Семінари 8-9. Різні аспекти множинної регресії

В цьому розділі ми розглянемо різні види регресій, а також таких, що можуть бути зведені до стандартної лінійної регресії:

1. Визначення сезонних коливань.
2. Функція Кобба-Дугласа.
3. Порівняння факторів за ступенем їхнього впливу
4. Перевірка наявності мультиколінеарності
5. Перевірка гіпотези про пропущені змінні
6. Перевірка гіпотези про зайві змінні
7. Процедура покрокового відбору змінних
8. Перевірка функціональної форми моделі

Визначення сезонних коливань

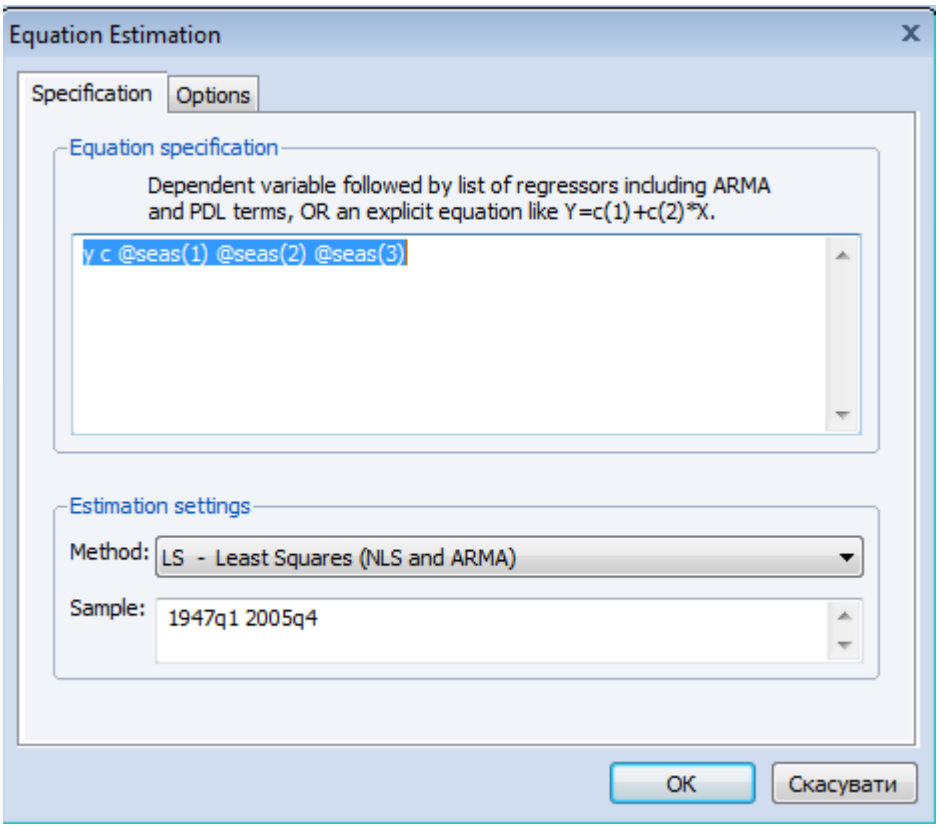
Нехай потрібно проаналізувати часовий ряд y_t на наявність сезонних коливань.

Одним з розповсюджених способів моделювання сезонності є використання фіктивних змінних. Для побудови регресійної моделі слід застосувати функцію *@seas(N).* Як і всі функції системи її запис починається зі знаку @. Ця функція має параметр *N* – номер сезону у році. Наприклад, для оцінки регресії вигляду

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \varepsilon_t$$

у вікні специфікації регресії вказується[1]:

[1] Файл macromod.wf1



В даному випадку в моделі використано 3 фіктивні змінні, оскільки періодичність даних була квартальною.

У випадку місячної структури даних необхідно було б використати 11 фіктивних змінних.

Розглянемо наступні моделі:

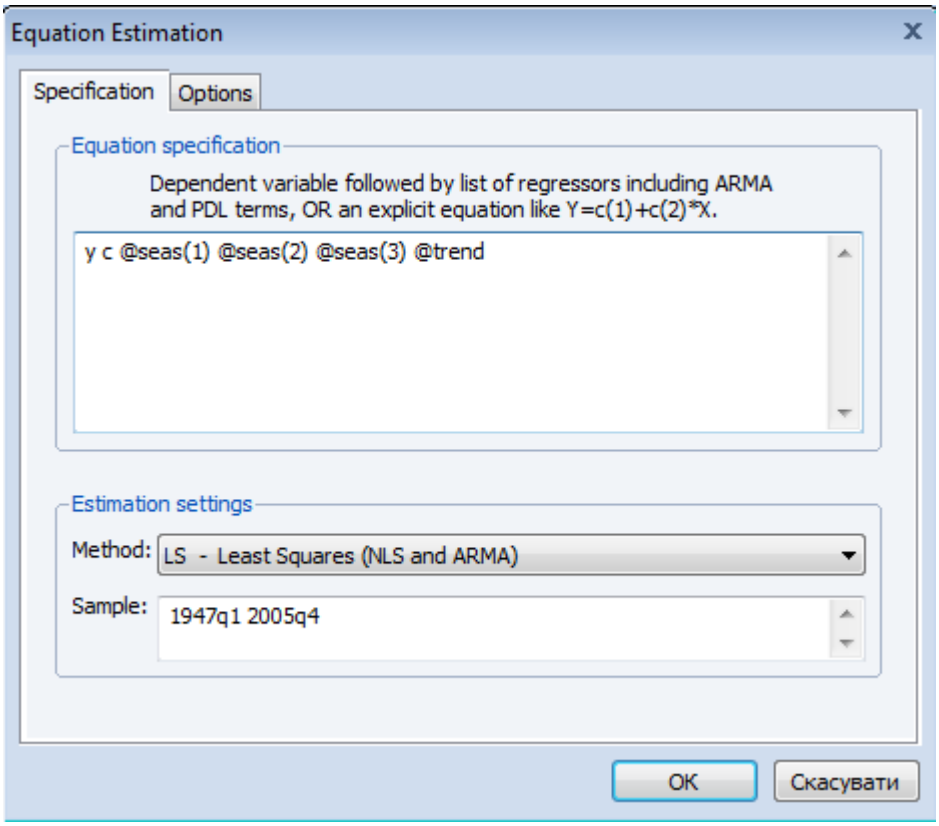
- Сезонна модель без тренду:

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \varepsilon_t$$

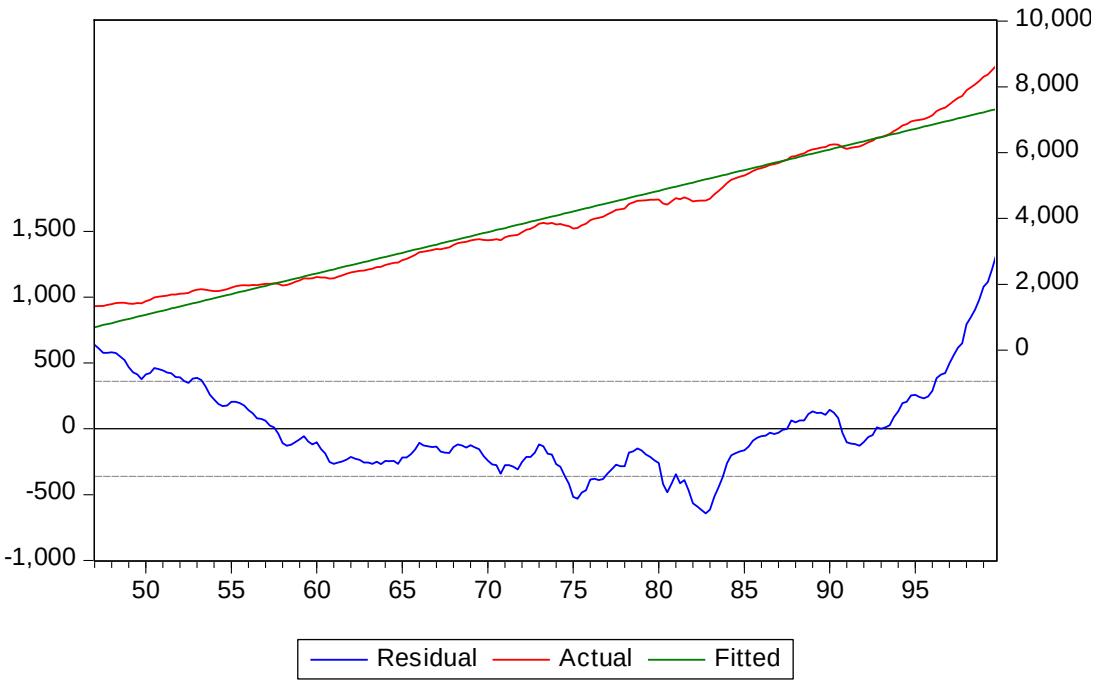
- Сезонна модель з трендом:

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \beta_4 t + \varepsilon_t$$

Для першої моделі специфікація моделі аналогічна попередній. Для другої моделі у вікні специфікації запишемо:



Для ілюстрації побудуємо графік останньої регресії:



Функція Кобба-Дугласа[2]

$$Y = \beta_0 K^{\beta_1} L^{\beta_2} + \varepsilon_t ,$$

[2] Файл cobb_duglas.wf1

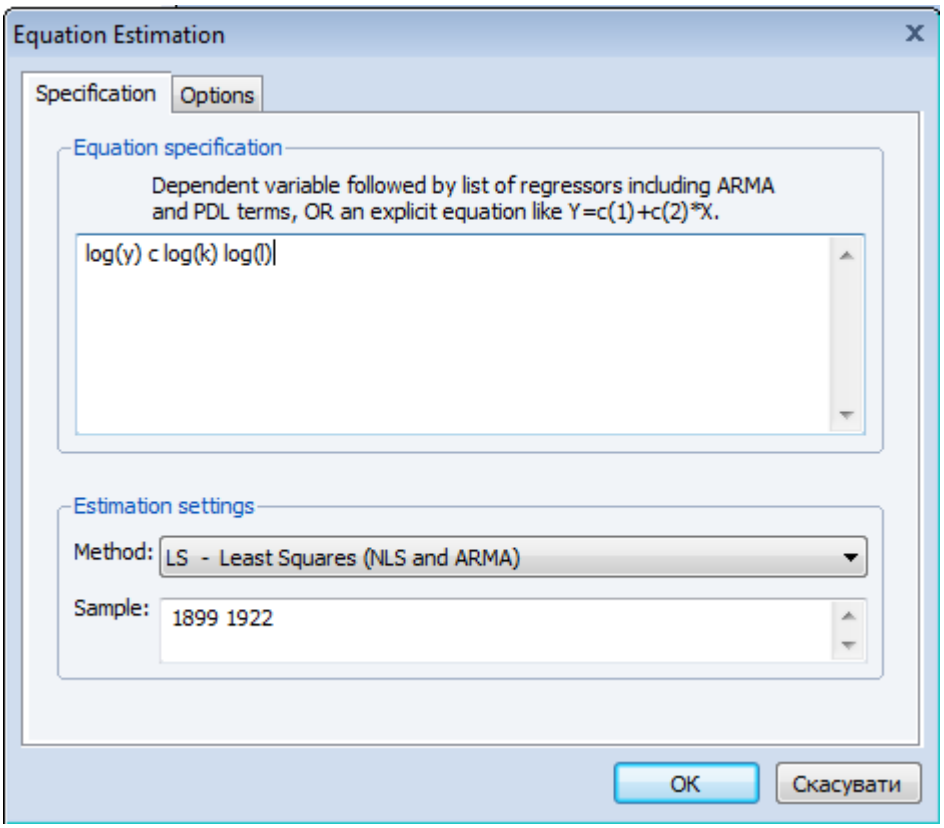
де Y - індекс реального обсягу виробництва

K - індекс реальних капітальних витрат

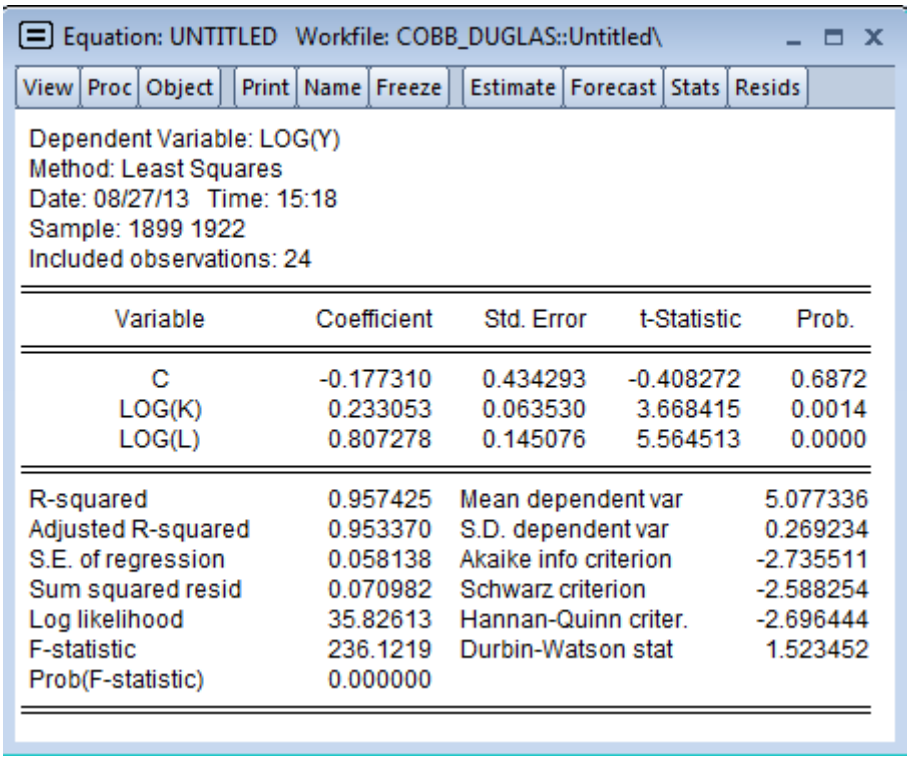
L - індекс реальних витрат праці

Для того, щоб звести модель до лінійної регресії, необхідно прологарифмувати задану функцію.

Для цього у вікні специфікації записуємо:



Оцінимо модель:



Equation: UNTITLED Workfile: COBB_DUGLAS::Untitled\

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

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 08/27/13 Time: 15:18
Sample: 1899 1922
Included observations: 24

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.177310	0.434293	-0.408272	0.6872
LOG(K)	0.233053	0.063530	3.668415	0.0014
LOG(L)	0.807278	0.145076	5.564513	0.0000

R-squared	0.957425	Mean dependent var	5.077336
Adjusted R-squared	0.953370	S.D. dependent var	0.269234
S.E. of regression	0.058138	Akaike info criterion	-2.735511
Sum squared resid	0.070982	Schwarz criterion	-2.588254
Log likelihood	35.82613	Hannan-Quinn criter.	-2.696444
F-statistic	236.1219	Durbin-Watson stat	1.523452
Prob(F-statistic)	0.000000		

Тепер знаходимо кінцеву відповідь початкової моделі:

$$\beta_0 = e^{-0.177310}$$
$$\beta_1 = 0.233053$$
$$\beta_2 = 0.807278$$

Для підрахунку першого значення в робочому файлі можна потрібно записати:

```
series b0=exp(-0.177310)
```

Порівняння факторів за ступенем їх впливу

Розглянемо множинну регресію, для якої вже отримані статистично значимі оцінки коефіцієнти регресії. В такому разі вибіркова регресійна функція може бути записана у вигляді:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \ldots + \hat{\beta}_{k-1} x_{k-1,t} \qquad (1)$$

Регресійні коефіцієнти не можна використовувати для порівняння дії різних факторів, тому найчастіше при цьому використовують два методи:

- порівняння коефіцієнтів в регресії відносно нормалізованих змінних;
- порівняння коефіцієнтів еластичності.

Зазначимо, що для порівняння не існує критерію, придатного у всіх ситуаціях. При виборі критерію треба враховувати мету дослідження, використовувати знання з тієї галузі економічної теорії, яка вивчає досліджуваний об'єкт.

Регресія відносно нормалізованих змінних

Основна ідея цього методу – позбавлення змінних від різних одиниць виміру. Розглянемо застосування цього методу на прикладі. Нехай потрібно оцінити модель лінійної регресії:

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

Уведемо наступні позначення:

$\overline{y} = \frac{\sum_{t=1}^{n} y_t}{n}$ – середнє значення залежної змінної ,

$\overline{x}_j = \frac{\sum_{t=1}^{n} x_{tj}}{n}$, $j = \overline{1,k-1}$ – середнє значення j-ї незалежної змінної ,

$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} (y_t - \overline{y})^2}{n-1}}$ – середньоквадратичне відхилення залежної змінної ,

$\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^{n} (x_{tj} - \overline{x}_j)^2}{n-1}}$, $j = \overline{1,k-1}$ – середньоквадратичне відхилення j-ї незалежної змінної ,

$y_t^* = \frac{y_t - \overline{y}}{\sigma_y}$, $i = \overline{1,n}$ – значення стандартизованої залежної змінної в t -му спостереженні

$x_{tj}^* = \frac{x_{tj} - \overline{x}_j}{\sigma_{x_j}}$, $t = \overline{1,n}$, $j = \overline{1,k-1}$ – значення стандартизованої j-ї незалежної змінної в t-му спостереженні.

Розрахунок величин y_t^* та x_{tj}^* називається нормалізацією змінних, що пов'язано з тим, що нові змінні асимптотично повинні мати стандартний нормальний розподіл.

Також варто відмітити, що середнє значення всіх нормалізованих дорівнює нулю. З цього випливає, що модель регресії відносно нормалізованих змінних записується у такому вигляді:

$$y_t^* = \beta_1^* x_{t,1}^* + \ldots + \beta_{k-1}^* x_{t,k-1}^* + \varepsilon_t, t = \overline{1,n}. \qquad (2)$$

Як відомо, регресія завжди проходить через точку середніх значень залежної і незалежної змінних. Оскільки середні значення всіх нормалізованих змінних дорівнюють нулю, то модель не містить константи.

Оскільки середньоквадратичні відхилення мають ті самі розмірності, що і змінні, нормалізовані змінні є безрозмірними величинами, а тому коефіцієнти регресії (2) можна інтерпретувати як міру впливу незалежних змінних на залежну змінну.

Значення коефіцієнтів регресії (2) можна знайти без безпосереднього застосування методу найменших квадратів, скориставшись формулою:

$$\hat{\beta}_j^* = \frac{\hat{\beta}_j \sigma_{x_j}}{\sigma_y}, \; j = \overline{1,k-1}.$$

Після знаходження величин $\hat{\beta}_j^*$ всі фактори можна ранжирувати за абсолютною величиною відповідного коефіцієнта.

Коефіцієнти еластичності

Як і в інших дисциплінах, **коефіцієнт еластичності** показує на скільки відсотків зміниться значення залежної змінної *при* зростанні однієї незалежної змінної на 1 відсоток за умови, що значення всіх інших змінних не зміниться.

Для довільної залежності виду

$$y_t = f\left(x_{1,t}, x_{2,t}, \ldots, x_{k-1,t}\right)$$

коефіцієнт еластичності змінної y_t відносно x_j визначається як:

$$\varepsilon_j = \frac{\partial(\ln f(x_1, x_2, ..., x_{k-1})}{\partial(\ln x_j)} = \frac{\partial f}{\partial x_j} \frac{x_j}{f(x_1, x_2, ..., x_{k-1})}, j = \overline{1,k-1} \qquad (3)$$

Якщо вигляд залежності задано явно, наприклад, за допомогою регресії (1), то значення вибіркового коефіцієнта еластичності можна розрахувати за формулою:

$$\varepsilon_j = \hat{\beta}_j \frac{x_j}{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_{k-1} x_{k-1}}, \quad j = \overline{1, k-1} \qquad (4)$$

З формули випливає, що коефіцієнти еластичності залежать від того, при якому значенні змінної вони обчислюються. Стандартним є обчислення коефіцієнтів еластичності при середніх значеннях змінних. Тоді формула (4) приймає вид:

$$\varepsilon_j = \hat{\beta}_j \frac{\overline{x}_j}{\overline{y}}, \quad j = \overline{1, k-1} \qquad (5)$$

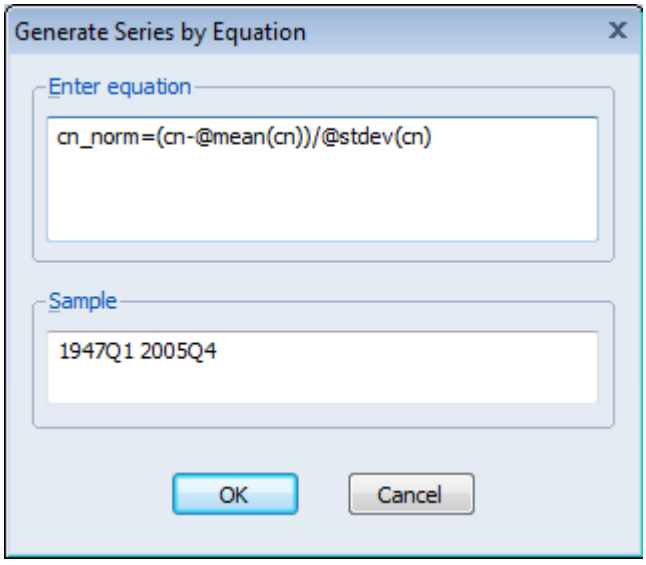
Для ранжування факторів регресії за ступенем їхнього впливу використовують абсолютне значення коефіцієнта еластичності.

Приклад

Нехай потрібно ранжирувати фактори за ступенем їхнього впливу у регресії індивідуального споживання від державних видатків та грошової маси[3]:

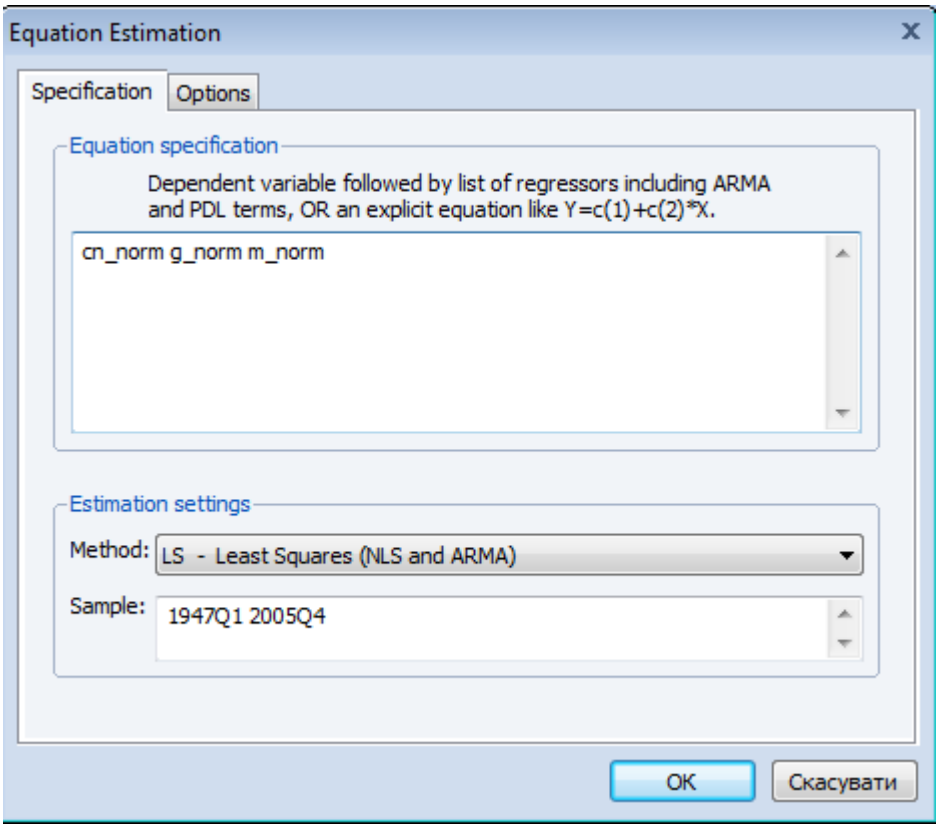
$$cn_t = \beta_0 + \beta_1 G_t + \beta_2 M_t + \varepsilon_t$$

Для створення нормалізованих змінних скористаємося наступною формулою:

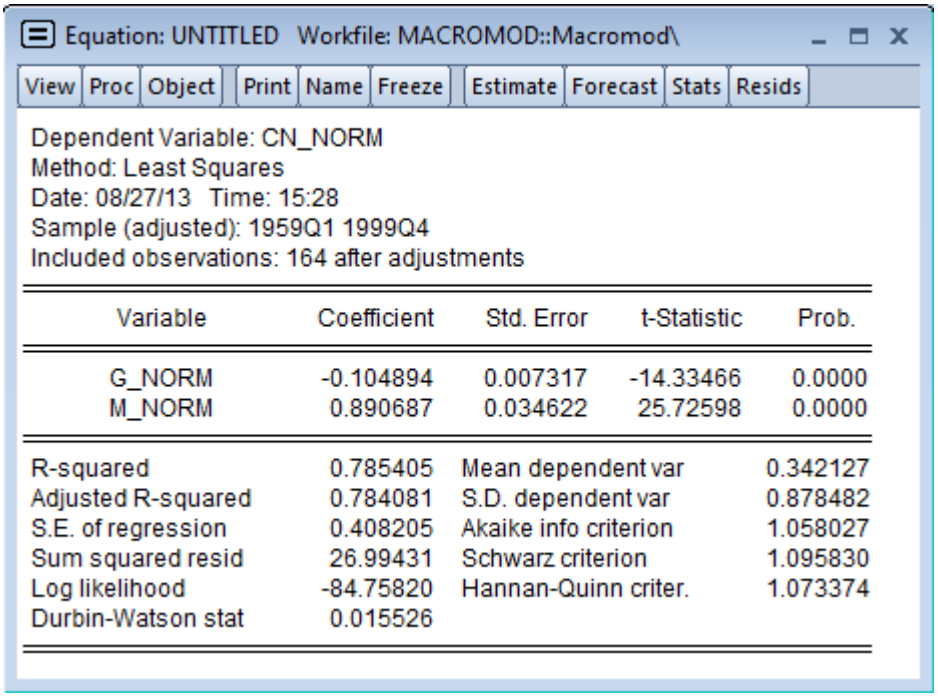


Поступивши аналогічно з усіма змінними, побудуємо регресію:

[3] Файл macromod.wf1

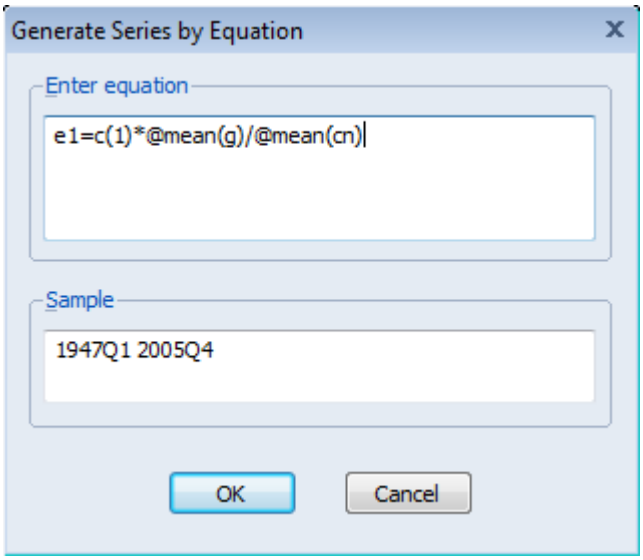


Зверніть увагу, що регресія будується без константи.

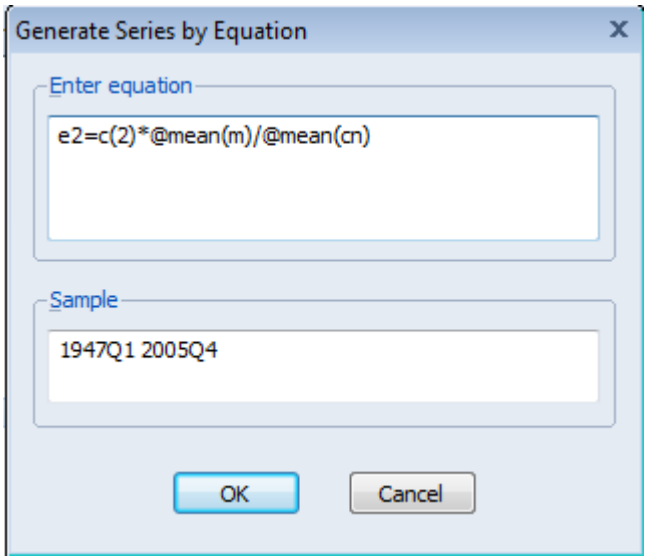


Оцінка моделі показує, що фактор державних видатків впливає негативно на споживання на рівні -0,10, а грошова маса – позитивно на рівні -0,89. За абсолютною величиною найбільш важливим є вплив саме вплив грошової маси.

Для розрахунку коефіцієнтів еластичності скористаємося наступною формулою:



Аналогічно розраховується показник еластичності для фактора грошової маси:



Відкривши значення змінних e1 та e2, бачимо, що коефіцієнти еластичності дорівнюють відповідно -0,03 та 0,27. Таким чином, найбільш впливовим є фактор грошової маси.

Перевірка наявності мультиколінеарності

Нехай розглядається модель виду:

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

Нехай є припущення про те, що найбільш домінантною, тобто найбільш впливовою на інші змінні є змінна x_j. В такому разі будується регресія цієї змінної від всіх інших змінних без константи:

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

Для цієї регресії знаходиться коефіцієнт детермінації R_j^2, на основі якого розраховується *VIF*-індекс:

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

Якщо значення VIF_j перевищує 5, це привід замислитися про наявність мультиколінеарності. В той же час деякі дослідники визначають наявність мультиколінеарності при $VIF_j > 10$.

Розрахуємо VIF-індекс для моделі:

Equation: EQ01 Workfile: CHICKEN::Chicken\

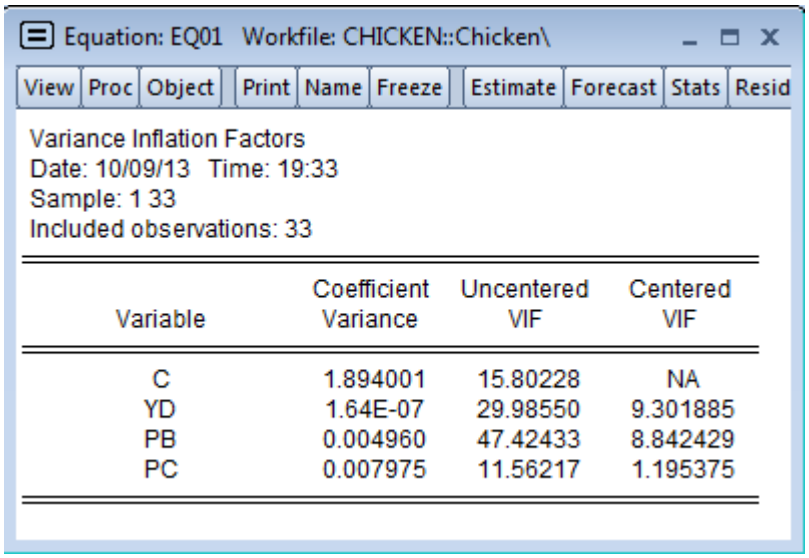
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

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		

У статистичному пакеті EViews для розрахунку цього індексу необхідно використати меню **View-Coefficient Diagnostics-Variance Inflation Factors**:


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

Необхідні для перевірки значення VIF_j знаходяться у стовпчику «Centered VIF". Як бачимо, всі величини у стовпчику не перевищують 10, а тому сильної мультиколінеарності немає. В той же час два значення VIF_j перевищують 5, що свідчить про ознаки мультиколінеарності, зокрема про те, що одна зі змінних *YD* чи *PB* може бути виключеною з регресії.

Перевірка гіпотези про пропущені змінні

Якщо потрібно перевірити за один раз декілька змінних, які слід включити або виключити з моделі, то можна скористатися узагальненням наведеного вище тесту.

Нехай реальна модель має вигляд $y = X\beta + Z\gamma + \varepsilon$, а дослідник оцінює модель $y = X\beta + \varepsilon$. Тоді потрібно перевірити, чи слід включити до регресії змінні, що містяться у матриці Z. Це можна зробити за допомогою тесту на пропущені змінні (Omitted Variables Test). За гіпотези H_0 всі коефіцієнти γ дорівнюють 0, сигналізуючи про недоцільність включення нових змінних до моделі.

Якщо дослідник вважає, що у моделі не виконується припущення про нормальність збурень, то неможна користуватися статистикою Фішера. Замість неї слід використовувати LR-критерій (likelihood ratio). Тоді

$$LR = -2\left(l_r - l_u\right),$$

де l_r та l_u – максимальні значення логарифмів функції правдоподібності для «короткої» та «довгої» регресії відповідно. Отримана величина має χ^2 – розподіл з m степенями свободи, де m – кількість стовпчиків матриці Z.

Якщо $LR < \chi^2(m)$, то гіпотеза H_0 приймається, тобто ніяких додаткових змінних до моделі включати не потрібно.

Оцінимо модель $y_t = \beta_0 + \beta_1 YD_t + \varepsilon_t$:

Equation: EQ02 Workfile: CHICKEN::Chicken\

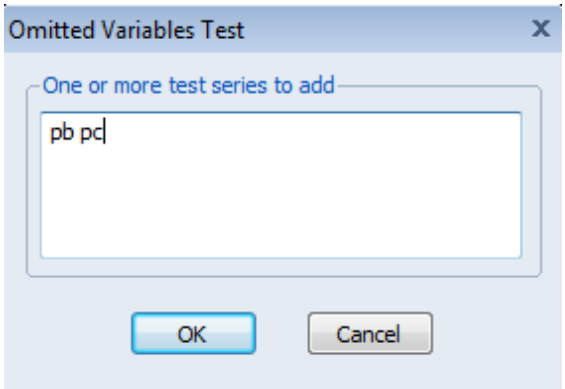
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

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

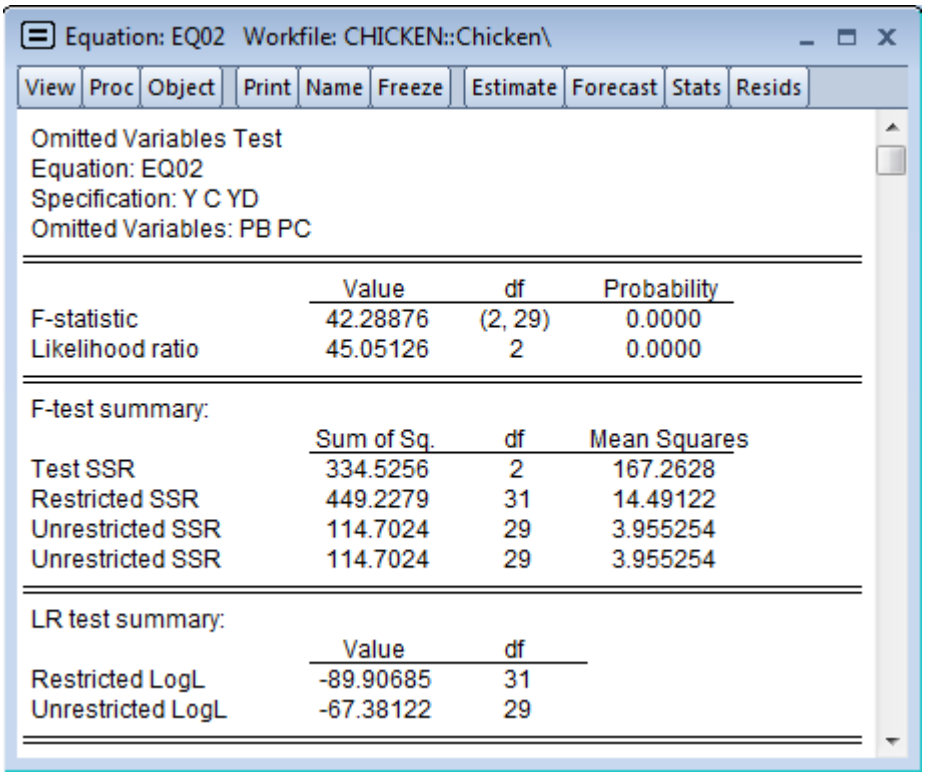
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.37888	1.189777	18.80930	0.0000
YD	0.003474	0.000254	13.66179	0.0000

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		

Тепер перевіримо, чи слід включити до моделі ще дві змінні *PB* та *PC*. Для цього скористаємося меню **View-Coefficient Diagnostics-Omitted Variables-Likelihood Ratio...** та введемо назви змінних, доцільність включення до моделі слід перевірити:



В результаті отримаємо розрахунки за гіпотезою:


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 H_0 відхиляється, тобто модель має бути розширена за рахунок двох змінних.

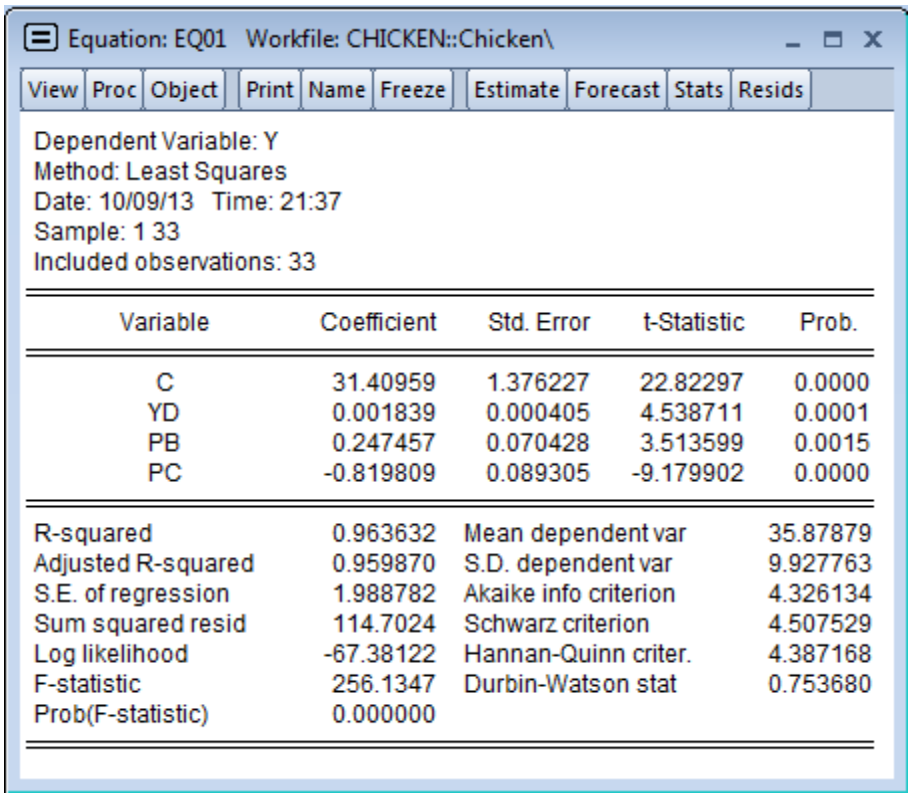
Перевірка гіпотези про зайві змінні

Аналогічно до попереднього тесту маємо перевірити, чи слід з певної регресії виключити деякі змінні. Для цього розроблений тест на зайві змінні (Redundant Variables Test).

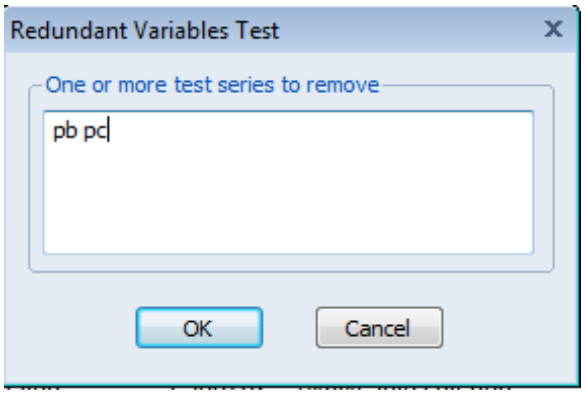
Нехай реальна модель має вигляд $y = X\beta + \varepsilon$, а дослідник оцінює модель $y = X\beta + Z\gamma + \varepsilon$. Постає питання, чи не слід виключити з моделі певну групу змінних, записаних у матриці Z. Таким чином, за гіпотези H_0 всі коефіцієнти γ дорівнюють 0, сигналізуючи про доцільність виключення змінних з моделі.

Очевидно, що цей тест є лише модифікацією попереднього тесту, а тому всі формули залишаються незмінними. Якщо залишки моделі розподілені нормально, то слід використовувати стандартну статистику Фішера, якщо ж залишки не мають нормального розподілу, то застосовується *LR*-критерій.

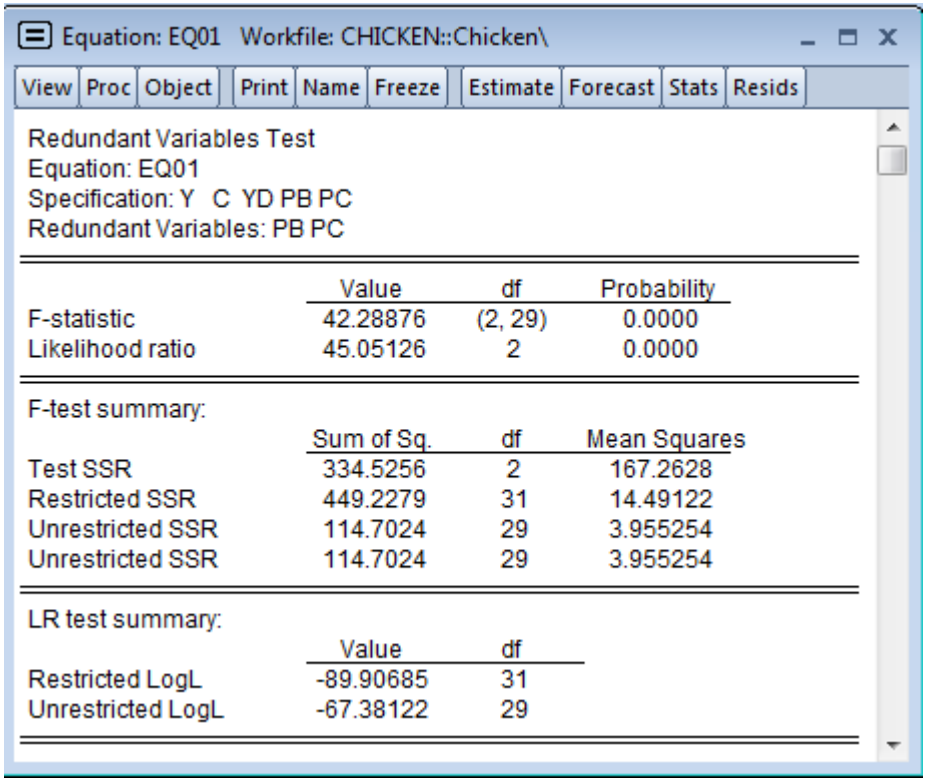
Нехай оцінена повна модель:



Для цієї моделі слід перевірити, чи доцільно залишати в ній змінні *PB* та *PC*. Скористаємося меню **View-Coefficient Diagnostics-Redundant Variables-Likelihood Ratio...** та введемо назви відповідних змінних:



В результаті отримаємо таке вікно з перевіркою гіпотези:


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 H_0 відхиляється, тобто модель не слід скорочувати за рахунок двох наведених змінних.

Процедура покрокового відбору змінних

Звичайно, при роботі з багатьма факторами досить важко вибрати саме необхідний набір змінних, адже потрібно проводити достатньо багато тестів щодо пропущених чи зайвих змінних. У цьому контексті може допомогти процедура покрокового відбору змінних (stepwise regression). Основна ідея цього методу полягає у переборі всіх можливих варіантів побудови регресії та виборі найкращого варіанту. В більшості випадків такий відбір здійснюється комп'ютером, а тому на виході дослідник отримує лише оптимальний вигляд регресії.

Розрізняють декілька алгоритмів для відбору змінних:

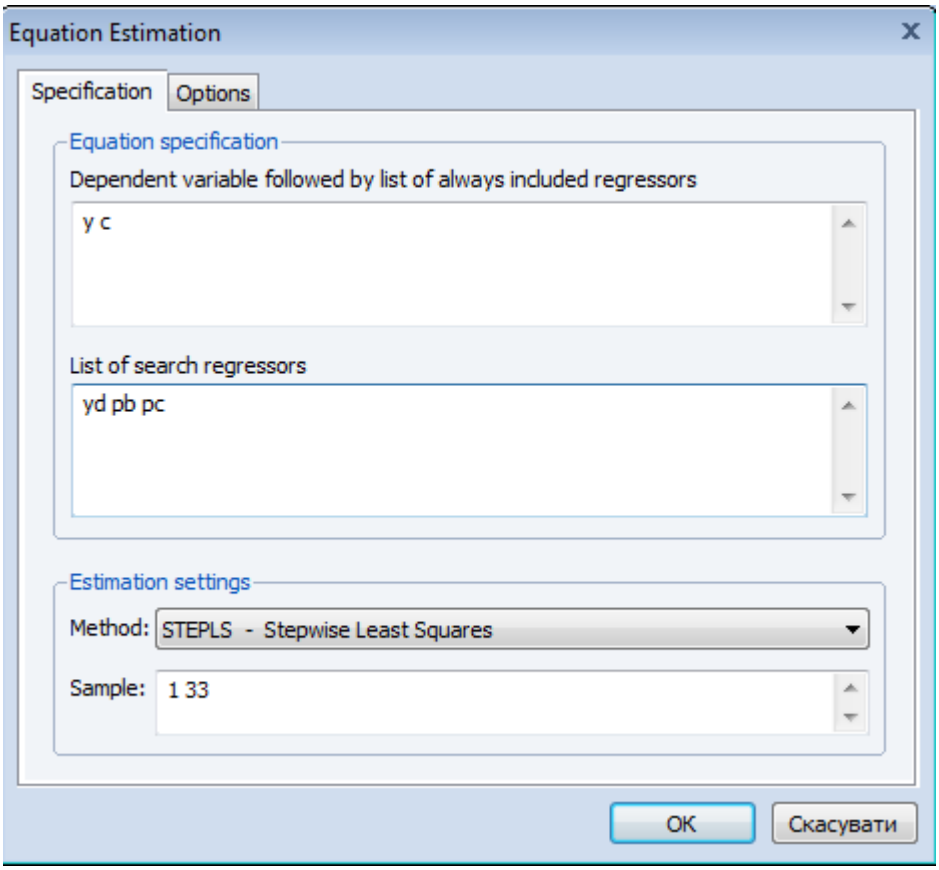
1) Додавання змінних. Спочатку розглядається модель без змінних з однією константою. На кожному кроці програма намагається додати змінну, яка покращить модель. Цей процес повторюється, доки жодне додавання змінної не покращить модель.

2) Виключення змінних. Спочатку розглядається модель, до якої включено всі можливі змінні. На кожному кроці здійснюється виключення однієї змінної, що не погіршує статистичні властивості моделі.

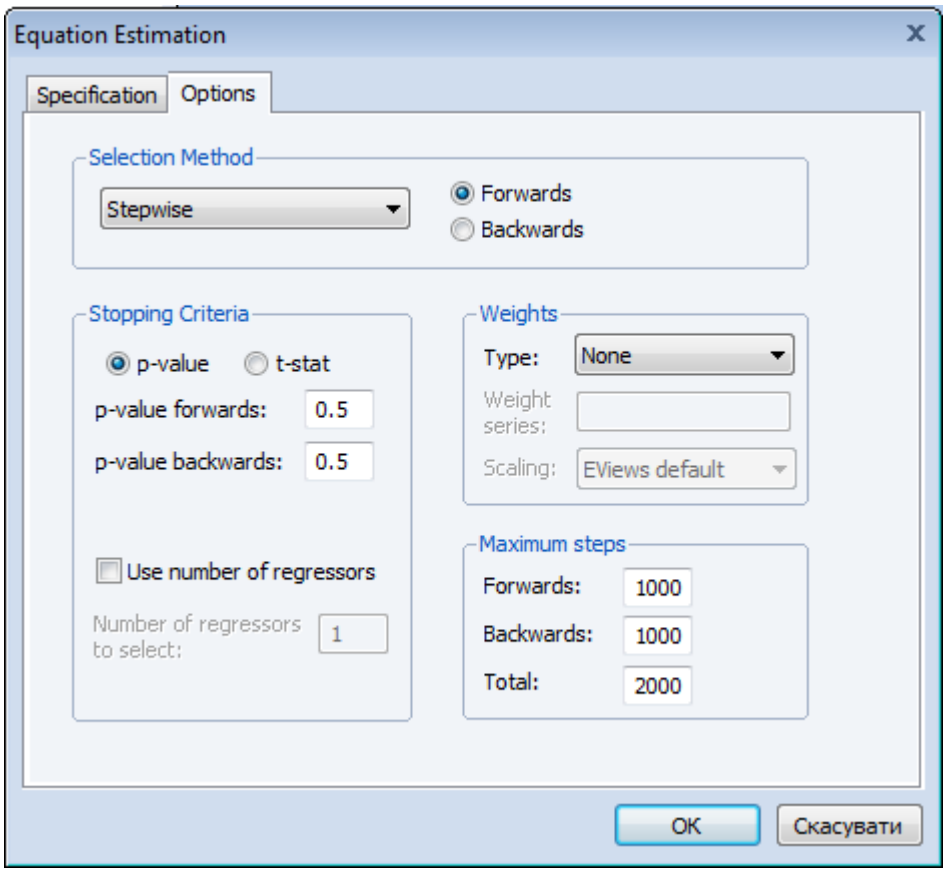
3) Одночасне включення та виключення змінних. На кожному кроці програма намагається одночасно додати та виключити певні змінні для покращення статистичних властивостей моделі.

В той же час слід зазначити, що деякі вчені критикую процедуру покрокового відбору змінних через низькі асимптотичні якості та не оптимальність отриманих моделей.

Скористаємося розглянутою процедурою для оптимального вибору змінних для моделі. Нехай потрібно пояснити змінну y за допомогою набору змінних yd, pb, pc. При побудові моделі напишемо залежну змінну, а також константу, яку обов'язково включимо до моделі. При необхідності можна вказати цілий набір змінних, які обов'язково мають бути присутніми в моделі.



Одним з основних етапів запуску процедури є налаштування опцій.



У даному вікні слід вибрати методу вибору оптимальної регресії. У EViews запрограмовано 4 способи:

1. Метод односпрямованого руху (Uni-directional)

1.1. Forwards. Цей метод починає будувати регресію без факторів, а потім додає змінні з найнижчими значеннями p-value у випадку їх додавання до регресії. Процес повторюється, поки нове значення p-value не стане більшим величини, заданої користувачем.

1.2. Backwards. Цей метод починає будувати регресію з усіма факторами, а потім виключає змінні з найвищими значеннями p-value. Процес повторюється, поки нове значення p-value не стане меншим величини, заданої користувачем.

2. Метод покрокового відбору (Stepwise)

2.1. Forwards. Цей метод в цілому аналогічний до односпрямованого пошуку, однак при додаванні змінних також відбувається процес виключення змінних з найвищим p-value.

2.2. Backwards. У цьому методі змінні з регресії виключаються на основі p-value, але при цьому можуть додаватися нові змінні.

3. Метод покрокової заміни (Swapwise)

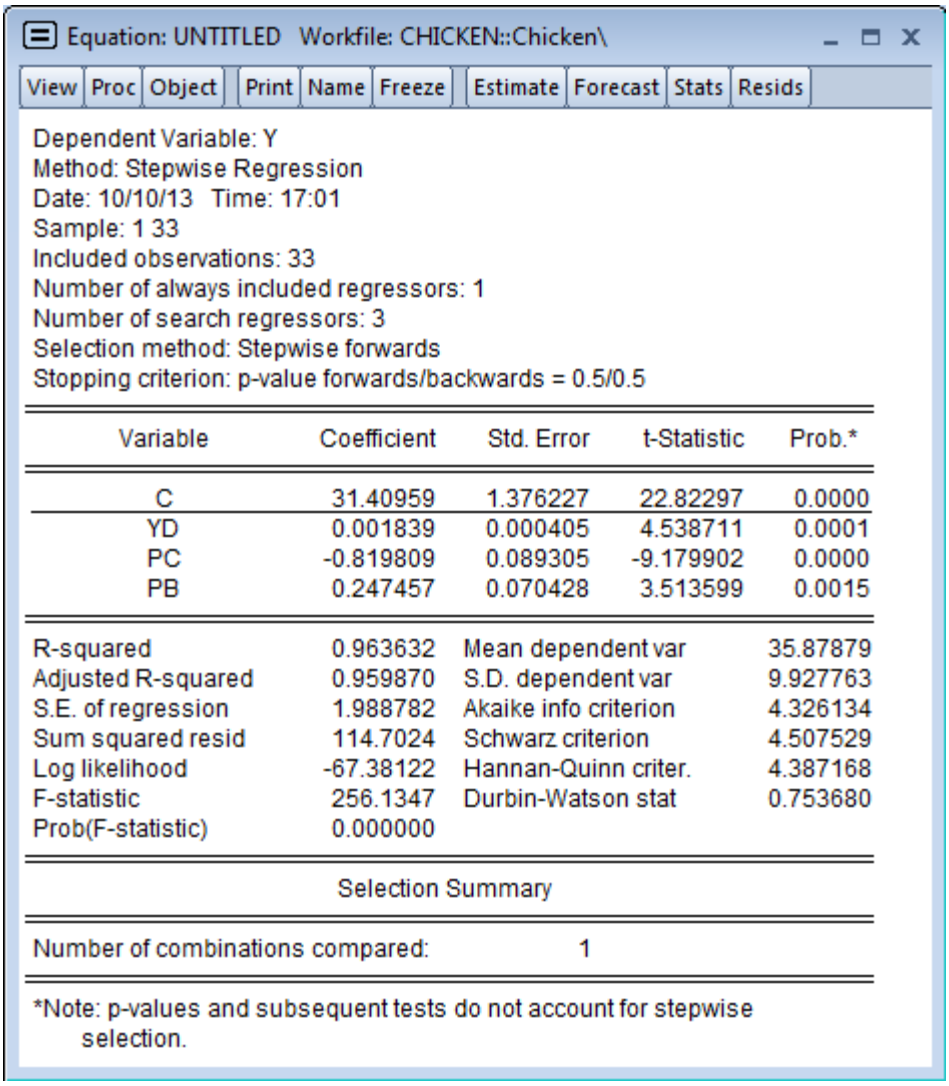
3.1. Max R-Squared Increment. Спочатку розглядається регресія без факторів. На кожному кроці додається змінна, яка максимізує коефіцієнт детермінації. Далі кожні дві змінні регресії порівнюються з іншими змінними, які не включено до моделі, щодо доцільності їх заміни у моделі на основі величини R-squared. Потім додається третя змінна, порівнюються всі трійки змінних включених до регресії з усіма трійками змінних, що не включені до регресії. Процес повторюється, поки не буде знайдена найкраща регресія.

3.2. Swapwise-Min R-Squared Increment. Цей метод майже аналогічний попередньому, але процес заміни змінних відбувається не на основі R-squared, а на основі найменшого приросту R-squared

4. Метод повного перебору (Combinatorial). Цей метод здійснює повний перебір варіантів та вибирає модель з найбільшим R-squared.

Для кожного методу є можливість вказати критерій порівняння (p-value або статистика Стьюдента), максимальну кількість кроків, максимальну кількість регресорів у кінцевій моделі.

Застосуємо для наших даних метод покрокового відбору. В результаті бачимо, що найкращою буде модель, до якої включені всі змінні.



Equation: UNTITLED Workfile: CHICKEN::Chicken\

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

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

Раніше, при побудові моделі, використовувалися переважно лінійні доданки. Якщо модель була записана у нелінійній формі, то за допомогою різномантіних перетворень, вона зводилася до лінійного вигляду. Проте на практиці не завжди потрібно переходити до лінійності. Якщо справжня залежність між змінними є нелінійною, то оцінювання лінійної моделі призводить до зміщених оцінок. Таким чином, набагато важливіше встановити коректну функціональну форму моделі, тобто за допомогою яких функцій мають бути записані регресори.

Найбільш уживаним критерієм правильності функціональної форми є критерій RESET (*Regression Specification Error Test*).

Для перевірки лінійності в моделі

$y_t = X\beta + \varepsilon_t$,

слід оцінити вихідну модель звичайним методом найменших квадратів, а потім в допоміжній регресії

$$y_t = X\beta + \alpha_2 \hat{y}_t^2 + \alpha_3 \hat{y}_t^3 + ... + a_q \hat{y}_t^q + \upsilon_t$$

перевірити гіпотезу

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

Прийняття H_0 означає коректність лінійної моделі. Можна використати стандартний $-F-$ критерій для перевірки гіпотези про лінійні обмеження на коефіцієнти моделі або загальний критерій Вальда зі статистикою

$$W = (q-1)F,$$

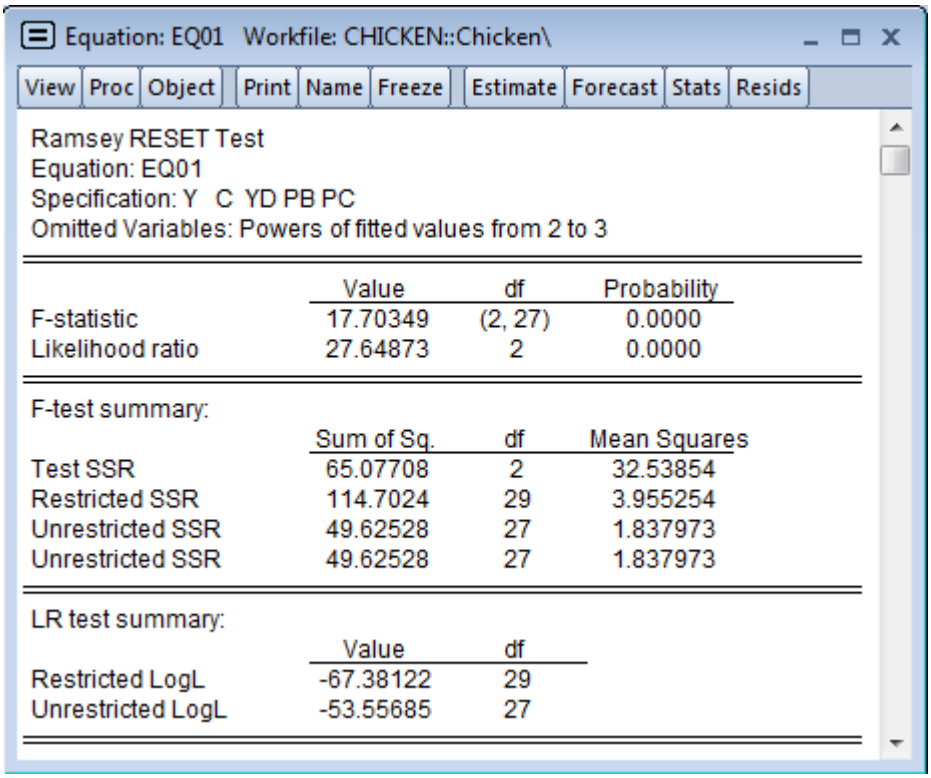
яка асимптотично має розподіл χ-квадрат з $q-1$ степенями свободи.

У випадку відхилення гіпотези про нормальність збурень у моделі, застосовують LR-критерій.

Якщо нульова гіпотеза відхиляється в модель потрібно включити відповідні степені та добутки вихідних змінних. На практиці достатньо обмежитись вибором $q = 3$ та $q = 2$.

Зауважимо також, що існує велика кількість свідчень на користь того, що модель, в яку входять логарифми змінних, а також доданки другого порядку відносно логарифмів є загалом кращою апроксимацією у випадку нелінійної залежності.

Для нашої моделі скористаємося тестом RESET (**View-Stability Diagnostics-Ramsey RESET Test...**) з параметром $q = 2$:


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

З нього видно, що оскільки значення Probability<0,05, то гіпотеза H_0 відхиляється, а значить, лінійна форма для моделі підібрана невірно. Аналогічний результат дає і варіант тесту з $q=3$. Таким чином, лінійна форма моделі не є коректною.

Самостійна робота

1. У файлі COBB.XLS знаходяться дані, використані Коббом і Дугласом у їхній праці з дослідження виробничої функції. Оцініть виробничу функцію Кобба-Дугласа, припускаючи відсутність технічного прогресу і вважаючи еластичність праці і капіталу незалежними параметрами. Оцініть параметри виробничої функції, використовуючи її вираз в термінах капіталоозброєності. Наскільки правомірно використовувати гіпотезу про постійність ефекту масштабу? Перевірте гіпотезу за допомогою критерію Вальда. Визначити, який з факторів є більш важливим для випуску продукції.

2. За допомогою комп'ютера розв'язати задачі: **2.21, 2.22, 2.25, 2.26, 2.28.**

3. Розв'язати задачі: **2.13, 2.14.**