Економетрика

ЛЕКЦІЯ 4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Д.Е.Н., ПРОФЕСОР СТАВИЦЬКИЙ А.В.

Модель

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1,n}$$

Перевірка гіпотез

- Статистичною гіпотезою називається будь-яке твердження про вигляд або властивості регресії, її коефіцієнтів чи її параметрів.
- Спочатку формулюється гіпотеза H_0 та зазначається величина α – похибка першого роду, тобто ймовірність прийняти альтернативну гіпотезу H_1, коли правильна H_0.
- На основі побудованої регресії розраховується деяке практичне значення, яке має порівнятися з теоретичним значенням, яке береться зі статистичних таблиць. Якщо практичне значення менше, ніж теоретичне, то вважається, що гіпотеза H_0 є вірною.

Перевірка (не)адекватності регресії

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_{k-1} = 0 \text{ або } H_0: R^2 = 0$$

$$F_{pr} = \frac{\dfrac{R^2}{k-1}}{\dfrac{1-R^2}{n-k}} \sim F_{teor}\left(1-\alpha, k-1, n-k\right)$$

- Якщо $F_{pr} > F_{teor}$, то модель вважається **адекватною**. У протилежному випадку – неадекватною.

Приклад

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 08/27/13 Time: 13:36
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.40959	1.376227	22.82297	0.0000
YD	0.001839	0.000405	4.538711	0.0001
PB	0.247457	0.070428	3.513599	0.0015
PC	-0.819809	0.089305	-9.179902	0.0000

R-squared	0.963632	Mean dependent var	35.87879
Adjusted R-squared	0.959870	S.D. dependent var	9.927763
S.E. of regression	1.988782	Akaike info criterion	4.326134
Sum squared resid	114.7024	Schwarz criterion	4.507529
Log likelihood	-67.38122	Hannan-Quinn criter.	4.387168
F-statistic	256.1347	Durbin-Watson stat	0.753680
Prob(F-statistic)	0.000000		

- Оскільки Prob. (F-statistic)=0.0000, що менше ніж 0,05 (рівень похибки), а тим більше 0,01, тоді можна зробити висновок про адекватність моделі при рівні значущості 0,05 та 0,01.

Перевірка гіпотези про значимість коефіцієнта регресії

$$H_0 : \beta_j = 0$$

$$t_{pr} = \left| \frac{\hat{\beta}_j}{s.e.\left(\hat{\beta}_j\right)} \right| \sim t_{teor}\left(1-\alpha, n-k\right)$$

- Якщо $t_{pr} < t_{teor}$, то гіпотеза H_0 приймається, тобто коефіцієнт β_j є **незначимим**.

Перевірка гіпотези про значення коефіцієнта регресії

$$H_0 : \beta_j = m$$

$$t_{pr} = \frac{\left|\hat{\beta}_j - m\right|}{s.e.(\hat{\beta}_j)} \sim t_{teor}\left(1-\alpha, n-k\right)$$

- Якщо t_{pr}<t_{teor}, то гіпотеза H_0 приймається, тобто коефіцієнт β_j можна прийняти рівним *m*.

Приклад

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 08/27/13 Time: 13:36
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.40959	1.376227	22.82297	0.0000
YD	0.001839	0.000405	4.538711	0.0001
PB	0.247457	0.070428	3.513599	0.0015
PC	-0.819809	0.089305	-9.179902	0.0000

R-squared	0.963632	Mean dependent var	35.87879
Adjusted R-squared	0.959870	S.D. dependent var	9.927763
S.E. of regression	1.988782	Akaike info criterion	4.326134
Sum squared resid	114.7024	Schwarz criterion	4.507529
Log likelihood	-67.38122	Hannan-Quinn criter.	4.387168
F-statistic	256.1347	Durbin-Watson stat	0.753680
Prob(F-statistic)	0.000000		

- H_0: $\beta_2=2$
- H_1: $\beta_2\neq2$

$$t_{pr} = \frac{|0{,}247457 - 2|}{0{,}070428} = 23{,}5937$$

- Теоретичне значення t-статистики для рівня надійності 0,95 та 29 степенів свободи дорівнює 2,045.

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Перевірка статистичних гіпотез без комп'ютера

На основі 30 спостережень була оцінена така регресія:

$$\underset{(3,14)}{y = 0,25} + \underset{(1,82)}{1,14}\, x_1 - \underset{(0,92)}{2,45}\, x_2 \qquad RSS = 1,16 \qquad TSS = 8,67$$

(у дужках наведено t-значення для коефіцієнтів моделі).

1. Визначити, які з коефіцієнтів регресії є значимими з рівнем надійності 0,95.
2. Перевірити гіпотезу $\beta_1=1$ з рівнем надійності 0,95.
3. Підрахувати коефіцієнт детермінації та скоригований коефіцієнт детермінації.
4. Перевірити модель на адекватність з рівнем надійності 0,9.

Розв'язок - 1

$$y = \underset{(3,14)}{0,25} + \underset{(1,82)}{1,14}\,x_1 - \underset{(-0,92)}{2,45}\,x_2$$

- Для перевірки значимості коефіцієнтів слід порівняти практичні значення t-статистик, що розташовані під коефіцієнтами моделі, з теоретичним значенням

$$t_{teor} = t\left(1-\alpha; n-k\right) = t\left(0,95; 27\right) = 2,052$$

- Таким чином, коефіцієнти β_1 та β_2 є статистично незначимими, а коефіцієнт β_0 – статистично значимим.

Розв'язок - 2

$$H_0 : \beta_1 = 1 \qquad y = \underset{(3,14)}{0,25} + \underset{(1,82)}{1,14}\, x_1 - \underset{(-0,92)}{2,45}\, x_2$$

- Визначимо стандартне відхилення для коефіцієнта β_1:

$$1,82 = t_{pr} = \frac{\hat{\beta}_1}{s.e.\left(\hat{\beta}_1\right)} = \frac{1,14}{s.e.\left(\hat{\beta}_1\right)} \qquad s.e.\left(\hat{\beta}_1\right) = \frac{1,14}{1,82} = 0,626$$

- Тоді маємо: $t_{pr} = \left|\frac{\hat{\beta}_1 - 1}{s.e.\left(\hat{\beta}_1\right)}\right| = \frac{1,14-1}{0,626} = 0,22$

що менше за теоретичне значення $t_{teor} = t\left(0,95; 27\right) = 2,052$

Таким чином, значення коефіцієнта β_1 можна прийняти рівним 1.

Розв'язок - 3

$$y = \underset{(3,14)}{0,25} + \underset{(1,82)}{1,14}\, x_1 - \underset{(-0,92)}{2,45}\, x_2 \qquad RSS = 1,16 \qquad TSS = 8,67$$

- Коефіцієнт детермінації дорівнює

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{1,16}{8,67} = 0,866$$

- Скоригований коефіцієнт детермінації:

$$R^2_{adj} = 1 - \frac{\frac{RSS}{n-k}}{\frac{TSS}{n-1}} = 1 - \frac{\frac{1,16}{27}}{\frac{8,67}{29}} = 0,856$$

Розв'язок - 4

$H_0 : R^2 = 0$

- Перевіримо модель на адекватність:

$$F_{pr} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{0,866}{2}}{\frac{1-0,866}{27}} = 87,246$$

$$F_{teor} = F\left(0,9; 2; 27\right) = 2,51$$

- Таким чином, оскільки практичне значення більше за теоретичне, то модель виявилася адекватною.

Перевірка гіпотези про лінійні обмеження на коефіцієнти регресії

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1,n}$$

$$y = X\beta + \varepsilon$$

$$H_0: \begin{cases} \theta_{1,0}\beta_0 + \theta_{1,1}\beta_1 + \ldots + \theta_{1,k-1}\beta_{k-1} = r_1, \\ \theta_{2,0}\beta_0 + \theta_{2,1}\beta_1 + \ldots + \theta_{2,k-1}\beta_{k-1} = r_2, \\ \ldots \\ \theta_{J,0}\beta_0 + \theta_{J,1}\beta_1 + \ldots + \theta_{J,k-1}\beta_{k-1} = r_J. \end{cases}$$

$$H_0 : \Theta\beta = r$$

- Тест Вальда

$$F_{pr} = \frac{\dfrac{\left(\Theta\hat{\beta} - r\right)^T \left(\Theta\left(X^T X\right)^{-1}\Theta^T\right)^{-1} \left(\Theta\hat{\beta} - r\right)}{J}}{\dfrac{RSS}{n-k}}$$

$$F_{teor} = F\left(1-\alpha; J;\ n-k\right)$$

- Якщо $F_{pr} < F_{teor}$, то гіпотеза H_0 приймається

Приклад

- Відома інформація по деяких підприємствах України по випуску продукції Y (млн. грн.), основному капіталу K (млн. грн.), чисельності працюючих L (тис.люд.-год.).

№	Y	K	L	№	Y	K	L
1	64,3	42,4	13,5	**19**	79	47,6	23,6
2	47,2	30,7	21	**20**	62,9	41,6	9,2
3	63,6	48,9	16,1	**21**	62,8	42,1	13,4
4	117,9	61,3	25,9	**22**	77,7	41,6	22,2
5	111,3	60,4	22,5	**23**	106,5	62,1	35,2
6	123	66,8	32,1	**24**	96,1	45	24,5
7	26,5	19,8	5,7	**25**	83,9	43,8	24,1
8	71,9	43,2	18,8	**26**	61,8	39,8	8,5
9	118,1	59,1	29,1	**27**	119,4	69,7	28,2
10	77,3	33	21,9	**28**	65	56,4	16
11	69,2	37,7	26,7	**29**	95,6	74,5	22,1
12	48,4	25,7	24	**30**	51,8	38,1	12,9
13	42,1	23,8	11,3	**31**	137,9	65,9	37,7
14	53,5	34,3	10,4	**32**	50,2	26,7	21,4
15	46,8	36,7	16,3	**33**	64	52,5	13,5
16	42,5	20,6	9	**34**	84,8	56,8	16,2
17	84	39,2	29,3	**35**	119,1	69	24,9
18	69,5	41,2	29				

Приклад

- Необхідно оцінити виробничу функцію Кобба-Дугласа

$$Y_t = \beta_0 K_t^{\beta_1} L_t^{\beta_2} + \varepsilon_t$$

- та перевірити гіпотезу з рівнем надійності 90%:

$$\begin{cases} \beta_1 + \beta_2 = 1, \\ \beta_0 = 2. \end{cases}$$

Приклад

- Для оцінювання виробничу функцію слід перетворити до множинної лінійної регресії шляхом логарифмування:

$$\ln Y_t = \ln \beta_0 + \beta_1 K_t + \beta_2 L_t + \varepsilon_t$$

- Оцінюємо отриману регресію звичайним методом найменших квадратів:

$$\ln Y = 0,63 + 0,72 \ln K + 0,32 \ln L \qquad R^2 = 0,94 \qquad RSS = 0,5847$$

Приклад

- Гіпотеза: $\begin{cases} \beta_1 + \beta_2 = 1, \\ \beta_0 = 2, \end{cases}$ або

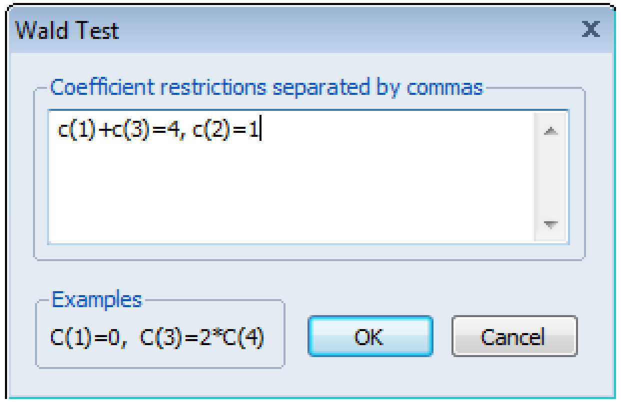
$$\Theta = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad r = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad J = 2 \qquad n - k = 32$$

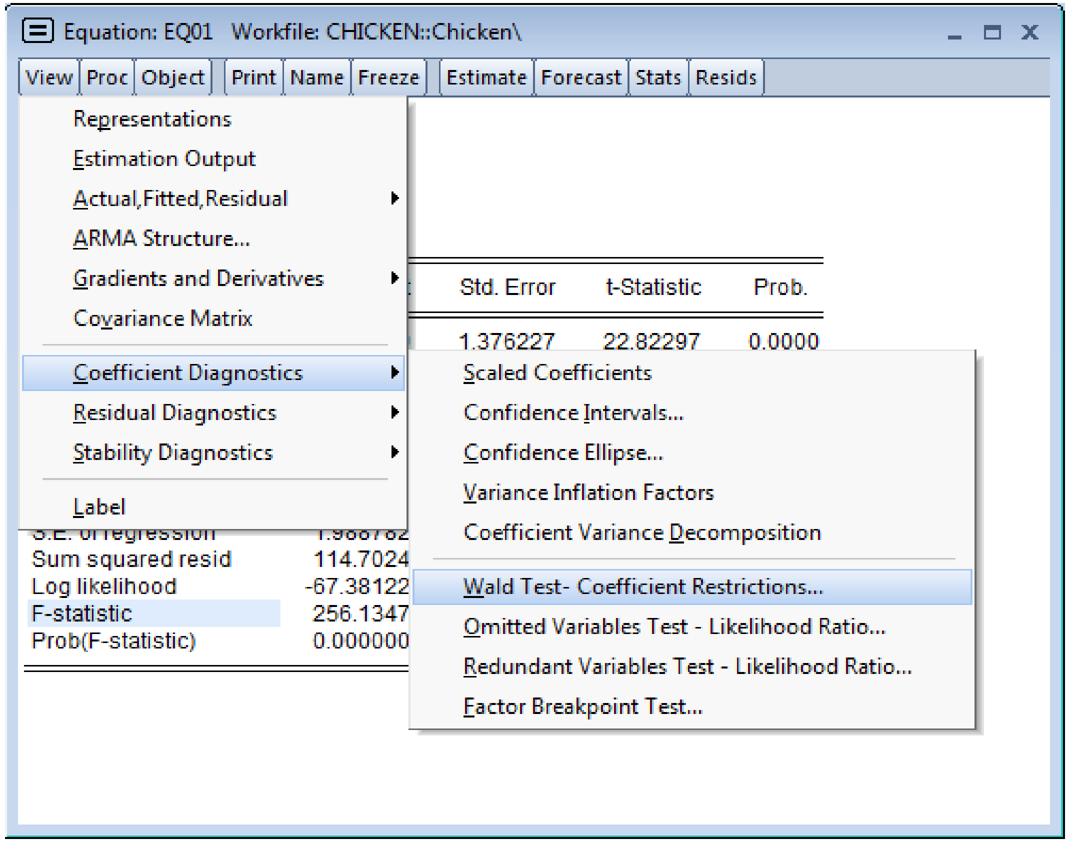
$$F_{pr} = \frac{\dfrac{\left(\Theta\hat{\beta} - r\right)^T \left(\Theta\left(X^T X\right)^{-1} \Theta^T\right)^{-1} \left(\Theta\hat{\beta} - r\right)}{J}}{\dfrac{RSS}{n-k}} = 1,22 \qquad F_{teor} = F\left(0,9; 2; 32\right) = 2,48$$

Оскільки $F_{pr}<F_{teor}$, то гіпотеза про лінійні обмеження приймається, тобто підприємства мають постійну віддачу від масштабу.

Приклад

$$H_0:\begin{cases}\beta_0+\beta_2=4,\\ \beta_1=1.\end{cases}$$





- Зверніть увагу, що нумерація коефіцієнтів починається з 1, тому c(1) відповідає β_0, c(2) – β_1 тощо).

Приклад

- Результати перевірки свідчать про те, що нульова гіпотеза має бути відхиленою, оскільки значення Probability менше за 0,05 (рівень похибки).

- Значенням *F*-статистики можна користуватися лише за припущення про нормальний розподіл збурень регресії, в іншому випадку необхідно використовувати значення χ^2.

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | R

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	3037415.	(2, 29)	0.0000
Chi-square	6074830.	2	0.0000

Null Hypothesis: C(1)+C(3)=4, C(2)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-4 + C(1) + C(3)	27.65704	1.362296
-1 + C(2)	-0.998161	0.000405

Restrictions are linear in coefficients.

Приклад

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$H_0 : \begin{cases} \beta_1 + \beta_2 + \beta_3 = 2, \\ \beta_1 - 3\beta_3 = 0. \end{cases}$$

- З гіпотези: $\beta_1 = 3\beta_3$ та $\beta_2 = 2 - 4\beta_3$
- Підставимо ці співвідношення до початкового рівняння:

$$y = \beta_0 + 3\beta_3 x_1 + (2 - 4\beta_3)x_2 + \beta_3 x_3 + \varepsilon$$

$$y - 2x_2 = \beta_0 + (3x_1 - 4x_2 + x_3)\beta_3 + \varepsilon$$

Приклад

- Оцінюємо початкову модель $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

та знаходимо URSS

- Оцінюємо нову модель $y - 2x_2 = \beta_0 + (3x_1 - 4x_2 + x_3)\beta_3 + \varepsilon$

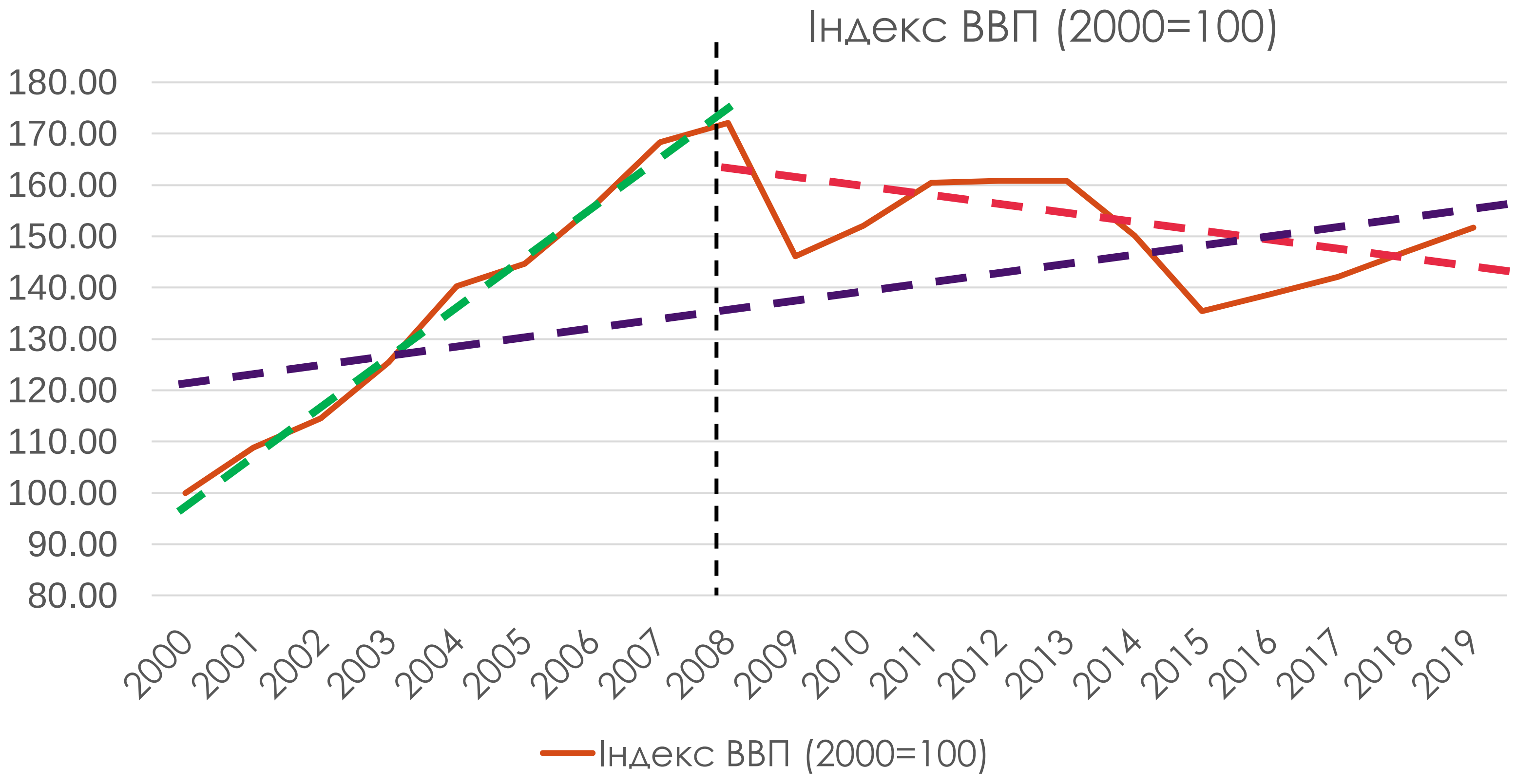
та знаходимо RRSS

$$F_{pr} = \frac{\frac{RRSS - URSS}{J}}{\frac{URSS}{n-k}}$$

$$F_{teor} = F\left(1-\alpha; J; n-k\right)$$

Якщо $F_{pr} < F_{teor}$, то гіпотеза про лінійні обмеження приймається

Перевірка гіпотез про стійкість моделі



Гіпотеза

- Модель буде називатися **стійкою**, якщо коефіцієнти моделей, що побудовані за різними вибірками, були статистично рівними:

$$H_0 : \forall \beta_j^{(I)} = \beta_j^{(II)} = \beta_j^{(III)}$$

Припущення

- Є n спостережень, які розбито на дві групи з n_1 та n_2 спостережень відповідно ($n_1+n_2=n$) .
- *RSS* – сума квадратів залишків у моделі, яка оцінена за всіма спостереженнями,
- RSS_1 – сума квадратів залишків у моделі, яка оцінена за першими спостереженнями
- RSS_2 – сума квадратів залишків у моделі, яка оцінена за останніми спостереженнями.

Тест Чоу

$$H_0 : \forall \beta_j^{(I)} = \beta_j^{(II)} = \beta_j^{(III)}$$

$$F_{pr} = \frac{\frac{RSS - (RSS_1 + RSS_2)}{k}}{\frac{RSS_1 + RSS_2}{n-2k}} \sim F_{teor}\left(1-\alpha; k; n-2k\right)$$

- Якщо практичне значення менше теоретичного $F_{pr} < F_{teor}$, то гіпотеза про стійкість приймається.

Прогнозний тест Чоу

- Використовується, коли по n_2 спостереженнях неможливо оцінити регресію.

$$F_{pr} = \frac{\dfrac{RSS - RSS_1}{n_2}}{\dfrac{RSS_1}{n_1 - k}} \sim F_{teor}\left(1-\alpha; n_2, n_1 - k\right)$$

- Якщо практичне значення менше теоретичного $F_{pr}<F_{teor}$, то гіпотеза про стійкість приймається.

Приклад

Equation: EQ01 Workfile: CHICKEN::Chicken\

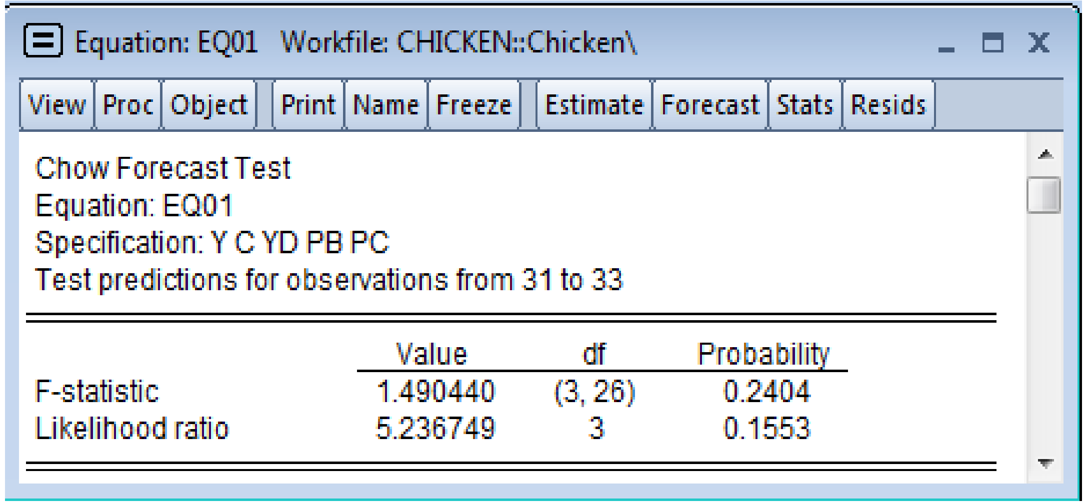
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Chow Breakpoint Test: 17
Null Hypothesis: No breaks at specified breakpoints
Varying regressors: All equation variables
Equation Sample: 1 33

F-statistic	27.51585	Prob. F(4,25)	0.0000
Log likelihood ratio	55.66666	Prob. Chi-Square(4)	0.0000
Wald Statistic	110.0634	Prob. Chi-Square(4)	0.0000

- Оскільки Probability < 0,05 відхиляється гіпотеза про відсутність структурних змін, що відбулися в 17-му спостереженні. Таким чином, модель є нестійкою, а 17-й період – є точкою перелому.
- У вибірці може бути не одна точка перелому.

Приклад


Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Chow Forecast Test
Equation: EQ01
Specification: Y C YD PB PC
Test predictions for observations from 31 to 33

	Value	df	Probability
F-statistic	1.490440	(3, 26)	0.2404
Likelihood ratio	5.236749	3	0.1553

- Оскільки Probability > 0,05 відхиляється альтернативна гіпотеза про структурні зміни, що відбулися в 31-ому спостереженні. Таким чином, модель є стійкою.

Перевірка гіпотези про нормальність збурень

- Для будь-якого розподілу величини можна розрахувати:
  - стандартне відхилення

$$s = \sqrt{\frac{\sum_{t=1}^{n}\left(y_t - \bar{y}\right)^2}{n-1}}$$

  - зміщену оцінку стандартного відхилення: $\hat{\sigma} = s\sqrt{\frac{n-1}{n}}$
  - коефіцієнт асиметрії (третій момент): $S = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t - \bar{y}}{\hat{\sigma}}\right)^3$
  - коефіцієнт ексцесу $K = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t - \bar{y}}{\hat{\sigma}}\right)^4$

Тест Харке-Бера

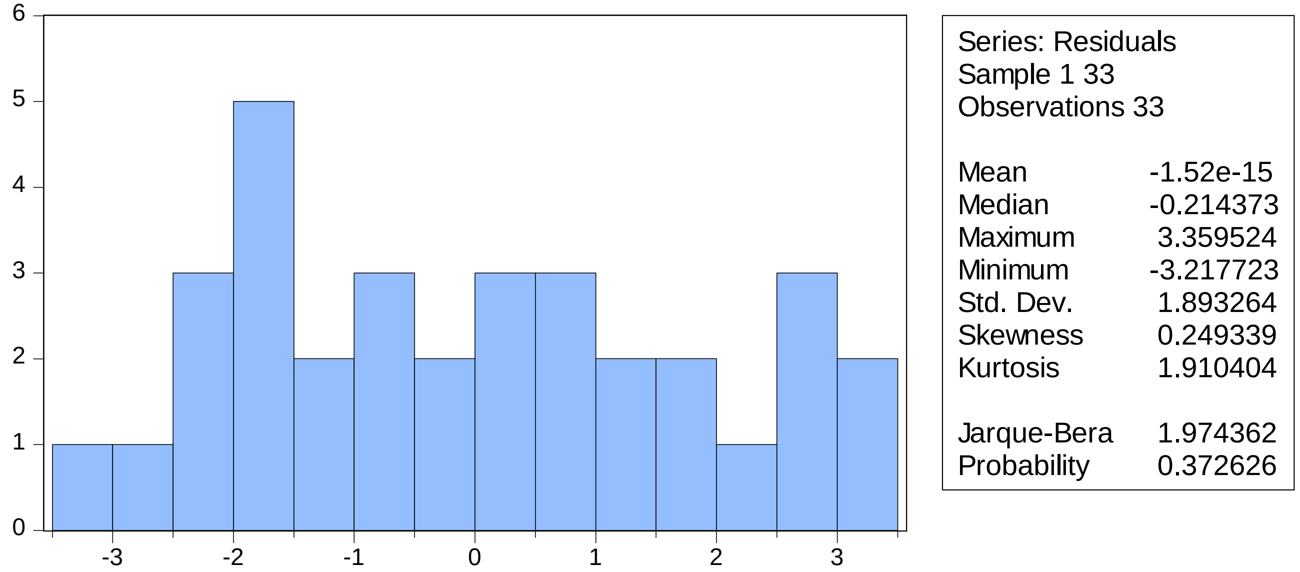
$$H_0 : S = 0, K = 3$$
$$H_1 : S \neq 0, K \neq 3$$

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \sim \chi^2(1-\alpha; 2)$$

- Якщо практичне значення *JB* не перевищую табличне значення χ^2, то приймається гіпотеза про нормальність розподілу.

Приклад

- *JB*=1,97. Табличне значення χ^2 з 2 степенями свободи та рівнем надійності 0,95 дорівнює 5,99, таким чином JB< $\chi^2(2)$, а значить гіпотеза приймається, тобто залишки мають нормальний розподіл.



Перевірка гіпотези про наявність мультиколінеарності

- **Мультиколінеарність** - наявність лінійної залежності між двома або більше факторними змінними у регресійній моделі.

Приклад

$$C_t = \beta_0 + \beta_1 S_t + \beta_2 N_t + \beta_3 T_t + \varepsilon_t$$

- де C – споживання,
- S – зарплата,
- N – дохід, отриманий поза роботою,
- T – повний дохід.

Оскільки з економічної точки зору виконується рівність T=S+N, то вихідну регресію можна переписати у вигляді

$$C_t = \beta_0 + \left(\beta_1 + \beta_3\right) S_t + \left(\beta_2 + \beta_3\right) N_t + \varepsilon_t$$

Причини мультиколінеарності

- На певні показники впливають однакові фактори. Це приводить до того, що вони відображають широкий спектр моделей однакової економічної ситуації.
- Широке використання в економетричних моделях лагових значень однієї змінної також призводить до виникнення мультиколінеарності. Наприклад, у функціях споживання витрати на споживання у попередньому періоді вводяться в модель поряд з величиною поточного рівня доходу.
- Неправильна специфікація моделі.

Ознаки мультиколінеарності

- Невелика зміна початкових даних (наприклад, додавання нових спостережень) призводить до істотної зміни оцінок коефіцієнтів моделі.
- Оцінки мають великі стандартні похибки, малу значимість, у той час як модель у цілому є значимою (високе значення коефіцієнта детермінації та відповідної *F*-статистики).
- Оцінки коефіцієнтів мають неправильні з погляду теорії знаки або незрозуміло великі значення.

Лікування мультиколінеарності

- Відкидання фіктивних змінних, що можуть створювати ідеальну мультиколінеарність
- Збільшення розміру вибірки.
- Відкидання певних змінних.
- Використання стандартизованих змінних.
- Використання гребневої регресії.

VIF-тест

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots \beta_{k-1} x_{k-1,t} + \varepsilon_t$$

- H_0: мультиколінеарність створює x_j

$$x_{j,t} = \gamma_1 x_{1,t} + \gamma_2 x_{2,t} + \dots + \gamma_{j-1} x_{j-1,t} + \gamma_{j+1} x_{j+1,t} + \dots + \gamma_{k-1} x_{k-1,t} + \varepsilon_t$$

$$VIF_j = \frac{1}{1 - R_j^2}$$

- Якщо $VIF_j > 5$, мультиколінеарність можлива.
- Якщо $VIF_j > 10$, мультиколінеарність сильна.

Приклад

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/13 Time: 19:28
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.40959	1.376227	22.82297	0.0000
YD	0.001839	0.000405	4.538711	0.0001
PB	0.247457	0.070428	3.513599	0.0015
PC	-0.819809	0.089305	-9.179902	0.0000

R-squared	0.963632	Mean dependent var	35.87879
Adjusted R-squared	0.959870	S.D. dependent var	9.927763
S.E. of regression	1.988782	Akaike info criterion	4.326134
Sum squared resid	114.7024	Schwarz criterion	4.507529
Log likelihood	-67.38122	Hannan-Quinn criter.	4.387168
F-statistic	256.1347	Durbin-Watson stat	0.753680
Prob(F-statistic)	0.000000		

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resid

Variance Inflation Factors
Date: 10/09/13 Time: 19:33
Sample: 1 33
Included observations: 33

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	1.894001	15.80228	NA
YD	1.64E-07	29.98550	9.301885
PB	0.004960	47.42433	8.842429
PC	0.007975	11.56217	1.195375

Два значення VIF перевищують 5, що свідчить про ознаки мультиколінеарності, зокрема про те, що одна зі змінних YD чи PB може бути виключеною з регресії.

Специфікація моделі

- **Специфікація моделі** – відбір потрібних та значимих для моделі факторів.

$$H_0: Y = X\beta + \varepsilon,$$
$$H_1: Y = X\beta + Z\gamma + \varepsilon.$$

- Незміщені МНК-оцінки коефіцієнтів регресії.
- Неефективні МНК-оцінки.

$$H_0: Y = X\beta + Z\gamma + \varepsilon,$$
$$H_1: Y = X\beta + \varepsilon.$$

- Зміщені МНК-оцінки коефіцієнтів регресії.
- Оцінка дисперсії у короткій регресії має невід’ємний зсув.

Тест на специфікацію

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1,n}$$

- H_0: вірна специфікація моделі
- H_1: треба додати змінну x_k

$$F_{pr} = \frac{\dfrac{R_2^2 - R_1^2}{1}}{\dfrac{1 - R_2^2}{n-k}} \sim F\left(1-a;1;n-k\right)$$

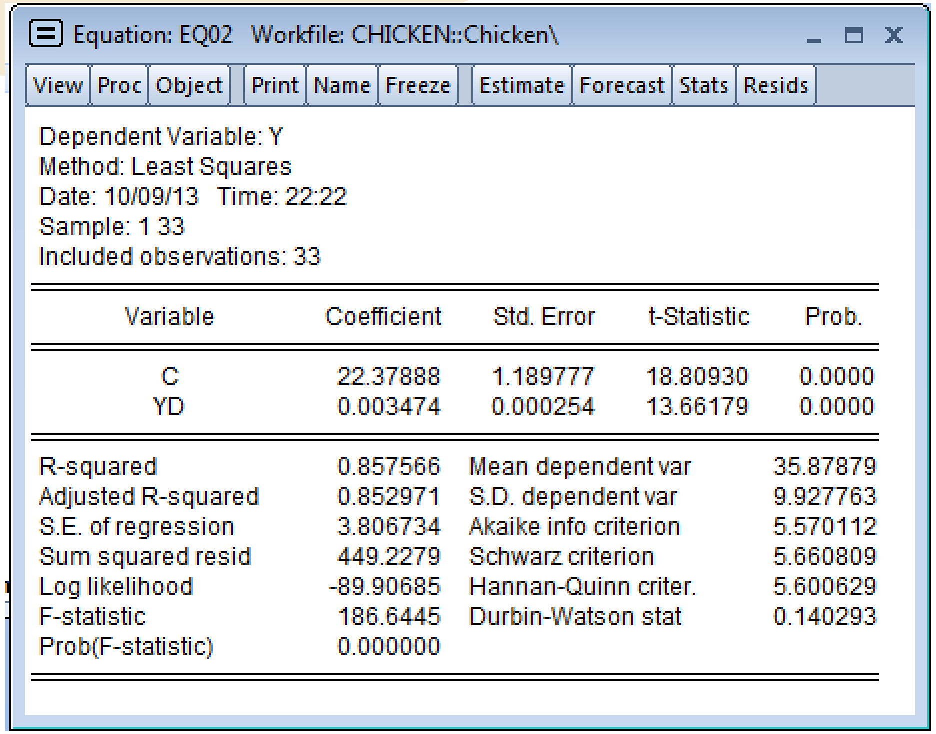
Перевірка гіпотези про пропущені змінні

$$H_0 : Y = X\beta + \varepsilon,$$
$$H_1 : Y = X\beta + Z\gamma + \varepsilon.$$

$$LR = -2\left(l_r - l_u\right) \sim \chi^2\left(1-\alpha; m\right)$$

- l_r, l_u - максимальні значення логарифмів функції правдоподібності для «короткої» та «довгої» регресії відповідно;
- m - кількість стовпчиків матриці Z.

Приклад



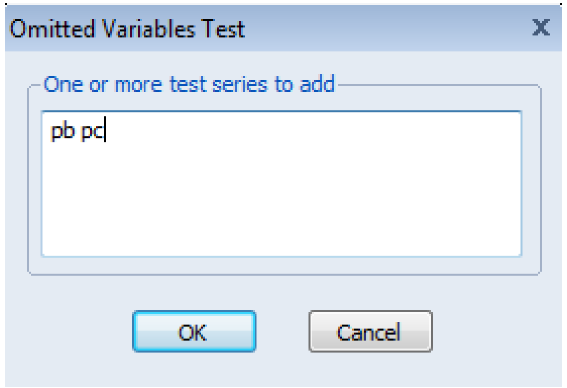
Equation: EQ02 Workfile: CHICKEN::Chicken\

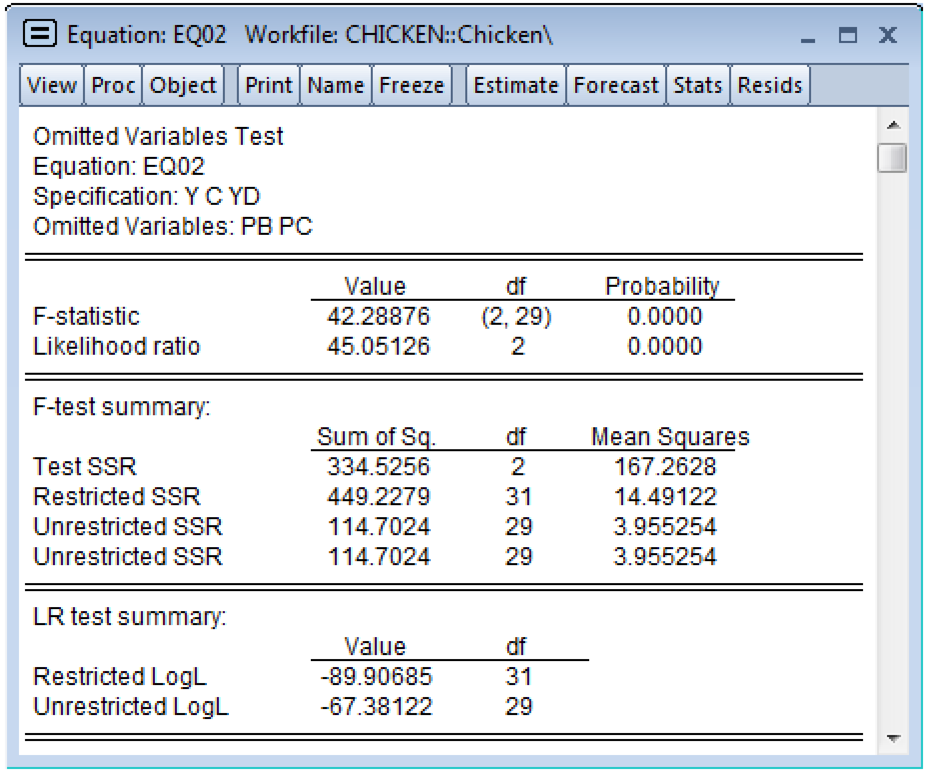
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/13 Time: 22:22
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.37888	1.189777	18.80930	0.0000
YD	0.003474	0.000254	13.66179	0.0000

R-squared	0.857566	Mean dependent var	35.87879
Adjusted R-squared	0.852971	S.D. dependent var	9.927763
S.E. of regression	3.806734	Akaike info criterion	5.570112
Sum squared resid	449.2279	Schwarz criterion	5.660809
Log likelihood	-89.90685	Hannan-Quinn criter.	5.600629
F-statistic	186.6445	Durbin-Watson stat	0.140293
Prob(F-statistic)	0.000000		





Equation: EQ02 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Omitted Variables Test
Equation: EQ02
Specification: Y C YD
Omitted Variables: PB PC

	Value	df	Probability
F-statistic	42.28876	(2, 29)	0.0000
Likelihood ratio	45.05126	2	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	334.5256	2	167.2628
Restricted SSR	449.2279	31	14.49122
Unrestricted SSR	114.7024	29	3.955254
Unrestricted SSR	114.7024	29	3.955254

LR test summary:

	Value	df
Restricted LogL	-89.90685	31
Unrestricted LogL	-67.38122	29

- Величини Probability для F-статистики та LR–критерію є меншими рівня похибки (0,05), а тому гіпотеза відхиляється, тобто модель має бути розширена за рахунок двох змінних.

Перевірка гіпотези про зайві змінні

$$H_0 : Y = X\beta + Z\gamma + \varepsilon,$$
$$H_1 : Y = X\beta + \varepsilon.$$

$$LR = -2\left(l_r - l_u\right) \sim \chi^2\left(1-\alpha; m\right)$$

- l_r, l_u - максимальні значення логарифмів функції правдоподібності для «короткої» та «довгої» регресії відповідно;
- m - кількість стовпчиків матриці Z.

Приклад

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/13 Time: 21:37
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.40959	1.376227	22.82297	0.0000
YD	0.001839	0.000405	4.538711	0.0001
PB	0.247457	0.070428	3.513599	0.0015
PC	-0.819809	0.089305	-9.179902	0.0000

R-squared	0.963632	Mean dependent var	35.87879
Adjusted R-squared	0.959870	S.D. dependent var	9.927763
S.E. of regression	1.988782	Akaike info criterion	4.326134
Sum squared resid	114.7024	Schwarz criterion	4.507529
Log likelihood	-67.38122	Hannan-Quinn criter.	4.387168
F-statistic	256.1347	Durbin-Watson stat	0.753680
Prob(F-statistic)	0.000000		

Redundant Variables Test

One or more test series to remove

pb pc

OK Cancel

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Redundant Variables Test
Equation: EQ01
Specification: Y C YD PB PC
Redundant Variables: PB PC

	Value	df	Probability
F-statistic	42.28876	(2, 29)	0.0000
Likelihood ratio	45.05126	2	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	334.5256	2	167.2628
Restricted SSR	449.2279	31	14.49122
Unrestricted SSR	114.7024	29	3.955254
Unrestricted SSR	114.7024	29	3.955254

LR test summary:

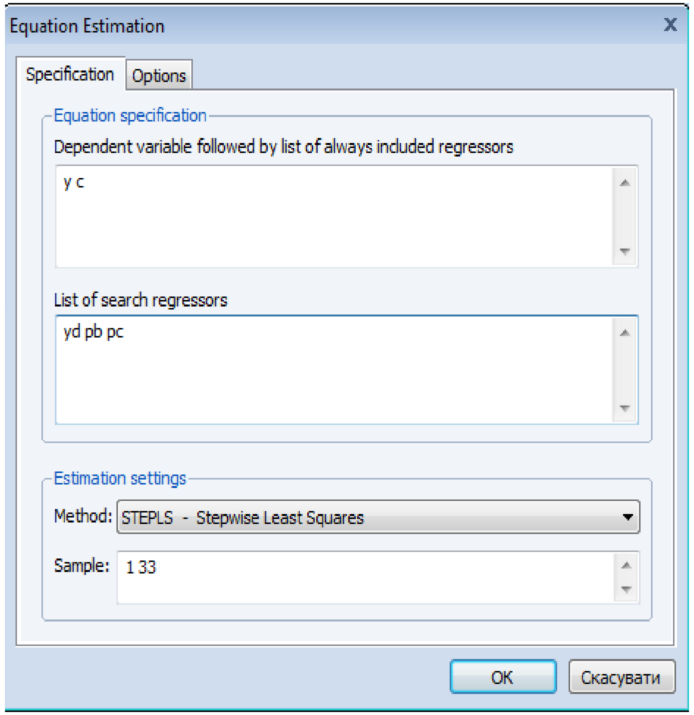
	Value	df
Restricted LogL	-89.90685	31
Unrestricted LogL	-67.38122	29

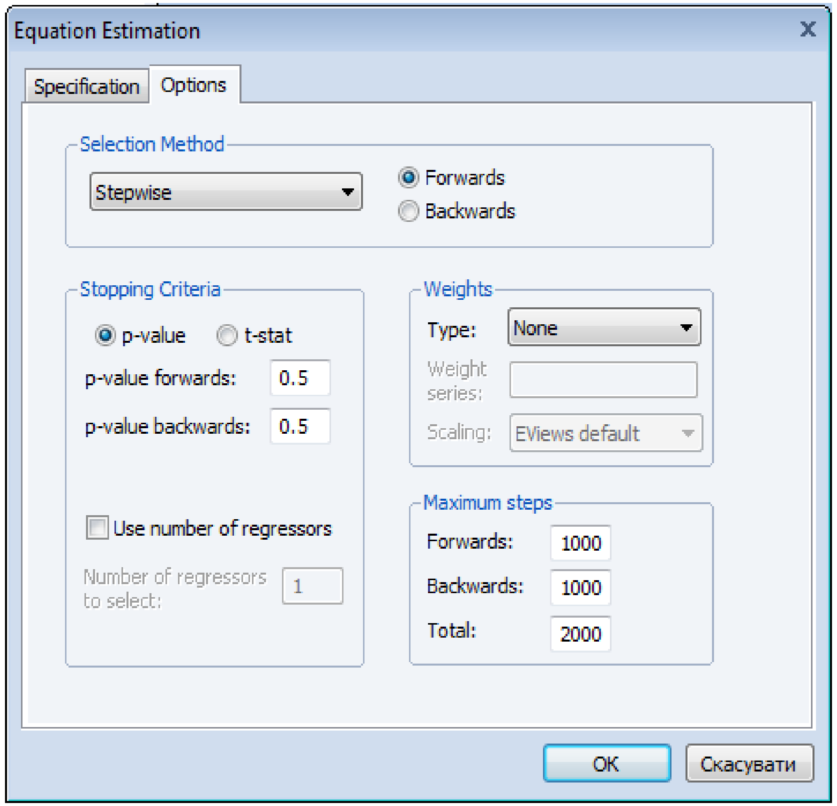
- Оскільки значення Probability для F-статистики та LR–критерію є меншими рівня похибки (0,05), то гіпотеза відхиляється, тобто модель не слід скорочувати за рахунок двох наведених змінних.

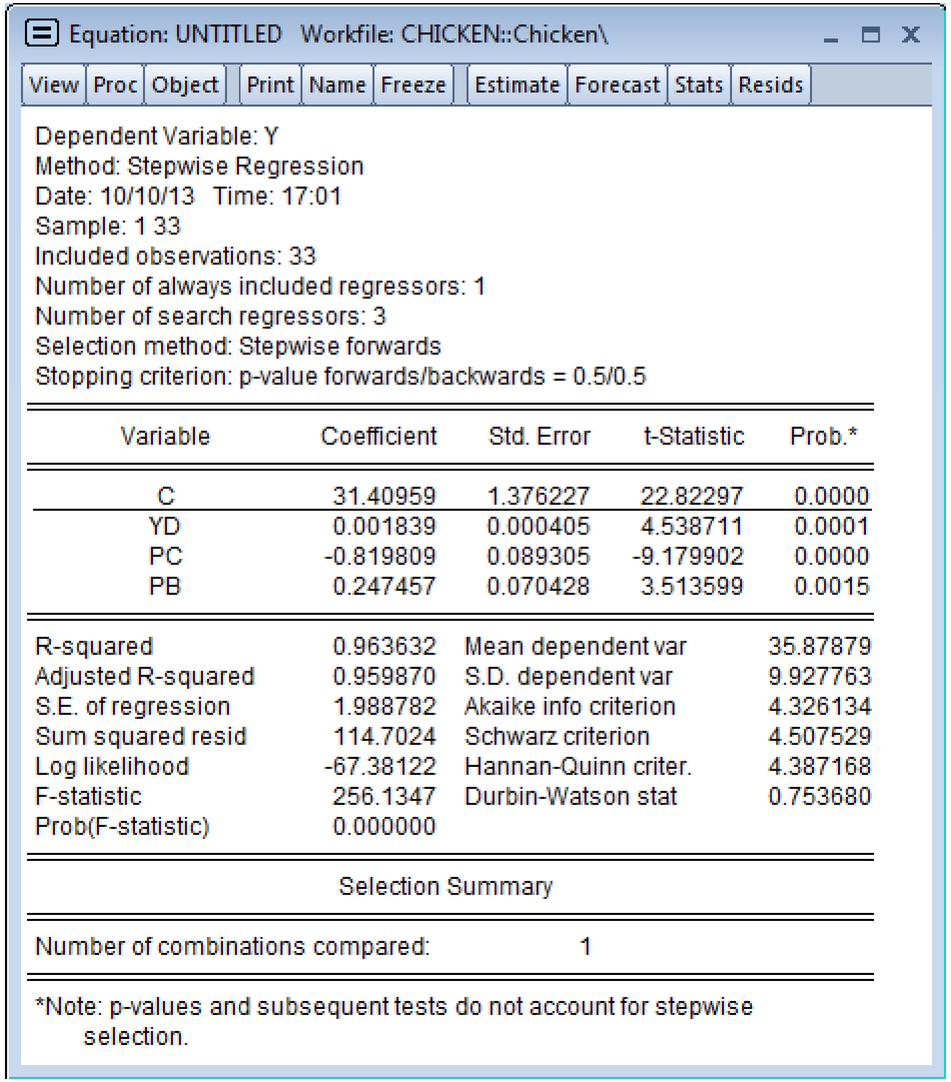
Процедура покрокового відбору змінних

- Додавання змінних. Спочатку розглядається модель без змінних з однією константою. На кожному кроці програма намагається додати змінну, яка покращить модель. Цей процес повторюється, доки жодне додавання змінної не покращить модель.
- Виключення змінних. Спочатку розглядається модель, до якої включено всі можливі змінні. На кожному кроці здійснюється виключення однієї змінної, що не погіршує статистичні властивості моделі.
- Одночасне включення та виключення змінних. На кожному кроці програма намагається одночасно додати та виключити певні змінні для покращення статистичних властивостей моделі.

Приклад







Dependent Variable: Y
Method: Stepwise Regression
Date: 10/10/13 Time: 17:01
Sample: 1 33
Included observations: 33
Number of always included regressors: 1
Number of search regressors: 3
Selection method: Stepwise forwards
Stopping criterion: p-value forwards/backwards = 0.5/0.5

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
C	31.40959	1.376227	22.82297	0.0000
YD	0.001839	0.000405	4.538711	0.0001
PC	-0.819809	0.089305	-9.179902	0.0000
PB	0.247457	0.070428	3.513599	0.0015

R-squared	0.963632	Mean dependent var	35.87879
Adjusted R-squared	0.959870	S.D. dependent var	9.927763
S.E. of regression	1.988782	Akaike info criterion	4.326134
Sum squared resid	114.7024	Schwarz criterion	4.507529
Log likelihood	-67.38122	Hannan-Quinn criter.	4.387168
F-statistic	256.1347	Durbin-Watson stat	0.753680
Prob(F-statistic)	0.000000		

Selection Summary

Number of combinations compared: 1

*Note: p-values and subsequent tests do not account for stepwise selection.

Перевірка функціональної форми моделі

- H_0: зв'язок лінійний
- H_1: зв'язок нелінійний

$$y_t = X\beta + \alpha_2 \hat{y}_t^2 + \alpha_3 \hat{y}_t^3 + \ldots + \alpha_q \hat{y}_t^q + \varepsilon_t$$

$$H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_q = 0$$

Приклад

- Оскільки значення Probability<0,05, то гіпотеза H_0 відхиляється, а значить, лінійна форма для моделі підібрана невірно.

Equation: EQ01 Workfile: CHICKEN::Chicken\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Ramsey RESET Test
Equation: EQ01
Specification: Y C YD PB PC
Omitted Variables: Powers of fitted values from 2 to 3

	Value	df	Probability
F-statistic	17.70349	(2, 27)	0.0000
Likelihood ratio	27.64873	2	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	65.07708	2	32.53854
Restricted SSR	114.7024	29	3.955254
Unrestricted SSR	49.62528	27	1.837973
Unrestricted SSR	49.62528	27	1.837973

LR test summary:

	Value	df
Restricted LogL	-67.38122	29
Unrestricted LogL	-53.55685	27

Питання?