4075
-------------+------------------------------           Adj R-squared =  0.4073
       Total |  21187.1619 19733  1.07369188           Root MSE      =   .7977

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1325051    .002059    64.36   0.000     .1284693    .1365408
  university |   .3974043    .007653    51.93   0.000     .3824039    .4124047
         exp |   .0120163   .0006381    18.83   0.000     .0107656     .013267
      tenure |   .0311069   .0008276    37.59   0.000     .0294847    .0327291
       _cons |   .2394071   .0278982     8.58   0.000     .1847243    .2940898
------------------------------------------------------------------------------
. gen l_newwage = lwage - 0.01*exp
(42455 missing values generated)

. gen newschool = school+3*university
(290 missing values generated)

. reg l_newwage newschool tenure if e(sample)==1

      Source |       SS       df       MS              Number of obs =   19734
-------------+------------------------------           F(  2, 19731) = 7184.58
       Model |   9147.8608     2   4573.9304           Prob > F      =  0.0000
    Residual |  12561.3759 19731   .63663149           R-squared     =  0.4214
-------------+------------------------------           Adj R-squared =  0.4213
       Total |  21709.2367 19733  1.10014882           Root MSE      =  .79789

------------------------------------------------------------------------------
   l_newwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   newschool |   .1299335   .0011791   110.20   0.000     .1276223    .1322446
      tenure |   .0325503   .0007075    46.01   0.000     .0311637     .033937
       _cons |   .2995577    .013601    22.02   0.000     .2728985    .3262168
------------------------------------------------------------------------------
```

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)} = \frac{(12561.3759 - 12554.1738)/2}{12554.1738/19729} = 5.66$$

$$> F_{2;19729;0.99} = 4.61 \Rightarrow \text{Reject } H_0$$

General Restrictions in STATA...

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT — forward together sonke siya phambili saam vorentoe

```
. reg lwage school university exp tenure

      Source |       SS       df       MS              Number of obs =   19734
-------------+------------------------------           F(  4, 19729) = 3391.71
       Model |  8632.98807     4  2158.24702           Prob > F      =  0.0000
    Residual |  12554.1738 19729  .636330974           R-squared     =  0.4075
-------------+------------------------------           Adj R-squared =  0.4073
       Total |  21187.1619 19733  1.07369188           Root MSE      =   .7977

------------------------------------------------------------------------------
      lwage1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      school |   .1325051    .002059    64.36   0.000     .1284693    .1365408
  university |   .3974043    .007653    51.93   0.000     .3824039    .4124047
         exp |   .0120163   .0006381    18.83   0.000     .0107656     .013267
      tenure |   .0311069   .0008276    37.59   0.000     .0294847    .0327291
       _cons |   .2394071   .0278982     8.58   0.000     .1847243    .2940898
------------------------------------------------------------------------------

. test (exp=0.01) (university=3*school)

 ( 1)  exp = .01
 ( 2)  - 3*school + university = 0

       F(  2, 19729) =    5.66
            Prob > F =    0.0035
```

What to report?

- Coefficients
  - With economic interpretation
    - Elasticity, semi-elasticity, unit changes
  - Add significance stars for easy reading: $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$
- Standard errors (usually in parentheses below coefficients)
  - Help us compute confidence intervals, T-stats and p-values
- R^2 and n
- In Stata
  - `ssc install esttab`
  - `help esttab`