Multiple Regression Analysis: The Problem of Inference

Hypothesis Testing in Multiple Regression

1. Hypothesis Testing about Individual Regression Coefficients
2. Testing the Overall Significance of the Sample Regression
3. Testing the Equality of Two Regression Coefficients
4. Restricted Least Squares: Testing Linear Equality Restrictions
5. Testing for Structural or Parameter Stability of Regression Models: The Chow Test

Hypothesis Testing about Individual Regression Coefficients

- T-test
- P-value
- CI

Hypothesis Testing about Individual Regression Coefficients

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

(1)
- State the hypothesis

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

(2)
- t-value

$$t = \frac{(\hat{\beta}_j - \beta_j)}{se(\hat{\beta}_j)}, df = n - 3$$

(3)
- critical value

$$t > t_{\alpha/2} \quad or \quad t < t_{\alpha/2}$$

(4)
- Conclusion
  Reject the null hypothesis if $t > t_{\alpha/2}$
  Not Reject the null hypothesis if $t < t_{\alpha/2}$

(5) Conclusion ______

Example

$$CM = \beta_1 + \beta_2 PGNP + \beta_3 FLFP + u_i$$

TABLE 6.4 Fertility and Other Data for 64 Countries

Observation	CM	FLFP	PGNP	TFR	Observation	CM	FLFP	PGNP	TFR
1	128	37	1870	6.66	33	142	50	8640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1620	3.91
5	96	76	2050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1180	3.93	43	191	31	1010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1730	3.50	45	37	88	1730	3.46
14	165	31	1150	7.41	46	103	35	780	5.66
15	94	77	1160	4.21	47	67	85	1300	4.82
16	96	80	1270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1340	7.17
24	12	81	4240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4370	2.86
27	107	87	3020	6.66	59	121	41	1310	4.88
28	72	63	1420	7.28	60	115	62	1470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19830	5.23	62	47	85	3630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

Note: CM = Child mortality, the number of deaths of children under age 5 in a year per 1000 live births.
FLFP = Female literacy rate, percent.
PGNP = per capita GNP in 1980.
TFR = total fertility rate, 1980–1985, the average number of children born to a woman, using age-specific fertility rates for a given year.

Source: Chandan Mukherjee, Howard White, and Marc Whyte, *Econometrics and Data Analysis for Developing Countries,* Routledge, London, 1998, p. 456.

$Y_i = \beta_1 + \beta_2 PGNP + \beta_3 FLR + u_i$

Source	SS	df	MS
Model	257362.373	2	128681.187
Residual	106315.627	61	1742.87913
Total	363678	63	5772.66667

Number of obs	=	64
F(2, 61)	=	73.83
Prob > F	=	0.0000
R-squared	=	0.7077
Adj R-squared	=	0.6981
Root MSE	=	41.748

cm	Coef.	Std. Err.	t ①	P>|t| ③	[95% Conf. Interval] ②	
pgnp	-.0056466	.0020033	-2.82	0.006	-.0096524	-.0016408
flr	-2.231586	.2099472	-10.63	0.000	-2.651401	-1.81177
_cons	263.6416	11.59318	22.74	0.000	240.4596	286.8236

① $H_0: \beta_2 = 0$

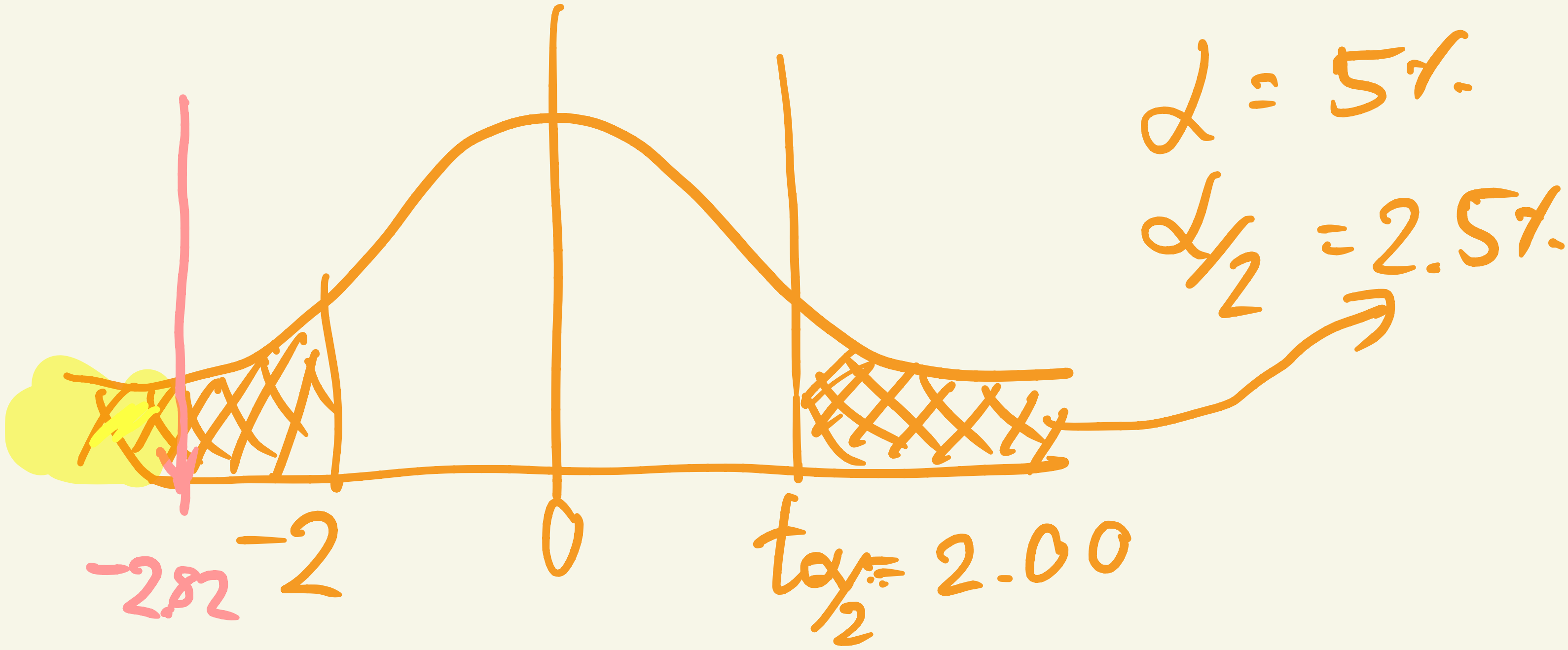
$H_1: \beta_2 \neq 0$

② p-value < $\alpha = 0.05$

$0.006 < \alpha = 0.05$

③ Reject H_0

④ Conclusion: There is enough evidence that $\beta_2 \neq 0$.



t = -2.82

p-value = 0.006

$1) H_0 : \beta_2 = 0 \qquad H_1 : \beta_2 \neq 0$

$$2) t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{-0.0056}{0.0020} = -2.8187$$

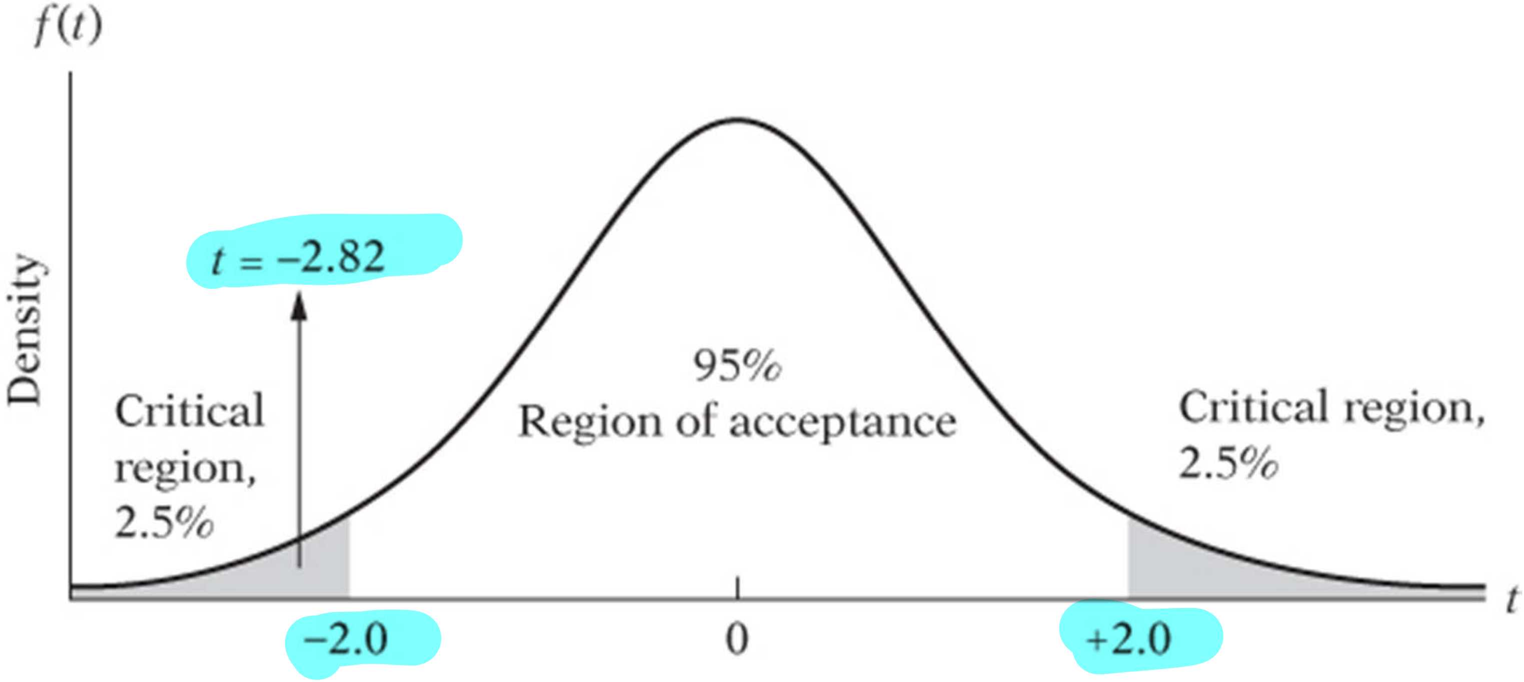
$3) df = 64 - 3 = 61$

The critical t value is 2 for a two-tail test $(\alpha = 0.05)$

4) *Since the computed t value of* 2.8187 *exceeds the critical t value of* 2

5) *We can reject the* H_0 *that PGNP has no effect on child mortality*

The female literacy rate held constant, per capita GNP has a significant negative effect on child mortality



f(t)
Density
t = −2.82
Critical
region,
2.5%
95%
Region of acceptance
Critical region,
2.5%
t
−2.0
0
+2.0

$\alpha = 5\%$

$df = 64 - 3 = 61$

$\pm t_{\alpha/2} = 2.000$

(1)

$$H_0 : \beta_2 = 0 \qquad H_1 : \beta_2 \neq 0$$

(2), (3)

$$\hat{\beta}_2 - t_{\alpha/2} se(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2} se(\hat{\beta}_2)$$

$$-0.0056 - 2(0.0020) \leq \beta_2 \leq -0.0056 + 2(0.0020)$$

$$-0.0096 \leq \beta_2 \leq -0.0016$$

(4) Since the interval does not include the null-hypothesized value of zero, we can reject the null hypothesis. The female literacy rate held constant, per capita GNP has a significant negative effect on child mortality

Testing the Overall Significance of the Sample Regression

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_1 : otherwise$$

Null hypothesis is a joint hypothesis that β_2 and β_3 are jointly or simultaneously equal to zero

Analysis of Variance (ANOVA)

- Total Sum of Square (TSS) consists of Explained Sum of Squares (ESS) and Residual Sum of Squares (RSS)

$$\sum y_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \sum \hat{u}_i^2$$

$$TSS \quad = ESS \qquad\qquad\qquad\qquad + RSS$$

k = # parameters

n = # obs.

$$F = \frac{(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})/2}{\sum \hat{u}_i^2 /(n-3)} = \frac{ESS/df}{RSS/df}$$

F distribution with degree of freedom k-1, n-k

TABLE 8.1
ANOVA Table for the Three-Variable Regression

Source of Variation	SS	df	MSS
Due to regression (ESS)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\dfrac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$
Due to residual (RSS)	$\sum \hat{u}_i^2$	$n - 3$	$\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n - 3}$
Total	$\sum y_i^2$	$n - 1$	

Example

TABLE 6.4 Fertility and Other Data for 64 Countries

Observation	CM	FLFP	PGNP	TFR	Observation	CM	FLFP	PGNP	TFR
1	128	37	1870	6.66	33	142	50	8640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1620	3.91
5	96	76	2050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1180	3.93	43	191	31	1010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1730	3.50	45	37	88	1730	3.46
14	165	31	1150	7.41	46	103	35	780	5.66
15	94	77	1160	4.21	47	67	85	1300	4.82
16	96	80	1270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1340	7.17
24	12	81	4240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4370	2.86
27	107	87	3020	6.66	59	121	41	1310	4.88
28	72	63	1420	7.28	60	115	62	1470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19830	5.23	62	47	85	3630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

Note: CM = Child mortality, the number of deaths of children under age 5 in a year per 1000 live births.
FLFP = Female literacy rate, percent.
PGNP = per capita GNP in 1980.
TFR = total fertility rate, 1980–1985, the average number of children born to a woman, using age-specific fertility rates for a given year.

Source: Chandan Mukherjee, Howard White, and Marc Whyte, *Econometrics and Data Analysis for Developing Countries,* Routledge, London, 1998, p. 456.

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ ①

$H_0: \beta_2 = \beta_3 = 0$

$CM = \beta_1 + \beta_2 PGNP + \beta_3 FLFP + u_i$

H_1: Otherwise

TABLE 8.3
ANOVA Table for the Child Mortality Example

Source of Variation	SS	df	MSS
Due to regression	257,362.4	2	128,681.2
Due to residuals	106,315.6	61	1742.88
Total	363,678	63	

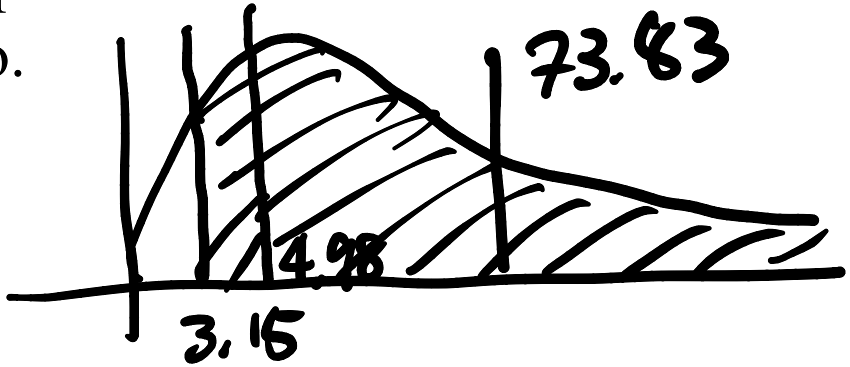
②

$$F = \frac{128,681.2}{1742.88} = 73.8325$$

F df: 2,61

α = 1%, 5%, 10%

The critical F value for 2 df in the numerator and 61 df in the denominator is 3.15 (5 % level of significance) or 4.98 (1% level of significance). Reject null hypothesis. There is evidence that not all parameter is equal to zero.



Class exercise

Find the critical F value

$$F_{0.05}(2,4)$$
$$F_{0.01}(2,4)$$
$$F_{0.05}(6,9)$$
$$F_{0.01}(10,20)$$
$$F_{0.05}(8,40)$$
$$F_{0.01}(4,120)$$

Class exercise

Find the Critical F value

$$F_{0.05}(2,4) = 6.94$$

$$F_{0.01}(2,4) = 18.0$$

$$F_{0.05}(6,9) = 3.37$$

$$F_{0.01}(10,20) = 3.37$$

$$F_{0.05}(8,40) = 2.18$$

$$F_{0.01}(4,120) = 3.48$$

Testing the Overall Significance of the Sample Regression

Testing the Overall Significance of a Multiple Regression-The F Test

Given the k-variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$$

H_1 : Not all slope coefficients are simultaneously zero

$$F = \frac{ESS / df}{RSS / df} = \frac{ESS / (k-1)}{RSS / (n-k)}$$

$$If\ \ F > F_\alpha(k-1, n-k), reject\ \ H_0$$

If F > critical region $F_{(1-\alpha);k-1,n-k}$ Reject H_0

If F < critical region $F_{(1-\alpha);k-1,n-k}$ Not reject H_0

An important relationship between R-squared and F

Assuming the normal distribution for the disturbances and the null hypothesis that $\beta_2 = \beta_3 = 0$

$$F = \frac{ESS/df}{RSS/df} = \frac{ESS/2}{RSS/(n-3)}$$

is distributed as the F distribution with 2 and n-3 df

k- variable case (including intercept)

Assuming the normal distribution for the disturbances and the null hypothesis that

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$$

$$F = \frac{ESS/df}{RSS/df} = \frac{ESS/(k-1)}{RSS/(n-k)}$$

is distributed as the F distribution with k-1 and n-k df

R-Squared and F

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} \quad (1)$$

$$= \frac{(n-k)}{(k-1)} \frac{ESS}{RSS} \quad (2)$$

$$= \frac{(n-k)}{(k-1)} \frac{ESS}{TSS - ESS} \quad (3)$$

$$= \frac{(n-k)}{(k-1)} \frac{ESS/TSS}{1-(ESS/TSS)} \quad (4)$$

$$= \frac{(n-k)}{(k-1)} \frac{R^2}{1-R^2} \quad (5)$$

$$= \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \quad (6)$$

$$TSS = ESS + RSS$$

$$RSS = TSS - ESS$$

$$\frac{TSS}{TSS}$$

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

Testing the Overall Significance of a Multiple Regression in Terms of R-Squared

Given the k-variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

①

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$$

H_1 : Not all slope coefficients are simultaneously zero

②

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

④

$$If \;\; F > F_\alpha(k-1, n-k), reject \;\; H_0$$

③

⑤ Conclusion

Example

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

$$H_1 : otherwise$$

TABLE 8.4
ANOVA Table in Terms of R^2

Source of Variation	SS	df	MSS*
Due to regression	$R^2(\sum y_i^2)$	2	$R^2(\sum y_i^2)/2$
Due to residuals	$(1 - R^2)(\sum y_i^2)$	$n - 3$	$(1 - R^2)(\sum y_i^2)/(n - 3)$
Total	$\sum y_i^2$	$n - 1$	

*Note that in computing the F value there is no need to multiply R^2 and $(1 - R^2)$ by $\sum y_i^2$ because it drops out, as shown in Eq. (8.4.12).

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$= \frac{0.7077/2}{(1-0.7077)/61} = 73.8726$$

The critical F value for 2 df in the numerator and 61 df in the denominator is 3.15 (5 % level of significance) or 4.98 (1% level of significance). Reject the null hypothesis. There is enough evidence to say that at least one parameter is not equal to zero.

- Overall significant test (F-test)
- Individual test (t-test)

Class practice

$$\log(wage) = \beta_1 + \beta_2 educ + \beta_3 Experience + \beta_4 Experience^2$$

Mincer model

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : otherwise$$

Source	SS	df	MS
Model	44.5393713	3	14.8464571
Residual	103.79038	522	.198832146
Total	148.329751	525	.28253286

Number of obs	= 526
F(3, 522)	= 74.67
Prob > F	= 0.0000
R-squared	= 0.3003
Adj R-squared	= 0.2963
Root MSE	= .44591

logwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0903658	.007468	12.10	0.000	.0756948	.1050368
exper	.0410089	.0051965	7.89	0.000	.0308002	.0512175
exper2	-.0007136	.0001158	-6.16	0.000	-.000941	-.0004861
_cons	.1279975	.1059323	1.21	0.227	-.0801085	.3361035

$$F = \frac{ESS/df}{RSS/df} = \frac{44.53931713/3}{103.79038/522} = 74.67$$

$$74.67 > Critical\ \ F\ \ value$$

$$reject\ \ H_0$$

There is evidence that not all parameter is equal to zero.

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$= \frac{0.3003/3}{(1-0.3003)/522} = 74.678$$

74.678 > Critical F-value

Reject null hypothesis.

There is evidence that not all parameter is equal to zero.

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$

The "incremental" or Marginal contribution of an explanatory variable

In most empirical investigations the researcher may be completely sure whether it is worth adding an X variable to the model knowing that several other X variables are already present in the model

One does not wish to include a variable (s) that contributes very little toward ESS.

One does not want to exclude a variable (s) that substantially increases ESS

How does one decide whether an X variable significantly reduces RSS?

Simple Regression model.

$H_0: \beta_j = 0$
$H_1: \beta_j \neq 0$

$$\hat{CM}_i = 157.4244 - 0.114 PGNP$$
$$t = (15.9894) \quad (-3.5156)$$
$$p\ value = (0.0000) \quad (0.0008)$$
$$r^2 = 0.1662$$

1. What is the marginal, or incremental, contribution of FLR, knowing that PGNP is already in the model and it is significantly related to CM?
2. Is the incremental contribution of FLR statistically significant?
3. What is the criterion for adding variables to the model?

$Y_i = \beta_1 + \beta_2 PGNP_i + \beta_3 FLR_i + u_i$

$\hat{Y}_i = \beta_1 + \beta_2 PGNP_i + u_i$

$$F = \frac{(ESS_{new} - ESS_{old}) / number\ of\ new\ regressors}{RSS_{new} / df(= n - number\ of\ parameters\ \text{in}\ the\ new\ Model)}$$

number of new regressors → 1

$64 - 3 = 61$ ($\beta_1, \beta_2, \beta_3$)

$$F = \frac{196{,}912.9}{1742.8786} = 112.9814$$

$$F > critical\ F_{\alpha}(number\ of\ new\ regressors, n - number\ of\ parameters\ \text{in}\ the\ new\ Model)$$

F value is highly significant, suggesting that the addition of FLR to the model significantly increases ESS and hence the R-square value

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

old

Source	SS	df	MS
Model	60449.4605	1	60449.4605
Residual	303228.539	62	4890.78289
Total	363678	63	5772.66667

Number of obs	=	64
F(1, 62)	=	12.36
Prob > F	=	0.0008
R-squared	=	0.1662
Adj R-squared	=	0.1528
Root MSE	=	69.934

old

cm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pgnp	-.0113645	.0032325	-3.52	0.001	-.0178262	-.0049027
_cons	157.4244	9.845583	15.99	0.000	137.7434	177.1055

new

Source	SS	df	MS
Model	257362.373	2	128681.187
Residual	106315.627	61	1742.87913
Total	363678	63	5772.66667

Number of obs	= 64
F(2, 61)	= 73.83
Prob > F	= 0.0000
R-squared	= 0.7077 new
Adj R-squared	= 0.6981
Root MSE	= 41.748

cm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pgnp	-.0056466	.0020033	-2.82	0.006	-.0096524	-.0016408
flr	-2.231586	.2099472	-10.63	0.000	-2.651401	-1.81177
_cons	263.6416	11.59318	22.74	0.000	240.4596	286.8236

$$F = \frac{(R^2_{new} - R^2_{old}) / number\ of\ new\ regressors}{(1 - R^2_{new}) / df (= n - number\ of\ parameters\ \text{in}\ the\ new\ Model)}$$

$$F = \frac{(0.7077 - 0.1662)/1}{(1 - 0.7077)/61} = 113.05$$

$$F > critical\ F_{\alpha}(number\ of\ new\ regressors, n - number\ of\ parameters\ \text{in}\ the\ new\ Model)$$

Class practice

$$\log(wage) = \beta_1 + \beta_2 educ$$

$$\log(wage) = \beta_1 + \beta_2 educ + \beta_3 Experience + \beta_4 Experience^2$$

1. What is the marginal, or incremental, contribution of Experience and Experience squared, knowing that educ is already in the model and it is significantly related to log(wage)?
2. Is the incremental contribution of Experience statistically significant?

Source	SS	df	MS
Model	27.5606288	1	27.5606288
Residual	120.769123	524	.230475425
Total	148.329751	525	.28253286

Number of obs	=	526
F(1, 524)	=	119.58
Prob > F	=	0.0000
R-squared	=	0.1858
Adj R-squared	=	0.1843
Root MSE	=	.48008

logwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0827444	.0075667	10.94	0.000	.0678796	.0976091
_cons	.5837727	.0973358	6.00	0.000	.3925563	.7749891

Source	SS	df	MS
Model	44.5393713	3	14.8464571
Residual	103.79038	522	.198832146
Total	148.329751	525	.28253286

Number of obs	=	526
F(3, 522)	=	74.67
Prob > F	=	0.0000
R-squared	=	0.3003
Adj R-squared	=	0.2963
Root MSE	=	.44591

logwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0903658	.007468	12.10	0.000	.0756948	.1050368
exper	.0410089	.0051965	7.89	0.000	.0308002	.0512175
exper2	-.0007136	.0001158	-6.16	0.000	-.000941	-.0004861
_cons	.1279975	.1059323	1.21	0.227	-.0801085	.3361035

$$F = \frac{(R^2_{new} - R^2_{old})/number\ of\ new\ regressors}{(1-R^2_{new})/df(= n - number\ of\ parameters\ \text{in}\ the\ new\ Model)}$$

$$F = \frac{(0.3003 - 0.1858)/2}{(1-0.3003)/522} = 42.71045$$

$$F > critical\ F_{\alpha}(number\ of\ new\ regressors, n - number\ of\ parameters\ \text{in}\ the\ new\ Model)$$

F value is highly significant, suggesting that the addition of Experience and Experience Squared to the model significantly increases ESS and hence the R-square value

Testing the Equality of Two Regression Coefficients

Testing the Equality of Two Regression Coefficients

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

$$H_0 : \beta_3 = \beta_4 \ or \ (\beta_3 - \beta_4) = 0$$

$$H_1 : \beta_3 \neq \beta_4 \ or \ (\beta_3 - \beta_4) \neq 0$$

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{se(\hat{\beta}_3 - \hat{\beta}_4)}$$

Degree of freedom = n-k

$se(\hat{\beta}_3)$ } STATA output table

$se(\hat{\beta}_4)$

$$se(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\operatorname{var}(\hat{\beta}_3) + \operatorname{var}(\hat{\beta}_4) - 2\operatorname{cov}(\hat{\beta}_3, \hat{\beta}_4)}$$

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4)}{\sqrt{\operatorname{var}(\hat{\beta}_3) + \operatorname{var}(\hat{\beta}_4) - 2\operatorname{cov}(\hat{\beta}_3, \hat{\beta}_4)}}$$

$\operatorname{var}(\hat{\beta}_3) = se(\hat{\beta}_3)^2$

$\operatorname{var}(\hat{\beta}_4) = se(\hat{\beta}_4)^2$

Example

TABLE 7.4
Total Cost (Y) and Output (X)

Output	Total Cost, $
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_i^2 + \hat{\beta}_4 X_i^3 \quad (*)$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3$$

Source	SS	df	MS
Model	38918.1562	3	12972.7187
Residual	64.7438228	6	10.7906371
Total	38982.9	9	4331.43333

Number of obs	= 10
F(3, 6)	= 1202.22
Prob > F	= 0.0000
R-squared	= 0.9983
Adj R-squared	= 0.9975
Root MSE	= 3.2849

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x ($\hat{\beta}_2$)	63.47766	4.778607	13.28	0.000	51.78483	75.17049
x2 ($\hat{\beta}_3$)	-12.96154	.9856646	-13.15	0.000	-15.37337	-10.5497
x3 ($\hat{\beta}_4$)	.9395882	.0591056	15.90	0.000	.794962	1.084214
_cons ($\hat{\beta}_1$)	141.7667	6.375322	22.24	0.000	126.1668	157.3665

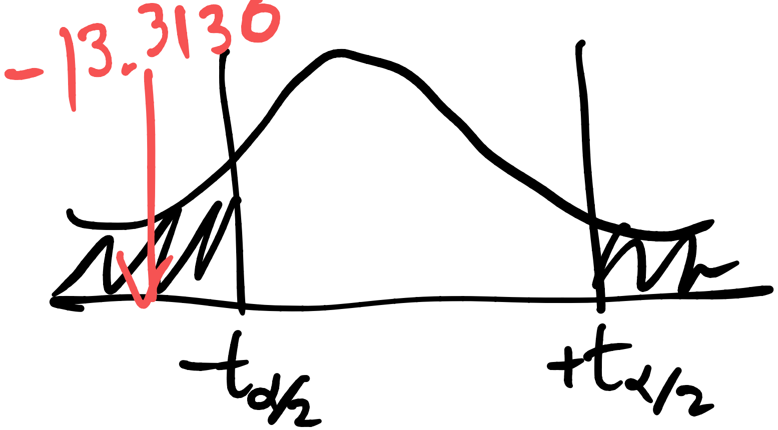
$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 X_i^2 + \hat{\beta}_4 X_i^3$$

$$\hat{Y}_i = 141.7667 + 63.4777 X_i - 12.9615 X_i^2 + 0.9396 X_i^3$$

$$se = (6.3753) \quad (4.7786) \quad (0.9857) \quad (0.0591)$$

$$COV(\hat{\beta}_3, \hat{\beta}_4) = -0.0576$$

$$R^2 = 0.9983$$



$t_{df=6}$

$\alpha = 1\%, 5\%, 10\%$

$$H_0 : \beta_3 = \beta_4$$

$$H_1 : \beta_3 \neq \beta_4$$

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\,\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}}$$

$$= \frac{-12.9615 - 0.9396}{\sqrt{(0.9867)^2 + (0.0591)^2 - 2(-0.0576)}} = \frac{-13.9011}{1.0442} = -13.3130$$

Degree of freedom = n-k=10-4=6   Check critical value

Reject the null hypothesis

There is not enough evidence to say that $\beta_3 = \beta_4$

$H_0: \beta_3 - \beta_4 = 0 \; ; \beta_3 = \beta_4$

$H_1: \beta_3 - \beta_4 \neq 0 \; ; \beta_3 \neq \beta_4$

~~$H_0: \beta_3 = \beta_4 = 0.8$~~

$H_0: \beta_3 - \beta_4 = 0$

$H_1: \beta_3 - \beta_4 \neq 0$

Restricted Least Squares: Testing Linear Equality Restrictions

Example
Cobb- Douglas production function

unrestricted model.

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{\mu_i}$$

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

where $\beta_0 = \ln \beta_1$

Is this restriction valid?

$$\beta_2 + \beta_3 = 1$$

The t-test approach

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{se(\hat{\beta}_2 + \hat{\beta}_3)}$$

$$= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{\sqrt{\operatorname{var}(\hat{\beta}_2) + \operatorname{var}(\hat{\beta}_3) + 2\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3)}}$$

The F-test approach: Restricted Least Squares

From $\beta_2 + \beta_3 = 1$

we see that

$$\beta_2 = 1 - \beta_3$$

$$\beta_3 = 1 - \beta_2$$

Therefore, using either of these equalities, we can eliminate one of the β coefficients in

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

where $\beta_0 = \ln \beta_1$

and estimate the resulting equation

$$\ln Y_i = \beta_0 + (1-\beta_3)\ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$= \beta_0 + \ln X_{2i} + \beta_3 (\ln X_{3i} - \ln X_{2i}) + u_i$$

$$\ln Y_i - \ln X_{2i} = \beta_0 + \beta_3 (\ln X_{3i} - \ln X_{2i}) + u_i$$

$$\ln\left(\frac{Y_i}{X_{2i}}\right) = \beta_0 + \beta_3 \ln\left(\frac{X_{3i}}{X_{2i}}\right) + u_i$$

$\frac{Y_i}{X_{2i}}$ = output labor ratio

$\frac{X_{2i}}{X_{3i}}$ = Capital labor ratio

Once we estimate β_3 from the previous equation, β_2 can be easily estimated from $\beta_2 = 1 - \beta_3$

The procedure outlined in

$$\ln\left(\frac{Y_i}{X_{2i}}\right) = \beta_0 + \beta_3 \ln\left(\frac{X_{3i}}{X_{2i}}\right) + u_i$$

is known as **restricted least squares (RLS)**

This procedure can be generalized to models containing any number of explanatory variables and more than one linear equality restriction

The F-Test Approach: Restricted Least Squares

$\sum \hat{u}_{UR}^2$ RSS of the unrestricted regression

$\sum \hat{u}_{R}^2$ RSS of the restricted regression

m Number of linear restrictions

k Number of parameters in the unrestricted regression

n Number of observations

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$$

$$= \frac{(\sum \hat{u}_R^2 - \sum \hat{u}_{UR}^2)/m}{\sum \hat{u}_{UR}^2/(n-k)}$$

$$\sum \hat{u}_{UR}^2 \leq \sum \hat{u}_R^2$$

F distribution with degree of freedom m, n-k

$$F = \frac{(R_{UR}^2 - R_R^2)/m}{(1 - R_{UR}^2)/(n-k)}$$

$$R_{UR}^2 \geq R_R^2$$

F distribution with degree of freedom m, n-k

Example

TABLE 8.8
Real GDP, Employment, and Real Fixed Capital—Mexico

Source: Victor J. Elias, *Sources of Growth: A Study of Seven Latin American Economies,* International Center for Economic Growth, ICS Press, San Francisco, 1992. Data from Tables E5, E12, and E14.

Year	GDP*	Employment†	Fixed Capital‡
1955	114043	8310	182113
1956	120410	8529	193749
1957	129187	8738	205192
1958	134705	8952	215130
1959	139960	9171	225021
1960	150511	9569	237026
1961	157897	9527	248897
1962	165286	9662	260661
1963	178491	10334	275466
1964	199457	10981	295378
1965	212323	11746	315715
1966	226977	11521	337642
1967	241194	11540	363599
1968	260881	12066	391847
1969	277498	12297	422382
1970	296530	12955	455049
1971	306712	13338	484677
1972	329030	13738	520553
1973	354057	15924	561531
1974	374977	14154	609825

*Millions of 1960 pesos.
†Thousands of people.
‡Millions of 1960 pesos.

$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{\mu_i}$$

$$\ln GDP_t = -1.6524 + 0.3397 \ln Labor_t + 0.8460 \ln Capital_t$$

$$t \quad = (-2.7259) \quad (1.8295) \qquad (9.0625)$$

$$p \ value = (0.0144) \quad (0.0849) \qquad (0.0000)$$

$$R^2 = 0.9951 \qquad RSS_{UR} = 0.0136$$

"unrestricted model"

$$\ln \widehat{GDP}_t = -1.6524 + 0.3397 \ln Labor_t + 0.8460 \ln Capital_t$$

$$t = (-2.7259) \qquad (1.8295) \qquad (9.0625)$$

$$p\ value = (0.0144) \qquad (0.0849) \qquad (0.0000)$$

$$R^2 = 0.9951 \qquad RSS_{UR} = 0.0136$$

As you can see, the output/labor elasticity is about 0.34 and the output/capital elasticity is about 0.85. If we add these coefficients, we obtain 1.19, suggesting that perhaps the Mexican economy during the stated time period was experiencing increasing returns to scale.

Restriction statement "$\beta_2 + \beta_3 = 1$"

Let us impose the restriction of constant returns to scale

$$\widehat{\ln(\frac{GDP}{Labor})_t} = -0.4947 + 1.0153 \ln(\frac{Capital}{Labor})_t$$

$$t = \quad (-4.0612) \qquad (28.1056)$$

$$p\ value = \quad (0.0007) \qquad (0.0000)$$

$R^2{}_R = 0.9777$ $\quad RSS_R = 0.0166$

H0: β2 + β3 = 1 "CRTS"

H1: otherwise

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$$

$$= \frac{(0.0166 - 0.0136)/1}{0.0136/(20-3)} = 3.75$$

m n−k.

F-distribution with degree of freedom 1, 17

F-value is not significant at the 5% level

The conclusion is that the Mexican economy was probably characterized by constant returns to scale over the sample period and therefore there may be no harm in using the restricted regression

Example
The demand for Chicken in the United States, 1960-1982

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 \ln X_{5t} + u_i$$

$where\ Y = per\ capita\ consumption, lb$

$X_2 = real\ disposable\ per\ capita\ income, \$$

$X_3 = real\ retail\ price\ of\ chicken\ per\ lb, cents$

$X_4 = real\ retail\ price\ of\ pork\ per\ lb, cents$

$X_5 = real\ retail\ price\ of\ beef\ per\ lb, cents$

$$\beta_2 > 0$$

$$\beta_3 < 0$$

$$\beta_4 > 0, \textit{if chicken and pork are competing products}$$

$$< 0, \textit{if chicken and pork are complementary products}$$

$$= 0, \textit{if chicken and pork are unrelated products}$$

$$\beta_5 > 0, \textit{if chicken and beef are competing products}$$

$$< 0, \textit{if chicken and beef are complementary products}$$

$$= 0, \textit{if chicken and beef are unrelated products}$$

Suppose someone maintains that chicken and pork and beef are unrelated products in the sense that chicken consumption is not affected by the prices of pork and beef.

$$H_0 : \beta_4 = \beta_5 = 0$$

$$H_1 : otherwise$$

restricted model / regression

Therefore, the constrained regression becomes

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_i$$

Unconstrained regression:

$$\widehat{\ln Y_t} = 2.1898 + 0.3425 \ln X_{2t} - 0.5046 \ln X_{3t} + 0.1485 \ln X_{4t} + 0.0911 \ln X_{5t}$$

$$(0.1557) \quad (0.0833) \quad (0.1109) \quad (0.0997) \quad (0.1007)$$

$$R^2_{UR} = 0.9823$$

Constrained regression:

$$\widehat{\ln Y_t} = 2.0328 + 0.4515 \ln X_{2t} - 0.3772 \ln X_{3t}$$

$$(0.1162) \quad (0.0247) \quad (0.0635)$$

$$R^2_{R} = 0.9801$$

① $H_0: \beta_3 = \beta_4 = 0$
H_1: otherwise

②

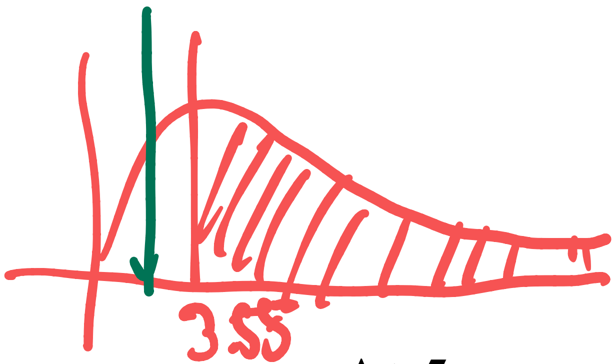
$$F = \frac{(R^2_{UR} - R^2_{R})/m}{(1 - R^2_{UR})/(n-k)}$$

$$= \frac{(0.9823 - 0.9801)/2}{(1 - 0.9801)/18} = 1.1224$$

$n-k$.

$F = 1.1224$

price of pork is correlated to price of beef



F df 2, 18
α = 1%.
5%.
10%

At 5 percent significance level, Critical F is 3.55.

Cannot reject the null hypothesis.

We can accept the constrained regression as representing the demand function for chicken.

Testing for Structural or Parameter Stability of Regression Models: The Chow Test

Testing for Structural or Parameter Stability of Regression Models: The Chow Test

"Structural Change" mean that the values of parameters of the model do not remain the same through the entire period

$t = 100$

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + u_t$$

TABLE 8.9
Savings and Personal Disposable Income (billions of dollars), United States, 1970–1995

Source: *Economic Report of the President*, 1997, Table B-28, p. 332.

Observation	Savings	Income	Observation	Savings	Income
1970	61.0	727.1	1983	167.0	2522.4
1971	68.6	790.2	1984	235.7	2810.0
1972	63.6	855.3	1985	206.2	3002.0
1973	89.6	965.0	1986	196.5	3187.6
1974	97.6	1054.2	1987	168.4	3363.1
1975	104.4	1159.2	1988	189.1	3640.8
1976	96.4	1273.0	1989	187.8	3894.5
1977	92.5	1401.4	1990	208.7	4166.8
1978	112.6	1580.1	1991	246.4	4343.7
1979	130.1	1769.5	1992	272.6	4613.7
1980	161.8	1973.3	1993	214.4	4790.2
1981	199.1	2200.2	1994	189.4	5021.7
1982	205.5	2347.3	1995	249.3	5320.8

$$Saving_t = \hat{\beta}_1 + \hat{\beta}_2 Income_t + \hat{u}_t$$

- This table gives data on disposable personal income and personal savings, in billions of dollars, the U.S. for the period 1970-1995
- We want to estimate a simple savings function that relates savings (Y) to disposable personal income DPI (X)
- In 1982 the United States suffered its worst peacetime recession –unemployment rate reached 9.7%

- Divide sample data into two time periods:
- 1970-1981 and 1982-1995

Three possible regressions:

Time period 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$ (1)

Time period 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14$ (2)

Time period 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_t \quad n = 26$ (3)

$RSS_{UR} = RSS_1 + RSS_2$

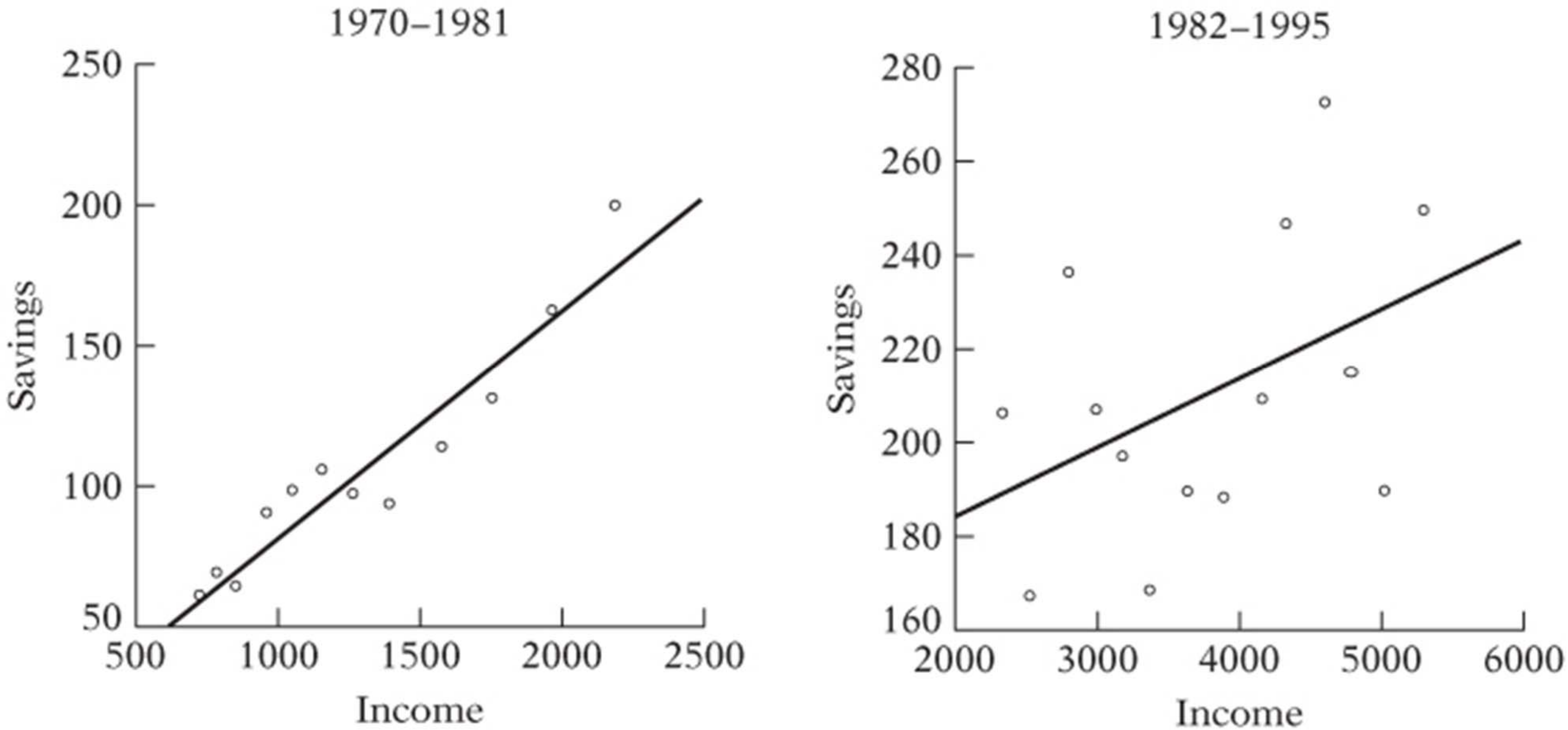
$$\hat{Y}_t = 1.0161 + 0.0803X_t$$
$$t = (0.0873)\ \ (9.6015)$$
$$R^2 = 0.9021 \quad RSS_1 = 1785.032 \quad df = 10$$

$$\hat{Y}_t = 153.4947 + 0.0148X_t$$
$$t = (4.6922)\ \ (1.7707)$$
$$R^2 = 0.2971 \quad RSS_2 = 10{,}005.22 \quad df = 12$$

$$\hat{Y}_t = 62.4226 + 0.0376X_t + \ldots$$
$$t = (4.8917)\ \ (8.8937) + \ldots$$
$$R^2 = 0.7672 \quad RSS_3 = 23{,}248.30 \quad df = 24$$

RSS_R

- The slope in the preceding savings-income regressions represents the **marginal propensity to save (MPS),** the mean change in savings as a result of a dollar's increase in disposable personal income


1970–1981
250
200
150
100
50
Savings
500
1000
1500
2000
2500
Income
1982–1995
280
260
240
220
200
180
160
Savings
2000
3000
4000
5000
6000
Income

Chow test

Assumption

1. $u_{1t} \sim N(0, \sigma^2)$ and $u_{2t} \sim N(0, \sigma^2)$ - The error terms in the subperiod regressions are normally distributed with the same (homoscedastic) variance
2. The two error terms are independently distributed

The mechanics of the Chow test

1. Estimate $Y_t = \alpha_1 + \alpha_2 X_t + u_t \quad n = 26$, which is appropriate if there is no parameter instability, and obtain RSS_3 with df = $(n_1 + n_2 - k)$. We call RSS_3 the restricted residual sum of squares (RSS_R)

2. Estimate $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$ and obtain its residual sum of squares, RSS_1, with df = $(n_1 - k)$

3. Estimate $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14$ and obtain its residual sum of squares, RSS_2, with df = $(n_2 - k)$

4. Since the two sets of samples are deemed independent, we can add RSS_1 and RSS_2 to obtain what may be called the **unrestricted residual sum of squares** (RSS_{UR})

$$RSS_{UR} = RSS_1 + RSS_2 \quad with\ df = (n_1 + n_2 - 2k)$$

$$RSS_{UR} = (1785.032 + 10,005.22) = 11,790.252$$

5. If there is no structural change, then the RSS_{UR} and RSS_{R} should not be statistically different.

$$F = \frac{(RSS_R - RSS_{UR})/k}{(RSS_{UR})/(n_1 + n_2 - 2k)} \sim F_{[k,(n_1+n_2-2k)]}$$

then the Chow has shown that under the null hypothesis the regression $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}$ $n_1 = 12$ and $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $n_2 = 14$ are statistically the same

H_0: $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$

H_1: otherwise

6. We find that for 2, 22 df the 1 percent critical F value is 5.72.

$$F = \frac{(23{,}248.30 - 11{,}790.252)/2}{(11{,}790.252)/22} = 10.69$$

Therefore, the probability of obtaining F value of as much as or greater than 10.69. We reject the null hypothesis of parameter stability and conclude that the regressions $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$ and $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14$ are different

Source

Gujarati, D.N. (2009) Basic Econometrics. 5th ed. Singapore, McGraw-Hill.