نشریه علوم دانشگاه خوارزمی جلد ۱۲، شماره ۱
تاریخ چاپ زمستان ۱۳۹۱

فاصله‌های پیش‌بینی بوت‌استرپ برای احتمال مرگ و میر ایران

*نصراله ایران‌پناه، هدی توسلی: دانشگاه اصفهان، دانشکده علوم، گروه آمار

چکیده

بررسی روند و احتمال مرگ و میر برای برنامه‌ریزان، کارشناسان جمعیت، سازمان‌های بازنشستگی و شرکت‌های بیمه اهمیت زیادی دارند. تحقیقات نشان می‌دهد جدول‌های مرگ و میر ایستا، احتمالات مرگ و میر را زیاد نشان می‌دهند. علت بیش‌برآوردی احتمالات مرگ و میر در جدول‌های ایستا را می‌توان به نادیده گرفتن تغییرات مرگ و میر بر حسب زمان نسبت داد. جدول‌های مرگ و میر پویا با در نظر گرفتن اثر زمان در سنین مختلف بر روند مرگ و میر از دقت بیش‌تری برخوردار هستند. در این مقاله، برای برآورد روند مرگ و میر ابتدا با استفاده از روش لی- کارتر مدلی برحسب زمان و رده‌های سنی به داده‌های نرخ مرگ و میر برازش و پارامترهای مدل را برآورد می‌کنیم. سپس با استفاده از روش‌های بوت استرپ پارامتری و نیم پارامتری فاصله‌های پیش‌بینی را برای احتمال‌های مرگ و میر محاسبه می‌کنیم. در انتها این مراحل را برای تحلیل داده‌های مرگ و میر ایران به‌طور مجزا برای مردان و زنان استفاده می‌کنیم.

مقدمه

در دهه‌های اخیر در اکثر تحقیقات آماری تغییرات مرگ و میر بر حسب زمان مورد توجه قرار گرفته است. این تغییرات برای گروه‌های سنی متفاوت بر حسب زمان و هم‌چنین احتمالات مرگ و میر در دوره‌های زمانی متفاوت برای سنین مشابه یکسان نیست. بررسی این تغییرات برای برنامه‌ریزان، کارشناسان جمعیت، سازمان‌های بازنشستگی و شرکت‌های بیمه از اهمیت خاصی برخوردار است. معمولاً تحلیل داده‌های مرگ و میر با استفاده از روش‌های آمار پارامتری و ناپارامتری براساس جدول‌های عمر ایستا انجام می‌گیرد، ولی این روش‌ها تغییرات مرگ ومیر برحسب زمان را در نظر نمی‌گیرند که باعث بیش برآوردی پیش‌بینی‌ها و هم‌چنین خطا در تحلیل‌ها می‌شود. برای این منظور از جدول‌های عمر پویا در تحلیل‌ها استفاده می‌شود که در این جدول‌های اثر زمان بر روی مرگ و میر بررسی می‌شود.

لی و کارتر [۱۲] برای اولین بار جدول‌های عمر پویا را پیشنهاد و از آن برای بررسی میزان مرگ ومیر آمریکا استفاده کردند. سپس مقالات متعددی برای بررسی روند مرگ ومیر کشورهای توسعه یافته از جمله کانادا (نولت [۱۴])، شیلی (لی و رافمن [۱۳])، ژاپن (ویلمث [۱۸])، بلژیک (براهانس و همکاران [۳])، انگلیس

واژه‌های کلیدی: جدول‌های عمر پویا، مدل لی- کارتر، آمار فضایی، تغییرنگار، بوت‌استرپ پارامتری، بوت‌استرپ نیم پارامتری.
دریافت ۹۰/۱۱/۱۵ پذیرش ۹۱/۸/۳۰

*نویسنده مسئول iranpanah@sci.ui.ac.ir

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

(رنشو وهابرمن [۱۶])، استرالیا (بوس و تیکل [۲])، فرانسه (هایدمن و شهیداله [۱۰]) و اسپانیا (دبن و همکاران [۵]) منتشر شد. یک بازنگری از مدل‌های پویا در مقالات بنجامین و پلارد [۱]، تابیو و همکاران [۱۷] و پیتاکو [۱۵] صورت گرفته است.

دبن و همکاران [۵] روش بوت‌استرپ را با بازنمونه‌گیری مستقل از باقیمانده‌های مدل لی- کارتر (LC) برای محاسبه فاصلهٔ پیش‌بینی احتمال مرگ و میر اسپانیا ارائه کرده است. روش بوت‌استرپ در این مقاله به علت وجود ساختار هم‌بستگی باقیمانده‌های مدل LC به اشتباه استفاده شده است. دبن و همکاران[۶] با در نظر گرفتن ساختار هم‌بستگی فضایی برای باقیمانده‌های مدل LC تحلیل فضایی را برای داده‌های مرگ و میر اسپانیا استفاده کرئده‌اند.

در این مقاله ضمن معرفی مدل LC، آمار فضایی و روش بوت‌استرپ، کاربرد این روش‌ها را برای تحلیل داده‌های مرگ و میر کشور ایران از سال ۱۳۷۵ تا ۱۳۸۹ و گروه‌های سنی مختلف استفاده می‌کنیم. در روش بوت‌استرپ پارامتری با در نظر گرفتن توزیع دوجمله‌ای برای تعداد مرگ و میر در زمان‌ها و گروه‌های سنی مختلف فاصله‌های پیش‌بینی احتمال مرگ و میر معرفی می‌گردد. هم‌چنین در روش بوت‌استرپ نیم پارامتری با در نظر گرفتن ساختار هم‌بستگی فضایی برای باقیمانده‌های مدل LC و حذف این هم‌بستگی فضایی از باقیمانده‌ها به روش تجزیه چولسکی، بازنمونه‌گیری از باقیمانده‌های ناهم‌بسته انجام و سپس فاصله‌های پیش‌بینی احتمال مرگ و میر ارائه می‌گردد.

برای این منظور، ابتدا ضمن معرفی مدل LC، برآورد پارامترهای مدل و پیش‌بینی احتمال مرگ و میر ارائه می‌شود. مقدمه‌ای از آمار فضایی و تعیین ساختار هم‌بستگی فضایی داده‌ها در بخش بعد ارائه می شود. در ادامه ضمن معرفی روش بوت‌استرپ ناپارامتری دو روش بوت‌استرپ پارامتری و نیم پارامتری برای تحلیل داده‌های مرگ و میر ارائه می‌شود. سپس تحلیل داده‌های مرگ و میر ایران بر اساس روش‌های ارائه شده در این مقاله برای تعیین فاصلهٔ پیش‌بینی احتمال مرگ و میر برای زمان‌های آینده در گروه‌های سنی متفاوت به‌طور مجزا برای مردان و زنان ارائه می‌شود. در انتها بحث و نتیجه‌گیری آورده شده است.

مدل لی- کارتر

لی و کارتر [۱۲] مدل LC را معرفی کردند و آن را به داده‌های نرخ مرگ و میر آمریکا برازش دادند. این مدل به‌صورت

$$m_{xt} = exp(a_x + b_x k_t);\ x=1,\ldots,n;\quad t=1,\ldots,T$$

یا به‌طور معادل به‌صورت

$$\mathrm{Ln}(m_{xt}) = a_x + b_x k_t + \varepsilon_{xt}$$

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

است، که در آن‌ها پارامترهای a_x و b_x به سن x وابسته، k_t شاخص مرگ و میر در زمان t و ε_{xt} باقی‌مانده‌های مدل هستند. برای مدل‌بندی احتمال‌های مرگ و میر q_{xt} از یک مدل تعمیم یافته LC (دبن و همکاران [۵]) با استفاده از الگوی لوجیت به‌صورت

$$\log(q_{xt}) = \ln(\frac{q_{xt}}{1-q_{xt}}) = \alpha_x + b_x k_t + \varepsilon_{xt} \qquad (۱)$$

استفاده می‌کنیم. برآورد پارامترهای مدل LC به روش‌های مختلفی در آمار صورت می‌گیرد. در این مقاله روش تجزیهٔ مقدار ویژه SVD ارائه می‌گردد که در آن پارامترها و مقادیر زیر در نظر گرفته می‌شوند:

a_x: میانگین $log_{it}(q_{xt})$ بر روی زمان است و $exp(a_x)$ شکل کلی جدول‌های مرگ و میر را نشان می‌دهد.

b_x: الگوی تغییرات مرگ و میر در سنین مشخص است.

k_t: روند مرگ و میر بر حسب زمان را نشان می‌دهد.

ε_{xt} : باقی‌مانده‌ها یا خطای مدل با میانگین صفر و واریانس

$\sigma^2\varepsilon$ است که اثرات سنین خاص در زمان‌های متفاوت که توسط مدل در نظر گرفته نشده را تعیین می‌کند.

۱. برآورد پارامترهای مدل LC به روش SVD

مراحل مختلف برآورد پارامترهای مدل به روش SVD بدین‌صورت است:

۱. ابتدا a_x را از رابطهٔ

$$\breve{\alpha}_x = \frac{1}{T}\sum_t \log it(q_{xt})$$

برآورد می‌کنیم، که در آن T تعداد کل زمان‌هاست.

۲. مقادیر b_x و k_t از عبارت اول SVD از معادلهٔ

$$logit(q_{xt}) - \breve{\alpha}_x = \sum_{i=1}^{min\ (n,T)} S^i\, U_x^i V_t^i$$

محاسبه می‌شوند، که در این معادله n تعداد گروه‌های سنی و S^i ، U_x^i و V_t^i به‌ترتیب مقادیر منفرد مرتب شده و بردارهای منفرد چپ و راست هستند (رنشاو و هابرمن ۲۰۰۳). بنا بر این $\hat{k}_t^{svd} = S^1 \mathrm{V}_\mathrm{t}^1$ و $\hat{b}_x^{svd} = \mathrm{U}_\mathrm{x}^1$. برای یکتایی جواب‌ها قیود $\sum_x b_x = 1$ و $\sum_t k_t = 0$ پیشنهاد می‌گردد، ولی چون $\sum_x \hat{b}_x^{svd} = c \neq 1$، تبدیلات $\hat{b}_x = \frac{\hat{b}_x^{svd}}{c}$ و $\breve{k}_t = c\hat{k}_t^{svd}$ در نظر گرفته می‌شود.

۳. با برآورد پارامترها در مراحل ۱ و ۲ اختلافی بین احتمال مرگ و میر مشاهده و برآورد شده وجود دارد که علت آن در نظر گرفتن الگوی لوجیت احتمال‌های مرگ و میر است. بنا بر این با استفاده از روش‌های عددی و مقدار اولیه $\tilde{k}_t$ مقدار $\breve{k}_t$ را از رابطهٔ

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

$$D_t = \sum_x \left(E_{xt} \frac{\exp(\check{\alpha}_x + \hat{b}_x k_t)}{1 + \exp(\check{\alpha}_x + \hat{b}_x k_t)} \right); \qquad t = 1, \dots, T$$

برآورد می‌کنیم، که در آن E_{xt} تعداد افراد در معرض مرگ و میر در سن x و زمان t و $D_t = \sum_x D_{xt}$ تعداد کل مرگ‌ها در زمان t است.

۴. با استفاده از تبدیلات $\hat{k}_t = \tilde{k}_t - mean(\tilde{k}_t)$ و $\hat{a}_x = \tilde{a}_x + \hat{b}_x mean(\tilde{k}_t)$ پارامترها را برآورد می‌کنیم.

۲. پیش‌بینی احتمال مرگ و میر q_{xt}

برای پیش‌بینی احتمال مرگ و میر q_{xt} با استفاده از مدل LC، ابتدا شاخص $\hat{k}_t$ را با استفاده از روش‌های برازش مدل (دیگل [۷]) میانگین متحرک جمع بسته اتورگرسیو (*ARIMA*) در سری زمانی مدل‌بندی می‌کنیم. در مدل *ARIMA* بعد از برازش یک مدل *ARIMA*(p,d,q) ابتدا مرتبه‌های اتورگرسیو (p)، تفاضل‌گیری (d) و میانگین متحرک (q) برآورد و سپس پارامترهای مدل به روش باکس- جنکینز برآورد می‌گردند. حال با استفاده از مدل فوق پیش‌بینی $\hat{k}_{t+h}$ را برای زمان‌های آینده محاسبه می‌کنیم. در نهایت پیش‌بینی احتمال مرگ و میر در سن x و در زمان $t+h$ به‌صورت

$$\hat{q}_{x(t+h)} = antilogit(\hat{a}_x + \hat{b}_x \hat{k}_{(t+h)}) \qquad (۲)$$

محاسبه می‌شود.

آمار فضایی

در تحلیل آمار فضایی لازم است یک مدل آماری در نظر گرفته شود. معمولاً یک میدان تصادفی به‌عنوان مدل آماری برای داده‌های فضایی در نظر گرفته می‌شود (کرسی [۴]). میدان تصادفی مجموعه‌ای از متغیرهای تصادفی $\{z(s): s\epsilon D \subseteq R^d; d \geq 1\}$ است که در آن D مجموعه‌ای اندیس‌گذار است. هر میدان تصادفی را می‌توان به‌صورت

$$z(s)=\mu(s) + \delta(s)$$

تجزیه، که در آن $\mu(s)$ تغییرات مقیاس بزرگ یا روند و $\delta(s)$ تغییرات مقیاس کوچک یا فرایند خطای میدان تصادفی است. اگر توزیع توام هر تعداد متناهی از متغیرهای تصادفی یک میدان دارای توزیع نرمال باشند، میدان تصادفی گوسی خواهد بود. اگر میانگین میدان تصادفی ثابت و به موقعیت s بستگی نداشته باشد، یعنی $E[z(s)]=\mu$ و واریانس عبارت $[z(s)-z(s+h)]$ فقط تابعی از فاصلهٔ موقعیت‌ها باشد، یعنی $var[z(s)-z(s+h)]=2\gamma(h)$، آن‌گاه میدان تصادفی را مانای ذاتی می‌نامند. اگر علاوه بر مانایی در میانگین، کوواریانس بین $z(s)$ و $z(s+h)$ فقط تابعی از فاصلهٔ موقعیت‌ها باشد، یعنی $cov[z(s),z(s+h)]=\sigma(h)$ آن‌گاه میدان تصادفی را مانای مرتبهٔ دوم می‌نامند. توابع $2\gamma(h)$ و $\sigma(h)$ که ساختار هم‌بستگی را مشخص می‌کنند، به‌ترتیب تغییرنگار و هم‌تغییرنگار نامیده می‌شوند و رابطهٔ $\gamma(h)= \sigma(\circ)-\sigma(h)$ بین آن‌ها برقرار است. توابع تغییرنگار و هم‌تغییرنگار به‌ترتیب

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

همیشه منفی شرطی و همیشه مثبت هستند. تغییرنگار و هم‌تغییرنگار علاوه بر مقدار h، به جهت آن نیز بستگی دارند. اگر مقادیر $2\gamma(h)$ و $\sigma(h)$ در جهت‌های مختلف یک‌سان و به جهت بستگی نداشته باشند، به‌عبارت دیگر فقط تابعی از اندازهٔ فاصلهٔ h یعنی $|h| = |s_i - s_j|$ باشند، آن‌ها (یا میدان تصادفی) را هم‌سانگرد گویند.

۱- برآورد ساختار هم‌بستگی

در آمار فضایی میزان هم‌بستگی داده‌ها با تغییرنگار یا هم‌تغییرنگار اندازه‌گیری می‌شود. افزایش تغییرنگار بیان‌گر کاهش هم‌بستگی داده‌ها است. در عمل تغییرنگار نامعلوم است و باید براساس داده‌های فضایی برآورد شود. یک برآورد تجربی برای تغییرنگار براساس مشاهدات $\{Z(s_i); i = 1, \dots, n\}$ به‌صورت

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(s_i + h) - Z(s_i)]^2$$

تعریف می‌شود، که در آن $N(h)$ تعداد زوج نقاط متمایزی است که به فاصلهٔ h از یک‌دیگر قرار دارند. $2\hat{\gamma}(h)$ یک برآوردگر ناریب است، ولی مستقیماً نمی‌توان آن را در روش‌های مختلف تحلیل داده‌های فضایی از جمله پیش‌بینی به‌کار برد. برای این منظور لازم است یک مدل به تغییرنگار تجربی مشاهدات در فواصل مختلف محاسبهٔ و نمودار آن رسم می‌شود. سپس با استفاده از روش‌های برازش یک مدل پارامتری مناسب برای تغییرنگار یا هم‌تغییرنگار انتخاب و پارامترهای آن برآورد می‌گردد. مدل‌های مختلفی برای هم‌تغییرنگار از جمله کروی، نمایی، خطی، گوسی، توانی و موجی استفاده می‌شود. به‌عنوان مثال مدل کروی به‌صورت

$$\sigma(h) = \begin{cases} c_\circ + c & h = \circ \\ c_\circ + c\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a}\right) & 0 < h \le a \\ 0 & h \ge a \end{cases} \qquad (۲)$$

است، که در آن $c_\circ$ اثر قطعه‌ای، $c_\circ + c$ آستانه و a دامنه است.

روش‌های باز نمونه‌گیری بوت‌استرپ

افرون [۸] روش بوت‌استرپ را بر اساس ایدهٔ بازنمونه‌گیری از داده‌ها برای برآورد اندازه‌های دقت برآوردگرها مانند اریبی و واریانس، توزیع برآوردگرها و همچنین فاصلهٔ اطمینان و آزمون فرض برای پارامترها ارائه کرد. این روش بر اساس مشاهدات مستقل است و برای داده‌های وابسته مانند سری‌های زمانی، داده‌های فضایی و داده‌های مرگ و میر کاربرد ندارد. در زیر علاوه بر معرفی روش بوت‌استرپ ناپارامتری برای مشاهدات مستقل، دو روش بوت‌استرپ پارامتری و نیم پارامتری را برای تحلیل داده‌های مرگ و میر ارائه می‌شود.

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

۱. روش بوت استرپ ناپارامتری

فرض کنید $Z=\{Z_1,\ldots,Z_n\}$ مجموعه‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع تجمعی F_θ و برآوردگر یا آمارهٔ مورد نظر، $T=t(Z)$ باشد. روش بوت‌استرپ بر اساس ایده بازنمونه‌گیری از داده‌ها برای تعیین مشخصات توزیع نمونه‌ای T بدون فرض معلوم بودن توزیع F است. افرون [۸] الگوریتمی برای برآورد مشخصات توزیع T به‌عنوان برآوردگر پارامتر مورد نظر θ و بر اساس مشاهدات مستقل به‌صورت این مراحل پیشنهاد کرد:

۱. تابع توزیع تجربی $F_n(z)=\frac{1}{n}\sum_{i=1}^{n} I(Z_i \le z)$ را تعیین کنید.

۲. نمونه بوت‌استرپ $\boldsymbol{Z}^*=\{Z_1^*,\ldots,Z_n^*\}$ از F_n به روش نمونه‌گیری تصادفی ساده با جای‌گذاری از $\boldsymbol{Z}$ به‌دست آورید.

۳. برآوردگر یا آماره بوت‌استرپ $T^*=t(\boldsymbol{Z}^*)$ را محاسبه کنید.

٤. برآورد بوت‌استرپ نظری اریبی، واریانس و توزیع T به‌ترتیب به‌صورت

$$Bias_*(T^*)=E_*(T^*)-T;\quad Var_*(T^*)=E_*[T^*-E_*(T^*)]^2;\quad G_*(t)=P_*(T^*\le t)$$

هستند، که در آن‌ها E_*، Var_* و P_* به‌ترتیب امید ریاضی، واریانس و احتمال شرطی بوت‌استرپ به شرط Z است.

۵. اگر $Bias_*(T^*)$، $Var_*(T^*)$ و $G_*(t)$ جواب‌های صریحی نداشته باشند، با استفاده از شبیه‌سازی مونت کارلو و تکرار B بار مراحل ۲ و ۳ و محاسبه $T_1^*,\ldots,T_B^*$، برآورد بوت‌استرپ تجربی اریبی، واریانس و توزیع T به‌ترتیب به‌صورت

$$\widehat{Bias}_B(T^*)=\overline{T^*}-T;\quad \widehat{Var}_B(T^*)=\frac{1}{B}\sum_{i=1}^{B}(T_i^*-\overline{T^*})^2;\quad \hat{G}_B(t)=\frac{1}{B}\sum_{i=1}^{B} I(T_i^*\le t)$$

برآورد می شوند که در آن‌ها $\overline{T^*}=\frac{1}{B}\sum_{i=1}^{B} T_i^*$ است. هم‌چنین فاصلهٔ اطمینان صدکی بوت‌استرپ با ضریب α-1 برای پارامتر θ به‌صورت $\left[T^*_{\left[B.\frac{\alpha}{2}\right]},T^*_{\left[B.(1-\frac{\alpha}{2})\right]}\right]$ ارائه می گردد که در آن $T^*_{[b]}$ آماره مرتب شدهٔ bام، $T_1^*,\ldots,T_B^*$ است (افرون و تیب‌شیرانی [۹]). در اغلب تحلیل‌های نظری بوت‌استرپ برای F_θ و T شرایطی منظور می‌شود، به‌طوری‌که با افزایش n توزیع و گشتاورهای T^* به توزیع و گشتاورهای مجانبی T میل کند.

۲. روش بوت استرپ پارامتری

در روش بوت استرپ پارامتری برای محاسبهٔ فاصلهٔ پیش‌بینی احتمال مرگ و میر q_{xt} به ازای سن x و زمان t، یک توزیع دوجمله‌ای برای تعداد مرگ‌های D_{xt} به‌صورت

$$D_{xt} \sim b(E_{xt}, q_{xt});\quad x=1,\ldots,n;\quad t=1,\ldots,T$$

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

در نظر می‌گیریم. در این رابطه E_{xt} تعداد افراد در معرض مرگ و میر هستند و احتمال مرگ و میر q_{xt} از رابطهٔ

$$q_{xt} = \frac{\frac{1}{2}(d_{xt} + d_{x(t+1)})}{P_{xt} + \frac{1}{2} d_{xt}} \quad (۴)$$

محاسبه می‌شود که در آن P_{xt} و d_{xt} به‌ترتیب تعداد جمعیت و فوت در گروه سنی x و در زمان t است. حال الگوریتم بوت‌استرپ پارامتری را به‌صورت این مراحل زیر معرفی می‌کنیم:

۱. نمونهٔ بوت‌استرپ $D^* = (d^*_{xt})$ را از توزیع دوجمله‌ای $b(E_{xt},q_{xt})$ تولید می‌کنیم. در واقع D^* یک نمونه تصادفی به‌صورت یک ماتریس $n\times T$ است که عنصر ijام آن یک متغیر تصادفی از توزیع دوجمله‌ای $b(E_{ij,}\ q_{ij})$ است.

۲. با استفاده از نمونهٔ بوت‌استرپ D^* که شامل تعداد مرگ‌هاست، ماتریس بوت‌استرپ احتمال‌های مرگ و میر $Q^* = \left(q^*_{xt}\right)$ را با جای‌گذاری d^*_{xt} به‌جای d_{xt} در رابطهٔ (۴) به‌دست می‌آوریم.

۳. با استفاده از احتمال‌های مرگ و میر بوت‌استرپ q^*_{xt} برآورد بوت‌استرپ پارامترهای مدل LC را به‌روش SVD ارائه شده در بخش ۲.۱ به‌صورت (a^*_x, b^*_x, k^*_t) و همچنین پیش‌بینی بوت استرپ k^*_{t+h} و $q^*_{x(t+h)}$ ارائه شده در بخش ۲.۲ را محاسبه می‌کنیم.

۴. با تکرار B بار مراحل ۳-۱ اندازه‌های دقت اریبی و واریانس و توزیع برآوردگرهای پارامترهای مدل LC و همچنین فاصلهٔ پیش‌بینی صدکی بوت‌استرپ k_{t+h} و $d_{x(t+h)}$ را مشابه مرحلهٔ ۵ روش بوت‌استرپ ناپارامتری محاسبه می‌کنیم.

۳. روش بوت استرپ نیم ناپارامتری

ایران پناه و همکاران [۱۱] روش بوت‌استرپ نیم پارامتری را در آمار فضایی معرفی کردند. در این بخش از روش بوت‌استرپ نیم پارامتری برای محاسبهٔ فاصله پیش‌بینی احتمال مرگ و میر q_{xt} استفاده می‌کنیم. در این روش ابتدا باقی‌مانده‌ها را از رابطهٔ

$$\varepsilon_{xt} = logit(q_{xt}) - logit(\hat{q}_{xt})$$

به‌دست می‌آوریم که در آن $\hat{q}_{xt}$ و q_{xt} به‌ترتیب از روابط ۲ و ۴ محاسبهٔ می‌شوند. با رسم نقش نمای باقی‌مانده‌های مدل LC در شکل ۱ برای داده‌های مرگ و میر مردان و زنان ایران و مقایسهٔ آن با باقی‌مانده‌های مستقل، مشاهده می‌گردد که یک هم‌بستگی فضایی بین باقی‌مانده‌های مدل LC وجود دارد. بنا بر این با استفاده از تجزیهٔ چولسکی، هم‌بستگی بین باقی‌مانده‌های مدل LC را حذف و بازنمونه‌گیری از باقی‌مانده‌های ناهم‌بسته را انجام می‌دهیم.

با فرض $N=nT$، گیریم $\widehat{\sum}_{N\times N} = (\hat{\sigma}(h))$ برآورد ساختار هم‌بستگی فضایی بردار باقی‌مانده‌های $\varepsilon =(\varepsilon_1,\ldots,\varepsilon_N)$ براساس برآورد هم‌تغییرنگار پارامتری $\hat{\sigma}(h)$ باشد که یک ماتریس متقارن و معین مثبت است.

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

آن‌گاه $\boldsymbol{e} = (e_1, \dots, e_N) = \hat{L}^{-1}\varepsilon$ یک بردار از باقی‌مانده‌های ناهم‌بسته است، که در آن $\hat{L}$ یک ماتریس پایین مثلثی حاصل از تجزیهٔ چولسکی $\hat{\sum} = \hat{L}\hat{L}^T$ است. حال بردار باقی‌مانده‌های ناهم‌بسته e را به‌صورت $\tilde{e}_i = e_i - N^{-1}\sum_{i=1}^{N} e_i$ مرکزی و سپس روش بوت‌استرپ پارامتری بخش ۴.۲ را روی بردار باقی‌مانده‌های ناهم‌بستهٔ مرکزی $\tilde{\boldsymbol{e}} = (\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_N)$ به‌صورت زیر اجرا می‌کنیم:

۱. نمونهٔ بوت‌استرپ باقی‌مانده‌های $\boldsymbol{e}^* = (e^*{}_1, \dots, e^*{}_N)$ را به‌روش نمونه‌گیری تصادفی ساده باجای‌گذاری از بردار باقی‌مانده‌های ناهم‌بسته مرکزی $\tilde{\boldsymbol{e}}$ تولید و با استفاده از تبدیل عکس، باقی‌مانده‌های $\varepsilon^* = \hat{L}e^*$ را محاسبه می‌کنیم.

۲. بردار بوت‌استرپ $\boldsymbol{\varepsilon}^* = (\varepsilon^*{}_1, \dots, \varepsilon^*{}_N)$ را به‌صورت ماتریس $n\times T$ در نظر گرفته و سپس $logit(q^*{}_{xt}) = logit(q_{xt}) - \varepsilon^*{}_{xt}$، را محاسبه می‌کنیم. حال عکس تابع لوجیت را به‌کار برده و احتمال‌های بوت استرپ مرگ و میر $q^*{}_{xt}$ را به‌دست می‌آوریم.

مراحل ۳ و ۴ مشابه مراحل ۳ و ۴ روش بوت‌استرپ پارامتری بخش ۴.۲ است.



شکل ۱. نقش‌نمای باقی‌مانده‌های مدل LC برای مردان و زنان و باقی‌مانده‌های مستقل.

تحلیل داده‌های مرگ و میر ایران

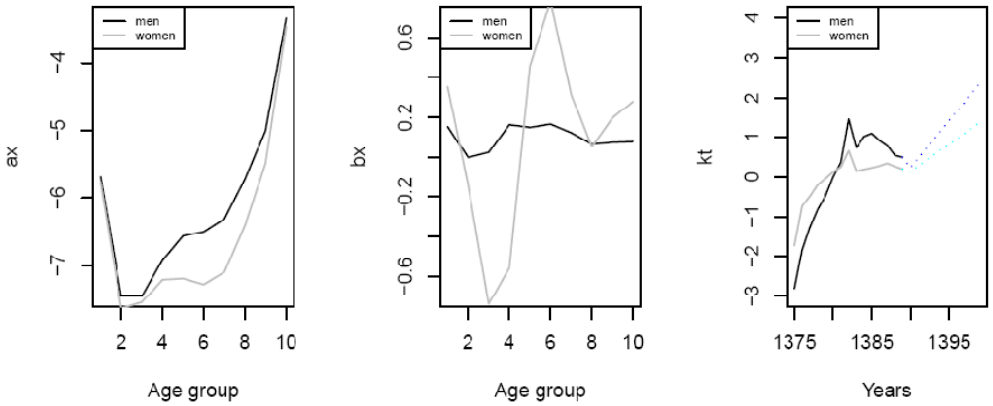
در این بخش، روش‌های ارائه شده در مقاله را برای محاسبهٔ فاصلهٔ پیش‌بینی احتمال مرگ و میر ایران استفاده می‌کنیم. داده‌های استفاده شده به‌صورت جدول‌های تعداد جمعیت و تعداد مرگ و میر به‌ترتیب در ماتریس‌های $P=(P_{xt})$ و $D=(d_{xt})$ برای گروه‌های سنی ۴-۰، ۹-۵، ۱۴-۱۰، ۱۹-۱۵، ۲۴-۲۰، ۲۹-۲۵، ۳۹-۳۰، ۴۹-۴۰، ۵۹-۵۰ و ۶۰ به بالا و زمان‌های۱۳۷۵ تا ۱۳۸۸ به‌صورت سالیانه به‌طور مجزا برای مردان و زنان است. ابعاد ماتریس‌های P و D به‌صورت ۱۰×۱۴ است. برای برآورد پارامترهای مدل LC (بخش ۲.۱) ابتدا احتمال‌های مرگ و میر q_{xt} را از رابطهٔ

$$q_{xt} = \frac{\frac{1}{2}(d_{xt} + d_{x(t+1)})}{P_{xt} + \frac{1}{2}d_{xt}} \qquad x = 1, \dots, 10; \quad t = 1375, \dots, 1388$$

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

محاسبه می‌کنیم. در این رابطه d_{xt} تعداد مرگ‌های گروه سنی x در زمان t، $d_{x(t+1)}$ تعداد مرگ‌های گروه سنی x در زمان $t+1$ و P_{xt} جمعیت گروه سنی x در روز ۳۱ام شهریور ماه سال t است. برآورد پارامترهای a_x، b_x و k_t به روش SVD مدل LC در شکل ۲ نشان داده شده است. در نمودار سوم علاوه بر برآورد $\hat{k}_t$ پیش‌بینی $\hat{k}_{t+h}$ بخش (۲.۲) برای زمان‌های $t=$ ۱۳۸۹ تا $t=$ ۱۳۹۸ محاسبه شده است. این پیش‌بینی‌ها براساس مدل $ARIMA$ (۱،۱،۰) در سری‌های زمانی براساس روش باکس-جنکینز محاسبه شده‌اند.

سپس برای اجرای روش بوت‌استرپ نیم پارامتری (بخش ۴.۳) ابتدا باقی‌مانده‌های مدل LC را به‌صورت

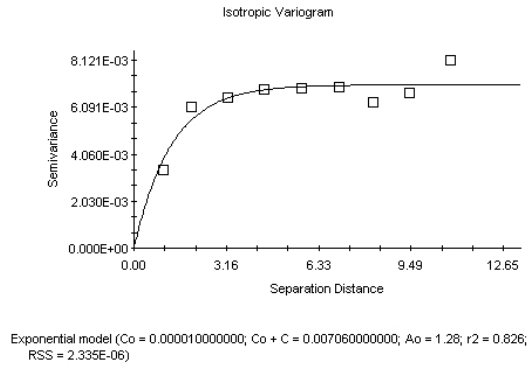


شکل ۲. برآورد پارامترهای مدل LC به‌روش SVD برای مردان و زنان.

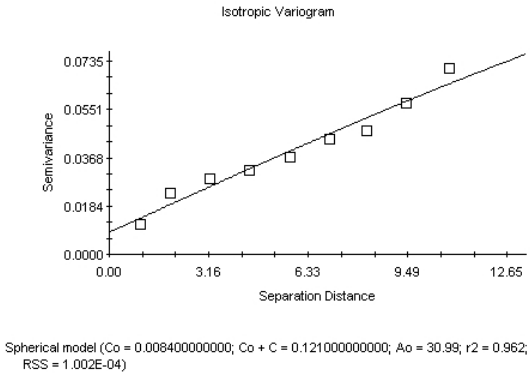
برآورد می‌کنیم. با توجه به توضیحات بخش ۴.۳ و شکل ۱ باقی‌مانده‌های برآورد شده مدل LC برای مردان و زنان دارای ساختار فضایی وابسته است و برآورد این ساختار هم‌بستگی فضایی در تحلیل آمار فضایی مهم است. با توجه به مطالب بخش ۳ برای تعیین هم‌بستگی فضایی ماتریس باقی‌ماندهای برآورد شده مدل LC ابتدا تغییرنگار تجربی باقی‌مانده‌ها را رسم و سپس از توابع تغییرنگار پارامتری برای برازش مدل استفاده و پارامترهای مدل تغییرنگار پارامتری برآورد می‌گردند. برای باقی‌مانده‌های برآورد شده مدل LC داده‌های مرگ و میر ایران، تغییرنگار نمایی (۳) برای داده‌های گروه مردان برازش و پارامترهای این مدل به روش حداقل مربع‌های خطا به‌صورت $\hat{a}=$ ۱/۲۸، $\hat{c}_0=$ ۰/۰۰۰۰۱ و $\hat{c}=$ ۰/۰۰۷۰۵ برآورد می‌گردند. شکل ۳ تغییرنگار تجربی و پارامتری نمایی باقی‌مانده‌های مدل LC برای داده‌های مرگ و میر گروه مردان ایران را نشان می‌دهد. هم‌چنین تغییرنگار کروی (۳) برای داده‌های گروه زنان برازش و پارامترهای این مدل به روش حداقل مربع‌های خطا به‌صورت$\hat{a}=$۳۰/۹۹، $\hat{c}_0=$ ۰/۰۰۸۴ و $\hat{c}=$ ۰/۱۱۲۶ برآورد می‌گردند. شکل ۴ تغییرنگار تجربی و پارامتری کروی باقی‌مانده‌های مدل LC برای داده‌های مرگ و میر گروه زنان ایران را نشان می‌دهد.

حال با اجرای روش بوت‌استرپ نیم پارامتری بخش ۴.۳ فاصلهٔ اطمینان صدکی بوت‌استرپ برای پارامترهای مدل محاسبه می‌گردد. شکل‌های ۵ و ۶ فاصلهٔ اطمینان ۰/۹۵ صدکی بوت‌استرپ نیم پارامتری را برای پارامترهای a_x، b_x و k_t به‌ترتیب برای مردان و زنان نشان می‌دهد.

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]



شکل ۳. تغییرنگار تجربی و پارامتری نمایی باقی‌مانده‌های مدل LC (مردان).

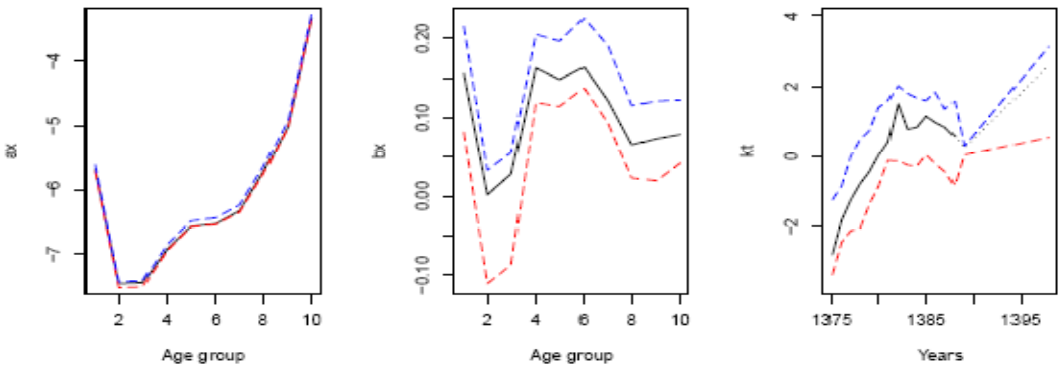


شکل ۴. تغییرنگار تجربی و پارامتری کروی باقی‌مانده‌های مدل LC (زنان)

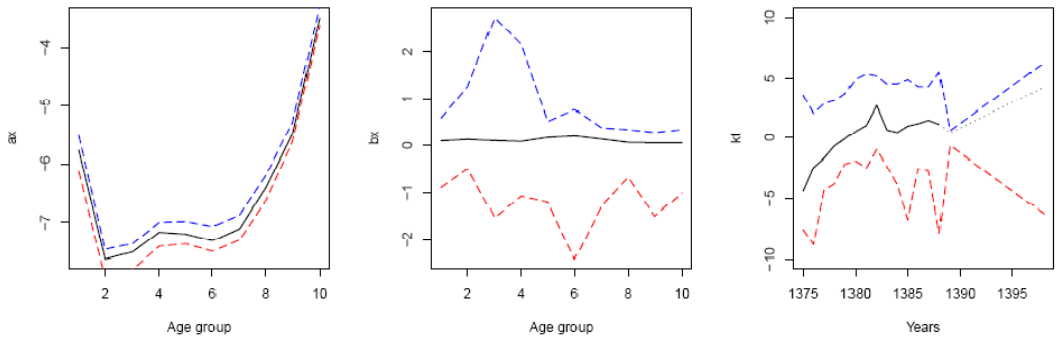
در شکل ۷ و ۸ نیز فاصلهٔ اطمینان ۰/۹۵ صدکی بوت‌استرپ برای پارامترهای مدل LC براساس روش بوت‌استرپ پارامتری بخش ۴.۲ به‌ترتیب برای مردان و زنان نشان داده شده است. در بخش سوم شکل‌های ۵ تا ۸ همچنین فاصله‌های پیش‌بینی بوت‌استرپ ۰/۹۵ نیم پارامتری و پارامتری برای k_t در زمان‌های آینده $t=$ ۱۳۸۹ تا $t=$ ۱۳۹۸ ارائه شده است. قابل ذکر است در هر دو روش بوت‌استرپ پارامتری و نیم‌پارامتری تعداد تکرارهای الگوریتم بوت‌استرپ $B=$ ۱۰۰۰ در نظر گرفته شده است.

با اجرای دو روش بوت‌استرپ پارامتری و نیم پارامتری، فاصلهٔ پیش‌بینی صدکی بوت‌استرپ نیز برای احتمال مرگ و میر q_{xt} به ازای گروه‌های متفاوت سنی در زمان‌های آینده $t=$ ۱۳۸۹ تا $t=$ ۱۳۹۸ ارائه می‌گردد. در شکل‌های ۹ و ۱۰ مقادیر پیش‌بینی و فاصله‌های پیش‌بینی ۰/۹۵ بوت‌استرپ صدکی پارامتری و نیم‌پارامتری برای احتمال مرگ و میر q_{xt} به ازای ۵ گروه سنی متفاوت (۹-۵)، (۲۴-۲۰)، (۲۹-۲۵)، (۴۹-۴۰) و (۵۹-۵۰) برای ۱۰ سال آینده ۱۳۸۹ تا ۱۳۹۸ ارائه شده است.

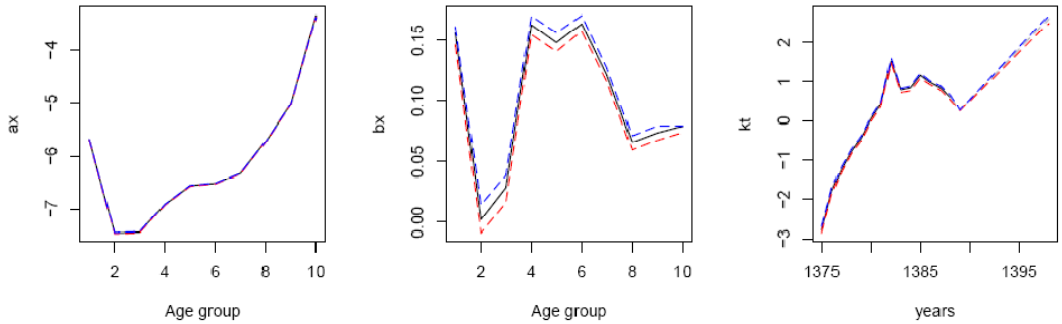
[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]



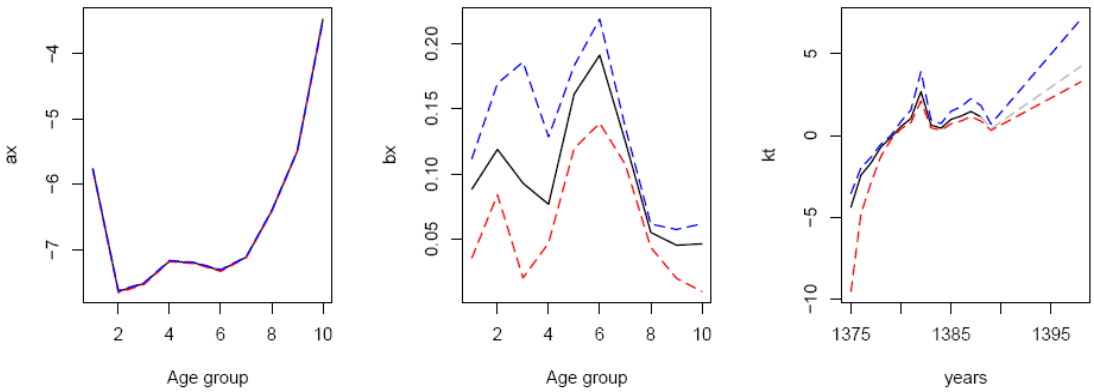
شکل ۵. فاصله‌های اطمینان ۰/۹۵ بوت‌استرپ نیم پارامتری پارامترهای مدل LC (مردان)



شکل ۶. فاصله‌های اطمینان ۰/۹۵ بوت‌استرپ نیم پارامتری پارامترهای مدل LC (زنان)

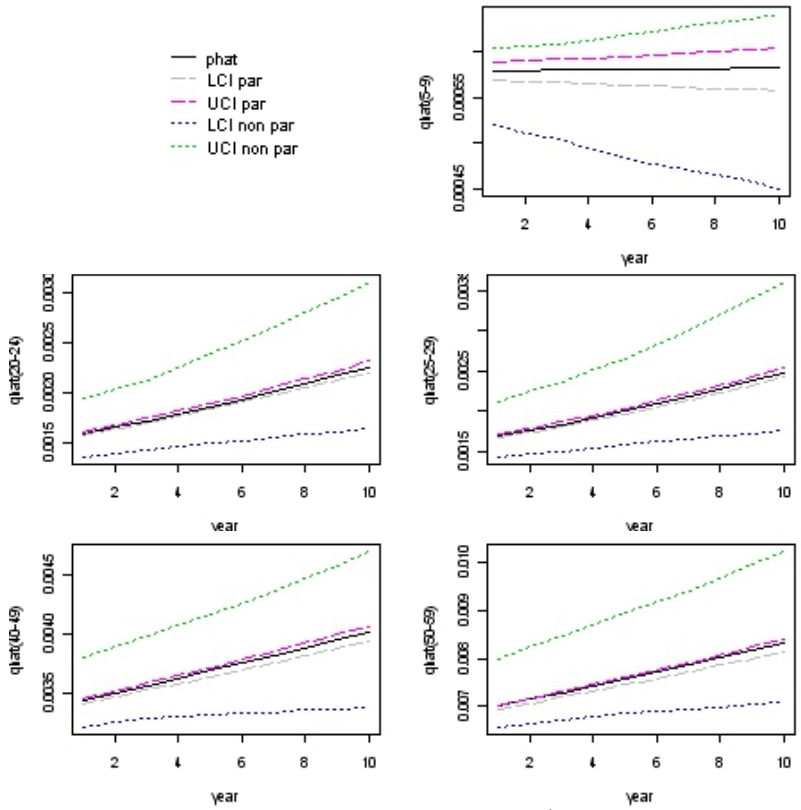


شکل ۷. فاصله‌های اطمینان ۰/۹۵ بوت‌استرپ پارامتری پارامترهای مدل LC (مردان)

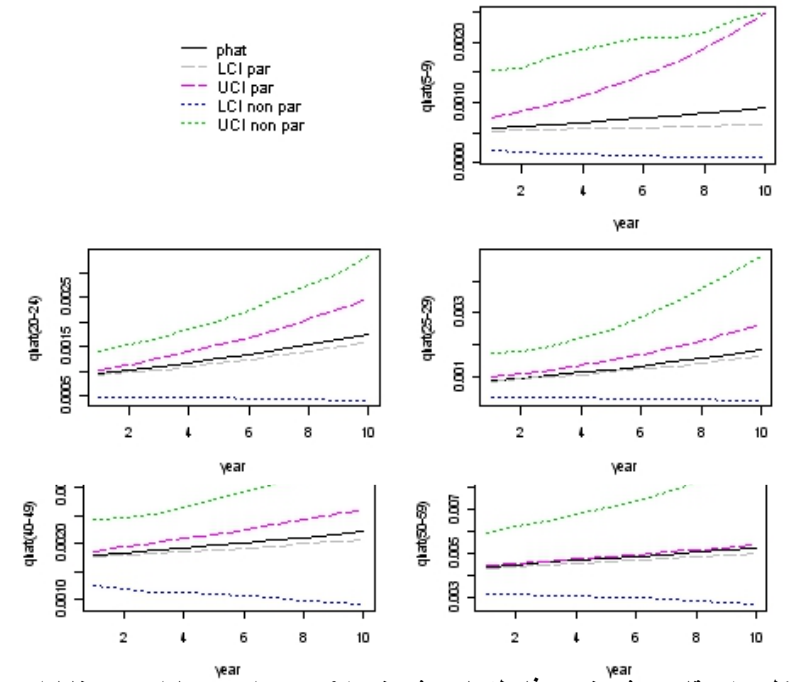


شکل ۸. فاصله‌های اطمینان ۰/۹۵ بوت‌استرپ پارامتری پارامترهای مدل LC (زنان)

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]



شکل ۹. مقادیر پیش‌بینی و فاصله‌های پیش‌بینی ۰/۹۵ بوت استرپ پارامتری و ناپارامتری برای احتمال مرگ و میر q_{xt} در گروه‌های سنی ۲، ۵، ۶، ۸ و ۹ (مردان)



شکل ۱۰. مقادیر پیش‌بینی و فاصله‌های پیش‌بینی ۰/۹۵ بوت‌استرپ پارامتری و ناپارامتری برای احتمال مرگ و میر q_{xt} در گروه‌های سنی ۲، ۵، ۶، ۸ و ۹ (زنان)

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

بحث و نتیجه‌گیری

در این مقاله مدل LC به همراه برآورد پارامترهای مدل به روش SVD در تحلیل احتمال‌های مرگ و میر ارائه شد. با توجه به وجود وابستگی باقی‌مانده‌های مدل LC این ساختار هم‌بستگی در آمار فضایی برآورد و از باقی‌مانده‌ها به روش تجزیهٔ چولسکی حذف می‌گردد. سپس می‌توان روش بوت‌استرپ نیم‌پارامتری را براساس بازنمونه‌گیری از باقی‌مانده‌های ناهم‌بسته اجرا کرد. دو روش بوت‌استرپ پارامتری و نیم‌پارامتری برای احتمال‌های مرگ و میر معرفی شد. فاصلهٔ پیش‌بینی صدکی بوت‌استرپ پارامتری و نیم‌پارامتری در شکل‌های ۹ و ۱۰ برای احتمال‌های مرگ و میر نشان می‌دهد که فاصلهٔ پیش‌بینی صدکی بوت‌استرپ پارامتری به‌علت در نظر گرفتن توزیع مناسب دوجمله‌ای برای تعداد مرگ و میر دارای طول فاصلهٔ کوتاه‌تر و در نتیجه از دقت بیش‌تری نسبت به نیم پارامتری برخوردار است.

مقادیر a_x در شکل ۳، یک افزایش مرگ را از گروه سنی ۱۴-۱۰ به بعد نشان می‌دهد. هم‌چنین نمودار k_t در این شکل، روندی صعودی را در طول زمان نشان می‌دهد که با توجه به مقادیر مثبت b_x این روند بیان‌گر افزایش تقریبی مرگ و میر در طول زمان است. در شکل های ۹ و ۱۰، می‌بینیم مقادیر پیش‌گویی احتمال مرگ و میر در گروه‌های سنی پایین، بسیار کم است ولی با افزایش گروه‌های سنی این احتمال افزایش می‌یابد که این بیان‌گر کاهش مرگ و میر در سال‌های جوانی و افزایش مرگ و میر در سنین پیری است.

منابع

1. B. Benjamin, J. Pollard, "The analysis of mortality and other actuarial Statistics", Sixth ed. Butterworth-Heinemann, London (1992).
2. H. Booth, L. Tickle, "The future aged: New projections of Australia's elderly population", Population Studies 22 (4) (2003) 38-44.
3. N. Brouhns, M. Denuit, J. Vermunt, "A Poisson log bilinear regression approach to the construction of projected life tables", Insurance: Mathematics & Economics 31 (3) (2002) 373-393.
4. N. Cressie, "Statistics for spatial data", John Wiley (1993).
5. A. Debón, F. Montes, F. Puig, "Modeling and forecasting mortality in Spain", Eur. J. Oper. Res. 189 (2008) 624-637.
6. A. Debón, F. Montes. J. Mateuc, E. Porcuc, M. Bevilacquad, "Modeling residuals dependence in dynamic life tables: A geostatistical approach", Computational Statistics and Data Analysis 52 (2008) 3128-3147.

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]

7. P.J. Diggle, "Time series", A Biostatistical introduction, Oxford University Press (1990).

8. B. Efron, "Bootstrap method; another look at the jackknife", Annals of Statistics 7 (1979) 1-26.

9. B. Efron, R. Tibshirani, "Introduction to the bootstrap", Chapman and Hall, London (1993).

10. R. J. Hyndman, Md. Shahid Ullah, "Robust forecasting of mortality and fertility rates: a functional data approach", Computational Statistics and Data Analysis 51(2007) 4942-4956.

11. N. Iranpanah, M. Mohammadzadeh, C. C. Taylor, "A comparison of block and semi-parametric bootstrap methods for variance estimation in spatial statistics", Computational Statistics and Data Analysis 55 (2011) 578-587.

12. R. Lee, L. Carter, "Modeling and forecasting U.S. mortality", J. Amer. Stat. Assoc. 87 (419) (1992) 659-671.

13. R. Lee, R. Rofman, "Modelaion y proyeccion de la mortalidad en Chile", Notas Poblacio´n 22 (59) (1994) 182-213.

14. R.L. F. Nault, "Modeling and forecasting provincial mortality in Canada", Montreal. World Congress of the International Union for Scientific Study of Population.

15. E. Pitacco, "Survival models in dynamic context: A survey", Mathematics & Economics 35 (2) (2004) (1993) 279-298.

16. A. Renshaw, S. Haberman, "Lee-Carter mortality forecasting with age specific enhancement", Insurance: Mathematics & Economics 33 (2) (2003b) 255-272.

17. E. Tabeau, A. van den Berg Jeths, C. (Eds.) Heathcode, "A Review of Demographic Forecasting Models for Mortality", Forecasting in Developed Countries: From description to explanation, Kluwer Academic Publishers (2001).

18. J. Wilmoth, "Health and Mortality among Elderly Populations Chapter: Mortality projections for Japan: A Comparison of Four Methods", Oxford University Press (1996) 266-287.

[Downloaded from c4i2016.khu.ac.ir on 2024-03-28]