



ORIGINAL RESEARCH PAPER

Comparison of singular value set analysis method in predicting mortality rate with Lee-Carter family methods

M. Zokaei, F.S. Alehoseini\*

Department of Insurance Statistics, Shahid Beheshti University, Tehran, Iran

ARTICLE INFO

Article History

Received: 20 January 2013

Revised: 23 February 2013

Accepted: 26 August 2013

Keywords

Mortality rate; Lee-Carter Method; Hyndman-Ullah Method; Singular Value Decomposition; Singular Spectrum Analysis.

ABSTRACT

Considering the importance of accurate modeling and forecasting of mortality rate in many decisions in the field of demography and statistical insurance, in this article, for the first time, the ability of singular value set analysis method (a non-parametric method in time series analysis) in modeling and forecasting the rate We will pay death.

This review is based on comparing the results of this method with several members of one of the most famous and widely used families in this field, known as Lee-Carter. So far, several articles have been presented regarding the capability of the single value set analysis method in modeling and forecasting many time series; This work will be another attempt to investigate another capability of this method in modeling and predicting mortality rate.

From the investigations, it can be concluded that in many cases, the singular value set analysis method is more accurate than the studied members of the Lee-Carter family; Therefore, due to the non-parametric nature of this method, if its other theoretical aspects are expanded, it can be used as a powerful method in modeling and predicting the mortality rate.

\*Corresponding Author:

Email: [f\\_alehoseini@yahoo.com](mailto:f_alehoseini@yahoo.com)

DOI: 10.22056/ijir.2014.01.02



## مقایسه روش تحلیل مجموعه مقادیر تکین در پیش‌بینی نرخ مرگ‌ومیر با روش‌هایی از خانواده لی - کارتر

محمد ذکایی، فاطمه سادات آل حسینی\*

گروه آمار بیمه، دانشگاه شهید بهشتی، تهران، ایران

### چکیده:

نظر به اهمیت مدل‌بندی و پیش‌بینی دقیق نرخ مرگ‌ومیر در بسیاری از تصمیم‌گیری‌های حوزه جمعیت‌شناسی و بیمه آمار، در این مقاله، برای نخستین بار، به بررسی توانایی روش تحلیل مجموعه مقادیر تکین (یک روش ناپارامتری در تحلیل سری‌های زمانی) در مدل‌بندی و پیش‌بینی نرخ مرگ‌ومیر خواهیم پرداخت.

این بررسی بر اساس مقایسه نتایج حاصل از این روش با چند عضو یکی از مشهورترین و پرکاربردترین خانواده‌ها در این حوزه، موسوم به لی-کارتر، انجام شده است. تاکنون، مقاله‌های متعددی در خصوص توانمندی روش تحلیل مجموعه مقادیر تکین در مدل‌بندی و پیش‌بینی بسیاری از سری‌های زمانی ارائه شده است؛ این اثر، تلاش دیگری در جهت بررسی یکی دیگر از قابلیت‌های این روش در مدل‌بندی و پیش‌بینی نرخ مرگ‌ومیر خواهد بود.

از بررسی‌های انجام‌شده می‌توان نتیجه گرفت که در بسیاری از موارد، روش تحلیل مجموعه مقادیر تکین از دقت بالاتری نسبت به اعضای تحت مطالعه خانواده لی-کارتر برخوردار است؛ از این رو، با توجه به ماهیت ناپارامتری این روش، در صورت بسط دیگر جنبه‌های نظری آن، می‌توان از آن به عنوان روشی توانمند در مدل‌بندی و پیش‌بینی نرخ مرگ‌ومیر استفاده کرد.

### اطلاعات مقاله

تاریخ دریافت: ۰۱ بهمن ۱۳۹۱  
تاریخ داوری: ۰۵ اسفند ۱۳۹۱  
تاریخ پذیرش: ۰۴ شهریور ۱۳۹۲

### کلمات کلیدی

نرخ مرگ‌ومیر  
روش لی-کارتر  
روش هیندمن-اولا  
تجزیه ویژه مقدار  
تحلیل مجموعه مقادیر تکین

\*نویسنده مسئول:

ایمیل: [f\\_alehosseini@yahoo.com](mailto:f_alehosseini@yahoo.com)

DOI: 10.22056/ijir.2014.01.02

شاخص‌های جمعیتی (همچون نرخ باروری، نرخ مرگ‌ومیر، نرخ رشد جمعیت) همواره نقش بسزایی در تصمیم‌گیری‌های کلان جمعیتی داشته‌اند. از این بین، نرخ مرگ‌ومیر مهم‌ترین عامل در تغییر الگوی جمعیت کشورها (به‌ویژه، کشورهای با سطح باروری پایدار) به حساب می‌آید.

داشتن تصویری درست از آینده جمعیت برای برآورد صحیح جمعیت سالخورده در سال‌های آتی به منظور برنامه‌ریزی‌های دقیق مالی جهت تأمین آینده این قشر از جامعه، ساخت جداول آتی عمر مطابق با محتمل‌ترین الگوی مرگ‌ومیر و قیمت‌گذاری صحیح سالانه‌های عمر (همچون حقوق بازنشستگی)، تنها بخشی از کاربردهای پیش‌بینی مقادیر آتی نرخ مرگ‌ومیر در جمعیت‌شناسی و بیمه آمار است. در این راستا، مدل‌بندی و پیش‌بینی نرخ خام مرگ در هر سن و سال خاص، به‌عنوان یک شاخص کلی از نرخ مرگ‌ومیر، از دیرباز، مورد توجه جمعیت‌شناسان قرار گرفته است. این مطلب ناشی از امکان ساده و سریع تبدیل این کمیت به سایر شاخص‌های مرگ‌ومیر و عناصر یک جدول عمر است. از این رو، هر تلاشی در جهت افزایش دقت پیش‌بینی الگوی آتی مرگ‌ومیر، تأثیر مستقیمی در بهبود نتایج در هر یک از کاربردهای فوق خواهد داشت.

این مقاله تلاشی در جهت بررسی توانایی روش تحلیل مجموعه مقادیر تکین<sup>۱</sup>، به عنوان یک روش ناپارامتری در تحلیل سری‌های زمانی، در مدل‌بندی و پیش‌بینی نرخ مرگ‌ومیر است. این بررسی بر اساس مقایسه دقت نتایج این روش با چند عضو از یک خانواده مشهور و پرکاربرد در این حوزه (لی-کارت<sup>۲</sup> صورت گرفته است (Lee and Carter, 1992).

از این رو، ابتدا، به معرفی روش لی-کارت<sup>۲</sup> و در ادامه، توسعه هیندمن-اولا<sup>۳</sup> خواهیم پرداخت. روش هیندمن-اولا با حفظ ساختار مدل لی-کارت<sup>۲</sup>، با نگاهی تابعی سعی در مدل‌بندی و پیش‌بینی بلندمدت نرخ خام مرگ کل جمعیت دارد. در ادامه، بر اساس آنچه که گولیندینا و همکاران<sup>۴</sup> در خصوص مبانی نظری و کاربردی روش تحلیل مجموعه مقادیر تکین بیان کرده‌اند، به شرح این روش خواهیم پرداخت. هیندمن و اولا، بر اساس یک ملاک مقایسه‌ای انتخابی، نشان دادند که توسعه پیشنهادی آنها از دقت بیشتری نسبت به روش لی-کارت<sup>۲</sup> و چند توسعه آن برخوردار است. از این رو، به ارزیابی دقت عملکرد روش تحلیل مجموعه مقادیر تکین در مقایسه با روش هیندمن-اولا در این حوزه خواهیم پرداخت.

از آنجاکه یک بانک اطلاعاتی جامع از داده‌های مرگ‌ومیر در ایران وجود ندارد و الگوی مرگ‌ومیر کشور فرانسه نیز، به عنوان نماینده‌ای از الگوی مرگ‌ومیر حاکم در کشورهای توسعه‌یافته، به الگوی مرگ‌ومیر ایران نزدیک است، ارزیابی عملکرد روش تحلیل مجموعه مقادیر تکین بر اساس نرخ خام مرگ کل جمعیت این کشور انجام خواهد شد.

#### ادبیات موضوع

مدل‌بندی نرخ مرگ‌ومیر پیشینه‌ای طولانی دارد. برجسته‌ترین محققین در این حوزه، دموآور<sup>۵</sup>، گومپرتس<sup>۶</sup>، می‌کام<sup>۷</sup>، سانگ<sup>۸</sup> و وایبول<sup>۹</sup> هستند. از سوی دیگر، پیش‌بینی نرخ مرگ‌ومیر، پیشینه کوتاه‌تری دارد. روش‌هایی که تا حدود بیست‌وپنج سال قبل به منظور پیش‌بینی نرخ مرگ‌ومیر مورد استفاده قرار می‌گرفتند، نسبتاً ساده و تا حدودی مبتنی بر سلیقه فردی بودند. اما طی بیست سال گذشته، روش‌های پیشرفته‌تری در این حوزه ارائه شده است. این روش‌ها کارشناسان بیمه آمار و جمعیت‌شناسان را به استفاده بیشتر از روش‌های آماری سوق داده‌اند. از میان این روش‌ها، روش لی-کارت<sup>۲</sup> با توجه به ساختار ساده و کاربرد موفق آن در بسیاری از کشورها، بیشتر مورد استفاده جمعیت‌شناسان و کارشناسان بیمه آمار قرار گرفته است.

1. Singular Spectrum Analysis (SSA)

2. Lee- Carter (LC)

3. Hyndman – Ullah (HU)

4. Golyandina et al., 2001

5. De Moivre, 1725

6. Gompertz, 1825

7. Makeham, 1867

8. Sang, 1868

9. Weibull, 1939

پیدایش روش لی-کارت به تغییر الگوی امید به زندگی از سال ۱۹۰۰ در ایالات متحده آمریکا برمی‌گردد. این اتفاق زمینه‌ای برای کار تحقیقاتی لی و کارت روی مدل‌بندی و پیش‌بینی بلندمدت نرخ مرگ‌ومیر کل جمعیت ایالات متحده به نام "مدل‌بندی و پیش‌بینی سری‌های زمانی جمعیتی" شد. گرچه این روش بر اساس داده‌های مرگ‌ومیر ایالات متحده آمریکا، داده‌های ۱۹۰۰ تا ۱۹۸۷، طراحی شده بود، عملکرد بسیار خوبی در مدل‌بندی و پیش‌بینی بلندمدت نرخ مرگ‌ومیر بسیاری از کشورهای توسعه‌یافته از جمله شیلی (۱۹۹۴)، ژاپن (۱۹۹۶)، کشورهای عضو G7<sup>۱</sup> (۲۰۰۰)، انگلستان (۲۰۰۳)، سوئد (۲۰۰۴)، ایتالیا (۲۰۰۵) داشته است.

یکی از توسعه‌های قابل توجه از مدل لی-کارت را هیندمن و اولاً<sup>۲</sup> ارائه کردند. روش‌های بسیاری در حوزه پیش‌بینی هموار<sup>۳</sup> مرگ‌ومیر ارائه شده است؛ روش هیندمن-اولا نمونه‌ای از این رده است که علاوه بر اینکه عضوی از خانواده لی-کارت به حساب می‌آید، با رویکردی تابعی سعی در پیش‌بینی استوار لگاریتم نرخ خام مرگ کل جمعیت هر سن دارد.

از سوی دیگر، روش تحلیل مجموعه مقادیر تکین، به عنوان یک روش ناپارامتری نسبتاً جدید، با استفاده از روش تجزیه ویژه مقدار به کاهش سطح نوفه، مدل‌سازی و پیش‌بینی سری‌های زمانی می‌پردازد. سادگی، عدم وابستگی به پیش‌فرض‌های محدودکننده رایج در سایر روش‌ها (همچون مانایی، خطی بودن سری و نرمال بودن باقی‌مانده‌ها) و کارایی مناسب در تحلیل سری‌های زمانی با طول کم از جمله دلایل‌های گسترش روزافزون استفاده از این روش در علوم مختلف است. از این رو، با توجه به توانایی‌های این روش، بررسی عملکرد آن برای اولین بار، در مدل‌بندی و پیش‌بینی نرخ خام مرگ، هدف این مقاله بوده است.

#### روش لی-کارت

فرض کنید،  $d_{x,t}$  و  $e_{x,t}$ ، به ترتیب، بیانگر تعداد افراد فوت‌شده و تعداد افراد در معرض مخاطره فوت در سن  $X$  و سال  $t$  باشند. نرخ خام مرگ در سن  $X$  و سال  $t$  در یک جامعه با  $m(x, t)$  نمایش داده شده و به کمک رابطه زیر محاسبه می‌شود:

$$m(x, t) = \frac{d_{x,t}}{e_{x,t}}$$

که در آن،  $t_1, t=t_1, t_1+1, \dots, t_1+h-1=t_n$  نخستین زمان تحت مطالعه و  $x=x_1, \dots, x_k$  است (Wilmoth, 2002). ساختار مدل پیشنهادی لی و کارت به شکل زیر است:

$$\ln m(x, t) = a_x + b_x k_t + \epsilon_{x,t} \quad (1)$$

-  $\ln m(x, t)$ : لگاریتم طبیعی نرخ خام مرگ مشاهده‌شده در سن  $X$  و سال  $t$ ؛

-  $a_x$ : متوسط لگاریتم طبیعی نرخ خام مرگ در سن  $X$ ؛

-  $k_t$ : روند اصلی موجود در لگاریتم طبیعی نرخ‌های خام مرگ تمامی سن‌ها در طول زمان؛

- ضرایب  $b_x$ : بیانگر این مطلب هستند که با تغییر در سطح مرگ‌ومیر ( $k_t$ ) مرگ‌ومیر در سن  $X$  به چه میزان تمایل به تغییر دارد. برای مثال، مقدار این پارامتر برای گروه سنی نوزادان و کودکان بزرگ‌تر از گروه سنی سالمندان است (در این مدل، مقادیر پارامتر  $b_x$  در طول زمان ثابت در نظر گرفته می‌شود)؛

-  $\epsilon_{x,t}$ : مؤلفه خطا در سن  $X$  و سال  $t$  است. این مؤلفه با میانگین صفر و واریانس  $\sigma_\epsilon^2$  بیانگر آن دسته از تغییرهای خاص سنی در نرخ مرگ‌ومیر است که به کمک این مدل منعکس نمی‌شود.

<sup>۱</sup> گروه G7 متشکل از وزرای دارایی و رئیس بانک‌های مرکزی هفت کشور پیشرفته اقتصادی شامل فرانسه، کانادا، آلمان، ایتالیا، ژاپن، انگلستان و آمریکا بوده که همه ساله در جلسات خود به بحث در خصوص مسائل اقتصادی رایج در دنیا می‌پردازند.

<sup>۲</sup> Hyndman and Ullah, 2007

<sup>۳</sup> رویکرد پیش‌بینی هموار مرگ‌ومیر، اغلب در حوزه تحلیل تابعی داده‌های مرگ‌ومیر مطرح شده که در آن به کمک یک روش هموارسازی مناسب، داده‌های دورافتاده مرگ‌ومیر در مدل‌بندی حذف می‌شوند. در نتیجه، از آنجاکه داده‌های دورافتاده در این رویکرد شناسایی و حذف شده‌اند، مدل‌بندی حاصل هموار خواهد بود.

مدل (۱) را در نظر بگیرید. از آنجاکه هیچ متغیر توضیحی در سمت راست این رابطه وجود ندارد، مدل را نمی‌توان به کمک روش‌های رایج رگرسیونی برازش داد. از طرف دیگر، ساختار مدل تحت هر یک از تبدیل‌های زیر ثابت است:

$$\forall c \in \mathbb{R}, c \neq 0; \{a_x, b_x, k_t\} \rightarrow \left\{ a_x, cb_x, \frac{k_t}{c} \right\}$$

$$\forall c \in \mathbb{R}; \{a_x, b_x, k_t\} \rightarrow \{a_x - cb_x, b_x, k_t + c\}$$

از این رو، پارامترهای مدل به منظور یافتن یک مجموعه جواب یکتا برای هر سن به کمک دو قید زیر استاندارد می‌شوند:

$$(2) \sum_{t=t_1}^{t_n} k_t = 0, \quad \sum_{x=x_1}^{x_k} b_x = 1$$

دو قید فوق مفهوم جمعیتی نیز دارند. قید اول بیانگر مجموع انحراف‌ها از روند کلی مرگ‌ومیر در طول بازه  $[t_1, t_n]$  است؛ بنابراین، مجموع این انحراف‌ها صفر است. از سوی دیگر،  $b_x$  نیز بیانگر پاسخ سن  $X$  به تغییر در سطح  $k_t$  در طول دوره برازش داده شده است؛ بنابراین، مجموع این پارامتر برابر یک (یا هر واحد انتخابی دیگر) خواهد بود (Shenas Setareh, 2011). لی و کارتز، به کمک دو قید فوق دو روش برای برآورد پارامترها معرفی کرده‌اند:

- روش کمترین توان‌های دوم عادی خطا؛

- روش تقریبی.

از آنجاکه هیندمن و اولاد در مقایسه توسعه خود با روش لی-کارتز از برآوردهای حاصل از روش کمترین توان‌های دوم عادی خطا استفاده کرده‌اند، در ادامه، تنها به بحث در خصوص برآوردهای حاصل از این روش خواهیم پرداخت<sup>۱</sup>. فرض کنید، نرخ خام مرگ کل جمعیت را در هر سن خاص و برای تمام سال‌های  $t_1$  تا  $t_n$  در اختیار داریم. در این صورت، به کمک رابطه‌های (۱) و (۲) برآورد کمترین توان‌های دوم عادی خطا برای پارامتر  $a_x$  به شکل زیر به دست می‌آید:

$$\hat{a}_x = \frac{1}{h} \sum_{t=t_1}^{t_n} \ln m(x, t), \quad t=t_1, \dots, t_n, \quad h=t_n-t_1+1$$

لی و کارتز به منظور برآورد کمترین توان‌های دوم عادی خطا برای پارامترهای  $b_x$  و  $k_t$  از تقریب درجه اول تجزیه ویژه مقدار استفاده کردند. از این رو، برآوردهای حاصل به این شکل به دست می‌آید:

$$\hat{b}_x = U_{x \times 1}, \quad \hat{k}_t = I_1 V_{t \times 1}$$

که در آن:

$$\hat{b}_x = \begin{pmatrix} u_{x_1,1} \\ \vdots \\ u_{x_k,1} \end{pmatrix}, \quad \hat{k}_t = \begin{pmatrix} v_{x_1,1} \\ \vdots \\ v_{x_k,1} \end{pmatrix}$$

در رابطه فوق،  $I_1$  بزرگ‌ترین ویژه مقدار ماتریس  $ZZ'$ ،  $\hat{b}_x$  ویژه بردار چپ متناظر و  $\hat{k}_t$  ویژه بردار راست متناظر با این ویژه مقدار خواهند بود.

پیش‌بینی شاخص مرگ‌ومیر به کمک برون‌یابی

پس از برآورد پارامترهای مدل، پارامتر  $k_t$  به کمک مدل‌های سری‌های زمانی مدل‌بندی شده و مقادیر آتی آن پیش‌بینی می‌شود. مدل‌بندی نرخ خام مرگ بیشتر کشورهای توسعه‌یافته، مدل قدم‌زدن تصادفی با رانش  $k_t$  را به خوبی مدل‌بندی کرده و بیشتر تغییرهای موجود در داده‌های مرگ‌ومیر هر سن و سال خاص به کمک این مدل پوشش داده می‌شود.

<sup>۱</sup> برای مطالعه بیشتر ر.ک: آل حسینی، ۱۳۹۱

<sup>۲</sup> Eigenvalue

توسیع‌هایی از روش لی-کارتر

توسیع‌های بسیاری برای روش لی-کارتر ارائه شده است. روش‌های لی-میلر<sup>۱</sup> (Lee and Miller, 2001) و بوث و همکاران<sup>۲</sup> از جمله مشهورترین این توسیع‌هاست. از سوی دیگر، روش هیندمن-اولا با معرفی رویکردی کاملاً متفاوت در مدل‌بندی داده‌های مرگ‌ومیر، به نوعی آخرین توسیع بر روش لی-کارتر است که در کنار دقت بالاتر، ایرادهای موجود در این روش را نیز نداشته و از ویژگی‌های برتری نسبت به این روش برخوردار است. هیندمن و اولا با مقایسه عملکرد روش خود با روش‌های لی-کارتر، لی-میلر و بوث و همکاران، در مدل‌بندی و پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت فرانسه، نشان دادند که روش هیندمن-اولا از دقت بالاتری نسبت به سایر توسیع‌های لی-کارتر برخوردار است.

روش هیندمن-اولا

رویکرد اتخاذی در روش هیندمن-اولا مدل‌بندی نرخ مرگ‌ومیر کل جمعیت با نگاهی تابعی به لگاریتم نرخ خام مرگ هر سال است. در این روش، نمودار لگاریتم نرخ خام مرگ کل جمعیت در هر سال به شکل تابعی از سن  $x_1, \dots, x_p$  در نظر گرفته می‌شود. این روش، در کنار استفاده از تابع‌های همواری از سن، نسبت به سال‌های دورافتاده‌ای همچون وقوع جنگ‌ها یا شیوع بیماری‌های خاص استوار بوده و امکان اعمال قیدهایی جمعیتی مختلفی نیز در این مدل وجود دارد. لازم به ذکر است که هموارسازی داده‌های مرگ‌ومیر با توجه به مشکلات و خطاهای موجود در جمع‌آوری و ثبت این داده‌ها امری ضروری به نظر می‌رسد؛ به‌عنوان مثال، تمایل افراد به گرد کردن سن در هنگام بیان سن خود (همچون ۲۰، ۲۵ و ۳۰) منجر به تراکم توزیع سنی در چنین سن‌هایی می‌شود.

ساختار مدل

فرض اصلی در روش هیندمن-اولا بر این است که اگر  $y_t(x)$  بیانگر لگاریتم نرخ خام مرگ مشاهده‌شده در سن  $x$  و زمان  $t$  باشد، تابع پیوسته همواری همچون  $f_t(x)$  در زمان  $t$  وجود دارد، به طوری که با اندکی خطا در نقاط گسسته سن  $(x_1, \dots, x_p)$  قابل مشاهده است. بنابراین، مشاهده‌ها در این روش به شکل  $\{x_i, y_t(x_i)\}$ ،  $(i=1, \dots, p)$  و  $(t=1, \dots, n)$  نمایش داده می‌شوند، که در آن:

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\epsilon_{t,i} \quad (3)$$

$\epsilon_{t,i}$ : متغیرهای تصادفی مستقل و هم‌توزیع با توزیع نرمال استاندارد؛

$\sigma_t(x_i)$ : آن مقدار از نوفه موجود در سری است که با تغییر سن، تغییر می‌کند. (وجود این مؤلفه در خطای مدل منجر به تصحیح فرض هم‌واریانسی در مدل لی-کارتر می‌شود). از این رو، ساختار روش هیندمن-اولا در مدل‌بندی و پیش‌بینی لگاریتم نرخ خام مرگ را می‌توان در پنج گام بیان کرد:

- گام اول: داده‌های تحت مطالعه، برای یک سال مشخص، به کمک یک روش ناپارامتری هموار می‌شوند؛

- گام دوم: خم برازنده‌شده در گام قبل، به کمک مدل زیر و با استفاده از یک بسط تابعی پایه تجزیه می‌شود:

$$f_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x) \quad (4)$$

$\mu(x)$ : معیار مکانی از  $f_t(x)$ ؛

$\{\phi_k(x)\}$ : یک مجموعه از تابع‌های پایه‌ای یک‌معامد<sup>۳</sup>؛

$K$ : تعداد مؤلفه‌های اصلی به‌کاررفته در مدل پویایی این فرایند به کمک ضریب‌های  $\{\beta_{t,k}\}$  (با فرض وجود استقلال در میان هر یک از

این سری‌های زمانی) مورد مطالعه قرار می‌گیرد؛

<sup>1</sup>. Lee - Miller

<sup>2</sup>. Booth et al., 2005

<sup>3</sup>. Orthonormal

-  $e_t(x)$ : خطای مدل و دارای توزیع نرمال با میانگین صفر و واریانس  $\theta(x)$  است (در مدل هیندمن - اولاً، فرض می‌کنیم که خطاهای مدل به طور پیاپی ناهمبسته هستند).

- گام ۳: به هر یک از ضریب‌های  $\{\beta_{t,k}\}$  یک مدل سری زمانی تک متغیره مناسب برازش داده می‌شود؛

- گام ۴: با استفاده از مدل انتخابی در گام قبل، مقادیر آتی ضریب‌های  $\{\beta_{t,k}\}$  برای یک دوره  $h$  ساله پیش‌بینی می‌شوند؛

- گام ۵: به کمک ضریب‌های پیش‌بینی شده در گام چهارم و با استفاده از رابطه (۴) مقادیر آتی  $f_t(x)$  و در ادامه، به کمک رابطه (۳)، مقادیر آتی  $y_t(x)$  برای  $t=n+1, \dots, n+h$  محاسبه می‌شوند.

روش‌های متعددی در هموارسازی ناپارامتری داده‌های تابعی (گام ۱) وجود دارد؛ هیندمن و اولاً از روش وود<sup>۱</sup> استفاده کرده‌اند. از سوی دیگر، در این روش، به منظور حذف اثر سال‌های دورافتاده از برآوردهای استوار در محاسبه  $\mu(x)$  و  $\phi_k(x)$ ، به ترتیب، روش هوسیر و کروکس<sup>۲</sup> و رمزی و سیلورمن<sup>۳</sup>، استفاده می‌شود (گام ۲). همچنین، سری زمانی  $\{\beta_{t,k}\}$  به شکل استوار برآورد نمی‌شود (گام ۳) تا امکان نمایش آشکار هر سال دورافتاده فراهم شده و بتوان چنین داده‌ای را به شکل یک داده دورافتاده در سری زمانی مدل‌بندی کرد (گام ۴). در ادامه، سری مدل‌بندی شده به روش استوار چن و لیو<sup>۴</sup> پیش‌بینی می‌شود (گام ۵).

پیش‌بینی

به کمک رابطه‌های (۳) و (۴)، مدل لگاریتم نرخ خام مرگ به شکل رابطه ۵ به دست می‌آید:

$$y_t(x_i) = \mu(x_i) + \sum_{k=1}^K \beta_{t,k} \phi_k(x_i) + e_t(x_i) + \sigma_t(x_i) \epsilon_{t,i} \quad (5)$$

بنابراین، با شرطی کردن رابطه بالا روی دو مجموعه  $I = \{y_t(x_i); t=1, \dots, n; i=1, \dots, p\}$  (داده‌های مشاهده شده) و  $\Psi$  (تابع‌های پایه)، پیش‌بینی  $h$  گام آتی برای  $y_{n+h}(x)$  به کمک رابطه ۶ به دست خواهد آمد:

$$\hat{y}_{n,h}(x) = E[y_{n,h}(x) | I, \Psi] = \mu(x) + \sum_{k=1}^K \hat{\beta}_{k, n+h} \hat{\phi}_k(x) \quad (6)$$

(در رابطه فوق،  $\hat{\beta}_{k, n+h}$  همان پیش‌بینی  $h$  گام آتی  $\beta_{k, n+h}$  به کمک سری زمانی برآورد شده از  $\{\hat{\beta}_{t1,k}, \dots, \hat{\beta}_{n,k}\}$  است).

دقت پیش‌بینی و انتخاب مرتبه

فرض کنید:

$$e_{n,h} = y_{n+h}(x) - \hat{y}_{n,h}(x)$$

بیانگر خطای پیش‌بینی حاصل از رابطه (۶) باشد.

بنابراین، خطای توان دوم جمع‌بسته پیش‌بینی را به شکل رابطه (۷) تعریف می‌کنیم:

$$ISFE_n(h) = \int_x e_{n,h}^2(x) dx \quad (7)$$

هیندمن و اولاً از رابطه فوق به عنوان یک معیار در انتخاب مرتبه مدل ( $K$ ) استفاده کرده‌اند. روش کار چنین است: ابتدا، مدل را به داده‌های  $t$  سال انتخابی برازانده و یک دوره  $m$  ساله را به کمک این مدل پیش‌بینی می‌کنیم. سپس، به کمک رابطه (۷)، خطای توان دوم جمع‌بسته پیش‌بینی متناظر با هر یک از افق‌ها را محاسبه می‌کنیم. اکنون، به کمک مقادیر به دست آمده، مقداری از  $K$  را که منجر به کمترین

مقدار  $\sum_{t=N}^{n-h} \sum_{h=1}^m ISFE_t(h)$  می‌شود ( $N$ ، کمترین سال مورد استفاده در برازش مدل است) برمی‌گزینیم.

<sup>1</sup>. Wood, 2003

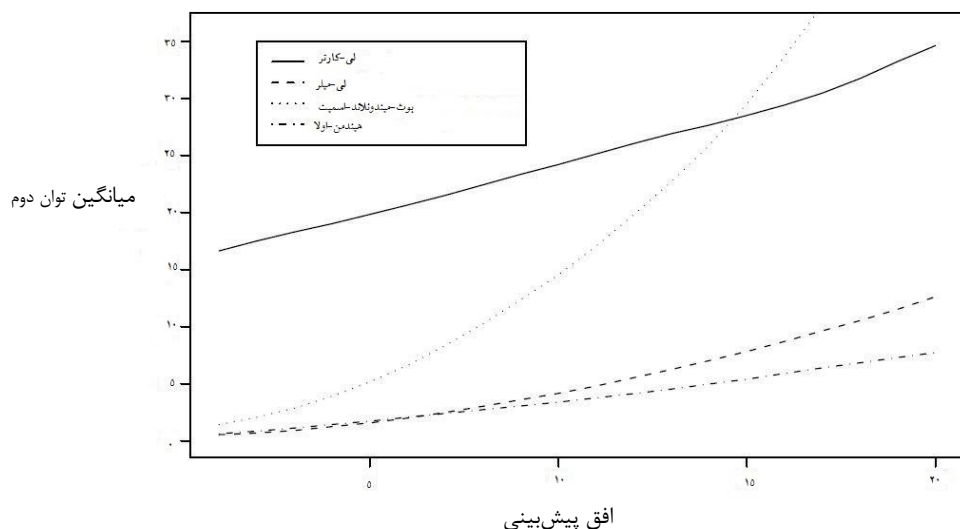
<sup>2</sup>. Hossjer and Croux, 1995

<sup>3</sup>. Ramsay and Silverman, 2005

<sup>4</sup>. Chen and Liu, 1993

مقایسه دقت عملکرد روش هیندمن-اولا با لی-کارتز

هیندمن و اولا بر اساس روش مقایسه‌ای در بخش پنجم این مقاله و با انتخاب دوره‌های برازش  $m=1899, \dots, 2000$  و دوره پیش‌بینی  $(2001, m+20)$  به بررسی دقت عملکرد روش خود با روش لی-کارتز، لی-میلر و بوث و همکاران (دو توسیع مشهور روش لی-کارتز) در مدل‌بندی و پیش‌بینی نرخ خام مرگ کل جمعیت فرانسه پرداختند. نمودار ۱ نتیجه این مقایسه را نشان می‌دهد. برای افق ۷ سال و کمتر، دو روش هیندمن-اولا و لی-میلر تقریباً دقت پیش‌بینی برابری دارند؛ اما، در افق‌های بزرگ‌تر، برتری استفاده از تابع‌های پایه بیشتر آشکار می‌شود.



(Hyndman and Ullah, 2007)

نمودار ۱: مقایسه دقت پیش‌بینی چهار روش لی-کارتز، لی-میلر، بوث و همکاران و هیندمن-اولا در مدل‌بندی نرخ خام مرگ کل جمعیت بر اساس میانگین توان دوم خطا<sup>۱</sup>

هیندمن و اولا با مقایسه دقت سه روش، لی-کارتز، لی-میلر و بوث و همکاران در پیش‌بینی نرخ خام مرگ کل جمعیت ۲۰ کشور نشان دادند که دو روش آخر نسبت به روش لی-کارتز از دقت بسیار بیشتری برخوردار هستند؛ با این وجود، عملکرد نسبتاً ضعیف روش بوث و همکاران در نمودار ۱ ناشی از دوره ارزیابی بزرگ‌تر در مطالعه هیندمن و اولا (1899-2001) نسبت به بوث و همکاران است. بنابراین، با توجه به دقت بیشتر و برتری‌های آشکار روش هیندمن-اولا در پیش‌بینی نرخ خام مرگ کل جمعیت، در ادامه، به ارزیابی دقت عملکرد دو روش مجموعه مقادیر تکین و هیندمن-اولا در پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت خواهیم پرداخت.

روش تحلیل مجموعه مقادیر تکین

در تحلیل سری‌های زمانی، مهم‌ترین مسئله، یافتن روشی است که در کنار توانایی بالا، نسبت به تغییرهای تصادفی نامنظم (نوفه) استوار باشد. تلاش‌های بسیاری در این حوزه انجام گرفته است با این حال، بسیاری از این روش‌ها مبتنی بر فرض‌های محدودکننده‌ای همچون مانایی، خطی‌بودن سری و نرمال‌بودن باقی‌مانده‌هاست. یک راه حل برای رفع این مشکل، تلاش در جهت یافتن روشی ناپارامتری است که در کنار برخورداری از توانمندی بالا، نسبت به نوفه موجود در سری زمانی مورد مطالعه نیز استوار باشد. روش‌های بسیاری به منظور کاهش سطح نوفه وجود دارد که در میان آنها، روش‌های مبتنی بر تجزیه ویژه مقدار<sup>۲</sup> از توانمندی بیشتری برخوردار هستند.

روش SSA، یک روش ناپارامتری نسبتاً جدید است که به کمک تجزیه ویژه مقدار به کاهش سطح نوفه، مدل‌سازی و پیش‌بینی سری‌های زمانی می‌پردازد. سادگی، عدم وابستگی به پیش‌فرض‌های محدودکننده رایج در سایر روش‌ها (همچون مانایی، خطی‌بودن سری و نرمال‌بودن

<sup>1</sup>. Mean Square Error

<sup>2</sup>. Singular Value Decomposition



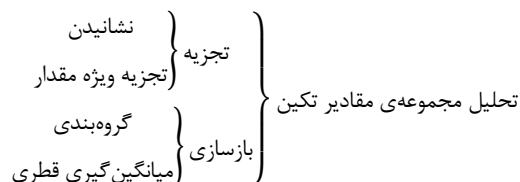
باقی مانده‌ها)، کارایی مناسب در تحلیل سری‌های زمانی با طول کم از جمله دلیل‌های گسترش روزافزون استفاده از این روش در علوم مختلف است.

پیدایش روش SSA اغلب به مقاله‌های بروم‌هد و کینگ<sup>۱</sup> نسبت داده می‌شود. تحقیق‌های این دو نفر و جذابیت روش SSA منجر به کاربرد آن در حوزه‌های مختلفی از علوم همچون علوم فیزیکی، اقتصادی، زیستی و مهندسی شده است. در این خصوص، بیش از صدها مقاله به نگارش درآمده است که بر اساس نتایج آنها، روش SSA در مدل‌بندی بسیاری از داده‌ها در مقایسه با سایر روش‌های موجود دارای توانمندی بسیار بالاتری است.<sup>۲</sup>

## روش شناسی پژوهش

### ساختار روش

هدف اصلی SSA تجزیه سری اصلی به تعدادی زیرسری است، به طوری که هر زیرسری را بتوان به عنوان روند، دوره، مؤلفه‌های فصلی یا نوفه سری مورد مطالعه در نظر گرفت. ویرایش اصلی تحلیل مجموعه مقادیر تکین شامل دو مرحله است؛ هر یک از مرحله‌های فوق نیز شامل دو مرحله است.



### نشاندن

نشاندن را می‌توان به عنوان یک نگاهت به منظور تبدیل یک سری زمانی حقیقی مقدار، غیر صفر و یک بعدی با طول  $N(N > 2)$ ،  $Y_N = (y_1, \dots, y_N)^T$  به سری‌های زمانی چند بعدی  $X_1, \dots, X_k$  (هر کدام به طول  $L$ ) با بردارهای زیر در نظر گرفت:

$$i = 1, \dots, k, X_i = (y_i, \dots, y_{i+L-1})^T \in R^L$$

که در آن عدد صحیح  $L(2 \leq L \leq N-1)$  طول پنجره نامیده می‌شود و:

$$K = N - L + 1$$

نتیجه گام نشاندن، این ماتریس است:

$$X = [X_1, \dots, X_k]$$

که بردارهای تأخیر، ستون‌های این ماتریس را تشکیل می‌دهند:

$$X = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} y_1 & y_2 & y_3 & \dots & y_k \\ y_2 & y_3 & y_4 & \dots & y_{k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_N \end{bmatrix}$$

شکل ۲: ماتریس مسیر

<sup>۱</sup>. Broomhead and King, 1986 a and b

<sup>۲</sup>. برای مطالعه بیشتر ر.ک: محمودوند، ۱۳۹۱

ماتریس  $X$  را می‌توان به عنوان یک مجموعه داده چند متغیره با  $L$  مشخصه و  $k$  مشاهده در نظر گرفت. با توجه به شکل ۲، واضح است که  $X_{ij} = Y_{i+j-1}$  چنین ماتریسی که عناصر روی قطرهای فرعی آن با هم برابرند، ماتریس هنکل می‌نامند. با توجه به ساختار این ماتریس، به ازای هر  $L$  و  $N$  ثابت، همواره رابطه یک به یکی میان ماتریس مسیر و سری زمانی اصلی وجود دارد. در حقیقت، درایه‌های ستون اول و سطر آخر (یا سطر اول و ستون آخر) ماتریس مسیر همان سری زمانی تک متغیره اصلی است.

#### تجزیه ویژه مقدار

دومین گام از مرحله نخست، تجزیه ماتریس مسیر بر اساس تجزیه ویژه مقدار است. در این گام، ماتریس مسیر به کمک تجزیه ویژه مقدار به مجموع ماتریس‌های مقدماتی متعامد (هر یک با رتبه یک) تجزیه می‌شوند. ماتریس  $S = X'X$  با بعد  $L \times L$  را در نظر بگیرید. فرض کنید،  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$  بیانگر ویژه مقادیر این ماتریس و  $U_1, \dots, U_L$  مجموعه یک‌متعامد از ویژه بردارهای متناظر با این ویژه مقادیر باشند. قرار دهید:

$$d = \max\{i; \lambda_i > 0\}$$

که  $d$  همان رتبه ماتریس  $X$  است. در این صورت، تجزیه ویژه مقدار ماتریس  $X$  به شکل زیر خواهد بود:

$$X = X_1 + \dots + X_d \quad (۸)$$

که در آن به ازای  $i = 1, \dots, d$ ،  $V_i = \frac{X'U_i}{\sqrt{\lambda_i}}$  و  $X_i = \sqrt{\lambda_i}U_iV_i^T$  است. اگر ویژه مقادیر غیر تکراری باشند، بسط (۸) به شکلی یکتا تعریف می‌شود.

#### گروه‌بندی

در گروه‌بندی،  $d$  ماتریس موجود در رابطه (۸) به چند زیرگروه تقسیم می‌شوند. فرض کنید،  $i = \{1, \dots, d\}$  به  $m$  زیرگروه  $I_1, \dots, I_m$  افزایش شود؛ در این صورت، رابطه (۸) را می‌توان به شکل زیر بازسازی کرد:

$$X = X_{I_1} + \dots + X_{I_m} \quad (۹)$$

که در این رابطه:

$$X_{I_j} = \sum_{i \in I_j} X_i, \quad j=1, \dots, m$$

در ساده‌ترین حالت ممکن ( $m=2$ )، گروه نخست همان مؤلفه اصلی سری (سیگنال) و گروه باقی‌مانده، نوفه در نظر گرفته می‌شود. در این حالت، ۲ مقدار از بزرگ‌ترین مقادیر تکیین (ریشه دوم ویژه مقادیر) و ویژه بردارهای متناظر آنها به منظور تقریب سری اصلی انتخاب شده و سایر مقادیر تکیین به عنوان مؤلفه‌های نوفه در نظر گرفته می‌شوند (در ادبیات SSA، پارامتر  $r$  را نقطه برش می‌نامند).

#### میانگین‌گیری قطری

وجود رابطه یک‌به‌یک میان ماتریس هنکل حاصل از SSA و سری تحت مطالعه، یکی از ویژگی‌های مفید در این روش است. از سوی دیگر، ماتریس‌هایی که در مرحله گروه‌بندی حاصل می‌شود، دارای خاصیت هنکلی نیستند و از این رو، نمی‌توان به کمک آنها تقریب یکتایی از سری اصلی به دست آورد. یک روش بهینه برای حل این مشکل استفاده از میانگین‌گیری قطری است (Golyandina, et al. 2001). هدف از میانگین‌گیری قطری، تبدیل یک ماتریس به ماتریس هنکلی است که در ادامه، می‌توان آن را به یک سری زمانی برگرداند. اگر،  $Z_{ij}$  درایه  $ij$ امین جمله از سری حاصل با میانگین‌گیری  $Z_{ij}$  روی تمام  $i$  و  $j$ هایی که  $i+j=k+1$  به دست می‌آید. این رویه، میانگین‌گیری قطری یا هنکل‌سازی ماتریس  $Z$  نامیده می‌شود؛ بنابراین، می‌توان هنکل‌سازی را به این شکل فرمول‌بندی کرد:

فرض کنید،  $Y$  یک ماتریس  $L \times K$  با عناصر  $Y_{ij}$ ،  $1 \leq j \leq k$  و  $1 \leq i \leq L$  باشد، قرار دهید:

$$s_2 = \max\{1, i + j - N - 1 + L\} \text{ و } s_1 = \min\{L, i + j - 1\}$$

از این رو، عنصر  $\tilde{Y}_{ij}$  ماتریس  $HY$  (عملگر هنکل‌سازی است) برابر است با:

$$\tilde{Y}_{ij} = \frac{1}{s_2 - s_1 + 1} \sum_{l=s_1}^{s_2} Y_{l, i+j-l+1}$$

(محمودوند، ۱۳۹۱). بنابراین، با هنکلسازی رابطه (۹)، بسط  $X = \tilde{X}_{I_1} + \dots + \tilde{X}_{I_m}$  به دست می‌آید که در آن به ازای  $j = 1, \dots, m$   $\tilde{X}_{I_j} = HX_{I_j}$  است. با انتخاب مناسب  $r$  مقدار از  $m$  زیرسری، سری بازسازی شده از مرحله میانگین‌گیری قطری، یک سری با سطح نوفه اندک خواهد بود و از این‌رو، می‌توان از آن برای پیش‌بینی مقادیر آتی سری استفاده کرد.

پیش‌بینی با استفاده از روش تحلیل مجموعه مقادیر تکین

در ادامه، به بیان دو روش اصلی در پیش‌بینی بر اساس SSA خواهیم پرداخت.

پیش‌بینی تحلیل مجموعه مقادیر تکین به روش بازگشتی<sup>۱</sup>

از آنجایی که این الگوریتم پیش‌بینی مبتنی بر یک رابطه بازگشتی است، عنوان RSSA برای آن در نظر گرفته شده است. این روش را دانیلو<sup>۲</sup> و دانیلو و ژیکلافسکی<sup>۳</sup> معرفی کردند.

اساس این روش مبتنی بر این فرض است که اگر زیرفضای  $r$ -بعدی  $\mathcal{E}_r = \{U_1, \dots, U_r\}$  از  $R^L$  یک فضای عمودی نباشد ( $U_i$ ، ویژه بردارهای چپ ماتریس مسیر  $X$  هستند)، آنگاه هر بردار در این زیرفضا در یک رابطه بازگشتی صدق می‌کند، به طوری که آخرین مؤلفه آن را می‌توان به شکل یک ترکیب خطی منحصر به فرد از سایر مؤلفه‌ها نوشت؛ بنابراین، هر یک از گروه‌های ایجاد شده در مرحله گروه‌بندی را با فرض وجود شرط بالا می‌توان به کمک این رابطه بازگشتی ادامه داد. اکنون، اگر سری زمانی به گونه‌ای باشد که بتوان چنین رابطه بازگشتی را برای آن توصیف نمود، آنگاه  $M$  جمله نهایی سری به کمک یک رابطه بازگشتی منحصر به فرد به شکل زیر برآورد می‌شود:

$$g_i = \begin{cases} \tilde{y}_i & i=1, \dots, N \\ \sum_{j=1}^{L-1} a_j g_{i-j} & i=N+1, \dots, N+M \end{cases} \quad (10)$$

در رابطه فوق، بردار ضرایب،  $R = (a_1, \dots, a_{L-1})$ ، به کمک رابطه زیر محاسبه می‌شود:

$$R = \frac{1}{1 - (\pi_1^2 + \dots + \pi_r^2)} \sum_{i=1}^r \pi_i P_i^{\nabla}$$

که در آن  $\{P_1, \dots, P_r\}$  یک پایه متعامد یکه برای زیرفضای  $\mathcal{E}_r$ ،  $P_i^{\nabla}$  برداری شامل  $L-1$  مؤلفه نخست  $P_i$ ،  $\pi_i$  مؤلفه آخر بردار  $P_i$  و  $r$  نقطه برش تعیین شده در مرحله گروه‌بندی است. فرمول بازگشتی (۱۰) در حقیقت، بیانگر، پیش‌بینی سری زمانی تا  $M$  گام جلوتر است. نکته‌ای که در این روش حائز اهمیت است، عدم وابستگی پیش‌بینی به یک الگوی از پیش تعیین شده است؛ در حقیقت، در این الگوریتم، پیش‌بینی‌ها تنها با استفاده از رابطه‌های موجود میان داده‌های تحت مطالعه به دست می‌آیند.

پیش‌بینی تحلیل مجموعه مقادیر تکین به روش برداری<sup>۵</sup>

در این روش، به منظور پیش‌بینی  $M$  گام جلوتر یک سری، به کمک رابطه بازگشتی (۱۱)، ماتریس مسیر بازسازی شده را با اضافه کردن تعداد  $M+L-1$  ستون جدید برآورد می‌کنیم.

$$Z_i = \begin{cases} \tilde{X}_i & i=1, \dots, k \\ \rho^{(\theta)} Z_{i-1} & i=k+1, \dots, k+M+L-1 \end{cases} \quad (11)$$

در رابطه فوق:

$$\rho^{(\theta)} = \begin{bmatrix} \Pi \\ R \end{bmatrix} \quad \Pi = \sum_{i=1}^r P_i^{\nabla} P_i^{\nabla T} + \frac{1}{1 - (\pi_1^2 + \dots + \pi_r^2)^2} R R^T$$

<sup>۱</sup>. Recurrent Singualr Spectrum Analysis (RSSA)

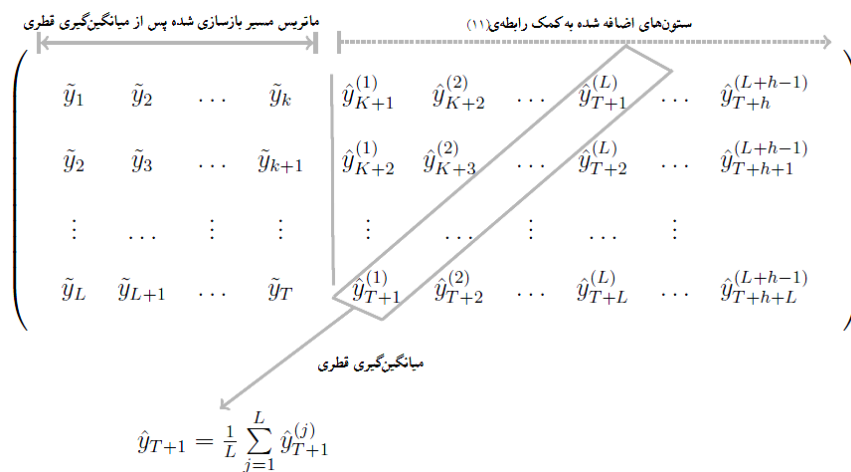
<sup>۲</sup>. Danilov, 1997

<sup>۳</sup>. Danilov and Zhigljavsky, 1997

<sup>۴</sup>. برای اثبات ر.ک: Golyandina et al., 2001

<sup>۵</sup>. Vertical Singular Spectrum Analysis

اکنون، با اعمال میانگین‌گیری قطری بر ماتریس حاصل (Z)، مؤلفه‌های  $N+1$  تا  $M+N$  سری، بیانگر پیش‌بینی سری زمانی تا  $M$  گام بعد خواهند بود. شکل ۳ پیش‌بینی برداری SSA را در یک گام جلوتر نمایش می‌دهد. پیش‌بینی به روش برداری برای یک افق یک ساله بر اساس  $L$  مقدار است؛ این در حالی است که با توجه به آنچه که گفته شد، این پیش‌بینی در روش بازگشتی بر اساس یک مقدار (اولین  $L$  در روش برداری) انجام می‌گیرد.



شکل ۳: پیش‌بینی یک گام جلوتر یک سری زمانی مفروض به روش VSSA

#### انتخاب پارامترها در روش تحلیل مجموعه مقادیر تکین

تاکنون، تلاش‌های بسیاری در زمینه تعیین مقدار بهینه دو پارامتر  $L$  و  $r$  انجام گرفته است. هر چند با توجه به ماهیت سری زمانی تحت مطالعه، مقدار این دو پارامتر نیز تغییر می‌یابد، انتخاب میانه طول سری تحت مطالعه به عنوان پارامتر  $L$ ، در بیشتر مطالعه‌های انجام گرفته، منجر به نتایج بهینه شده است.<sup>۱</sup>

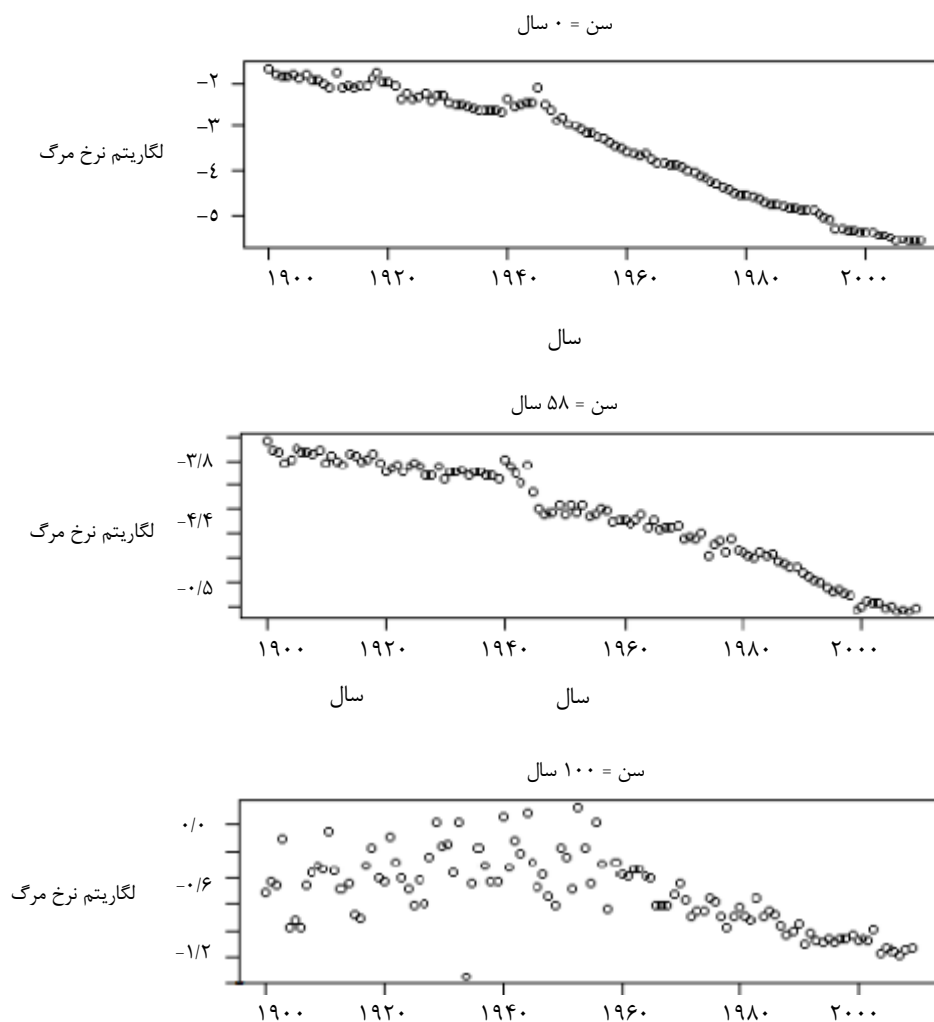
#### مقایسه روش تحلیل مجموعه مقادیر تکین و هیندمن-اولا

داده‌های این مقاله شامل نرخ خام مرگ کل جمعیت فرانسه در سنین ۰ تا ۱۰۰ سال و در سال‌های ۲۰۰۹-۱۸۹۹ بوده و از منظرگاه  $HMD^2$  تهیه شده است. در این مقایسه، پارامتر  $K$  مدل هیندمن-اولا در هر دوره برازش بر اساس رابطه (۷) تعیین شده است. از سوی دیگر، با انتخاب پارامتر  $L$ ، سری زمانی یک بعدی به یک مجموعه داده  $L$  متغیره با  $K$  تکرار تبدیل می‌شود. از این رو، در کنار میانه طول سری تحت مطالعه، این پارامتر را می‌توان با توجه به میزان تغییرپذیری الگوی موجود در این سری (در اینجا، لگاریتم نرخ خام مرگ هر سن خاص) نیز انتخاب کرد. با توجه به نمودار ۲، نمودار لگاریتم نرخ خام مرگ در سن صفر تقریباً ماهیتی خطی دارد؛ لازم به ذکر است که دو برآمدگی موجود در این نمودار ناشی از اثرهای جنگ جهانی اول و دوم در این کشور است. چنین الگویی با افزایش سن، تقریباً پایدار بوده و تنها سطح دو برآمدگی یادشده متغیر خواهند بود. از سوی دیگر، این الگو برای افراد بالای ۹۰ سال پراکنندگی بیشتری دارد. از این رو، انتخاب ما برای پارامتر طول پنجره در مدل‌بندی داده‌های مرگ‌ومیر فرانسه ۱۰ خواهد بود. در حقیقت، با بررسی این نمودار و ارزیابی نمونه‌ای خطای حاصل از انتخاب مقادیر مختلف طول پنجره (از مقداری برابر میانه داده‌های تحت مطالعه تا مقدار ۱۰) این نتیجه حاصل شده است که حداکثر استفاده از ۱۰ مؤلفه اصلی در مدل‌بندی این داده‌ها، به طور متوسط، منجر به نتایج مناسبی خواهد شد.

<sup>۱</sup> ر.ک: محمودوند، ۱۳۹۱

مدل‌بندی و پیش‌بینی نرخ خام مرگ کل جمعیت فرانسه به سه روش لی-کارتر، هیندمن-اولا و تحلیل مجموعه مقادیر تکین در آل حسینی (۱۳۹۱) بحث شده است.

<sup>۲</sup> Human Mortality Database, <<http://www.mortality.org>>



نمودار ۲: لگاریتم نرخ خام مرگ کل جمعیت فرانسه برای سن‌های ۵۸، ۰ و ۱۰۰

رویکرد اتخاذی در تعیین پارامتر  $r$  نیز بدین شکل است: لگاریتم نرخ خام مرگ هر سن را روی یک دوره انتخابی برازش داده و برای یک افق پیش‌بینی  $h$  ساله این کمیت را پیش‌بینی می‌کنیم. به آخرین سال دوره برازش فوق، یک سال دیگر اضافه کرده و یکبار دیگر، لگاریتم نرخ خام مرگ هر سن خاص را بر اساس این دوره برازش جدید برای همان دوره  $h$  ساله پیش‌بینی خواهیم کرد. این رویه را تا زمانی که آخرین سال دوره برازش انتخابی یک سال کمتر از آخرین سال داده‌های تحت مطالعه باشد، ادامه می‌دهیم. در این رهیافت، میانگین توان دوم خطای پیش‌بینی حاصل را به ازای مقادیر مختلف  $(1 < r < L - 1)$  و برای هر افق پیش‌بینی انتخابی محاسبه کرده و مقدار بهینه نقطه برش برای هر سن و هر افق پیش‌بینی را با توجه به کمترین میانگین توان دوم خطای پیش‌بینی به دست آمده، انتخاب می‌کنیم.

معیار مقایسه دقت عملکرد دو روش تحلیل مجموعه مقادیر تکین و هیندمن-اولا

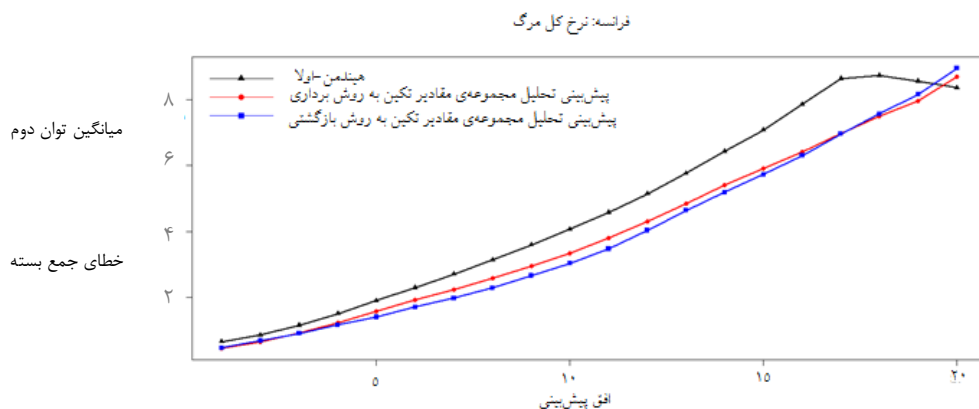
یک دوره مناسب برازش (از نظر طول دوره) را انتخاب کنید. هر یک از دو مدل  $SSA$  و هیندمن-اولا را بر این دوره برازش داده و برای یک افق پیش‌بینی  $h$  ساله، لگاریتم نرخ خام مرگ در هر سن را پیش‌بینی کنید. به آخرین سال دوره برازش فوق یک سال دیگر اضافه کرده و یکبار دیگر، لگاریتم نرخ خام مرگ هر سن خاص را برای همان دوره  $h$  ساله پیش‌بینی کنید. این رویه را تا زمانی که آخرین سال دوره برازش انتخابی یک سال کمتر از آخرین سال داده‌های تحت مطالعه است، ادامه دهید. اکنون، در هر دوره برازش و افق پیش‌بینی، توان دوم خطای

### مقایسه روش تحلیل مجموعه مقادیر تکین در پیش‌بینی نرخ مرگ‌ومیر با روش‌هایی از خانواده لی-کارتر

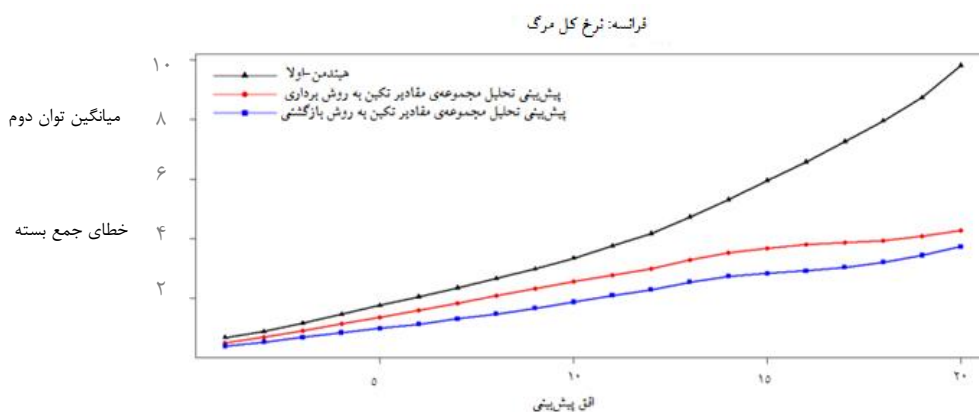
حاصل از پیش‌بینی لگاریتم نرخ خام مرگ در هر سن را با هم جمع کنید. بنابراین، به ازای هر دوره برازش و افق پیش‌بینی، توان دوم خطای جمع‌بسته پیش‌بینی را در اختیار خواهید داشت. در پایان، به منظور داشتن یک معیار کلی از خطای پیش‌بینی در هر افق، میانگین توان دوم خطاهای جمع‌بسته پیش‌بینی در هر افق خاص را محاسبه کنید. بنابراین، به تعداد افق‌های پیش‌بینی، میانگین توان دوم جمع‌بسته خطا در اختیار خواهید داشت.

#### مقایسه دقت عملکرد روش تحلیل مجموعه مقادیر تکین و هیندمن-اولا در دوره‌های برازش مختلف

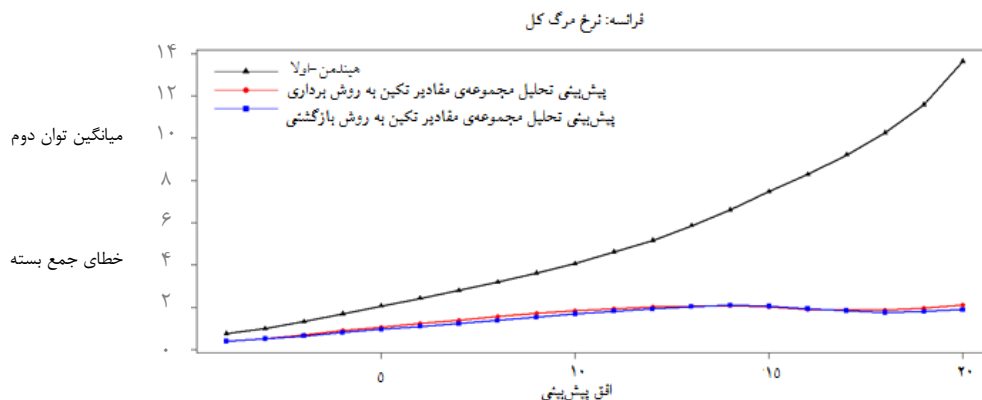
نمودار ۳ نتایج حاصل از مقایسه دقت پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت فرانسه بر اساس معیار مقایسه‌ای زیربخش قبل، دوره برازش  $m=1899, \dots, 2000$  و  $m=1959, \dots, 2000$  را به دو روش هیندمن-اولا و SSA نمایش می‌دهد. به جز افق ۲۰، دقت هر دو روش پیش‌بینی SSA از روش هیندمن-اولا بالاتر است؛ به منظور بررسی بیشتر توانایی SSA، سال پایانی در اولین دوره برازش را به سال‌های ۱۹۶۹ و ۱۹۷۹ افزایش داده و سال پایانی در کل مطالعه را ۲۰۰۹ انتخاب می‌کنیم. الگوی مرگ‌ومیر در دو دوره ۱۹۵۰-۱۹۰۰ و ۱۹۵۰ به بعد متفاوت از یکدیگر است؛ بنابراین، این افزایش منجر به مدل‌بندی پویایی غالب بر داده‌های مرگ‌ومیر در سال‌های اخیر خواهد شد. ازسوی دیگر، از آنجاکه در صورت افزایش سال پایانی در اولین دوره برازش به ۱۹۸۹، تنها برای اولین دوره، خطای توان دوم جمع‌بسته پیش‌بینی برای تمام افق‌ها قابل محاسبه خواهد بود، به منظور ارزیابی این دو روش بر اساس نمونه‌های بیشتر، این افزایش را به سال ۱۹۷۹ محدود می‌کنیم. نمودارهای ۴ و ۵، به ترتیب، نتایج این مقایسه‌ها را نمایش می‌دهند.



نمودار ۳: مقایسه دقت عملکرد روش SSA و هیندمن-اولا در پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت براساس معیار MISE (دوره‌های برازش: 1899- $m=1959, \dots, 2000$ ، دوره‌های پیش‌بینی  $m=1969, \dots, 2000$ ،  $K=3$  و  $m+1, \dots, \min(2001, m+20)$ )



نمودار ۴: مقایسه دقت عملکرد روش SSA و هیندمن-اولا در پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت براساس معیار MISE (دوره‌های برازش: 1899- $m=1969, \dots, 2000$ ، دوره‌های پیش‌بینی  $m=1969, \dots, 2000$ ،  $K=3$  و  $m+1, \dots, \min(2001, m+20)$ )



نمودار ۵: مقایسه دقت عملکرد روش SSA و هیندمن-اولا در پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت براساس معیار MISE (دوره‌های برازش: 1899- $m=1979, \dots, 2000$ , دوره‌های پیش‌بینی  $m=1, \dots, \min(2001, m+20)$  و  $K=3$ )

هر یک از نمودارهای فوق گویای دقت بالاتر روش SSA در مدل‌بندی و پیش‌بینی لگاریتم نرخ خام مرگ کل جمعیت کشور فرانسه هستند. درحقیقت، با استفاده از داده‌های جدیدتر، دقت SSA در مدل‌بندی و پیش‌بینی این کمیت نیز بیشتر می‌شود. یکی از مهم‌ترین دلایل این نتیجه ناشی از ماهیت روش SSA است؛ در این روش، هیچ مدل پارامتری خاصی بر داده‌های برازش داده نشده است و آینده سری زمانی تنها بر اساس ویژگی‌های داده‌های تحت مطالعه پیش‌بینی می‌شود.<sup>۱</sup>

#### نتایج و بحث

بر اساس یافته‌های این پژوهش می‌توان نتیجه گرفت که در بسیاری از موارد، روش SSA از دقت بالاتری نسبت به اعضای تحت مطالعه خانواده لی-کارتر برخوردار است؛ از این رو، با توجه به ماهیت ناپارامتری این روش، در صورت بسط دیگر جنبه‌های نظری آن، می‌توان از آن به عنوان روشی توانمند در مدل‌بندی و پیش‌بینی نرخ مرگ‌ومیر استفاده کرد.

تعیین مقدار بهینه پارامتر طول پنجره با توجه به پویایی موجود در داده‌های مرگ‌ومیر در هر دوره برازش، استفاده از سایر رویکردها در تعیین پارامتر بهینه نقطه برش، استفاده از دیگر ویرایش‌های قدرتمند SSA (همچون SSA بر مبنای برآوردگر کمترین واریانس و نظریه پرشیدگی)، ارائه الگوریتم‌های پیش‌بینی جدیدی در SSA که منجر به استفاده از داده‌های بیشتر و اختصاص وزن‌های بیشتر به داده‌های جدیدتر می‌شود، از جمله پیشنهادهایی است که می‌توان در راستای دستیابی به نتایج دقیق‌تر از آنها بهره جست.

#### منابع و ماخذ

آل حسینی، ف.س.، (۱۳۹۱). مقایسه روش تحلیل مجموعه مقادیر تکین در پیش‌بینی نرخ مرگ‌ومیر با روش‌هایی از خانواده لی-کارتر. پایان‌نامه کارشناسی ارشد، دانشگاه شهید بهشتی، دانشکده علوم ریاضی.  
محمودوند، ر.، (۱۳۹۱). گسترش برخی از مبانی نظری روش تحلیل مجموعه مقادیر تکین. رساله دکتری، دانشگاه شهید بهشتی، دانشکده علوم ریاضی.

Broomhead, D.S.; King, G.P., (1986 a). Extracting qualitative dynamics from experimental data. *Physica D*, 20, pp. 217-36.

Broomhead, D.S.; King, G.P., (1986 b). On the qualitative analysis of experimental dynamical systems. *Nonlinear Phenomena and Chaos*, Adam Hilger, Bristol, 12, pp. 113-44.

<sup>۱</sup> چنین مقایسه‌ای در خصوص عملکرد SSA و هیندمن-اولا در افق‌های پیش‌بینی و سنین مختلف انجام گرفته است. رک: آل حسینی، ۱۳۹۱

- Booth, H.; Tickle, L.; Smith, L., (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: A multi-country comparison. *New Zealand Population Review*, 31(1), pp. 13-43.
- Chen, C.; Liu, L.M., (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88, pp. 284-97.
- Danilov, D., (1997). Principal components in time series forecast. 6(1), pp. 112-21.
- Danilov, D.; Zhigljavsky, A., (1997). Principal components in time series forecast, *Journal of Computational and Graphical Statistics*, 6(1), pp.112-121.
- De Moivre, A., (1725). *Annuities upon lives, or, the valuation of annuities upon any number of lives, as also, of reversions to which is added, an appendix concerning the expectations of life, and probabilities of survivorship*, London: W.P.
- Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A., (2001). *Analysis of time series structure: SSA and related techniques*. C, New York: Chapman Hall/CR.
- Gompertz, B., (1825). On the nature of the fuction expressive of the law of human mortality, and on a new mode if deternining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115, pp. 513-83.
- Hyndman, R.J.; Ullah, M.S., (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10), pp. 4942-56.
- Hössjer, O.; Croux, C., (1995). Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter. *Nonparametric Statistics*, 4, pp.293-308.
- Lee, R.D.; Carter, L., (1992). Modeling and forecasting the time series of US mortality. *Journal of the American Statistics Association*, 87, pp. 659-71.
- Lee, R.D.; Miller, T., (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4), pp. 537-49.
- Makeham, W.M., (1867). On the law of mortality. *Journal of the Institute of Actuaries*, 13(6), pp. 325-58.
- Ramsay, J.O.; Silverman, B.W., (2005). *Functional data analysis*, New York: Springer-Verlag, 2<sup>nd</sup> ed .
- Sang, E., (1868). *Tables for the mutual conversion of solar and sidereal time*. Edinburgh: Blackwood.
- SetarehShenas, S.S., (2011). *Mortality forecasting based on Lee-Carter model*. Allameh Tabatabaie University, E.C.O College Of Insurance, MA.
- Weibull, W., (1939). *The phenomenon of rupture in solids*. Stockholm: Generalstabena Litografiska Anstalts Forlag.
- Wilmoth, J.R., (2002). *Methods protocol for the human mortality database*, <[http://www.mortality.org/Public/ Docs/MethodsProtocol. pdf](http://www.mortality.org/Public/Docs/MethodsProtocol.pdf)>[Accessed 2013/1/16].
- Wood, S.N., (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B*, 65(1), pp.95-114.