МОДЕЛИРАНЕ И АНАЛИЗ НА ВРЕМЕВИ РЕДОВЕ

MODELING AND ANALYSIS OF TIME SERIES

Petar Tomov, Vladimir Monov
Institute of Information and Communication Technologies - Bulgarian Academy of Sciences
akad. Georgi Bonchev Str., block. 2, office 514, 1113 Sofia, Bulgaria
+ 35929793237, p.tomov@iit.bas.bg

Abstract
A time series is a sequence of numerical data points in successive order. Most commonly, it is a sequence, taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. A time series tracks the movement of the chosen data points, such as a security's price, over a specified period of time with data points recorded at regular intervals. The analysis of a time series is used to see how a given asset, security or economic variable changes over time and examine how the changes, associated with the chosen data point, compare to shifts in other variables over the same time period. There are different basic models, involved in modeling of a time series, such as additive model, multiplicative models, autoregressive model, moving-average model, ARIMA (AutoRegressive Integrated Moving Average) and others.

Keywords: time series; sequence; analysis; model; period.

ВЪВЕДЕНИЕ

Съществуват голям брой величини, които се променят с течение на времето. В теорията на вероятностите такъв тип величини се наричат случайни процеси или случайни функции. Най-често случайният процес е непрекъснат във времето и променя стойностите си в определен интервал от време. Промените обикновено зависят от други величини (фактори), по един или друг начин – с трайна тенденция, периодично, циклично, сезонно и т.н. Анализът на такива случайни процеси с помощта на статистиката обикновено се извършва след като те се дискретизират и представят като серия от данни. Дискретизирането е операция по регистрирането или разпределянето на стойностите на случайния процес през определен период от време. Периодът е константна величина, определя се по определени правила и се нарича време на дискретизация.

Дискретизираните случайни процеси се наричат времеви редове.

Времевият ред е множество от статистически наблюдения над някаква случайна величина, подредени в хронологичен ред [1].

В почти всички области на човешката дейност се наблюдават явления, които могат да се характеризират като времеви редове. Например, ежедневните средни стойности на индексите NASDAQ и Dow Jones за период от един месец, три месеца, една година, ежедневният брой на продажби в голям супермаркет, продажбите на компютри от дадена фирма за всеки месец, цената на отделните валути варира във всеки един момент от времето, под влиянието на различни фактори на пазара [2] и т.н.

Класическият анализ на времевите редове е метод, при който изследваният ред се представя като съставен от отделните компоненти, представящи различните ефекти от групиране и влияние на фактори [3].

ИЗЛОЖЕНИЕ

Понятието времеви ред може да бъде дефинирано като последователност от данни, отнасящи се за определен финансов или икономически обект [4], които описват неговото състояние през N различни времеви контура (N момента или N периода, $N > 1$) [5]. Тези данни са представени като

y_{t_i}, където индексът t_i показва времевия контур. Един времеви ред може да бъде представен като:

$$y(t) = \{y_{t_1}, y_{t_2}, \dots, y_{t_N}\} \qquad (1)$$

В зависимост от това дали стойностите на променливата y се отнасят до моменти, или до периоди, се обособяват моментни и хронологични редове. В случаите, когато се работи с моментни наблюдения, те могат предварително да бъдат преобразувани до хронологичен ред.

Основни компоненти на времевите редове са:

- Дългосрочен тренд – T;
- Цикличен флуктуации (ефекти) – C;
- Сезонни вариации (ефекти) – S;
- Случайни изменения – I.

Трендът е такава част от реда, която описва тенденцията на наклона (нарастваща или намаляваща), по която се разполагат стойностите на времевия ред в достатъчно дълъг период от време. Трендовете могат да се представят с различни математически функции - линейни, полиномни, експоненциални, логаритмични, степенно-показателни и др.

Цикличният ефект е тази част от реда, която представлява периодично нарастване и намаляване на стойностите му, наблюдавани в един и същ времеви период на повторение. Те се описват най-често с помощта на елементарни тригонометрични функции.

Сезонните ефекти също са периодични колебания на стойностите на времевите редове, свързани с общоприетите сезони.

Случайни изменения са изменения в стойностите на времевите редове, които нямат конкретно обяснение. Те се дължат на случайни въздействия върху обекта на изследването и не подлежат на друго описание освен вероятностно.

Трендовете и цикличните ефекти образуват гладки изменения във времевите редове. Затова те често се разглеждат като тяхна обща съставка. Един времеви ред може да съдържа няколко тренда, няколко циклични и няколко сезонни изменения.

Цикличните ефекти или колебания са колебания около (над и под) трендовата линия, които се повтарят през близки по дължина интервали в рамките на продължителен период. Циклични колебания често се представят с вълнообразна крива. В икономиката са характерни така наречените „бизнес – цикли”, които изразяват цикличните закономерности в протичането на стопанската дейност.

В статистиката има голям брой разнообразни методи за изследване и моделиране на времеви редове. Построяването на математически модел на реда означава да се намери формула (формули), която достатъчно точно възпроизвежда поведението в промяната на реда. Тя се използва да се предсказват стойностите на реда в известен или неизвестен период от време, като се изчислява стойността с помощта на моделната формула.

Класическите модели са параметрични. При тях времевият ред се описва с една приближаваща функция, зависеща от времето, в която се определят известен брой коефициенти (параметри) на модела.

Различават се следните основни видове класически модели: адитивен, мултипликативен и смесен.

Адитивен модел

$$Y = T + C + S + I \qquad (2)$$

Този модел е сума на компонентите на времевия ред, описани по-горе. Той е сравнително най-лесен за анализиране, лесно се разлага на отделните си компоненти и поради това обикновено се търси възможност за преобразуване и на други модели към този тип.

Мултипликативен модел

$$Y = T\,.\,C\,.\,S\,.\,I \qquad (3)$$

Моделът е съставен като произведение на компонентите. Когато по някаква причина някой от компонентите отсъства от времевия ред, в произведението той се счита за равен на единица. Ако всички компоненти са положителни, след логаритмуване той се свежда към адитивен:

или приема вида

$$Y1=T1+C1+S1+I1 \qquad (5)$$

Смесен модел

Обикновено в смесения модел само случайните изменения са адитивна компонента. Изброените три вида модели не са единствени.

Всеки времеви ред е резултат от изследването на някакъв обект във времето. Моделът на времевия ред е резултат от изследване на поведението на обекта. Той по същество е описание на динамични процеси, протичащи в обекта.

Класическите модели не са единствените, които се използват за описание на времевите редове. В съвременната литература са описани множество други модели. Сравнително добро описание на времевите редове може да се направи с така наречените модели на авторегресия, пълзящо средно и авторегресия-пълзящо средно (ARMA). Те се построяват чрез добре известния метод на най-малките квадрати, като се отчитат специфичните особености за всеки модел.

Модел на авторегресия

Моделът може да се представи по следния начин:

В този модел стойността yi се представя като функция на n предхождащи я стойности xi-j (j=1, 2, … , n). Той може да се използва в изследването на различни иконометрични величини, като борсови индекси например.

Модел на пълзящо средно

$$y_i = b(0).\,x_i + b(1).\,x_{i-1} + b(2).\,x_{i-2} + \cdots + b(m).\,x_{i-m} \qquad (8)$$

В този модел стойността yi се представя като функция на m предхождащи стойности xi-j (j=1, 2, … , m) на величина, която може да бъде пореден номер на период на наблюдение или друга величина, от която зависи y.

Модел на авторегресия – пълзящо средно

Моделът на авторегресия – пълзящо средно се записва във вида:

$$y_i = a(1).\,y_{i-1} + \cdots + a(n).\,y_{i-n} + b(0).\,x_{i-1} + \cdots + b(m).\,x_{i-m} \qquad (9)$$

Величините n и m се наричат порядъци на модела. Той позволява да се опишат сравнително сложни времеви редове чрез прилагане на метод на най-малките квадрати. Получават се матрици, но в случая регресори са yi-1 , … , yi-n , xi , yi-p , … , yi-m

Те образуват матрицата за всички наблюдения. След прилагане на метод на най-малките квадрати, моделът има вида:

Времевите редове се делят на стационарни и нестационарни.

Стационарните редове се основават на предположението, че процесите които те описват, остават в равновесие относно едно постоянно средно ниво. Графиката на един такъв процес представлява начупена линия, която описва флуктоациите (случайни отклонения от постоянната стойност) около една постоянна средна величина.

При дефинирането на автоковариацията и автокорелацията се използват индекси за средните величини (μ_t) и стандартното отклонение (σ_t). Това показва, че съществува възможност и двете величини да се променят във времето, а да не бъдат константи.

Често в индустрията и икономиката, където прогнозирането има особено важно значение, много времеви редове са нестационарни, тъй като при тях няма постоянна средна величина. При изследване на такива редове се наблюдава периодично изменение на средните стойности, постоянна тенденция на нарастване или на намаляване на средната стойност или други.

Стационарен процес е такъв тип развитие на една величина, при който вероятността стойностите да попаднат в определен интервал не се променя с времето. Записано формално, това означава, че ако функцията на кумулативната вероятност е F, то трябва да бъде изпълнено:

$$F\left(y_{t_1}, y_{t_2}, \dots, y_{t_k}\right) = F\left(y_{t_{1+i}}, y_{t_{2+i}}, \dots, y_{t_{k+i}},\right)$$
(11)

за $\forall$ k и i, имащи смисъл за реда.

Ако горното равенство е изпълнено наистина за всяка стойност на k и i, то това ще доведе до две важни следствия за анализа на такъв тип процеси:

- Функцията на кумулативната вероятност няма да зависи от времето, следователно, ако веднъж се установи нейния вид, може да се прилага за целия времеви ред;

- Стойностите на средната величина (μ) и стандартното отклонение (σ) за признака y във всеки времеви контур ще бъдат константи и също няма да зависят от времето.

И двете следствия водят до значително опростяване при изследването на стационарни процеси.

Освен посочената дефиниция се използва още една форма на стационарност, която, макар и с отслабени изисквания, позволява да се опрости анализът и намира важно практическо приложение. Слаба форма на стационарност, при която е необходимо да бъдат изпълнени само изискванията за константна стойност на средната величина (μ) и стандартното отклонение (σ).

Освен че разширява обхвата на процесите, които могат да бъдат наречени стационарни, тази форма има още едно важно следствие, а именно – че автоковариацията зависи само от лага във времето и е функция само на този лаг. Това е резултат от използването на функцията на автоковариация, при условие на константна средна величина:

$$E\left[\left(y_{t_1} - \mu_{t_1}\right)\left(y_{t_2} - \mu_{t_2}\right)\right] =$$
$$yti1 - \mu yti2 - \mu = CYYti1,ti2 \quad (12)$$

Общата дефиниция на ковариацията е функция от две променливи – t_{i_1} и t_{i_2}, докато при слабата форма на стационарност тя е функция само на тяхната разлика. Тъй като стандартното отклонение също е константа от връзката:

$$R_{YY}(t_1, t_2) = \frac{C_{YY}(t_1, t_2)}{\sigma_{t_1}\sigma_{t_2}} = \frac{C_{YY}(t_{i_1} - t_{i_2})}{\sigma^2},$$

(13)

следва, че автокорелацията също е функция само на лага. Това може да бъде използвано за съкратено записване на операциите с процеси, отговарящи на условията на слабата форма на стационарност. Често като синоним на слабата форма на стационарност се използва „стационарност от втори ред” или наречена още „слаба форма на стационарност“. Ако един процес е стационарен, то това може да се тълкува, че в основата му стоят едни и същи движещи сили. Посредством анализа на времеви редове може да се установи кои са те и да използват практически получените резултати. Например, ако цената на една ценна книга следва стационарен процес, то може да се търсят неговите характеристики и оттам да се установи кое всъщност определя цената ѝ и кара тази цена да се променя.

Анализът на времеви редове с ARIMA [6] се състои от следните модели - авторегресия, интегрирана, с пързяща средна - ARIMA (AutoRegressive Integrated Moving Average) [7]. Записват се като ARIMA (p, d, q) модели.

Авторегресионният елемент p представлява влиянието на данните от p предишни моменти в модела. Интегрираният елемент d представлява тенденции (трендът) в данните, а елементът q показва колко члена се използват за изглаждане на малки флуктуации с помощта на пързящо средно. Стандартно времевият анализ на данните се провежда в следните три стъпки: идентифициране, оценка и диагностика [8].

Първата стъпка е идентификация на времевия ред, която включва изследване на данните с изчисляване и начертаване на графиката, на графиките на автокорелационните функции (ACFs) и частичните автокорелационни функции (PACFs) [9].

Автокорелациите са самостоятелни корелации на серия от резултати със себе си, като се прескачат един или повече периоди назад във времето (lag). Частичните автокорелации са самостоятелни корелации с междинни частни автокорелации. Различни авторегресионни с пълзящо средни шаблони (подмножества на данните с близко поведение) често имат влияние за специфични промени в автокорелациионните и частични автокорелационни функции.

Когато времевият ред е дълъг, може да има тенденции, показващи периодични промени, наричани сезонност, периодичност, или цикличност. Например, увеличени вирусни инфекции по време на зимните месеци, завишено електропотребление за отопление в бита и др. По този начин, сезонността е друга форма на автокорелация, която често се наблюдава в масивите от данни. Периодична промяна може да възниква и за по-кратки периоди от време. Например, качеството на производството може да се различава по деня от седмицата, достигайки в средата на седмицата максимум. Или потреблението на електроенергия се увеличава през почивните дни. Тези модели могат да се идентифицират с помощта ACFs и PACFs преди построяване на модела и да помогнат за предварително определяне на p, d, q.

Анализът на времеви редове е по-подходяща техника за моделиране на данни с автокорелация, отколкото, например линейна регресия. Най-честата причина да не се получават модели с линейна регресия и класическите методи е нарушаването на допускането за независимост на грешките. Грешките са също автокорелиращи, което също трябва да се отчита от модела.

Идентификацията на времевите редове е процес на намиране на параметрите на ARIMA (p, d, q). Принципно се търсят възможно най-малките стойности на параметрите, които обикновено са 0, 1, 2... Когато стойността е 0, елементът не е необходим в този модел. Средният елемент, d (тренд), се изследва преди p и q. Целта е да се определи дали процесът е стационарен, а ако не, да се преобразува към стационарен, чрез отстраняване на тренда преди определянето на стойностите на p и q.

Стационарният процес има постоянна средна стойност и малка дисперсия през целият времеви период на изследването, т.е. за променливата Y за ред с n наблюдения:

$$\overline{Y} = \frac{Y_1 + Y_2 + \cdots + Y_{1n}}{n} \approx \text{const}, \tag{14}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \tag{15}$$

Втората стъпка при моделиране на времеви серии от данни е построяването на модел и оценка на неговите параметри, тествани срещу нулевата хипотеза, че са равни на нула.

Третата стъпка е диагностиката, в която се изследват остатъците (резидиумите). Остатъците са разликите между предсказаните (изчислени) по модела стойности и наблюдаваните данни. Теоретично допускането е, че остатъците са случайни и имат нормално разпределение. Ако това не е така, вероятно има още шаблони в данните, които не са отчетени. Ако всички шаблони от данни са отчетени в модела, остатъците са случайни. В много приложения на времевите редове, идентифицирането и моделирането са достатъчни, за да се намери уравнение, което да се използва за предсказване на бъдещето на процеса. Това се нарича прогнозиране, което е целта на много приложения на времевите редове в много области като икономика, екология, социология, техника и др.

ЗАКЛЮЧЕНИЕ

Времевите редове представляват множество от статистически наблюдения над някаква случайна величина, подредени в хронологичен ред.

В почти всички области на човешката дейност могат да се открият и наблюдават явления и събития, които могат да се определят като времеви редове.

В основите на разбирането на времевите редове е тяхното анализиране и моделиране.

Огромен интерес представлява възможността тези редове да бъдат предсказвани и прогнозирани.

Един от възможните подходи, посредством който да бъдат прогнозирани бъдещите стойности на един времеви ред [10], е използването на изкуствени неврони мрежи. Поради наличност на зависимости от много процеси може да бъде полезен метод за избягване на локални оптимуми, представен в [11]. Това ще бъде повод за по-нататъшни изследвания.

ЛИТЕРАТУРА

[1] S. Ilieva, Econometrics Lectures 6-8 Time Series, 2014.

[2] Gershenfeld and Neil, "The nature of mathematical modeling," Cambridge: Cambridge Univ. Press, ISBN 978-0521570954, OCLC 174825352, 2000.

[3] D. Damgaliev and J. Tellalyan, "Business statistics, Publisher NBU, 2006, subject 11".

[4] C. Granger, "Some Properties of Time Series Data and Their Use in Econometric Model Specification," in University of California at San Diego, La Jolla, CA 92093, USA.

[5] S. Kabaivanov, Econometrics for financiers, first edition, Plovdiv, 2014.

[6] G. P. Zhang, "Time Series Forecasting Using a Hybrid Arima and Neural Network Model" in Department of Management, J. Mack Robinson College of Business, Georgia State University, University Plaza, Atlanta, GA 30303, USA Received 16 July 1999, accepted 23 November 2001.

[7] "Autoregressive integrated moving average," [Online].Available: https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average.

[8] B. G. Tabachnik and L. S. Fidell, "Using multivariate statistics, 5th edition, Pearson".

[9] C. W. J. Granger and R. Joyeux, "An Introduction to Long-memory Time Series Models and Fractional Differencing" in University of California, San Diego and Cornell University.

[10] T. Balabanov, Forecasting with Heuristic Approaches in a Distributed Environment, Proceedings of Anniversary Scientific Conference 40 Years Department of Industrial Automation, ISBN:978-954-465-043-8, 163-166, 2011.

[11] I. Zankinski, "Effects of the Neuron Permutation Problem on Training Artificial Neural Networks with Genetic Algorithms". Sixth Conference on Numerical Analysis and Applications (NAA'16), 10187, Springer International Publishing Switzerland, 2017, ISBN:978-3-319-57098-3, DOI:10.1007, 777-782. SJR:0.339.