

УДК 65.012.122:330.3

ВПРОВАДЖЕННЯ НАУКОВО-ПРАКТИЧНОГО ДОСЛІДЖЕННЯ ТА УДОСКОНАЛЕННЯ ЙОГО МЕТОДИЧНОГО ІНСТРУМЕНТАРІЮ: ОЦІНКА ПОКАЗНИКІВ ДОСЛІДЖЕННЯ НА АНОМАЛЬНІСТЬ

Євсєєва О.О., к.е.н, доцент (УкрДАЗТ)

В статті детально розглянуто методологічні основи, описано використаний методологічний математичний інструментарій та порядок проведення дослідження на етапі «вхідні дані», т.т. передосновного етапу дослідження, здійснення якого дає наукове обґрунтування використання чи невикористання в передбачуваному дослідженні відповідних показників.

Ключові слова: дослідження, показники, методичний інструментарій, оцінка показників, аномальність, критерії аномальності.

Постановка проблеми та її зв'язки з науковими та практичними завданнями. Будь-яке наукове дослідження в спрощеному вигляді проводиться за схемою: «вхідні дані – обробка та дослідження даних за пропонованою методикою – аналіз результатів дослідження». Такий, на перший погляд, простий підхід до втілення наукового дослідження, не є таким насправді.

Визначивши на теоретичному рівні етапи дослідження, необхідно перейти до їх втілення на практиці.

Аналіз останніх досліджень і публікацій, на які спирається автор з посиланням на джерела. Вже на етапі «вхідні дані» виникають питання щодо способу їх отримання. Такими способами є або накопичення інформації за необхідними відповідними показниками за результатами фактичного спостереження за об'єктом дослідження, або проведення наукових досліджень в лабораторії, або проведення так званих імітаційних досліджень та фіксування отриманих результатів імітаційної моделі тощо. Навіть після отриманих значень показників, не можна одразу розпочинати основне наукове дослідження, оскільки на цьому підготовчому етапі необхідно оцінити отриману інформацію з точки зору таких критеріїв, як суміжності часових періодів, аномальності [3] тощо. Тобто знизити вірогідність використання некоректної інформації в майбутньому дослідженні, оскільки це може призвести до невірних результатів та висновків, а в деяких випадках і до небажаних наслідків. Також на цьому етапі відбувається критичний аналіз існуючих даних та ситуації, що вони характеризують, а, відповідно, доводиться необхідність проведення нових обґрунтованих досліджень.

На етапі «обробка та дослідження даних за пропонованою методикою» відбувається обробка

даних або за пропонованою методикою, яка була розроблена іншими науковцями та втілюється (перекладається, адаптується та удосконалюється) в новій сфері наукової чи/та практичної діяльності, або за методикою, що розробляється автором наукового дослідження самостійно та застосовується в досліджуваній науковій чи/та практичній сфері. На цьому етапі реалізуються та використовуються результати попередніх наукових досліджень, які нібито «переломлюються» в іншій якості на дійсному науковому дослідженні, та вагомий науковий досвід і потенціал самого науковця. Це відбувається через необмежений вибір різноманітних інструментів дослідження [2] саме тих, за допомогою яких воно й відбувається.

На етапі «аналіз результатів дослідження» доводиться доцільність проведеного дослідження, розкриваються позитивні та негативні сторони, планування та прогнозування вагомих показників дослідження, проводиться аналіз вихідних результатів на адекватність, т.т. проводиться перевірка пропонованої методики або моделі на адекватність за відповідними параметрами адекватності [3], перспективи подальших удосконалення та проведення дослідження [1], наводять висновки та пропонують відповідні пропозиції тощо.

Виділення невирішених частин загальної проблеми, котрим присвячується стаття. Враховуючи вищенаведене, виникає необхідність розробки та побудови логіко-концептуальної моделі дослідження (рисунок 1).

Розроблена модель в подальшому може бути покладена в дослідження оцінки ефективності діяльності та управління будь-якого галузевого національного комплексу України (машинобудівельного, енергетичного, залізничного, сільськогосподарського тощо).

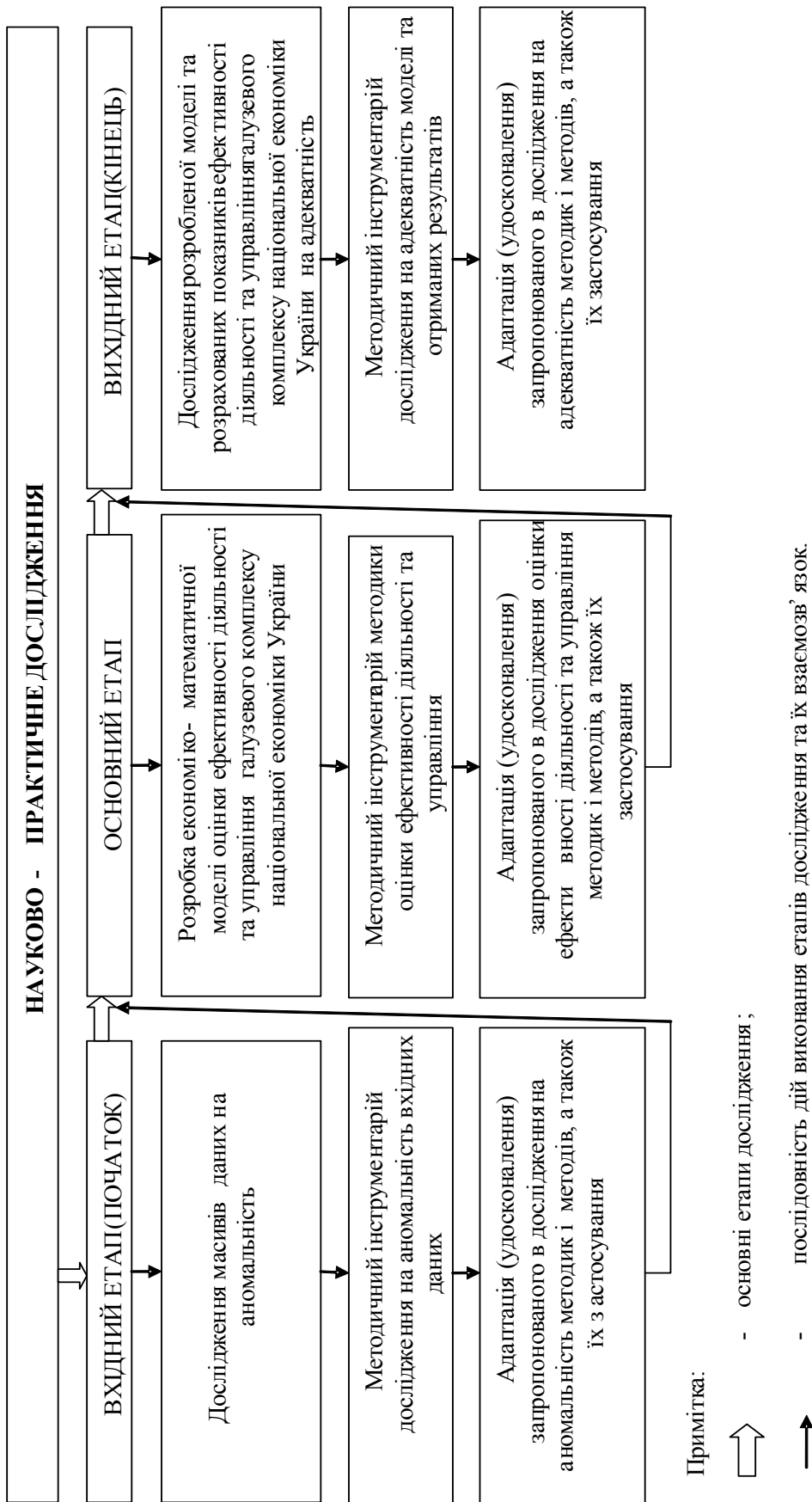


Рисунок 1 – Логіко - концептуальна модель проведення наукового практичного дослідження

Формулювання цілей статті (постановка завдання). В дійсній статті нами детально буде розглянуто методологічні основи, описано використаний методологічний математичний інструментарій та порядок проведення дослідження на етапі «вхідні дані», т.т. передосновного етапу дослідження, здійснення якого дає наукове обґрунтування використання чи невикористання в передбачуваному дослідженні відповідних показників. Визначена мета обумовлена тим, що в сучасних умовах, на жаль, методичним і методологічним основам відбору показників в дослідження приділяють недостатньо уваги або використовують загальноприйняті кореляційні та конкордаційні прийоми вивчення часових рядів. Такі прийоми хоча і вважаються універсальними, але мають свої недоліки, наприклад, вони тільки констатують факт виявлення аномальності, але не усувають її.

Виклад основного матеріалу дослідження з обґрунтуванням отриманих наукових результатів.

Впровадження вхідного етапу. Для будь-якого виду оціночної діяльності, а тим більше наукового економічного дослідження, повинна використовуватись відповідна достовірна вихідна база показників, яка характеризує певні процеси цифровими даними та масивами як в статичній формі, так і за періоди в динаміці, тобто не тільки моментні дані або дані за короткий час, а й за відповідні часові проміжки (поєси). Такий масив даних, як правило, використовується в рамках обраного методу без урахування потенційних наслідків у разі некоректного використання даних. В такому випадку мова йде вже просто про розробку суворой математичної моделі, яка на практиці не буде спрацьовувати ефективно внаслідок перекручування показників, що використовуються в такому випадку за ради «вписання» їх в складну математичну модель.

Враховуючи вищесказане, нами запропоновано перед безпосереднім застосуванням тих чи інших масивів даних за відповідні періоди, які можуть характеризувати будь-які сторони розвитку процесу підприємства, застосовувати процедуру виявлення так званих аномальних значень в вихідних масивах даних з метою зниження й, в кінцевому разі, усунення ризику як некоректного використання показників в обраній математичній моделі (усунення об'єктивних

причин), так і усунення людського фактору щодо надання помилкових або неточних даних в масиві показників (усунення суб'єктивних причин). Також треба мати на увазі, що, наприклад, для прогнозування та будь-якої іншої аналітичної роботи важливе значення мають й інші чинники, наприклад, сам проміжок часових рядів та їх співставність за одиницями вимірювання та методикою підрахунку.

«Важливе значення для прогнозування має вибір інтервалів між сусідніми рівнями ряду. При занадто великому інтервалі часу можуть бути упущені деякі закономірності в динаміці показника. При занадто малому – збільшується обсяг обчислень, можуть появлятися неістотні деталі в динаміці процесу. Вибір інтервалу часу між рівнями ряду повинен вирішуватися конкретно для кожного процесу, причому зручно мати рівновіддалені один від одного рівні.» [2, с. 404-405]

«Важливою умовою правильного відображення часовим рядом реального процесу розвитку є порівнянність рівнів ряду. Непорівнянність частіше всього зустрічається в вартісних характеристиках, змінах цін, територіальних змінах, укрупнення підприємства та інше. Для величин показника, що не порівнюються, неправомірно проводити його прогнозування.

Для успішного вивчення динаміки процесу необхідно, щоб інформація була повною на прийнятому рівні спостережень, часовий ряд мав достатню довжину, були відсутні пропущені спостереження.

Рівні часових рядів можуть мати аномальні значення. Поява таких значень може бути визвано помилками при зборі, записі або передачі інформації – це помилки технічного порядку або помилки першого ряду. Однак, аномальні значення можуть відображати реальні процеси, наприклад, стрибок курсу долара або падіння курсу цінних паперів на фондовому ринку та інше; такі аномальні значення відносять до помилок другого роду, вони не підлягають усуненню.

Для виявлення аномальних рівнів часових рядів можна використовувати метод Ірвіна.» [2, с. 405]

Нехай y_t є часовий ряд, формула (1) [2, с. 405, формула (18.1)]:

$$y = y_1, y_2, \dots, y_t, \dots, y_n, \quad t = \overline{1, n} \quad (1)$$

Метод Ірвіна припускає використання наступної формули, формула (2) [2, с. 405, формула (18.2)]:

$$\lambda_t |y_t - y_{t-1}| / \sigma_y, \quad t = \overline{1, n}, \quad (2)$$

де: σ_y – середньоквадратичне відхилення часового ряду y_1, y_2, y_3, y_n .
 Середньоквадратичне відхилення часового ряду y_1, y_2, y_3, y_n визначається за формулою (3) [2, с. 394, формула (17.9)]:

$$\sigma_y^2 = \sum (y_i - \bar{y})^2 / (n - 1), \tag{3}$$

де: i – є фактором та може приймати значення $i = 1, n$ [2, с. 393, формула (17.7)].
 Фактор i в формулі показує вже не скільки
 аномальність показника, скільки його суттєвість на етапі вводу.
 Тобто σ буде дорівнювати, формула (4):

$$\sigma = \sqrt{\sigma_y^2} \tag{4}$$

«Розрахункові значення $\lambda_2, \lambda_3 \dots$ порівнюються з табличними значеннями критерію Ірвіна λ_n , якщо яке-небудь з них виявляється більше табличного, то відповідне значення y_i рівня ряду вважається аномальним.» [2, с. 405]
 Значення критерію Ірвіна для рівня значимості $\alpha = 0,05$ наведено в таблиці 1 [2, с. 406, таблиця (18.1)]:

Таблиця 1

Значення критерію Ірвіна для рівня значимості $\alpha = 0,05$

n	2	3	10	20	30	50	100
λ_n	2,8	2,3	1,5	1,3	1,2	1,1	1,0

«Після виявлення аномальних рівнів необхідно визначити причини їх виникнення. Якщо вони викликані помилками технічного порядку, то вони усуваються або заміною аномальних рівнів відповідними значеннями по кривій, по якій апроксимує часовий ряд, або заміною рівнів середньої арифметичної двох сусідніх рівнів ряду.

Помилки, що виникають за причиною впливу факторів, які мають об'єктивний характер, усуненню не підлягають» [2, с. 406]

Удосконалення та адаптація. З урахуванням попередньо викладеного матеріалу про сутність методу Ірвіна, наведені в джерелі [2, с. 406] значення критерію Ірвіна, на нашу думку, є обмежувачами з точки зору надання їх під визначені періоди дослідження, тобто дослідження може проводитись лише за 2, 3, 10, 20, 30, 50 та 100 періодів. Але дослідження не може завжди обмежуватись тільки визначеними періодами. Більше того, дослідження завжди є цікавим як з точки зору змін в динаміці, так і з точки зору накопичення статистичних даних за відповідні періоди із ходом часу. Таким чином, виникає питання визначення значення критерію Ірвіна для періодів, що припадають в межі [4-9], [11-19], [21-29], [31-49], [51-99]. В той же час може виникнути питання, наприклад, „Навіщо розраховувати значення критерію Ірвіна для періоду [51-99]?”. Тут можна зауважити, що все залежить від характеру дослідження та сфери виникнення обстежуваних показників. Так, наприклад, в дослідженні технічної сфери є цікавим велика кількість досліджень одного і того ж явища з

метою зниження ймовірності помилок невизначеного характеру. В дослідженні соціально-економічної сфери з метою вивчення того чи іншого явища є доцільним обстеження об'єкту дослідження та подальшого прогнозування його стану в майбутньому через проведення так званих імітавань, від кількості яких залежить точність розповсюдження результатів та висновків на дослідження, а, відповідно, виникає необхідність проведення таких імітацій в більш великій кількості, ніж запропоновано авторами в джерелі [2]. Також в дослідженні соціально-економічної сфери є цікавим вивчення певних показників за відповідні звітні періоди, наприклад роки, які в подальшому можуть деталізуватись показниками більш менших за термінами звітних періодів, наприклад, півріччям, кварталом, місяцем, і в такому випадку, ці періоди за відповідні роки можуть дати велику кількість досліджуваних періодів з метою вивчення факторів, що впливають на діючу модель, та побудови оновленої факторної моделі. В такому випадку, наприклад, 7 років дають 12 місяців, відповідно виникає необхідність мати значення критерію Ірвіна при n -періодів, що дорівнює 84.

Враховуючи вищевказане, виникає необхідність довести значення критерію Ірвіна для рівня значимості λ_n до універсального вигляду, тобто так, як це розроблено для різних методик, де є порівняння розрахованих показників із якимось критерієм або який показник використовується для подальших розрахунків, наприклад, як: функція Пуассона [2, с. 450-451], критичні точки розподілення Стьюдента [2, с. 453; 1, с.460],

критичні точки розподілення Фішера-Снедекора [2, с. 454-455; 1, с. 461], експоненціальні значення [2, с. 456], порядкові номери днів в році [2, с. 458-459; 3, с. 307], розподілення критерію Дурбіна-Уотсона [2, с. 460], складні відсотки [3, с. 308-310], дисконтні складні відсотки [3, с. 3-313], коефіцієнти приросту річної ренти [3, с. 314-319], кванти лі x_2 [1, с. 462], таблиця x_2 розподілу [1, с. 463], критичні значення кореляційного відношення η^2 і коефіцієнта детермінації R^2 [1, с. 463], критичні значення F-критерію [1, с. 464], рівень значимості [1, с. 465] тощо.

Все це доводить необхідність розрахувати значення критерію Ірвіна для рівня значимості $\alpha = 0,05$ та довести отримані результати до читабельно-універсального вигляду. Отримані результати отримані зворотно-пропорційним методом через визначений „шаг” (Ш) відстані між

відомими показниками відповідних періодів з урахуванням крайніх періодів, які в даному випадку складають нуль 0. Так, наприклад, для $n=[4-9]$, розраховано відхилення для n від 3 до 10, яке складає $\Delta n = 10-3=7$ та модульне відхилення для n від 3 до 10, яке складає $\Delta \lambda_n = |1,5-2,3|=0,8$, після чого був визначений „шаг” (Ш) для проміжку від 4 до 9, який складає $0,8/7=0,1142857$. Значення „шагу” (Ш) додається, починаючи від нижчого граничного значення n та доводиться „по-шагово” до значення верхнього значення n у заданому проміжку (розрахунок також можна вести шляхом віднімання від верхнього значення n у заданому проміжку до нижнього граничного значення n). Таким чином, значення критерію Ірвіна для n розраховано в таблиці 2.

Таблиця 2

Значення критерію Ірвіна для n 4 до 9

N	3	4	5	6	7	8	9	10
λ_n	2,300	2,186	2,071	1,957	1,843	1,729	1,614	1,500

Таким чином, виявлення аномальних рівнів допомагає уникнути помилки так званого технічного рівня, тобто такого, що має суб'єктивний характер та помилки, які в свою чергу можуть призвести дослідника до

неправильних висновків. Такі помилки усуваються взагалі або шляхом заміни аномальних рівнів відповідними значеннями шляхом апроксимації часового ряду, або заміною рівнів середньої арифметичної двох сусідніх рівнів ряду.

Таблиця 3

Значення критерію Ірвіна для $n = [1-100]$

n	λ_n	n	λ_n	n	λ_n	n	λ_n	n	λ_n
1	-	21	1,290	41	1,145	61	1,078	81	1,038
2	-	22	1,280	42	1,140	62	1,076	82	1,036
3	2,300	23	1,270	43	1,135	63	1,074	83	1,034
4	2,186	24	1,260	44	1,130	64	1,072	84	1,032
5	2,071	25	1,250	45	1,125	65	1,070	85	1,030
6	1,957	26	1,240	46	1,120	66	1,068	86	1,028
7	1,843	27	1,230	47	1,115	67	1,066	87	1,026
8	1,729	28	1,220	48	1,110	68	1,064	88	1,024
9	1,614	29	1,210	49	1,105	69	1,062	89	1,022
10	1,500	30	1,200	50	1,100	70	1,060	90	1,020
11	1,480	31	1,195	51	1,098	71	1,058	91	1,018
12	1,460	32	1,190	52	1,096	72	1,056	92	1,016
13	1,440	33	1,185	53	1,094	73	1,054	93	1,014
14	1,420	34	1,180	54	1,092	74	1,052	94	1,012
15	1,400	35	1,175	55	1,090	75	1,050	95	1,010
16	1,380	36	1,170	56	1,088	76	1,048	96	1,008
17	1,360	37	1,165	57	1,086	77	1,046	97	1,006
18	1,340	38	1,160	58	1,084	78	1,044	98	1,004
19	1,320	39	1,155	59	1,082	79	1,042	99	1,002
20	1,300	40	1,150	60	1,080	80	1,040	100	1,000

Аналогічні процедури розрахунку визначення критерію Ірвіна можна провести і для періодів [11-19], [21-29], [31-49], [51-99]. Тоді $Ш_{[11-19]}=0,02$, $Ш_{[21-29]}=0,01$, $Ш_{[31-49]}=0,005$, $Ш_{[51-99]}=0,002$. Відповідна матриця значень критерію

Ірвіна для n розраховано та представлено в таблиці 3.

Слід відзначити, що до цього часу, метод Ірвіна не викривується в економічних дослідженнях з метою оцінки даних на

аномальність статистичних рядів, таке дослідження проводиться вперше. Як правило, дослідники в своїх працях на етапі попереднього аналізу даних на аномальність обмежувались найбільш розповсюдженим прийомом усунення аномальних значень та відсутності тенденції часового ряду за допомогою згладжування часового ряду по: простій ковзаючій середній [1, с. 412], або зваженій ковзаючій середній [1, с. 413] або експоненціальному значенню [1, с. 415].

Після усунення аномальних значень у часовому ряді нами також нами запропоновано з метою підвищення якості дослідження перевірити наявність в вихідних даних ознак тренду та перевірити гіпотезу щодо існування тенденції.

Існують й інші способи та прийоми, які дозволяють на етапі попереднього аналізу даних усунути їх аномальність та відсутності тенденції часового ряду, такі як згладжування часового ряду по: простій ковзаючій середній [1, с. 412-413], зваженій ковзаючій середній [1, с. 413-415] або експоненціальному значенню [1, с. 415-416].

Висновки даного дослідження і перспективи подальших робіт у цьому напрямку.

1 Будь-яке наукове дослідження в спрощеному вигляді проводиться за схемою: «вхідні дані – обробка та дослідження даних за запропонованою методикою – аналіз результатів дослідження». Виникає необхідність розробки та побудови логіко-концептуальної моделі дослідження оцінки ефективності діяльності та управління галузевого комплексу України.

2 Перед проведенням основного дослідження оцінки ефективності діяльності та управління запропоновано провести відповідні процедури щодо оцінки використовуваних в дослідженні інформації на аномальність (провадження вхідного етапу). Проведення

дослідження виявлення аномальних значень відібраних в дослідження показників дозволяє знизити та, в кінцевому разі, усунути ризик як некоректного використання показників в обраній математичній моделі (усунення об'єктивних причин), так і усунення людського фактору щодо надання помилкових або неточних даних в масиві показників (усунення суб'єктивних причин).

3 Під час проведення дослідження виникає необхідність довести значення критерію Ірвіна для рівня значимості λ_n до універсального вигляду, для чого нами були розраховані значення критерію Ірвіна для рівня значимості $\alpha = 0,05$ зворотньо-пропорційним методом та отримані результати доведені до читабельно-універсального вигляду.

4 Отримані результати дослідження дозволяють проводити такі науково-дослідні процедури, як імітування, прогнозування, перспективне планування показників тощо в рамках розробленої моделі не тільки момент часу або за короткий час, але й оцінювати економічні процеси в динаміці та на перспективу.

5 Прийнятий в дійсне дослідження метод Ірвіна до цього часу не викривлювався в економічних дослідженнях з метою оцінки даних на аномальність статистичних рядів, таке дослідження було проведено нами вперше.

СПИСОК ЛІТЕРАТУРИ

1 Захожай В.Б., Литвиненко Я.В., Захожай К.В., Литвиненко Р.Я. Система оподаткування та податкова політика: Навч. посібник / Під заг. ред. В.Б.Захожая та Я.В.Литвиненка. – К.: Центр навчальної літератури, 2006. – 468 с.

2 Красс М.С., Чупрынов Б.П. Математика для экономистов. – СПб.: Питер, 2006. – 464 с.: ил. – (Серия «Учебное пособие»).

Аннотация. В статье детально рассмотрены методологические основы, описан методический инструментальный и порядок проведения исследования на этапе «входные данные», т.е. начального этапа исследования, осуществление которого дает научное обоснование использования или неиспользования в предполагаемом исследовании соответствующих показателей.

Ключевые слова: исследование, показатели, методический инструментальный, оценка показателей, аномальность, критерии аномальности.

Summary. In the article we present a detailed review of methodological foundations, describe the methodological mathematical toolset that we used as well as the research procedure at input data stage, i.e., preliminary stage of research which gives scientific grounds for use or non-use in provided research of corresponding indices.

Keywords: research, indices, methodological toolset, evaluation of indices, anomaly, anomaly criteria.

Рецензент д.е.н., професор УкрДАЗТ Чебанова Н.В.
Експерт редакційної колегії к.е.н., доцент УкрДАЗТ Шраменко О.В.