УДК 519.2

К.т.н., доцент Олефір О.С., студентка Савчук А.Б.

Національний технічний університет України «Київський політехнічний інститут»

МЕТОД ЛІ-КАРТЕРА ДЛЯ ПРОГНОЗУВАННЯ СМЕРТНОСТІ НАСЕЛЕННЯ ТА ЙОГО МОДИФІКАЦІЯ

Abstract

Alexander S. Olefir, assoc. prof., PhD; Savchuk Alla, student
The Lee-Carter method for forecasting mortality and its updating

This paper describes the basic Lee-Carter method and discusses the forecasts to which it has led. It then discusses extensions, applications, and methodological improvements that have been made in recent years; considers shortcomings of the method. The ways for further research are proposed as well.

Вступ

Управління розвитком країни або регіону, зокрема планування його бюджету, вимагає знання перспективної чисельності та особливостей вікової структури населення. Тому все більшої актуальності набувають дослідження особливостей відтворення населення, виявлення причинно-наслідкових зв'язків цього процесу з економічним розвитком окремих територій.

Смерть є однією із двох складових процесу природного руху населення. Зменшення смертності має вплив на підвищення природного, як наслідок, загального приросту населення [1]

Ця підвищена цікавість до прогнозу смертності супроводжувалась розвитком нових і більш складних методів [1-4]. Так в США на при кінці 1980-х – 1990-х помітно підвищилась смертність молодих чоловіків (15 – 44 років). Цей ріст був пов'язаний з епідемією СНІДу. Саме в цей період була розроблена модель Лі-Картера, однією з її цілей є адекватно відобразити в прогнозі даний підйом смертності. Метод вважається одним із кращих в наш час і широко використовується. В даній статті пропонується огляд методу Лі-Картера і його модифікація, яка дозволяє збільшити точність прогнозу.

Постановка задачі

Задача полягає в застосуванні метода Лі-Картера [2-3] для прогнозування смертності населення, що враховує всі можливі причини смертності у різних вікових групах і різних статях, та його вдосконалення, яке не ускладнює реалізацію в порівняні зі звичайним, але при цьому покращує результати.

Термінологія

Екстраполяція - наближене знаходження за рядом даних значень функції інших її значень, що містяться поза цим рядом [4].

Стохастичний процес – процес називається стохастичним, якщо він описується випадковими змінними, значення яких змінюються з часом [4].

Сингулярний розклад матриці A розміром $m \times n$ називається її представлення у вигляді $A = UWV^T$, де U – ортогональна матриця розміром $m \times m$, V – ортогональна матриця розміром $n \times n$, W – матриця розміром $m \times n$, на головній діагоналі якої знаходяться невід'ємні числа, розташовані в порядку спадання, а всі недіагональні елементи рівні нулю [4].

Авторегресійне інтегроване ковзаюче среднє (autoregressive integrated moving average, *ARIMA*) є узагальненням моделі авто регресійного ковзаючого середнього. Ці моделі використовуються при роботі з часовими рядами для більш глибокого розуміння даних або прогнозування майбутніх точок ряду. Зазвичай модель згадується, як *ARIMA (p, d, q)*, де p, d і q - цілі невід'ємні числа, що характеризують порядок для частин моделі (відповідно авторегресійной, інтегрованої і ковзаючого середнього) [1].

Опис вдосконаленого методу Лі-Картера

Переважна більшість сучасних демографічних прогнозів смертності розраховані методом екстраполяції. Гіпотеза про те, що майбутня зміна смертності є продовженням змін, які відбувалися і відбуваються, здається доволі природною, особливо, якщо врахувати те, що смертність населення більшості країн світу стійко знижується. Тим не менше, і в найбільш успішних країнах інколи виникають зворотні тенденції, які ставлять під сумнів можливості звичайного екстраполяційного прогнозування. Метод Лі-Картера є суттєвим відходом від попередніх, особливо це пов'язано з двома факторами (вік і час) та використанням матричного розкладу для того, щоб отримати єдиний змінний за часом коефіцієнт рівня смертності.

Метод був розроблений для довгострокового прогнозу, основаного на тривалому часовому ряді історичних даних [2].

По суті модель Лі-Картера описується логарифмічно перетвореною центральною нормою смертності ($m_{x,t}$ – показник смертності для віку x в час t), як сума вікового компонента, який незалежний від часу (a_x), і продукт змінного часом параметра (k_t, також відомий як індекс смертності), який підсумовує загальний рівень смертності і додаткового вікового компонента (b_x), який показує, як швидко або повільно змінюється смертність в кожному віці, коли індекс смертності змінюється. Математично

$$ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$$

Останній компонент $\varepsilon_{x,t}$ є залишковим членом, який відображає вікові впливи, що не захоплені моделлю.

Щоб отримати унікальне рішення, ми беремо a_x як середнє арифметичне $ln(m_{x,t})$ протягом довгого часу, тоді як суми b_x і k_t нормалізовані і дорівнюють одиниці та нулю відповідно:

$$a_x = \sum_x b_x^2 = 1, \qquad \sum_t k_t = 0$$

Використання звичайного метода найменших квадратів неможливе. Аби подолати цю перешкоду використовується двох-етапна процедура оцінки, яка дає точні рішення. У першій стадії до матриці $\{\ln(m_{x,t}) - a_x\}$ застосовується сингулярний розклад, аби отримати оцінки основних компонент b_x і k_t. На другій стадії часовий ряд k_t повторно оцінюється, таким чином, що

$$D_t = \Sigma\{exp(a_x + b_x k_t) N_{x,t}\}$$

де D_t - загальна кількість смертельних випадків в часі t, і N_x є популяційний вік x в часі t.

Є декілька переваг в тому, щоб здійснювати таким чином другу оцінку параметра k. Таким чином, емпіричний часовий ряд k може бути розширений, щоб включити роки, на протязі яких вікові дані про смертність не доступні.

На наступному кроці потрібно змоделювати k як стохастичний процес часового ряду. Найчастіше використовується стандартна одновимірна модель часового ряду ARIMA (0,1,0).

$$k_t = \theta + k_{t-1} + \varepsilon_t$$

де θ параметр дрейфу і

$$\theta = \frac{k_t - k_{t-1}}{T - 1}$$

це означає, що θ залежить лише від перших і останніх оцінок k_t; ε_t - залишковий член. Потім, щоб передбачити на два періоди вперед, ми лиш замінюємо k_{t-1}:

$$k_t = \theta + k_{t-1} + \varepsilon_t = (\theta + k_{t-2} + \varepsilon_{t-1}) + \theta + \varepsilon_t = k_{t-2} + 2\theta + (\varepsilon_{t-1} + \varepsilon_t)$$

Щоб передбачити k_t в час $T+(\Delta t)$ з даними відомими для періоду T, ми слідуємо за тією ж самою процедурою і отримуємо:

$$k_{T+\Delta t}=k_T+(\Delta t)\theta+\sum_{n}^{(\Delta t)}\varepsilon_{T+n-1} = k_T+(\Delta t)\theta+\sqrt{(\Delta t)}\varepsilon_t$$

Якщо проігнорувати залишковий член, ми можемо отримати пряму, яка є функцією від (Δt), з нахилом θ: $\frac{k_T - k_1}{T-1}$

$$k_{T+\Delta t}=k_T+(\Delta t)\theta= k_T+(\Delta t)$$

Недоліком методу э те, що він пропонує визначений спосіб зміни смертності у розподілі за віком, таким чином, що норми зниження у різному віці завжди підтримують одне і те ж відношення на протязі довгого часу. Але на практиці, відносна швидкість зниження в різному віці може змінюватись. Наприклад у Швеції смертність в старості мала тенденцію зменшуватись більш повільно, ніж в іншому віці. Метод не може врахувати такі зміни [3].

Проблема оригінального методу Лі-Картера полягає в тому, що модель на початку прогнозу може не відповідати конкретним віковим даним про смертність точно, тобто $m_{x,T} \neq \exp\left(a_x + b_x k_T\right)$. Така ситуація неминуче призведе до помилки, яка буде особливо важливою в перші роки прогнозу. Ця проблема може бути вирішена, якщо замість того, щоб оцінювати коефіцієнт a_x як середнє арифметичне $ln\ (m_{x,t})$, встановити a_x рівне логарифму нових показників смертності. Це гарантує, що перший рік прогнозу близько співпаде з показниками смертності, які спостерігались востаннє. В якості початкової оцінки потрібно взяти останні показники. Тоді прогноз отримаємо за допомогою формули:

$$ln(m_{x,t+s}) = ln(m_{x,t}) - b_x(k_{t+s} - k_t).$$

Висновки

Модель Лі-Картера є однією з найбільш популярних методологій для прогнозування смертності. Модель є простою і дуже успішно використовується в США та інших економічно розвинених країнах. Вдосконалення методу, гарантує, що основна оцінка точно відповідає віковим показникам смертності та тривалості життя, що спостерігаються, і що прогнози вікових норм плавно розвиваються від цих значень.

Є багато розширень основного методу: розбиття за статтю, регіонами і т.д. В подальшому було б доцільно визначити достоїнства і недоліки усіх розширень, та вибрати найкраще поєднання.

Література

1. *Елисеева И. И., Васильева Э. К.* Демография и статистика населения – М.: Финансы и статистика, 2006. - С.189-273.
2. *Rob J. Hyndman, Leonie Tickle, Booth H.* Demographic research: Article 9, Rostock , GERMANY, 2006. – P.289 – 310.
3. *Zheng Wang J.* Fitting and forecasting mortality for Sweden: applying the Lee-Carter model – Stockholm, 2007. – P.8-47.
4. *Бестужев-Лада И.В.* Рабочая книга по прогнозированию. М.: Мысль, 1982. – С.6-24.