Семінар 11. Регресія з автокорельованими збуреннями

В цьому розділі ми розглянемо можливість виявлення автокореляції збурень, а також методи оцінки регресій за умови автокореляції збурень:

1. Виявлення автокореляції.
2. Узагальнений метод найменших квадратів.

Виявлення автокореляції

Статистика Дарбіна-Уотсона

Найчастіше для виявлення автокорельованості збурень користуються критерієм Дарбіна–Уотсона. При застосуванні цього критерію нульовою гіпотезою є некорельованість збурень, а альтернативою є те, що збурення підпорядковані процесу авторегресії першого порядку. Позначимо через $u_i, 1 \le i \le n$ залишки методу найменших квадратів у моделі

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \upsilon_i, i = \overline{1,n} \quad (1)$$

Значення статистики Дарбіна–Уотсона знаходиться за наступною формулою:

$$d = \frac{\sum_{i=2}^{n} (u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2}. \quad (2)$$

Можливі значення *d* належать інтервалу (0; 4). Розподіл статистики Дарбіна–Уотсона приблизно симетричний відносно двійки. Значення *d*, близькі до 2, вказують на відсутність автокореляції. Значення, близькі до 0, вказують на наявність автокореляції з додатнім ρ, значення, близькі до 4, вказують на наявність автокореляції з від’ємним ρ. Параметрами розподілу статистики Дарбіна–Уотсона є кількість спостережень та регресорів. В таблицях приводяться такі пари критичних значень, що для будь-якого вигляду матриці **X** точне критичне значення лежить між табличними. Алгоритм застосування критерію Дарбіна–Уотсона полягає у наступному.

1. Оцінюємо модель (1) за допомогою звичайного методу найменших квадратів.

2. За формулою (2) обчислюємо значення статистики Дарбіна–Уотсона.

3. Вибираємо рівень значущості α і за таблицею критичних значень статистики Дарбіна–Уотсона знаходимо верхнє і нижнє критичні значення d_u та d_l, а також обчислюємо $4 - d_u$ та $4 - d_l$. Зауважимо, що $0 < d_l < d_u < 2 < 4 - d_u < 4 - d_l < 4$.

4. Робимо висновок за таким правилом:

1) Якщо $d < d_l$, то має місце автокореляція з додатнім ρ.

2) Якщо $d_l < d < d_u$, то ми не можемо зробити ніякого висновку, і цей інтервал називається областю невизначеності.

3) Якщо $d_u < d < 4 - d_u$, то автокореляція відсутня.

4) Якщо $4 - d_u < d < 4 - d_l$, то ми не можемо зробити ніякого висновку. Цей інтервал також є областю невизначеності.

5) Якщо $4 - d_l < d < 4$, то має місце автокореляція з від'ємним ρ.

Щодо областей невизначеності можна дати таку практичну рекомендацію: якщо вибіркове значення d потрапляє до інтервалу невизначеності, то вважають, що має місце автокореляція.

Нехай ми оцінили наступну регресію[1]:

[1] Файл chicken.wf1

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: 21:20
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		

Значення статистики Дарбіна-Уотсона складає 0,753680.

Критичні значення для 95% надійного інтервалу дорівнюють (див. табл. DW.pdf) 1,26 та 1,65. Таким чином, робимо висновок, що присутня додатня автокореляція.

Статистика Бройша-Годфрі (LM-метод)

Для тестування вибираємо **View → Residual Diagnostics → Serial Correlation LM Test… .** Вводимо потрібну кількість лагів (як правило, 2), та отримуємо результат:

Equation: EQ01 Workfile: CHICKEN::Chicken\

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

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	9.910265	Prob. F(2,27)	0.0006
Obs*R-squared	13.96989	Prob. Chi-Square(2)	0.0009

Оскільки всі ймовірності не перевищують 0,05, то гіпотеза H_0 відхиляється, тобто за критерієм Бройша-Годфрі автокореляція присутня.

Узагальнений метод найменших квадратів

Спочатку оцінюємо модель (1) за методом найменших квадратів, потім обчислюємо статистику Дарбіна–Уотсона і приймаємо рішення про наявність чи відсутність автокореляції. При наявності автокореляції використовуємо вибірковий коефіцієнт кореляції залишків методу найменших квадратів як оцінку параметра ρ:

$$\hat{\rho} = 1 - \frac{d}{2}, \tag{3}$$

де d – статистика Дарбіна-Уотсона. Далі за формулами (4) – (9), в яких параметр ρ замінено його оцінкою (3), знаходимо $\mathbf{y}^*$ та $\mathbf{X}^*$**.**

На останньому етапі ми оцінюємо модель $\boldsymbol{y} = \boldsymbol{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ використовуючи звичайний метод найменших квадратів.

Елементи вектора $\mathbf{y}^*$дорівнюють

$$y_1^* = \sqrt{1-\rho^2}\, y_1, \tag{4}$$

$$y_i^* = y_i - \rho y_{i-1},\ 2 \le i \le n. \tag{5}$$

Елементи *j*-го ($0 \le j \le k-1$) стовпчика матриці $\mathbf{X}^*$ знаходяться аналогічно:

$$x_{1j}^* = \sqrt{1-\rho^2}\, x_{1j}, \tag{6}$$

$$x_{ij}^* = x_{ij} - \rho x_{i-1,j},\ 2 \le i \le n. \tag{7}$$

Якщо у вихідній моделі є постійний доданок, то перетворена модель не матиме константи. Замість неї з’явиться змінна x_0^*, значення якої дорівнюють

$$x_{10}^* = \sqrt{1-\rho^2}, \tag{8}$$

$$x_{i0}^* = 1 - \rho,\ 2 \le i \le n. \tag{9}$$

Зауважимо, що оцінка β_0- коефіцієнта при змінній x_0^* є оцінкою постійного доданку у вихідній моделі.

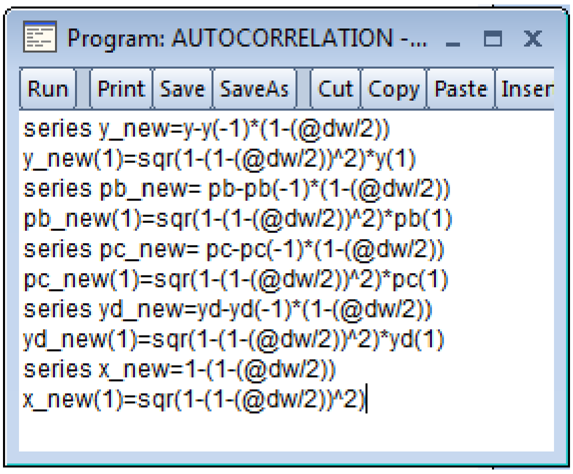
Отже, знаходимо значення параметра ρ за формулою (3): ρ=0,62316.

Далі створюємо нові змінні за формулами (4)-(9), які ми записуємо в командному рядку:

```
series y_new=y-y(-1)*0.62316
y_new(1)=sqr(1-0.62316^2)*y(1)
series pb_new= pb-pb(-1)*0.62316
pb_new(1)=sqr(1-0.62316^2)*pb(1)
series pc_new= pc-pc(-1)*0.62316
pc_new(1)=sqr(1-0.62316^2)*pc(1)
series yd_new=yd-yd(-1)*0.62316
yd_new(1)=sqr(1-0.62316^2)*yd(1)
series x_new=1-0.62316
x_new(1)=sqr(1-0.62316^2)
```

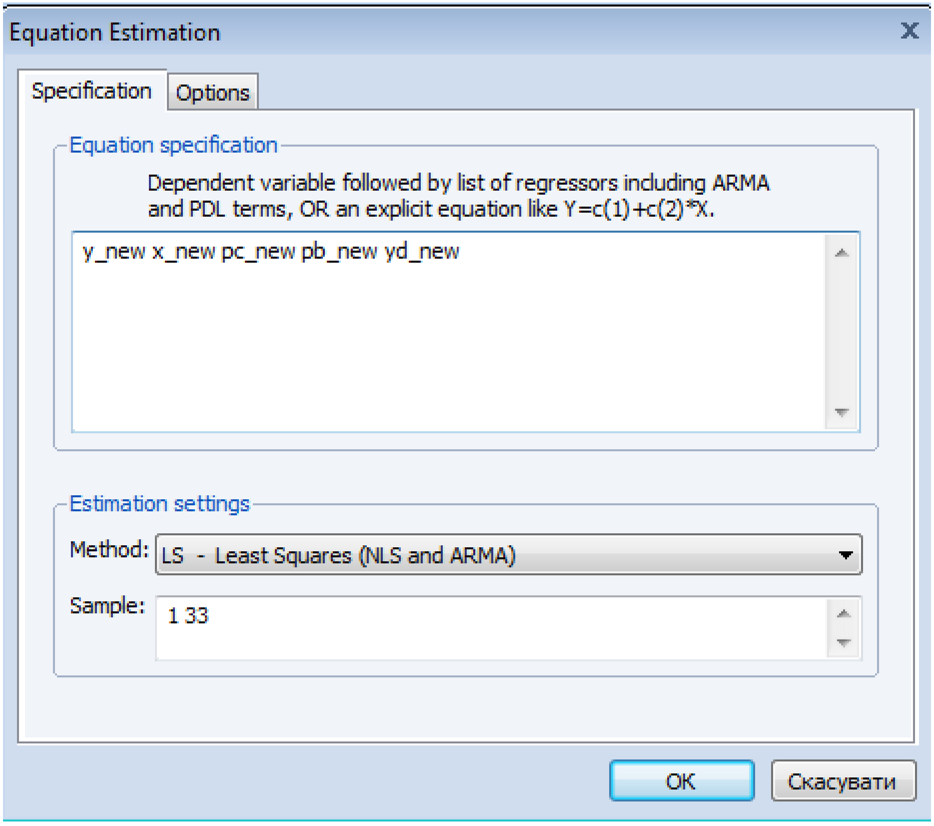
Або за допомогою створення наступної програми (autocorrelation.prg):

```
series y_new=y-y(-1)*(1-(@dw/2))
y_new(1)=sqr(1-(1-(@dw/2))^2)*y(1)
series pb_new= pb-pb(-1)*(1-(@dw/2))
pb_new(1)=sqr(1-(1-(@dw/2))^2)*pb(1)
series pc_new= pc-pc(-1)*(1-(@dw/2))
pc_new(1)=sqr(1-(1-(@dw/2))^2)*pc(1)
series yd_new=yd-yd(-1)*(1-(@dw/2))
yd_new(1)=sqr(1-(1-(@dw/2))^2)*yd(1)
series x_new=1-(1-(@dw/2))
x_new(1)=sqr(1-(1-(@dw/2))^2)
```



Для її створення в меню обирається: **File→New→Program**. Для збереження даних програми необхідно натиснути кнопку **Save**, для запуску програми – кнопку **Run**.

Отже, будуємо нову регресію вигляду:



Слід звернути увагу на те, що в регресії за перетвореними даними на одну незалежну змінну більше, ніж в початковій регресії, та вона не містить константу.

Оцінена регресія за перетвореними даними має наступний вигляд (після застосування УМНК перевірку на наявність автокореляції робити не потрібно!):

Equation: EQ05 Workfile: CHICKEN::Chicken\

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

Dependent Variable: Y_NEW
Method: Least Squares
Date: 08/28/13 Time: 10:17
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X_NEW	30.58421	1.462484	20.91251	0.0000
PC_NEW	-0.750298	0.094056	-7.977146	0.0000
PB_NEW	0.235184	0.067729	3.472418	0.0016
YD_NEW	0.001925	0.000391	4.918569	0.0000

R-squared	0.949474	Mean dependent var	26.62504
Adjusted R-squared	0.944248	S.D. dependent var	7.169532
S.E. of regression	1.692867	Akaike info criterion	4.003937
Sum squared resid	83.10820	Schwarz criterion	4.185332
Log likelihood	-62.06496	Hannan-Quinn criter.	4.064971
Durbin-Watson stat	1.071670		

Самостійна робота

1. Використовуючи дані з файлу expend.wf1, побудуйте регресії рівня витрат на продукти харчування по особистому доходу у розпорядженні, індексу відносних цін, а також множинну регресію по обох показниках. Для кожної з регресій спробуйте відповісти на запитання: чи присутня автокореляція? У випадку наявності автокореляції використайте узагальнений метод найменших квадратів для позбавлення від неї.

2. Менеджер фірми протягом року акуратно вів записи своїх доходів ***DPI*** і витрат ***CONS***:

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
DPI	2500	3000	3500	4000	4500	5000	5500	6000	6200	7000	7200	8000
CONS	2000	2300	2500	3800	3200	5000	4000	5300	4200	6000	4800	7000

Оскільки його робота була успішною, це позитивно відбивалося на його доході (у карбованцях), при цьому і витрачати він став набагато більше. Щоб знайти цей зв'язок, побудуйте регресію рівня споживання по доходу. Чи наявна автокореляція в регресії? Позбудьтеся її за допомогою узагальненого методу найменших квадратів.

3. Використовуючи дані з файлу M2.xls побудувати регресійну модель та перевірити її на наявність автокореляції.

4. За допомогою комп'ютера розв'язати задачі: 4.9-4.17.

5. Розв'язати задачі: 4.1, 4.6, 4.7.