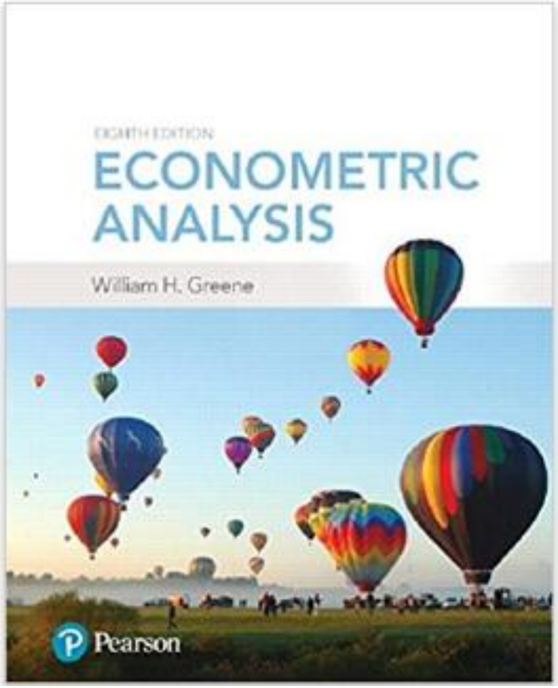
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 5 – Regression Algebra and Fit; Restricted Least Squares

Minimizing $\mathbf{e}'\mathbf{e}$

$\mathbf{b}$ minimizes $\mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$.

Any other coefficient vector has a larger sum of squared residuals.

Proof:

$\mathbf{d}$ = the vector, not equal to $\mathbf{b}$; $\mathbf{u} = \mathbf{Xd}$

$$\mathbf{u} = \mathbf{y} - \mathbf{Xd} = \mathbf{y} - \mathbf{Xb} + \mathbf{Xb} - \mathbf{Xd}$$
$$= \mathbf{e} - \mathbf{X}(\mathbf{d} - \mathbf{b}).$$

Then, $\mathbf{u}'\mathbf{u} = (\mathbf{y} - \mathbf{Xd})'(\mathbf{y}-\mathbf{Xd})$ = sum of squares using $\mathbf{d}$

$$= [(\mathbf{y} - \mathbf{Xb}) - \mathbf{X}(\mathbf{d} - \mathbf{b})]'[(\mathbf{y} - \mathbf{Xb}) - \mathbf{X}(\mathbf{d} - \mathbf{b})]$$
$$= [\mathbf{e} - \mathbf{X}(\mathbf{d} - \mathbf{b})]'[\mathbf{e} - \mathbf{X}(\mathbf{d} - \mathbf{b})]$$

Expand to find $\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} + (\mathbf{d}-\mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{d}-\mathbf{b})$

(The cross product term is $2\mathbf{e}'\mathbf{X}(\mathbf{d}-\mathbf{b})=\mathbf{0}$ as $\mathbf{X}'\mathbf{e} = \mathbf{0}$.)

$$= \mathbf{e}'\mathbf{e} + \mathbf{v}'\mathbf{v} \geq \mathbf{e}'\mathbf{e}$$

Dropping a Variable

An important special case. Suppose

$\mathbf{b}_{X,z}$ = [**b**,c]

= the regression coefficients in a regression of **y** on [**X**,**z**]

$\mathbf{b}_{X}$ = [**d**,0]

= is the same, but computed to force the coefficient on **z** to equal 0. This removes **z** from the regression.

We are comparing the results that we get with and without the variable **z** in the equation. Results which we can show:

- **Dropping a variable(s)** cannot improve the fit - that is, it cannot reduce the sum of squared residuals.

- **Adding a variable(s)** cannot degrade the fit - that is, it cannot increase the sum of squared residuals.

Adding a Variable Never Increases the Sum of Squares

Theorem 3.5 on text page 40.

u = the residual in the regression of **y** on [**X**,**z**]

e = the residual in the regression of **y** on **X** alone,

$$\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} - c^2(\mathbf{z}^{*\prime}\mathbf{z}^*) \leq \mathbf{e}'\mathbf{e}$$

where $\mathbf{z}^* = \mathbf{M}_X\mathbf{z}$ and c is the coefficient on **z** in the regression of **y** on [**X**,**z**].

The Fit of the Regression

- "**Variation**:" In the context of the "model" we speak of *covariation* of a variable as movement of the variable, usually associated with (not necessarily caused by) movement of another variable.
- **Total variation** = $\sum_{i=1}^{n}(y_i - \bar{y})^2$ = $\mathbf{y'M^0y}$.
- $\mathbf{M^0 = I - i(i'i)^{-1}i'}$
  = the **M** matrix for
  X = a column of ones.

Decomposing the Variation

$$y_i = \mathbf{x}_i'\mathbf{b} + e_i$$

$$y_i - \bar{y} = \mathbf{x}_i'\mathbf{b} - \bar{\mathbf{x}}'\mathbf{b} + e_i$$

$$= (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b} + e_i$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} [(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}]^2 + \sum_{i=1}^{n} e_i^2$$

(Sum of cross products is zero.)

Total variation = regression variation + residual variation

Recall the decomposition:

Var[y] = **Var [E[y|x]]** + **E[Var [y | x]]**

= Variation of the conditional mean around the overall mean

+ Variation around the conditional mean function.

Decomposing the Variation of Vector **y**

Decomposition: (This all assumes the model contains a constant term. one of the columns in **X** is **i**.)

$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ so

$\mathbf{M}^0\mathbf{y} = \mathbf{M}^0\mathbf{Xb} + \mathbf{M}^0\mathbf{e} = \mathbf{M}^0\mathbf{Xb} + \mathbf{e}$.

(Deviations from means.)

$\mathbf{y}'\mathbf{M}^0\mathbf{y} = \mathbf{b}'(\mathbf{X}'\mathbf{M}^0)(\mathbf{M}^0\mathbf{X})\mathbf{b} + \mathbf{e}'\mathbf{e}$

$= \mathbf{b}'\mathbf{X}'\mathbf{M}^0\mathbf{Xb} + \mathbf{e}'\mathbf{e}.$

($\mathbf{M}^0$ is idempotent and $\mathbf{e}'\mathbf{M}^0\mathbf{X} = \mathbf{e}'\mathbf{X} = \mathbf{0}$.)

Total sum of squares = **Regression Sum of Squares (SSR)+**

Residual Sum of Squares (SSE)

A Fit Measure

$R^2 = \mathbf{b'X'M^0Xb/y'M^0y}$

$$= 1 - \frac{\mathbf{e'e}}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\text{Regression Variation}}{\text{Total Variation}}$$

(Very Important Result.) R^2 ***is bounded by zero and one only if:***

(a) There is a constant term in X and

(b) The line is computed by linear least squares.

Adding Variables

- R^2 never falls when a variable **z** is added to the regression.
- **A useful general result for adding a variable**

$R^2_{\mathbf{Xz}}$ with both **X** and variable **z** equals

$R^2_{\mathbf{X}}$ with only **X** plus the increase in fit due to **z**

after **X** is accounted for:

$$R^2_{\mathbf{Xz}} = R^2_{\mathbf{X}} + (1 - R^2_{\mathbf{X}}) r^{*2}_{\mathbf{yz|X}}$$

- **Useful practical wisdom**: It is not possible meaningfully to accumulate R^2 by adding variables in sequence. The incremental fit added by each variable depends on the order. The increase in R^2 that occurs from x_3 in (x_1 then x_2 then x_3) is different from that in (x_1 then x_3 then x_2).

Adding Variables to a Model

What is the effect of adding PN, PD, PS?

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=G       Mean                  =     226.09444
            Standard deviation    =      50.59182
----------  No. of observations   =            36  DegFreedom   Mean square
Regression  Sum of Squares        =       88338.9           3   29446.30141
Residual    Sum of Squares        =       1244.71          32      38.89734
Total       Sum of Squares        =       89583.6          35    2559.53197
----------  Standard error of e   =       6.23677  Root MSE         5.88009
Fit         R-squared             =        .98611  R-bar squared     .98480
Fit with PN, PD, PS added to the regression
Fit         R-squared             =        .99259  R-bar squared     .99106
Model test  F[  3,    32]         =     757.02619  Prob F > F*       .00000
Model was estimated on Jul 21, 2012 at 08:40:19 AM
Effects of additional variables on the regression below: -------------
Variable Coefficient   New R-sqrd  Chg.R-sqrd   Partial-Rsq   Partial F
PN          -24.7246      .9882       .0021        .1515        5.534
PD          -19.8131      .9868       .0006        .0465        1.511
PS          -18.3479      .9907       .0046        .3310       15.340
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
       G|  Coefficient       Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   4154.60**      1748.656     2.38  .0237     727.29   7581.90
    YEAR|  -2.19582**        .90680    -2.42  .0213   -3.97311   -.41853
      PG|  -11.9827***      2.18235    -5.49  .0000   -16.2600   -7.7053
       Y|    .04782***       .00467    10.25  .0000     .03867    .05696
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

A Useful Result

Squared partial correlation of an x in **X** with *y* is

$$\frac{\text{squared t - ratio}}{\text{squared t - ratio} + \text{degrees of freedom}}$$

We will define the 't-ratio' and 'degrees of freedom' later. Note how it enters:

$$R^2_{Xz} = R^2_X + (1 - R^2_X) r^{*2}_{yz} \Rightarrow r^{2*}_{yz} = \left(R^2_{Xz} - R^2_X\right) / \left(1 - R^2_X\right)$$

Partial Correlation

Partial correlation is a difference in R^2s.

For PS in the example above,

G	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-8520.85**	3547.533	-2.40	.0225	-15473.89	-1567.81
YEAR	4.34824**	1.83288	2.37	.0241	.75586	7.94062
PG	-14.0226***	1.88682	-7.43	.0000	-17.7207	-10.3245
Y	.02429***	.00715	3.40	.0019	.01028	.03830
PS	-18.3478***	4.68454	-3.92	.0005	-27.5294	-9.1663

R^2 without PS = .9861, R^2 with PS = .9907

(.9907 - .9861) / (1 - .9861) = .331

$3.92^2 / (3.92^2 + (36-5))$ = .331

Comparing fits of regressions

Make sure the denominator in R^2 is the same - i.e., same left hand side variable. Example, linear vs. loglinear. Loglinear will almost always appear to fit better because taking logs reduces variation.

```
|-> regr;lhs=g;rhs=x$
--------------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=G       Mean                 =      226.09444
            Standard deviation   =       50.59182
Fit         R-squared            =         .98611  R-bar squared      .98480
Model test  F[  3,    32]        =      757.02619  Prob F > F*        .00000
--------+-----------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
       G|  Coefficient      Error       t    |t|>T*          Interval
--------+-----------------------------------------------------------------------
Constant|    4154.60**    1748.656     2.38  .0237      727.29   7581.90
    YEAR|   -2.19582**      .90680    -2.42  .0213    -3.97311   -.41853
      PG|   -11.9827***    2.18235    -5.49  .0000    -16.2600   -7.7053
       Y|     .04782***     .00467    10.25  .0000      .03867    .05696
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
|-> regr;lhs=log(g);rhs=one,year,log(pg),log(y)$
--------------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=logG    Mean                 =        5.39299
            Standard deviation   =         .24878
Fit         R-squared            =         .98754  R-bar squared      .98637
Model test  F[  3,    32]        =      845.23602  Prob F > F*        .00000
--------+-----------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
    logG|  Coefficient      Error       t    |t|>T*          Interval
--------+-----------------------------------------------------------------------
Constant|    6.60154       4.18578     1.58  .1246    -1.60244  14.80551
    YEAR|   -.00967***      .00255    -3.78  .0006     -.01467   -.00466
   logPG|   -.11268***      .02228    -5.06  .0000     -.15634   -.06902
    logY|   1.97383***      .11248    17.55  .0000     1.75337   2.19428
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

(Linearly) Transformed Data

- How does linear transformation affect the results of least squares? $\mathbf{Z} = \mathbf{XP}$ for K×K nonsingular $\mathbf{P}$
- Based on $\mathbf{X}$, $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$.
- Based on $\mathbf{Z}$, $\mathbf{c} = (\mathbf{Z'Z})^{-1}\mathbf{Z'y} = (\mathbf{P'X'XP})^{-1}\mathbf{P'X'y}$
  $$= \mathbf{P}^{-1}(\mathbf{X'X})^{-1}\mathbf{P'}^{-1}\mathbf{P'X'y} = \mathbf{P}^{-1}\mathbf{b}$$
  - "Fitted value" is $\mathbf{Zc} = (\mathbf{XP})(\mathbf{P}^{-1}\mathbf{b}) = \mathbf{Xb}$. The same!!
- Residuals from using $\mathbf{Z}$ are $\mathbf{y} - \mathbf{Zc} = \mathbf{y} - \mathbf{Xb}$ (we just proved this.). The same!!
  - Sum of squared residuals must be identical, as $\mathbf{y}-\mathbf{Xb} = \mathbf{e} = \mathbf{y}-\mathbf{Zc}$.
  - R^2 must also be identical, as $R^2 = 1 - \mathbf{e'e}/\mathbf{y'M}^0\mathbf{y}$ (!!).

Linear Transformation

What are the practical implications of this result?

(1) Transformation *does not* affect the fit of a model to a body of data.

(2) Transformation *does* affect the "estimates." If **b** is an estimate of something (β), then **c** cannot be an estimate of β - it must be an estimate of $\mathbf{P}^{-1}\beta$, which might have no meaning at all.

Xb is the projection of **y** into the column space of **X**. **Zc** is the projection of **y** into the column space of **Z**. But, since the columns of **Z** are just linear combinations of those of **X**, the column space of **Z** must be identical to that of **X**. Therefore, the projection of **y** into the former must be the same as the latter, which now produces the other results.)

Principal Components

- **Z** = **XC**
  - Fewer columns than **X**
  - Includes as much ‘variation’ of **X** as possible
  - Columns of **Z** are orthogonal
- Why do we do this?
  - Collinearity
  - Combine variables of ambiguous identity such as test scores as measures of ‘ability’

What is a Principal Component?

- **X** = a data matrix (deviations from means)
- **z** = **Xp** = a linear combination of the columns of **X**.
- Choose **p** to maximize the variation of **z**.

How? **p** = *eigenvector* that corresponds to the largest *eigenvalue* of **X'X**. (Notes 7:41-44.)

MediaPost News Events Awards Members More

MarketingDaily

Movies' Online Buzz Indicative Of Box Office Success

by Aaron Baar, February 19, 2015, 3:47 PM

Comment Recommend (1)

18 SHARES

As much as anything else, a big film's online buzz has a direct correlation to how successful that film will be on its opening weekend.

According to a new study from two NYU Stern School of Business professors and a business co-author, the amount of chatter a film is getting three weeks prior to its release on film-specific websites (such as Fandango and TrailerAddict), is a decent indicator of the film's opening weekend grosses. Conversely, the study found that star power and MPAA rating had less to do with the opening weekend revenues.

Movie Regression. Opening Week Box for 62 Films			
Ordinary	least squares regression		
LHS=LOGBOX	Mean	=	16.47993
	Standard deviation	=	.9429722
	Number of observs.	=	62
Residuals	Sum of squares	=	20.54972
	Standard error of e	=	.6475971
Fit	R-squared	=	.6211405
	Adjusted R-squared	=	.5283586

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	12.5388***	.98766	12.695	.0000	
LOGBUDGT	.23193	.18346	1.264	.2122	3.71468
STARPOWR	.00175	.01303	.135	.8935	18.0316
SEQUEL	.43480	.29668	1.466	.1492	.14516
MPRATING	-.26265*	.14179	-1.852	.0700	2.96774
ACTION	-.83091***	.29297	-2.836	.0066	.22581
COMEDY	-.03344	.23626	-.142	.8880	.32258
ANIMATED	-.82655**	.38407	-2.152	.0363	.09677
HORROR	.33094	.36318	.911	.3666	.09677
4 INTERNET BUZZ VARIABLES					
LOGADCT	.29451**	.13146	2.240	.0296	8.16947
LOGCMSON	.05950	.12633	.471	.6397	3.60648
LOGFNDGO	.02322	.11460	.203	.8403	5.95764
CNTWAIT3	2.59489***	.90981	2.852	.0063	.48242

The fit goes down when the 4 buzz variables are reduced to a single linear combination of the 4.

```
+----------------------------------------------------+
| Ordinary least squares regression                  |
| LHS=LOGBOX   Mean                 =   16.47993     |
|              Standard deviation   =   .9429722     |
|              Number of observs.   =         62     |
| Residuals    Sum of squares       =   25.36721     |
|              Standard error of e  =   .6984489     |
| Fit          R-squared            =   .5323241     |
|              Adjusted R-squared   =   .4513802     |
+----------------------------------------------------+
```

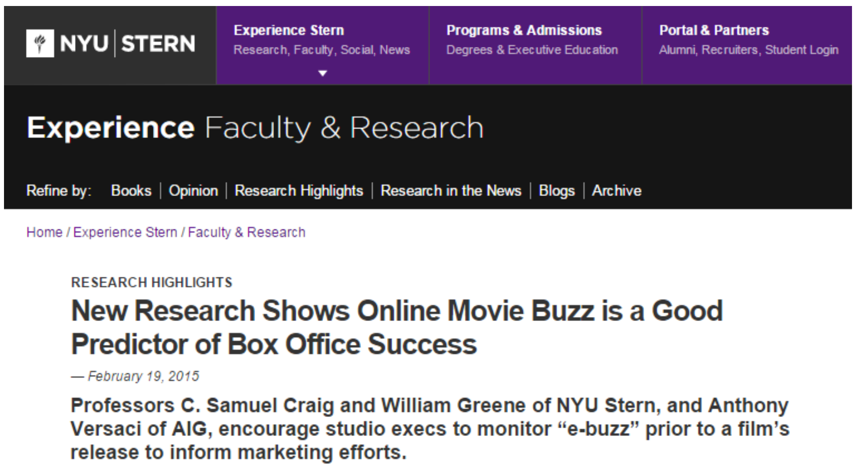
$$\mathbf{P}=\begin{bmatrix} p_{11} & 0 & 0 & 0 \\ p_{21} & 0 & 0 & 0 \\ p_{31} & 0 & 0 & 0 \\ p_{41} & 0 & 0 & 0 \end{bmatrix}$$

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	11.9602***	.91818	13.026	.0000	
LOGBUDGT	.38159**	.18711	2.039	.0465	3.71468
STARPOWR	.01303	.01315	.991	.3263	18.0316
SEQUEL	.33147	.28492	1.163	.2500	.14516
MPRATING	-.21185	.13975	-1.516	.1356	2.96774
ACTION	-.81404**	.30760	-2.646	.0107	.22581
COMEDY	.04048	.25367	.160	.8738	.32258
ANIMATED	-.80183*	.40776	-1.966	.0546	.09677
HORROR	.47454	.38629	1.228	.2248	.09677
PCBUZZ	.39704***	.08575	4.630	.0000	9.19362

Table 2
Predicting First Weekend Box Office
(Log of Total Box Office)

Variable	Model I Coefficient (Standard Error)	Model II Coefficient (Standard Error)
Constant	14.06^{a} (.940)	12.86^{a} (.899)
Log of Film Budget	.70^{a} (.204)	.26 (.178)
Star Power	.007 (.015)	.001 (.013)
Sequel	.65^{c} (.326)	.25 (.265)
MPAA Rating	-.13 (.163)	-.18 (.129)
Genre		
Action	-.30 (.337)	-.90^{b} (.286)
Comedy	.003 (.298)	-.016 (.236)
Animation	-.74 (.479)	-.74 (.379)
Horror	1.03^{c} (.432)	.41 (.359)
Internet Buzz (Awareness)	--	.23^{c} (.097)
Internet Buzz (Intention)	--	2.74^{b} (.881)
R^2	.34	.61
Adjusted R^2	.24	.53
F (8, 53)	3.4^{b}	
F (10, 51)		7.8^{a}

$^{a}p < .001$, $^{b}p < .01$, $^{c}p < .05$


NYU | STERN

Experience Stern
Research, Faculty, Social, News

Programs & Admissions
Degrees & Executive Education

Portal & Partners
Alumni, Recruiters, Student Login

Experience Faculty & Research

Refine by: Books | Opinion | Research Highlights | Research in the News | Blogs | Archive

Home / Experience Stern / Faculty & Research

RESEARCH HIGHLIGHTS

New Research Shows Online Movie Buzz is a Good Predictor of Box Office Success

— February 19, 2015

Professors C. Samuel Craig and William Greene of NYU Stern, and Anthony Versaci of AIG, encourage studio execs to monitor “e-buzz” prior to a film’s release to inform marketing efforts.

Adjusted R Squared

- Adjusted R^2 (adjusted for degrees of freedom)

$$\overline{R}^2 = 1 - \left(\frac{n-1}{n-K}\right)\left(1 - \overline{R}^2\right)$$

- Degrees of freedom" adjustment.
- $\overline{R}^2$ includes a penalty for variables that don't add much fit. Can fall when a variable is added to the equation.

NYU | STERN

Experience Stern — Research, Faculty, Social, News

Programs & Admissions — Degrees & Executive Education

Portal & Partners — Alumni, Recruiters, Student Login

Experience Faculty & Research

Refine by: Books | Opinion | Research Highlights | Research in the News | Blogs | Archive

Home / Experience Stern / Faculty & Research

RESEARCH HIGHLIGHTS

New Research Shows Online Movie Buzz is a Good Predictor of Box Office Success

— February 19, 2015

Professors C. Samuel Craig and William Greene of NYU Stern, and Anthony Versaci of AIG, encourage studio execs to monitor "e-buzz" prior to a film's release to inform marketing efforts.

$$.24 = 1 - \left(\frac{62-1}{62-9}\right)(1-.34)$$

Table 2
Predicting First Weekend Box Office
(Log of Total Box Office)

	Model I	Model II
Variable	Coefficient (Standard Error)	Coefficient (Standard Error)
Constant	14.06^a	12.86^a
	(.940)	(.899)
Log of Film Budget	.70^a	.26
	(.204)	(.178)
Star Power	.007	.001
	(.015)	(.013)
Sequel	.65^c	.25
	(.326)	(.265)
MPAA Rating	-.13	-.18
	(.163)	(.129)
Genre		
Action	-.30	-.90^b
	(.337)	(.286)
Comedy	.003	-.016
	(.298)	(.236)
Animation	-.74	-.74
	(.479)	(.379)
Horror	1.03^c	.41
	(.432)	(.359)
Internet Buzz (Awareness)	--	.23^c
		(.097)
Internet Buzz (Intention)	--	2.74^b
		(.881)
R^2	.34	.61
Adjusted R^2	.24	.53
F (8, 53)	3.4^b	
F (10, 51)		7.8^a

$^a p<.001$, $^b p<.01$, $^c p<.05$

Adjusted R^2

What is being adjusted?

The penalty for using up degrees of freedom.

$$\bar{R}^2 = 1 - [\mathbf{e'e}/(n - K)]/[\mathbf{y'M^0y}/(n-1)]$$

$$= 1 - [(n-1)/(n-K)(1 - R^2)]$$

Will $\bar{R}^2$ rise when a variable is added to the regression?

$\bar{R}^2$ is higher with *z* than without *z* if and only if the t ratio on *z* is in the regression when it is added is larger than one in absolute value. (See p. 46 in text.)

Full Regression (Without PD)

```
----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=G        Mean                  =      226.09444
             Standard deviation    =       50.59182
             Number of observs.    =             36
Model size   Parameters            =              9
             Degrees of freedom    =             27
Residuals    Sum of squares        =      596.68995
             Standard error of e   =        4.70102
Fit          R-squared             =         .99334  <**********
             Adjusted R-squared    =         .99137  <**********
Info criter. LogAmemiya Prd. Crt. =        3.31870  <**********
             Akaike Info. Criter. =        3.30788  <**********
Model test   F[  8,    27] (prob) =   503.3(.0000)
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
Constant|   -8220.38**      3629.309       -2.265   .0317
      PG|   -26.8313***      5.76403       -4.655   .0001      2.31661
       Y|     .02214***       .00711        3.116   .0043      9232.86
     PNC|    36.2027        21.54563        1.680   .1044      1.67078
     PUC|   -6.23235         5.01098       -1.244   .2243      2.34364
     PPT|    9.35681         8.94549        1.046   .3048      2.74486
      PN|    53.5879*       30.61384        1.750   .0914      2.08511
      PS|   -65.4897***     23.58819       -2.776   .0099      2.36898
    YEAR|    4.18510**       1.87283        2.235   .0339      1977.50
--------+-------------------------------------------------------------
```

PD added to the model. R^2 rises, Adjusted R^2 falls

```
----------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=G       Mean                 =      226.09444
            Standard deviation   =       50.59182
            Number of observs.   =             36
Model size  Parameters           =             10
            Degrees of freedom   =             26
Residuals   Sum of squares       =      594.54206
            Standard error of e  =        4.78195
Fit         R-squared            =         .99336  Was 0.99334
            Adjusted R-squared   =         .99107  Was 0.99137
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
Constant|   -7916.51**      3822.602       -2.071   .0484
      PG|   -26.8077***      5.86376       -4.572   .0001      2.31661
       Y|     .02231***       .00725        3.077   .0049      9232.86
     PNC|    30.0618        29.69543        1.012   .3207      1.67078
     PUC|   -7.44699         6.45668       -1.153   .2592      2.34364
     PPT|    9.05542         9.15246         .989   .3316      2.74486
      PD|    11.8023        38.50913         .306   .7617      1.65056 (NOTE LOW t ratio)
      PN|    47.3306        37.23680        1.271   .2150      2.08511
      PS|   -60.6202**      28.77798       -2.106   .0450      2.36898
    YEAR|    4.02861*        1.97231        2.043   .0514      1977.50
--------+-------------------------------------------------------------
```

Linear Least Squares Subject to Restrictions

Restrictions: Theory imposes certain restrictions on parameters.

Some common applications

- Dropping variables from the equation = certain coefficients in b forced to equal 0. (Probably the most common testing situation. "Is a certain variable significant?")
- Adding up conditions: Sums of certain coefficients must equal fixed values. Adding up conditions in demand systems. Constant returns to scale in production functions.
- Equality restrictions: Certain coefficients must equal other coefficients. Using real vs. nominal variables in equations.

General formulation for linear restrictions:

Minimize the sum of squares, **e′e,** subject to the linear constraint **Rb** = **q**.

Restricted Least Squares

In practice, restrictions can usually be imposed by solving them out.

1. Force a coefficient to equal zero. Drop the variable from the equation

   Problem: Minimize for $\beta_1, \beta_2, \beta_3 \sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3})^2$ subject to $\beta_3 = 0$

   Solution: Minimize for $\beta_1, \beta_2 \sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2})^2$

2. Adding up restriction. Impose $\beta_1 + \beta_2 + \beta_3 = 1$. Strategy: $\beta_3 = 1 - \beta_1 - \beta_2$.

   Solution: Minimize for $\beta_1, \beta_2 \sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - (1 - \beta_1 - \beta_2) x_{i3})^2$

   $$= \sum_{i=1}^{n}[(y_i - x_{i3}) - \beta_1 (x_{i1} - x_{i3}) - \beta_2 (x_{i2} - x_{i3})]^2$$

3. Equality restriction. Impose $\beta_3 = \beta_2$

   Minimize for $\beta_1, \beta_2, \beta_3 \sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3})^2$ subject to $\beta_3 = \beta_2$

   Solution: Minimize for $\beta_1, \beta_2 \sum_{i=1}^{n}[y_i - \beta_1 x_{i1} - \beta_2 (x_{i2} + x_{i3})]^2$

In each case, least squares using transformations of the data.

Restricted Least Squares Solution

- General Approach: Programming Problem
  Minimize for β $\quad L = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$
  subject to $\quad \mathbf{R}\beta = \mathbf{q}$
  Each row of **R** is the K coefficients in a restriction.
  There are J restrictions: J rows
- $\beta_3 = 0$: $\quad \mathbf{R} = [0,0,1,0,\ldots] \quad \mathbf{q} = (0)$. J=1
- $\beta_2 = \beta_3$: $\quad \mathbf{R} = [0,1,-1,0,\ldots] \quad \mathbf{q} = (0)$. J=1
- $\beta_2 = 0, \beta_3 = 0$: $\quad \mathbf{R} = \begin{bmatrix} 0,1,0,0,\ldots \\ 0,0,1,0,\ldots \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ J=2

Solution Strategy

- Quadratic program: Minimize quadratic criterion subject to linear restrictions
- All restrictions are binding
- Solve using Lagrangean formulation
- Minimize over (β,λ)

$$L^* = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) + 2\lambda'(\mathbf{R}\beta\mathbf{-q})$$

(The 2 is for convenience – see below.)

Restricted LS Solution

Necessary Conditions

$$\frac{\partial L^*}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + 2\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0}$$

$$\frac{\partial L^*}{\partial \boldsymbol{\lambda}} = 2(\mathbf{R}\beta - \mathbf{q}) = \mathbf{0}$$

Divide everything by 2. Collect in a matrix form

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{q} \end{pmatrix} \text{ or } \mathbf{A}\boldsymbol{\theta} = \mathbf{w}. \text{ Solution } \hat{\theta} = \mathbf{A}^{-1}\mathbf{w}$$

Does not rely on full rank of **X**.

Relies on column rank of $\mathbf{A} = K + J$.

Restricted Least Squares

If **X** has full rank, there is a partitioned solution for β* and λ*

$$\boldsymbol{\beta}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'](\mathbf{Rb}-\mathbf{q})$$

$$\boldsymbol{\lambda}^* = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'](\mathbf{Rb}-\mathbf{q})$$

where $\mathbf{b}$ = the simple least squares coefficients, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Note $\boldsymbol{\beta}^* = \mathbf{b}$ and $\boldsymbol{\lambda}^* = \mathbf{0}$ if $\mathbf{Rb}-\mathbf{q} = \mathbf{0}$.

Aspects of Restricted LS

1. $\mathbf{b}^* = \mathbf{b} - \mathbf{Cm}$ where

   $\mathbf{m}$ = the "discrepancy vector" $\mathbf{Rb} - \mathbf{q}$.

   Note what happens if $\mathbf{m} = \mathbf{0}$.

   What does $\mathbf{m} = \mathbf{0}$ mean?

2. $\lambda = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q}) = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{m}$.

   When does $\lambda = 0$. What does this mean?

3. Combining results: $\mathbf{b}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\lambda$.

   How could $\mathbf{b}^* = \mathbf{b}$?

Restrictions and the Criterion Function

Assume full rank **X** case. (The usual case.)

$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ uniquely minimizes $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathrm{y}-\mathbf{X}\beta) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$.

$(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) < (\mathbf{y}-\mathbf{Xb}^*)'(\mathbf{y}-\mathbf{X}\mathrm{b}^*)$ for any $\mathbf{b}^* \neq \mathbf{b}$.

Imposing restrictions cannot improve the criterion value.

It follows that $R^2{*} < R^2$. Restrictions must degrade the fit.