

Вступ до прикладної економетрики

К.е.н., доц. Ставицький А.В.

www.andriystav.cc.ua

Література

- Greene W. H., Econometric Analysis, 7th ed., Prentice Hall, 2011.
- Murray M. P. (2005) Econometrics: A Modern Introduction. Prentice Hall.
- Stock James H., Mark W. Watson (2010) Introduction to Econometrics (3rd Edition)
- Черняк О.І., Комашко О.В., Ставицький А.В., Баженова О.В. Економетрика (2010)

План

- Основні аспекти економетрики
- Приклади нелінійних моделей

ОСНОВНІ АСПЕКТИ ЕКОНОМЕТРИКИ



Економетричний аналіз

- Теоретичний підхід
- Практичний підхід

Типи даних

- ✓ Часові ряди (Time series data)
- ✓ Перехресні дані (Cross-sectional data)
- ✓ Панельні дані (Panel data)

Часові ряди

- **Види**

- кількісні (наприклад, обмінний курс, ціни акцій, випуск продукції),
- якісні (наприклад, день тижня, стать людини).

- **Приклади часових рядів**

Ряди

GNP, безробіття

дефіцит бюджету

Пропозиція грошей

Біржовий індекс

Частота

місячні чи квартальні

квартальні чи річні

тижневі чи місячні

щохвилини

Приклади задач для часових рядів

1. Як біржова капіталізація пов'язана з основними макроекономічними показниками країни?
2. Як ціна акцій змінюється в залежності від оголошеного розміру дивідендів?
3. Як зміна ставок НБУ впливає на курс валюти?

Перехресні дані

- **Перехресні дані** – дані, у яких розглядаються значення декількох змінних у один період часу:
 - Набір обсягу продажів всіх брокерів біржі за один день
 - Дивіденди всіх акцій на біржі за поточний рік
 - Набір даних про кредитні рейтинги банків

Приклади задач для перехресних даних

- Залежність між розміром компанії та дивідендами
- Залежність між ВВП та ймовірністю дефолту за державними боргами

Панельні дані

- Панельні дані узагальнюють часові ряди та перехресні дані, використовуючи спостереження за декількома змінними декілька періодів.

Приклади задач для панельних даних

- Залежність між ВВП та обсягом державного боргу у всіх країнах ЄС
- Залежність між обсягом продажів та кількістю персоналу у всіх філіях компанії

Етапи моделювання

- Визначення проблеми чи гіпотези
- Розробка моделі
- Збір статистики
- Проведення описового аналізу даних
- Оцінка невідомих коефіцієнтів
- Оцінка якості та придатності моделі
- Використання моделі для прогнозу

Мета

- Розробка статистичної моделі, яка дозволить прогнозувати значення залежної змінної на основі незалежних змінних.

Лінійна регресія

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

y_t - залежна змінна;

$x_{1t}, x_{2t}, \dots, x_{k-1t}$ незалежні змінні;

ε_t - збурення.

Припущення

- **Лінійність** - Y лінійно залежить від набору X .
- **Незалежність похибок** – збурення незалежні з X .
- **Гомоскедастичність** – дисперсія збурень є константою для всіх X .
- **Нормальність** – збурення мають нормальний розподіл.

Метод найменших квадратів

- Визначає коефіцієнти регресії, при яких різниця між реальними даними (Y) та прогнозними (\hat{Y}) буде найменшою:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \rightarrow \min$$

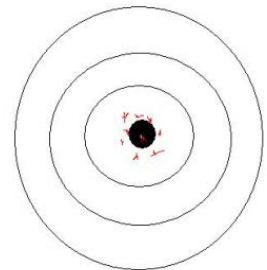
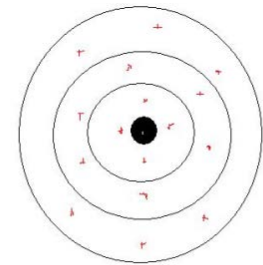
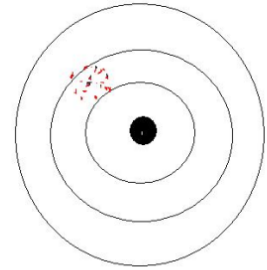
Бажані характеристики

- Незміщеність $E(\hat{\beta}) = \beta$
- Ефективність
 - Стандартна похибка/дисперсія має бути мінімальною:

$$\text{var}(\hat{\beta}) = \frac{1}{\sum x_i^2} \sigma^2 = \frac{\sigma^2}{\sum x_i^2}$$

МНК мінімізує σ^2 (дисперсію збурень)

- Консистентність
 - При збільшенні N стандартна похибка зменшується



Перетворення до лінійного вигляду

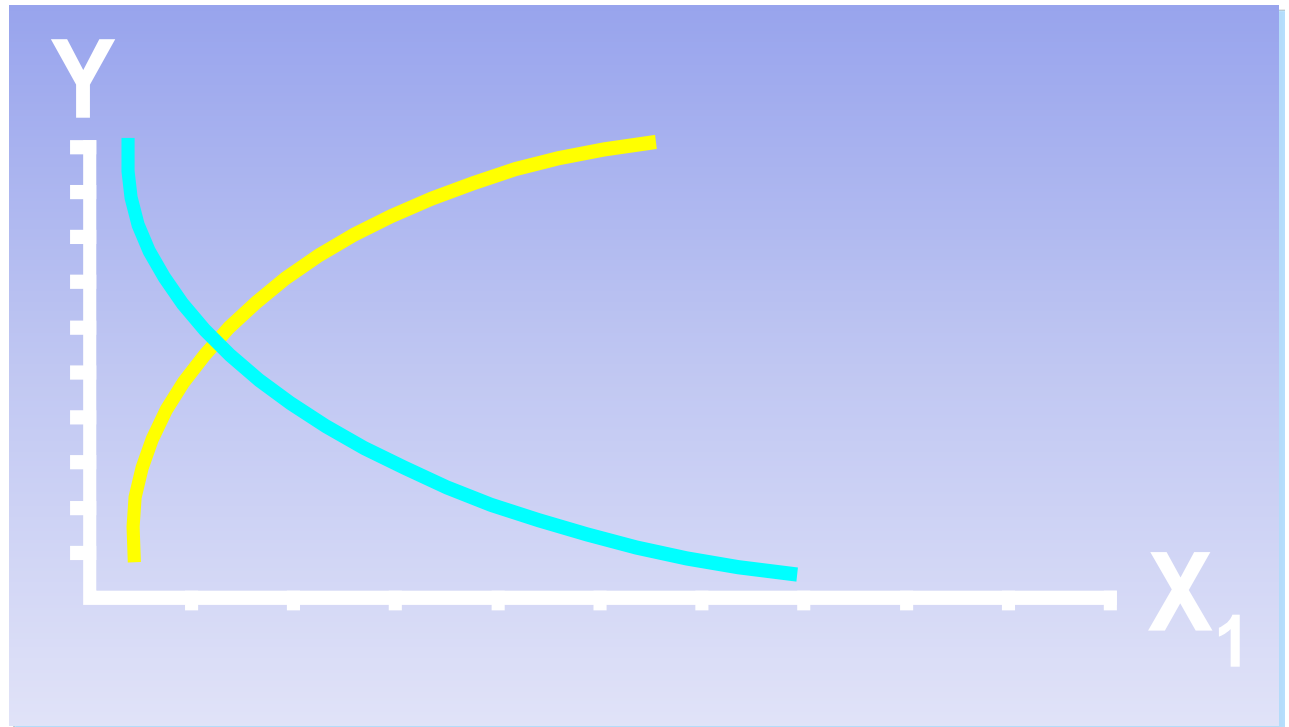
- Нелінійні моделі можуть бути виражені та оцінені у лінійній формі
- Необхідна трансформація даних

Логарифмчна трансформація

$$Y = \beta + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \varepsilon$$

$\beta_1 > 0$

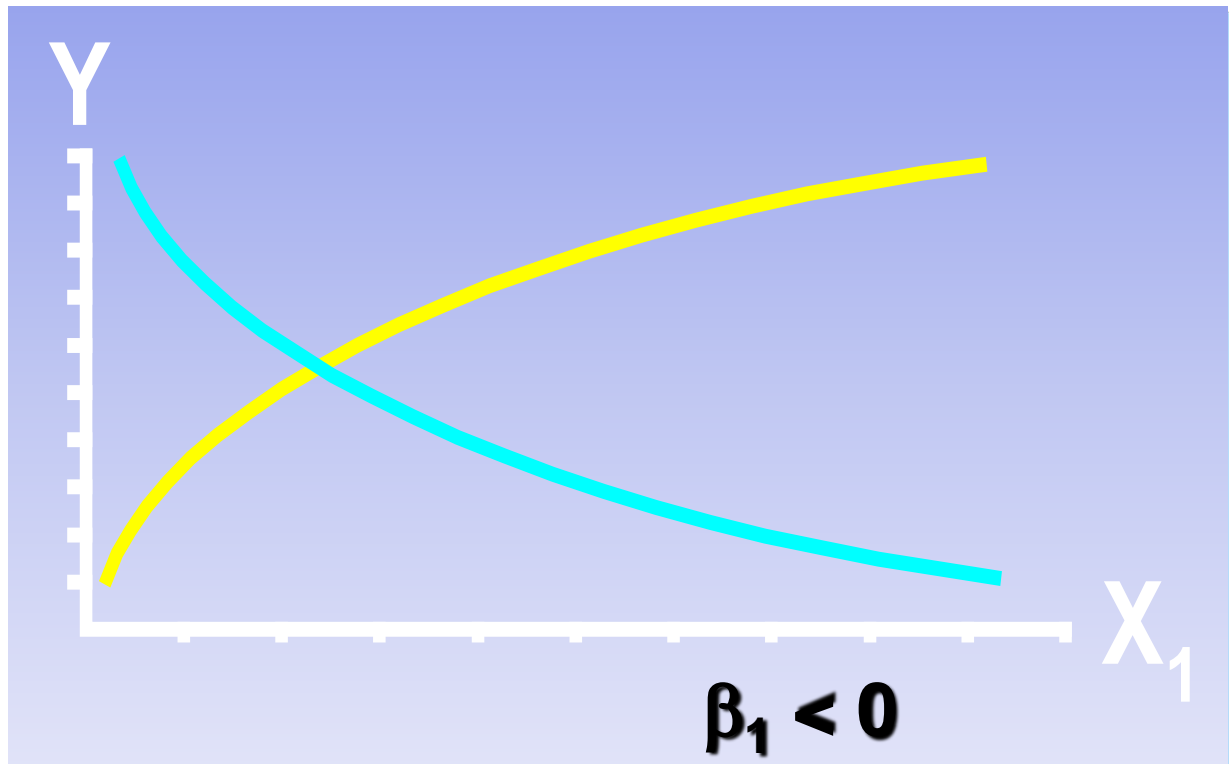
$\beta_1 < 0$



Степенева трансформація

$$Y_i = \beta_0 + \beta_1 \sqrt{X_{1i}} + \beta_2 \sqrt{X_{2i}} + \varepsilon_i$$

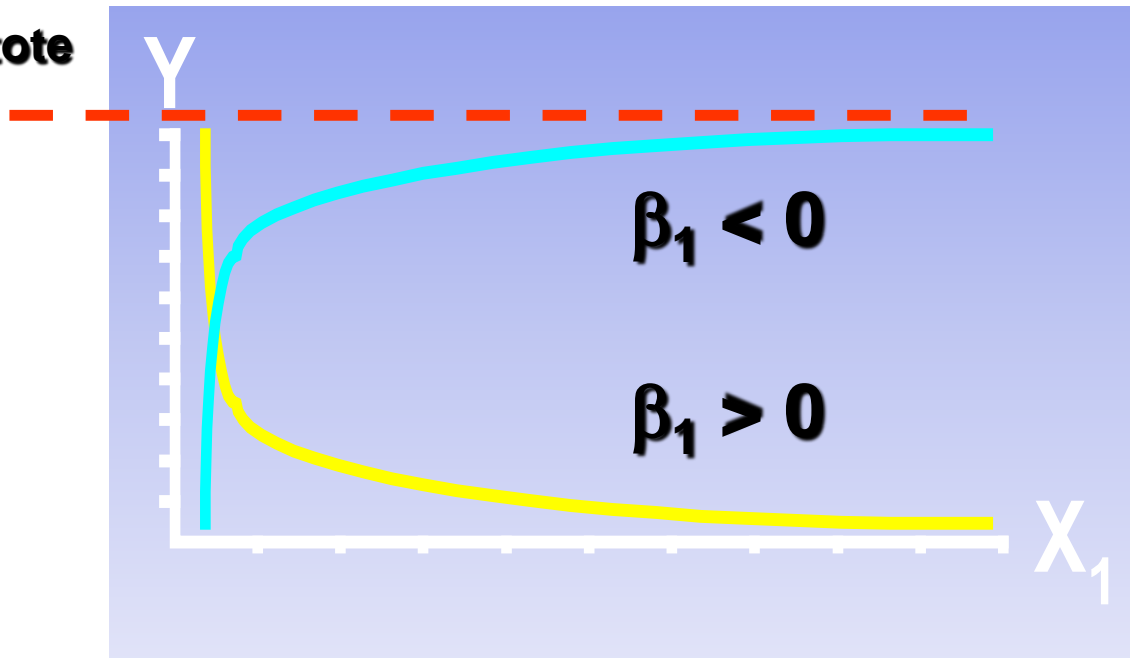
$\beta_1 > 0$



Обернена трансформація

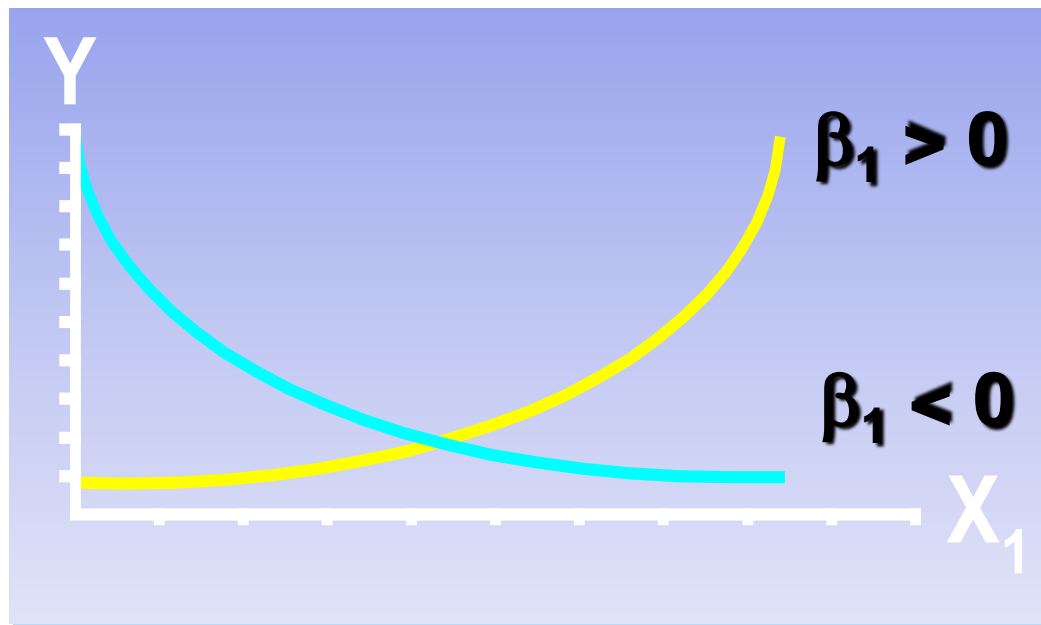
$$Y_i = \beta_0 + \beta_1 \frac{1}{X_{1i}} + \beta_2 \frac{1}{X_{2i}} + \varepsilon_i$$

Asymptote



Експоненційна трансформація

$$Y_i = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}} \varepsilon_i$$



Моделі з фіктивними змінними

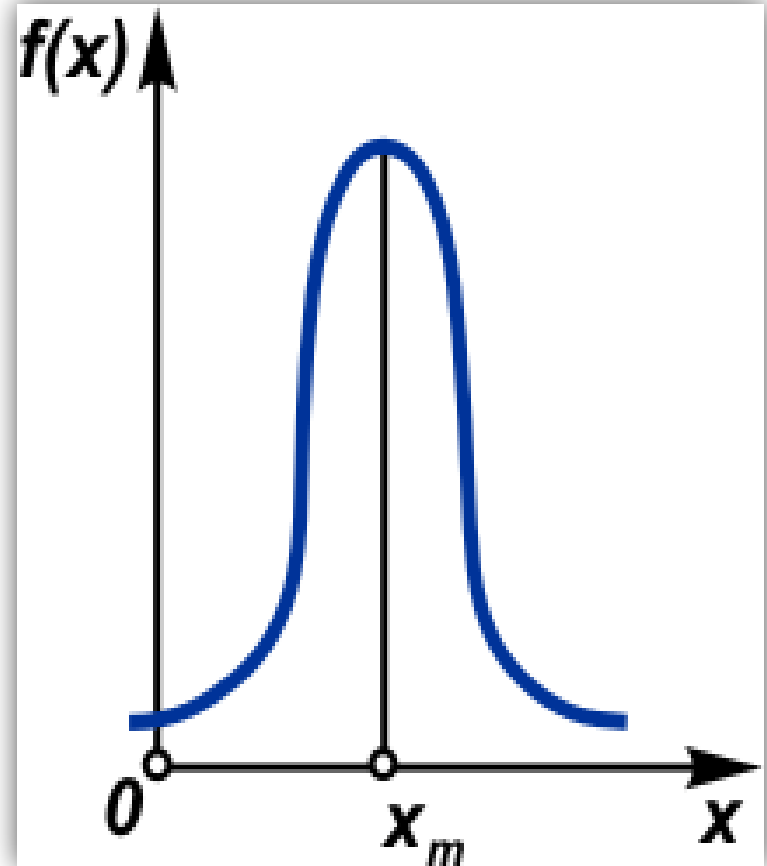
- Використовує незалежні змінні з 2 значеннями:
 - *наприклад, жінки-чоловіки, працює-безробітний тощо.*
- Змінні приймають значення 0 та 1
- Передбачається, що константа моделі різна, а коефіцієнти регресії – однакові.

Кроки для оцінки якості регресійної моделі

- Тест залишків на нормальність
- Тест значимості коефіцієнтів
- Тест адекватності моделі
- Тест на мультиколінеарність
- Тест на стійкість моделі
- Тест на автокореляцію залишків
- Тест на гетероскедастичність залишків
- Тест на специфікацію моделі

Тест залишків на нормальність

Критерій Жарке-Бера (Jarque-Bera statistics)



Критерій Жарке-Бера

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

$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$	$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$
---	---

- S - коефіцієнт асиметрії,
- K - ексцес.

Приклад

Dependent Variable: TAX_ENT

Method: Least Squares

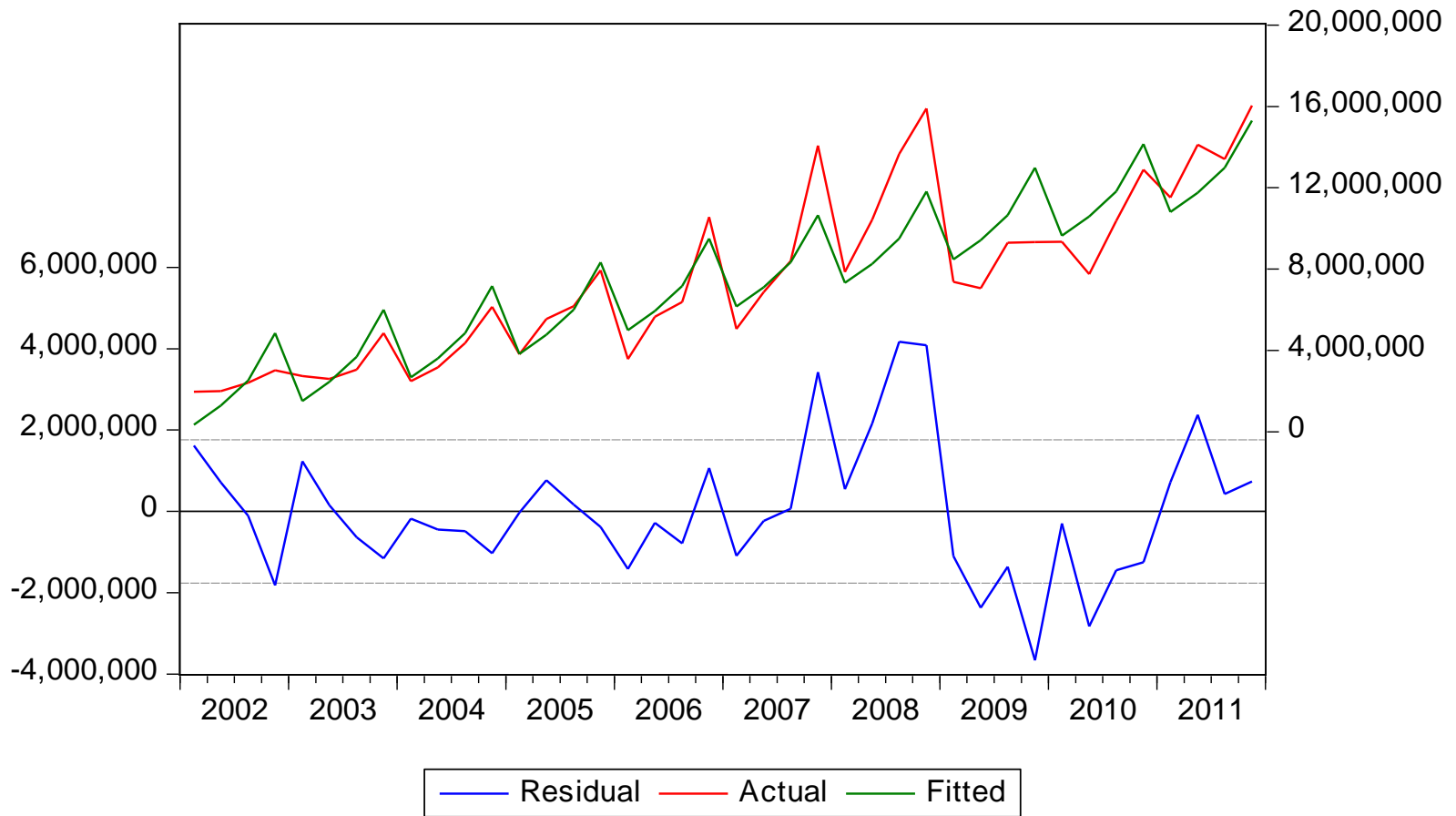
Date: 12/09/12 Time: 20:49

Sample: 2002Q1 2011Q4

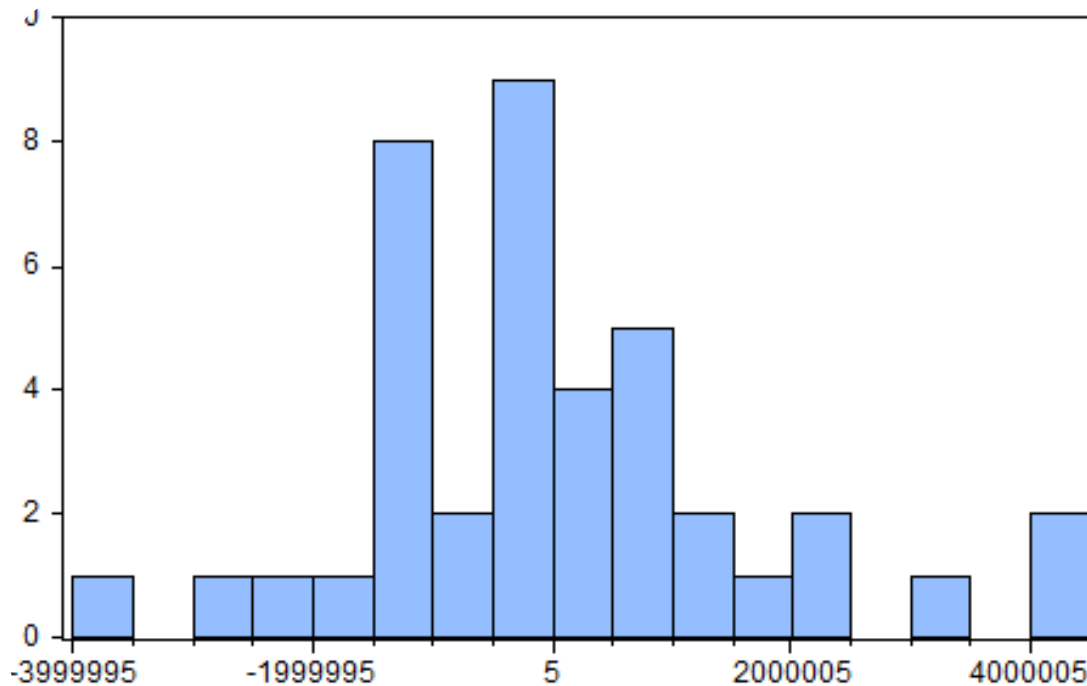
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3973770.	754540.7	5.266475	0.0000
@TREND	290525.1	24239.34	11.98568	0.0000
@SEAS(1)	-3627516.	791034.8	-4.585786	0.0001
@SEAS(2)	-2975920.	789175.7	-3.770922	0.0006
@SEAS(3)	-2032456.	788058.1	-2.579068	0.0143
R-squared	0.837415	Mean dependent var		7480035.
Adjusted R-squared	0.818834	S.D. dependent var		4138083.
S.E. of regression	1761318.	Akaike info criterion		31.71749
Sum squared resid	1.09E+14	Schwarz criterion		31.92860
Log likelihood	-629.3498	Hannan-Quinn criter.		31.79382
F-statistic	45.06800	Durbin-Watson stat		1.123746
Prob(F-statistic)	0.000000			

Залишки



Перевірка на нормальність



Series: Residuals
Sample 2002Q1 2011Q4
Observations 40

Mean	-3.49e-11
Median	-204126.5
Maximum	4171075.
Minimum	-3659494.
Std. Dev.	1668551.
Skewness	0.590368
Kurtosis	3.696860

Jarque-Bera	3.132917
Probability	0.208783

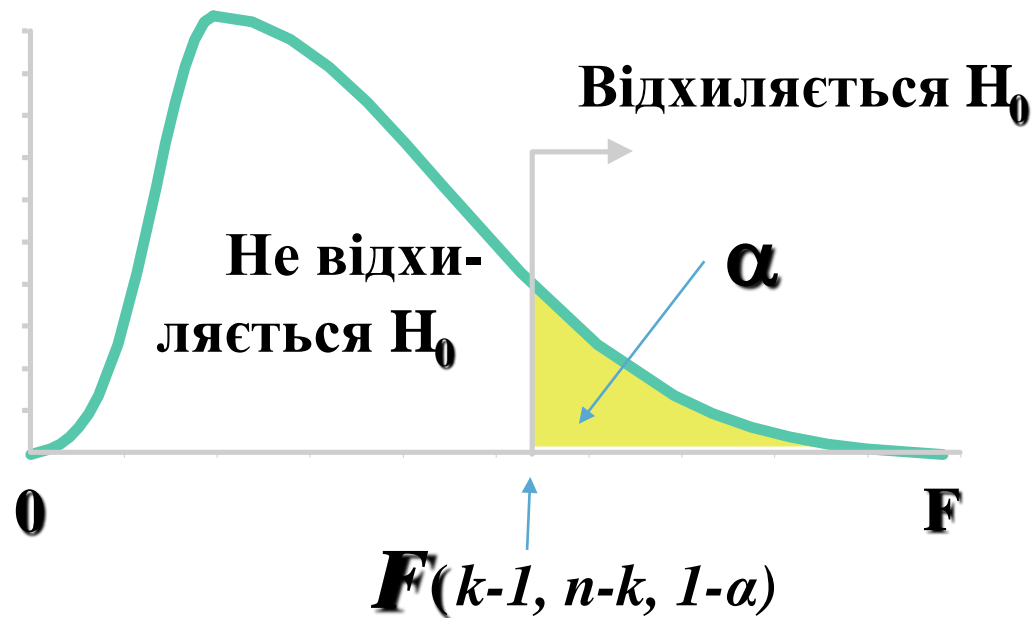
Перевірка моделі на адекватність

- Гіпотеза
 - $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$
 - *Немає зв'язку*
 - H_a : Хоча б один коефіцієнт не рівний 0
 - *Хоча б одна змінна впливає на Y*

$$F = \frac{RSS / (k - 1)}{ESS / (n - k)} = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} \stackrel{H_0}{\sim} F_{k, n-k}$$

Правило відхилення гіпотези

- Відхиляється H_0 на користь H_a , якщо F_{calc} попадає у зафарбовану область



- Відхилити H_0 , якщо $P\text{-value} = P(F > F_{\text{calc}}) < \alpha$

Приклад

Dependent Variable: TAX_ENT

Method: Least Squares

Date: 12/09/12 Time: 20:49

Sample: 2002Q1 2011Q4

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3973770.	754540.7	5.266475	0.0000
@TREND	290525.1	24239.34	11.98568	0.0000
@SEAS(1)	-3627516.	791034.8	-4.585786	0.0001
@SEAS(2)	-2975920.	789175.7	-3.770922	0.0006
@SEAS(3)	-2032456.	788058.1	-2.579068	0.0143

R-squared	0.837415	Mean dependent var	7480035.
Adjusted R-squared	0.818834	S.D. dependent var	4138083.
S.E. of regression	1761318.	Akaike info criterion	31.71749
Sum squared resid	1.09E+14	Schwarz criterion	31.92860
Log likelihood	-629.3498	Hannan-Quinn criter.	31.79382
F-statistic	45.06800	Durbin-Watson stat	1.123746
Prob(F-statistic)	0.000000		

Тест на значення коефіцієнта

- Гіпотеза
 - $H_0: \beta_i = m$
 - $H_a: \beta_i \neq m$

Статистика

$$t = \frac{\hat{\beta}_i - m}{S_{\hat{\beta}_i}}$$

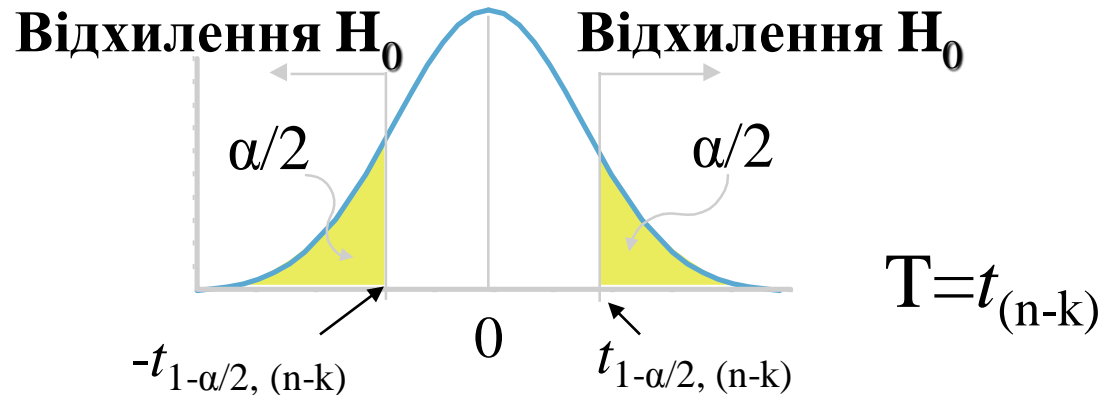
$$\text{де } S_{\hat{\beta}_i} = \frac{S}{\sqrt{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}}$$

$$S = \hat{\sigma} = \sqrt{\frac{ESS}{n-k}}$$

$$ESS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \left[Y_i - \left(\hat{\beta}_0 + \sum_{i=1}^{k-1} \hat{\beta}_i X_i \right) \right]^2$$

Правило

- Відхилити H_0 на користь H_a , якщо t попадає в зафарбовану область



- Відхилити H_0 , якщо $P\text{-value} = P(T > |t|) < \alpha$

Частковий випадко: значимість коефіцієнтів

- Гіпотеза
 - $H_0: \beta_i = 0$
 - $H_a: \beta_i \neq 0$

$$t = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}}$$

Приклад

Dependent Variable: TAX_ENT

Method: Least Squares

Date: 12/09/12 Time: 20:49

Sample: 2002Q1 2011Q4

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3973770.	754540.7	5.266475	0.0000
@TREND	290525.1	24239.34	11.98568	0.0000
@SEAS(1)	-3627516.	791034.8	-4.585786	0.0001
@SEAS(2)	-2975920.	789175.7	-3.770922	0.0006
@SEAS(3)	-2032456.	788058.1	-2.579068	0.0143
R-squared	0.837415	Mean dependent var		7480035.
Adjusted R-squared	0.818834	S.D. dependent var		4138083.
S.E. of regression	1761318.	Akaike info criterion		31.71749
Sum squared resid	1.09E+14	Schwarz criterion		31.92860
Log likelihood	-629.3498	Hannan-Quinn criter.		31.79382
F-statistic	45.06800	Durbin-Watson stat		1.123746
Prob(F-statistic)	0.000000			

Тест Вальда (Wald test)

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$H_1 : \beta_1$ або β_2 або β_3 - не дорівнюють 0.

В матричній формі:

$$\text{Гіпотеза: } Rb = r \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Статистика:

$$F = \frac{(Rb - r)' [R \text{ cov}(b) R']^{-1} (Rb - r)}{J}$$

Розрахувати F та порівняти з критичним $F(J, n-k)$ з таблиці Фішера.

Мультиколінеарність

- Висока кореляція між змінними X
- Призводить до нестабільності коефіцієнтів моделі
- Завжди існує, питання лише у степені
- *Приклад*: Використання кількості кімнат та спалень як незалежних змінних в одній моделі

Визначення мультиколінеарності

- Критерій Фара-Глаубера
- VIF-тест

- Способи позбавлення
 - Збільшення кількості спостережень
 - Видалення однієї з корельованих змінних
 - Нормалізація змінних.

Приклад

$$\hat{s}_t = 0.4 + 0.8y_t + 0.2li_t - 0.1si_t$$

(0.9) (1.2) (0.4) (0.1)

$\bar{R}^2 = 0.98$, (стандартні похибки у дужках)

(n = 60), де:

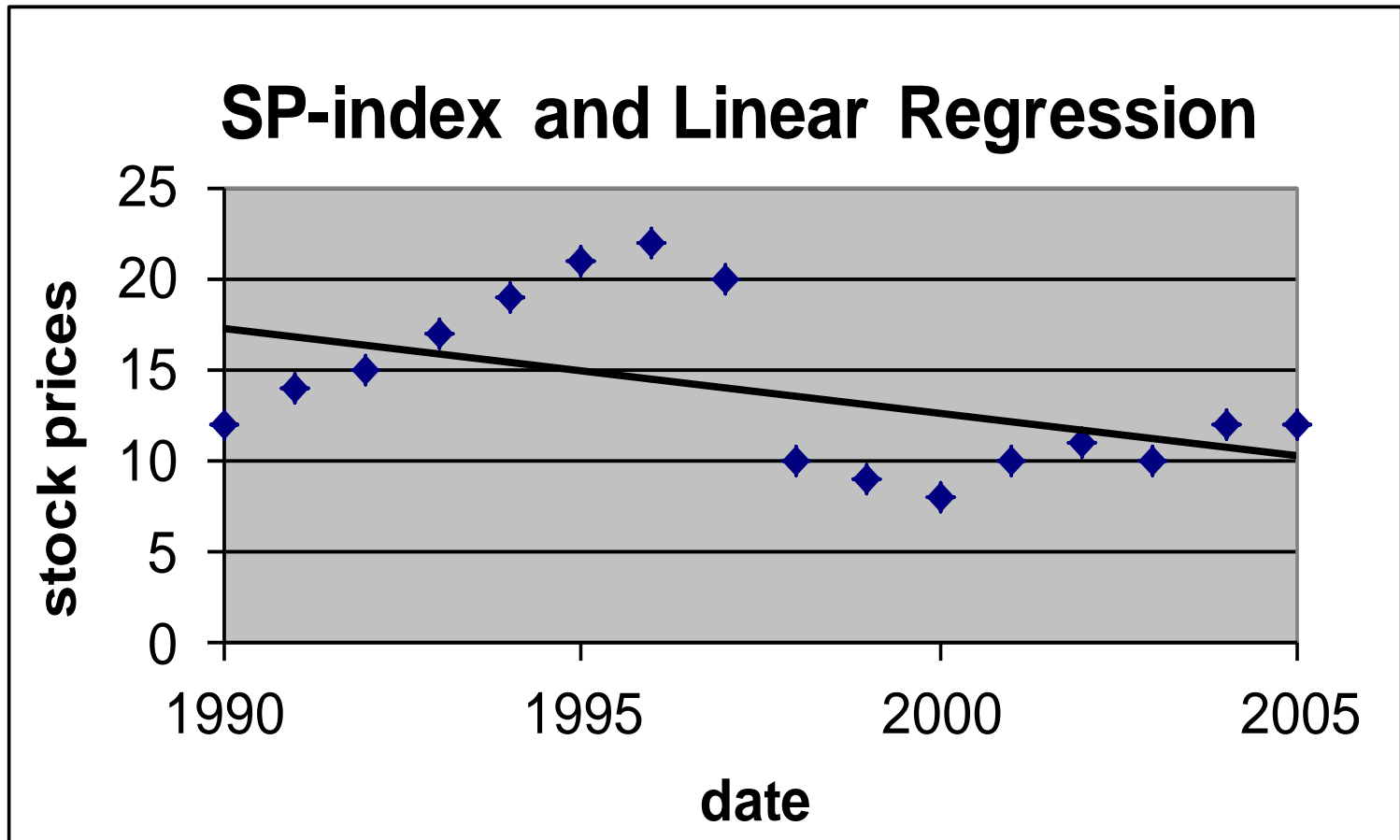
s_t – ціна акцій

y_t – ВВП

li_t – довгострокові ставки

si_t – короткострокові ставки

Тест на стабільність



Тест Чоу

- Перевіряє рівність коефіцієнтів у двох регресіях різних вибірок.

$$F = \frac{RSS_c - (RSS_1 + RSS_2) / k}{RSS_1 + RSS_2 / n - 2k} \sim F_{k, n-2k}$$

RSS_c – combined _RSS

RSS_1 – pre – break _RSS

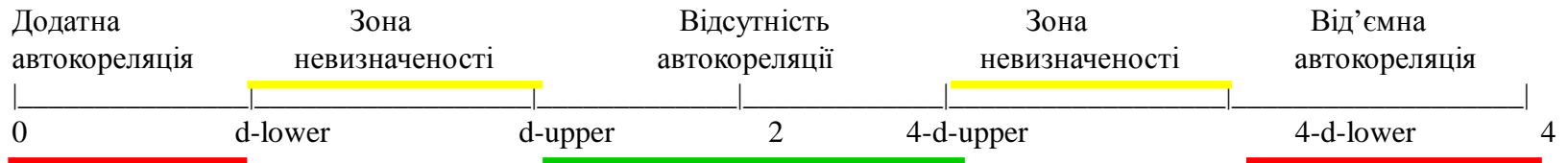
RSS_2 – post – break _RSS

Тести на автокореляцію

- **Durbin-Watson test** (перевіряє автокореляцію першого порядку)
- **Breusch-Godfrey Test** (перевіряє автокореляцію порядку q)

Статистика Дурбіна-Уотсона

$$d = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2}, \text{ для } n \text{ і } K-1 \text{ d.f.}$$



- Явна автокореляція
- Автокореляція під питанням
- Відсутня автокореляція

Критерій Бройша-Годфрі

Процес авторегресії: AR(p)

$$\mu_t = \rho_1 \mu_{t-1} + \rho_2 \mu_{t-2} + \dots + \rho_p \mu_{t-p} + \varepsilon_t$$

Гіпотеза

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$$

Тест моделі:

$$\hat{\mu}_t = \delta_1 + \delta_2 X_{2t} + \dots + \delta_k X_{kt} + \lambda_1 \hat{\mu}_{t-1} + \dots + \lambda_p \hat{\mu}_{t-p} + \omega_t$$

Статистика

$$LM = (n-p) * R_{\text{aux}}^2 \sim \chi_p^2$$

Тести на гетероскедастичність

- Типи тестів:
 - Регресійні: *White test, Breusch–Pagan tests тощо.*
 - Загальні: *the Goldfeld–Quandt test*

Тест Уайта

- Оцінка регресії Y відносно всіх змінних, розрахунок залишків $\varepsilon_1, \dots, \varepsilon_n$
- Регресія ε_i^2 відносно константи, всіх змінних, їх квадратів та попарних добутків. Розрахунок R^2 .
- Порівняння nR^2 з теоретичним Хі-квадрат з p степенями свободи.

Тест Бройша-Пагана

- The Breusch–Pagan test is very similar to the White test.
- The White test specifies exactly which explanators to include in the auxiliary equation. Because the test includes cross-terms, the number of slopes (p) increases very quickly.
- In the Breusch–Pagan test the econometrician selects which explanators to include. Otherwise, the tests are the same.

Тест Бройша-Пагана

- Економетрист вибирає m змінних у допоміжній регресії.
- В іншому тест схожий на тест Уайта

Тест Голдфелда-Квондта –

1

Розділити n спостережень на 2 групи розмірами n_1 та n_2

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ проти}$$

$$H_a : \sigma_1^2 \neq \sigma_2^2$$

Оцінити регресію Y відносно всіх змінних за вибіркою групи 1.

Оцінити регресію Y відносно всіх змінних за вибіркою групи 2.

Тест Голдфелда-Квондта – 2

$$F_{calc} = \frac{\frac{RSS_1}{n_1 - k}}{\frac{RSS_2}{n_2 - k}} > 1$$

Порівняти F_{calc} з теоретичною F -статистикою з $(n_1 - k)$ та $(n_2 - k)$ степенями свободи.

Тест на специфікацію

$$F_{n-m-k+1}^k \sim \frac{\frac{R_1^2 - R_0^2}{k}}{\frac{1 - R_1^2}{n - m - k}}$$

Тест Рамсея (RESET)

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \delta_1 \hat{y}_2 + \delta_2 \hat{y}_3 + \varepsilon$$

$$H_0: \delta_1 = 0, \delta_2 = 0$$

де $F \sim F_{2, n-k-3}$ або $LM \sim \chi^2(2)$.

Популярні нелінійні регресійні моделі

- Експоненціальна: $(y = ae^{bx})$
- Степенева: $(y = ax^b)$
- Модель постійного росту: $\left(y = \frac{ax}{b+x}\right)$
- Поліноміальна: $(y = a_0 + a_1x + \dots + a_mx^m)$

ОГЛЯД



Лінійна регресія

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

y_t - залежна змінна;

$x_{1t}, x_{2t}, \dots, x_{k-1t}$ незалежні змінні;

ε_t - збурення.

Припущення

- **Лінійність** - Y лінійно залежить від набору X .
- **Незалежність похибок** – збурення незалежні з X .
- **Гомоскедастичність** – дисперсія збурень є константою для всіх X .
- **Нормальність** – збурення мають нормальний розподіл.

Метод найменших квадратів

- Визначає коефіцієнти регресії, при яких різниця між реальними даними (Y) та прогнозними (\hat{Y}) буде найменшою:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \rightarrow \min$$

Кроки для оцінки якості регресійної моделі

- Тест залишків на нормальність
- Тест значимості коефіцієнтів
- Тест адекватності моделі
- Тест на мультиколінеарність
- Тест на стійкість моделі
- Тест на автокореляцію залишків
- Тест на гетероскедастичність залишків
- Тест на специфікацію моделі



ПИТАННЯ?



ДЯКУЮ ЗА УВАГУ!