

## ЗАЛЕЖНІСТЬ ЕФЕКТИВНОСТІ ПРОГНОЗУВАННЯ ПОДАТКОВИХ НАДХОДЖЕНЬ ВІД ДОВЖИНИ СТАТИСТИЧНОЇ ВИБІРКИ

Степашко В.С., Коппа Ю.В.

Мета проведених досліджень - вивчити ефективність застосування системи АСТРІД, в основі якої лежать алгоритми МГУА, для побудови прогнозних моделей в задачах моніторингу економічних - у даному разі податкових - процесів. Основні відмінності від попередніх досліджень [1]: по-перше, розглядаються умови наявності значно *довших статистичних рядів*; по-друге, вивчаються можливості застосування побудованих моделей для *довгострокового прогнозування* у залежності від довжини інтервалу "навчання" моделі.

Як приклад наведемо результати побудови за допомогою МГУА моделей для прогнозування зміни обсягу *надходжень до зведеного державного бюджету*. Для цієї задачі було обрано 14 показників (факторів), відібраних за методикою [1] які можуть впливати на цільовий показник - доходи консолідованого бюджету. Щомісячні значення всіх показників було взято з бюлетеня Держкомстату за період з листопада 1995 по вересень 1999 рр. (тобто довжина початкової вибірки – 47 точок). За цими даними із застосуванням МГУА будувались оптимальні моделі залежності доходу бюджету від поточних значень найістотніших факторів.

Отже, за даними [2-4] було сформовано вибірку для таких вхідних змінних (відбирались ті статистичні параметри, де не було пропусків даних):

- X1 - грошові доходи населення, всього (млн.грн.);
- X2 - грошові витрати та заощадження населення, всього (млн.грн.);
- X3 - індекс споживчих цін (%);
- X4 - індекс цін виробників легкої промисловості (%);
- X5 - індекс ВВП (%);
- X6 - роздрібний товарообіг всього (млн.грн.);
- X7 - загальна зайнятість: чисельність робітників і службовців (тис.чол);
- X8 - зайнятість у легкій промисловості (промислово-виробничий персонал, тис.чол);
- X9 - заробітна плата (середньомісячна нарахована на одного працівника, номінальна, грн.);
- X10 - індекс реальної зарплати ("нетто", %);
- X11 - грошова маса готівки в обігу (M0) (млн.грн.);
- X12 - офіційний курс гривні до 100 доларів США (грн.);
- X13 - кредиторська заборгованість між підприємствами України (млн.грн.);
- X14 - видатки зведеного бюджету, всього (млн.грн.).

В якості цільової (вихідної) змінної Y було прийнято *доходи зведеного бюджету*, всього (млн.грн.). Загальна довжина вибірки становить 47 місяців.

При цьому зазначимо такі особливості прийнятої постановки задачі досліджень:

- для простоти розглядається лише клас лінійних моделей;
- будуються моделі для визначення обсягу доходу зведеного бюджету поточного місяця за значеннями вхідних змінних цього ж місяця;
- аналізується вплив збільшення довжини навчальної частини вибірки на структуру моделі оптимальної складності, тобто на кількість і склад найістотніших факторів;
- аналізуються довгострокові прогнозні властивості одержаних моделей за таких умов: із загальної довжини вибірки  $N=47$  точок (місяців) для *навчання* (побудови) моделі використовуються  $N1 < N$  *перших* точок, а на решті  $N-N1$  точках перевіряється ефективність моделей у режимі *контрольного прогнозування*.

Наведемо одержані за допомогою системи АСТРІД результати для кількох значень N1, а саме: N1 = 14; 18; 24; 29; 39 точок (місяців).

**1. N1 = 14**, тобто для побудови моделі використовуються перші 14 точок (листопад 1995 р. - грудень 1996 р.), а для її перевірки (контрольного прогнозування) - решта N-N1=33 точки.

Одержана модель має вигляд:

$$y = -6007 + 0.5776 * x^2 + 19.435 * x^5 - 5.1717 * x^9 + 20.385 * x^{12} + 0.67185 * x^{14} . \quad (1)$$

Модель має такі характеристики: СКО=89.57 - середньоквадратична похибка моделювання; R=200.16 - максимальна абсолютна різниця між фактичним і модельним значеннями; R<sub>відн</sub>=6.76% - максимальна різниця у процентах від розмаху вихідної змінної (усі характеристики обчислені на навчальній вибірці даних). На рис. 1 показано графік зміни оцінок бюджетних надходжень за цією моделлю у порівнянні з табличними даними (дійсною зміною надходжень).

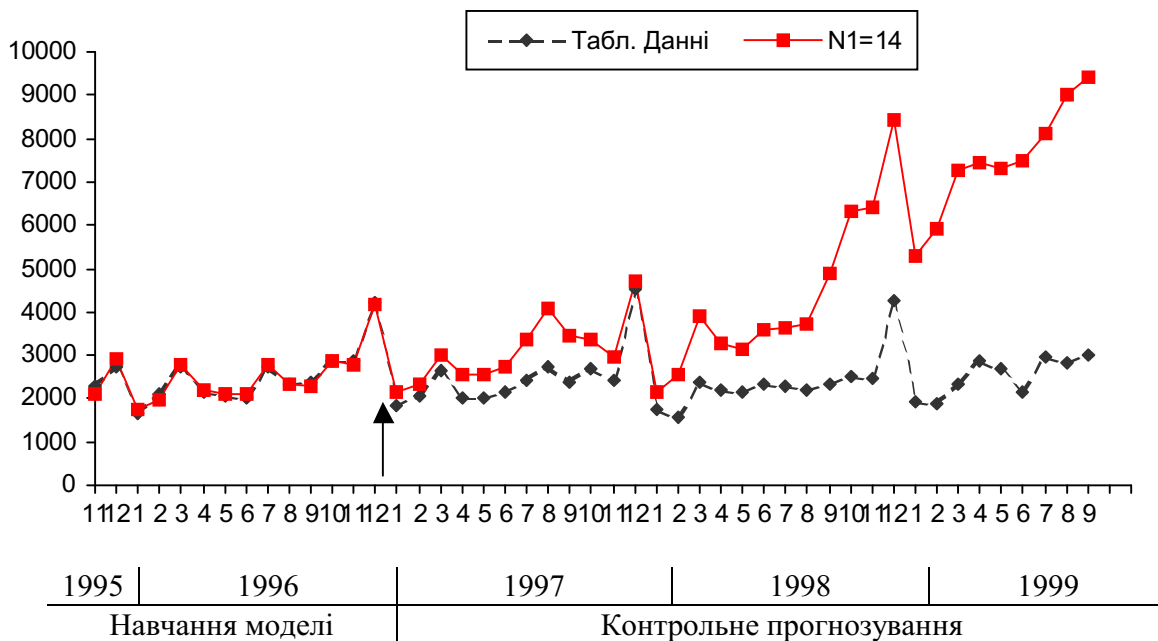


Рис.1 Графіки зміни доходів зведеного бюджету та його оцінок за моделлю (1), побудованою за першими 14 (з 47 наявних) точками вибірки (листопад 1995 р. - грудень 1996 р.)

З графіка видно, що побудована модель (1) добре описує модельований процес на інтервалі навчання і на перших трьох точках (січень-березень 1997 р.) контрольного інтервалу, а далі оцінки дедалі більше відхиляються від реального ходу процесу. Отже, за таким доволі малим числом даних (14 точок за умови наявності 14 факторів) можна побудувати модель лише для короткострокового прогнозування.

**2. N1 = 18**, тобто для побудови моделі використовуються перші 18 точок (листопад 1995 р. - квітень 1997 р.), а для контрольного прогнозування - решта N-N1=29 точок.

Одержано модель:

$$y = 0.75931 * x_2 - 42.771 * x_4 + 0.32946 * x_7 - 15.127 * x_8 - 0.56748 * x_{11} + 25.444 * x_{12} + 0.65745 * x_{14} \quad (2)$$

з характеристиками: СКО=86.57; R=231.52; R<sub>відн</sub>=7.6% . На рис. 2 показано графік зміни оцінок бюджетних надходжень за цією моделлю у порівнянні з табличними даними.

Видно, що модель (2) досить добре описує модельований процес на інтервалі навчання і принаймні на перших шести точках (травень-жовтень 1997 р.) контрольного інтервалу, а далі оцінки знов дедалі більше відхиляються від реального ходу процесу. Отже, збільшення вибірки навіть на 4 точки (за наявності 14 факторів) збільшило прогнозу ефективність побудованої моделі, але її все одно можна застосовувати лише для короткострокового прогнозування.

Зазначимо, що зі збільшення навчальної вибірки зросло число членів моделі, причому фактори X<sub>2</sub>, X<sub>12</sub> та X<sub>14</sub> входять до обох моделей, тобто є найістотнішими.

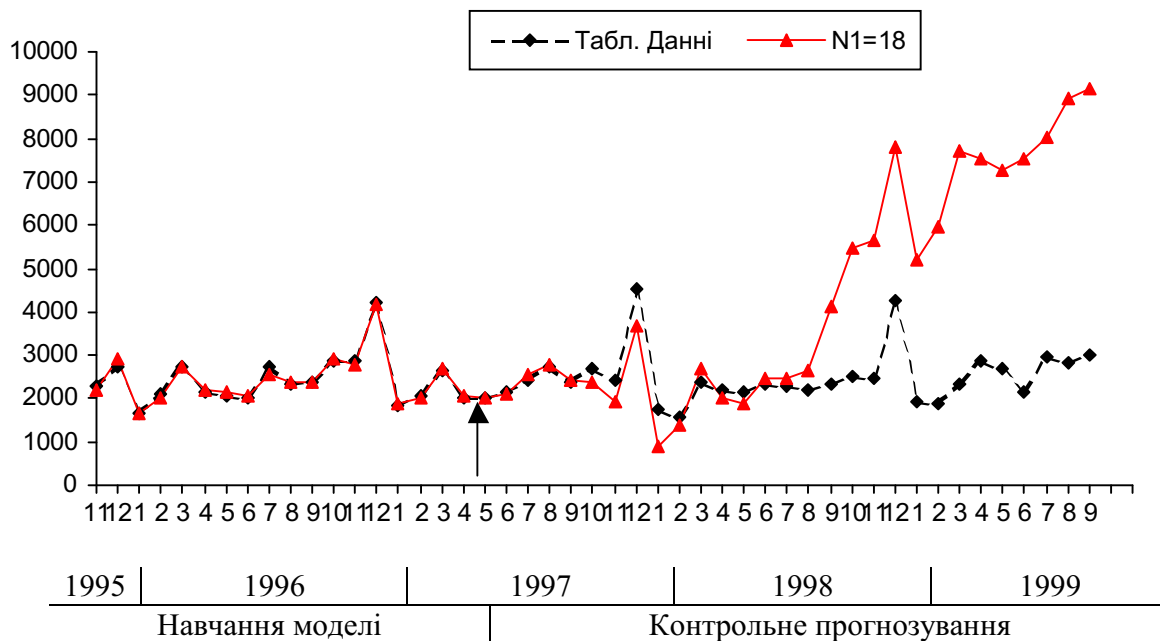


Рис. 2 Графіки зміни доходів зведеного бюджету та його оцінок, одержаних за моделлю (2), побудованою за першими 18 (з 47 наявних) точками статистичної вибірки

**3. N<sub>1</sub> = 24**, тобто для побудови моделі використовуються перші 24 точки (листопад 1995 р. - жовтень 1996 р.), а для контрольного прогнозування - решта N-N<sub>1</sub>=23 точки.

Одержана модель має вигляд:

$$y = 279.41 - 0.084185 * x_1 + 0.73103 * x_2 - 38.409 * x_4 - 0.75364 * x_6 - 0.8337 * x_8 + 9.0591 * x_{10} - 0.31535 * x_{11} + 19.041 * x_{12} + 0.66703 * x_{14} \quad (3)$$

з характеристиками: СКО=185.95; R=363.64; R<sub>відн</sub>=12.28% .

На рис. 3 показано графік зміни оцінок бюджетних надходжень за цією моделлю у порівнянні з табличними даними (дійсною зміною надходжень).

З графіка видно, що модель (3) досить добре описує модельований процес на інтервалі навчання і принаймні на перших десяти точках (листопад 1997 р. - серпень 1998 р.) контрольного інтервалу, а далі оцінки знову дедалі більше відхиляються від реального ходу процесу. Отже, подальше збільшення числа даних ще на 6 точок (за наявності 14 факторів) збільшило прогнозу ефективність побудованої моделі, і її вже можна застосовувати для коротко - і середньострокового прогнозування.

Зазначимо, що після нового збільшення навчальної вибірки знову зросло число членів моделі, причому фактори  $X_2$ ,  $X_{12}$  та  $X_{14}$  входять до всіх трьох моделей, тобто є найістотнішими.

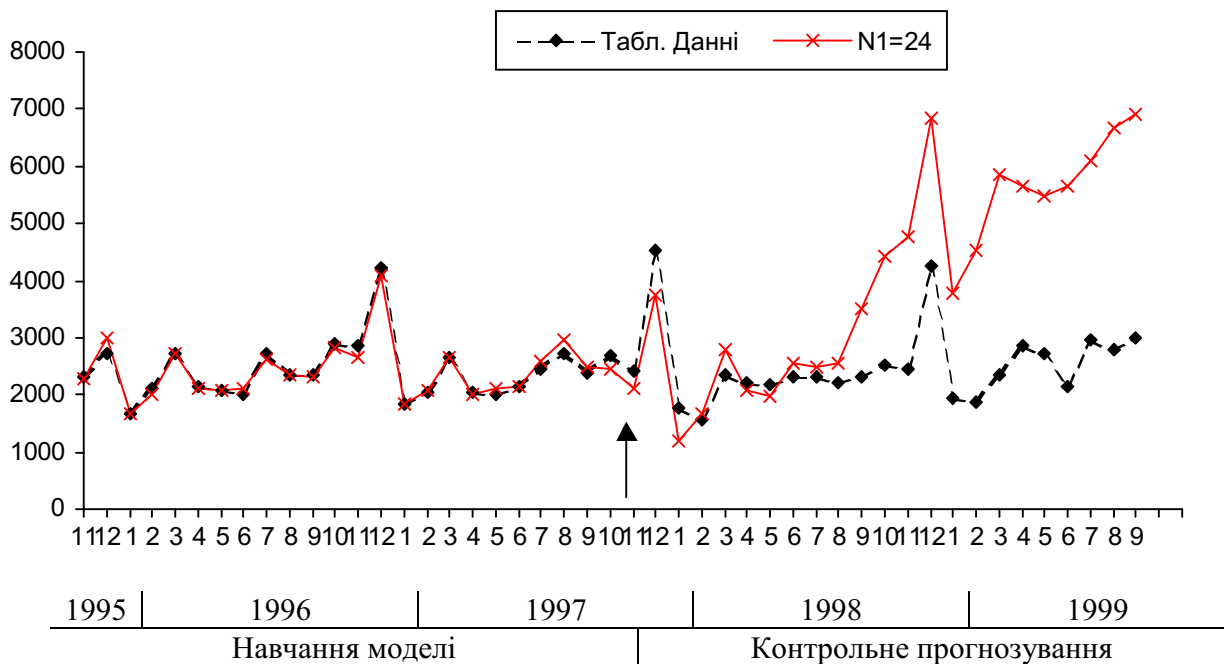


Рис. 3 Графіки зміни доходів зведеного бюджету та його оцінок, одержаних за моделлю (3), побудованою за першими 24 (з 47 наявних) точками статистичної вибірки

**4.  $N_1 = 29$** , тобто для побудови моделі використовуються перші 29 точок (листопад 1995 р. - березень 1998 р.), а для контрольного прогнозування - решта  $N-N_1=33$  точки.

Одержана модель має вигляд:

$$y = -3161.1 - 0.76472 \cdot x_1 + 0.66463 \cdot x_2 + 12.55 \cdot x_3 + 0.0069 \cdot x_5 + 3.9681 \cdot x_9 + 10.593 \cdot x_{10} - 0.028068 \cdot x_{11} + 3.791 \cdot x_{12} + 0.9073 \cdot x_{14} \quad (4)$$

з характеристиками:  $СКО=167.28$ ;  $R=382.55$ ;  $R_{відн}=12.91\%$ . На рис. 4 показано графік зміни оцінок бюджетних надходжень за цією моделлю в порівнянні з табличними даними (дійсною зміною надходжень).

З графіка видно, що модель (4) досить добре описує модельований процес на інтервалі навчання і принаймні на перших десяти точках (квітень 1998 р. - січень 1999р.) контрольного інтервалу, а далі оцінки дещо більше відхиляються від реального ходу процесу, та все ж якісно

його відтворюють. Отже, вже за такого співвідношення числа навчальних (29) і контрольних (18) точок даних (за наявності 14 факторів) прогнозна ефективність побудованої моделі значно зросла, її структура наблизилась до відтворення закономірностей модельованого процесу і її вже можна застосовувати для коротко-, середньо- та оціночного довгострокового прогнозування.

Зазначимо, що нове збільшення навчальної вибірки вже не призвело до подальшого зростання числа членів моделі, причому включені в модель (4) фактори змінились неістотно у порівнянні з моделлю (3).

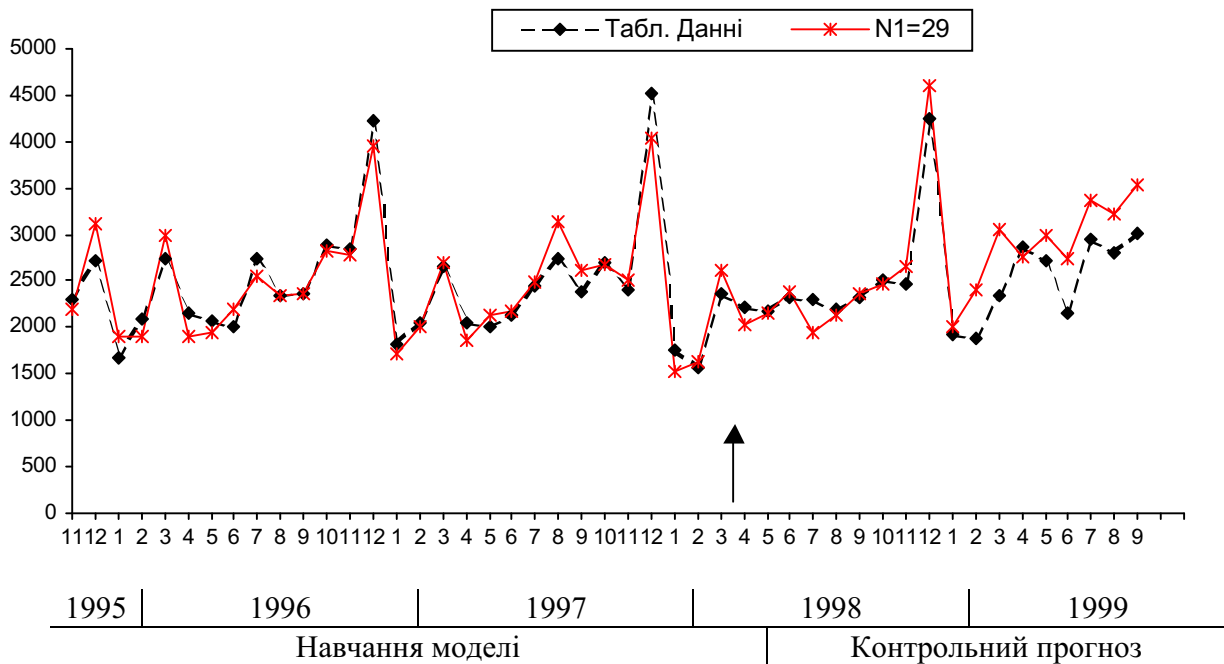


Рис. 4 Графіки зміни доходів зведеного бюджету та його оцінок, одержаних за моделлю (4), побудованою за першими 29 (з 47 наявних) точками статистичної вибірки

**5.  $N_1 = 38$ ,** тобто для побудови моделі використовуються перші 38 точок (листопад 1995 р. - грудень 1998 р.), а для контрольного прогнозування - решта  $N-N_1=9$  точок.

Одержана модель має вигляд:

$$y = -4895.9 - 0.67339 \cdot x_1 + 0.51653 \cdot x_2 + 31.106 \cdot x_3 - 0.80295 \cdot x_8 + 7.2679 \cdot x_9 + 13.296 \cdot x_{10} + 1.7832 \cdot x_{12} + 0.83299 \cdot x_{14}. \quad (5)$$

Модель має такі характеристики:  $СКО=206.45$  ;  $R=460.71$  ;  $R_{\text{відн}}=16.93\%$ . На рис. 5 показано графік зміни оцінок бюджетних надходжень за цією моделлю в порівнянні з табличними даними (дійсною зміною надходжень).

З графіка видно, що модель (5) досить добре описує модельований процес як на інтервалі навчання, так і на контрольному інтервалі, тобто оцінки відтворюють реальний хід процесу і якісно, і кількісно. Отже, можна зробити висновок, що за такого співвідношення числа навчальних (38) і контрольних (9) точок даних (за наявності 14 факторів) структура побудованої моделі відтворює закономірності модельованого процесу, і її прогнозна ефективність дозволяє застосовувати її для кількісного коротко-, середньо- та довгострокового прогнозування.

Зазначимо, що нове збільшення навчальної вибірки знову не призвело до зростання числа членів моделі (5) у порівнянні з моделлю (4), але включені в модель (5) фактори досить істотно змінилися, і така структура оптимальної моделі потребує аналізу з боку відповідних фахівців.

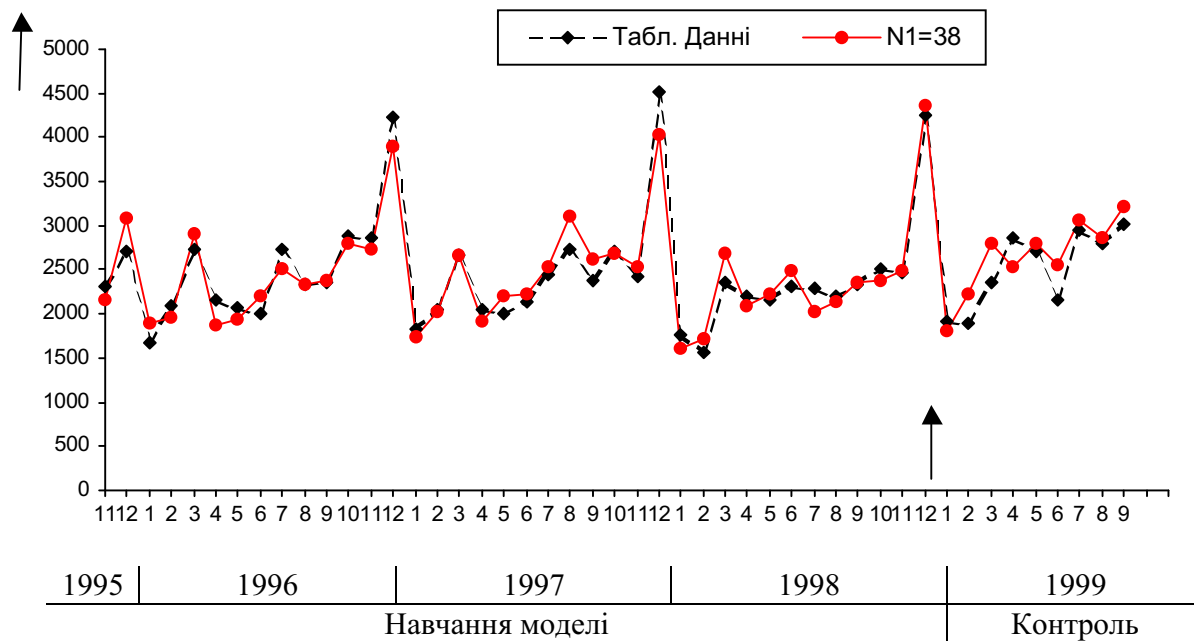


Рис.5. Графіки зміни доходів зведеного бюджету та його оцінок, одержаних за моделлю (5), побудованою за першими 38 (з 47 наявних) точками статистичної вибірки

**Висновок.** Побудовані моделі, з одного боку, свідчать про наявність у статистичних даних про модельовані процеси певних закономірностей, а з другого – про ефективність застосування індуктивного підходу на основі МГУА для автоматичного виявлення цих закономірностей і для розв’язання задач прогнозування в системі моніторингу економічних - зокрема, податкових - процесів.

#### Література

1. Степашко В.С., Коппа Ю.В. Опыт применения системы АСТРИД для моделирования экономических процессов по статистическим данным // Кибернетика и вычислительная техника. -1998. - вып.117. - С.24-31.
2. Бюлетень економічної кон'юнктури України.- Київ: НДІ статистики Мінстату України.- 1997.- випуск N 4 - 12-89с.
3. Бюлетень економічної кон'юнктури України.- Київ: НДІ статистики Мінстату України.- 1998.- випуск N 4(8)- 12-93с.
4. Бюлетень економічної кон'юнктури України.- Київ: НДІ статистики Мінстату України.- 1999.- випуск N 4(12)- 12-91с.