УДК 519.21+685.115

Кутовий В. О., канд. фіз.-мат наук, доцент
Катуніна О. С., к.е.н., доцент
доценти кафедри економіко-математичного моделювання,
ДВНЗ «Київський національний економічний університет
імені Вадима Гетьмана»

Kutovyi V. O., Ph.D., associate professor
Katunina O. S., Ph.D., associate professor
Associate Professors of the Department of Economics
and Mathematical Modeling
Kyiv National Economic University named after Vadym Hetman

ПРОЕКТУВАННЯ ПРЕДИКТОРІВ ДЛЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ З МАТРИЦЕЮ СПОСТЕРЕЖЕНЬ НЕПОВНОГО РАНГУ

PROJECTING PREDICATORS FOR ECONOMETRIC MODELS WITH MATRIX OF SUPERVISORY RANGE OBSERVATIONS

АНОТАЦІЯ. Розглянуто питання проектування предикторних просторів і розроблення ефективних процедур оцінювання параметрів економетричних моделей із мультиколінеарними змінними. Проведено дослідження альтернативних підходів до формування сукупності ознак у моделях взаємозалежностей. Запропоновано математичний інструментарій для обчислення параметрів лінійної економетричної моделі у випадку матриці спостереження неповного рангу, що базується на дослідженні сингулярних чисел.

КЛЮЧОВІ СЛОВА: Проектування ознак, економетрична модель, мультиколінеарність, матриця спостережень неповного рангу, сингулярний розклад, власні числа

АННОТАЦИЯ. Рассмотрены вопросы проектирования предикторных пространств и разработки эффективных процедур оценивания параметров эконометрических моделей с мультиколлинеарными переменными. Проведено исследование альтернативных подходов к формированию совокупности признаков в моделях взаимозависимостей. Предложен математический инструментарий вычисления параметров линейной эконометрической модели в случае матрицы наблюдений неполного ранга, основанный на исследовании сингулярных чисел.

КЛЮЧЕВЫЕСЛОВА:Проектирование признаков, эконометрическая модель, мультиколлинеарность, матрица наблюдений неполного ранга, сингулярное разложение, собственные числа

ABSTRACT.The questions of designing of predictor spaces and development of effective procedures for estimating the parameters of econometric models with multi-collinear variables are considered. The research of alternative approaches to the formation of a set of features in the models of interdependencies has been carried out. A mathematical tool for calculating the

parameters of a linear econometric model in the case of a matrix of observation of an incomplete rank based on the study of singular numbers is proposed.

KEYWORDS:Characterization design, econometric model, multicollinearity, matrix of observations of incomplete rank, singular expansion, eigenvalues

Вступ. У даний час інструментарій дослідницької аналітики інтенсивно нарощується новітніми процедурами, що забезпечують високу прогностичну ефективність. В економетричному моделюванні актуальною є проблема визначення інформативних предикторів, формування раціональної сукупності екзогенних змінних, обґрунтування розмірності та структури предикторних просторів. Сучасні модельні технології DataScience, DataMining, MachineLearning (ML) надають широкий арсенал підходів до проектування ознак (featureengineering).Метою проектування та відбору ознак є запобігання ефекту перенавчання, досягнення більшої компактності моделі за рахунок видалення надлишкових регресорів, зниження розмірності в процесах навчання без вчителя, побудова класифікаторів, відображення процесу розбиття даних на класи та визначення межі розв'язків у редукованому просторі, а також обґрунтована інтерпретація, забезпечення поглибленого розуміння моделі та навчальних даних, візуалізація у просторах, мірність яких сприйматиметься дослідником.

Аналіз публікацій. Базовими підходами до вибору оптимальної множини ознак, зокрема, є перебір усіх комбінацій, прямий відбір та обернене виключення, надання ваг ознакам в алгоритмах лінійної та логістичної регресії, надання важливості ознакам у деревах рішень і варіантах ансамблів, наприклад, «випадковому лісі» тощо [1, 7, 18].

У результаті проектування ознак початкові дані співвідносяться із цільовою змінною, додаються зовнішні джерела даних, залучаються джерела неструктурованих даних. Потокові дані, що містять тексти, часові ряди, зображення, відео, відомості з журналів подій, історії відвідування сайтів тощо агрегуються в процесі генерування векторів ознак для подальшого використання у ML-фреймворках [2, 4–6]. Крім того, поглиблюється інтерпретованість і корисність даних для подальшого дослідження, збільшується потенціал та інформативність великих сукупностей ознак. Після навчання моделі на великій кількості ознак можна відібрати їхню множину, що має найбільшу прогностичну цінність. Зокрема, можна видобути нові тенденції та шаблони в даних, хешувати ознаки тексту з метою їхнього перетворення в індекси тощо.

Постановка проблеми. Водночас для забезпечення раціональної специфікації моделі та перспектив її використання в режимі реального часу, дослідник доволі часто потребує можливості виходу за межі усталених модельних технологій у напрямку певного «регресу» від автоматизованих (принаймні, добре алгоритмізованих) процедур пошуку кращих предикторів до керування цим процесом у «ручному режимі». Останній передбачає, зокрема, використання більш індивідуалізованого та адресного налаштування моделі щодо вибору предикторів.

Зауважимо, що навіть з урахуванням представницької сукупності алгоритмів проектування ознак, що реалізовано на базі платформ MLтипу R, Python [3, 8], у дослідника доволі часто виникають труднощі при проектуванні ознакових просторів, а вибір концептуального підходу та інструментарію є неоднозначним. За необхідності врахування специфіки походження та форматів даниху технологіях BigData, поряд із їхньою подальшою уніфікацією та ансамблевізацією, існує потреба у керованому дослідником «втручанні» у швидкі ансамблеві технології аналізу великих даних.

Водночас, за усталеною практикою, вибір предикторів здійснюється переважно без належного обґрунтування, здебільшого консервативно, найчастіше за аналогією із попередньо застосованими підходами, що показали раніше певну ефективність. Отже, проектування ознак є дієвим інструментом перевірки стереотипів — доволі поширеного наслідку суб'єктивного відтворення простору ознак. Досліджувану гіпотезу включають, зокрема, як ознаку ML — моделі і за умови досягнення суттєвого покращення прогнозних властивостей моделі, доводять інформативність і прогностичну корисність даної ознаки та доцільність набуття нею прогностичного статусу.

На думку авторів, процес вибору ознак і конструювання предикторних просторів доцільно подати ітеративним [13, 14]. Для побудови економетричних моделей із мультиколінеарними змінними та матрицею спостережень неповного рангу ефективним є підхід до вибору пояснюючих змінних на базі дослідження сингулярних розкладів [15–17]. Незважаючи на помірне збільшення часу для прийняття рішень щодо структури економетричної моделі, «персоналізація» побудови моделей взаємообумовленості факторів сприятиме зростанню їхньої адекватності.

Основна частина. Нехай існує лінійне співвідношення між змінною Y і n пояснюючими змінними $X_1, X_2, \ldots, X_m$ та збуренням ε [9].

Якщо ми маємо вибірку з n спостережень над змінними Y і $X_j, j = 1,2,\dots,m$, то можна записати

$$y_i = \sum_{j=1}^{m} x_{ij}\beta_j + \varepsilon_i, \quad i = 1,2,\dots,n. \tag{1}$$

Рівняння (1) можуть бути записані в матричній формі

$$Y = X\beta + \varepsilon, \tag{2}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Через X' і ε' позначимо матриці транспоновані до X і ε, відповідно. Нехай виконуються умови:

1. $M(\varepsilon) = 0$; (3)

2. $M(\varepsilon\varepsilon') = \sigma^2 E$, E — одинична матриця (4)

3. X — матриця, елементи якої – детерміновані числа; (5)

4. $rang\, X = m$ (матриця X- повного рангу.) (6)

В системах лінійних алгебраїчних рівнянь, якщо матриця має повний ранг, то при відповідному обмеженні рівня помилок вхідних даних та збільшенні точності обчислень розв'язок (псевдорозв'язок) системи можна одержати із будь-якою заданою точністю. Але ці ж дослідження показують і інше. Якщо рівень помилок вхідних даних та точність обчислень фіксовані, то завжди знайдуться системи з настільки великими значеннями чисел обумовленості, що для них неможна гарантувати в розв'язку (псевдорозв'язку) прийнятної точності[12].

Такі системи називаються *нестійкими* або *погано обумовленими*. В цілому вони характеризуються тим, що незначна зміна умов обчислень може призвести до неприпустимо великих помилок в розв'язку. Причина цього явища, здебільшого, є наступною: в межах зміни помилок вхідних даних або еквівалентних збурень матриця системи або стає матрицею не повного рангу, або дуже

близькою до такої. Всі складнощі розв'язання нестійких систем пов'язано, по-суті, лише із складнощами розв'язання систем з матрицями неповного рангу в умовах збурення вхідних даних та впливу помилок заокруглення. Таких складнощів є чимало.

Теоретично розв'язок систем з матрицяминеповного рангу не відрізняється від розв'язку систем із прямокутними матрицями повного рангу. Нехай задано будь-яку систему (1,2) лінійних алгебраїчних рівнянь. Шукатимемо її нормальний псевдорозв'язок, тобто такий вектор β, який має найменшу евклідову норму серед векторів, що мінімізують функціонал нев'язки:

$$S = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} \hat{\beta}_j x_{ij})^2 \to \min \tag{7}$$

Відомо [3], що в цьому випадку

$$\hat{B} = X^+ Y \tag{8}$$

$$\hat{\beta} = \{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m\}$$

де X^+ — ліва псевдозворотняматриця для матриці X.

Однак зовнішня схожість між системами з матрицями повного і неповного рангів є оманливим. В дійсності між ними існує принципова відмінність.

Якщо матриця системи має повний ранг, то в деякому околі вхідних даних нормальний псевдорозв'язок є неперервним. Якщо ж матриця системи не має повного рангу, то в будь-якому околі вхідних даних нормальний псевдорозв'язок є розривним [11].

Ця відмінність настільки важлива, що обумовлює вважати дослідження залежності похибки нормального псевдорозв'язку від збурень вхідних даних і помилок заокруглення невід'ємною та обов'язковою складовими будь-якого чисельного методу розв'язання систем з матрицями неповного рангу. Однак таке дослідження здійснюється дуже рідко, що, можливо, пояснюється тією удаваною легкістю, із якою математика точних обчислень пропонує «ефективні» методи розв'язання систем лінійних алгебраїчних рівнянь.

Розглянемо докладніше оцінювання параметрів лінійної економетричної моделі у випадку матриці спостереження неповного рангу.

Нехай має місце (1-2). Тоді

$$\begin{cases} X\hat{\beta}=Y \\ X\beta+\varepsilon=Y \end{cases} \Rightarrow X'X\hat{\beta}=X'Y \Rightarrow X'X\hat{\beta}=X'(X\beta+\varepsilon)$$

Нехай $V \neq 0$

Якщо $(X'X)V = 0 \Rightarrow \hat{\beta} + Vk$ — розв'язок (1) (2), $\hat{\beta}$ — не однозначний.

Означення 1. Матриця X^+ з розмірами $m \times n$ називається псевдозворотньою Мура-Пенроуза для матриці X, якщо вона задовільняє наступні чотири умови:

1. $X^+XX^+ = X^+$
2. $X\,X^+X = X$
3. X^+X —симетрична
4. $X\,X^+$ — симетрична (9)

Можна довести, що така матриця X^+ завжди існує та єдина [10]. Якщо X — невироджена квадратна матриця, то $X^+ = X^{-1}$, очевидно, задовольняє умови(1-4) якщо X — прямокутна і має повний ранг, то $X^+ = (X'X)^{-1}X'$. Можна перевірити, що псевдозворотьою до діагональної $n \times m$ матриці

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \sigma_m \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

є діагональна $m \times n$матриця.

$$\Sigma^+ = \begin{pmatrix} \sigma_1^+ & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^+ & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_m^+ & 0 & \cdots & 0 \end{pmatrix} \qquad (10)$$

де

$$\sigma_j^+ = \begin{cases} \dfrac{1}{\sigma_j}, \textit{якщо } \sigma_j \neq 0 \\ 0\ , \textit{якщо } \sigma_j = 0 \end{cases}$$

Далі використаємо сингулярний розклад матриці X [11].

$$X = U \sum V' \tag{11}$$

де U — ортогональна $n \times m$ матриця, V' — ортогональна $m \times m$ матриця, а $\sum$ — діагональна $n \times m$ матриця, у якої $\sigma_{ij} = 0$ при $i \neq j \sigma_{ii} = \sigma_i \geq 0$. Стовпчики матриці U є власними векторами матриці XX', а стовпчики матриці V є власними векторами матриці $X'X$. Використовуючи (9), отримаємо $X^+ = V \sum^+ U'$

Оцінки за методом найменших квадратів параметра β в (1) визначаються як значення $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m$, що мінімізують суму квадратів (7) по всіх наборах $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m$. У випадку, коли виконується (6), точка мінімуму визначається формулою $\hat{\beta} = (X'X)^{-1} X'Y$, таких точок мінімуму виявляється багато. Будемо вимагати, щоб сума

$$L = \sum_{j=1}^{m} \beta_j^2 \to \min \tag{12}$$

була мінімальною. Тоді розв'язок (7),(12) однозначний[3] і дорівнює (8).

Означення 2. Оцінка $\hat{\beta}$ параметра β називається ***X***-незміщеною, якщо $MX\ \hat{\beta} = X\ \beta$, тобто якщо $X\ \hat{\beta}$ — незміщена оцінка $X\ \beta$.

Лема 1. Оцінка $\hat{\beta} = X^+ Y$ є X — незміщеною оцінкою β.

Довеведення. Використаємо (2). Тоді

$$\hat{\beta} = X^+ Y = X^+ (X\ \beta + \varepsilon) = X^+ X\ \beta + X^+ \varepsilon$$

отже з (3) і (9)

$$M(X\hat{\beta}) = M(XX^+ X\beta + XX^+ \varepsilon) = XX^+ X\ \beta = X\ \beta$$

Лема 2. Матриця коваріацій $D(X\hat{\beta})$ параметра $X\hat{\beta}$ — оцінки $X\beta$ в моделі (2) дорівнює

$D(X\hat{\beta}) = \sigma^2 X(X'X)^+X'^+$, де $(X'X)^+$ — псевдозворотня до $X'X$.

Доведення. Використаємо $X^+ = (X'X)^+X'$, $X^+Y = X^+X\beta + X^+\varepsilon$ [2].

Тоді $X\hat{\beta} = XX^+Y = XX^+Y = X\ X^+X\beta + XX^+\varepsilon = X\beta + XX^+\varepsilon$. Отже, використовуючи (3),(4), (9) отримаємо

$$D(X\hat{\beta}) = M(X\hat{\beta} - X\beta)\ (X\hat{\beta} - X\beta)' = M(XX^+\varepsilon\ \varepsilon'\ (XX^+)') =$$

$$= M(X(X'X)^+X'\ \varepsilon\varepsilon' X(X'X)^+X') = \sigma^2 X(X'X)^+(X'X\ )(X'X)^+X' =$$

$$= \sigma^2 X(X'X)^+X'$$

Нехай $C = (C_1, C_2, \ldots, C_m)$

Нехай $L(X_1, X_2, \ldots, X_n)$ лінійна оболонка рядків матриці X

Теорема 1. $C \in L(X_1, X_2, \ldots, X_n) \Rightarrow M(c\hat{\beta}) = c\beta$

Доведення. $C \in L(X_1, X_2, \ldots, X_n) \Rightarrow C = \gamma X$, де $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}$

$M(c\hat{\beta}) = M(\gamma X\hat{\beta})$

Згідно леми 2, $M(X\hat{\beta}) = X\beta$

Тому $M(c\hat{\beta}) = \gamma\ X\beta = c\beta$

Тобто $c\hat{\beta}$ — лінійна незміщена оцінка $c\beta$.

Обчислимо

$$D(c\hat{\beta}) = M(c\hat{\beta} - c\beta)(c\hat{\beta} - c\beta)' = M(\gamma X\hat{\beta} - \gamma X\beta)(\gamma X\hat{\beta} - \gamma X\beta)' =$$

$$= M(\gamma XX^+\varepsilon\varepsilon' X^{+\prime} X'\gamma') = M(\gamma X(X'X)^+X'\varepsilon\varepsilon' X(X'X)^+X'\gamma' =$$

$$= \sigma^2(\gamma X(X'X)^+X'X(X'X)^+X'\gamma' = \sigma^2\gamma X(X'X)^+X'\gamma' = c(X'X)^+c'$$

Обчислимо дисперсію σ^2.

Нехай $\begin{cases} Y=X\beta+\varepsilon \\ Y=X\hat{\beta}+e \end{cases}$

тоді $\begin{cases} Y=X\beta+\varepsilon \\ e=Y-X\hat{\beta} \end{cases} \Rightarrow$

$$e=X\beta+\varepsilon-X\hat{\beta}=X\beta+\varepsilon-XX^{+}Y=X\beta+\varepsilon-XX^{+}(X\beta+\varepsilon)=$$

$$=X\beta+\varepsilon-XX^{+}X\beta-XX^{+}\varepsilon=X\beta+\varepsilon-X\beta-XX^{+}\varepsilon=(E-XX^{+})\varepsilon$$

Врахуємо, що $X^{+}=(X'X)^{+}X'$

і отримаємо $e=(E-X(X'X)^{+}X')\varepsilon$.

Маємо:

$$(E-X(X'X)^{+}X')'(E-X(X'X)^{+}X')=$$

$$=E-X(X'X)^{+}X'-X(X'X)^{+}X'+X(X'X)^{+}X'X(X'X)^{+}X'=$$

$$=E-X(X'X)^{+}X'-X(X'X)^{+}X'+X(X'X)^{+}X'X(X'X)^{+}X=$$

$$=E-X(X'X)^{+}X'-X(X'X)^{+}X'+X(X'X)^{+}X'=(E-X(X'X)^{+}X')$$

Позначимо $P=E-X(X'X)^{+}X'$

Отримаємо P — симетрична матриця, причому $P'P=P \Leftrightarrow P^2=P$.

Отже, P — ідемпотентна матриця.

Нехай $P=\left\| p_{ij} \right\|$ - матриця $m\times n$ розміру.

Позначимо $spA=\sum_{i=1}^{n} a_{ii}$, де A — квадратна матриця.

Обчислимо $(e'e)$. Маємо

$$(e'e)=\varepsilon' P\varepsilon=\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\varepsilon_i\varepsilon_j=$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}((\varepsilon_i\varepsilon_j-M\,\varepsilon_i M\,\varepsilon_j)+\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}M\,\varepsilon_i M\,\varepsilon_j$$

В силу припущень (3) $M\,\varepsilon_i=M\,\varepsilon_j=0$.

Отже

$$(e'e)=\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}(\varepsilon_i\varepsilon_j-M\,\varepsilon_i M\,\varepsilon_j)=$$

$$=\sum_{i=j}p_{ii}(\varepsilon_i{}^2-(M\,\varepsilon_i)^2)+\sum_{i\neq j}p_{ij}(\varepsilon_i\varepsilon_j-M\,\varepsilon_i M\,\varepsilon_j)$$

Тому $M(e'e)=\sum_{i=1}^{n}p_{ii}\sigma^2+\sum_{i\neq j}p_{ij}\operatorname{cov}(\varepsilon_i\varepsilon_j)=\sigma^2(spP)$,
так як $\operatorname{cov}(\varepsilon_i\varepsilon_j)=0$. Але

$$spP=sp(E_n-X(X'X)^+X')=spE_n-spX(X'X)^+X'=$$

$$n-sp(X'X)^+X'X=n-spE_r=n-2$$

де $r=rangX<m$. Отже

$$M\frac{\sum_{i=1}^{n}e_i^2}{n-2}=\sigma^2$$

і означає, що $\frac{\sum_{i=1}^{n}e_i^2}{n-r}$ — незміщена оцінка σ^2 .

В силу припущень в лінійній моделі (1-5), $rang\,X<m$. Матриця $(X'X)$має порядок $m\times m$ і вона симетрична і невід'ємно визначена. Тому у неї m невід'ємних власних значень ρ_1^2, ρ_2^2,ρ_3^2…ρ_m^2, таких що

$$\rho_1^2\geq\rho_2^2\geq\cdots\geq\rho_t^2>\rho_{t+1}^2=\cdots=\rho_m^2=0$$

Тоді має місце (11), де U — ортогональна $n\times n$матриця, V — ортогональна $m\times m$ матриця. Стовпчики матриці U є власними векторами матриці XX', а стовпчики матриці V є власними векторами матриці $X'X$. Тобтоматриця $V=(v_1,v_2,\ldots,v_m)$, утворена власними векторами, де v_i позначається вектор-стовпчик, що відповідає власному значенню ρ_i^2. Маємо

$$(X'X)V_i=\rho_i^2v_i$$

і крім того $V'V=E_m$.

Нехай $L(V_1, V_2, V_3, \dots, V_t)$ — лінійна оболонка векторів $V_1, V_2, \dots, V_t$ і нехай $C' = (C_1, C_2, \dots, C_m)' \in L(V_1, V_2, \dots, V_t)$. Тоді

$c_i' = \sum_{i=1}^{t} v_j \ \alpha_j \ \Leftrightarrow C = \alpha V'$,

де $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_t\} = (\alpha_1, \alpha_2, \dots, \alpha_t, 0, 0, \dots, 0) \in R^m$.

Теорема 2. Матриця коваріацій $D(\hat{\beta}) = \sigma^2 (X'X)^+$.

Доведення. Нехай $\hat{\beta} = X^+ Y$ нормальний псевдорозв'язок. Також має місце (2), звідки $\hat{\beta} = X^+(X\beta + \varepsilon) = X^+ X\beta + X^+ \varepsilon$.

Із (3) математичне сподівання $M\hat{\beta} = X^+ X\beta \neq \beta$

тобто

$M\hat{\beta} \neq \beta$

$D(\hat{\beta}) = M(\hat{\beta} - X^+ X\beta)(\hat{\beta} - X^+ X\beta)' = M(X^+ \varepsilon, \varepsilon' X^{+\prime}) = \sigma^2 M(X^+ X^{+\prime})$

$X^+ = (X'X)^+ X'$ [2], тому з (9)

$D(\hat{\beta}) = \sigma^2 (X'X)^+ X'X(X'X)^+ = \sigma^2 (X'X)^+$

Маємо з теореми 2 і з $c = \alpha V'$

$M\alpha \ V'\hat{\beta} = \alpha V' X^+ X\beta$

та $c\hat{\beta} - Mc\hat{\beta} = \alpha V' X^+ \varepsilon$

$D(c\hat{\beta}) = M\alpha V' X^+ \varepsilon\varepsilon' X^{+\prime} V\alpha' = \sigma^2 \alpha V' X^+ X^{+\prime} V'\alpha' =$

$= \sigma^2 \alpha V' (X'X)^+ X'X(X'X)^+ V'\alpha' = \sigma^2 \alpha V' (X'X)^+ V'\alpha' =$

$= \sigma^2 \alpha V' (X'X)^+ V'\alpha'$

Тоді згідно (10-11)

$$D(c\hat{\beta}) = \sigma^2 \alpha \ \Sigma^+ \alpha'$$

де

$$\Sigma^+ = \begin{pmatrix} \left(\rho_1^2\right)^+ & \cdots & (0) \\ \vdots & \left(\rho_2^2\right)^+ & \vdots \\ (0) & \cdots & \left(\rho_m^2\right)^+ \end{pmatrix}$$

де

$$\left(\rho_i^2\right)^+ = \begin{cases} \dfrac{1}{\rho_i^2} & при\ \rho_i^2 \neq 0 \\ 0 & при\ \rho_i^2 = 0 \end{cases}$$

так що

$$D(c\hat{\beta}) = \sigma^2\left(\frac{\alpha_1^2}{\rho_1^2} + \frac{\alpha_1^2}{\rho_2^2} + \cdots + \frac{\alpha_t^2}{\rho_t^2}\right) \qquad (13)$$

Бачимо, що дисперсія лінійної комбінації $c\hat{\beta}$ залежить від обернених власних значень і внаслідок менший по абсолютній величині характеристичний корінь виявляє найбільший вплив на цю дисперсію.

Нас цікавить окремий коефіцієнт $\hat{\beta}_i$. Нехай вектор $c=e_i=(0,0,....,1,...,0)$ буде мати одиницю на і-й позиції, а решта координат дорівнюватиме нулю. При цьому вимагаємо: $\mathbf{c' \in L(V_1, V_2, \ldots, V_t)}$.

Із $\mathbf{c = \alpha V'}$ отримаємо

$$\alpha = e_i V' = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{it} \end{pmatrix}'' \quad \text{де } v_{i1}, v_{i2}, \ldots, v_{it} \text{ — елементи рядка матриці V.}$$

Наслідок 1.

$$D(\hat{\beta}_i) = \sigma^2\left(\frac{v_{i1}^2}{\rho_1^2} + \frac{v_{i2}^2}{\rho_1^2} + \cdots + \frac{v_{it}^2}{\rho_t^2}\right)$$

$i=1,...,m$

при $\mathbf{e_i \in L(V_1, V_2, \ldots, V_t)}$.

Нехай $C = \{C_1, C_2, \ldots, C_m\}$

$\{\rho_1^2 \geq \rho_2^2 \geq .... \rho_t^2 > 0, 0, ..., 0\}$ — власні значення $X'X$.

$L(V_1, V_2, ..., V_t)$ — лінійна оболонка $V_1, V_2, ..., V_t$ для яких відповідно

$$\rho_1^2 \geq \rho_2^2 \geq .... \rho_t^2 > 0$$

Нехай

$$C' = \{C_1, C_2, \dots, C_m\}' \in L\ (V_1, V_2, \dots, V_t) \Rightarrow c = \alpha V' \qquad (14)$$

де

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_t) = (\alpha_1, \alpha_2, \dots, \alpha_t, 0, 0, \dots, 0) \in R^m$$

і розв'язок системи

$$\alpha_1(V_1) + \alpha_2(V_2) + \dots + \alpha_t(V_t) = c' \ , \qquad (15)$$

або в матричному вигляді $C = \alpha V'$

Нехай X^+Y — нормальний псевдорозв'язок задачі (2-5). Як показано вище, має місце (13)

Покладемо в (14-15)

$$\alpha_{k+1}^2 + \alpha_{k+2}^2 + \dots + \alpha_t^2 = 0 \qquad (16)$$

тоді

$\hat{c}_k = \sum_{i=1}^{k} \alpha_i V_i$ — зміщена оцінка c

$\alpha_{k+1}(V_{k+1}) + \alpha_2(V_{k+2}) + \dots + \alpha_t(V_t)$ — зміщення, а

$N_k = \alpha_{k+1}^2 + \alpha_{k+2}^2 + \dots + \alpha_t^2$ — величина зміщення $\hat{C}_k$.

Тоді з (13)-(16)

$$D(\hat{C}_k'\hat{\beta}) = \sigma^2 \left(\frac{\alpha_1^2}{\rho_1^2} + \frac{\alpha_2^2}{\rho_2^2} + \dots + \frac{\alpha_k^2}{\rho_k^2} \right)$$

Нехай

$$M_k = \text{зміщення}\, \hat{C}_k + D(\hat{C}_k\hat{\beta}) \approx N_k + D(\hat{C}_k\hat{\beta})$$

Тоді

$$M_k = \begin{cases} \sum_{i=1}^{t} \alpha_i^2 & \text{при } k=0 \\ \sigma^2 \sum_{j=1}^{k} \frac{\alpha_j^2}{\rho_j^2} + \sum_{j=k+1}^{t} \alpha_j^2 & \text{при } 1 \le k \le t-1 \\ \sigma^2 \sum_{j=1}^{t} \frac{\alpha_j^2}{\rho_j^2} & \text{при } k=t \end{cases}$$

Шукаємо $\min\limits_{0\le k\le t} M_k$

Отже

$$M_{i-1}=\sigma^2\sum_{j=1}^{i-1}\frac{\alpha_j^2}{\rho_j^2}+\sum_{j=i}^{t}\alpha_j^2$$

$$M_i=\sigma^2\sum_{j=1}^{i}\frac{\alpha_j^2}{\rho_j^2}+\sum_{j=i+1}^{t}\alpha_j^2$$

$$M_{i+1}=\sigma^2\sum_{j=1}^{i+1}\frac{\alpha_j^2}{\rho_j^2}+\sum_{j=i+2}^{t}\alpha_j^2$$

$$M_i-M_{i-1}=\frac{\alpha_i^2\sigma^2}{\rho_i^2}-\alpha_i^2=\alpha_i^2\left(\frac{\sigma^2-\rho_i^2}{\rho_i^2}\right)$$

$$M_{i+1}-M_i=\frac{\alpha_{i+1}^2\sigma^2}{\rho_{i+1}^2}-\alpha_{i+1}^2=\alpha_{i+1}^2\left(\frac{\sigma^2-\rho_i^2}{\rho_{i+1}^2}\right)$$

Мінімум $M_i \Rightarrow \min$ коли M_i-M_{i-1} змінює знак «мінус» на «плюс».

$$\begin{cases}\sigma^2-\rho_i^2\le 0\\ \sigma^2-\rho_{i+1}^2\ge 0\end{cases}\Rightarrow\begin{cases}\rho_i^2\ge\sigma^2\\ \rho_{i+1}^2\le\sigma^2\end{cases}\Rightarrow\rho_{i+1}^2\le\sigma^2\le\rho_i^2$$

1) випадок

$\sigma^2\ge\rho_1^2\Rightarrow M_i-M_{i-1}>0$

$\Rightarrow M_i$ зростає. Мінімум M_i при $i=0$

і дорівнює

$M_0=\alpha_1^2+\alpha_2^2+\alpha_3^2\cdots+\alpha_t^2\to\min$

2) випадок

$\rho_{i+1}^2<\sigma^2<\rho_i^2$

Мінімум

$N_k+D(\hat{C}_k)$ $1\le k\le t-1$ дорівнює

$N_i+D(\hat{C}_i\hat{\beta})=$

$$= \sigma^2\left(\frac{\alpha_1^2}{\rho_1^2} + \frac{\alpha_2^2}{\rho_2^2} + \cdots + \frac{\alpha_i^2}{\rho_i^2}\right) + \sum_{j=i+1}^{t} \alpha_j^2 \to \min_{0 \le k \le t}$$

3) випадок

$\sigma^2 < \rho_t^2$,

M_i спадає. Мінімум при $i=t$ і дорівнює

$$M_t = \sigma^2\left(\frac{\alpha_1^2}{\rho_1^2} + \frac{\alpha_2^2}{\rho_2^2} + \cdots + \frac{\alpha_t^2}{\rho_t^2}\right) \to \min$$

Наслідок 2.

При умовах 1) $\sigma^2 < \rho_t^2$; 2) $e'_i = (0,...,1,0,...)' \in L(V_1, V_2, ..., V_t)$

$$\Rightarrow D(\hat{\beta}_i) = \sigma^2\left(\frac{v_{i1}^2}{\rho_1^2} + \frac{v_{i2}^2}{\rho_2^2} + \cdots + \frac{v_{it}^2}{\rho_t^2}\right) \to \min$$

Наслідок 2 справедливий не для всіх $\hat{\beta}_i$, тому щовсі e_i не можуть належати до $L(V_1, V_2, ..., V_t)$ так як e_i — лінійнонезалежні, $(i = 1,2,...,m)$, але $rangX < m$.

Висновки. В економетричному моделюванні проектування дослідником структури предикторного простору є інструментом одержання прогностично ефективної специфікації екзогенних змінних. Визначення підмножини ознактакож є важливою складовою машинного навчання та значно впливає на точність ML-моделей.Концепція машинного навчання максимізує можливість визначення шаблонів в даних, що досягається, зокрема, шляхом агрегування множини ознак. Інформативно слабка ознака може значно збільшити власну прогностичну корисність і стати сильною за присутності іншої ефективної множини ознак.

Реконструювання та вибір ознак сприяє збільшенню автоматизованості процесу навчання. При навчанні ML-моделей на великих даних виникає небезпека перенавчання, але чіткі правила відбору дозволяють зменшити їхню кількість. Отже, цілеспрямоване варіювання кількістю ознак має використовуватись дослідниками для калібрування та навчання моделі, оскільки дає можливість обґрунтувати вибір раціональної множини незалежних змінних, що зумовлюють структури в даних і в подальшому успішно прогнозують тенденції поводження економічних систем.

Література

1. *Джеймс Г., Уиттон Д., Хасти Т.* Введение в статистическое обучение с примерами на языке R / Пер. с англ. С.Э. Мастицкого. — М.: ДМК Пресс, 2016. — 450 с.

2. *Ю. Лесковец, А. Раджараман*. Анализ больших наборов данных / Пер. с англ. Слинкин А.А. — М.: ДМК Пресс, 2016. — 498 с.

3. *Мастицкий С.Э., Шитиков В.К*. Статистический анализ и визуализация данных с помощью R / С.Э. Мастицкий, В.К. Шитиков. — М.: ДМК Пресс, 2015. — 496 с.

4. *Флах П*. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания их данных / пер. с англ. А.А. Слинкина. — М.: ДМК Пресс, 2015. — 400 с.

5. *Гудфеллоу Я., Бенджио И., Курвилль А*. Глубокое обучение / пер. с анг. А.А. Слинкина. — М.: ДМК Пресс, 2017. — 652 с.

6. *Силен Д., Мейсман А., Али М*. Основы DataScience и BigData. Python и наука о данных. — СПб.: Питер, 2017. — 336 с.

7. *Бринк Х., Ричардс Дж., Феверолф М*. Машинное обучение. — СПб.: Питер, 2017. –336 с.

8. *Рашка С*. Python и машинное обучение / пер. с англ. А.В. Логунова. — М.: ДМК Пресс, 2017. — 418 с.

9. *Джонстон Дж*. Эконометрические методы. — М.: Наука. 1986. — 230 с.

10. *Лоусон И., Хенсен*. Численное решение задач метода наименьших квадратов. — М.: Наука, 1986. — 232 с.

11. *Воеводин В.В*. Вычислительные основы линейной алгебры. — М.: Наука, 1977. — 304 с.

12. *Кутовий В.О*. Про теорему Гаусса-Маркова у випадку виродженої матриці спостережень // Доповіді АН України. — №5. — 2001. — С. 19–22.

13. *Кутовий В.О*. Про застосування інструментальних змінних для визначення параметрів загальної лінійної моделі // Моделювання та інформаційні системи в економіці. — Випуск 64. — 2000.

14. *Кутовий В.О., Роскач О.С*. Математико-статистичне узагальнення покрокових методів побудови предикторних просторів // Машинна обробка інформації. — Київ: КДЕУ, 1997. — Випуск 59. — С. 140–149.

15. *Кутовий В.О., Роскач О.С*. Про застосування на ЕОМ алгоритму Фаррара-Глаубера // Машинна обробка інформації. — К.: КНЕУ, 1997. — Вип. 61. — С. 142–149.

16. *Кутовий В.О*. Про умови застосування теореми Гаусса-Маркова // Вчені записки. — К.: КНЕУ. — Вип. 2. — С. 206–208.

17. *Кутовий В.О*. Про ефективність зміщених оцінок параметрів економетричних моделей // Вчені записки. — К.: КНЕУ, 2001. — Випуск 3. — С. 324–326.

18. *Вітлінський В.В., Катуніна О.С*. Тенденції розвитку економіко-математичного моделювання / В.В. Вітлінський О.С. Катуніна // Мате-

ріали першої національної конференції «Економіко-математичне моделювання». — К.: КНЕУ, 2016 — С. 54–56.

References

1. Dzhejms H., Uytton D., Khasty T., Vvedenye v statystycheskoe obuchenye s prymeramy na iazyke R./ Per. s anhl. S.E.Mastytskoho. — M.: DMK Press, 2016. — 450 s.
2. Yu. Leskovets, A. Radzharaman Analyz bol'shykh naborov dannykh./ Per. s anhl. Slynkyn A.A. — M.: DMK Press, 2016. — 498 s.
3. Mastytskyj S.E., Shytykov V.K. Statystycheskyj analyz y vyzualyzatsyia dannykh s pomosch'iu R / S.E. Mastytskyj, V.K. Shytykov. — M.: DMK Press, 2015. — 496 s.
4. Flakh P. Mashynnoe obuchenye. Nauka y yskusstvo postroenyia alhorytmov, kotorye yzvlekaiut znanyia ykh dannykh / per. s anhl. A.A. Slynkyna. — M.: DMK Press, 2015. — 400 s.
5. Hudfellou Ya., Bendzhyo Y., Kurvyll' A. Hlubokoe obuchenye / per. s anh. A.A.Slynkyna. — M.: DMK Press, 2017. — 652 s.
6. Sylen D., Mejsman A., Aly M. Osnovy DataSciencey BigData. Python y nauka o dannykh. — SPb.: Pyter, 2017. — 336 s.
7. Brynk Kh., Rychards Dzh., Feverolf M. Mashynnoe obuchenye. — SPb.: Pyter, 2017. –336 s.
8. Rashka S. Python y mashynnoe obuchenye / per. s anhl. A.V. Lohunova. — M.: DMK Press, 2017. — 418 s.
9. Dzhonston Dzh. Ekonometrycheskye metody. –тM.: Nauka. 1986. — 230 s.
10. Louson Y., Khensen Chyslennoe reshenye zadach metoda naymen'shykh kvadratov. — M.: Nauka, 1986. — 232 s.
11. Voevodyn V.V. Vychyslytel'nye osnovy lynejnoj alhebry. — M.: Nauka, 1977. — 304 s.
12. Kutovyj V.O. Pro teoremu Haussa-Markova u vypadku vyrodzhenoi matrytsi sposterezhen' // Dopovidi AN Ukrainy. — №5. — 2001. — S. 19–22.
13. Kutovyj V.O. Pro zastosuvannia instrumental'nykh zminnykh dlia vyznachennia parametriv zahal'noi linijnoi modeli // Modeliuvannia ta informatsijni systemy v ekonomitsi. Vypusk 64. — 2000.
14. Kutovyj V.O.,Roskach O.S. Matematyko-statystychne uzahal'nennia pokrokovykh metodiv pobudovy predyktornykh prostoriv // Mashynna obrobka informatsii. — Kyiv: KDEU, 1997. — Vypusk 59. — S. 140–149.
15. Kutovyj V.O.,Roskach O.S. Pro zastosuvannia na EOM alhorytmu Farrara-Hlaubera // Mashynna obrobka informatsii. — K.: KNEU, 1997. — Vyp. 61. — S. 142–149.
16. Kutovyj V.O. Pro umovy zastosuvannia teoremy Haussa-Markova // Vcheni zapysky. — K.: KNEU. — Vyp. 2. — S. 206–208.
17. Kutovyj V.O. Pro efektyvnist' zmischenykh otsinok parametriv ekonometrychnykh modelej // Vcheni zapysky. — K.: KNEU, 2001. — Vypusk 3. — S. 324–326.

18. Vitlins'kyj V.V., Katunina O.S. Tendentsii rozvytku ekonomiko-matematychnoho modeliuvannia / V.V. Vitlins'kyj, O.S. Katunina // Materialy pershoi natsional'noi konferentsii «Ekonomiko-matematychne modeliuvannia» — K.: KNEU, 2016. — S. 54–56.

УДК 502/504

Галіцина О. В., к.е.н., доцент,
Київський національний економічний університет
імені Вадима Гетьмана

Galitsyna O. K., Ph.D., assistant professor
Kyiv National Economic University
named after Vadym Hetman

СТАТИСТИЧНИЙ МОНІТОРИНГ НАДЗВИЧАЙНИХ ЕКОЛОГІЧНИХ СИТУАЦІЙ ТЕХНОГЕННОГО ХАРАКТЕРУ В УКРАЇНІ

STATISTICAL MONITORING OF ECOLOGICAL EMERGENCIES BY TECHNOGENIC CHARACTER IN UKRAINE

АНОТАЦІЯ. Обґрунтовано роль і значення статистичного моніторингу надзвичайних екологічних ситуацій, як важливого стратегічного ресурсу управління. Проведено статистико-аналітичну оцінку наслідків надзвичайних екологічних ситуацій на різних рівнях їх прояву, проаналізовано дані статистичного моніторингу надзвичайних екологічних ситуацій, запропоновано показник оцінки результатів НЕС.

КЛЮЧОВІ СЛОВА: моніторинг, проблема, надзвичайна ситуація, екологія, техногенний характер.

АННОТАЦИЯ. Обоснована роль и значение статистического мониторинга чрезвычайных экологических ситуаций , как важного стратегического ресурса управления. Проведена статистико-аналитическая оценка последствий чрезвычайных экологических ситуаций на разных уровнях ее проявления.

КЛЮЧЕВЫЕ СЛОВА: мониторинг, проблема, чрезвычайная ситуация, экология, техногенный характер.

ANNOTATION. A role and value of the statistical monitoring of ecological emergencies are reasonable, as an important strategic resource of management. Thestatistical and analyticalestimation of consequences of ecological emergencies is conducted on the different levels of their display, data of the statistical monitoring of ecological emergencies are analysed, the index of estimation of results of ecological emergenciesis offered.

KEY WORDS: monitoring, problem, emergency, ecology, technogenic character.