

Testing Multiple Hypotheses

(I) Testing exclusion restrictions

(II) Overall F Test

III General linear Hypotheses Restrictions.

MLB

$$\log(\text{salary}) = \beta_1 + \beta_2 \text{years} + \beta_3 \text{gomeyr} + \beta_4 \text{bAvg}$$

$$+ \beta_5 \text{HRUNSYR} + \beta_6 \text{rbisyr} + \text{ee.}$$

1993 salary

years # years in league

gomeyr average gomee per year.

Performance

bAvg	career BA
HR	HR/sea
RBI	

H_0 : performance has no effect

$$\beta_4 = \beta_5 = \beta_6 = 0$$

H_A : H_0 not true

Individual signif of Variables
is not the same as joint signif.

$$\begin{aligned} \log(\text{Salary}) = & 11.10 + .0689 \text{ years} + .0126 \text{ gamesper} \\ & (-.0121) \quad (.0026) \\ & + .00098 \text{ BAvg} + .0144 \text{ HRunyr} + .0108 \text{ RBExper} \\ & (-.00110) \quad (-.0161) \quad (-.0108) \\ n = 353 \quad R^2 = .6278 \quad SSR = 183.186 \end{aligned}$$

none of the ~~variables~~ Performance vars
one individually significant

$$H_0: \beta_4 = \beta_5 = \beta_6 = 0 \quad (3 \text{ restrictions or 14 hypotheses})$$

$$H_A: \beta_4 \neq 0 \text{ and/or } \beta_5 \neq 0 \text{ and/or } \beta_6 \neq 0$$

Homoscedasticity only:

7

Test stat is based on the
difference in fit between the
2 models

(1) I with no restrictions

(2) I with all " imposed

Restricted Model

$$\ln(\text{salary}) = \beta_1 + \beta_2 \text{years} + \beta_3 \text{gmcyear.} + u$$

note this model only has
3 persons whereas the other had 6.
 \Rightarrow 3 rest. imposed.

$$\ln(\text{salary}) = 11.22 + .0713 \text{ year} + .0102 \text{ gmcyear}$$
$$(.0125) \qquad \qquad (.0013)$$

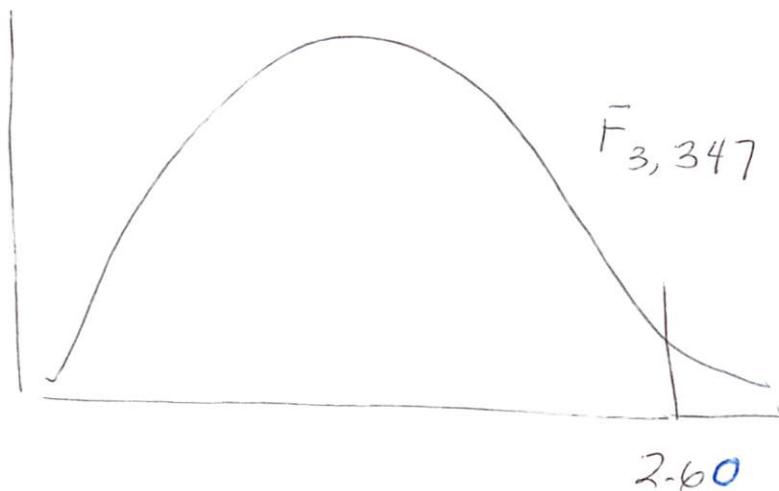
$$n = 353, R^2 = .5971 \qquad SSR = 198.311$$

fech stat

$$\frac{(SSR_R - SSR_u) / \# \text{ rech.}}{SSR_u / (n - k)}$$

$n = \# \text{ parons im unref model.}$

$$F = \frac{(198.311 - 183.186) / 3}{(183.186) / (353 - 6)} = 9.55$$



$9.95 > 2.60$ Rejek.
Performance Matthes.

SER - *

For Hetero, use Robust option
and test statements in program

get1 -

OLS y const x2 x3 x4 x5, robust
omit x3 x4.

~~Stata~~

restrict

$$b_2 = 0$$

$$b_3 + b_4 = 1$$

end restrict

Stata

regress y x2 x3 x4 x5, vce(robust) vce(robust)

Example test statements:

testparm x2 x3

test x2+x3=1

test (x2=0)(x3+x4=1)

testparm x2 x3

test x2 + x3 = 1

test (x2=0)(x3+x4=1)

Overall F-Test

$$(1) \quad Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{in} + u_i$$

If none of the x_i 's have any effect on y , then we could say that there is no model.

$$(2) \quad Y_i = \beta_1 + u_i$$

The overall F-test tests the no model null against the full model.

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

$$H_A: \text{At least } 1 \beta_i \text{ not equal zero.}$$

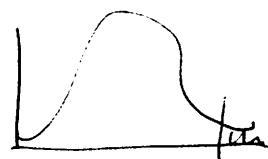
$$i = 2, \dots, k.$$

Homoskedasticity only statistic:

$$F = \frac{\frac{(SSR_n - SSR_u)}{k-1}}{SSR_u / n-k} \sim F_{k-1, n-k} \text{ if } H_0 \text{ true}$$

Reject if $F \geq c_\alpha$.

or if P-value < desired α .



General linear restrictions

$$(1) \quad y_i = \beta_1 + \beta_2 \gamma_{i2} + \dots + \beta_5 \gamma_{i5} + u_i$$

Suppose:

$$H_0: \beta_2 + \beta_3 = 1 \text{ AND } \beta_4 = 3\beta_5$$

$$H_A: \text{not } H_0.$$

NOTE: 2 hypotheses.

Again, compare SSR from a restricted and unrest. model.

Restricted:

$$\beta_2 = .1 - \beta_3$$

$$\beta_4 = 3\beta_5$$

Sub into model & solve.

$$y_i = \beta_1 + (.1 - \beta_3)\gamma_{i2} + \beta_3 \gamma_{i3} + 3\beta_5 (\gamma_{i4})$$

$$+ \beta_5 \gamma_{i5} + u_i$$

$$= \beta_1 + .1 \gamma_{i2} + \beta_3 (\gamma_{i3} - \gamma_{i2}) + \beta_5 (3\gamma_{i4} + \gamma_{i5}) + u_i$$

$$y_i - \gamma_{i2}^{(1)} = \beta_1 + \beta_3 (\gamma_{i3} - \gamma_{i2}) + \beta_5 (3\gamma_{i4} + \gamma_{i5}) + u_i$$

Create these new variables

$$y_i - \gamma_{i2} = y_i^*$$

$$\gamma_{i3} - \gamma_{i2}/10 = \gamma_{i2}^*$$

$$3\gamma_{i4} + \gamma_{i5} = \gamma_{i3}^*$$

$$y_i^* = \beta_1 + \beta_3 \gamma_{i2}^* + \beta_5 \gamma_{i3}^* + u$$

and eliminate $\beta_1, \beta_3, \beta_5$ using LS.

(Stat is Homoskedastic only--otherwise use vce(robust) option in Stata)

$$F = \frac{(SSR_k - SSR_u)/g}{SSR_u/(n-k)} \sim F_{g, n-k} \text{ if null hypotheses are true.}$$

Reject if $F \geq c_\alpha$

```

. open "C:\Program Files (x86)\gretl\data\wooldridge\mlb1.gdt"
. # model 1
. ols lsalary const years gamesyr bavg hrunsyr rbisyrs
. omit bavg hrunsyr rbisyrs
-
. # Overall-F for regression significance
. ols lsalary const years gamesyr bavg hrunsyr rbisyrs
. omit years gamesyr bavg hrunsyr rbisyrs --test-only
.
10 # model 2
. ols lsalary const years gamesyr bavg hrunsyr rbisyrs
. restrict
.   2*b[years] - b[gamesyr] = 0
.   b[bavg]+b[hrunsyr]+b[rbisyrs] = 0.05
-   end restrict
.
. ****
. gretl version 1.9.9cvs
. Current session: 2012-10-22 13:08
20 ? open "C:\Program Files (x86)\gretl\data\wooldridge\mlb1.gdt"
.
. Read datafile C:\Program Files (x86)\gretl\data\wooldridge\mlb1.gdt
. periodicity: 1, maxobs: 353
. observations range: 1-353
-
. Listing 48 variables:
. 0) const      1) salary      2) teamsal      3) nl      4) years
. 5) games      6) atbats      7) runs      8) hits      9) doubles
. 10) triples     11) hruns      12) rbis      13) bavg      14) bb
30 15) so          16) sbases      17) fldperc      18) frstbase      19) scndbase
. 20) shrtstop     21) thrdbase     22) outfield      23) catcher      24) yrsallst
. 25) hispan      26) black      27) whitepop      28) blackpop      29) hisppop
. 30) pcinc      31) gamesyr      32) hrunsyr      33) atbatsyr      34) allstar
. 35) slugavg     36) rbisyrs      37) sbasesyr      38) runsyr      39) percwhte
- 40) percblck    41) perchisp     42) blckpb      43) hispph      44) whtepw
. 45) blckph      46) hisppb      47) lsalary
.
. # model 1
. ? ols lsalary const years gamesyr bavg hrunsyr rbisyrs
40
. Model 1: OLS, using observations 1-353
. Dependent variable: lsalary
.
.      coefficient      std. error      t-ratio      p-value
. -----
. const      11.1924      0.288823      38.75      4.19e-128 *** 
. years      0.0688626      0.0121145      5.684      2.79e-08 *** 
. gamesyr     0.0125521      0.00264676      4.742      3.09e-06 *** 
. bavg      0.000978594      0.00110351      0.8868      0.3758
50 hrunsyr     0.0144295      0.0160570      0.8986      0.3695
. rbisyrs     0.0107657      0.00717496      1.500      0.1344
.
. Mean dependent var      13.49218      S.D. dependent var      1.182466
. Sum squared resid      183.1863      S.E. of regression      0.726577
- R-squared      0.627803      Adjusted R-squared      0.622440
. F(5, 347)      117.0603      P-value(F)      2.94e-72
. Log-likelihood      -385.1076      Akaike criterion      782.2152
. Schwarz criterion      805.4141      Hannan-Quinn      791.4463
.
60 Log-likelihood for salary = ^'5147.85
.
. Excluding the constant, p-value was highest for variable 13 (bavg)
.
. ? omit bavg hrunsyr rbisyrs
- Test on Model 1:

```

```

.
.
.
Null hypothesis: the regression parameters are zero for the variables
    bavg, hrunsyr, rbisyrr
Test statistic: F(3, 347) = 9.55027, p-value 4.47358e-006
70 Omitting variables improved 0 of 3 model selection statistics.

.
.
.
Model 2: OLS, using observations 1-353
Dependent variable: lsalary

.
.
.
      coefficient   std. error   t-ratio   p-value
-----
const      11.2238     0.108312   103.6   9.97e-265 *** 
years      0.0713180   0.0125050   5.703   2.50e-08  *** 
gamesyr    0.0201745   0.00134287  15.02    1.02e-039 *** 

80 Mean dependent var  13.49218   S.D. dependent var  1.182466
Sum squared resid  198.3115    S.E. of regression  0.752731
R-squared          0.597072    Adjusted R-squared  0.594769
F(2, 350)         259.3203    P-value(F)        8.22e-70
Log-likelihood     -399.1103   Akaike criterion   804.2206
Schwarz criterion  815.8200    Hannan-Quinn     808.8361

.
.
.
Log-likelihood for salary = -5161.85

.
.
.
90 ? ols lsalary const years gamesyr bavg hrunsyr rbisyrr

.
.
.
Model 3: OLS, using observations 1-353
Dependent variable: lsalary

.
.
.
      coefficient   std. error   t-ratio   p-value
-----
const      11.1924     0.288823   38.75    4.19e-128 *** 
years      0.0688626   0.0121145   5.684   2.79e-08  *** 
gamesyr    0.0125521   0.00264676  4.742   3.09e-06  *** 
bavg       0.000978594  0.00110351  0.8868  0.3758
hrunsyr    0.0144295   0.0160570   0.8986  0.3695
rbisyrr    0.0107657   0.00717496  1.500   0.1344

.
.
.
100 Mean dependent var  13.49218   S.D. dependent var  1.182466
Sum squared resid  183.1863    S.E. of regression  0.726577
R-squared          0.627803    Adjusted R-squared  0.622440
F(5, 347)         117.0603    P-value(F)        2.94e-72
Log-likelihood     385.1076    Akaike criterion   782.2152
Schwarz criterion  805.4141    Hannan-Quinn     791.4463

.
.
.
110 Log-likelihood for salary = -5147.85

.
.
.
Excluding the constant, p-value was highest for variable 13 (bavg)

.
.
.
? omit years gamesyr bavg hrunsyr rbisyrr --test-only
Test on Model 3:

.
.
.
Null hypothesis: the regression parameters are zero for the variables
    years, gamesyr, bavg, hrunsyr, rbisyrr
120 Test statistic: F(5, 347) = 117.06, p-value 2.93802e-072

.
.
.
# model 2
? ols lsalary const years gamesyr bavg hrunsyr rbisyrr

.
.
.
Model 4: OLS, using observations 1-353
Dependent variable: lsalary

.
.
.
      coefficient   std. error   t-ratio   p-value
-----
const      11.1924     0.288823   38.75    4.19e-128 *** 

```


R²

fraction of sample variance
explained by the regressors.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSE}{TSS}$$

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2$$

$$SSE = \sum \hat{e}_i^2$$

$$TSS = \sum (Y_i - \bar{Y})^2$$

R^2 — increases as regressors are added.

5.1 - 5.5

$$\text{Adjusted } R^2 = \bar{R}^2$$

$$1 - \frac{n-1}{n-k} \frac{SSR}{TSS}$$

$k = \# \text{ of parameters in linear model.}$

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_n + u.$$

$$\textcircled{1} \quad \frac{n-1}{n-k} > 1 \Rightarrow R^2 > \bar{R}^2$$

\textcircled{2} Adding a regressor has 2 effects

(a) it reduces SSR - better fit

(b) it increases $\frac{n-1}{n-k-1}$ - penalty for more regressors.

The net effect determine

whether $\bar{R}^2 \uparrow$ or \downarrow as regressors added.

\textcircled{3} \bar{R}^2 can be < 0 !

$R^2 -$

- ① Just Because \hat{R}^2 \neq p
does not mean that the added variable is not stat. signif.
- ② Just Because R^2 high
it does not mean that the regressors are the true
cause of y.
- ③ High R^2 $\not\Rightarrow$ no omitted Variable Bias
- ④ High R^2 $\not\Rightarrow$ correct regression
Low R^2 $\not\Rightarrow$ BAD Model.

Recap

Omitted Variable Bias occurs

When you omit a variable from a model that affects γ and is also correlated with any of the other variables included in the model.

Base Regression

- Expert knowledge
- Econ Theory
- knowledge of data and how collected

lead you to a Base spec.

Included variables of interest and any control vars suggested by above.

Then consider alternative spec.

- other theories
- other experts

Compare results - If large t 's occur

That could be evidence of
omitted Var. BiAs in smaller
models.