

برآزش مدل‌های نیمه پارامتری بقا با اثرات وابسته به زمان برای داده‌های بازگردنده با روش هسته

احسان اسحقى، حسین باغیشنى، داود شاهسونى

گروه ریاضی کاربردی، دانشگاه صنعتی شاهرود

تاریخ دریافت: ۱۳۹۲/۳/۱۴ تاریخ آخرین بازنگری: ۱۳۹۲/۶/۲۶

چکیده: در برخی از مدل‌های نیمه پارامتری بقا که برای مدل‌بندی داده‌های بازگردنده بقا منعطف و مفید هستند، ضرایب متغیرهای موجود در مدل، وابسته به زمان هستند. در این مدل‌ها برآوردگرها به صورت بسته و دقیق به دست نمی‌آیند و باید از روش‌های تقریبی برای محاسبه آن‌ها استفاده شود. شکل پیچیده این برآوردگرها، به دست آوردن توزیع آن‌ها را ناممکن می‌سازد. در این موارد معمولاً از نظریه مجانبی توزیع‌ها برای بررسی ویژگی‌های برآوردگرها استفاده می‌شود. در این مقاله، ضمن معرفی این مدل‌ها به تشریح برآورد پارامترهای آن، به کمک بسط تیلور و روش هسته، پرداخته، سازگاری و نرمال مجانبی بودن توزیع برآوردگرها نشان داده می‌شود. سپس عملکرد مدل و روش برآورد در یک مطالعه شبیه‌سازی ارزیابی می‌شود. در پایان کاربرد مدل با تحلیل داده‌های مربوط به شوک‌های وارد شده به بیماران قلبی در یکی از بیمارستان‌های شهر مشهد نشان داده می‌شود.

واژه‌های کلیدی: روش هسته، توزیع مجانبی، داده‌های بازگردنده، سازگاری، مدل نیمه پارامتری بقا، نرخ همگرایی.

در بسیاری از علوم مانند پزشکی، مهندسی و محیط زیست، گاهی پیشامد مورد نظر ممکن است بیش از یک بار رخ دهد. این قبیل پیشامدها را پیشامدهای بازگردنده^۱ و به داده‌های حاصل از آن‌ها، داده‌های بازگردنده گویند (کوک و لااولس، ۲۰۰۷). برای مثال، سکته قلبی و عود بیماری در پزشکی و خرابی یک دستگاه تولیدی در مهندسی، از جمله پیشامدهای بازگردنده هستند. یکی از ویژگی‌های داده‌های بازگردنده آن است که معمولاً نمی‌توان پیشامدهای متفاوت برای یک فرد را مستقل از هم در نظر گرفت.

از جمله رهیافت‌های مدل‌بندی این گونه داده‌ها استفاده از مدل‌های رگرسیون نیمه پارامتری است (کوک و لااولس، ۲۰۰۷). 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 در طول زمان تغییر کنند. بنابراین شناخت و سنجش اثرات موقتی متغیرها در زمان‌های رخداد پیشامدهای بازگردنده برای محققان مهم است. در نتیجه، در نظر گرفتن مدل‌هایی با ترکیب متغیرهای تبیینی وابسته و مستقل از زمان، از اهمیت بالایی برخوردار است.

چندین مدل رگرسیونی نیمه پارامتری برای تحلیل داده‌های بازگردنده در خانواده‌های مدل‌های شرطی (وی و همکاران، ۱۹۸۹) و کناری پیشنهاد شده‌اند (اندرسن و گیل، ۱۹۸۲، پرنیس و همکاران، ۱۹۸۱). مدل‌های رگرسیونی مخاطره شرطی (مارتینوسن و شیکه، ۲۰۰۶) به طور گسترده در تحلیل بقا برای توصیف وابستگی زمان‌های بقا به متغیرهای تبیینی مورد استفاده قرار می‌گیرند. مدل مخاطره نسبی کاکس^۲ (کاکس، ۱۹۷۲)، یکی از معروف‌ترین این مدل‌هاست. البته برای داده‌های بازگردنده نمی‌توان مدل کاکس را به طور مستقیم به کار برد، اما پایه مدل‌های مختلفی در این زمینه است. در عمل، مدل‌بندی میانگین تعداد پیشامدها در مقابل مدل مخاطره نسبی کاکس، برای داده‌های بازگردنده، قابل فهم‌تر است (سان و همکاران، ۲۰۱۱). لین و همکاران (۲۰۰۰) مدلی را برای میانگین و نرخ حاشیه‌ای بر

^۱ Recurrent event

^۲ Cox proportional hazard model

اساس یک تابع پیوند از نوع کاکس معرفی کردند و استنباطها را با فرض پیوستگی زمان، تعمیم دادند. مارتینسون و همکاران (۲۰۰۲)، مطالعاتی را برای برآورد ضرایب وابسته و مستقل از زمان در مدل‌های کاکس انجام دادند. اخیراً نیز سان و همکاران (۲۰۱۱)، مدل رگرسیونی حاشیه‌ای را با ترکیبی از ضرایب وابسته و مستقل از زمان برای داده‌های بازگردنده مورد استفاده قرار داده‌اند. در بخش ۲ مدل مورد نظر معرفی می‌شود و با استفاده از روش هسته^۳ و الگوریتم نیوتون-رافسون، رهیافتی برای برازش مدل ارائه می‌گردد. در بخش ۳ سازگاری، نرخ همگرایی و توزیع مجانبی برآوردگرها بررسی می‌شوند. در بخش ۴ کارایی مدل معرفی شده و روش برازش در کوچک نمونه، با یک مطالعه شبیه‌سازی ارزیابی می‌شود. در بخش ۵ مدل به داده‌های مربوط به ۶۵ بیمار مبتلا به بیماری قلبی که در بازه زمانی ابتدای سال ۱۳۸۴ لغایت پایان نیمه اول سال ۱۳۸۷ به مرکز قلب بیمارستان قائم (عج) مشهد مراجعه کردند، برازش داده می‌شود.

۲ مدل نیمه پارامتری 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 زمانی $[0, t]$ برای فرد i ام رخ می‌دهد و $X_i(\cdot)$ و $Z_i(\cdot)$ بردارهایی به ترتیب p و q بعدی برای فرد i ام باشند. در اکثر کاربردها، بازه زمانی مورد بررسی محدود است و ممکن است برخی از پیشامدها سانسور شوند. از این رو، $N_i^*(t)$ به‌طور کامل مشاهده نمی‌شود. اگر C_i به عنوان زمان سانسور تعریف شود، آن‌گاه فرآیند قابل مشاهده به صورت $N_i(t) = N_i^*(t \wedge C_i)$ تعریف می‌شود، که در آن $a \wedge b = \min(a, b)$. فرض کنید با شرط داشتن $X_i(\cdot)$ و $Z_i(\cdot)$ ، متغیرهای C_i و $N_i^*(\cdot)$ مستقل باشند و $Y_i(t) = I(C_i \geq t)$ که در آن $I(\cdot)$ تابع نشان‌گر است. بنابراین، مجموعه مشاهدات به صورت $\{N_i(\cdot), Y_i(\cdot), X_i(\cdot), Z_i(\cdot); i = 1, \dots, n\}$ هستند.

^۳ Kernel method

۴ برآزش مدل‌های نیمه پارامتری بقا با اثرات وابسته به زمان

مدل رگرسیون‌ی نرخ حاشیه‌ای با ترکیبی از ضرایب وابسته و مستقل از زمان،
به صورت

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

تعریف می‌شود (سان و همکاران، ۲۰۱۱)، 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

$$N(t) = (N_1(t), \dots, N_n(t))^T, \quad N_0(t) = n^{-1} \sum_{i=1}^n N_i(t)$$

$$X(t) = (X_1(t), \dots, X_n(t))^T, \quad Z(t) = (Z_1(t), \dots, Z_n(t))^T,$$

و

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

تحت مدل (۱)، $M_i(t)$ ها فرآیندهایی با میانگین صفر هستند. بنابراین به ازای مقادیر $\beta(t)$ و γ ، برآوردگری منطقی برای $\mu_0(t)$ از حل عبارت

$$\sum_{i=1}^n [dN_i(t) - Y_i(t) \exp\{\beta_0(t)^T X_i(t) + \gamma_0^T Z_i(t)\} d\mu_0(t)] = 0, \quad 0 \leq t \leq \tau,$$

به دست می‌آید، که در آن τ مقدار تعیین شده‌ای است که $P(C_i \geq \tau) > 0$ در نتیجه برآوردگر $\mu_0(t)$ به صورت

$$\hat{\mu}_0(t; \beta, \gamma) = \int_0^t S_0(u; \beta, \gamma)^{-1} dN_0(u), \quad (2)$$

محاسبه می‌شود. به طوری که

$$S_0(t; \beta, \gamma) = \frac{1}{n} \sum_{i=1}^n \phi_i(t; \beta, \gamma),$$

$$\phi_i(t) = \phi_i(t; \beta, \gamma) = Y_i(t) \exp\{\beta(t)^T X_i(t) + \gamma^T Z_i(t)\}.$$

برای برآورد $\beta_0(t)$ و γ_0 ، با استفاده از معادله‌های برآوردیابی تعمیم‌یافته^۴ (GEE) لیانگ و زگر (۱۹۸۶) و رابطه (۲)، معادله‌های برآوردیاب به صورت

$$\begin{aligned} X(t)^T [dN(t) - \phi(t; \beta, \gamma) S_0(u; \beta, \gamma)^{-1} dN_0(u)] &= 0, \quad 0 \leq t \leq \tau, \\ \int_0^\tau Z(t)^T [dN(t) - \phi(t; \beta, \gamma) S_0(u; \beta, \gamma)^{-1} dN_0(u)] &= 0, \end{aligned} \quad (3)$$

حاصل می‌شوند، که در آن $\phi(t; \beta, \gamma) = (\phi_1(t; \beta, \gamma), \dots, \phi_n(t; \beta, \gamma))^T$ معادله‌های (۳) با روش نیوتون-رافسون، به وسیله بسط تیلور تابع $\phi(t; \beta, \gamma) S_0(t; \beta, \gamma)^{-1}$ حول برآورد جاری $(\beta^\ell(t), \gamma^\ell)$ ، برای برآورد $\beta_0(t)$ و γ_0 مورد استفاده قرار می‌گیرند. معادله‌های به‌روزکننده به صورت

$$\begin{aligned} &\{\beta^{\ell+1}(t) - \beta^\ell(t)\} S_0^\ell(t)^{-1} dN_0(t) \\ &= n^{-1} E_{xx}^\ell(t)^{-1} \{X(t) - \bar{X}^\ell(t)\}^T [dN(t) - \Phi^\ell(t) \{Z(t) - \bar{Z}^\ell(t)\} \\ &\quad \times (\gamma^{\ell+1} - \gamma^\ell) S_0^\ell(t)^{-1} dN_0(t)], \end{aligned} \quad (4)$$

$$\begin{aligned} &n^{-1} \int_0^\tau \{Z(t) - \bar{Z}^\ell(t)\}^T dN(t) - \int_0^\tau E_{zx}^\ell(t) \{\beta^{\ell+1}(t) - \beta^\ell(t)\} \\ &\quad \times S_0^\ell(t)^{-1} dN_0(t) \\ &= \int_0^\tau E_{zz}^\ell(t) (\gamma^{\ell+1} - \gamma^\ell) S_0^\ell(t)^{-1} dN_0(t), \end{aligned} \quad (5)$$

تعریف می‌شوند، که در آن‌ها $\Phi^\ell(t)$ ماتریسی قطری از $i = 1, \dots, n$ $\phi_i(t; \beta, \gamma)$ است و سایر کمیت‌ها به صورت

$$\begin{aligned} E_{xx}(t; \beta, \gamma) &= \frac{1}{n} \{X(t) - \bar{X}(t; \beta, \gamma)\}^T \Phi(t; \beta, \gamma) \{X(t) - \bar{X}(t; \beta, \gamma)\}, \\ E_{zx}(t; \beta, \gamma) &= \frac{1}{n} \{Z(t) - \bar{Z}(t; \beta, \gamma)\}^T \Phi(t; \beta, \gamma) \{X(t) - \bar{X}(t; \beta, \gamma)\}, \\ E_{zz}(t; \beta, \gamma) &= \frac{1}{n} \{Z(t) - \bar{Z}(t; \beta, \gamma)\}^T \Phi(t; \beta, \gamma) \{Z(t) - \bar{Z}(t; \beta, \gamma)\}, \\ E_{xx}^\ell(t) &= E_{xx}(t; \beta^\ell, \gamma^\ell), \end{aligned}$$

^۴ Generalized Estimating Equation

$$E_{zx}^{\ell}(t) = E_{zx}(t; \beta^{\ell}, \gamma^{\ell}),$$

$$E_{zz}^{\ell}(t) = E_{zz}(t; \beta^{\ell}, \gamma^{\ell}),$$

$$S_{\circ}^{\ell}(t) = S_{\circ}(t; \beta^{\ell}, \gamma^{\ell}),$$

$$\Phi^{\ell}(t) = \Phi(t; \beta^{\ell}, \gamma^{\ell}),$$

محاسبه می‌شوند. ماتریس‌هایی $\bar{X}^{\ell}(t)$ و $\bar{Z}^{\ell}(t)$ نیز به ترتیب با ابعاد $n \times p$ و $n \times q$ با سطرهای $E_x^{\ell}(t)$ و $E_z^{\ell}(t)$ هستند به طوری که

$$E_x^{\ell}(t; \beta, \gamma) = \frac{S_x^{\ell}(t; \beta, \gamma)}{S_{\circ}^{\ell}(t; \beta, \gamma)}, \quad E_z^{\ell}(t; \beta, \gamma) = \frac{S_z^{\ell}(t; \beta, \gamma)}{S_{\circ}^{\ell}(t; \beta, \gamma)},$$

$$S_x^{\ell}(t; \beta, \gamma) = \frac{1}{n} \sum_{i=1}^n \phi_i^{\ell}(t; \beta, \gamma) X_i(t), \quad S_z^{\ell}(t; \beta, \gamma) = \frac{1}{n} \sum_{i=1}^n \phi_i^{\ell}(t; \beta, \gamma) Z_i(t).$$

همچنین وقتی $n \rightarrow \infty$ مقادیر حدی $S_z(t; \beta, \gamma)$, $S_x(t; \beta, \gamma)$, $S_{\circ}(t; \beta, \gamma)$

$E_{zx}(t; \beta, \gamma)$ و $E_{zz}(t; \beta, \gamma)$ به صورت

$$e_{zx}(t; \beta, \gamma), e_z(t; \beta, \gamma), e_x(t; \beta, \gamma), s_z(t; \beta, \gamma), s_x(t; \beta, \gamma), s_{\circ}(t; \beta, \gamma)$$

و به طور مشابه، نمادهای $e_{zz}(t; \beta, \gamma)$ و $e_{zx}(t; \beta, \gamma)$ تعریف می‌شوند. تعریف می‌کنیم

$s_{\circ}(t) = s_{\circ}(t; \beta_{\circ}, \gamma_{\circ})$ و به طور مشابه، نمادهای $e_{zz}(t)$ و $e_{zx}(t)$

تعریف می‌شوند. مقدار حدی $A(\tau)$ نیز به صورت

$$a(\tau) = \int_0^{\tau} [e_{zz}(t) - e_{zx}(t)e_{xx}(t)^{-1}e_{zx}(t)^T] d\mu_{\circ}(t),$$

تعریف می‌شود. همچنین فرض کنید $\bar{a}(t)$ مقدار حدی $\bar{X}(t)$ ، ماتریسی $n \times p$ با

سطرهای $e_x(t)$ و $\bar{z}(t)$ ، مقدار حدی $\bar{Z}(t)$ ، ماتریسی $n \times q$ با سطرهای $e_z(t)$ باشند.

با قرار دادن رابطه (۴) در (۵) و حل معادله بر اساس $\gamma^{\ell+1}$ ، رابطه تکراری

$$\gamma^{\ell+1} = \Psi_r(\gamma^{\ell})$$

$$\Psi_r(\gamma^{\ell}) = \gamma^{\ell} + \frac{A^{\ell}(\tau)^{-1}}{n} \int_0^{\tau} [\{Z(t) - \bar{Z}^{\ell}(t)\}^T - E_{zx}^{\ell}(t)E_{xx}^{\ell}(t)^{-1} \\ \times \{X(t) - \bar{X}^{\ell}(t)\}^T] dN(t), \quad (۶)$$

که در آن $A^{\ell}(\tau) = A(\tau; \beta^{\ell}, \gamma^{\ell})$ عبارت است از

$$A(\tau; \beta, \gamma) = \int_0^{\tau} [E_{zz}(t) - E_{zx}(t)E_{xx}(t)^{-1}E_{zx}(t)^T] \frac{dN_{\circ}(t)}{S_{\circ}(t)}.$$

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. مارتینوسن و همکاران (۲۰۰۲) و شیکه و مارتینوسن (۲۰۰۴)، پیشنهاد کردند به جای استفاده از $\beta_0(t)$ ، از ضرایب رگرسیون تجمعی $B_0(t) = \int_0^t \beta_0(s) ds$ استفاده شود که منجر به نتایج پایدارتری می‌شود. فرض کنید $\mu_0^\ell(t) = \hat{\mu}_0(t; \beta^\ell, \gamma^\ell)$ و

$$\lambda_0^\ell(t) = \int h^{-1} K\left(\frac{u-t}{h}\right) d\mu_0^\ell(u),$$

$$\beta^\ell(t) = \int h^{-1} K\left(\frac{u-t}{h}\right) dB^\ell(u),$$

برآوردگرهای هسته‌ای $\beta_0(t)$ و $\lambda_0(t) = \frac{d}{dt}\mu_0(t)$ بر پایه $B^\ell(t)$ و $\mu_0^\ell(t)$ با پهنای نوار h باشند، که در آن تابع هسته متقارن با دامنه فشرده است (شرط (C۴) را ملاحظه کنید). با استفاده از $\gamma^{\ell+1}$ و با قرار دادن $\frac{dN_0(t)}{S_0^\ell(t)}$ در رابطه (۴) به جای $\lambda_0^\ell(t) dt$ ، به رابطه تکراری $B^{\ell+1} = \Psi_b(B^\ell)$ دست می‌یابیم، به طوری که

$$\Psi_b(B^\ell(t)) = \int_0^t \beta^\ell(u) du + n^{-1} \int_0^t \lambda_0^\ell(u)^{-1} E_{xx}^\ell(u)^{-1} \{X(u) - \bar{X}^\ell(u)\}^T$$

$$\times [dN(u) - \Phi^\ell(u) \{Z(u) - \bar{Z}^\ell(u)\} (\gamma^{\ell+1} - \gamma^\ell) S_0^\ell(u)^{-1} dN_0(u)]. \quad (۷)$$

با استفاده متوالی از معادله‌های تکراری (۶) 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

فرض کنید $\|v\| = (v^T v)^{\frac{1}{2}}$ ، نرم اقلیدسی برای بردار v و شرایط نظم به صورت زیر باشند:

(C۱) $\beta_0(t)$ و $\mu_0(t)$ ، توابعی پیوسته از $t \in [0, \tau]$ و سه بار مشتق‌پذیر هستند.

(C۲) $X_i(t)$ و $Z_i(t)$ ، متغیرهایی کران‌دار در بازه $[0, \tau]$ هستند.

۸ برآزش مدل‌های نیمه پارامتری بقا با اثرات وابسته به زمان

(C۳) $a(\tau)$ و $e_{xx}(t)$ ، به ازای $t \in [0, \tau]$ ، ماتریس‌هایی نامنفرد هستند.

(C۴) $K(\cdot)$ ، یک تابع هسته پیوسته و متقارن با دامنه فشرده است، به طوری که

$$\int K(u)du = 1 \text{ و به ازای } \frac{1}{\lambda} < \alpha < \frac{1}{\lambda} \text{، } h \text{ برابر } O(n^{-\alpha}) \text{ است.}$$

(C۵) $s_j(t; \beta, \gamma)$ در $[0, \tau] \times \vartheta \times \Theta$ به طور یکنواخت پیوسته می‌باشد.

همچنین ϑ مجموعه‌ای فشرده از R^p است که شامل یک همسایگی برای $\beta_0(t)$ به

ازای مقادیر $t \in [0, \tau]$ است و Θ ، مجموعه‌ای فشرده از R^q و شامل γ_0 است. z نیز

مقادیر 0 ، x یا z را اختیار می‌کند.

(C۶) $\hat{\beta}(t)$ و $\hat{\gamma}$ ، به ترتیب مقادیر اولیه γ_0 و $B_0(t)$ هستند، به طوری که

$$\|\hat{\gamma} - \gamma_0\| = O_p(n^{-\frac{1}{4}}), \quad \sup_{0 \leq t \leq \tau} \|\hat{B}(t) - B_0(t)\| = O_p(n^{-\frac{1}{4}}).$$

قضیه ۱: تحت شرایط نظم (C۱) تا (C۶)، معادله $\Psi_r(\gamma) = \gamma$ دارای جواب $\hat{\gamma}$

است، به طوری که $\|\hat{\gamma} - \gamma_0\| = O_p(n^{-\frac{1}{4}})$ و $n^{\frac{1}{4}}(\hat{\gamma} - \gamma_0)$ به طور مجانبی دارای

توزیع نرمال با میانگین صفر و ماتریس کوواریانس است که به وسیله

$$\Sigma = n^{-1} \sum_{i=1}^n \hat{\xi}_i(\tau) \hat{\xi}_i(\tau)^T \text{ برآورد می‌شود، به طوری که}$$

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

برهان: ابتدا فرض کنید

$$\bar{\beta}_0(t) = \int \frac{1}{h} K\left(\frac{u-t}{h}\right) dB_0(u).$$

با استفاده از بسط تیلور و شرایط (C۱) و (C۴) می‌توان نوشت

$$\sup_{0 \leq t \leq \tau} \|\bar{\beta}_0(t) - \beta_0(t)\| = O(h^2). \quad (\Lambda)$$

با به کار بردن شرایط (C۴) و (C۶) می‌توان نشان داد

$$\hat{\beta}(t) - \beta_0(t) = O_p(h^{-1} \|\hat{B} - B_0\|) + O(h^2) = o_p(n^{-\frac{1}{4}}). \quad (۹)$$

در نتیجه با استفاده از (۹) می توان نتیجه گرفت

$$\hat{A}(\tau) - a(\tau) = o_p(n^{-\frac{1}{2}}). \quad (10)$$

با قرار دادن $dN(t) = dM(t) + \phi(t)d\mu_0(t)$ در رابطه (۶) داریم

$$\begin{aligned} \Psi_r(\hat{\gamma}) &= \hat{\gamma} + \frac{\hat{A}(\tau)^{-1}}{n} \int_0^\tau [\{Z(t) - \tilde{Z}(t)\}^T - \hat{E}_{zx}(t)\hat{E}_{xx}(t)^{-1} \\ &\times \{X(t) - \tilde{X}(t)\}^T] dM(t) \\ &+ \frac{\hat{A}(\tau)^{-1}}{n} \int_0^\tau [\{Z(t) - \tilde{Z}(t)\}^T - \hat{E}_{zx}(t)\hat{E}_{xx}(t)^{-1} \\ &\times \{X(t) - \tilde{X}(t)\}^T] \phi(t) d\mu_0(t), \end{aligned} \quad (11)$$

که در آن $M(t) = (M_1(t), \dots, M_n(t))^T$ با توجه به رابطه (۹)، جمله دوم سمت راست رابطه (۱۱) برابر است با

$$\begin{aligned} \frac{\hat{a}(\tau)^{-1}}{n} \int_0^\tau [\{Z(t) - \bar{z}(t)\}^T - e_{zx}(t)e_{xx}(t)^{-1} \{X(t) - \bar{x}(t)\}^T] dM(t) \\ + o_p(n^{-\frac{1}{2}}). \end{aligned} \quad (12)$$

مشابه اثبات رابطه (۱۰)، داریم

$$\int_0^\tau \hat{E}_{zz}(t) - \hat{E}_{zx}(t)\hat{E}_{xx}(t)^{-1}\hat{E}_{zx}(t)^T d\mu_0(t) - a(\tau) = o_p(n^{-\frac{1}{2}}). \quad (13)$$

با استفاده از بسط تیلور $\phi(t; \beta_0, \gamma_0)$ در $(\hat{\beta}(t), \hat{\gamma})$ و روابط (۱۰) و (۱۳)، جمله سوم رابطه (۱۱) به صورت

$$\begin{aligned} (a(\tau)^{-1} + o_p(n^{-\frac{1}{2}})) \{ [a(\tau) + o_p(n^{-\frac{1}{2}})] (\gamma_0 - \hat{\gamma}) + o_p(n^{-\frac{1}{2}}) \} \\ = (\gamma_0 - \hat{\gamma}) + o_p(n^{-\frac{1}{2}}), \end{aligned} \quad (14)$$

قابل بیان است. بنابراین رابطه (۱۱) را می توان با استفاده از روابط (۱۲) و (۱۴)، به صورت

$$\begin{aligned} \Psi_r(\hat{\gamma}) &= \gamma_0 + \frac{\hat{a}(\tau)^{-1}}{n} \int_0^\tau [\{Z(t) - \bar{z}(t)\}^T - e_{zx}(t)e_{xx}(t)^{-1} \\ &\times \{X(t) - \bar{x}(t)\}^T] dM(t) + o_p(n^{-\frac{1}{2}}) \\ &= \gamma_0 + O_p(n^{-\frac{1}{2}}), \end{aligned} \quad (15)$$

۱۰برآزش مدل‌های نیمه پارامتری بقا با اثرات وابسته به زمان

نوشت. در نتیجه $\|\hat{\gamma} - \gamma_0\| = O_p(n^{-\frac{1}{2}})$. از سوی دیگر، تساوی (۱۵) را می‌توان به صورت

$$\Psi_r(\hat{\gamma}) = \gamma_0 + n^{-1} \sum_{i=1}^n \xi_i(\tau) + o_p(n^{-\frac{1}{2}}),$$

بازنویسی کرد. بنابراین

$$n^{\frac{1}{2}}(\hat{\gamma} - \gamma_0) = n^{-\frac{1}{2}} \sum_{i=1}^n \xi_i(\tau) + o_p(1). \quad (16)$$

در نهایت با استفاده از قضیه حد مرکزی چندمتغیره، $n^{\frac{1}{2}}(\hat{\gamma} - \gamma_0)$ در توزیع به بردار تصادفی نرمال با میانگین صفر و ماتریس کوواریانس $\Sigma = E\{\xi_i(\tau)\xi_i(\tau)^T\}$ همگرا است که به وسیله $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \hat{\xi}_i(\tau)\hat{\xi}_i(\tau)^T$ برآورد می‌شود، به طوری که

$$\hat{\xi}_i(\tau) = \hat{A}(\tau)^{-1} \int_0^T [\{Z_i(t) - \hat{E}_z(t)\}^T - \hat{E}_{zx}(t)\hat{E}_{xx}(t)^{-1}\{X_i(t) - \hat{E}_x(t)\}^T] dM_i(t).$$

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

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

برهان: برای اثبات سازگاری و به دست آوردن نرخ همگرایی $\hat{B}(t)$ به $B_0(t)$ ، توجه به رابطه (۷)، از آنجایی که برآوردهای γ در دو مرحله متوالی پایانی اختلاف ناچیزی دارند و قابل صرف نظر کردن هستند، می‌توان نوشت

$$\Psi_b(\hat{B})(t) = \int_0^t \hat{\beta}(u) du + \frac{1}{n} \int_0^t \hat{\lambda}_0(u)^{-1} \hat{E}_{xx}(u)^{-1} \{X(u) - \tilde{X}(u)\}^T dN(u). \quad (17)$$

همانند اثبات سازگاری ضرایب مستقل از زمان، رابطه (۱۷) را می توان به صورت

$$\begin{aligned} \Psi_b(\hat{B})(t) &= B_o(t) + \frac{1}{n} \int_0^t \lambda_o(u)^{-1} e_{xx}(u)^{-1} \{X(u) - \bar{x}(u)\}^T dM(u) \\ &+ \int_0^t \lambda_o(u)^{-1} e_{xx}(u)^{-1} e_{zx}(u)^T d\mu_o(u) (\hat{\gamma} - \gamma_o) + o_p(n^{-\frac{1}{2}}) \\ &= B_o(t) + O(n^{-1}) + O_p(n^{-\frac{1}{2}}) + o_p(n^{-\frac{1}{2}}), \end{aligned} \quad (18)$$

بازنویسی کرد. با توجه به دلخواه بودن مقدار t در بازه $[\circ, \tau]$ ، داریم

$$\sup_{\circ \leq t \leq \tau} \|\hat{B}(t) - B_o(t)\| = O_p(n^{-\frac{1}{2}}).$$

از طرفی با توجه به رابطه (۱۶)، رابطه (۱۸) را می توان به صورت

$$\begin{aligned} \hat{B}(t) - B_o(t) &= n^{-1} \sum_{i=1}^n \left(\int_0^t \lambda_o(u)^{-1} e_{xx}(u)^{-1} \{X_i(u) - e_x(u)\}^T dM_i(u) \right. \\ &- \left. \int_0^t e_{xx}(u)^{-1} e_{zx}(u)^T du \xi_i(\tau) \right) + o_p(n^{-\frac{1}{2}}) \\ &= n^{-1} \sum_{i=1}^n \eta_i(t) + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

نوشت. با استفاده از قضیه حد مرکزی، $n^{\frac{1}{2}}(\hat{B}(t) - B_o(t))$ در توزیع به توزیع نرمال با میانگین صفر و تابع کوواریانس $\Gamma_b(s, t)$ همگرا است که به وسیله $\hat{\Gamma}_b(s, t) = \frac{1}{n} \sum_{i=1}^n \hat{\eta}_i(s) \hat{\eta}_i(t)^T$ که

$$\begin{aligned} \hat{\eta}_i(t) &= \int_0^t \hat{\lambda}_o(u)^{-1} \hat{E}_{xx}(u)^{-1} \{X_i(u) - \hat{E}_x(u)\}^T d\hat{M}_i(u) \\ &- \int_0^t \hat{E}_{xx}(u)^{-1} \hat{E}_{zx}(u)^T du \hat{\xi}_i(\tau). \end{aligned}$$

۱.۳ آزمون وابستگی به زمان ضرایب رگرسیونی

تشخیص اثرات رگرسیونی وابسته به زمان و مستقل از زمان مهم است. با این گونه دسته بندی، می توان به یک مدل مناسب تر برای تحلیل داده ها دست یافت. برای این منظور، می توان فرضیه $H_o: \beta_{oj}(t) = \beta_{oj}$ را آزمون نمود

به طوری که $\beta_{\circ j}(t)$ ، ضریب رگرسیونی زامین متغیر تبیینی است. با قرار دادن $V(t) = B_{\circ j}(t) - B_{\circ j}(\tau)t/\tau$ و $\hat{V}(t) = \hat{B}_j(t) - \hat{B}_j(\tau)t/\tau$ ، که در آن‌ها $B_{\circ j}(t)$ و $\hat{B}_j(t)$ به ترتیب زامین عضو $B_{\circ}(t)$ و $\hat{B}(t)$ هستند، تحت فرضیه صفر داریم

$$V(t) = \int_{\circ}^t \beta_{\circ j}(u) du - \frac{t}{\tau} \int_{\circ}^{\tau} \beta_{\circ j}(u) du = \beta_{\circ j}(t - \circ) - \frac{t}{\tau} \beta_{\circ j}(\tau - \circ) = \circ.$$

بنابراین اگر تحت این فرضیه، دو تابع $V(t)$ و $\hat{V}(t)$ نزدیک باشند، حاکی از آن است که زامین عضو $\hat{\beta}(t)$ ، مستقل از زمان است. از این رو، می‌توان از آزمون‌هایی که برای نیکویی برازش توزیع‌ها به کار می‌روند، بهره برد. دو آماره آزمون، مبتنی بر $\hat{V}(t)$ ، آماره آزمون کلموگروف-اسمیرنوف (بهبودیان، ۱۳۸۷)، $F_1 = \sup_{\circ \leq t \leq \tau} |n^{\frac{1}{2}} \hat{V}(t)|$ ، و آماره آزمون کرامر-ون میسز (استفنز، ۱۹۸۶)، $F_2 = \int_{\circ}^{\tau} n \hat{V}(t)^2 dt$ ، که برای مقایسه توابع توزیع ساخته شده‌اند. از آنجایی که $V(t)$ و $\hat{V}(t)$ توابع توزیع نیستند، باید از رهیافتی مشابه این دو آزمون استفاده شود. بنابراین با استفاده از $n^{\frac{1}{2}}\{\hat{V}(t) - V(t)\}$ آماره آزمون از نوع کلموگروف-اسمیرنوف F_1 و از نوع کرامر-ون میسز F_2 تحت فرضیه صفر به صورت

$$F_1 = \sup_{\circ \leq t \leq \tau} |n^{\frac{1}{2}} \hat{V}(t)|, F_2 = \int_{\circ}^{\tau} n \hat{V}(t)^2 dt,$$

ساخته می‌شوند. با توجه به قضیه ۲ می‌توان نوشت

$$n^{\frac{1}{2}}\{\hat{V}(t) - V(t)\} = n^{\frac{1}{2}} \sum_{i=1}^n (\eta_{ij}(t) - \eta_{ij}(\tau) \frac{t}{\tau}),$$

که در توزیع معادل با

$$\hat{W}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \{\hat{\eta}_{ij}(t) - \hat{\eta}_{ij}(\tau)t/\tau\},$$

است و از یک فرآیند نرمال با میانگین صفر پیروی می‌کند، که در آن $\hat{\eta}_{ij}(t)$ ، زامین عضو $\hat{\eta}_i(t)$ است. ساختار تابع کوواریانس $\hat{W}(t)$ پیچیده و استفاده از آن مشکل می‌باشد. بدین منظور، لین و همکاران (۲۰۰۰) نشان دادند که توزیع فرآیند $\hat{W}(t)$ می‌تواند با فرآیند

$$\tilde{W}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \{\hat{\eta}_{ij}(t) - \hat{\eta}_{ij}(\tau)t/\tau\} G_i,$$

که دارای توزیع نرمال با میانگین صفر است تقریب زده شود، به طوری که (G_1, \dots, G_n) متغیرهای تصادفی مستقل نرمال استاندارد هستند. بنابراین توزیع های F_1 و F_2 به ترتیب با $\tilde{F}_1 = \sup_{0 \leq t \leq \tau} |\tilde{W}(t)|$ و $\tilde{F}_2 = \int_0^\tau \tilde{W}(t)^2 dt$ برآورد می شوند.

۴ مطالعه شبیه سازی

در این بخش با انجام شبیه سازی، ویژگی های مجانبی برآوردگرهای ارایه شده در بخش ۲ برای حجم های نمونه کوچک و متوسط در قالب دو مثال ارزیابی می شوند. در مثال اول، به ازای $i = 1, \dots, n$ ، زمان های رخداد از یک فرآیند پواسن با مدل نرخ کناری

$$E\{dN_i^*(t) | X_i(t), Z_i(t)\} = \exp\{-\circ/5 + \circ/5 \cos(2t - 1/75)X_i(t) + \circ/5 Z_i(t)\} dt, \quad (19)$$

تولید شده است، به طوری که Z_i و X_i به ترتیب دارای توزیع نرمال استاندارد و توزیع برنولی با احتمال موفقیت $\circ/5$ هستند. زمان های سانسور C_i نیز از توزیع یکنواخت $U(2, 5)$ تولید شده اند که تقریباً به طور متوسط ۳ رخداد برای هر فرد روی می دهد. برای دو حجم نمونه ($n = 100, 200$) و با استفاده از چندین پهنای نوار، همواری های متفاوتی بررسی شده اند. با در نظر گرفتن $\tau = 4/5$ ، برآورد پارامتر مستقل از زمان γ و برآورد ضریب رگرسیونی تجمعی وابسته به زمان $B(t)$ ، در مدل (۱۹) به ازای دو هسته نرمال و اپانچنیکوف (ریزو، ۲۰۰۸) به دست آورده شده اند.

جدول ۱ شامل اریبی (Bias)، میانگین نمونه ای خطاهای استاندارد برآورد شده $\hat{\gamma}$ (SEE)، خطای استاندارد نمونه ای $\hat{\gamma}$ (SSE) و احتمال پوشش تجربی با سطح ۹۵ درصد برای γ_0 (CP) است، که بر مبنای ۵۰۰ مجموعه داده شبیه سازی شده محاسبه شده اند. مقادیر اریبی به گونه ای است که به نظر می رسد برآوردها ناریب هستند. مقادیر احتمال پوشش تجربی با افزایش حجم نمونه از ۱۰۰ تا ۲۰۰ به $\circ/95$ نزدیک تر می شوند. همچنین خطای برآوردگر کمتر می شود که این نتایج تاییدی بر

ویژگی‌های مجانبی برآوردگر $\hat{\gamma}$ می‌باشد. مقایسه نتایج دو هسته، بیان‌گر آن است که عملکرد آن‌ها تقریباً یکسان است. مشاهده می‌شود که روش برآورد پیشنهاد شده به ازای حجم‌های نمونه متفاوت و همچنین نوارهایی با پهناهای متفاوت، بسیار خوب عمل کرده است.

جدول ۲، مقادیر اریبی جمع‌بسته توان دوم^۶ (ISB) و میانگین جمع‌بسته توان دوم خطاها^۷ (MISE) را برای برآورد ضریب رگرسیون تجمعی $B_0(t)$ در مدل (۱۹) با دو هسته نرمال و اپانچنیکوف نشان می‌دهد. مقادیر کوچک ISB و MISE، نشان می‌دهند که برآوردگر پیشنهادی بسیار خوب عمل می‌کند و تقریباً برآوردی نااریب از ضریب رگرسیونی تجمعی ارائه می‌دهد.

برای بررسی عملکرد نقطه‌ای برآوردگر $B_0(t)$ ، $\hat{B}(t)$ و خطاهای استاندارد مجانبی، برآوردها و احتمال‌های پوشش برای $\hat{B}(t)$ در ۱۰۰ نقطه $k = 1, \dots, 100$ ، محاسبه شده‌اند. در شکل ۱-الف، حجم نمونه $n = 100$ و پهناهای نوار $h = 0.4$ و در شکل ۱-ب $n = 200$ و $h = 0.4$ در نظر گرفته شده‌اند. منحنی توپر نشان‌دهنده نمودار تابع رگرسیونی تجمعی واقعی $B_0(t) = \int_0^t \beta_0(s) ds = \int_0^t 0.5 \cos(2s - 1/\sqrt{5}) ds$ و منحنی خط چین، میانگین برآوردهای $\hat{B}(t)$ در ۵۰۰ تکرار است. همانطور که ملاحظه می‌شود با افزایش حجم نمونه، نمودار برآورد به نمودار واقعی نزدیک‌تر می‌شود که این حاکی از سازگاری برآوردگر است. شکل ۲-الف برای حجم نمونه $n = 100$ و پهناهای نوار $h = 0.4$ و ۲-ب برای $n = 200$ و $h = 0.4$ رسم شده‌اند. همانطور که ملاحظه می‌شود، خطاهای استاندارد مجانبی در بازه اطمینان واقع شده توصیف بسیار خوبی از تغییرات $\hat{B}(t)$ ارائه می‌دهند و نشان می‌دهد نتایج مجانبی برای حجم‌های کوچک و متوسط نیز قابل استفاده هستند. این دو شکل با استفاده از هسته اپانچنیکوف ایجاد شده‌اند و برای هسته نرمال نیز نتایج مشابهی حاصل شده‌اند. با توجه به دو قضیه ۱ و ۲، توزیع مجانبی برآوردهای $\hat{\gamma}$ و $\hat{B}(t)$ نرمال هستند. برای بررسی این ویژگی، نمودارهای چگالی آن‌ها برای حجم نمونه ۲۰۰، پهناهای نوار ۰/۴ و هسته اپانچنیکوف

^۶ Integrated Squared Bias

^۷ Mean Integrated Squared Error

جدول ۱: مقادیر اریبی، خطای استاندارد مجانبی و تجربی و نرخ پوشش برای پارامتر مستقل از زمان به ازای $\gamma_0 = 0/5$

هسته	n	h	Bias	SEE	SSE	CP
اپانچنیکوف	۱۰۰	۰/۴	۰/۰۰۵۸	۰/۱۱۹۵	۰/۱۲۲۱	۰/۹۵۷
		۰/۵	-۰/۰۰۲۴	۰/۱۱۹۵	۰/۱۲۲۳	۰/۹۶۰
		۰/۶	۰/۰۰۳۴	۰/۱۱۹۱	۰/۱۲۱۸	۰/۹۴۳
		۰/۷	-۰/۰۰۰۷	۰/۱۱۹۷	۰/۱۱۸۸	۰/۹۴۷
	۲۰۰	۰/۴	۰/۰۰۸۹	۰/۰۸۴۵	۰/۰۸۳۵	۰/۹۶۰
		۰/۵	۰/۰۰۰۵	۰/۰۸۴۵	۰/۰۸۲۸	۰/۹۶۳
		۰/۶	۰/۰۰۲۹	۰/۰۸۵۸	۰/۰۸۴۳	۰/۹۶۵
		۰/۷	-۰/۰۰۳۹	۰/۰۸۴۹	۰/۰۸۶۶	۰/۹۴۳
نرمال	۱۰۰	۰/۴	۰/۰۰۸۸	۰/۱۲۰۳	۰/۱۱۷۹	۰/۹۴۰
		۰/۵	۰/۰۱۲۹	۰/۱۲۰۲	۰/۱۲۰۶	۰/۹۳۷
		۰/۶	۰/۰۰۴۵	۰/۱۱۹۲	۰/۱۲۵۴	۰/۹۴۳
		۰/۷	۰/۰۰۷۵	۰/۱۱۸۶	۰/۱۳۵۸	۰/۹۰۳
	۲۰۰	۰/۴	-۰/۰۰۰۱	۰/۰۸۴۷	۰/۰۸۷۴	۰/۹۴۴
		۰/۵	۰/۰۰۱۸	۰/۰۸۴۷	۰/۰۸۴۵	۰/۹۵۳
		۰/۶	-۰/۰۰۰۵	۰/۰۸۴۷	۰/۰۸۴۰	۰/۹۵۰
		۰/۷	-۰/۰۰۴۰	۰/۰۸۴۸	۰/۰۸۰۳	۰/۹۷۳

رسم شده‌اند. این نتایج برای هسته نرمال نیز مشابه هستند. شکل ۳، چگالی‌های $\hat{B}(t)$ را در ۴ نقطه زمانی متفاوت نشان می‌دهد. نمودارهای الف، ب، ج و د به ترتیب چگالی‌های حاصل از ۵۰۰ برآورد استاندارد شده $\hat{B}(t)$ ، در نقاط زمانی ۸، ۴۰، ۸۰ و ۱۰۰ می‌باشند. در این نمودارها، منحنی توپر چگالی $\hat{B}(t)$ استاندارد شده و منحنی خط چین چگالی نرمال استاندارد می‌باشند. با توجه به نمودارها، این ضریب، در نقاط زمانی دلخواه انتخاب شده، دارای توزیع نرمال هستند. نتیجه مشابهی نیز برای برآوردگر $\hat{\gamma}$ برقرار است که در شکل ۴ نمایش داده شده است.

اکنون عملکرد مدل پیشنهادی در حالتی بررسی می‌شود که تعداد پارامترهای وابسته به زمان افزایش یافته و ترکیبی از شکل‌های سینوسی و غیرسینوسی می‌باشد.

جدول ۲: مقادیر اربیی جمع‌بسته توان دوم و میانگین جمع‌بسته توان دوم خطاها برای ضریب رگرسیون تجمعی وابسته به زمان $\hat{B}(t)$

هسته اپانچنیکوف		هسته نرمال		h	n
ISB	MISE	ISB	MISE		
۰/۰۰۰۳	۰/۱۰۸۴	۰/۰۰۰۶	۰/۱۰۹۲	۰/۴	۱۰۰
۰/۰۰۱۱	۰/۰۹۶۵	۰/۰۰۱۸	۰/۱۲۷۵	۰/۵	
۰/۰۰۱۳	۰/۱۰۱۹	۰/۰۰۳۱	۰/۱۱۵۶	۰/۶	
۰/۰۰۲۷	۰/۰۹۹۸	۰/۰۰۵۸	۰/۱۱۴۲	۰/۷	
۰/۰۰۱۶	۰/۰۵۸۹	۰/۰۰۰۹	۰/۰۵۰۹	۰/۴	۲۰۰
۰/۰۰۰۸	۰/۰۵۰۵	۰/۰۰۱۵	۰/۰۴۹۶	۰/۵	
۰/۰۰۱۷	۰/۰۵۲۹	۰/۰۰۳۱	۰/۰۵۴۶	۰/۶	
۰/۰۰۲۶	۰/۰۴۹۳	۰/۰۰۵۵	۰/۰۵۵۹	۰/۷	

بدین منظور مدل

$$E\{dN_i^*(t)|X_i(t), Z_i(t)\} = \exp\{-0.5 + 0.5 \cos(2t - 1/\sqrt{5})X_{i1}(t) + 0.7(\sqrt{t} - 1)X_{i2}(t) + 0.5Z_i(t)\}dt, \quad (20)$$

در نظر گرفته شده است، که در آن متغیرهای X_1 و X_2 از توزیع نرمال استاندارد تولید شده و سایر متغیرها مانند مثال اول در نظر گرفته شده‌اند.

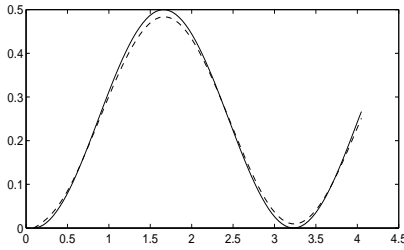
با توجه به جدول ۳، استفاده از دو هسته نرمال و اپانچنیکوف، برآوردهایی نااریب برای پارامتر مستقل از زمان ایجاد کرده‌اند که قدرت مدل نیمه پارامتری معرفی شده را برای برآورد پارامتر مستقل از زمان هنگامی که بیش از یک متغیر وابسته به زمان وجود دارد، نشان می‌دهد. مقادیر کوچک ISB و MISE برای هر دو متغیر وابسته به زمان در جدول ۴، حاکی از کارآمد بودن روش برآورد در مدل پیشنهادی است. این نتیجه از اهمیت بسزایی برخوردار است، زیرا در واقعیت ممکن است تعداد متغیرهای وابسته به زمان بیش از یکی باشند.

جدول ۳: مقادیر اریبی، خطای استاندارد مجانبی و تجربی و نرخ پوشش برای پارامتر مستقل از زمان به ازای $\gamma_0 = 0.5$

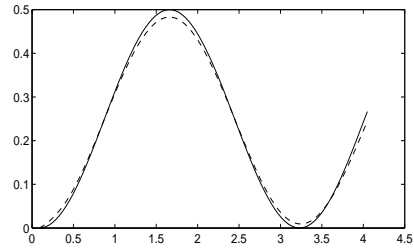
هسته	n	h	Bias	SEE	SSE	CP
پانچنیکوف	۱۰۰	۰/۴	۰/۰۰۹۰	۰/۱۴۱۲	۰/۱۴۴۷	۰/۹۳۵
	۰/۵	۰/۵	۰/۰۰۸۷	۰/۱۴۱۶	۰/۱۳۹۳	۰/۹۴۰
	۰/۶	۰/۶	۰/۰۰۶۱	۰/۱۳۹۰	۰/۱۲۸۸	۰/۹۶۵
	۰/۷	۰/۷	-۰/۰۰۹۶	۰/۱۴۰۲	۰/۱۳۷۸	۰/۹۵۵
۲۰۰	۰/۴	۰/۴	۰/۰۰۸۳	۰/۱۰۰۱	۰/۰۹۷۰	۰/۹۶۵
	۰/۵	۰/۵	-۰/۰۰۴۳	۰/۰۹۹۴	۰/۰۹۹۶	۰/۹۶۵
	۰/۶	۰/۶	-۰/۰۰۷۳	۰/۰۹۸۳	۰/۱۰۳۳	۰/۹۵۰
	۰/۷	۰/۷	۰/۰۹۸۰	۰/۰۹۶۶	۰/۰۹۶۶	۰/۹۶۵
نرمال	۱۰۰	۰/۴	۰/۰۱۹۹	۰/۱۴۰۴	۰/۱۴۷۰	۰/۹۱۰
	۰/۵	۰/۵	-۰/۰۰۴۱	۰/۱۳۹۲	۰/۱۴۵۷	۰/۹۱۵
	۰/۶	۰/۶	-۰/۰۰۱۳۹	۰/۱۳۸۹	۰/۱۳۷۸	۰/۹۶۵
	۰/۷	۰/۷	-۰/۰۰۵۶۶	۰/۱۴۰۳	۰/۱۵۲۶	۰/۹۳۰
۲۰۰	۰/۴	۰/۴	-۰/۰۰۹۰	۰/۰۹۸۲	۰/۰۹۷۰	۰/۹۵۰
	۰/۵	۰/۵	-۰/۰۰۲۲	۰/۰۹۷۹	۰/۰۹۹۵	۰/۹۳۵
	۰/۶	۰/۶	۰/۰۰۲۵	۰/۰۹۸۱	۰/۰۹۸۲	۰/۹۶۵
	۰/۷	۰/۷	۰/۰۱۰۸	۰/۰۹۷۶	۰/۰۹۸۶	۰/۹۵۰

جدول ۴: مقادیر اریبی جمع بسته توان دوم و میانگین جمع بسته توان دوم خطاها برای ضرایب $\hat{B}_1(t)$ و $\hat{B}_2(t)$

$\hat{B}_1(t)$		$\hat{B}_2(t)$		h	n	هسته
ISB	MISE	ISB	MISE			
۰/۰۰۲۰	۰/۳۳۶۱	۰/۰۲۵۲	۰/۳۰۷۸	۰/۴	۱۰۰	پانچنیکوف
۰/۰۰۵۱	۰/۳۰۲۲	۰/۰۵۸۲	۰/۳۴۵۲	۰/۵		
۰/۰۰۴۲	۰/۳۴۳۰	۰/۰۶۵۹	۰/۳۲۶۳	۰/۶		
۰/۰۰۷۷	۰/۲۸۵۲	۰/۰۵۱۳	۰/۳۳۰۹	۰/۷		
۰/۰۰۱۰	۰/۱۵۱۴	۰/۰۲۶۴	۰/۱۶۴۶	۰/۴	۲۰۰	
۰/۰۰۱۵	۰/۱۶۲۰	۰/۰۳۳۲	۰/۱۶۲۹	۰/۵		
۰/۰۰۴۴	۰/۱۲۰۸	۰/۰۶۸۳	۰/۱۹۸۲	۰/۶		
۰/۰۰۵۱	۰/۱۳۴۰	۰/۰۶۹۸	۰/۱۷۸۷	۰/۷		
۰/۰۱۲۱	۰/۲۹۶۰	۰/۰۳۴۷	۰/۳۵۰۹	۰/۴	۱۰۰	نرمال
۰/۰۱۸۸	۰/۲۶۲۷	۰/۰۶۰۲	۰/۳۴۳۵	۰/۵		
۰/۰۰۵۰	۰/۳۴۳۳	۰/۰۸۰۴	۰/۳۴۲۹	۰/۶		
۰/۰۰۷۸	۰/۲۶۴۹	۰/۰۷۷۳	۰/۳۳۳۰	۰/۷		
۰/۰۰۲۲	۰/۱۳۱۱	۰/۰۲۴۶	۰/۱۴۹۱	۰/۴	۲۰۰	
۰/۰۰۵۵	۰/۱۶۰۱	۰/۰۳۸۲	۰/۱۷۸۸	۰/۵		
۰/۰۰۵۲	۰/۱۳۶۲	۰/۰۹۱۸	۰/۲۱۷۲	۰/۶		
۰/۰۰۷۴	۰/۱۲۷۶	۰/۱۰۸۵	۰/۲۱۷۹	۰/۷		

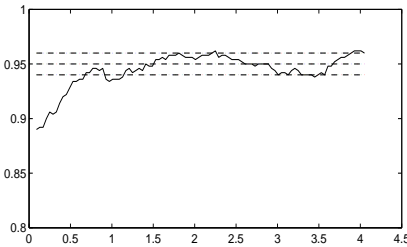


(ب)

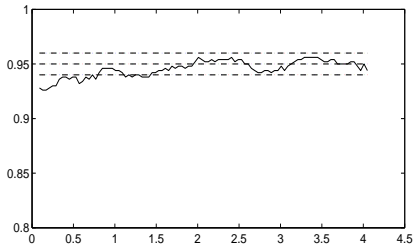


(الف)

شکل ۱: نمودارهای منحنی ضریب رگرسیونی تجمعی و برآورد آن‌ها: الف) برای $n = 100$ (ب) برای $n = 200$



(ب)



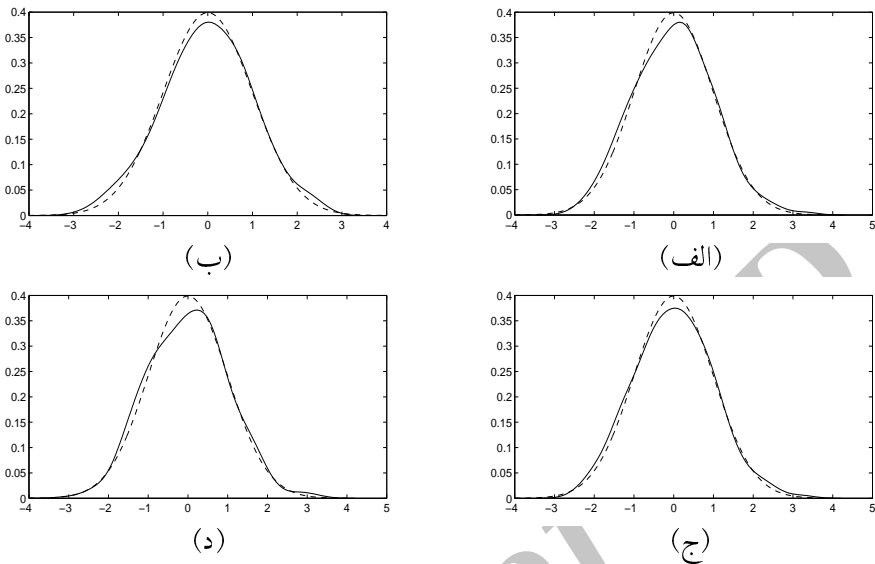
(الف)

شکل ۲: احتمال پوشش برای برآورد ضریب رگرسیونی تجمعی: الف) برای $n = 100$ (ب) برای $n = 200$

۵ مثال کاربردی

بیماری‌های قلبی عروقی، علت اصلی مرگ و میر در سراسر جهان هستند که هر ساله ۱۷/۳ میلیون نفر را به کام مرگ می‌فرستند و اگر تا سال ۲۰۳۰ اقدام پیشگیرانه‌ای نشود، این تعداد به ۲۳/۳ میلیون نفر خواهند رسید (مدرس و لونکار، ۲۰۰۶). یکی از درمان‌های بیماران قلبی پرخطر و به‌ویژه نارسایی قلبی، تعبیه دستگاه کنترل کننده ضربان قلب دفیبریلاتور کاشتنی قلبی^۸ (ICD) برای پیشگیری از مرگ ناگهانی است. ICD در صورتی که تشخیص دهد منشا آریتمی، بطنی است، با دادن شوک آن را خاتمه خواهد داد (شوک بجای)، و در صورتی که منشا آن را

^۸ Implantable Cardioverter Defibrillator



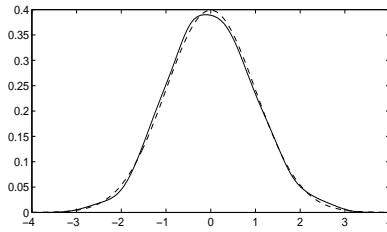
شکل ۳: نمودارهای چگالی ضریب رگرسیونی تجمعی در زمان‌های: الف) هشتم (ب) چهارم (ج) هشتماد و (د) صدم

فوق بطنی تشخیص دهد، اقدامی برای قطع آن نخواهد کرد. در صورتی که یک تاکی کاردی^۹ فوق بطنی توسط ICD به اشتباه تاکی کاردی بطنی تشخیص داده شود و منجر به تخلیه شوک شود، به آن شوک نابجا گفته می‌شود.

مطالعات متعددی نشان داده‌اند که مدت QRS (زمان برای دیپلاریزاسیون بطنی^{۱۰}) طولانی، همراه با افزایش خطر آریتمی و مرگ ناگهانی قلبی است. بنابراین به نظر می‌رسد یکی از پیشگویی‌های مناسب برای نرخ شوک، مدت زمان QRS است. برای بیماران تحت مطالعه، مدت زمان QRS در نوار قلبی بیماران قبل از تعبیه ICD با دقت یک هزارم میلی‌متر با استفاده از کولیس دیجیتال اندازه‌گیری گردید. برای افزایش دقت مطالعه، در دو نوبت نوار فوق اندازه‌گیری و میانگین دو نوبت ثبت گردید. بعلاوه سن بیماران و نوع دستگاه ICD برای هر بیمار نیز ثبت شدند. پس از تعبیه ICD، بیماران به‌طور منظم تحت پیگیری یک، سه و شش ماهه قرار

^۹ Tachycardia

^{۱۰} Ventricular depolarization



شکل ۴: نمودار چگالی ضریب رگرسیونی مستقل از زمان

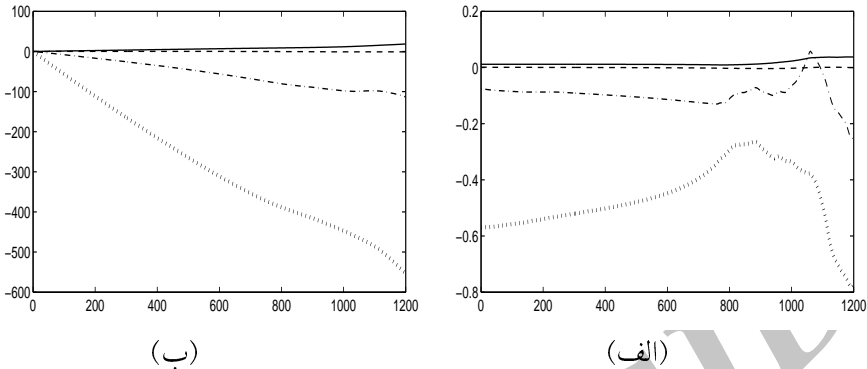
گرفتند و تعداد و زمان شوک‌های واردشده به آن‌ها ثبت شدند. میانه زمان پیگیری بیماران ۹۰ روز است.

افرادی که زمان دریافت شوک آن‌ها مقادیری گمشده بود، از مطالعه خارج شدند و در نتیجه تعداد بیماران به ۴۴ نفر تقلیل یافت. سه متغیر سن، QRS و نوع دستگاه به عنوان عوامل موثر بر نرخ شوک در نظر گرفته شدند. برای تعیین چگونگی تاثیر سه نوع دستگاه ICD بر نرخ شوک، متغیر ظاهری X_i برابر ۱ تعریف شد، اگر دستگاه نوع اول باشد و متغیر Z_i برابر ۱ تعریف شد، اگر دستگاه نوع دوم باشد. متغیرهای W_i و V_i نیز به ترتیب متغیرهای سن و QRS انتخاب شدند. برای تعیین وابسته یا مستقل از زمان بودن متغیرها، ابتدا همه متغیرها را وابسته به زمان تعریف نموده و از مدل

$$E\{dN_i^*(t)|X_i, Z_i, W_i, V_i\} = \exp\{\beta_1(t)X_i + \beta_2(t)Z_i + \beta_3(t)W_i + \beta_4(t)V_i\}d\mu_0(t), \quad (21)$$

استفاده شد. همچنین، با استفاده از هسته اپانچنیکوف، پهنای نوار برابر $h = 0.76$ روز انتخاب شد. برای آزمون وابسته به زمان بودن این چهار متغیر، آماره‌های دو آزمون F_1 و F_2 برای $B_1(t)$, $B_2(t)$, $B_3(t)$ و $B_4(t)$ با استفاده از روش لین و همکاران (۲۰۰۰) بر مبنای ۱۰۰۰ تکرار، محاسبه شد. p -مقدارهای هر دو آزمون برای هر چهار متغیر، کمتر از ۰/۰۰۲ به دست آمدند. بنابراین اثر همه متغیرهای تبیینی، وابسته به زمان شناخته شدند. شکل ۵، ضرایب $\beta_0(t)$ و $B_0(t)$ را در مقابل

^{۱۱} Dummy variable



شکل ۵: الف) ضرایب رگرسیونی $\beta_0(t)$ ب) ضرایب رگرسیونی تجمعی $B_0(t)$ ICD1 (نقطه چین)، ICD2 (نقطه-خطچین)، QRS (خطچین)، سن (توپر).

زمان نشان می‌دهد، که در آن ICD1 نشان‌دهنده ضریب رگرسیونی متغیر X و ICD2 ضریب Z است. با توجه به شکل ۵-الف، دستگاه نوع اول نسبت به نوع سوم، به‌طور متوسط، دارای نرخ شوک کمتری است، اما اختلاف این نرخ بین دو دستگاه تا ۱۸۰۰امین روز، به‌طور پیوسته، در حال کاهش است و از روز ۱۸۰۰ به بعد این اختلاف افزایش می‌یابد. بنابراین، دستگاه نوع سوم نسبت به اول همواره نرخ شوک بالاتری داراست. برای دستگاه نوع دوم نسبت به نوع سوم نیز تقریباً نتیجه مشابهی ملاحظه می‌شود. البته اختلاف بین نرخ شوک در این دو دستگاه، نسبت به دستگاه‌های نوع اول و سوم، خیلی کمتر است و حتی در بازه کوتاهی، نرخ شوک دستگاه نوع دوم بیشتر از نوع سوم برآورد شده است. تغییرات دستگاه نوع اول نسبت به نوع سوم، در طول زمان، بسیار بیشتر از تغییرات دستگاه نوع دوم نسبت به نوع سوم است. یعنی دستگاه نوع اول بسیار متفاوت‌تر از دو نوع دیگر عمل می‌کند و با توجه به نمودار ۵-الف، این دستگاه تعداد شوک کمتری نسبت به دو نوع دیگر وارد می‌کند. این امر می‌تواند نشان از آن باشد که برای بیماران تحت مطالعه، این دستگاه تعداد شوک نابجای کمتری نسبت به دو نوع دیگر بر بیماران وارد می‌کند.

متغیر سن تا قبل از روز ۱۹۰۰ تقریباً ثابت است و پس از آن تا پایان مطالعه اثری افزایشی در نرخ دریافت شوک دارد. بنابراین با افزایش زمان و در نتیجه

افزایش سن، نرخ شوک دریافتی بیشتر می‌شود. متغیر QRS نیز، علی‌رغم آن‌که وابسته به زمان بودن اثرش معنی‌دار تشخیص داده شد، تغییر کمی در طول زمان دارد. این بدان معناست که اختلال در مدت زمان طی شده یک تحریک الکتریکی درون قلب که با QRS نشان داده می‌شود، در طول زمان، تغییرپذیری اندکی دارد. این نتیجه با واقعیت تطابق دارد زیرا در واقعیت، در این‌گونه بیماران، وقتی گره^{۱۲}های مولد تپش کارایی خود را از دست می‌دهند، این تغییر در مدت QRS مشاهده می‌شود و طبیعی است که در طول زمان، این گره‌ها به‌طور طبیعی بازسازی نمی‌شوند و فقط ممکن است که در گذشت زمان، این نقص تشدید شود و بنابراین اثر این متغیر تغییر کمتری نسبت به بقیه متغیرها در طول زمان دارد.

بحث و نتیجه‌گیری

مطالعه شبیه‌سازی و مقادیر کوچک اریبی و خطاها، بیان‌گر عموماً خوب برآوردگرهای پیشنهادی است. پایداری برآوردگرهای $\hat{B}(t)$ وقتی بیش از یک متغیر وابسته به زمان وجود دارد، گاهی دچار خدشه می‌شود. در این حالت، اریبی برآوردگرها کمی بیشتر می‌شود اما با افزایش حجم نمونه، مقدار آن قابل صرف‌نظر کردن است. استفاده از هسته‌های متفاوت نیز در موقعیت‌های مختلف، نتایج یکسانی را دربر دارد. برآوردگرهای ضرایب وابسته به و مستقل از زمان در مدل پیشنهادی از عملکرد مطلوبی برخوردارند. رفتار برآوردگرها را می‌توان با به‌دست آوردن توزیع مجانبی آن‌ها بررسی کرد. آزمون‌های تشخیص وابستگی به زمان ضرایب متغیرهای تبیینی، با کاهش حجم نمونه از توان کمتری برخوردارند. بنابراین، به دلیل کوچک بودن حجم داده‌های مربوط به بیماران قلبی، ممکن است اثر برخی از متغیرها، مانند QRS، به اشتباه وابسته به زمان معرفی شده باشند. وقتی ناهم‌واری ضرایب رگرسیونی زیاد است، استفاده از اسپلاین‌ها به جای هسته‌ها به عنوان رهیافتی جان‌شین توصیه می‌شود (هستی و همکاران، ۲۰۰۹). همچنین مدل‌بندی پیشامدهای بازگردنده با مدل‌های نیمه پارامتری بقا از دیدگاه بیزی می‌تواند جالب توجه باشد.

تقدیر و تشکر

نویسندگان مقاله از پیشنهادات ارزنده داوران محترم و ویراستار مجله که منتهی به بهبود مقاله شد، کمال تشکر را دارند. همچنین از آقایان دکتر علیرضا حیدری بکاولی، دکتر حسین نظری حیانو و وحید قوامی قنبرآبادی به خاطر در اختیار قرار دادن مجموعه داده بیماران قلبی بیمارستان قائم (عج) مشهد، بسیار سپاسگزارند.

مراجع

بهبودیان، ج. (۱۳۸۷)، روش‌های ناپارامتری، چاپ ششم، انتشارات دانشگاه پیام نور، تهران.

Andersen, P. K. and Gill, R. D. (1982), Cox's Regression Model for Counting Processes: A Large Sample Study, *The Annals of Statistics*, **10**, 1100-1120.

Cook, R. J. and Lawless, J. F. (2007), *The Statistical Analysis of Recurrent Event*, First Edition, Springer, New York.

Cox, D. R. (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society*, B, **34**, 187-220.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *Elements of Statistical Learning: Data Mining, Inference and Prediction*, Second Edition, Springer, New York.

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*, B, **62**, 711-730.

Martinussen, T. and Scheike, T. H. (2006), *Dynamic Regression Models for Survival Data*, First Edition, Springer, New York.

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.

Mathers, C. D. and Loncar, D. (2006), Projections of Global Mortality and Burden of Disease from 2002 to 2030, *PLOS Medicine*, **3**, 2011-2030.

Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981), On the Regression Analysis of Multivariate Failure Time Data, *Biometrika*, **68**, 373-379.

Rizzo, M. L. (2008), *Statistical Computing with R*, First Edition, Chapman & Hall/CRC, London.

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.

Stephens, M. A. (1986), *Tests Based On EDF Statistics*, Marcel Dekker, New York.

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.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989), Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions, *Journal of the American Statistical Association*, **84**, 1065-1073.