Тішков Максим Олегович
студент
Національний технічний університет України «Київський політехнічний інститут»
Тишков Максим Олегович
студент
Национальный технический университет Украины «Киевский политехнический институт»
Tishkov M.
Student
National technical university of Ukraine «Kyiv Polytechnic Institute»

МЕТОДИ І МОДЕЛІ ПРОГНОЗУВАННЯ СТРАХОВИХ ВИПАДКІВ НА ВИРОБНИЦТВІ

Анотація. *Дана робота присвячена дослідженню проблеми прогнозування фінансових витрат на виплати страхових відшкодувань потерпілим. В роботі розглядаються методи та моделі прогнозування. Пропонується будувати прогнози на основі авторегресійних моделей часових рядів. В роботі детально розглянуті моделі авторегресії, авторегресії з ковзним середнім та авторегресії з трендом. В результаті побудована інформаційно-аналітична система для моделювання та прогнозування фінансових процесів.*
Ключові слова: *часові ряди, авторегресія, тренд, прогноз, моделювання, фінансовий процес, ковзне середнє.*

Аннотация. *Данная работа посвящена исследованию проблемы прогнозирования финансовых затрат на выплаты страховых возмещений пострадавшим. В работе рассматриваются методы и модели прогнозирования. Предлагается строить прогнозы на основе авторегрессионных моделей временных рядов. В работе подробно рассмотрены модели авторегрессии, авторегрессии со скользящим средним и авторегрессии с трендом. В результате построена информационно-аналитическая система для моделирования и прогнозирования финансовых процессов.*
Ключевые слова: *временные ряды, авторегрессия, тренд, прогноз, моделирование, финансовый процесс, скользящее среднее.*

Abstract. *The paper is devoted to the problem of forecasting financial expenses for insurance claims. We consider the forecasting methods and models. The algorithm prediction based on autoregressive time series. The paper considered by the autoregression model, autoregression model with moving average and autoregression model with trend. In consequence of research, information-analytic system for forecasting financial processes was built.*
Keywords: *time series, autoregression, trend, forecast, modelling, financial process, moving average.*

Вступ

Соціальне страхування від нещасних випадків на виробництві проводить фонд соціального страхування. Щорічно в бюджет фонду виділяються великі суми грошей і, очевидно, було б добре передбачити та спрогнозувати об'єм необхідного фінансування даної організації.

У зв'язку з напруженою ситуацією на сході України та анексією Кримського півострова гостро постала проблема переселенців зі сходу та півдня України, що теж потребують соціального захисту та фінансової допомоги. Цей фактор призводять до необхідності зміни політики фінансування фонду. Тому зараз доцільність прогнозування нещасних випадків є незаперечною. Нещасні випадки на виробництві завжди мали своє місце. Зараз же у зв'язку з важким економічним становищем країни обладнання багатьох установ стає дедалі більш морально застарілим, що призводить до збільшення кількості нещасних випадків. За мету даної роботи було поставлено створення інформаційно-аналітичної системи, яку можна було б застосовувати для прогнозування та планування фінансування «Фонду соціального страхування».

Методи і моделі часових рядів

Часовий ряд — модель випадкового процесу, набір послідовних результатів спостережень, вимірів. Для прогнозування за часовим рядом необхідно побудувати його математичну модель. В даній роботі розглядаються наступні моделі часових рядів: модель авторегресії (АР), авторегресії з ковзним середнім (АРКС), авторегресії з трендом (АР + Тренд).

Модель АР(p) з ковзним середнім (1) має наступним вигляд:

$$y(k)=a_0+\sum_{i=1}^{p}a_i y(k-i). \quad (1)$$

Дана модель відображує залежність стану процесу в поточний момент часу від p значень в минулому [1].

Модель АРКС(p, q) з ковзним середнім (2) має наступним вигляд:

$$y(k)=a_0+\sum_{i=1}^{p}a_i y(k-i)+\sum_{j=1}^{q}b_j\varepsilon(k-j)+\varepsilon(k), \quad (2)$$

де ε — випадковий, нормально розподілений вектор.

Модель АРКС(p, q), так як і модель АР(p) відображує залежність від значень часового ряду в минулому і відрізняється від моделі АР(p) тим, що містить випадкову змінну, що відображує випадкові коливання значень процесу, що можливі в реальних умовах [1].

Модель АР + Тренд (p, q) (3) має наступним вигляд:

$$y(k)=a_0+\sum_{i=1}^{p}a_i y(k-i)+\sum_{j=1}^{q}b_j t^j(k)+\varepsilon(k). \quad (3)$$

Дана модель відрізняється від АР додаванням поліноміального тренду порядку q. Дана модель використовується для процесів з трендом, тобто процесів, що візуально можна оцінити як монотонно зростаючі, чи спадні [4].

Для визначення порядку авторегресійної частини використовуємо автокореляційну функцію (АКФ) та більш точну характеристику вибірки — часткову автокореляційну функцію — (ЧАКФ), що відображують статистичну залежність між випадковими величинами одного ряду, але взятих зі зміщенням. В результаті обрахунку АКФ та ЧАКФ ми отримуємо ряд значень і за кількістю значень, що є більшими за обраний нами рівень значимості, можемо визначити достатній порядок моделі.

Побудувати математичну модель означає знайти усі коефіцієнти рівняння. Це ми можемо зробити використовуючи метод найменших квадратів (МНК) (4) та рекурентний метод найменших квадратів (РМНК) (5).

$$\Theta=\left(X^T X\right)^{-1}XY, \quad (4)$$

де

$X=(1,y(k-1),y(k-2),\ldots,y(k-p),v(k),v(k-1,\ldots,v(k-q)$ — матриця вимірів;

$\Theta=\left(a_0,a_1,\ldots,a_p,1,b_1,b_2,\ldots b_q\right)$ для моделі АРКС(p, q);

$$\Theta(k)=\Theta(k-1)+\gamma(k)\left[y(k)-\Theta^T(k-1)\psi(k)\right];$$
$$\gamma(k)=\frac{P(k-1)\psi(k)}{\alpha_k^{-1}+\psi^T(k)P(k-1)\psi(k)}; \quad (5)$$
$$P(k)=P(k-1)-\frac{P(k-1)\psi(k)\psi^T(k)P(k-1)}{\alpha_k^{-1}+\psi^T(k)P(k-1)\psi(k)},$$

де

$\psi^T(k)=[-y(k-1),-y(k-2),\ldots,-y(k-p),v(k-1),$
$v(k-2),\ldots,v(k-q)$ — матриця вимірів;

$\Theta=\left(a_0,a_1,\ldots,a_p,1,b_1,b_2,\ldots b_q\right)$ для моделі АРКС(p, q);

α_k — вагові коефіцієнти [2].

Оцінювання якості прогнозу

Важливим моментом процесу прогнозування є об'єктивне визначення якості отриманого прогнозу. Оскільки прогнозовані значення — випадкові величини, то для оцінювання їх якості необхідно використовувати декілька статистичних критеріїв.

Коефіцієнт детермінації (6) — статистичний показник, що вказує наскільки варіація залежної змінної відрізняється від варіації незалежних змінних. Вказує наскільки точно побудована модель підтверджує спостереження. Оптимальне значення — 1.

$$R^2=\frac{\mathrm{var}(\hat{y})}{\mathrm{var}(y)}. \quad (6)$$

Сума квадратів похибок (7) — сума квадратів різниць значень знайдених за побудованою моделлю від значень спостережень за якими модель була побудована. Оптимальне значення — якомога менше.

$$\Sigma e^2=\sum_{k=1}^{N}\left(\hat{y}(k)-y(k)\right)^2. \quad (7)$$

Статистика Дарбіна-Уотсона (8) — вказує на кореляцію між похибками моделі. Чим менша кореляція — тим ближче значення статистики до 2.

$$DW=\frac{\sum_{k=2}^{N}\left(e(k)-e(k-1)\right)^2}{\sum_{k=1}^{N}e^2(k)}. \quad (8)$$

Коефіцієнт Тейла (9) — вказує на якість моделі та прогнозу. Набуває значень від 0 до 1. Оптимальне значення — 0.

$$U=\frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y(k)-\hat{y}(k)\right)^2}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}y^2(k)}+\sqrt{\frac{1}{N}\sum_{k=1}^{N}\hat{y}^2(k)}}. \quad (9)$$

Середньо-квадратична похибка (10) — середня похибка знайдених за моделлю значень.

$$\text{СКП} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y(k)-\hat{y}(k)\right)^2}\,. \qquad (10)$$

Середня абсолютна похибка у відсотках (11) — середнє абсолютних значень похибок оцінок прогнозу у відсотках відносно фактичного значення показника [3]. Інтерпретація значень наведена в таблиці 1.

$$\text{САПП} = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|y(k)-\hat{y}(k)\right|}{y(k)}\cdot 100\%\,. \qquad (11)$$

Таблиця 1

Інтерпретація типових значень критерію САПП

САПП,%	Інтерпретація
< 10	Висока точність
10–20	Хороша точність
20–50	Задовільна точність
> 50	Незадовільна (неприйнятна) точність

Обробка вибірок розробленим програмним продуктом

Вибірка даних — виплати потерпілим з 01.2002 по 06.2016 помісячно. Графік вибірки зображений на Рис. 1.

Рисунок 1. Графік вибірки

Як бачимо з графіка, вибірка містить тренд, тобто значення вибірки з часом збільшуються, тобто можна припустити, що оптимальною моделлю буде модель з трендом.

За допомогою ЧАКФ оцінимо необхідний максимальний порядок моделей — Рис. 2.

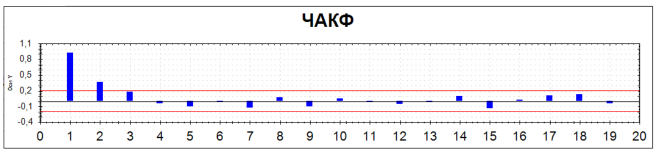
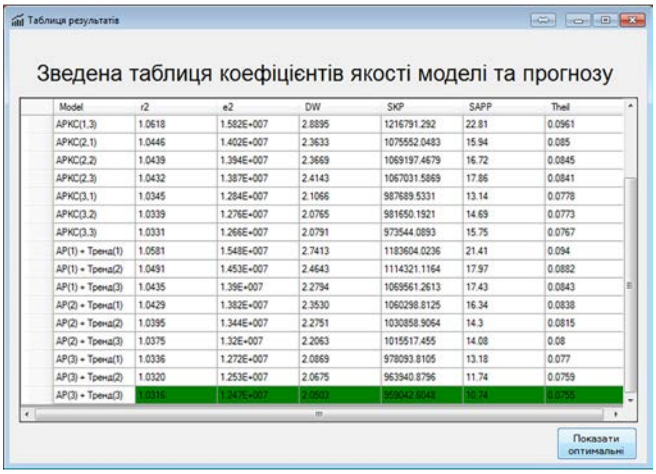


Рисунок 2. ЧАКФ — 20 лагів

При рівні значущості 0,2 можемо зробити висновок про доцільність побудови моделей до третього порядку.



Таблиця результатів

Зведена таблиця коефіцієнтів якості моделі та прогнозу

Model	r2	e2	DW	SKP	SAPP	Theil
АРКС(1,3)	1.0618	1.582E+007	2.8895	1216791.292	22.81	0.0961
АРКС(2,1)	1.0446	1.402E+007	2.3633	1075552.0483	15.94	0.085
АРКС(2,2)	1.0439	1.394E+007	2.3669	1069197.4679	16.72	0.0845
АРКС(2,3)	1.0432	1.387E+007	2.4143	1067031.5869	17.86	0.0841
АРКС(3,1)	1.0345	1.284E+007	2.1066	987689.5331	13.14	0.0778
АРКС(3,2)	1.0339	1.276E+007	2.0765	981650.1921	14.69	0.0773
АРКС(3,3)	1.0331	1.266E+007	2.0791	973544.0893	15.75	0.0767
АР(1) + Тренд(1)	1.0581	1.548E+007	2.7413	1183604.0236	21.41	0.094
АР(1) + Тренд(2)	1.0491	1.453E+007	2.4643	1114321.1164	17.97	0.0882
АР(1) + Тренд(3)	1.0435	1.39E+007	2.2794	1069561.2613	17.43	0.0843
АР(2) + Тренд(1)	1.0429	1.382E+007	2.3530	1060298.8125	16.34	0.0838
АР(2) + Тренд(2)	1.0395	1.344E+007	2.2751	1030858.9064	14.3	0.0815
АР(2) + Тренд(3)	1.0375	1.32E+007	2.2063	1015517.455	14.08	0.08
АР(3) + Тренд(1)	1.0336	1.272E+007	2.0869	978093.8105	13.18	0.077
АР(3) + Тренд(2)	1.0320	1.253E+007	2.0675	963940.8796	11.74	0.0759
АР(3) + Тренд(3)	1.0316	1.247E+007	2.0503	959042.6048	10.74	0.0755

Показати оптимальні

Рисунок 3. Таблиця результатів

Особливістю розробленої системи є те, що за досить короткий час вона здатна побудувати велику кількість моделей, які дослідник може порівняти за допомогою спеціального вікна, в якому зручно зображені результати оцінювання якості побудованих моделей. До того ж, на основі обрахованих коефіцієнтів якості програмний продукт пропонує користувачеві використовувати модель, що показала оптимальні результати. Так, на Рис. 3 ми можемо побачити вікно про яке говорилося вище.

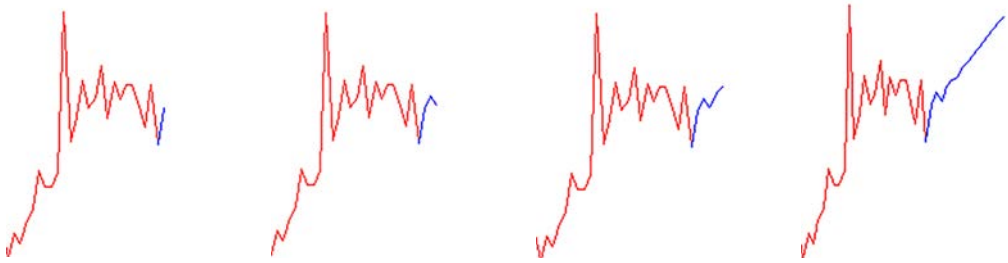
Зеленим підсвічені комірки, що містять найкращі результати по стовпчиках. В даному випадку, найкращою є модель АР + Тренд (3,3) — (12).

$$\begin{aligned} y(k) &= -32008{,}85 + 0{,}2168\cdot y(k-1) + \\ &+ 0{,}2139\cdot y(k-2) + 0{,}3439\cdot y(k-3) + \\ &+ 13462{,}46\cdot k - 184{,}22\cdot k^2 + 1.4426\cdot k^3\,. \end{aligned} \qquad (12)$$

На Рис. 4 зображені графіки прогнозів на 1, 3, 6 та 12 періодів.

Числові дані не будуть інформативними в даному випадку, тому їх не наводимо. Щодо результатів прогнозування: усі прогнози досить вдало відображують реальну поведінку процесу. Єдине, що можна виділити — прогноз на 12 місяців. Такий прогноз уже вважається довгостроковим. Методи часових рядів не можуть гарантувати такої ж високої точності прогнозування для довгострокових прогнозів, як і для коротко- чи середньострокових прогнозів. Тому, через певний час доцільно буде перерахувати отриманий прогноз, уже маючи дані за наступні кілька місяців, для покращення правдивості прогнозу.

Загалом, опираючись на показник САПП, який дорівнює 10,74 для цієї моделі, що є показником хорошої, і вже майже високої точності прогнозу, можемо стверджувати, що прогноз є вдалим. Як бачимо, згідно з прогнозом, протягом наступних шести місяців



а) б) в) г)

Рисунок 4. Графіки прогнозів. Модель АР + Тренд(3,3)

витрати Фонду соціального страхування на виплати страхових відшкодувань будуть триматися приблизно на тому ж рівні, що і протягом усього 2015 року. Тому, при плануванні бюджету на пів року його можна не збільшувати, а от якщо дивитися на рік вперед, то виплати рівномірно йдуть догори. Але, як вже було сказано вище, прогноз на 12 періодів може бути достатньо неточним, тому його необхідно переглянути через 3–6 місяців.

Висновки. На основі проведеного дослідження та отриманих результатів прогнозування можемо зробити наступні висновки: по-перше виконувати дослідження та прогнозувати використовуючи авторегресійні методи часових рядів дуже зручно та ефективно, адже досліднику немає необхідності проводити додатковий аналіз предметної області для пошуку конкретних зовнішніх чинників, що призводять до змін в предметній області; по-друге, на даний момент для прогнозування бюджету Фонду соціального страхування найбільш вдалими моделями авторегресійних часових рядів виявилися моделі з трендом, тому саме їх необхідно використовувати для прогнозування фінансової діяльності Фонду соціального страхування; по-третє, з часом динаміка виплат може змінитися, і вже з графіка розглянутої вибірки можна побачити, що починаючи з середини 2014 року у зв'язку з початком конфлікту на сході та півдні України об'єми виплат різко збільшилися, а от з першого кварталу 2015 року і до сьогодні ніяких різких стрибків у вибірці ми не спостерігаємо. Це говорить про те, що з часом моделі з трендом можуть стати не такими ефективними, як, наприклад, моделі авторегресії з ковзним середнім, тому необхідно періодично перераховувати моделі та критерії якості для визначення оптимальної, що дозволяє зробити розроблена система.

Література

1. Бідюк П.І Аналіз часових рядів: навч. посіб. / П. І. Бідюк, В. Д. Романенко, О. Л. Тимощук. — К.: НТУУ «КПІ», 2013. — 600 с. — 400 пр. — ISBN 978-966-622-588-0.

2. Бокс Дж. Анализ временных рядов: монографія / Дж. Бокс, Г. Дженкинс. — М.: Мир, 1974. — 406 с. — ISBN 0-13-060774-6.

3. Лук'яненко І. Г. Сучасні економетричні методи у фінансах: навч. посіб. / Лук'яненко І. Г., Н. Т. Кузовков. — М.: Машиностроение, 1976. — 184 с.

4. Бідюк П. І. Часові ряди: моделювання і прогнозування: монографія / П. І. Бідюк, О. І. Савенков, І. В. Баклан. — К.: ЕКМО, 2003. — 144 с.