**Лесько Ю. О. ЕП-33**

**5. МЕТОДИ ПОБУДОВИ ЗАГАЛЬНОЇ ЛІНІЙНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ**

***5.1. Характеристика основних методів побудови загальної лінійної економетричної моделі***

На кожнийекономічний показник впливає безліч різних факторів. При ідентифікації загальної лінійної економетричної моделі виникає питання, які саме з них потрібно увести до моделі у якості пояснюючих змінних. Для вирішення цього питання в принципі існує два наступні підходи.

**1-й підхід.** З точки зору надійності прогнозування, необхідно включати до моделі якомога більше факторів. Але при цьому, кожна „зайва” пояснююча змінна погіршує ситуацію ,оскільки вона зменшує надійність F-тесту на загальну статистичну значимість моделі і може призвести до невірних статистичних висновків.

**2-й підхід.** З точки зору отримання надійної статистичної інформації по кожному фактору слід прагнути, щоб модель мала якомога менше факторів, оскільки збір і обробка великих статичних масивів потребує великих затрат при недостатній надійності статистичних даних.

Компромісом між цими крайніми підходами є те, що називають вибором "найкращого рівняння" регресії. Для реалізації такого вибору немає єдиної статистичної процедури, єдиного методу. Існує декілька методів побудови "найкращої" лінійної регресії, найбільш поширеними серед яких є:

1. метод усіх можливих регресій ;
2. метод покрокової регресії ;
3. метод виключень .

**А) Метод усіх можливих регресій**

Метод усіх можливих регресій – історично перший метод побудови лінійних регресійних моделей і найбільш громіздкий серед усіх методів. Ідея методу полягає у побудові множини регресійних рівнянь, які містять усі можливі комбінації попередньо відібраних факторів, і у порівнянні цих рівнянь за трьома критеріями : коефіцієнтом детермінації R2, стандартною похибкою  і критерієм Меллоуза Ср. У загальному випадку для **m** відібраних факторів (пояснюючих змінних) можна побудувати 2m рівнянь регресії і виконати їх порівняння.

Побудова і аналіз усіх можливих регресійних рівнянь є доволі громіздка і ненадійна процедура, тому цей метод рекомендується використовувати при невеликій кількості відібраних факторів.

**Б) Метод покрокової регресії**

Цей метод є найпоширенішим на практиці і більш економним у порівнянні з попереднім. Ідея методу полягає у послідовному включенні до моделі факторів (пояснюючих змінних) до тих пір, поки модель не стане задовільною. Порядок включення факторів до моделі вибирається на основі значень коефіцієнтів парної кореляції між пояснюючими і залежною змінною моделі. Алгоритм методу покрокової кореляції можна подати у наступному вигляді :

***Алгоритм методу***

1. Розраховується кореляційна матриця **r** для усіх змінних моделі, які планується включити до моделі.
2. Спочатку з кореляційної матриці вибирається і включається до моделі той фактор , якому у кореляційній матриці відповідає найбільший за модулем коефіцієнт парної кореляції з залежною змінної моделі **у** (нехай це буде змінна х1). Будується регресійне рівняння з однією незалежною змінною  і для нього обчислюється коефіцієнт детермінації. Після цього перевіряється чи значима ця змінна за коефіцієнтом детермінації і за частковим F- критерієм. Якщо ні, то приймаємо  і процес побудови моделі припиняється. Якщо так, то переходимо до наступного кроку 3.
3. На основі аналізу кореляційної матриці серед тих пояснюючих змінних, що залишились, шукаємо нову змінну, яка має найбільший за модулем коефіцієнт кореляції з **у** і включаємо її до моделі (нехай це буде змінна х2) .
4. Будується нове рівняння регресії :



і для нього розраховується звичайний і оцінений коефіцієнт детермінації. Аналізується зміна цих показників у порівнянні з попередньою моделлю. Потім розраховуються часткові F- критерії для кожного фактора. Серед них обирається найменше значення і порівнюється із заздалегідь обраним критичним значенням F - критерію. В залежності від результатів перевірки додана на цьому кроці змінна або залишається у моделі, або відкидається.

1. Після цього модель перераховується в залежності від факторів, які залишились і здійснюється перехід до кроку 3.

Процес побудови моделі за наведеним алгоритмом припиняється, якщо жодний фактор, що знаходиться у поточному рівнянні , не вдається виключити, а новий претендент на включення не відповідає частковому F - критерію.

**В. Метод виключень**

Метод виключень діє у зворотному порядку порівняно з методом покрокової регресії і є також досить поширеним. Загальний алгоритм методу складається з 5 кроків.

1. Будується рівняння регресії, яке включає всі відібрані фактори . Якщо попередньо було відібрано m факторів , то вихідне базове рівняння має вигляд :

 .

1. Для кожного фактора (пояснюючої змінної)  обчислюється значення часткового F- критерію.
2. Серед розрахованих значень часткового F- критерію вибирається найменше Fmin і порівнюється із заздалегідь обраним критичним значенням розподілу Фішера Fкр .
3. Якщо Fmin < Fкр , то відповідний фактор виключається з рівняння. Проводиться новий розрахунок регресійного рівняння вже без виключеного фактора і здійснюється перехід знову до кроку 2.
4. Якщо Fmin > Fкр , то регресійне рівняння залишається без змін.

***5.2. Статистичні показники, які використовуються при побудові загальної лінійної економетричної моделі***

При розгляді методів побудови загальної лінійної економетричної моделі були використані такі нові поняття і статистичні показники, як **частковий** **F – критерій ,оцінений** **коефіцієнт** **детермінації** ,**кореляційна** **матриця**. Розглянемо ці показники більш детальніше.

**А. Частковий F- критерій.**

Одним з головних питань будь-якого методу побудови багатофакторної регресійної моделі є питання визначення суттєвості впливу на залежну змінну *у* окремих факторів. Таку оцінку можна зробити з використанням F- статистики на основі часткового F- критерію Фішера. Зміст часткового F- критерію розглянемо на наступному прикладі.

Нехай є економетрична модель ,яка враховує вплив **к** факторів на залежну змінну **у**, тобто :

 .

Припустимо, що з **к** факторів **p** факторів несуттєво впливають на показник **у**. Тоді побудуємо другу регресійну модель, в яку не включаємо ці **р** факторів.

 .

Позначимо суму квадратів залишків 1-ї моделі через **SSE1**, а 2-ї моделі - через **SSE2** . Тоді різниця **SSE1 - SSE2** дорівнює *додатковій* сумі квадратів залишків, яка пов'язана з включенням (або вилученням) до 1-ї моделі **p** додаткових факторів. Зазначимо, що ця додаткова сума квадратів буде мати ступінь вільності **p = k – q***.*

Знайдемо наступне розрахункове значення F- статистики :

. ( 62 )

Для заданого рівня значимості **α** і ступенів вільності  і за статистичними таблицями F- розподілу знаходимо критичне значення критерію Фішера Fкр*.* Якщо Fp,n-k > Fкр,то із надійністю 1 - α можна вважати, що вилучені фактори суттєво впливають на результуючий показник **у** і їх потрібно залишити у складі пояснюючих змінних моделі. У протилежному випадку (Fp,n-k < Fкр.)з надійністю 1-α можна вважати, що виключення із моделі **р** факторів несуттєво впливає на показник **у** і для моделювання можна вибрати другу модель з **q** пояснюючими змінними.

Якщо розглядати процес поступово включення (або вилучення) факторів до моделі, коли на кожному етапі до моделі включається (або вилучається) **тільки один фактор** за допомогою наведеного F- відношення ( 57 ) будемо мати критерій, який і визначатиме суттєвість впливу цього окремого фактора на залежну змінну у. Такий варіант F- критерію називається **частковим F-****критерієм** і визначається за формулою ( 57 ) для р=1 .

**Б. Оцінений коефіцієнт детермінації.**

Важливою властивістю коефіцієнта детермінації R2є те, що він – неспадна функція від кількості факторів, які входять до моделі. Якщо кількість факторів зростає, то R2 також зростає і ніколи не зменшується. Це ускладнює порівняння економетричних моделей і вибір серед них найкращої. Так наприклад, якщо ми порівнюємо дві економетричні моделі з однаковою залежною змінною, але різною кількістю пояснюючих змінних, ми звичайно віддаємо перевагу тій, яка має більше значення R2, хоча це може і не відповідати дійсності.

Тому щоб запобігти невиправданому розширенню моделі і **мати** **можливість** **порівнювати** **моделі** з різною кількістю факторів уводять так званий **оцінений коефіцієнт детермінації ***,*  який зменшує вплив зростання кількості факторів на коефіцієнт детермінації за рахунок поправки на ступені вільності. У практиці економетричного дослідження використовуються два різновиди оцінененого коефіцієнта детермінації :

* коефіцієнт детермінації, скоригований за Тейлом - ;
* коефіцієнт детермінації, скоригований за Амемією - .

Перший з них обчислюється за наступною залежністю :

, ( 63 )

а другий – за залежністю :

. ( 64 )

Коефіцієнт детермінації R2 і оцінені коефіцієнти детермінації  та  пов’язані між собою наступними співвідношеннями :

, ( 65 )

. ( 66 )

Вочевидь, для кожного оціненого коефіцієнта детермінації виконується нерівність , тобто зі збільшенням числа пояснюючих змінних моделі оцінені коефіцієнти детермінації зростають повільніше, ніж , зменшуючи таким чином вплив числа факторів на величину коефіцієнта детермінації. Крім того, якщо , то і . Якщо  прямує до нуля , оцінені коефіцієнти кореляції стають від’ємними. Така властивість скоригованих коефіцієнтів детермінації дає змогу більш об’єктивно оцінювати якість моделей з різним числом факторів.

**В. Кореляційна матриця.**

Кореляційна матриця дозволяє оцінити щільність лінійного кореляційного зв'язку між залежною змінною моделі і окремими факторами, а також між окремими незалежними змінними. У загальному випадку вона представляє собою квадратну симетричну матрицю, елементами якої є коефіцієнти парної кореляції між залежною змінною моделі і кожною пояснюючою змінною, а також коефіцієнти парної кореляції між самими пояснюючими змінними моделі.

Для випадку **m** пояснюючих змінних кореляційна матриця має наступний вигляд і структуру :

,. ( 67 )

Діагональні елементи матриці r дорівнюють 1. Коефіцієнти парної кореляції  та  обчислюються за відомою формулою визначення коефіцієнта парної кореляції, яка у даному випадку трансформується у наступні вирази ( 32 ):

, ( 68 )

, ( 69 )

де  - вибіркова дисперсія j – ї пояснюючої змінної,  - вибіркова дисперсія k – ї пояснюючої змінної,  - вибіркова дисперсія залежної змінної,  - вибірковий коефіцієнт коваріації між j – ю пояснюючою змінною і залежною змінною моделі. Ці величини обчислюються за відомими залежностями, як у виразі ( 32 ).

Як відомо, будь-який коефіцієнт парної кореляції з наведеної вище кореляційної матриці характеризує тісноту зв'язку між відповідними змінними за умови, що інші змінні ведуть себе „природним чином” – тобто також змінюють свої значення разом з тими, для яких обчислюється коефіцієнт парної кореляції. Це не дає можливість оцінити тісноту кореляційного зв’язку між двома змінними моделі так би мовити у „чистому вигляді”.

Тому при вивченні зв'язку між змінними економетричної моделі недостатньо спиратися тільки на кореляційну матрицю. Необхідно також проаналізувати **часткові коефіцієнти кореляції**. На відміну від парних часткові коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що інші змінні сталі. Такі коефіцієнти кореляції більш коректно вимірюють силу лінійного кореляційного зв’язку між змінними моделі і дають більш точну інформацію, необхідну при побудові загальної лінійної економетричної моделі. Формула для визначення часткового коефіцієнта кореляції між двома змінними моделі має вигляд :

, ( 70 )

де  – елементи матриці **С**, оберненої до кореляційної матриці **r**. За цією формулою визначається частковий коефіцієнт кореляції між економічним показником **j** і показником **k** при умові, що усі інші економічні показники, які фігурують в економетричні моделі є сталими. Слід зазначити, що у якості показника **j** може виступати як залежна змінна моделі так і незалежна, а у якості показника **k** – незалежна, пояснююча змінна.