

УДК 338.24:330.4

## КОМПОНЕНТНИЙ АНАЛІЗ СТАНУ СОЦІАЛЬНО-ЕКОНОМІЧНОГО РОЗВИТКУ УКРАЇНИ

А. Овчаренко

*Львівський національний університет імені Івана Франка  
79008, м. Львів, проспект Свободи, 18*

*Стаття присвячена оцінці стану соціально-економічного розвитку України за допомогою методу головних компонент. Факторний аналіз був застосований в якості методу обробки бази даних, що дало змогу спростити опис вхідних показників, дослідити структуру і характер взаємозв'язків між ними, виключити з аналізу неінформативні показники, усунути дублювання та надлишковість інформації. У роботі ґрунтовно опрацьовано статистичні дані, особливості і тенденції соціально-економічного розвитку держави, апробовано методикою комплексного статистичного аналізу показників з метою розробки рекомендацій, виявлення можливостей і перспективних напрямків вдосконалення подальшої соціальної політики України, зокрема заходів соціального захисту населення.*

*Ключові слова: соціальна політика, соціально-економічні показники, факторний аналіз, метод головних компонент, критерій Кайзера, критерій «кам'яного зсуву» Кеттела, дисперсія, факторні навантаження.*

**Постановка проблеми.** Побудова інноваційної економіки України є пріоритетним завданням на сьогодні. Для вироблення ефективної економічної політики та прийняття оптимальних управлінських рішень, для розуміння того, яким чином проходить становлення і розвиток економіки держави, для вирішення багатьох інших завдань у галузі економічних досліджень, необхідні знання математичної та економічної статистики, необхідне проведення комплексного статистичного дослідження, спрямованого на виявлення макроекономічних факторів. Статистика дозволяє адекватно оцінювати ситуацію, що складається і виявляти тенденції, приймати оперативні й стратегічні рішення. Тому пакети статистичного аналізу даних є настільним робочим інструментом фахівців будь-якого рівня. А для фахівця в галузі управління та економіки знання статистичних методів обробки інформації та сучасних комп'ютерних технологій, які дозволяють автоматизувати громіздкі розрахунки, є абсолютно необхідними. Перебіг та динаміка соціально-економічних процесів держави зазвичай обумовлені дією безлічі різноманітних факторів (чинників). При дослідженні таких багатопараметричних процесів завжди виникає питання: чи не можна відкинути частину параметрів або замінити їх меншим числом будь-яких функцій від них, зберігши при цьому всю інформацію? Зробити це дозволяють методи факторного

аналізу, що є одним з розділів сучасної багатовимірної статистики і широко використовуються в різних областях дослідницької діяльності [1]. У даній роботі проведено математичний аналіз статистичної інформації щодо динаміки соціально-економічного розвитку України на основі методу головних компонент. Це дає змогу досить просто і точно описати спостережувані вихідні дані, структуру і характер взаємозв'язків між ними. Стиснення інформації здійснюється за рахунок того, що число чинників або головних компонент (нових одиниць виміру) значно скорочується у порівнянні із кількістю вхідних ознак.

**Аналіз останніх досліджень та публікацій.** Метод факторного аналізу ґрунтовно досліджено багатьма науковцями, зокрема свої праці вивченню цього питання присвятили Стівенс, Кім, Мюллер, Клекка, Олденерфер і Блешфілд, Кулі і Лонес, Лоулі та Максвел, Харман, Ліндемман, Меренда і Голд, Моррісон і Мулейк, Веррі., Кеттел, Пірсон, Гальтон та багато інших. Однак проведений аналіз наукових публікацій показав, що потенціал статистичних методів при аналізі соціально-економічного стану України для вироблення ефективної економічної політики та прийняття оптимальних управлінських рішень використаний не повною мірою, що зумовило вибір теми даного дослідження.

#### **Формулювання цілей статті.**

Метою даної роботи є проведення економіко-статистичного аналізу стану, особливостей і тенденцій соціально-економічного розвитку держави, розробка та апробація методики комплексного статистичного аналізу соціально-економічних показників, виявлення взаємозв'язків між даними змінними.

**Виклад основного матеріалу.** Для виконання факторного аналізу початкові ознаки потрібно перш за все стандартизувати. Оскільки компоненти є гіпотетичними величинами, то виміряти їх можна лише опосередковано за допомогою спеціально сконструйованих моделей. Нехай  $Z_j$  – стандартизоване значення  $j$ -ї первинної ознаки з одиничною дисперсією. У моделі головних компонент зв'язок між цими первинними ознаками і компонентами  $F_k$  описується як лінійна комбінація [2]:

$$z_j = \sum_{k=1}^J a_{jk} F_k, \quad (1)$$

де  $a_{jk}$  – факторне навантаження  $k$ -ї компоненти на  $j$ -ту ознаку ( $j = \overline{1, J}$ ).

Сумарна дисперсія дорівнює кількості ознак  $J$ .

В повній моделі головних компонент (компонентному аналізі) чисельність компонент співпадає з кількістю первинних ознак. Звідси випливає, що всі  $J$  компонент повністю пояснюють сумарну дисперсію ознакової множини. Враховуючи те, що за умови ортогональності компонент квадрат факторного навантаження  $a_{jk}^2$  характеризує внесок  $k$ -ї компоненти у варіацію  $j$ -ї ознаки, повний внесок цієї компоненти в сумарну дисперсію  $J$  ознак становить

$\lambda_k = \sum_{j=1}^J a_{jk}^2$ . У термінах матричної алгебри дисперсії компонент  $\lambda_k$  - це влас- тиві числа кореляційної матриці  $R$ , недіагональні елементи якої представлені кое- фіцієнтами кореляції  $r_{ls}$ , які оцінюють супутні зв'язки між ознаками  $x_l$  та  $x_s$  [3].

У процесі компонентного аналізу сумарна варіація  $J$  первинних ознак  $x_j$  перерозподіляється між компонентами  $F_k$  з дисперсіями  $k_\lambda$ . Тобто сумарну дисперсію ознакової множини можна представити як суму дисперсій компонент  $\sum_{k=1}^J \lambda_k$  або через факторні навантаження:

$$j = \sum_{k=1}^J \lambda_k = \sum_{k=1}^J \sum_{j=1}^J a_{jk}^2. \quad (2)$$

Для проведення комплексного статистичного дослідження була побудована система показників на основі попереднього змістовного аналізу вже існуючих офіційних даних Держкомстату. Емпіричною базою для аналізу було обрано динамічні ряди значень 34 соціально-економічних показників за період 2000-2009 рр.[4, 5, 6]. При виборі цих показників враховувалася їх адекватність, точність, об'єктивність, однозначність тлумачення, порівнянність і унікальність.

Таблиця 1.

**База соціально-економічних показників, що були використані при компонентному аналізі**

Назва статистичного показника	Позначення, $x_j$ ( $j = 1;34$ )
Кількість населення, млн. чол. (на початок року)	x1
Природний приріст (скорочення)	x2
Кількість пенсіонерів, млн. чол. (на 1 січня)	x3
ВВП України у розрахунку на одну особу, грн.	x4
Експорт товарів та послуг, млн.дол. США	x5
Імпорт товарів та послуг, млн.дол. США	x6
Доходи Зведеного бюджету України, млрд. грн.	x7
Видатки Зведеного бюджету України, млрд. грн.	x8
Загальна сума державного та гарантованого державою боргу України, млрд.дол.США	x9
Прямі інвестиції в Україну, млн. дол. США (на початок року)	x10
Прямі інвестиції з України, млн. дол. США (на початок року)	x11
Офіційний курс нац. грошової одиниці до долара США, встановлений Нацбанком України (на кінець року)	x12
Індекс споживчих цін, % (рік до попереднього року)	x13
Обсяг реалізованої продукції промисловості (у фактичних цінах), млн.грн.	x14
Кількість найманих працівників на малих підприємствах, тис.	x15
Обсяг реалізованої продукції (робіт, послуг) малих підприємств, млн. грн.	x16
Вантажооборот, млрд. т/км	x17

Пасажирооборот, млрд. пас./км	x18
Економічно активне населення у віці 15-70 років в середньому, тис. осіб	x19
Зайняте населення у віці 15-70 років в середньому, тис. осіб	x20
Безробітне населення (за методологією МОП) у віці 15-70 років, в середньому тис.осіб	x21
Рівень безробіття (за методологією МОП) у віці 15-70 років, у % до економічно активного населення відпов. вікової групи	x22
Зареєстровані безробітні, тис. осіб	x23
Кількість незайнятих громадян, які скористалися послугами державної служби зайнятості (в цілому за рік), тис. осіб	x24
з них працевлаштовано, тис. осіб	x25
Навантаження незайнятого населення на одне вільне робоче місце, вакантну посаду, осіб (на кінець року)	x26
Середньомісячна номінальна заробітна плата, грн.	x27
Мінімальна заробітна плата, грн.	x28
Заборгованість із виплати заробітної плати, млрд. грн (на початок року)	x29
Середній розмір місячної пенсії, грн. (на початок року)	x30
Середньозважений за рік розмір прожиткового мінімуму для працездатної особи, грн.	x31
Межа бідності, грн. на 1 особу на місяць	x32
Рівень бідності, %	x33
Коефіцієнт Джині, %	x34

Сформований масив багатовимірних даних аналізувався статистичним пакетом Statistica 8, а також Statgraphics Plus. Метою аналізу було скорочення великої множини соціальних і економічних індикаторів та одержання невеликого числа лінійних комбінацій із 34 змінних. Це передбачає перетворення змінних у нову множину попарно некорельованих змінних, серед яких, перша відповідає напрямку максимально можливої дисперсії, друга – напрямку максимально можливої дисперсії в просторі, який ортогональний першому напрямку і т.д. На практиці визначити кількість головних компонент можна за допомогою кількох методичних процедур, найкращі результати з яких показують такі методи, як критерій Кайзера та критерій «кам'яного зсуву» Кеттела [2]. Відповідно до критерію Кайзера [7] для визначення достатнього числа факторів слід виділяти тільки головні компоненти із власними значеннями, більшими за одиницю (в даному випадку перші три) (рис. 1). Суть критерію «кам'яного зсуву» Кеттела полягає в тому, що власні значення, представлені на рисунку 1, зображено у вигляді простого графіка. Кеттел запропонував знайти таке місце на графіку, де спадання власних значень зліва направо максимально сповільнюється.

Отже, перед обертанням було використано графічний критерій «кам'яного зсуву» (кам'янистого осипу, *Scree plot*), (рис 2). Як бачимо, власні значення кореляційної матриці (вісь ординат) на початку стрімко зменшуються і лише три перших більші за одиницю. Внесок більшості компонент у сумарну варіацію незначний, тому для дослідження, зазвичай, беруть найважливіші (ті перші фактори, котрі ще не потрапляють на пологу частину графіка), які власне і називають головними.

Сподіваючись отримати результат, який можна інтерпретувати в предметній області, було застосовано метод обертання (поворот осей) *Varimax normalized*.

Таким чином було виділено 3 головні компоненти (рис. 3), що показують максимальний відсоток (93,216%) сумарної дисперсії множини вхідних даних, що

свідчить про високий ступінь факторизації. Зокрема перша компонента пояснює 56,811 % дисперсії, друга – 27,533% і третя – 8,852 % відповідно.

Eigenvalues of correlation matrix, and related statistics (Untitled1 (B2:K11))					
Active variables only					
Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %	
1	23,02914	67,73277	23,02914	67,7328	
2	6,39363	18,80479	29,42277	86,5376	
3	2,27067	6,67845	31,69344	93,2160	
4	0,98011	2,88266	32,67355	96,0987	
5	0,64136	1,88636	33,31491	97,9850	
6	0,40510	1,19147	33,72001	99,1765	
7	0,12517	0,36813	33,84518	99,5446	
8	0,10563	0,31067	33,95081	99,8553	
9	0,04919	0,14469	34,00000	100,0000	

Рис. 1. Власні значення кореляційної матриці та сумарна дисперсія головних компонент (до процедури обергання)

### Scree Plot

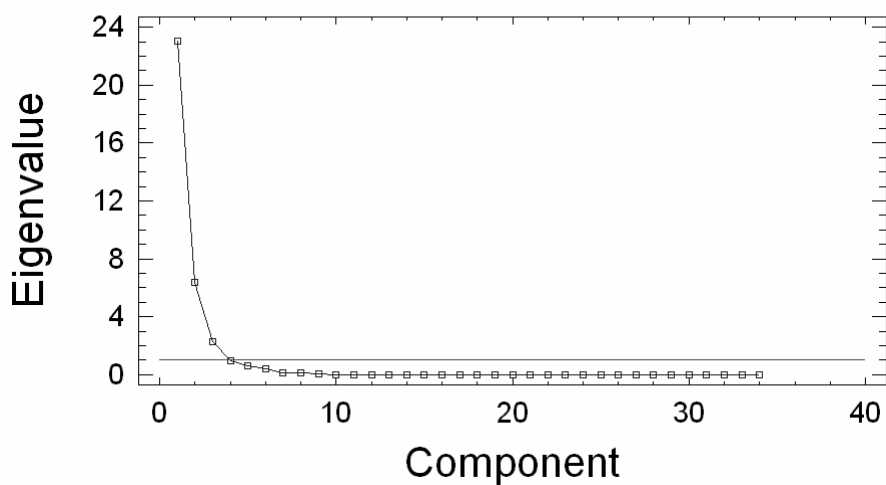


Рис.2. Критерій «кам'яного зсуву» (графік)

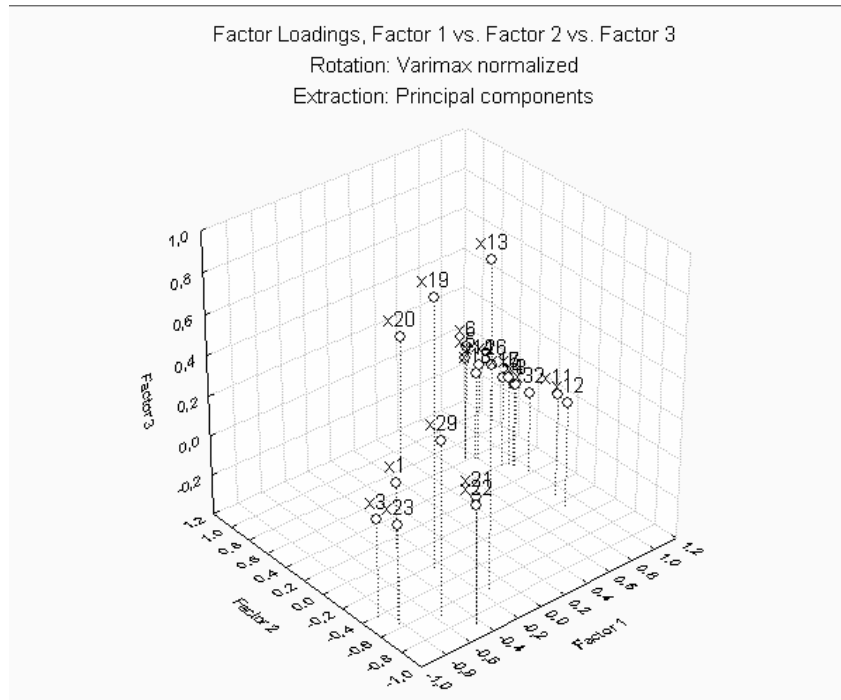


Рис. 3. Діаграма розсіювання факторних навантажень

Результати компонентного аналізу, зокрема факторні навантаження після обертання вказані у таблиці 2.

Таблиця 2

Значення факторних навантажень після обертання

Факторна ознака $x_j (j = \overline{1;34})$	Ваговий коефіцієнт		
	Z1	Z2	Z3
x1	-0,80079	-0,514388	0,275748
x2	0,97673	0,049832	-0,053634
x3	-0,88162	-0,416125	0,087322
x4	0,93483	0,353236	0,016293
x5	0,74314	0,637607	0,118787
x6	0,74025	0,629215	0,177077
x7	0,92032	0,385744	0,047823
x8	0,93926	0,339154	0,020376
x9	0,94300	-0,261798	-0,124101
x10	0,97637	0,191514	-0,014822
x11	0,92080	-0,104531	0,117306
x12	0,88796	-0,243696	0,122574
x13	0,37081	-0,062278	0,895109
x14	0,80989	0,576220	0,078959
x15	0,61451	0,558284	-0,309014
x16	0,87076	0,394215	0,060859

x17	0,09618	0,944759	-0,095722
x18	0,79170	0,583363	0,038206
x19	-0,25331	-0,229368	0,922249
x20	0,36901	0,884130	0,240267
x21	-0,43379	-0,863473	0,223976
x22	-0,43651	-0,871167	0,186439
x23	-0,82052	-0,553675	0,087526
x24	-0,92172	0,126235	-0,042685
x25	0,07361	0,968718	-0,192330
x26	-0,13293	-0,643657	0,701599
x27	0,94981	0,309216	-0,022553
x28	0,94709	0,277627	-0,108208
x29	-0,55375	-0,667836	0,460218
x30	0,97361	0,195641	-0,033068
x31	0,93872	0,321902	-0,094470
x32	0,96850	0,235869	0,002156
x33	-0,26789	0,703751	-0,289868
x34	-0,35510	0,447075	-0,251047
Expl. Var	19,31572	9,368079	3,009649
Prp. Totl	0,56811	0,275532	0,088519

Тепер знайдене рішення вже можна інтерпретувати. Фактори зазвичай інтерпретують по навантаженнях на ознаки. Згідно з отриманими результатами до структури першого фактора (компоненти) увійшло 22 соціально-економічних показника з коефіцієнтами факторних навантажень за модулем від 0,74025 до 0,97673. Першу компонентну можна інтерпретувати, як демографо-економічний фактор. До структури другої компоненти увійшло 6 індикаторів, а третьої – 3 (з коефіцієнтами факторних навантажень за модулем від 0,703751 до 0,968718 для другої та від 0,701599 до 0,922249 для третьої компоненти відповідно). Друга і третя компоненти подібні за своєю суттю. Друга навантажує ознаки, які характеризують стан зайнятості населення та рівень його благополуччя, а третя – ознаки трудового потенціалу.

Отже, загальне рівняння головних компонент можемо представити наступним чином:

$$z_{kt} = \sum_{j=1}^J w_{kj} y_{jt}, \quad (3)$$

де  $z_{kt}$  – значення  $l$ -ої головної компоненти у період  $t$  ( $k = \overline{1; K}$ ;  $t = \overline{1; T}$ )

$w_{kj}$  – ваговий коефіцієнт, що оцінює значущість впливу  $j$ -ого показника на  $k$ -ту компоненту ( $j = \overline{1; J}$ ;  $k = \overline{1; K}$ )

$y_{jt}$  – значення  $j$ -ого соціально-економічного показника у період  $t$  ( $j = \overline{1; J}$ ;  $t = \overline{1; T}$ ).

$K$  – кількість виділених головних компонент;

$T$  – кількість досліджуваних періодів;

$J$  – кількість вхідних показників.

Використовуючи отримані дані конкретизуємо формулу (3). Тоді залежність кожної головної компоненти від вхідних показників можна записати у вигляді наступних рівнянь:

$$\begin{aligned}
 z_{1t} = & -0,80079 * Y_{1t} + 0,97673 * Y_{2t} - 0,88162 * Y_{3t} + 0,93483 * Y_{4t} + \\
 & + 0,74314 * Y_{5t} + 0,74025 * Y_{6t} + 0,92032 * Y_{7t} + 0,93926 * Y_{8t} + \\
 & + 0,94300 * Y_{9t} + 0,97637 * Y_{10t} + 0,92080 * Y_{11t} + 0,88796 * Y_{12t} + \\
 & + 0,37081 * Y_{13t} + 0,80989 * Y_{14t} + 0,61451 * Y_{15t} + 0,87076 * Y_{16t} + \\
 & + 0,09618 * Y_{17t} + 0,79170 * Y_{18t} - 0,25331 * Y_{19t} + 0,36901 * Y_{20t} - \\
 & - 0,43379 * Y_{21t} - 0,43651 * Y_{22t} - 0,82052 * Y_{23t} - 0,92172 * Y_{24t} + \\
 & + 0,07361 * Y_{25t} - 0,13293 * Y_{26t} + 0,94981 * Y_{27t} + 0,94709 * Y_{28t} - \\
 & - 0,55375 * Y_{29t} + 0,97361 * Y_{30t} + 0,93872 * Y_{31t} + 0,96850 * Y_{32t} - \\
 & - 0,26789 * Y_{33t} - 0,35510 * Y_{34t} \quad (4)
 \end{aligned}$$

$$\begin{aligned}
 z_{2t} = & -0,514388 * Y_{1t} + 0,049832 * Y_{2t} - 0,416125 * Y_{3t} + 0,353236 * Y_{4t} + \\
 & + 0,637607 * Y_{5t} + 0,629215 * Y_{6t} + 0,385744 * Y_{7t} + 0,339154 * Y_{8t} - \\
 & - 0,261798 * Y_{9t} + 0,191514 * Y_{10t} - 0,104531 * Y_{11t} - 0,243696 * Y_{12t} - \\
 & - 0,062278 * Y_{13t} + 0,576220 * Y_{14t} + 0,558284 * Y_{15t} + 0,394215 * Y_{16t} + \\
 & + 0,944759 * Y_{17t} + 0,583363 * Y_{18t} - 0,229368 * Y_{19t} + 0,884130 * Y_{20t} - \\
 & - 0,863473 * Y_{21t} - 0,871167 * Y_{22t} - 0,553675 * Y_{23t} + 0,126235 * Y_{24t} + \\
 & + 0,968718 * Y_{25t} - 0,643657 * Y_{26t} + 0,309216 * Y_{27t} + 0,277627 * Y_{28t} - \\
 & - 0,667836 * Y_{29t} + 0,195641 * Y_{30t} + 0,321902 * Y_{31t} + 0,235869 * Y_{32t} + \\
 & + 0,703751 * Y_{33t} + 0,447075 * Y_{34t} \quad (5)
 \end{aligned}$$

$$\begin{aligned}
 z_{3t} = & 0,275748 * Y_{1t} - 0,053634 * Y_{2t} + 0,087322 * Y_{3t} + 0,016293 * Y_{4t} + \\
 & + 0,118787 * Y_{5t} + 0,177077 * Y_{6t} + 0,047823 * Y_{7t} + 0,020376 * Y_{8t} - \\
 & - 0,124101 * Y_{9t} - 0,014822 * Y_{10t} + 0,117306 * Y_{11t} + 0,122574 * Y_{12t} + \\
 & + 0,895109 * Y_{13t} + 0,078959 * Y_{14t} - 0,309014 * Y_{15t} + 0,060859 * Y_{16t} - \\
 & - 0,095722 * Y_{17t} + 0,038206 * Y_{18t} + 0,922249 * Y_{19t} + 0,240267 * Y_{20t} +
 \end{aligned}$$



$$\begin{aligned}
& +0,223976* Y_{21t} +0,186439* Y_{22t} +0,087526* Y_{23t} -0,042685* Y_{24t} - \\
& -0,192330* Y_{25t} +0,701599* Y_{26t} -0,022553* Y_{27t} -0,108208* Y_{28t} + \\
& +0,460218* Y_{29t} -0,033068* Y_{30t} -0,094470* Y_{31t} +0,002156* Y_{32t} - \\
& -0,289868* Y_{33t} -0,251047* Y_{34t} , \quad (6)
\end{aligned}$$

де  $t = \overline{1; 10}$ .

**Висновки.** Прийняття будь-якого рішення неприпустиме без статистичного аналізу інформації, оскільки виділяти закономірності із випадковостей, порівнювати ймовірні альтернативи вибору, будувати прогнози розвитку процесів, виявляти зв'язки і відмінності безлічі об'єктів можливе тільки і виключно засобами математичної статистики. При вивченні складних соціально-економічних процесів факторний аналіз дозволяє глибше зрозуміти їх сутність, що є надзвичайно важливим при розробці стратегії розвитку соціальної політики держави, особливо за умов перехідної економіки.

Проведений аналіз в подальшому дозволяє математично обґрунтувати рекомендації щодо вдосконалення заходів соціального захисту. Даний підхід значно спрощує розрахунки, не впливаючи на точність результатів аналізу, дає змогу досить просто інтерпретувати вихідні дані, структуру і характер взаємозв'язків між ними, використати отримані результати при розробці методики моделювання основних показників соціально-економічного розвитку.

1. А.И. Бахтин, Н.М. Низамутдинов, Н.М. Хасанова, Е.М. Нуриева. Факторный анализ в геологии. - Казань 2007
2. Голубник О. Компонентний аналіз чинників трудової міграції населення України. // Вісник Львівського університету. Серія економічна. – 2009. – С.434-442.
3. Приймак В.І. Трудовий потенціал і механізм його реалізації в регіоні: Монографія. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2002. – С. 124.
4. <http://www.ukrstat.gov.ua/> – офіційний сайт Державного комітету статистики України.
5. Статистичний щорічник України за 2008 рік / За ред. О. Г. Осауленка. – К. – 2009.
6. Домогосподарства Львівської області у 2003-2008 роках. Статистичний щорічник. – Львів, Головне управління статистики у Львівській області. – 2010.
7. Минаева О.Н., Тихов М.С. Построение модели измерения экономики знаний с использованием метода главных компонент // Обозрение прикладной и промышленной математики, 2010. Т. 17. № 2. – С. 286—288.

**COMPONENT ANALYSIS OF THE SOCIO-ECONOMIC DEVELOPMENT OF UKRAINE****A. Ovcharenko***Ivan Franko National University of Lviv,  
Svobody Ave., 18, UA – 79008, Lviv*

The article is devoted to investigation of Ukrainian socio-economic development by the method of principal components. Factor analysis was applied as a database reduction and structure detection method, allowing us to simplify the description of input parameters, to investigate the structure and nature of the relationships between them, eliminate duplication and redundancy of information. The author investigates statistic data, features and trends of socio-economic development, approves the methods of complex statistical analysis to develop recommendations, identify opportunities for further improvement of social policy of Ukraine, in particular the priorities of social protection of population.

Keywords: social policy, socio-economic indicators, factor analysis, principal components method, the criterion of Kaiser, the scree criterion of Cattell, dispersion, factor loadings..

**КОМПОНЕНТНЫЙ АНАЛИЗ СОСТОЯНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ УКРАИНЫ****А. Овчаренко***Львовский национальный университет имени Ивана Франко  
79008, г. Львов, проспект Свободы, 18*

Статья посвящена оценке состояния социально-экономического развития Украины с помощью метода главных компонент. Факторный анализ был применен в качестве метода обработки базы данных, что позволило упростить описание входных показателей, исследовать структуру и характер взаимосвязей между ним, исключить из анализа неинформативные показатели, устранить дублирование и избыточность информации. В работе основательно изучены статистические данные, особенности и тенденции социально-экономического развития государства, апробирована методика комплексного статистического анализа показателей с целью разработки рекомендаций, выявления возможностей и перспективных направлений совершенствования дальнейшей социальной политики Украины, в частности мер социальной защиты населения.

**Ключевые слова:** социальная политика, социально-экономические показатели, факторный анализ, метод главных компонент, критерий Кайзера, критерий «каменистой осыпи» Кеттела, дисперсия, факторные нагрузки.