مدل‌های نیمه‌پارامتری برای داده‌های بازگشتی بقا با ناهمواری‌های متفاوت برای ضرایب وابسته به زمان

احسان اسحقی — حسین باغیشنی — داود شاهسونی

دانشگاه شاهرود

چکیده: در برخی از مدل‌های نیمه‌پارامتری بقا که برای مدل‌سازی داده‌های بازگشتی به کار می‌روند، ضرایب متغیرهای موجود در مدل وابسته به زمان هستند. در این مدل‌ها، برآوردگرهای حاصل از روش‌های مختلف معمولاً، به صورت بسته و دقیق به‌دست نمی‌آیند و تکرار معادله‌های برآوردیاب، به ازای هر نقطه زمانی، برآوردهایی ناهموار برای ضرایب وابسته به زمان ایجاد می‌کند. هموارساز هسته می‌تواند راه حلی برای رفع این ناهمواری‌ها باشد. رهیافتی سنتی برای هموار کردن ضرایب با استفاده از این روش، انتخاب پهنای نوارهای یکسان برای تمام ضرایب است. ناهمواری متفاوت ضرایب، کارآمدی چنین رهیافتی را کاهش می‌دهد. ما در این مقاله، با انتخاب پهنای نوارهای متفاوت برای ضرایب متفاوت، به برازش مدل می‌پردازیم. همچنین ویژگی‌های مجانبی برآوردگرهای حاصل، از جمله سازگاری، توزیع مجانبی و نرخ همگرایی، را بررسی می‌کنیم. در پایان نیز با یک مطالعه شبیه‌سازی، عملکرد روش پیشنهادی و ویژگی‌های نظری مطرح‌شده را، در کوچک‌نمونه، ارزیابی می‌کنیم.

واژه‌های کلیدی: پهنای نوارهای متفاوت، تابع هسته، داده‌های بازگشتی، سازگاری، مدل نیمه‌پارامتری بقا، نرخ همگرایی.

۱ مقدمه

در بسیاری از علوم مانند پزشکی، مهندسی و محیط زیست، گاهی اوقات پیشامد مورد نظر ممکن است بیش از یک بار رخ دهد. این قبیل پیشامدها را پیشامدهای بازگشتی [۱] و به داده‌های حاصل از آن‌ها، داده‌های بازگشتی می‌گویند. برای مثال، بیماران مبتلا به فیبروز کیستی [۲]، اغلب از تشدید علایم تنفسی، که تکرارشونده هستند، رنج می‌برند و بیماران ایدزی ممکن است عفونت‌های متوالی را تجربه کنند. انفارکتوس قلبی و عود بیماری دیگر مثال‌های پیشامدهای بازگشتی هستند. در داده‌های بازگشتی، معمولاً نمی‌توان پیشامدهای متفاوت برای یک فرد را مستقل از هم در نظر گرفت. اخیراً تحلیل این داده‌ها، موضوعی برای گسترش مطالعات روان‌شناختی شده است و محققان، اغلب، علاقه‌مند به ارزیابی اثرات متغیرها روی ویژگی‌های این نوع داده‌ها هستند.

از جمله رهیافت‌های مدل‌بندی این‌گونه داده‌ها، استفاده از مدل‌های رگرسیونی نیمه‌پارامتری است. در اغلب این مدل‌ها، فرض بر این است که ضرایب رگرسیونی در طول زمان ثابت هستند، در حالی که در واقعیت، ممکن است پارامترهای رگرسیونی وابسته به زمان باشند. بنابراین شناخت و سنجش اثرات موقتی متغیرها در زمان‌های رخداد پیشامدهای بازگشتی برای محققان مهم است. در نتیجه، در نظر گرفتن مدل‌هایی با ترکیب متغیرهای تبیینی وابسته و مستقل از زمان، از اهمیت بالایی برخوردار است. مدل‌های رگرسیونی نیمه‌پارامتری مختلفی برای تحلیل داده‌های بازگشتی در خانواده‌های مدل‌های شرطی و حاشیه‌ای پیشنهاد شده‌اند. مدل‌های رگرسیونی تابع مخاطره شرطی به طور گسترده در تحلیل بقا برای توصیف وابستگی زمان‌های بقا به متغیرهای تبیینی مورد استفاده قرار می‌گیرند. مدل مخاطره نسبی کاکس [۳]، یکی از معروف‌ترین این مدل‌هاست. در عمل، مدل‌بندی میانگین تعداد پیشامدها در مقابل مدل مخاطره نسبی کاکس، برای این نوع داده‌ها، قابل فهم‌تر است (سان و همکاران، ۲۰۱۱). برخی از محققان، از مدل‌های رگرسیونی برای مدل‌بندی تابع میانگین و تابع نرخ استفاده کرده‌اند. لین و همکاران (۲۰۰۰)، مدلی را برای میانگین و نرخ حاشیه‌ای بر اساس یک تابع پیوند از نوع کاکس معرفی کردند و استنباط‌ها را با فرض پیوستگی زمان، تعمیم دادند.

---
[۱] Recurrent event
[۲] Cystic fibrosis
[۳] Cox proportional hazard model

مارتینسون و همکاران (۲۰۰۲)، مطالعاتی را برای برآورد ضرایب وابسته و مستقل از زمان در مدل‌های کاکس انجام دادند. اخیرا نیز سان و همکاران (۲۰۱۱)، مدل رگرسیونی حاشیه‌ای را با ترکیبی از ضرایب وابسته و مستقل از زمان برای داده‌های بازگشتی مورد استفاده قرار داده‌اند. سان و همکاران (۲۰۱۱) برای ضرایب وابسته به زمان متفاوت، پارامترهای پهنای نوار[۴] یکسانی را در نظر گرفتند در حالی که ضرایب متفاوت، ممکن است ناهمواری‌های متفاوتی داشته باشند. از این رو، با در نظر گرفتن پهنای نوارهای متفاوت، می‌توان برآوردهای دقیق‌تر و در نتیجه مدلی کاراتر داشت.

بر روی این مدل‌ها با پهنای نوارهای متفاوت، مطالعات زیادی انجام نشده‌اند. فان و ژانگ (۱۹۹۹) یک روش برآورد دو مرحله‌ای را پیشنهاد دادند که از پهنای نوارهای نابرابر برای برآورد ضرایب رگرسیونی وابسته به زمان استفاده می‌کند. البته رهیافت آن‌ها تنها برای مدل‌های خطی ارایه شد و برای p ضریب رگرسیونی تنها از دو پهنای نوار مختلف، یکی برای $p-1$ ضریب و دیگری برای یک ضریب، استفاده کردند.

در این مقاله، روش برآورد سان و همکاران (۲۰۱۱) را تعمیم داده و برای هموار کردن ضرایب وابسته به زمان، از پهنای نوارهای متفاوت استفاده می‌کنیم. این رهیافت جدید و در خور توجه، محدودیت‌های روش فان و ژانگ (۱۹۹۹) را ندارد و برای هر ضریب، امکان در نظر گرفتن پهنای نوار متفاوت وجود دارد. تعاریف، شرایط نظم و نمادها مشابه موارد متناظر در سان و همکاران (۲۰۱۱) هستند.

در ادامه، در بخش ۲ به معرفی مدل مورد نظر می‌پردازیم و با استفاده از روش هسته[۵] و الگوریتم نیوتون—رافسون، رهیافتی برای برازش مدل ارایه می‌کنیم. سپس ویژگی‌های مجانبی برآوردگرهای به‌دست آمده را در بخش ۳ مورد بررسی قرار می‌دهیم. در بخش ۴، با اجرای یک مطالعه شبیه‌سازی، عملکرد روش پیشنهادی را مورد ارزیابی قرار می‌دهیم و با حالتی که پهنای نوارها یکسان باشند، مقایسه می‌کنیم. در پایان، بحث و نتیجه‌گیری مقاله را مطرح می‌کنیم.

۲ مدل نیمه‌پارامتری بقا

برای ساخت مدل، n نفر را در طول زمان مورد بررسی قرار می‌دهیم. فرض کنید $N_i^*(t)$ تعداد پیشامدهایی باشد که در طول بازه $[0,t]$ برای فرد iاُم رخ می‌دهد. همچنین فرض کنید $X_i(\cdot)$ و $Z_i(\cdot)$ بردارهای فرآیندهایی به ترتیب p و q بعدی برای فرد iاُم باشند. در اکثر کاربردها، بازه زمانی مورد بررسی محدود است و ممکن است برخی از پیشامدها سانسور شوند. از این رو، $N_i^*(t)$ به طور کامل مشاهده نمی‌شود. اگر C_i را به عنوان زمان سانسور تعریف کنیم، آن‌گاه فرآیند قابل مشاهده به صورت $N_i(t) = N_i^*(t \wedge C_i)$ تعریف می‌شود، به طوری که $a \wedge b = \min(a,b)$. فرض می‌کنیم با شرط داشتن $X_i(\cdot)$ و $Z_i(\cdot)$، کمیت‌های C_i و $N_i^*(\cdot)$ مستقل از هم می‌باشند. تعریف می‌کنیم $Y_i(t) = I(C_i \geq t)$، به طوری که $I(\cdot)$ تابع نشان‌گر است. در نتیجه، مجموعه داده مشاهده‌شده به صورت $\{N_i(\cdot), Y_i(\cdot), X_i(\cdot), Z_i(\cdot)\}, i = 1,\ldots,n$ می‌باشد.

مدل رگرسیونی نرخ حاشیه‌ای با ترکیبی از ضرایب وابسته به و مستقل از زمان، به صورت زیر تعریف می‌شود:

$$E\{dN_i^*(t)|X_i(t), Z_i(t)\} = \exp\{\beta_0(t)^T X_i(t) + \gamma_0^T Z_i(t)\} d\mu_0(t), \qquad (1)$$

که در آن $\beta_0(t)$ بردار p بعدی ضرایب رگرسیونی وابسته به زمان، γ_0 بردار q بعدی ضرایب رگرسیونی مستقل از زمان و $\mu_0(t)$ تابع میانگین پایه‌ای است. فرض کنید

$$N(t) = (N_1(t),\ldots,N_n(t))^T, \quad X(t) = (X_1(t),\ldots,X_n(t))^T,$$
$$Z(t) = (Z_1(t),\ldots,Z_n(t))^T,$$

و $N_0(t) = n^{-1}\sum_{i=1}^n N_i(t)$. تعریف می‌کنیم

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\exp\{\beta_0(s)^T X_i(s) + \gamma_0^T Z_i(s)\} d\mu_0(s), \; i = 1,\ldots,n.$$

[۴] Bandwidth
[۵] Kernel method

تحت مدل (؟؟)، $M_i(t)$ها فرآیندهایی با میانگین صفر هستند. صفر بودن میانگین $M_i(t)$ها با توجه به تعریف $Y_i(t)$ و مدل (۱)، به سادگی قابل درک است. بنابراین به ازای مقادیر $\beta(t)$ و γ، برآوردگری منطقی برای $\mu_\circ(t)$ از حل عبارت

$$\sum_{i=1}^{n}[dN_i(t) - Y_i(t)\exp\{\beta_\circ(t)^T X_i(t) + \gamma_\circ^T Z_i(t)\}d\mu_\circ(t)] = \circ,\ \circ \le t \le \tau,$$

به‌دست می‌آید که در آن τ مقدار تعیین‌شده‌ای است که $P(C_i \ge \tau) > \circ$. در نتیجه برآوردگر $\mu_\circ(t)$ به صورت

$$\hat{\mu}_\circ(t;\beta,\gamma) = \int_\circ^t S_\circ(u;\beta,\gamma)^{-1} dN_\circ(u), \tag{۲}$$

محاسبه می‌شود.

۱.۲ برازش مدل

برای برآورد $\beta_\circ(t)$ و $\gamma_\circ$، با استفاده از معادله‌های برآوردیاب تعمیم‌یافته[۶](GEE) لیانگ و زگر (۱۹۸۶) و رابطه (۲)، معادله‌های برآوردیاب زیر را داریم:

$$\begin{aligned} & X(t)^T[dN(t) - \phi(t;\beta,\gamma)S_\circ(u;\beta,\gamma)^{-1}dN_\circ(u)] = \circ, \qquad \circ \le t \le \tau, \\ & \int_\circ^\tau Z(t)^T[dN(t) - \phi(t;\beta,\gamma)S_\circ(u;\beta,\gamma)^{-1}dN_\circ(u)] = \circ. \end{aligned} \tag{۳}$$

معادله‌های (۳) را با به کار بردن روش نیوتون—رافسون، به وسیله بسط تیلور تابع $\phi(t;\beta,\gamma)S_\circ(t;\beta,\gamma)^{-1}$ حول برآورد جاری $(\beta^l(t),\gamma^l)$، برای برآورد $\gamma_\circ$ و $\beta_\circ(t)$ استفاده می‌کنیم. معادله‌های به‌روزکننده به صورت زیر تعریف می‌شوند:

$$\begin{aligned} & \{\beta^{l+1}(t) - \beta^l(t)\}S_\circ^l(t)^{-1}dN_\circ(t) \\ & \quad = n^{-1}E_{xx}^l(t)^{-1}\{X(t) - \bar{X}^l(t)\}^T[dN(t) - \Phi^l(t)\{Z(t) - \bar{Z}^l(t)\}(\gamma^{l+1} - \gamma^l) \\ & \quad \times S_\circ^l(t)^{-1}dN_\circ(t)], \end{aligned} \tag{۴}$$

$$\begin{aligned} & n^{-1}\int_\circ^\tau \{Z(t) - \bar{Z}^l(t)\}^T dN(t) - \int_\circ^\tau E_{zx}^l(t)\{\beta^{l+1}(t) - \beta^l(t)\}S_\circ^l(t)^{-1}dN_\circ(t) \\ & \quad = \int_\circ^\tau E_{zz}^l(t)(\gamma^{l+1} - \gamma^l)S_\circ^l(t)^{-1}dN_\circ(t). \end{aligned} \tag{۵}$$

با قرار دادن رابطه (۴) در (۵) و حل معادله بر اساس γ^{l+1}، به رابطه تکراری $\gamma^{l+1} = \Psi_r(\gamma^l)$ دست می‌یابیم، به طوری که

$$\begin{aligned} \Psi_r(\gamma^l) = \gamma^l + \frac{A^l(\tau)^{-1}}{n}\int_\circ^\tau & [\{Z(t) - \bar{Z}^l(t)\}^T - E_{zx}^l(t)E_{xx}^l(t)^{-1} \\ & \times \{X(t) - \bar{X}^l(t)\}^T]dN(t), \end{aligned} \tag{۶}$$

که در آن $A^l(\tau) = A(\tau;\beta^l,\gamma^l)$ و

$$A(\tau;\beta,\gamma) = \int_\circ^\tau [E_{zz}(t) - E_{zx}(t)E_{xx}(t)^{-1}E_{zx}(t)^T]\frac{dN_\circ(t)}{S_\circ(t)}.$$

---

۶ Generalized Estimating Equation

برای برآورد $\beta_0(t)$ با استفاده از معادله (۴)، تکرار معادله به ازای هر t باعث ناهمواری منحنی $\beta(t)$ می‌شود و حتی ممکن است موجب عدم سازگاری برآوردگر شود. بنابراین از هموارساز هسته استفاده می‌شود. مارتینسون و همکاران (۲۰۰۲) و شیکه و مارتینسون (۲۰۰۴)، پیشنهاد کردند به جای استفاده از $\beta_0(t)$، از ضرایب رگرسیونی تجمعی $B_0(t) = \int_0^t \beta_0(s)ds$ استفاده کنیم که منجر به نتایج پایدارتری می‌شود.

فرض کنید $\mu_0^l(t) = \hat{\mu}_0(t; \beta^l, \gamma^l)$. در این مقاله روش متداول، یعنی در نظر گرفتن پهنای نوارهای یکسان، را تعمیم داده و برای ضرایب متفاوت، پهنای نوارهای متفاوتی در نظر می‌گیریم. برای این منظور، ماتریسی قطری از پهنای نوارها، H، را در نظر می‌گیریم. بنابراین فرض کنید $\beta^l(t)$ و $\lambda_0^l(t)$ به ترتیب برآوردگرهای هسته $\beta_0(t)$ و $\lambda_0(t) = \frac{d}{dt}\mu_0(t)$ بر پایه $B^l(t)$ و $\mu_0^l(t)$ بر مبنای ماتریس پهنای نوار H باشند. یعنی

$$\lambda_0^l(t) = \int H^{-1} K d\mu_0^l(u), \quad \beta^l(t) = \int H^{-1} K dB^l(u),$$

که در آن K و H به صورت

$$K = \begin{pmatrix} K(\frac{u-t}{h_1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K(\frac{u-t}{h_p}) \end{pmatrix}, \quad H = \begin{pmatrix} h_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_p \end{pmatrix},$$

تعریف می‌شوند به طوری که، توابع هسته، $K(\cdot)$، متقارن با دامنه فشرده هستند. با استفاده از γ^{l+1} و با قرار دادن $\frac{dN_0(t)}{S_0^l(t)}$ در رابطه (۴) به جای $\lambda_0^l(t)dt$، به رابطه تکراری $B^{l+1} = \Psi_b(B^l)(t)$ دست می‌یابیم، به طوری که

$$\Psi_b(B^l)(t) = \int_0^t \beta^l(u)du + n^{-1}\int_0^t \lambda_0^l(u)^{-1} E_{xx}^l(u)^{-1}\{X(u) - \bar{X}^l(u)\}^T$$
$$\times [dN(u) - \Phi^l(u)\{Z(u) - \bar{Z}^l(u)\}(\gamma^{l+1} - \gamma^l)S_0^l(u)^{-1}dN_0(u)]. \quad (۷)$$

با استفاده متوالی از معادله‌های تکراری (٦) و (۷)، برآوردگرهای $B_0(t)$ و γ_0 به‌روز می‌شوند. مراحل تا جایی ادامه می‌یابند که اختلاف بین برآوردها در دو مرحله متوالی، کمتر از یک حد آستانه مشخص کوچک، مثلا 10^{-6}، شود.

۳ ویژگی‌های مجانبی برآوردگرها

در این بخش، سازگاری، توزیع مجانبی و نرخ همگرایی برآوردگرهای نهایی به‌دست آمده از معادله‌های تکراری (٦) و (۷)، $\hat{\gamma}$ و $\hat{B}(\cdot)$، را بررسی می‌کنیم. فرض کنید $\|v\| = (v^T v)^{\frac{1}{2}}$، نرم اقلیدسی برای بردار v باشد.

قضیه ۱ تحت شرایط نظم (C1) تا (C6) در سان و همکاران (۲۰۱۱)، معادله $\gamma = \Psi_r(\gamma)$ دارای جواب $\hat{\gamma}$ است، به‌طوری که $\|\hat{\gamma} - \gamma_0\| = O_p(n^{-\frac{1}{2}})$ و $n^{\frac{1}{2}}(\hat{\gamma} - \gamma_0)$ به‌طور مجانبی دارای توزیع نرمال با میانگین صفر و ماتریس کوواریانسی است که به‌وسیله $\hat{\Sigma} = n^{-1}\sum_{i=1}^n \hat{\xi}_i(\tau)\hat{\xi}_i(\tau)^T$ برآورد می‌شود، به‌طوری که

$$\xi_i(\tau) = a(\tau)^{-1}\int_0^\tau [\{Z_i(t) - e_z(t)\}^T - e_{zx}(t)e_{xx}(t)^{-1}\{X_i(t) - e_x(t)\}^T]dM_i(t).$$

قضیه ۲ تحت شرایط نظم (C۱) تا (C۶) در سان و همکاران (۲۰۱۱)، معادله $\Psi_b(B)(t) = B(t)$ دارای جواب $\hat{B}(t)$ است، به‌طوری که $\sup_{0 \le t \le \tau} \|\hat{B}(t) - B_0(t)\| = O_p(n^{-\frac{1}{2}})$ و $n^{\frac{1}{2}}(\hat{B}(t) - B_0(t))$ به‌طور مجانبی دارای توزیع نرمال با میانگین صفر و تابع کوواریانسی در (s,t) است که به‌وسیله $\hat{\Gamma}_b(s,t) = n^{-1}\sum_{i=1}^{n} \hat{\eta}_i(s)\hat{\eta}_i(t)^T$ برآورد می‌شود، به طوری که

$$\eta_i(t) = \int_0^t \lambda_0(u)^{-1} e_{xx}(u)^{-1}\{X_i(u) - e_x(u)\}^T dM_i(u) - \int_0^t e_{xx}(u)^{-1} e_{zx}(u)^T du\, \xi_i(\tau).$$

۴ مطالعه شبیه‌سازی

برای مطالعه شبیه‌سازی، مدلی با دو ضریب وابسته به زمان در نظر گرفتیم و ضرایب را با استفاده از دو روش پیشنهادی ما و سان و همکاران (۲۰۱۱)، برآورد کردیم. در این مدل، به ازای $i = 1, \ldots, n$، زمان‌های رخداد از یک فرآیند پواسن با مدل نرخ حاشیه‌ای

$$E\{dN_i^*(t)|X_i(t), Z_i(t)\} = \exp\{-0/5 + 0/5\cos(2t - 1.75)X_i(t) + 0/5 Z_i(t)\}dt, \tag{8}$$

تولید شدند، به طوری که $X_i \sim N(0,1)$ و $Z_i \sim ber(0/5)$. زمان‌های سانسور C_i از توزیع یکنواخت $U(2,5)$ تولید شدند به‌طوری که متوسط پیشامدهای بازگشتی برای هر فرد تقریبا ۳ به دست آمد. برای انجام شبیه‌سازی، حجم نمونه $n = 200$ را در نظر گرفتیم و با استفاده از چندین پهنای نوار، همواری‌های متفاوتی را بررسی کردیم. جدول ۱ نتایج شبیه‌سازی برای ضرایب $B(\cdot)$ به ازای پهنای نوارهای یکسان و متفاوت و هسته اپانچنیکوف[7] است. در این

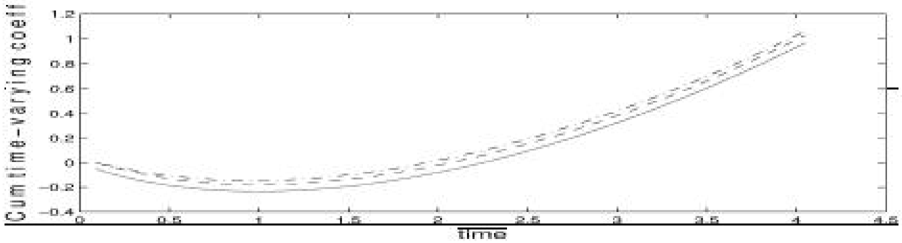
جدول ۱: نتایج شبیه‌سازی برای برآورد ضرایب رگرسیونی تجمعی وابسته به زمان در مدل (۸) با دو روش پهنای نوارهای یکسان و متفاوت

$MISE_2$	$MISE_1$	ISB_2	ISB_1	h_2	h_1	n
۰/۱۷۷۵	۰/۱۴۰۷	۰/۰۱۳۲	۰/۰۰۳۳	۰/۲	۰/۵	۲۰۰
۰/۱۷۷۵	۰/۱۳۹۴	۰/۰۳۲۰	۰/۰۰۳۱	۰/۵	۰/۵	

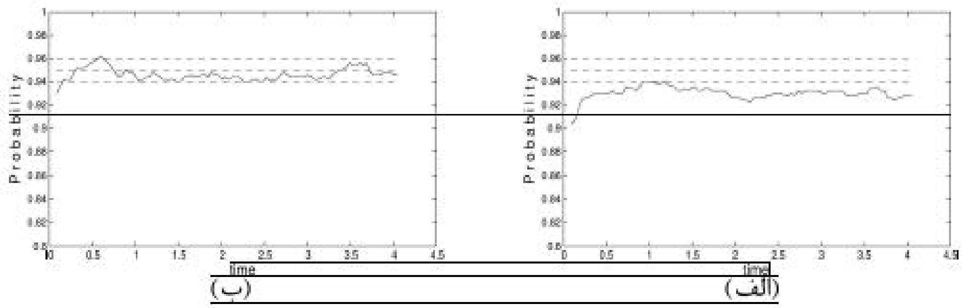
جدول، ISB_1 و $MISE_1$ به ترتیب مقادیر اریبی جمع‌بسته توان دوم و میانگین جمع‌بسته توان دوم خطاها برای ضریب $B_1(t) = \int_0^t 0/5\cos(2u - 1/75)du$ و ISB_2 و $MISE_2$، مقادیر مذکور برای ضریب $B_2(t) = \int_0^t 0/7(\sqrt{u} - 1)du$ هستند. همانطور که در جدول مشاهده می‌شود، با کاهش پهنای نوار مربوط به ضریب $B_2(\cdot)$، اریبی کوچکتر و در نتیجه برآورد دقیق‌تر می‌شود؛ ضمن آن‌که خطا نیز تغییر نکرده است. این نتایج، برتری رهیافت جدید پیشنهادی ما را بر رهیافت سنتی، که پهنای نوارها را یکسان در نظر می‌گیرد، نشان می‌دهد.

با توجه به شکل ۱، برآورد با پهنای نوارهای متفاوت به منحنی واقعی نزدیک‌تر و اریبی کمتر است. در نمودار ۲–(ب) نیز استفاده از پهنای نوارهای متفاوت، باعث قرار گرفتن نرخ‌های پوشش در بازه اطمینان شدند، در حالی که با استفاده از پهنای نوارهای یکسان، نمودار ۲–(الف)، این مقادیر خارج از بازه اطمینان قرار گرفتند.

[7] Epanechnikov



شکل ۱: نمودارهای ضرایب رگرسیونی تجمعی (منحنی توپر) و برآورد آن‌ها با پهنای نوارهای یکسان (منحنی نقطه خط‌چین) و پهنای نوارهای متفاوت (منحنی خط‌چین) برای $B_۲(\cdot)$



شکل ۲: مقادیر احتمال پوشش برای ضریب $B_۲(\cdot)$ با روش (الف) پهنای نوارهای یکسان (ب) پهنای نوارهای متفاوت

بحث و نتیجه‌گیری

مدل‌های نیمه‌پارامتری بقا با ترکیبی از متغیرهای تبیینی وابسته و مستقل از زمان، دارای انعطاف بالایی برای به‌کارگیری در موقعیت‌های عملی مختلف هستند. در مدل نیمه‌پارامتری انتخاب‌شده، با استفاده از روش عددی نیوتون—رافسون و هموارسازی هسته، رهیافتی منطقی و مناسب برای برازش مدل، بیان کردیم. با استفاده از نتایج به‌دست آمده نشان دادیم استفاده از پهنای نوارهای متفاوت، برآوردهای دقیق‌تری نسبت به استفاده از پهنای نوارهای یکسان نتیجه می‌دهد.

از نقطه‌نظر کاربردی، در نظر گرفتن پهنای نوارهای متفاوت موضوع با اهمیتی است، زیرا در عمل ضرایب رگرسیونی وابسته به زمان دارای درجه‌های ناهمواری یکسانی نیستند. این اهمیت، به‌ویژه، زمانی که تعداد ضرایب وابسته به زمان افزایش یابد، پررنگ می‌شود.

مراجع

Fan, J. Q. and Zhang, W. Y. (1999), Statistical Estimation in Varying Coefficient Models, *The Annals of Statistics*, **27**, 1491-1518.

Liang, K. Y. and Zeger, S. L. (1986), Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, **73**, 13-22.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000), Semiparametric Regression for the Mean and Rate Function of Recurrent Events, *Journal of the Royal Statistical Society, Series B*, **62**, 711-730.

Martinussen, T., Scheike, T. H. and Skovgaard, I. M. (2002), Efficient Estimation of Fixed and Time-Varying Covariate Effects in Multiplicative Intensity Models, *Scandinavian Journal of Statistics*, **29**, 57-74.

Scheike, T. H. and Martinussen, T. (2004), On Estimation and Tests of Time-Varying Effects in the Proportional Hazards Models, *Scandinavian Journal of Statistics*, **31**, 51-62.

Sun, L., Zhou, X. and Guo, S. (2011), Marginal Regression Models with Time-Varying Coefficients for Recurrent Event Data, *Statistics in Medicine*, **30**, 2265-2277.