

Прогнозування часових рядів урожайності зернових культур

В статті досліджується питання прогнозування урожайності зернових культур та вибору для цього найбільш оптимального методу.

The article examines the problems of forecasting grain yield and choice for the most efficient method.

Ключові слова: прогноз, урожайність, часовий ряд, економіко-математичне моделювання.

Вступ. Ефективне функціонування господарюючого суб'єкта залежить від якості прогнозування динаміки економічних процесів. З метою дослідження економічних або природно-економічних часових рядів необхідне отримання їх типових характеристик, які дозволили б вибрати для прогнозування найбільш адекватний метод, тобто передпрогнозний аналіз.

Постановка завдання. Ефективність економіко-математичного моделювання значно підвищується при використанні спеціальних алгоритмів, реалізованих у комп'ютерних програмах. У процесі побудови моделі з використанням таких алгоритмів передбачається врахування істотних особливостей прогнозованих економічних ситуацій. Зміна чинників, за допомогою регулювання параметрів програм, дозволяє вибрати оптимальний варіант на основі порівняльного аналізу результативних ознак.

Результати. Традиційні статистичні підходи до прогнозування часових рядів базуються на трендах і регресії, при цьому передбачається виконання умови незалежності рівнів (спостережень), що становлять розглянуті часові ряди. Ця умова для економічних часових рядів найчастіше не виконується в силу того, що для них характерний довготривалий прогноз.

У загальному вигляді часовий ряд може бути представлений у вигляді

$$y = f(T, S, C, E),$$

де T - основна тенденція, тренд; S , C - сезонна і циклічна компоненти; E - випадкові варіації.

Лінійний або нелінійний тренд є стійкою зміною рівня показника протягом тривалого часу. Сезонність характеризує стійкі міжкрокові коливання рівня показника, період яких відомий. Циклічна компонента описує циклічні процеси, параметри яких апріорі невідомі. Залежно від взаємозв'язку цих компонент може бути побудована як адитивна, наприклад тренд-сезонна, так і мультиплікативна модель часового ряду, або їх комбінація.

З арсеналу сучасних методів прогнозування економічних часових рядів досить часто затребуваними є:

- 1) поліноміальне або тригонометричне згладжування;
- 2) апроксимація сплайнами;
- 3) використання процесів авторегресії і ковзного середнього;
- 4) використання процесів авторегресійного ковзного середнього (ARIMA).

Процес авторегресії і ковзного середнього складається з лінійної функції від попереднього спостереження плюс незалежний випадковий шум мінус деяка частка попереднього випадкового шуму, тобто ARIMA - це поєднання авторегресії з процесом ковзного середнього. Ці процеси забезпечують високу точність прогнозування таких часових рядів, які є достатньо гладкими з повільними змінами напрямку.

Застосування цих процесів для прогнозування часто і різко мінливих часових рядів, до яких можна віднести і часові ряди динаміки врожайності, приводить до появи досить значних помилок прогнозування.

Відзначимо, що в разі різких викидів в аналізованих часових рядах і відхилення їх розподілу від нормального закону, використання методу найменших квадратів (МНК) стає неефективним [1].

Таким чином, необхідно проведення передпрогнозного аналізу часових рядів врожайності з метою обґрунтованого вибору найбільш адекватного методу прогнозування. Одним з методів отримання відповідної інформації є R/S аналіз або його різновид фрактальний аналіз часових рядів.

Метою фрактального аналізу будь-якого часового ряду є виявлення наявності в ньому довготривалого прогнозу, оцінка його глибини, а також значення / H -показника Херста [2]. Крім того, передбачається виявлення такої характеристики, як стійкість або такої оберненої до неї властивості, як «повернення до середнього частіше, ніж у випадковій поведінці часових рядів» (часте реверсування спад-підйом). Дуже важливим для прогнозування є виявлення періодичних циклів, якщо такі є, або квазіциклів. Для позначення

останніх використовуються терміни «дробова квазіперіодичність або «хаотичні цикли». Знання перерахованих фрактальних характеристик розглянутого часового ряду надає аналітику передпрогнозу інформацію, тобто дозволяє йому оцінити перспективність надійного прогнозування часових рядів за допомогою клітинно-автоматної прогнозувальної моделі.

Термін «R/S-аналіз» часових рядів зустрічається під назвою «метод нормованого розмаху Херста». Метод нормованого розмаху починають з побудови часового ряду n послідовних величин x_1, \dots, x_n . В аналізі, який проводив Херст, кожне x_r , $r = \overline{1, n}$ визначає собою щорічну зміну врожайності зернових культур.

Середнє значення часового ряду x_m визначається як:

$$x_m = \sum_{r=1}^n x_r \cdot$$

Стандартне відхилення S_m розраховується за формулою:

$$S_m = \frac{1}{\sqrt{n}} \sum_{r=1}^n (x_r - x_m)^2 \cdot$$

Центрування даних здійснюється за формулою:

$$Z_r = x_r - x_m$$

із середнім значенням, рівним нулю.

Далі перейдемо до кумулятивних часових рядів:

$$Y_1 = Z_1 + Z_r, \quad r = \overline{2, n}.$$

При цьому останнє значення Y_n завжди буде дорівнювати нулю, так як $\overline{Z_r} = 0$.

Розмах визначається різницею між максимальним і мінімальним значеннями Y_r :

$$R_n = \max(Y_1, \dots, Y_n) - \min(Y_1, \dots, Y_n).$$

Внаслідок того, що Y відрегульованим до середнього значення, рівного нулю, максимальне значення Y завжди буде невід'ємним, а мінімальне значення - недодатним. При цій умові величина R_n буде завжди невід'ємною.

Херст знайшов більш загальну формулу:

$$(R/S)_n = cn^H, \quad (1)$$

де c - константа, H - показник Херста.

Величина R/S називається нормованим розмахом, так як має нульове середнє і виражається у вигляді локального середньоквадратичного відхилення. Нормування розмаху поділом на середньоквадратичне відхилення дозволяє порівнювати різні явища, а також періоди часу, які віддалені один від одного.

Після логарифмування виразу (1) одержуємо вираз

$$\log(R/S)_n = \log c + H \log n, \quad (2)$$

який може бути використано для побудови графіка $\log(R/S)_n$, в залежності від $\log n$ і знаходження нахилу за допомогою регресії. Кутовий коефіцієнт у виразі (2) дає оцінку показника Херста.

Однією з основних фрактальних характеристик часових рядів є колір шуму [2], який відповідає цьому ряду на тому чи іншому часовому відрізку.

Значення $H > 2/3$ визначає чорний колір шуму. Чим більше значення $H \in [2/3, 1]$ тим більша трендостійкість притаманна відповідному відрізку часового ряду.

Значення H в межах $\approx 0,5 \pm 0,1$ визначає область білого шуму, який відповідає «хаотичній поведінці часового ряду» і, отже, найменшій надійності прогнозу.

Значення H в межах $\approx 0,3 \pm 0,1$ визначає відрізок часового ряду в області рожевого шуму. Рожевий шум говорить про притаманність розглянутому відрізку часового ряду властивості антиперсистентності у випадку, коли часовий ряд реверсує частіше, ніж ряд випадковий (часте повернення до середнього).

Автокореляція рівнів часового ряду є кореляційною залежністю між послідовними рівнями часового ряду, графік якої утворює автокореляційну функцію.

Кількісно автокореляцію можна виміряти за допомогою лінійного коефіцієнта кореляції між рівнями вихідного часового ряду і рівнями цього ряду, зсунутими на кілька кроків у часі. Наприклад, коефіцієнт кореляції між рядами y_t і y_{t-1} вимірює тісноту зв'язку між урожайністю зернових культур поточного і попереднього років.

Одна з формул для розрахунку лінійного коефіцієнта кореляції має вигляд:

$$r_{xy} = \frac{\sum (x_j - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_j - \bar{x})^2 \sum (y_j - \bar{y})^2}}.$$

Якщо під першою змінною ми розглянемо ряд y_2, y_3, \dots, y_n , а під другою ряд y_1, y_2, \dots, y_{n-1} , то наведена вище формула прийме вигляд:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1) \sum_{t=2}^n (y_{t-1} - \bar{y}_2)}}$$

$$\text{де } \bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}, \bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1}.$$

Ця величина називається коефіцієнтом автокореляції рівнів ряду першого порядку, так як він вимірює залежність між сусідніми рівнями ряду при лагу 1. Число періодів, за якими розраховується коефіцієнт автокореляції, називають лагом. Зі збільшенням лагу число пар значень, за якими розраховується коефіцієнт автокореляції, зменшується. Максимальний лаг не повинен перевищувати величини $n/4$.

Аналогічно визначаються коефіцієнти автокореляції другого і більш високих порядків. Так, коефіцієнт автокореляції другого порядку характеризує тісноту зв'язку між рівнями y_t і y_{t-2} і визначається за формулою:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3) \sum_{t=3}^n (y_{t-2} - \bar{y}_4)}}$$

$$\text{де } \bar{y}_3 = \frac{\sum_{t=3}^n y_t}{n-2}, \bar{y}_4 = \frac{\sum_{t=3}^n y_{t-2}}{n-2}.$$

Коефіцієнт автокореляції характеризує тісноту тільки лінійного зв'язку, тому за коефіцієнтом автокореляції можна судити про наявність лінійної (або близької до лінійної) тенденції. Для деяких часових рядів, що мають сильну нелінійну тенденцію (наприклад, параболу або експоненту), коефіцієнт автокореляції рівнів вихідного ряду може наближатися до нуля.

Послідовність коефіцієнтів автокореляції рівнів першого, другого і вищих порядків утворюють автокореляційну функцію часового ряду. Графік залежності

її значень від величини лагу (порядку коефіцієнта автокореляції) називається корелограмою.

Аналіз автокореляційної функції за формою піків корелограми дозволяє виявити структуру часового ряду. При цьому визначають лаг, при якому автокореляція приймає найбільші значення.

Аналіз структури можна проводити за таким алгоритмом:

- якщо найбільш високим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить тільки тенденцію;
- якщо найбільш високим виявився коефіцієнт автокореляції порядку τ , ряд містить циклічні коливання з періодичністю в τ моментів часу;
- якщо жоден з коефіцієнтів автокореляції не є значущим, можна зробити одне з припущень щодо структури ряду: 1) ряд не містить тенденції і циклічних коливань, а включає тільки випадкову компоненту; 2) ряд містить сильну нелінійну тенденцію.

При практичному аналізі часових рядів врожайності зернових культур найбільш часто зустрічається поєднання згаданих особливостей рівнів коефіцієнтів автокореляції. Отже, необхідний детальний аналіз співвідношень величин піків, а можливо, і використання альтернативних підходів, наприклад методу фазових портретів, гармонійного аналізу тощо.

Висновки. Спільним для розглянутих способів практичного прогнозування врожайності є висока трудомісткість, недостатні точність і завчасність прогнозів. У зв'язку з цим, обґрунтування, адаптація і модифікація методів прогнозування врожайності, зокрема в умовах посушливого клімату, є актуальною задачею, затребуваною практикою сільськогосподарського виробництва.

У зв'язку зі специфічними особливостями часових рядів врожайності, можна вважати вельми перспективним застосування методів нелінійної динаміки для моделювання і прогнозування їх поведінки, а також для оцінки ризиків, обумовлених недосконалістю моделей і методичних підходів. Синергетичні підходи до прогнозування врожайності, в тому числі методами нелінійної економічної динаміки, розроблялися І.Г. Вінтізенко, В.А. Долятовським, В.І. Лебедевим, Р.М. Нижегородцевим, В.А. Перепелицею, Є.В. Поповою, А.І. Пригожиним, Т.Андерсені, Г. Дженкінсом, В.Б. Зангом, Е. Петерсом, Д. Фішером та іншими вітчизняними і зарубіжними вченими.

На наш погляд, найбільш перспективним підходом можна вважати дворівневий, коли на першому етапі виявляються такі прогнозні властивості, як

трендостійкість, персистентність (антиперсистентність), внутрішня структура, довготривала пам'ять та ін. Крім того, попереднє, передпрогнозне оцінювання циклічності, що реалізується не тільки автокореляційним аналізом, але й з використанням «істинних» фазових портретів, формованих на основі сплайн-апроксимації, істотно підвищує якість моделювання, роблячи його більш наочним і осмисленим.

Методику прогнозування оцінки врожайності на основі апарату нечітких множин і клітинних автоматів можна коротко сформулювати таким чином. Прогнозування врожайності сільськогосподарських культур, зокрема озимої пшениці, оброблюваної в умовах посушливого клімату, здійснюють з використанням багаторічних щорічних спостережень за фактичними значеннями врожайності. Застосовують перетворення фактичних значень врожайності до якісних величин, причому фактичні значення врожайності перетворюють в ряд лінгвістичних значень «Висока врожайність», «Середня врожайність», «Низька врожайність». Виділяють фрагменти передпрогнозного періоду перетвореного ряду різної тривалості (1,2 і більше років), а прогнозовану врожайність на наступний рік знаходять з урахуванням повторюваності фрагментів передпрогнозного періоду в перетвореному ряді значень врожайності за всі роки спостережень. Недоліком цього досить ефективного методу є необхідність попереднього навчання («розфарбовування») часового ряду з використанням експертного підходу.

Література

1. Грицюк П.М. Аналіз моделювання та прогнозування динаміки врожайності озимої пшениці в розрізі областей України. – Рівне: 2010 р. – 350 с.
2. Петерс Э. Фрактальный анализ финансовых рынков. – М: Интернет-трейдинг, 2004. — 304 с.