УДК 510.67

Т. Ф. МИХАЙЛОВА (ДІІТ), О. В. ПІСКУНОВА (ДФА)

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ОБСЯГІВ ПОДАТКОВИХ НАДХОДЖЕНЬ З УРАХУВАННЯМ РИЗИКУ

Запропоновано економетричну модель для прогнозування обсягів податкових надходжень та оцінки ризику недовиконання планових показників.

Предложена эконометрическая модель для прогнозирования объемов налоговых поступлений и оценки риска недовыполнения плановых показателей.

An econometric model for forecasting volumes of tax returns and estimating risk of not being completed plan indices have been proposed.

Побудова математичних моделей для часових рядів та їх застосування для прогнозу з урахуванням ризику розглядалась у роботах [1, 3]. Для дослідження нестаціонарних рядів доцільно застосувати ARIMA моделі, що детально розглянуті в [2].

Для того, щоб побудувати математичну модель для прогнозування обсягів податкових надходжень з урахуванням ризику докладно були проаналізовані часові ряди місячних даних щодо обсягів податкових надходжень по Дніпропетровській області у період з січня 2000 року по грудень 2005 року, а саме податку на прибуток підприємств, акцизного збору, плати за землю, податку з доходів фізичних осіб і податку на додану вартість. Дані наведено у фактичних цінах.

Розглянемо прогнозування находжень Y від податку з доходів фізичних осіб (одиниці виміру – млн грн). Для того, щоб уникнути інфляційного викривлення інформації, необхідно перейти до реальних одиниць виміру. Для переходу до цін грудня 2005 року необхідно всі спостереження ряду Y помножити на індекс CPI.

Побудуємо модель для прогнозу податку з доходів фізичних осіб Y у цінах грудня 2005 року Спочатку проаналізуємо динаміку показника Y у період, що розглядався, яку представлено на рис. 1. Аналіз структури часового ряду показує, по-перше, наявність лінійного тренду, по-друге, підвищення надходжень від податку з доходів фізичних осіб у грудні кожного року, по-третє, зміну тенденції показника Y з січня 2004 року, що відповідає зміні законодавчої бази щодо нарахування даного податку.

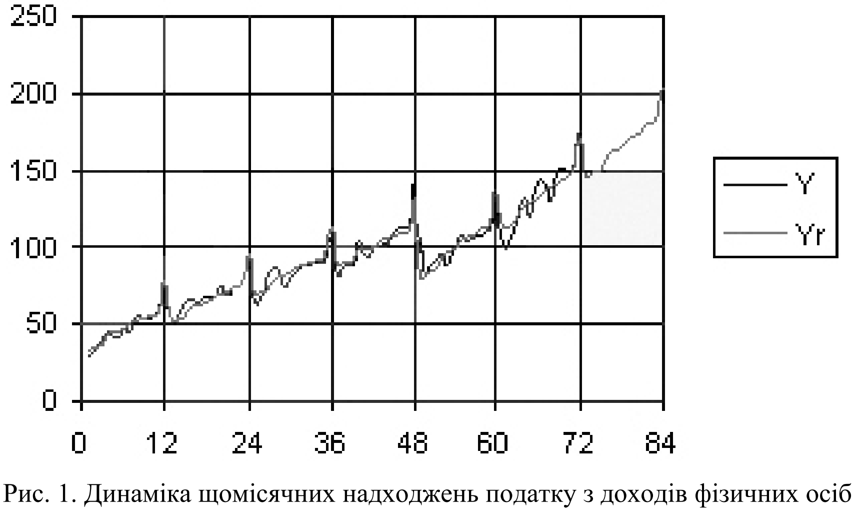

Рис. 1. Динаміка щомісячних надходжень податку з доходів фізичних осіб у порівняльних цінах у період з січня 2000 року по грудень 2005 року

Виходячи з даних спостережень побудуємо модель такого виду

$$y_t = a_0 + a_1 t + a_2 x_{1t} + a_3 t \cdot x_{1t} + a_4 x_{2t} + a_5 x_{3t} + \alpha_6 x_{4t} + \alpha_7 x_{5t} + l_t \quad (1)$$

де t – змінна часу (номер періоду), x_1 – фіктивна змінна, яка приймає значення 0 для періодів часу до 2004 року і значення 1 – з 1 січня 2004 року; x_2, x_3, x_4 – фіктивні змінні, що відображають сезонні коливання в надходженні даного податку та приймають значення 1 відповідно для I, II і III кварталів, і значення 0 для усіх інших періодів часу; x_5 – фіктивна змінна, яка приймає значення 1 для грудня місяця кожного року, та значення 0 для інших періодів.

Отримані за методом найменших квадратів за допомогою функції «ЛИНЕЙН» оцінки параметрів моделі (1) наведені у табл. 1.

Таблиця 1

Оцінки параметрів прогнозної моделі для податку з доходів фізичних осіб у порівняльних цінах

Залежна змінна Y				
незалежні змінні	коефіцієнт	стандартна похибка	t - статистика	P - рівень значимості
константа	52,2	2,17	24,0	0,000
t	1,68	0,068	24,6	0,000
x_1	-33,6	12,0	-2,8	0,006
$t \cdot x_1$	0,010	0,204	0,049	0,961
x_2	6,77	2,14	3,16	0,002
x_3	11,6	2,18	5,31	0,000
x_4	12,8	2,48	5,15	0,000
x_5	21,2	3,19	6,63	0,000

Як бачимо з табл. 1, оцінки майже усіх параметрів моделі (крім параметру при перехресному членові $t \cdot x_1$) можна вважати значимими. Значення коефіцієнта детермінації $R^2 = 0,95$ наближається до 1. Відносна похибка регресії), що відображає прогнозні властивості моделі, складає: $\frac{s}{\bar{y}} 100\,\% = 5,7\ \%$.

Розглянемо залишки моделі L, що побудовано. На основі аналізу графіку і гістограми залишків L (рис. 2), а також результатів перевірки наявності автокореляції в системі за тестом Дарбіна-Уотсона, який дозволив прийняти гіпотезу про відсутність автокореляції ($DW_p = 1,8 > DW_2 = 1,682$ для 1 %-го рівня значимості), можемо зробити припущення про нормальний закон розподілу залишків даної моделі.

За допомогою моделі, що побудовано, отримано прогнозні значення надходжень від податку з доходів фізичних осіб на 2006 рік.

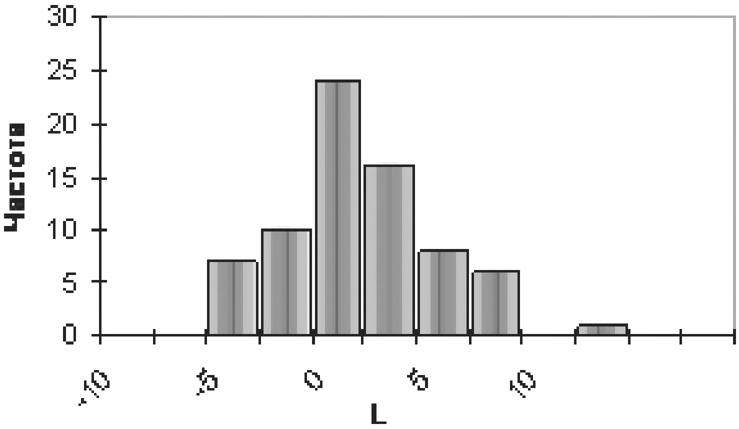

Рис. 2. Гістограма частот залишків L прогнозної моделі податку з доходів фізичних осіб у порівняльних цінах

Визначимо як одну з компонент вектора податкового ризику ймовірність можливого недовиконання планових податкових надходжень [1]. Ризик 10 %-го недовиконання планового показника за прогнозування на один крок дорівнюватиме:

$$P(Y \le 0,9\widehat{Y_p}) = 1 - \Phi\left(\frac{0,1}{V}\right) \tag{2}$$

де V – аналог коефіцієнта варіації:

$$V = \frac{s}{\widehat{Y_p}}; \tag{3}$$

s – похибка прогнозу, отримана на базовому інтервалі; $\widehat{Y_p}$ – прогнозоване значення податкових надходжень; $\Phi(Y)$ – функція Лапласа.

В електронних таблицях Excel функція Лапласа розраховується за допомогою статистичної функції НОРМРАСП(Y , 0, 1, 1).

Так, згідно отриманого прогнозу у січні 2006 року очікуються надходження від податку з доходів фізичних осіб в обсязі 141,6 млн грн, похибка прогнозу $s = 6{,}38$ млн грн, $V = 0{,}045$. Величина ступеня ризику як ймовірність недоотримання податкових надходжень від податку з доходів фізичних осіб по Дніпропетровській області становить:

$$P(Y \le 0{,}9\widehat{Y_p}) = 1 - \Phi\left(\frac{0{,}1}{0{,}045}\right) =$$

$$= 1 - \Phi(2{,}219) = 0{,}013 .$$

Тобто ймовірність 10 %-го недовиконання планового показника податку з фізичних осіб складає 0,013, або 1,3 %.

За збільшенням прогнозованого інтервалу похибка прогнозу також збільшується. На річному інтервалі похибка розраховується за її максимально можливим значенням $s^2(1+2+3+...+12)$.

Таким чином, ми отримали прогноз надходжень до бюджету податку з доходів фізичних осіб у цінах грудня 2005 року. В той же час у бюджетному процесі важливим є прогнозування податкових надходжень у номінальних цінах. Тому розглянемо тепер прогнозування обсягів податку з доходів фізичних осіб у фактичних цінах.Спочатку побудуємо модель регресії показника Y виду (1). Отримані у цьому випадку за методом найменших квадратів з використанням функції «ЛИНЕЙН» оцінки параметрів моделі (1) наведені у табл. 2. Таким чином, побудована така модель:

$$\hat{y}_t = 31{,}9 + 1{,}51t - 88{,}9x_{1t} +$$

$$+1{,}28t \cdot x_{1t} 5{,}30x_{2t} + 7{,}23x_{3t} +$$

$$+8{,}68x_{4t} + 18{,}7x_{5t} . \qquad (4)$$

Як бачимо з табл. 2, усі оцінки параметрів моделі можна вважати значимими на 1 % рівні. Значення коефіцієнта детермінації $R^2 = 0{,}97$ наближається до 1. Відносна похибка регресії складає: $\frac{s}{\overline{y}} 100\,\% = 6{,}5\,\%$.

Аналіз залишків моделі L на основі побудованої гістограми (рис. 3) та дослідження наявності автокореляції за допомогою статистики Дарбіна-Уотсона ($DW_2 = 1{,}284 < DW_p = 1{,}6 < < DW_2 = 1{,}682$ для 1-го % рівня значимості, конкретних висновків зробити не можемо) вже не дозволяє зробити висновок про їх нормальний розподіл. Крім того, слід взяти до уваги, що зазвичай рівень цін є нестаціонарним рядом, а оскільки ціна входить складовою в показники у номінальному виразі, це збільшує порядок інтеграції часового ряду. Отже ймовірно ми маємо справу з нестаціонарним часовим рядом.

Для дослідження часового ряду на стаціонарність необхідно побудувати та проаналізувати автокореляційну та часткову автокореляційну функції, а також застосувати тест Дікі–Фуллера на наявність одиничного кореня.[2, 3]. Виконати такі процедури дозволяють сучасні статистичні пакети програм, зокрема, STATA, E.VIEWS, SPSS, STATISTICA.

Таблиця 2

Оцінки параметрів прогнозної моделі для податку з доходів фізичних осіб у фактичних цінах

Залежна змінна: Y				
незалежні змінні	коефіцієнт	стандартна похибка	t - статистика	P - рівень значимості
константа	31,9	1,98	16,1	0,000
t	1,51	0,062	24,4	0,000
x_1	-88,9	10,9	-8,13	0,000
$t \cdot x_1$	1,28	0,186	6,89	0,000
x_2	5,30	1,95	2,72	0,008
x_3	7,23	1,99	3,64	0,000
x_4	8,68	2,26	3,85	0,000
x_5	18,7	2,91	6,41	0,008

Використовуючи можливості електронних таблиць Excel, змоделюємо залишки за допомогою ARIMA моделі. Ряд залишків позначимо YL.

Будемо вважати, що порядок інтеграції ряду

YL дорівнює 1, оскільки для економічних даних більш високі порядки інтеграції майже не зустрічаються.[2] Створимо ряд перших різниць $D1YL$:

$$d1yl_t = yl_t - yl_{t-1} \qquad (5)$$

та побудуємо модель для $D1YL$ виду (6).

$$d1yl_t = d_0 + \alpha_1 d1yl_{t-1} + \alpha_2 d1yl_{t-2} + \ldots +$$
$$+\alpha_p d1yl_{t-p} + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \ldots - \beta_q u_{t-q} + \varepsilon_t \qquad (6)$$

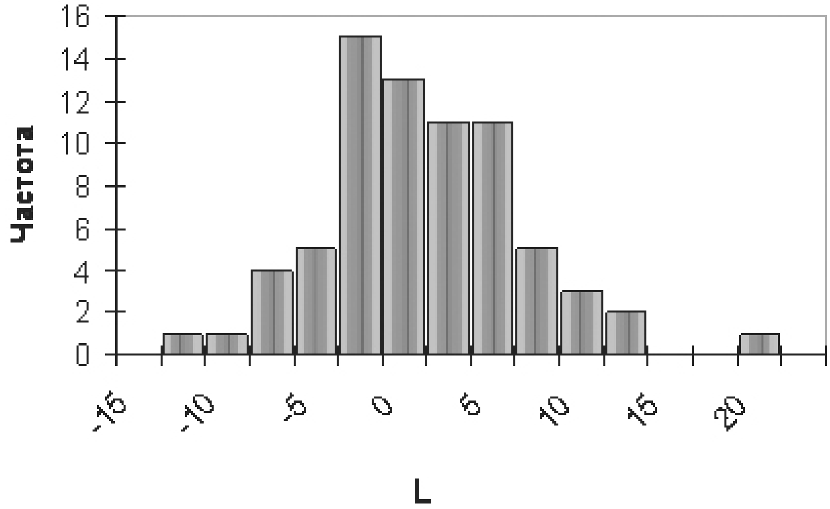

Рис. 3. Гістограма частот залишків L прогнозної моделі податку з доходів фізичних осіб у фактичних цінах

Для цього необхідно обрати специфікацію моделі, тобто визначити порядок p AR-складової моделі та порядок q MA-складової. Зауважимо, що не існує правила для знаходження ідеального порядку p і q. Один з найпростіших та найпоширеніших методів – аналіз корелограм. Його зручно застосовувати, використовуючи сучасні статистичні пакети програм. В електронних таблицях Excel будемо здійснювати вибір моделі на підставі зниження показника ризику, що визначається на основі значення абсолютної похибки регресії s та відносної похибки [1]. Крім того, під час побудови моделі будемо враховувати похибки оцінок коефіцієнтів моделі (6) α_j, β_j, тобто модель вважається адекватною, якщо всі її параметри будуть значимими на 5% рівні значимості. В результаті було обрано модель виду:

$$d1yl_t = \alpha_0 + \alpha_1 d1yl_{t-1} + \alpha_2 d1yl_{t-2} + \varepsilon_t \qquad (7)$$

тобто $p = 2$, $q = 0$. Відмітимо, що аналіз корелограм за допомогою статистичного пакету програм STATA підтвердив обрану нами специфікацію моделі $ARIMA(2,1,0)$ для ряду YL (моделі $ARMA(2,0)$ для ряду $D1YL$).

Для побудови моделі (7) в електронних таблицях Excel було створено часові ряди $D1YL\_1$ і $D1YL\_2$, які отримані з ряду $D1YL$ шляхом зсуву у часі відповідно на 1 і 2 періоди, та за допомогою функції «ЛИНЕЙН» знайдено оцінки параметрів моделі (7):

$$d1\hat{y}l_t = 0{,}141 - 0{,}588 \cdot d1yl_{t-1} - 0{,}441 \cdot a1yl_{t-2} \qquad (8)$$

Оцінки параметрів даної моделі можна вважати значимими на 1 % рівні. Аналіз залишків LL моделі (5.14) на основі побудованої гістограми (рис. 4) та статистики Дарбіна-Уотсона ($DW_p = 2{,}00$) дозволяє зробити припущення про їх нормальний розподіл.

Для отримання прогнозу для ряду $\widehat{YL}$ за допомогою побудованої моделі спочатку необхідно розрахувати за формулою (8) прогнозні значення для ряду перших різниць $\widehat{D1YL}$, а потім від прогнозних значень ряду перших різниць $\widehat{D1YL}$ перейти до прогнозних значень ряду $\widehat{YL}$:

$$\widehat{yl_t} = yl_{t-1} + \widehat{d1yl_t}$$

– для базового періоду;

$$\widehat{yl_t} = \widehat{yl_{t-1}} + \widehat{d1yl_t}$$

– для прогнозного періоду, (9).

Прогнозні значення надходжень від податку з доходів фізичних осіб $\hat{Y}_P$ розраховуються як сума прогнозних значень показника $\hat{Y}$, отриманих за формулою (4), та прогнозних значень ряду $\widehat{YL}$, отриманих з рівняння (9): $\widehat{Y_p} = \hat{Y} + \widehat{YL}$

Використання ARIMA моделей та фіктивних змінних для аналізу нестаціонарних часових рядів дозволило побудувати економетричні моделі для прогнозу обсягів податкових надходжень з фізичних осіб з урахуванням ризику та проаналізувати залишки моделей.

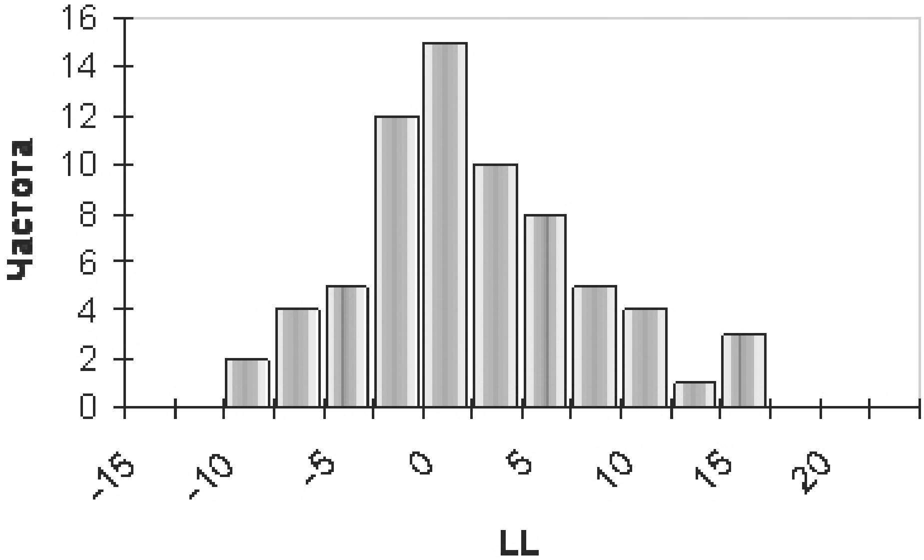

Рис. 4. Гістограма частот залишків *LL* для моделі часового ряду *YL*

БІБЛІОГРАФІЧНИЙ СПИСОК

1. Вітлінський В. В. Моделювання економіки – К.:КНЕУ, 2003. – 408 с.
2. Лукяненко І. Г. Сучасні економетричні методи у фінансах / І. Г. Лукяненко. Ю. О. Городніченко, - К.: Літера ЛТД, 2002. – 352 с.
3. Dickey D. A., Fuller W. A. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root // Econometrica, 1981, № 49. - P. 1057-1072.

Надійшла до редакції 18.09.2007.