Конспект лекції №4

Тема 4. Принципи побудови економетричних моделей. Парна лінійна регресія

Міжпредметні зв'язки: Зв'язок із елементами знань і умінь таких навчальних дисциплін як «Макроекономіка», «Мікроекономіка», «Інформатика», «Математика для економістів».

Мета лекції полягає у формуванні полягає у формуванні в студентів системи теоретичних знань та практичних навичок з питань побудови економетричних моделей та оцінки параметрів парної лінійної регресії

План лекції

1. Економетрична модель, її види.
2. Побудова та аналіз економетричної моделі з двома змінними.
3. Сутність методу найменших квадратів.
4. Особливості та етапи економетричного моделювання.

Опорні поняття: математичні методи, модель, економетрика, регресія, метод найменших квадратів, лінійна регресія.

Інформаційні джерела:

Основна та допоміжна література:

1. Васильків І. М., Карпінський Б. А., Максимук О. В., Шкулка С. К. Вступ до економетрики: Навч. посіб. – Львів: Львівський національний університет ім. І. Франка, 2015. – 280 с.

2. Сингаевская Г. И. Функции в Excel. Решение практических задач. М.: Издательский дом «Вильямс», 2009. – 880 с.

3. Наконечний С. І.. Терещенко Т.О. Економетрія: Навч.-метод, посібник для самост. вивч. дисц. – К.: КНЕУ. 2001. – 192 с.

4. Мороз В. С., Мороз В. В. Економетрія: Навч. посібник. – Хмельницький: ТУП, 2000. – 166с.

5. Корольов О.А. Економетрія: Лекції, питання, тести, залачі, ситуації, проблеми: Навч. посіб. – К.: КДТЕУ. 2000.

6. Кулинич О.І. Економетрія. Навчальний посібник. – Хм.: Видавництво «Поділля», 1997. – 115 с.

7. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Підручник. – К.: Товариство «Знання», КОО. 1998. – 494 с.

8. Марюта А. Н., Бойцун Н. Е. Статистические методы и модели в экономике. Монография. – Дніпропетровськ: Пороги, 2002. – 384с.

Інтернет сайти:

1. http://stud.com.ua/9254/ekonomika/ekonomiko-matematichni_metodi_i _ prikladni _modeli - Прикладні економіко-математичні моделі

2. http://www.uabs.edu.ua/images/stories/docs/K_F/Yepifanov_16.pdf – Сучасні та перспективні методи і моделі управління в економіці. Монографія.

3. ekhnuir.univer.kharkov.ua/handle/123456789/9599 - Моделювання світо господарських процесів: Підручник.

4.Теоретичні основи кількісних методів моделювання та прогнозування економічних процесів // http://bookss.co.ua/book_medoti-ekonomyko-statestichnih-doslidzhen_806/3_1.-teoretichn-osnovi-klksnih-metodv-modelyuvannya-ta-prognozuvannya-ekonomchnih-procesv.

5. Державний комітет статистики України – www. ukrstat. gov.ua

Навчальне обладнання, ТЗН, презентація тощо: ноутбук, проектор, мультимедійна презентація.

ВИКЛАД МАТЕРІАЛУ ЛЕКЦІЇ

1. Економетрична модель та її види.

При аналізі економічних явищ на основі економіко-математичних методів особливе місце займають моделі, що виявляють кількісні зв'язки між досліджуваними показниками і впливаючими на них чинниками. Науковою дисципліною, предметом якої є вивчення цієї кількісної сторони економічних явищ і процесів засобами математичного і статистичного аналізу, економетрика, в якій результати теоретичного аналізу економіки синтезуються з висновками математики і статистики. Основне завдання економетрики - перевірка економічних теорій на фактичному (емпіричному) матеріалі за допомогою методів математичної статистики.

Головним інструментом економетрики служить ***економетрична модель,*** тобто економіко-математична модель факторного аналізу, параметри якої оцінюються засобами математичної статистики. Ця модель виступає в якості засобу аналізу і прогнозування конкретних економічних процесів на основі реальної статистичної інформації.

Економетричні моделі включають достатньо широкий клас різноманітних економіко-математичних моделей. Приведемо одну із класифікацій економетричних моделей.

За засобом математичної формалізації економетричні моделі можна умовно розділити на прості та складні.

Прості економетричні моделі зображені одним рівнянням, однією залежністю, складні – декількама рівняннями, залежностями:

$$У = ax^{в} + \varepsilon; \quad (1)$$

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \varepsilon_{1,} \quad (2) \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \varepsilon_{2.} \end{aligned}$$

За кількістю факторів, що включаються в модель, прості економетричні моделі можна розділити на однофакторні та багатофакторні. Однофакторні моделі містять одну незалежну змінну, багатофакторні моделі – ряд незалежних змінних.

Однофакторні і багатофакторні моделі можуть бути зображені лінійними та нелінійними функціями.

Приклад нелінійної функції. Виробнича функція Кобба-Дугласа:

$$У = AK^{\alpha}L^{\beta} + \varepsilon,$$

де У – об’єм виробництва, К – витрати капіталу, L – витрати праці.

Складні економетричні моделі можуть бути зображені трьома видами систем одночасних рівнянь у залежності від форми включення у праву частину ендогенних змінних. Зазвичай виділяють три типи систем :

- системи, що розв’язуються відносно ендогенних змінних;
- рекурсивні системи;
- системи, що не розв’язуються відносно ендогенних змінних.

Особливістю систем одночасних рівнянь є те, що кожне з рівнянь системи, окрім «своїх» пояснюючих змінних, може включати пояснюючі змінні із інших рівнянь. Класичним прикладом такої системи являється модель попиту Q^d та пропозиції Q^l, відколи попит на товар визначається його ціною Р та доходом споживача І, пропозиція товару - його ціною Р, і досягається рівновага між попитом та пропозицією:

$$\begin{aligned} &Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1; \\ &Q^l = \beta_4 + \beta_5 I + \varepsilon_2; \\ &Q^d = Q^l \end{aligned} \qquad (3)$$

У цьому випадку спостережуване значення Р – це ціна рівноваги, яка формується одночасно із попитом та пропозицією. Таким чином, Q^d, Q^l,P – пояснювані змінні, І – пояснююча змінна.

Перші три змінні формують свої значення відповідно рівнянням (3), тобто усередині моделі. Такі змінні звуться ендогенними. Між тим змінна І вважається у рівняннях (3) заданою зовні, її значення формуються поза моделлю. Такі змінні звуться екзогенними.

У залежності від наявності (відмінності) у моделі фактору часу розрізняють динамічні та статичні моделі. Динамічні моделі: трендові моделі; моделі улагоджування нових рядів; моделі декомпозиції часового ряду і авторегресійні моделі, моделі ковзної середньої, лагові та регресійні динамічні моделі.

2. Побудова та аналіз економетричної моделі з двома змінними.

З курсу математики ви знайомі з функціональною залежністю, коли кожному значенню однієї змінної відповідає визначене значення іншої.

В економіці у багатьох випадках між змінними існують залежності, коли кожному значенню однієї змінної відповідає не деяке певне, а множина можливих значень іншої змінної. Інакше кажучи, кожному значенню однієї змінної відповідає певний (умовний) розподіл іншої змінної. Така залежність отримала назву статистичної (або імовірної).

Виникнення поняття статистичної залежності обумовлюється тим, що залежна змінна підпадає під вплив неконтрольованих або неврахованих факторів, а також тим, що вимірювання значень змінних неминуче супроводжується певними випадковими похибками. В силу невизначеності статистичної залежності між ***X*** та ***Y*** для дослідження представляє інтерес усереднена по ***X*** схема залежності. Тобто закономірність у вимірюванні умовного математичного сподівання $M_x(y)$.

Якщо залежність між двома змінними така, що кожному значенню однієї змінної відповідає певне умовне математичне сподівання іншої, то така статистична залежність називається кореляційною:

$$M_x(y)=F(x).$$

У регресійному аналізі розглядаються залежність випадкової змінної ***Y*** від однієї (або декількох) невипадкової незалежної змінної ***X***. Така залежність може виникнути у випадку, коли при кожному значенні змінної ***X*** відповідні значення ***Y*** підпадають під вплив неконтрольованих факторів. Така залежність ***Y*** від ***X*** (іноді її називають регресійною) також може бути представлена у вигляді модельного рівняння рівняння регресії або економетричної моделі.

Y - функція відгуку, пояснювальна, ендогенна, результативна ознака, вихідна ознака; ***X*** - пояснююча, екзогенна, предикторна, фактор, регресор, факторна ознака.

Найпростішою у використанні є парна регресія. Передумови застосування парної регресійної моделі є наступними. По-перше, її застосовують для вивчення, опису, формалізованого уявлення і оцінки такої залежності, яка виникає в процесі взаємодії всього ***двох ознак*** , або двох змінних.

По-друге, кожна з ознак грає цілком визначену роль у формуванні досліджуваної залежності, тобто розглядається процес, в якому можна точно вказати, яка з ознак є причиною виникнення інші, а яка - результатом зазначеного впливу. Ознака-причина (фактор), змінюючись, викликає зміни ознаки-результату, впливає на нього і формує його значення. Традиційно ознаку-причину позначають через ***x,*** а сформований нею зв'язок розглядають як причинно-наслідковий. Ознака-результат відчуває на собі вплив факторного чинника, залежить від цього впливу, її значення формуються під впливом фактора. Ознаку-результат; його, як правило, позначають ***y.***

По-третє, парну регресійну модель застосовують в тому випадку, коли вивчають залежності, які проявляються лише в масі однорідних подій, явищ, носять імовірнісний характер і мають форму не строго, однозначно сформульованої, яка завжди спостерігається, закономірності, а простежуються в формі найбільш імовірної паралельності змін фактора і результату. Подібні залежності визначаються як стохастичні, ймовірнісні, а для їх вивчення використовують методи теорії ймовірностей і математичної статистики.

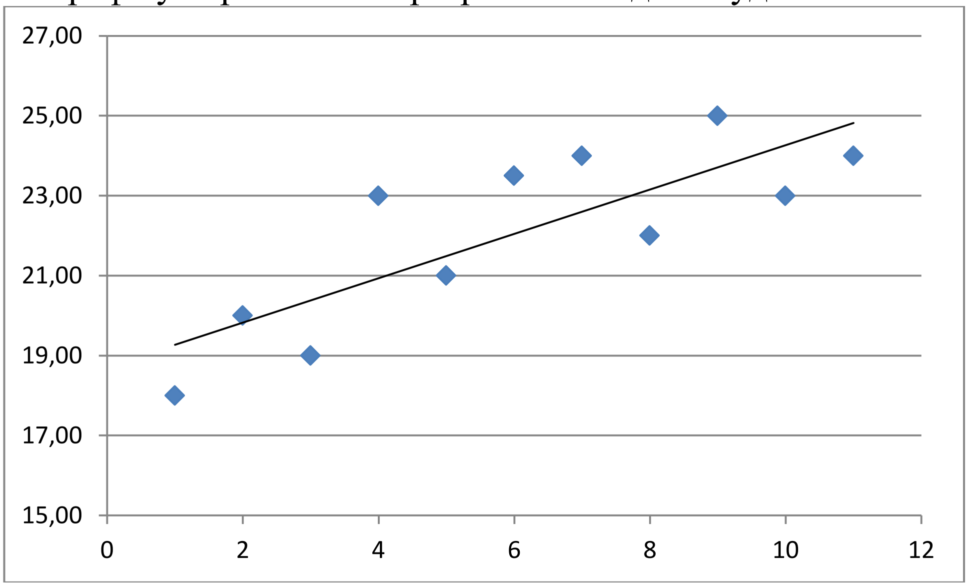
Найбільш простою і поширеною є парна лінійна регресія. Якщо за розташуванням точок даних, що характеризують певний економічний процес чи явище, можна припустити наявність лінійної регресіїної моделі, то рівняння регресії може бути записане у вигляді:

$$Y=\beta_0+\beta_1 X+\varepsilon, \qquad (4)$$

або

$$y_i=\beta_0+\beta_1 x_i+\varepsilon_i, i=\overline{1,n},$$

На графіку парна лінійна регресійна модель буде мати вигляд:



Рівняння регресії шукається у вигляді лінійного рівняння

$\hat{y}=b_0+b_1x$, (5)

де $\hat{y}$ - це оцінка $M_x(y)$, b_0 - оцінка β_0, b_1 - оцінка β_1.

3.Сутність методу найменших квадратів

Для знаходження параметрів економетричної моделі, може бути використаний метод найменших квадратів.

Згідно методу найменших квадратів (1МНК) невідомі параметри b_0 та b_1 обираються таким чином, щоб сума квадратів відхилень емпіричних значень y_i від теоретичних значень $\hat{y}_i$ була найменшою:

$$S=\sum_{i=1}^{n}(\hat{y}_i-y_i)^2=\sum_{i=1}^{n}(b_0+b_1x_i-y_i)\rightarrow \min$$

За необхідними умовами екстремуму:

$$\begin{cases}\dfrac{\partial s}{\partial b_0}=2\sum_{i=1}^{n}(b_0+b_1x_i-y_i)=0,\\[2mm] \dfrac{\partial s}{\partial b_1}=2\sum_{i=1}^{n}(b_0+b_1x_i-y_i)x_i=0.\end{cases}$$

Звідси після перетворень отримаємо систему нормальних рівнянь:

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ b_0 \sum_{i=1}^{n} x_i + b_0 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i; \end{cases}$$

$$\begin{cases} b_0 + b_1 \overline{x} = \overline{y}, \\ b_0 \overline{x} + b_1 \overline{x^2} = \overline{xy}; \end{cases}$$

де

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}, \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n}$$

Після розв'язання останньої системи отримуємо коефіцієнт регресії ***Y*** по ***X***:

$$b_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}. \qquad (5)$$

Коефіцієнт регресії ***Y*** по ***X*** показує, наскільки одиниць у середньому змінюються змінна ***Y*** при збільшені змінної ***X*** на одну одиницю.

Формула обчислення параметру b_0 (вільного члена):

$$b_0 = \overline{y} - \overline{b_1}\overline{x}. \qquad (6)$$

Для оцінки щільності кореляційного зв'язку використовується **коефіцієнт кореляції**:

$$r = \frac{\overline{xy} - \overline{x}\,\overline{y}}{S_x S_y},$$

$$S_x = \sqrt{\overline{x}^2 - \overline{x}^2}, S_y = \sqrt{\overline{y}^2 - \overline{y}^2}, \qquad (7)$$

де S_x, S_y - середньоквадратичні відхилення.

<u>Властивості коефіцієнта кореляції:</u>

1. Коефіцієнт кореляції приймає значення на відрізку ***[-1;1]***, тобто $-1 \le r \le 1$. Чим ближче $|r|$ до одиниці, тим тісніше зв'язок.

Дві кореляційні залежності наведені на рис. 1. Очевидно, що у випадку ***а*** залежність між змінними менш щільна, і коефіцієнт кореляції повинен бути менший, ніж у випадку ***б***, так як точки кореляційного поля ***а*** подальш відстоять від лінії регресії, ніж точки поля ***б***.

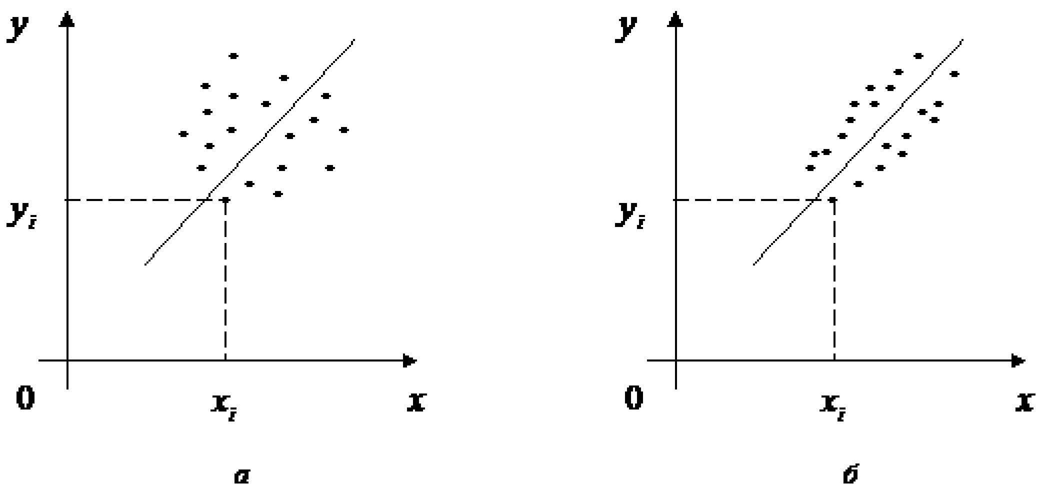


Рис. 1.

2. Якщо $r > 0 (b_1 > 0)$ то кореляційний зв'язок прямий (рис. 2.а), якщо $r < 0 (b_1 < 0)$, - обернений (рис. 2.б).

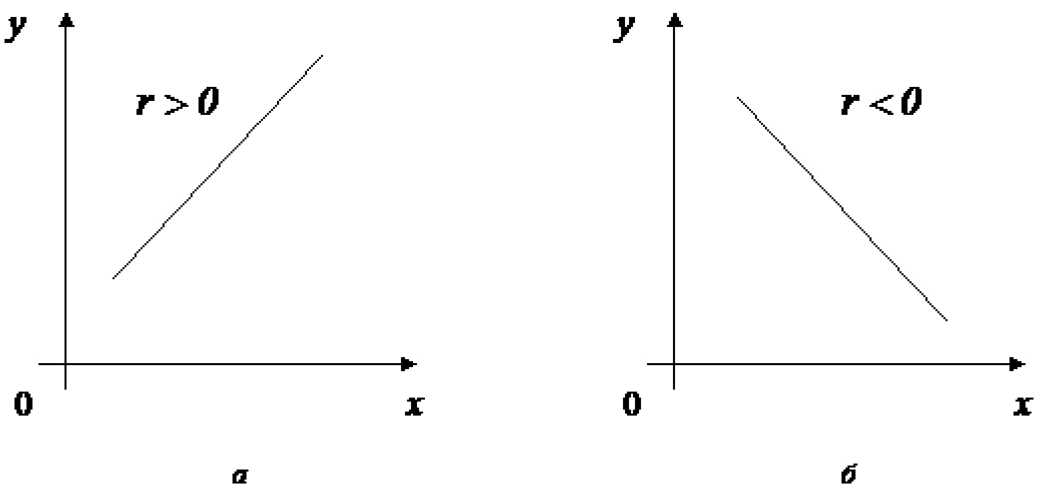


Рис. 2.

3. При $r = \pm 1$ кореляційна залежність являється лінійною функціональною залежністю. При цьому усі значення, що спостерігаються, розташовані на прямій лінії (рис. 3).

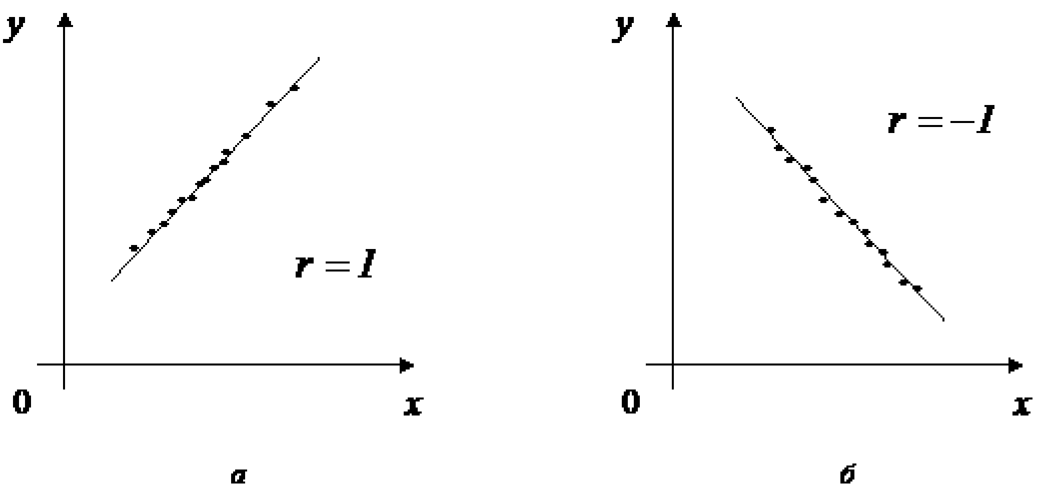


Рис. 3.

4. При $r=0$ лінійна кореляційна залежність відсутня. Це означає або відсутність будь-якої залежності між змінними x та y (рис. 4.а), або належність до певної нелінійної залежності (рис. 4.б).

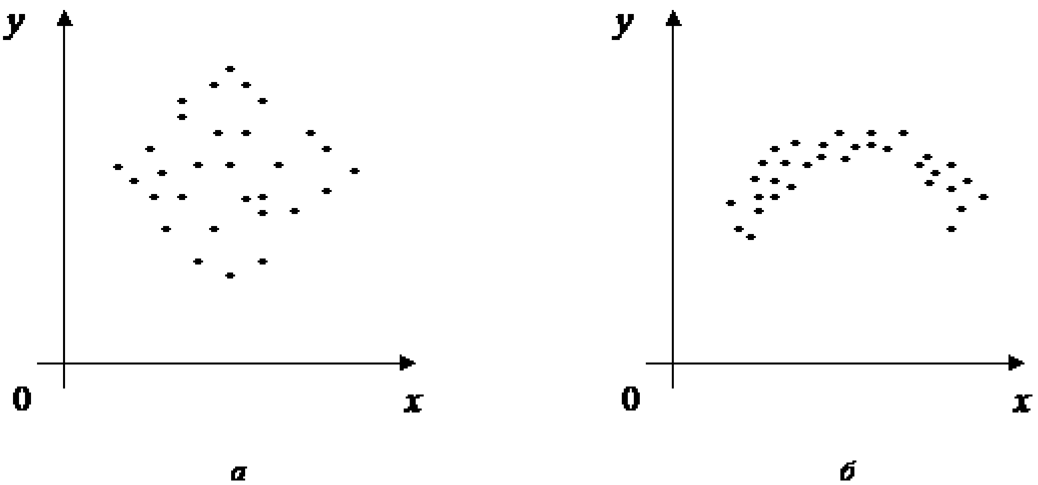


Рис. 4.

На практиці для оцінки ступені взаємозв'язку можна керуватись наступними емпіричними правилами:

1) $|r|>0{,}95$ - існує практично лінійна залежність;

2) $0{,}8<|r|<0{,}95$ - сильна ступінь лінійної залежності;

3) $0{,}6<|r|<0{,}8$ - належність лінійного зв'язку;

4) $|r|<0{,}4$ - лінійний зв'язок виявити не вдалося.

Оцінка адекватності регресійної моделі. Коефіцієнт детермінації.

Оцінка адекватності регресійної моделі робиться на підставі коефіцієнта детермінації:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}. \qquad (8)$$

Величина R^2 показує, яка частка варіації залежної змінної обумовлена варіацією фактора.

<u>Властивості коефіцієнта детермінації:</u>

1. Для ЛПР $R^2=r^2$.
2. Коефіцієнт детермінації приймає значення на відрізку *[0;1]*, тобто $0 \le R^2 \le 1$. Чим ближче R^2 до одиниці, тим краще регресія апроксимує емпіричні дані.
3. Якщо $R^2=1$, між змінними x та y існує лінійна функціональна залежність.
4. Якщо $R^2=0$, то варіація залежної змінної повністю обумовлена впливом випадкових та неврахованих у моделі змінних.

На практиці **для оцінки ступені апроксимації рівнянням регресії вихідних даних** використовують наступні емпіричні правила:

1). $R^2>0,95$ - висока точність апроксимації.
2). $0,8<R^2<0,95$ - задовільна апроксимація.
3). $R^2<0,6$ - незадовільна апроксимація.

Обчислення коефіцієнта еластичності:

Коефіцієнт еластичності E показує - наскільки відсотків (від середньої) змінюється у середньому y при зміненні x на *1%* та обчислюється за формулою:

$$E = b_1 \frac{\bar{x}}{\bar{y}}.$$

Перевірити значущість рівняння регресії - означає встановити, чи відповідає математична модель, що виражає залежність між змінними, експериментальним даним, чи достатньо залучених у рівняння факторів (одного або декількох) для опису залежної змінної.

Для оцінки значущості рівняння регресії використовується **F-тест**. Для цього виконується зрівняння фактичного $F_{факт}$ та критичного (табличного) $F_{табл}$ значення **F-критерію Фішера**.

$$F_{факт} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}(n-2) = \frac{r^2}{1-r^2}(n-2). \qquad (9)$$

$F_{табл}=F_{\alpha,k1,k2}$ - це максимально можливе значення критерію під впливом випадкових факторів при даних ступенях свободи $k_1=m-1$ і $k_1=m-n$ і рівні значущості α, де m - кількість параметрів, що оцінюються (для ЛПР $m=2$, так як оцінюються параметри b_0 та b_1), n - кількість спостережень, α - зазвичай приймається 0,05 (в економіці) або 0,01.

Знайти $F_{табл}$ можна у таблицях **F-розподілу Фішера-Снедекора** або за функцією **"FРАСП"** в MS Excel.

F-тест. Якщо $\mathbf{F_{факт}}>\mathbf{F_{табл}}$, рівняння регресії статистично значуще на рівні значущості $\boldsymbol{\alpha}$. Якщо $\mathbf{F_{факт}}<\mathbf{F_{табл}}$, то признається статистична незначущість рівняння регресії.

Для оцінки значущості коефіцієнтів регресії $\mathbf{b_0}$, $\mathbf{b_1}$ використовується **t-тест**. Для цього зрівнюються фактичне $\mathbf{t_{факт}}$ та критичне (табличне)$\mathbf{t_{табл}}$ значення **t-критерія** Стьюдента. $\mathbf{t_{факт}}$ для коефіцієнтів $\mathbf{b_0}$, $\mathbf{b_1}$ визначається за наступними формулами:

$$t_{b_0} = \frac{|b_0|}{S}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\overline{x^2}}}. \quad (10)$$

$\mathbf{t_{табл}=t_{1-\alpha;n-2}}$ - це максимально можливе значення критерію під впливом випадкових факторів при **n-2** ступенях свободи і рівні значущості $\boldsymbol{\alpha}$. Значення $\mathbf{t_{табл}}$ міститься у таблицях **t-розподілу** Стьюдента або визначається за функцією **"СТЬЮДРАСПОБР»** в MS Excel.

Якщо $\mathbf{t_{bi}>t_{1-\alpha;n-2}}$, коефіцієнт $\mathbf{b_i}$ – статистично значущий на рівні значущості $\boldsymbol{\alpha}$. Якщо $t_{bi}<t_{1-\alpha;n-2}$, то признається статистична незначущість.

3. Основні етапи економетричного дослідження

Можна виділити наступні основні етапи економетричного дослідження:
1) постановка проблеми;
2) отримання даних та аналіз їх якості;
3) специфікація моделі;
4) оцінка параметрів (ідентифікація) моделі;
5) перевірка адекватності (верифікація) моделі;
6) інтерпретація результатів.

1-й етап: формується мета дослідження, сукупність економічних змінних. За вибором економічних змінних необхідне обґрунтування кожної змінної (при цьому, рекомендується, щоб їх кількість не була великою).

2-й етап: здійснюється збір необхідної статистичної інформації – спостережливі значення економічних змінних:

$(x_{i1},x_{i2},\ldots,x_{ip};\ y_{i1},y_{i2},\ldots,y_{iq})$, $i = 1,n$, та їх аналіз.

3-й етап: здійснюється вибір загального типу моделі: вираз у математичній формі виявлених зв’язків та співвідношень; встановлення складу екзогенних та ендогенних змінних; формулювання вихідних посилань та обмежень моделі.

4-й етап: здійснюється статистичний аналіз моделі та оцінка її параметрів.

5-й етап: проводиться перевірка істинності, адекватності моделі. З’ясовується, наскільки відповідає побудована модель реальному економічному об’єкту або процесу.

6-й етап: економічна інтерпретація та практичне використання моделі.

Загальний висновок за темою лекції:

У лекції розглянуто концептуальні аспекти економетричного моделювання економічних процесів, особливості та принципи парної лінійної регресійної моделі, застосування методу найменших квадратів для оцінки параметрів лінійної регресії, етапи економетричного моделювання.

Питання для самоконтролю:
1. Дайте визначення економетричної моделі.
2. Назвіть основні види економетричних моделей.
3. Визначте відмінність між функціональною та кореляційною залежністю.
4. Опишіть характерні можливості використання парної регресії в економічних дослідженнях.
5. Охарактеризуйте зміст параметрів лінійної регресійної моделі.
6. Охарактеризуйте суть методу найменших квадратів.
7. Назвіть етапи екнометричного моделювання.

Укладачі: ____________ Стадник Ю.А., доцент, Мищишин О.Я., доцент
(підпис) (ПІБ, посада, науковий ступінь, вчене звання)

Конспект лекції №5

Тема 5. Моделі множинної регресії. Застосування нелінійних функцій.

Міжпредметні зв'язки: Зв'язок із елементами знань і умінь таких навчальних дисциплін як «Макроекономіка», «Мікроекономіка», «Інформатика», «Математика для економістів».

Мета лекції полягає у формуванні полягає у формуванні в студентів системи теоретичних знань та практичних навичок з питань побудови множинної лінійної регресійної моделі, нелінійних економетричних моделей та їх застосування для прогнозування.

План лекції
1. Класична лінійна модель множинної регресії, основні припущення.
2. Оцінка параметрів множинної моделі та перевірка її на адекватність.
3. Прогнозування на основі множинної лінійної регресійної моделі.
4. Побудова нелінійних економетричних моделей.

Опорні поняття: математичні методи, модель, множинна регресія, адекватність регресійної моделі, метод найменших квадратів, прогноз, нелінійна регресійна модель.

Інформаційні джерела:

Основна та допоміжна література:

1. Васильків І. М., Карпінський Б. А., Максимук О. В., Шкулка С. К. Вступ до економетрики: Навч. посіб. – Львів: Львівський національний університет ім. І. Франка, 2015. – 280 с.

2. Сингаевская Г. И. Функции в Excel. Решение практических задач. М.: Издательский дом «Вильямс», 2009. – 880 с.

3. Наконечний С. І.. Терещенко Т.О. Економетрія: Навч.-метод, посібник для самост. вивч. дисц. – К.: КНЕУ. 2001. – 192 с.

4. Мороз В. С., Мороз В. В. Економетрія: Навч. посібник. – Хмельницький: ТУП, 2000. – 166с.

5. Корольов О.А. Економетрія: Лекції, питання, тести, залачі, ситуації, проблеми: Навч. посіб. – К.: КДТЕУ. 2000.

6. Кулинич О.І. Економетрія. Навчальний посібник. – Хм.: Видавництво «Поділля», 1997. – 115 с.

7. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Підручник. – К.: Товариство «Знання», КОО. 1998. – 494 с.

8. Марюта А. Н., Бойцун Н. Е. Статистические методы и модели в экономике. Монография. – Дніпропетровськ: Пороги, 2002. – 384с.

Інтернет сайти:

1. http://stud.com.ua/9254/ekonomika/ekonomiko-matematichni_metodi_i _ prikladni _modeli - Прикладні економіко-математичні моделі

2. http://www.uabs.edu.ua/images/stories/docs/K_F/Yepifanov_16.pdf – Сучасні та перспективні методи і моделі управління в економіці. Монографія.

3. ekhnuir.univer.kharkov.ua/handle/123456789/9599 - Моделювання світо господарських процесів: Підручник.

4.Теоретичні основи кількісних методів моделювання та прогнозування економічних процесів // http://bookss.co.ua/book_medoti-ekonomyko-statestichnih-doslidzhen_806/3_1.-teoretichn-osnovi-klksnih-metodv-modelyuvannya-ta-prognozuvannya-ekonomchnih-procesv.

5.Державний комітет статистики України – www. ukrstat. gov.ua

Навчальне обладнання, ТЗН, презентація тощо: ноутбук, проектор, мультимедійна презентація.

ВИКЛАД МАТЕРІАЛУ ЛЕКЦІЇ

1. Класична нормальна лінійна модель множинної регресії, основні припущення.

Економічні явища, як правило, визначаються більш ніж одним з одночасно та сукупно діючих факторів. У зв'язку з цим виникає задача дослідження залежності однієї залежної змінної У від декількох пояснюючих змінних $X_1, X_2, \ldots, X_p$. Ця задача вирішується за допомогою множинного регресійного аналізу. Множинна регресія широко використовується при рішенні питань попиту, доходності акцій, при вивченні витрат виробництва, у макроекономічних розрахунках тощо.

Загальна множинна регресійна модель має наступний вигляд:

$$y = f(x_1, x_2, \ldots, x_p) + \varepsilon, \quad (1)$$

де y - залежна змінна;

$x_1, x_2, \ldots, x_p$ - фактори (незалежні змінні).

Якщо множинна регресійна модель є лінійною (ЛМР), то вона подається у вигляді:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon. \quad (2)$$

Позначимо i-е спостереження змінної y через y_i, а факторів – $x_{i1}, x_{i2}, \ldots, x_{ip}$. Відтоді модель (2) можна подати у вигляді:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i, \ i = \overline{1,n}, \quad (3)$$

або у матричній формі:

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

де $y = [y_1, y_2, \ldots, y_n]^T$ - вектор (матриця-стовпець) значень залежної змінної;

$\beta = [\beta_1, \beta_2, \ldots, \beta_p]^T$ - вектор (матриця-стовпець) коефіцієнтів регресійної моделі;

$\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]^T$ - вектор (матриця-стовпець) похибок;

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ - & - & - & \ldots & - \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}$$ - матриця значень факторів.

Відмітимо основні припущення регресійного аналізу:

1. В моделі (3) похибка ε_i (або залежна змінна y_i) є випадковою величиною, а фактори x_{ip} невипадкові величини ($i = \overline{1,n}$).

2. Математичне сподівання похибки ε_i дорівнює нулю:

$$M[\varepsilon_i] = 0, \ i = \overline{1,n}.$$

3. Дисперсія похибки ε_i (або залежної змінної y_i) постійна дія будь-якої i:

$$D[\varepsilon_i] = \sigma^2.$$

тобто виконується умова гомоскедастичності.

4. Похибки ε_i та ε_j не корельовані:

$$M[\varepsilon_i \varepsilon_j] = 0, \ i \neq j.$$

5. Похибка ε_i (або залежна змінна y_i) є нормально розподіленою випадковою величиною.

6. Матриця значень факторів невироджена, тобто її ранг дорівнює $p+1$:

$$rangX = p + 1 < n.$$

Модель, для якої виконуються припущення 1-6, називається класичною нормальною лінійною моделлю множинної регресії (*CNLMR*-model).

2. Оцінка параметрів множинної моделі та перевірка її на адекватність.

Оцінкою цієї моделі за вибіркою є рівняння регресії:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p, \qquad (4)$$

де $\hat{y}$ - оцінка математичного сподівання залежної змінної $M_x[y]$;

$b_i\left(i=\overline{0,p}\right)$ - оцінка коефіцієнтів $\beta_i\left(i=\overline{0,p}\right)$ регресійної моделі (або коефіцієнти регресії).

Як і раніше, для оцінки коефіцієнтів *CNLMR*-model використовують МНК:

$$S(b_0, b_1, \ldots, b_p) = \sum_{i=1}^{n}\left(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - \ldots - b_px_{ip}\right)^2 \to \min.$$

Після розв'язання системи нормальних рівнянь

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum\limits_{i=1}^{n}\left(y_i - b_0 - b_1x_{i1} - \ldots - b_px_{ip}\right) = 0, \\ \dfrac{\partial S}{\partial b_1} = -2\sum\limits_{i=1}^{n}x_{i1}\left(y_i - b_0 - b_1x_{i1} - \ldots - b_px_{ip}\right) = 0, \\ - \quad - \quad - \quad - \quad - \quad - \quad - \\ \dfrac{\partial S}{\partial b_p} = -2\sum\limits_{i=1}^{n}x_{ip}\left(y_i - b_0 - b_1x_{i1} - \ldots - b_px_{ip}\right) = 0, \end{cases}$$

отримаємо значення коефіцієнтів рівняння регресії, які в матричній формі мають вигляд:

$$b = \left(X^TX\right)^{-1}X^Ty, \qquad (5)$$

де $b = \left[b_1, b_2, \ldots, b_p\right]^T$ - вектор (матриця-стовпець) коефіцієнтів рівняння регресії.

Оцінки b_j є незміщеними, обґрунтованими та ефективними.

Оцінка дисперсії похибок

$$S^2 = \frac{\sum\limits_{i}^{n}\left(y_i - \hat{y}_i\right)^2}{n-p-1} \qquad (6)$$

є незміщеною та обґрунтованою.

<u>Коефіцієнт (індекс) множинної кореляції *R*</u> використовується для оцінки тісноти спільного впливу факторів на залежну змінну:

$$R = \sqrt{1 - \frac{\sum\limits_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum\limits_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum\limits_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}. \qquad (7)$$

<u>Властивості коефіцієнта множинної кореляції *R*:</u>

1. Коефіцієнт множинної кореляції приймає значення на відрізку $[0;1]$, тобто $0 \le R \le 1$.

Чим ближче R до одиниці, тим тісніше зв'язок між залежною y та факторами $x_1, x_2, \ldots, x_p$.

2. При $R=1$ кореляційний зв'язок є лінійною функціональною залежністю.

3. При $R=0$ лінійний кореляційний зв'язок відсутній.

Щодо оцінки ступеня взаємозв'язку, можна керуватись аналогічними емпіричними правилами, як і для випадку ЛПР (лекція 3.1).

Коефіцієнт (індекс) множинної детермінації R^2

Для оцінки адекватності регресії моделі, мірою якості рівняння регресії використовують коефіцієнт детермінації, який визначається, як і раніше, за формулою:

$$R^2 = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} . \qquad (12)$$

Нагадаємо, що R^2 характеризує частку варіації залежної змінної, що обумовлена варіаціями факторів.

Властивості коефіцієнта множинної детермінації R^2:

1. Коефіцієнт множинної детермінації приймає значення на відрізку $[0;1]$, тобто $0 \le R^2 \le 1$.

Чим ближче R^2 до одиниці, тим краще регресія апроксимує емпіричні дані.

2. Якщо $R^2=1$, між змінними y та $x_1, x_2, \ldots, x_p$ існує лінійна функціональна залежність.

3. Якщо $R^2=0$, то варіація залежної змінної повністю обумовлена виливом випадкових та неврахованих факторів.

Для оцінки ступеня апроксимації емпіричних даних рівнянням ЛМР можна керуватись аналогічними емпіричними правилами, як і для випадку ЛПР (лекція 3.1).

Зауваження

Недоліком коефіцієнта множинної детермінації R^2 являється те, що він, взагалі, збільшується при додаванні нових факторів, хоча це не обов'язково означає поліпшення якості регресійної моделі. Тому має сенс використовувати скоригований (адаптований, виправлений) коефіцієнт детермінації $\mathbf{R}^2$, який визначається за формулою:

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}\left(1 - R^2\right) . \qquad (13)$$

На відміну від R^2 скоригований коефіцієнт $\mathbf{R}^2$ може зменшуватись при введенні у модель нових факторів, які не чинять істотного впливу на залежну змінну.

Оцінка значущості ЛМР.

Значущість рівня ЛМР у цілому оцінюється за допомогою *F*-критерія Фішера

$$F_{факт} = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p} \qquad (8)$$

порівнянням його з табличним значенням

$$F_{табл} = F_{\alpha,p,n-p-1}. \qquad (9)$$

F-тест.

Якщо $F_{факт} > F_{\alpha,p,n-p-1}$, то рівняння ЛМР признається статистично значущим на рівні значущості α (зазвичай, $\alpha = 0{,}05$).

Якщо $F_{факт} < F_{\alpha,p,n-p-1}$, то рівняння ЛМР признається статистична незначущість ЛМР на рівні значущості α.

Другий варіант *F*- тесту: якщо рівень значущості фактичного *F*-критерію $\alpha_F < \alpha$, то рівняння ЛМР – статистичного значуще на рівні значущості α.

Якщо $\alpha_F > \alpha$, то ЛМР – статистичного незначуще на рівні значущості α.

Оцінка значущості коефіцієнтів рівняння ЛМР.

Оцінка значущості коефіцієнтів рівняння ЛМР здійснюються за допомогою *t*-критерію Ст'юдента:

$$t_{b_j} = \frac{|b_j|}{S_{b_j}} \qquad (10)$$

із зрівнянням його з табличним значенням

$$t_{табл} = t_{1-\alpha,n-p-1}, \qquad (11)$$

де $S_{b_j} = S\sqrt{\left(X^T X\right)^{-1}_{jj}} \left(j = \overline{0,p}\right)$ - середньоквадратичне відхилення (стандартна похибка) коефіцієнт а регресії b_j;

$S = \frac{\sum_i^n \left(y_i - \hat{y}_i\right)^2}{n-p-1}$ - оцінка середньоквадратичного відхилення похибок;

$\left(X^T X\right)^{-1}_{jj} \left(j = \overline{0,p}\right)$ - відповідний діагональний елемент матриці $\left(X^T X\right)^{-1}$.

t-тест.

Якщо $t_{b_j} > t_{1-\alpha,n-p-1}$, то коефіцієнт b_j признається статистично значущим; якщо $t_{b_j} < t_{1-\alpha,n-p-1}$, то b_j – статистично незначущий на рівні значущості α.

3.Прогнозування на основі множинної лінійної регресійної моделі.

Прогнозне значення $\hat{y}_П$ визначається за шляхом підстановки у рівняння регресії (4) відповідних значень факторів $x_{П1}, x_{П2}, \ldots, x_{Пp}$:

$$\hat{y}_П = \hat{y}(x_{П1}, x_{П2}, \ldots, x_{Пp}) = b_0 + b_1 x_{П1} + b_2 x_{П2} + \ldots + b_p x_{Пp}. \qquad (14)$$

Довірчий інтервал прогнозу обчислюється за слідуючими формулами:

$$M_{x_{П1}-x_{Пp}}[y] = \hat{y}_П \pm t_{1-\alpha;n-p-1} \cdot S_{\hat{y}}, \qquad (15)$$

де $M_{x_{П1}-x_{Пp}}[y]$ - умовне математичне сподівання залежної змінної в точці прогнозу;

$S_{\hat{y}}$ - оцінка стандартної похибки прогнозу, яка обчислюється за формулою

$$S_{\hat{y}} = S\sqrt{X_П^T (X^T X)^{-1} X_П}; \qquad (16)$$

X - матриця значень факторів;

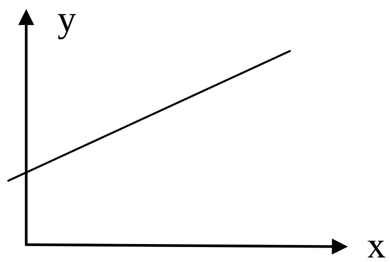
$X_П = [1\ x_{П1}\ x_{П2} \ldots x_{Пp}]^T$ - вектор (матриця-стовпець) значень факторів для прогнозу.

<u>Довірчі інтервали для коефіцієнтів регресійної моделі:</u>

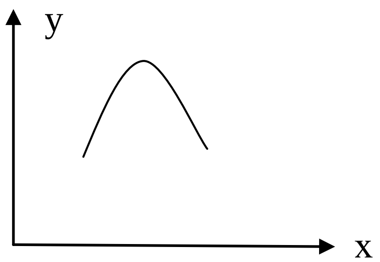
$$\beta_j = b_j \pm t_{1-\alpha;n-p-1} \cdot S_{b_j}, \quad j = \overline{0,p}. \qquad (17)$$

4.Побудова нелінійних економетричних моделей.

Окрім, лінійної при побудові одно факторних регресійних моделей використовуються і інші складніші типи функцій. Розглянемо основні типи кривих, що використовуються при кількісній оцінці в однофакторних моделях:

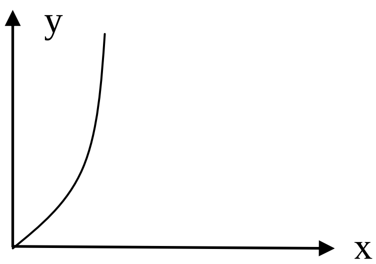


а

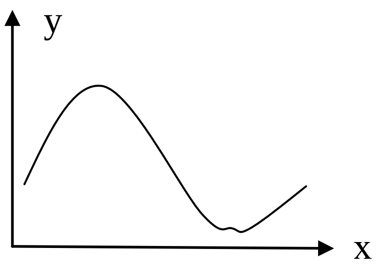


б

а)лінійна $y = \alpha+\beta x+\varepsilon$; б)поліноміальна $y = \alpha+\beta x+\gamma x^2 +\varepsilon$;

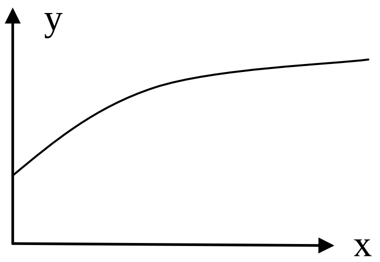


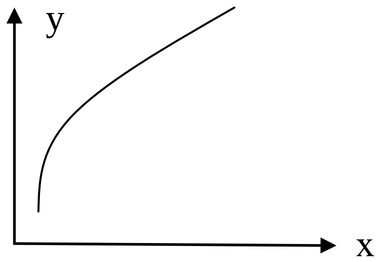
в



г

в)гіпербола $y = \alpha + \beta/x + \varepsilon$; г) $y = \alpha + \beta x + \gamma x^2 + \delta x^3 + \varepsilon$;





д е

д)степенева $y = \alpha x^{\beta}\, \varepsilon$; е)експоненціальна $y = \alpha \beta^{x}\, \varepsilon$.

Окрім розглянутих використовуються й інші типи кривих:

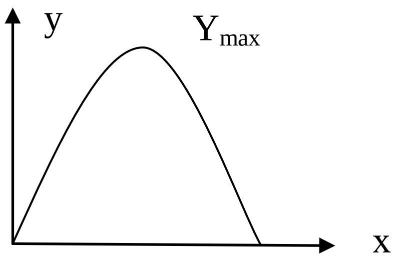
ж)$y = 1/(\alpha+\beta x)+\varepsilon$; з)$y = \alpha+\beta x+\gamma/x + \varepsilon$;

і)$y = \alpha+\beta \ell g x+\varepsilon$ – напівлогарифмічна;

ї)$y = 1/(\alpha+\beta x+\gamma x^2)$;

к)$y = \gamma/[1+\alpha EXP(-\beta x)]$ – логістична(використовується для опису виробництва нових товарів, росту чисельності населення);

л)$\ell g y = \alpha+\beta x+\gamma x^2$



м

м)Крива Лаффера ($y = a + вx + cx^2$, $c<0$) показує нелінійний зв'язок між податковою ставкою (на прибуток та заробітну плату) та надходжень від податків у бюджет.

Деякі з цих моделей можуть бути лінеаризовані (зведені до лінійних) і для знаходження їх параметрів застосовується метод найменших квадратів. Для інших моделей застосовуються складніші методи оцінки їх параметрів, наприклад метод трьох точок.

Основні типи багатофакторних економетричних моделей:

а)лінійна: $y = a + b_1x_1 + b_2x_2 + ... + b_px_p + \varepsilon$;

б)гіпербола $y = 1/(a + b_1x_1 + b_2x_2 + ... + b_px_p) + \varepsilon$;

в)степенева $y = a x_1^{в1} x_2^{в2} \dots x_p^{вp}\, \varepsilon$;

г)експоненціальна y = a EXP($b_1x_1 + b_2x_2 + ... + b_px_p$) .

Загальний висновок за темою лекції:

У лекції розглянуто класичну лінійну модель множинної регресії та методику оцінки її параметрів, показники, що використовуються для перевірки моделі на адекватність. Названі основні види нелінійних економетричних моделей, що використовуються у регресійному аналізі.

Питання для самоконтролю:

1. Дайте визначення класичної лінійної моделі множинної регресії.
2. Назвіть основні припущення регресійного аналізу.
3. Охарактеризуйте зміст параметрів лінійної моделі множинної регресії.
4. Дайте визначення коефіцієнта множинної кореляції.
5. Дайте визначення коефіцієнта множинної детермінації.
6. Для чого використовується *F*-критерій Фішера.
7. Для чого використовується *t*-критерій Стьюдента.
8. Що таке довірчі інтервали прогнозу?
9. Які нелінійні види функцій використовуються у регресійному аналізі?

Укладачі: ____________ Стадник Ю.А., доцент, Мищишин О.Я., доцент
(підпис) (ПІБ, посада, науковий ступінь, вчене звання)