

مجله علوم آماری، پاییز و زمستان ۱۴۰۰

جلد ۱۵، شماره ۲، ص ۵۰۵ - ۵۳۱

DOI: 10.29252/jss.15.2.505

مقاله پژوهشی

مدل‌بندی پاسخ‌های طولی آمیخته دو متغیره پیوسته-شمارشی و پیوسته-ترتیبی آماسیده با امکان گم‌شدگی غیر تصادفی

نسترن شریفیان، احسان بهرامی سامانی

گروه آمار، دانشکده علوم ریاضی، دانشگاه شهید بهشتی، تهران

تاریخ دریافت: ۱۳۹۹/۰۴/۲۶ تاریخ پذیرش و انتشار: ۱۳۹۹/۱۲/۲۴

چکیده: یکی از مسائل مهم مطالعات طولی، داده‌هایی هستند که به دلیل گم‌شدن ملاقات‌ها یا انصراف از دست می‌روند. در چنین حالاتی آزمودنی‌ها مجموعه زمان‌های مشاهده یکسانی ندارند. مسئله گم‌شدگی در تحلیل داده‌های آمیخته گسسته و پیوسته طولی نیز بسیار رایج است و ممکن است گم‌شدگی در یکی از پاسخ‌ها و یا هر دو رخ دهد. عدم توجه به دلیل گم‌شدن (سازوکار گم‌شدگی) منجر به برآوردها و استنباط‌های اریب می‌شود. بنابراین در این مقاله به بررسی سازوکار گم‌شدن غیرقابل چشم‌پوشی در پاسخ‌های آمیخته طولی مجموعه آماسیده پیوسته و آماسیده در صفر سری توانی و همچنین پاسخ‌های آمیخته پیوسته و k -آماسیده ترتیبی پرداخته می‌شود. با یک رویکرد درستی مبنای برآوردهای مدل‌های پیشنهادی را به دست می‌آید. عملکرد مدل‌ها را با مطالعات شبیه‌سازی و دو مجموعه داده واقعی تغییرات زندگی آمریکایی‌ها و آزمون موفقیت فردی پیبادی مورد بررسی قرار می‌گیرد.

واژه‌های کلیدی: داده‌های آمیخته طولی، آماسیدگی، اثرهای تصادفی، مقادیر گم‌شده.

۱ مقدمه

یکی از مسائل مهم و رایج در مسائل علمی از قبیل روان‌شناسی، اقتصاد، سلامت و پزشکی وجود پاسخ‌های گسسته و پیوسته برای هر واحد آزمایشی است. در سال‌های اخیر داده‌های آمیخته پیوسته و دودویی، داده‌های آمیخته پیوسته و ترتیبی، داده‌های آمیخته پیوسته و شمارشی و همچنین داده‌های آمیخته شمارشی و ترتیبی توسط تعدادی از محققان مورد مطالعه قرار گرفته است. یکی از رهیافت‌های مدل‌بندی توأم پاسخ‌های آمیخته همبسته، استفاده از اثرهای تصادفی مشترک یا همبسته و به عبارتی استفاده از رویکرد متغیر پنهان در مدل‌های هم‌زمان پاسخ‌ها برای به حساب آوردن همبستگی بین پاسخ‌های آمیخته است (مک‌کالوک، ۲۰۰۷؛ بهرامی سامانی و همکاران، ۲۰۰۸؛ تِکسیرا - پینتو و نُرمند، ۲۰۰۹؛ بهرامی سامانی و طهماسبی‌نژاد، ۲۰۱۱؛ بهرامی سامانی، ۲۰۱۳؛ شریفیان و همکاران، ۲۰۲۰).

در حالت کلی در مطالعات طولی، حتی در موقعیت‌هایی که به خوبی کنترل می‌شوند، داده‌های گم‌شده همواره رخ می‌دهند. لیتل و روبین (۲۰۰۲) با معرفی سازوکارهای گم‌شدگی، یک تابع درست‌نمایی برای داده‌های آمیخته پیوسته و گسسته با مقادیر گم‌شده ارائه کردند. یانگ و کانگ (۲۰۱۰) مدل‌های رگرسیونی برای تحلیل توأم داده‌های طولی پواسون و پیوسته با مقادیر گم‌شدگی غیرقابل چشم‌پوشی را تحت چارچوب کاملاً پارامتری معرفی کردند. بهرامی سامانی و همکاران (۲۰۱۰) مسئله گم‌شدگی در پاسخ‌های آمیخته طولی را با مدل‌های اثرهای تصادفی به دو صورت کلاسیک و بیزی مورد مطالعه قرار دادند. میری و بهرامی سامانی (۱۳۹۱) یک مدل تعمیم یافته برای توزیع توأم متغیرهای اسمی، ترتیبی و پیوسته در دو حالت داده‌های کامل و داده‌های گم‌شده معرفی کردند. فرم‌های بسته‌ای برای تابع درست‌نمایی مربوط به مدل‌های مکانی عام ارائه می‌شود. اولین روشی که با استفاده از رویکرد پارامتری برای داده‌های طولی ترتیبی و دوجمله‌ای منفی معرفی شد، توسط بهرامی سامانی و گنجعلی (۲۰۱۴) ارائه و به حالت مقادیر گم‌شده غیرقابل چشم‌پوشی در هر نوع پاسخ تعمیم داده شد و رضیئی و همکاران (۲۰۱۷) مدل آن‌ها را به حالت داده‌های آمیخته سری توانی و ترتیبی در حالت طولی و گم‌شدگی غیرقابل چشم‌پوشی، با رویکرد متغیر پنهان تعمیم دادند.

مدل رگرسیون پواسون آماسیده در صفر برای داده‌های شمارشی شامل تعداد صفرهای زیاد، نخستین بار توسط لمبرت (۱۹۹۲) معرفی شد. کلی و اندرسون (۲۰۰۸) یک مدل بخت‌های متناسب آماسیده در صفر را برای یک پاسخ ترتیبی بسط دادند. بهرامی سامانی و همکاران (۲۰۱۲) مدل رگرسیون سری توانی آماسیده صفر را برای تحلیل داده‌های شمارشی طولی با صفرهای زیاد، با و بدون داده‌های گم‌شده مورد استفاده قرار دادند. اسماعیل‌زاده و بهرامی سامانی (۱۳۹۸) داده‌های آمیخته شمارشی و پیوسته

دومتغیره‌ای را مورد تحلیل قرار دادند که متغیر پاسخ شمارشی دومتغیره در یک یا دو نقطه آماسیده شده است. شریفیان و همکاران (۲۰۲۱) متغیرهای مجموعه آماسیده پیوسته را معرفی کردند. آن‌ها با مدل‌بندی همزمان، متغیرهای مجموعه آماسیده پیوسته و مجموعه آماسیده شمارشی همبسته طولی را با استفاده از رویکرد اثرهای تصادفی همبسته یا مشترک مورد بررسی قرار دادند. بررسی و مطالعه آماسیدگی در پاسخ‌های طولی ترتیبی نیز توسط شریفیان و همکاران (۲۰۲۰) مورد مطالعه و بررسی قرار گرفت. ایشان با الگو گرفتن از مطالعه انجام شده توسط کیلی و اندرسون (۲۰۰۸) یک متغیر پاسخ طولی ترتیبی که در رده k ام آماسیده است را به همراه یک متغیر پاسخ همبسته پیوسته با استفاده از مدل‌های متناسب k -آماسیده^۱ ($k-IPO$) برای پاسخ ترتیبی و مدل اثرهای آمیخته برای پاسخ پیوسته با رویکرد اثرهای تصادفی همبسته به طور همزمان مدل‌بندی کردند. اجرای مدل‌های داده‌های طولی می‌تواند به سازوکار گم‌شدن داده‌ها وابسته باشد؛ بنابراین بررسی سازوکار گم‌شدگی امری ضروری در انتخاب یک تحلیل مناسب است. سازوکارهای گم‌شدگی داده‌ها عبارت است از گم‌شدگی کاملاً تصادفی^۲ (MCAR)، گم‌شدگی تصادفی^۳ (MAR) و گم‌شدگی غیر تصادفی^۴ (MNAR). در گم‌شدن غیرتصادفی بردار متغیر وابسته مشاهده‌نشده با گم‌شدگی در ارتباط است و از آن به عنوان گم‌شدگی غیرقابل چشم‌پوشی نام برده می‌شود.

در این مقاله به بررسی سازوکار گم‌شدگی غیرقابل چشم‌پوشی در پاسخ‌های آمیخته مجموعه آماسیده پیوسته و آماسیده در صفر شمارشی و پاسخ‌های آمیخته پیوسته و k -آماسیده ترتیبی پرداخته می‌شود. در این راستا، بخش ۲ شامل مدل‌های مورد بررسی برای داده‌های آمیخته با امکان گم‌شدگی غیرتصادفی، تابع درستنمایی بر اساس متغیرهای پاسخ و متغیرهای نشانگر گم‌شدگی، برآورد پارامترهای مدل و توضیح مختصری در رابطه با شناساپذیری مدل‌ها است. در بخش ۳ عملکرد مدل‌ها با استفاده از مطالعات شبیه‌سازی مورد بررسی قرار می‌گیرد. در بخش ۴، نوع گم‌شدگی در پاسخ‌های آمیخته مورد نظر در دو مجموعه داده واقعی تغییرات زندگی آمریکایی‌ها^۵ (ACL) و آزمون موفقیت فردی پیبادی^۶ (PIAT) بررسی می‌شود. در نهایت نتیجه‌گیری حاصل از مطالعه گم‌شدگی غیرتصادفی در مدل‌های مورد بررسی ارائه می‌شود.

¹k-Inflated Proportional odds

²Missing Completely at Random

³Missing at Random

⁴Missing Not at Random

⁵American's Changing Lives

⁶Peabody Individual Achievement Test

۲ مدل‌ها و درست‌نمایی

در این بخش امکان گم‌شدگی تصادفی در پاسخ‌های آمیخته طولی مجموعه آماسیده پیوسته و آماسیده در صفر سری‌های توانی^۷ ($AIN - ZIPS$) و (۲) نرمال و k -آماسیده ترتیبی^۸ ($N - kIO$) با مدل‌بندی همزمان پاسخ‌ها و متغیرهای نشانگر گم‌شدگی تعریف‌شده، مورد بررسی قرار می‌گیرد. به این منظور مدل‌های رگرسیونی تعمیم‌یافته آمیزه‌ای متناهی^۱ ($GRFMM$) با اثرهای تصادفی، با وارد کردن سازوکار گم‌شدگی در ساختار مدل ارائه می‌شود. در مسائل گم‌شدگی، روش‌های پیشینه درست‌نمایی قابلیت حذف یا کاهش آریبی برآوردهایی را که فقط بر اساس داده‌های مشاهده‌شده (داده‌های در دسترس) به دست آمده‌اند را دارند. از این رو برای به دست آوردن پارامترها در حالت داده‌های گم‌شده مورد توجه هستند (لیتل و اسکلوچر، ۱۹۸۵). روش‌های مختلفی برای دست‌یابی به برآوردهای پیشینه درست‌نمایی در مسائل گم‌شدگی در نوشتگان آماری مطرح و استفاده شده است. یکی از آنان به کارگرفتن روش متداول درست‌نمایی و به دست آوردن برآوردها به طور مستقیم با پیشینه کردن تابع درست‌نمایی است (دی‌لئون و چاف، ۲۰۱۳؛ بهرامی سامانی و گنجعلی، ۲۰۱۰؛ رضیعی و همکاران، ۲۰۱۹). لذا، تابع درست‌نمایی برای داده‌های مورد بررسی به منظور به دست آوردن برآوردهای پیشینه درست‌نمایی پارامترهای مدل به دست می‌آید. فرض کنید Y_{it} و N_{it} به ترتیب پاسخ‌های طولی مجموعه آماسیده پیوسته (AIN) و آماسیده در صفر سری توانی برای آزمودنی i ام، $i = 1, \dots, m$ ، در زمان $t = 1, \dots, T_i$ ، باشند. مجموعه‌ای که آماسیدگی پاسخ پیوسته در آن بررسی می‌شود، A ، به عنوان مثال می‌تواند یک بازه روی محور اعداد حقیقی باشد. همچنین Y_{it} و \tilde{Y}_{it} به ترتیب پاسخ‌های طولی پیوسته و k -آماسیده ترتیبی برای آزمودنی i ام، $i = 1, \dots, m$ ، در زمان $t = 1, \dots, T_i$ ، باشند. برای مدل‌بندی توأم این پاسخ‌های طولی همبسته فرض کنید برای $k = 1, 2, 3$ بردارهای ثابتی از متغیرهای کمکی $1 \times r_k$ بُعدی برای i مین آزمودنی در زمان t و W_{kit} زیربردارهایی $1 \times s_k$ بُعدی از X_{kit} ها باشند. همچنین X_{kit}^* ، $k = 1, 2$ ، نیز بردارهای ثابتی از متغیرهای کمکی $1 \times r_k^*$ بُعدی برای i مین آزمودنی در زمان t ، و W_{kit}^* زیربردارهایی $1 \times s_k^*$ بُعدی از X_{kit}^* ها باشند. فرض کنید $R_{Y_{it}}$ ، $R_{N_{it}}$ و $R_{\tilde{Y}_{it}}$ متغیرهای نشانگر مربوط به سازوکارهای گم‌شدگی باشند که اگر متغیرهای Y_{it} ، N_{it} و \tilde{Y}_{it} مشاهده شده باشند مقدار ۱ را اختیار می‌کنند، در غیر این صورت برابر با صفر هستند. M_{1it} و M_{2it} بردارهای خاصی از متغیرهای کمکی برای سازوکارهای گم‌شدگی هستند.

⁷ A-Inflated Normal-Zero Inflated Power Series

⁸Normal-k Inflated Ordinal

¹Generalized Regression Finite Mixture Model

۲.۱ مدل برای داده‌های $AIN - ZIPS$ با امکان گم‌شدگی غیر تصادفی

فرض می‌شود که $Y_{it}|b_i^{(1)}, b_i^{(2)}, b_i^{(3)}$ و $N_{it}|b_i^{(4)}, b_i^{(5)}$ به ترتیب دارای یک توزیع مجموعه آماسیده نرمال (AIN) و یک توزیع آماسیده در صفر سری توانی ($ZIPS$) هستند که $b_i^{(1)}, b_i^{(2)}, b_i^{(3)}, b_i^{(4)}, b_i^{(5)}$ و بردارهایی از اثرهای تصادفی آزمودنی-مشخص هستند. مدل توأم برای پاسخ‌های آمیخته طولی مجموعه آماسیده پیوسته، Y_{it} ، و آماسیده در صفر سری توانی، N_{it} ، وقتی که امکان گم‌شدگی غیرتصادفی وجود دارد به صورت

$$\begin{aligned} Y_{it}|b_i^{(1)}, b_i^{(2)}, b_i^{(3)}, \theta_{A_{it}}, \mu_{it}, \pi_{A_{it}} &\sim AIN(\theta_{A_{it}}, \mu_{it}, \sigma^2, \sigma_*^2, \pi_{A_{it}}), \\ q_1(\mu_{it}) &= \mathbf{X}'_{vit}\beta_t^{(1)} + \mathbf{W}'_{vit}b_i^{(1)}, \\ q_2(\theta_{A_{it}}) &= \mathbf{X}'_{vit}\beta_t^{(2)} + \mathbf{W}'_{vit}b_i^{(2)}, \\ q_3(\pi_{A_{it}}) &= \mathbf{X}'_{vit}\beta_t^{(3)} + \mathbf{W}'_{vit}b_i^{(3)}, \\ \text{logit}(Pr(R_{y_{it}} = 1|M_{vit}, y_{it})) &= \mathbf{M}'_{vit}\alpha_t^{(1)} + \alpha_{vt}^{(1)}y_{it}, \\ N_{it}|b_i^{(4)}, b_i^{(5)}, \nu_{it}, \pi_{\circ it} &\sim ZIPS(\nu_{it}, \pi_{\circ it}), \\ q_1^*(\nu_{it}) &= \mathbf{X}'_{vit}\gamma_t^{(1)} + \mathbf{W}'_{vit}b_i^{(4)}, \\ q_2^*(\pi_{\circ it}) &= \mathbf{X}'_{vit}\gamma_t^{(2)} + \mathbf{W}'_{vit}b_i^{(5)}, \\ \text{logit}(Pr(R_{n_{it}} = 1|M_{vit}, n_{it})) &= \mathbf{M}'_{vit}\alpha_t^{(2)} + \alpha_{vt}^{(2)}n_{it}. \end{aligned} \quad (1)$$

است، که در آن $\theta_{A_{it}}$ و μ_{it} پارامترهای متغیر پیوسته، ν_{it} پارامتر متغیر شمارشی، و $\pi_{A_{it}}$ و $\pi_{\circ it}$ به ترتیب پارامترهای آمیزنده متغیر پیوسته و شمارشی هستند. $q_1, q_2, q_3, q_1^*, q_2^*$ به ترتیب توابع ربط معلوم برای $\mu_{it}, \theta_{A_{it}}, \pi_{A_{it}}, \nu_{it}$ و $\pi_{\circ it}$ هستند. معمولاً q_1 و q_2 به ترتیب تابع ربط همانی برای μ_{it} و تابع ربط لوجیت روی یک تبدیل مکان-مقیاس برای $\theta_{A_{it}}$ ، q_1^* تابع ربط لگاریتمی برای ν_{it} ، و q_2^* و q_1^* به ترتیب تابع ربط لوجیت برای $\pi_{A_{it}}$ و $\pi_{\circ it}$ هستند. توابع ربط دیگری نیز مانند تابع پروبیت یا تابع لگ-لگ مکمل می‌توانند برای $\pi_{A_{it}}$ و $\pi_{\circ it}$ استفاده شوند. پارامترهای $\beta_t^{(1)}, \beta_t^{(2)}, \beta_t^{(3)}, \gamma_t^{(1)}, \gamma_t^{(2)}$ و ضرایب رگرسیونی هستند. پارامترهای $\alpha_{vt}^{(1)}$ و $\alpha_{vt}^{(2)}$ در بردارنده اطلاعات مربوط به سازوکار گم‌شدگی برای پاسخ‌ها در زمان t هستند به طوری که تایید صفر بودن هر دوی این ضرایب نشان می‌دهد که سازوکار گم‌شدگی تصادفی (MAR) و تایید صفر بودن مقدار یکی از آن‌ها حالتی را نشان می‌دهد که سازوکار گم‌شدگی غیرتصادفی است.

دهد که سازوکار گم‌شدگی غیر تصادفی (NMAR) است. برای مدل‌بندی طبیعت طولی پاسخ‌ها، یعنی وابستگی درون-آزمودنی در طول زمان، از بردارهای اثرهای تصادفی $b_i^{(1)}$ ، $b_i^{(2)}$ (برای پاسخ پیوسته) و $b_i^{(4)}$ (برای پاسخ شمارشی) استفاده شده است. علاوه بر این به شرط اثرهای تصادفی بردار پاسخ‌ها در موقعیت‌های مختلف برای هر آزمودنی مستقل شرطی هستند. بردارهای اثرهای تصادفی $b_i^{(2)}$ و $b_i^{(5)}$ برای وابستگی پارامترهای آمیزنده در طول زمان در نظر گرفته می‌شوند.

برای مدل‌بندی توأم پاسخ‌های آمیخته، با وارد کردن تعدادی اثرهای تصادفی همبسته به مدل (۱)، همبستگی بین Y_{it} و N_{it} در نظر گرفته شده است. فرض کنید $(b_i^{(1)'}, b_i^{(2)'}, b_i^{(r)'})$ ، $v_i^{(1)'}$ و $v_i^{(2)'}$ فرض می‌شود که Σ_j و G به ترتیب توسط λ_j و λ_G پارامتریده می‌شوند. علاوه بر این، $v_i^{(1)}$ و $v_i^{(2)}$ دارای توزیع گاوسی با میانگین 0 و ماتریس کوواریانس $\Sigma_{v_i^{(1)}, v_i^{(2)}} = \begin{pmatrix} \Sigma_{v_i^{(1)}} & G \\ G & \Sigma_{v_i^{(2)}} \end{pmatrix}$ هستند که $\Sigma_{v_i^{(1)}} = \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \Sigma_r & 0 \\ 0 & 0 & \Sigma_r \end{pmatrix}$ و $\Sigma_{v_i^{(2)}} = \begin{pmatrix} \Sigma_r & 0 \\ 0 & \Sigma_5 \end{pmatrix}$. بردارهای پاسخ برای i امین آزمودنی، $Y_i = (Y_{i1}, \dots, Y_{iT_i})'$ و $N_i = (N_{i1}, \dots, N_{iT_i})'$ ، به شرط $v_i^{(1)}$ و $v_i^{(2)}$ مستقل هستند.

فرض کنید $\beta = (\beta^{(1)'}, \beta^{(2)'}, \beta^{(r)'})'$ و $\gamma = (\gamma^{(1)'}, \gamma^{(2)'})'$ اگر $T_{(1)} \leq \dots \leq T_{(m)}$ شکل مرتب شده T_1, \dots, T_m باشد به طوری که برای هر $i = 1, \dots, m$ ، $1 \leq T_{(i)} \leq T$ ، آنگاه $\beta^{(j)} = (\beta_1^{(j)'}, \dots, \beta_{T_{(1)}}^{(j)'}, \dots, \beta_{T_{(m)}}^{(j)'})'$ و $\gamma^{(j)} = (\gamma_1^{(j)'}, \dots, \gamma_{T_{(1)}}^{(j)'}, \dots, \gamma_{T_{(m)}}^{(j)'})'$ که در آن $j = 1, 2, 3$. بردار $\sigma = (\sigma_*^2, \sigma^2)'$ برداری از پارامترهای پراکنندگی مربوط به چگالی‌های موجود در مدل آمیخته و $\lambda = (\lambda_1', \dots, \lambda_r', \lambda_G')$ برداری شامل پارامترهای واریانس-کوواریانس مربوط به اثرهای تصادفی است. فرض کنید $\xi = (\beta', \gamma', \sigma', \lambda')$ برداری مرکب از پارامترهای مدل باشد که متعلق به فضای پارامتری $\Omega \subseteq \mathbb{R}^d$ است.

تابع درستنمایی مدل (۱)، حاصل ضرب همه توابع درستنمایی تکی در نمونه به صورت

$$\begin{aligned}
 L(\xi) &= \prod_{i=1}^m \int_{\mathcal{D}_i} \prod_{t=1}^{T_i} f(y_{it}, n_{it}, r_{y_{it}}, r_{n_{it}} | \mathbf{x}_{it}, \mathbf{w}_{it}, \mathbf{x}_{it}^*, \mathbf{w}_{it}^*, \mathbf{m}_{\setminus it}, \mathbf{m}_{\setminus it}, \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) \\
 &\quad \times \psi(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) d\mathbf{v}_i^{(1)} d\mathbf{v}_i^{(2)} \\
 &= \prod_{i=1}^m \int_{\mathcal{D}_i} \prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \mathbf{w}_{it}, \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) \times f(n_{it} | \mathbf{x}_{it}^*, \mathbf{w}_{it}^*, \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) \\
 &\quad \times f(r_{y_{it}} | \mathbf{m}_{\setminus it}, y_{it}) \times f(r_{n_{it}} | \mathbf{m}_{\setminus it}, n_{it}) \times \psi(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) d\mathbf{v}_i^{(1)} d\mathbf{v}_i^{(2)}. \quad (2)
 \end{aligned}$$

است، که در آن $\mathcal{D}_i = \{(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}); \mathbf{v}_i^{(1)} \in \mathbb{R}^{d_1}, \mathbf{v}_i^{(2)} \in \mathbb{R}^{d_2}\}$ تابع چگالی متناظر با توزیع شیء تصادفی $(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)})$ را نمایش می‌دهد. بدیهی است که $\mathbf{x}_{it} = (\mathbf{x}'_{\setminus it}, \mathbf{x}'_{\setminus it}, \mathbf{x}'_{\setminus it})'$ به طور مشابه برای $(\mathbf{w}_{it}^* \text{ و } \mathbf{x}_{it}^*, \mathbf{w}_{it})$. چگالی شرطی $f(y_{it} | \mathbf{x}_{it}, \mathbf{w}_{it}, \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)})$ به صورت

$$\begin{aligned}
 & \left[\pi_{\mathcal{A}_{it}} g_{\mathcal{A}}(y_{it} | \theta_{\mathcal{A}_{it}}, \sigma_*^2) + (1 - \pi_{\mathcal{A}_{it}}) \phi(y_{it}; \mu_{it}, \sigma^2) \right]^{z_{\mathcal{A}_{it}}} \\
 & \quad \times \left[(1 - \pi_{\mathcal{A}_{it}}) \phi(y_{it}; \mu_{it}, \sigma^2) \right]^{(1-z_{\mathcal{A}_{it}})},
 \end{aligned}$$

و چگالی شرطی $f(n_{it} | \mathbf{x}_{it}^*, \mathbf{w}_{it}^*, \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)})$ به صورت

$$\left[\pi_{\circ_{it}} + (1 - \pi_{\circ_{it}}) \frac{a(n_{it}) \nu_{it}^{n_{it}}}{g(\nu_{it})} \right]^{z_{\circ_{it}}} \times \left[(1 - \pi_{\circ_{it}}) \frac{a(n_{it}) \nu_{it}^{n_{it}}}{g(\nu_{it})} \right]^{(1-z_{\circ_{it}})},$$

است، که در آن $g_{\mathcal{A}_{it}}(\cdot; \theta_{\mathcal{A}_{it}}, \sigma_*^2)$ یک تابع چگالی احتمال بریده روی مجموعه \mathcal{A} یا تابع چگالی احتمال یک توزیع دلخواه $\mathcal{G}_{\mathcal{A}}$ روی مجموعه \mathcal{A} با پارامترهای مکان و مقیاس $\theta_{\mathcal{A}_{it}}$ و σ_*^2 است و

$$z_{\mathcal{A}_{it}} = \begin{cases} 1 & ; y_{it} \in \mathcal{A} \\ \circ & ; y_{it} \notin \mathcal{A} \end{cases} \quad \text{و} \quad z_{\circ_{it}} = \begin{cases} 1 & ; n_{it} = \circ \\ \circ & ; n_{it} \neq \circ \end{cases}$$

توزیع شرطی $R_{y_{it}} | (m_{y_{it}}, y_{it})$ و $R_{n_{it}} | (m_{n_{it}}, n_{it})$ به ترتیب توزیع برنولی با احتمال‌های موفقیت

$$\frac{\exp(m'_{y_{it}} \alpha_{y_{it}}^{(1)} + \alpha_{y_{it}}^{(1)} y_{it})}{1 + \exp(m'_{y_{it}} \alpha_{y_{it}}^{(1)} + \alpha_{y_{it}}^{(1)} y_{it})} \quad \text{و} \quad \frac{\exp(m'_{n_{it}} \alpha_{n_{it}}^{(2)} + \alpha_{n_{it}}^{(2)} n_{it})}{1 + \exp(m'_{n_{it}} \alpha_{n_{it}}^{(2)} + \alpha_{n_{it}}^{(2)} n_{it})}$$

است. تابع درستنمایی (۲) در حالات مختلفی، بسته به اینکه کدام پاسخ‌ها در کدام زمان‌ها شامل مقادیر گم شده هستند، متفاوت خواهد بود. در حالت گم‌شدگی در یک پاسخ، در مجموع به تعداد $\sum_{i=1}^m T_i$ الگوی مختلف از مقادیر گم‌شده برای هر آزمودنی وجود دارد. با فرض $T_1 = \dots = T_m = T = 2$ ، این الگوها و شکل تابع درستنمایی مشاهده شده در یک حالت خاص ارائه می‌شود. فرض کنید که توزیع متغیر شمارشی، آماسیده در صفر دوجمله‌ای باشد. همچنین فقط متغیر شمارشی گم شده باشد و $T_1 = \dots = T_m = T = 2$. در این صورت برای هر فرد $2^T = 4$ الگو برای مقادیر گم‌شده وجود خواهد داشت. در این حالت، تابع درستنمایی به صورت

$$\begin{aligned} L(\xi) &= \prod_{\{i:r_{i1}=1\}} f_{Y_i, N_i, R_i | X_i, M_i}(y_i, n_i, r_i | x_i, m_i; \xi) \\ &\times \prod_{\{i:r_{i1}=0, r_{i2}=1\}} f_{Y_i, N_{i2}, R_i | X_i, M_i}(y_i, n_{i2}, r_i | x_i, m_i; \xi) \\ &\times \prod_{\{i:r_{i1}=1, r_{i2}=0\}} f_{Y_i, N_{i1}, R_i | X_i, M_i}(y_i, n_{i1}, r_i | x_i, m_i; \xi) \\ &\times \prod_{\{i:r_{i1}=0, r_{i2}=0\}} f_{Y_i, R_i | X_i, M_i}(y_i, r_i | x_i, m_i; \xi) \\ &= \prod_{i=1}^m [f(y_i, n_{i1}, n_{i2}; \xi)]^{r_{i1} r_{i2}} [f(y_i, NA, n_{i2}; \xi)]^{(1-r_{i1}) r_{i2}} \\ &\times [f(y_i, n_{i1}, NA; \xi)]^{r_{i1} (1-r_{i2})} [f(y_i, NA, NA; \xi)]^{(1-r_{i1})(1-r_{i2})}, \quad (3) \end{aligned}$$

است، که در آن R متغیر نشانگر مربوط به گم شدن متغیر شمارشی است و

$$\begin{aligned}
 f(y_i, n_{i1}, n_{i2}; \xi) &= \int_{\mathcal{D}_i} \prod_{t=1}^2 \left(f(y_{it} | \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}; \xi) f(n_{it} | \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}; \xi) \right. \\
 &\quad \times \left. Pr(R_{it} = 1 | y_{it}; \xi) \right) \psi(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) d\mathbf{v}_i^{(1)} d\mathbf{v}_i^{(2)}, \\
 f(y_i, NA, n_{i2}; \xi) &= \int_{\mathcal{D}_i} \left(\prod_{t=1}^2 f(y_{it} | \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}; \xi) \right) Pr(R_{i1} = 0 | \xi) \\
 &\quad \times f(n_{i2} | \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}; \xi) \psi(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) d\mathbf{v}_i^{(1)} d\mathbf{v}_i^{(2)}, \\
 f(y_i, NA, NA; \xi) &= \int_{\mathcal{D}_i} \left(\prod_{t=1}^2 f(y_{it} | \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}; \xi) \right) Pr(R_{i2} = 0 | \xi) \\
 &\quad \times Pr(R_{i1} = 0 | \xi) \psi(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}) d\mathbf{v}_i^{(1)} d\mathbf{v}_i^{(2)}.
 \end{aligned}$$

$f(y_i, n_{i1}, NA; \xi)$ به طور مشابه قابل محاسبه است. در این روابط اگر مثلاً متغیر شمارشی توزیع آماسیده در صفر دوجمله‌ای n و ν_{it} ، $ZIB(n, \nu_{it}, \pi_{oit})$ داشته باشد، آنگاه برای آزمودنی i ام که پاسخ شمارشی آن در زمان اول گم شده است رابطه

$$\begin{aligned}
 Pr(R_{i1} = 0 | \xi) &= \sum_{n_{i1}=0}^n \left(1 - Pr(R_{i1} = 1 | n_{i1}; \xi) \right) \left(\pi_{i1} + \frac{1 - \pi_{i1}}{(1 + \nu_{i1})^n} \right)^{z_{oi1}} \\
 &\quad \times \left((1 - \pi_{i1}) \binom{n}{n_{i1}} \left(\frac{\nu_{i1}}{1 + \nu_{i1}} \right)^{n_{i1}} \left(\frac{1}{1 + \nu_{i1}} \right)^{n - n_{i1}} \right)^{1 - z_{it}}.
 \end{aligned}$$

برقرار است. به همین شکل در حالت‌های مختلف این احتمال‌ها قابل محاسبه است. موارد دیگر نیز به شکل مشابه به دست می‌آیند.

انتگرال موجود در رابطه (۲) که نسبت به بردارهای اثرهای تصادفی است با روش مونت کارلو تقریب زده می‌شود. این روشی ساده و قوی برای تقریب انتگرالهای پیچیده است. برای برآورد پارامترهای مدل از روش بیشینه درستنمایی استفاده می‌شود. به دلیل پیچیدگی تابع درستنمایی، می‌توان با روش‌های عددی تابع درستنمایی را ماکسیم کرد. تابع `optim` و `nlminb` در نرم افزار R برای مینیم کردن توابع مختلف با روش‌های عددی استفاده می‌شوند. می‌توان با قرار دادن منفی لگاریتم تابع درستنمایی در این توابع به مقدار ماکسیم تابع درستنمایی دست پیدا کرد. روش پیش فرض برای تابع `optim` یک فرایند بهینه‌سازی

فارغ از مشتق‌گیری است که الگوریتم Nelder-Mead نامیده می‌شود. این روش فقط مقادیر تابع را به کار می‌برد، استوار اما نسبتاً کند است و به طور منطقی برای توابع مشتق ناپذیر خوب عمل می‌کند.

۲.۲ مدل برای داده‌های $N - kIO$ با امکان گم‌شدگی غیر تصادفی

فرض می‌شود که $Y_{it}|b_i^{(1)}$ و $\tilde{Y}_{it}|b_i^{(2)}, b_i^{(r)}$ به ترتیب دارای یک توزیع نرمال و یک توزیع k -آماسیده چندجمله‌ای ($kIMULT$) هستند که $b_i^{(1)}, b_i^{(2)}, b_i^{(r)}$ بردارهایی از اثرهای تصادفی آزمودنی-مشخص هستند. مدل توأم برای پاسخ‌های آمیخته طولی پیوسته، Y_{it} ، و رده-آماسیده ترتیبی، \tilde{Y}_{it} وقتی که امکان گم‌شدگی غیرتصادفی وجود دارد به صورت

$$\begin{aligned} Y_{it}|b_i^{(1)}, \mu_{it} &\sim N(\mu_{it}, \sigma^2), \\ q_1(\mu_{it}) &= \mathbf{X}'_{vit}\boldsymbol{\beta}_t^{(1)} + \mathbf{W}'_{vit}\mathbf{b}_i^{(1)}, \\ \text{logit}(Pr(R_{y_{it}} = 1|M_{vit}, y_{it})) &= \mathbf{M}'_{vit}\boldsymbol{\alpha}_t^{(1)} + \alpha_{vt}^{(1)}y_{it}, \\ \tilde{Y}_{it}|b_i^{(2)}, b_i^{(r)}, \eta_{j,it}, \pi_{it} &\sim kIMULT(1, \eta_{\circ,it}, \dots, \eta_{c-1,it}, \pi_{it}), \\ \text{logit}(\eta_{j,it}) &= \theta_j - (\mathbf{X}'_{vit}\boldsymbol{\gamma}^{(1)} + \mathbf{W}'_{vit}\mathbf{b}_i^{(r)}), \quad j = \circ, 1, \dots, c-1 \\ q_2(\pi_{it}) &= \mathbf{X}'_{vit}\boldsymbol{\gamma}^{(2)} + \mathbf{W}'_{vit}\mathbf{b}_i^{(r)} \\ \text{logit}(Pr(R_{\tilde{Y}_{it}} = 1|M_{vit}, \tilde{y}_{it})) &= \mathbf{M}'_{vit}\boldsymbol{\alpha}_t^{(2)} + \alpha_{vt}^{(2)}\tilde{y}_{it}. \end{aligned} \quad (۴)$$

است] که در آن μ_{it} میانگین توزیع نرمال متغیر پیوسته، برای هر j ، $j = 1, \dots, c-1$ ، $\eta_{j,it} =$ ، $Pr(\tilde{Y}_{it} \leq j)$ ، و π_{it} پارامتر آمیزنده متغیر ترتیبی است. q_1 و q_2 به ترتیب توابع ربط معلوم برای μ_{it} و π_{it} هستند. معمولاً q_1 و q_2 به ترتیب تابع ربط همانی و تابع ربط لوجیت هستند. توابع ربط دیگری نیز مانند تابع پروبیت یا تابع لگ-لگ مکمل می‌توانند برای π_{it} استفاده شوند. پارامترهای $\theta_{\circ}, \dots, \theta_{c-1}$ پارامترهای عرض از مبدأ در رگرسیون ترتیبی هستند به طوری که برای هر j ، $j = 1, \dots, c-1$ ، $\theta_j > \theta_{j-1}$ ، پارامترهای $\boldsymbol{\beta}_t^{(1)}, \boldsymbol{\gamma}_t^{(1)}, \boldsymbol{\gamma}_t^{(2)}, \boldsymbol{\alpha}_t^{(1)}, \boldsymbol{\alpha}_t^{(2)}$ ضرایب رگرسیونی هستند. پارامترهای $\alpha_{vt}^{(1)}$ و $\alpha_{vt}^{(2)}$ در بردارنده اطلاعات مربوط به سازوکار گم‌شدگی برای پاسخ‌ها در زمان t هستند به طوری که تایید صفر بودن هر دوی این ضرایب نشان می‌دهد که سازوکار گم‌شدگی تصادفی و تایید صفر بودن مقدار یکی از آن‌ها حالتی را نشان می‌دهد که سازوکار گم‌شدگی غیر تصادفی است. برای مدل‌بندی طبیعت طولی پاسخ‌ها، یعنی وابستگی درون-آزمودنی در طول زمان، از بردار اثرهای تصادفی $\mathbf{b}_i^{(1)}$ (برای پاسخ پیوسته)

و $\mathbf{b}_i^{(۴)}$ (برای پاسخ ترتیبی)، استفاده شده است. علاوه بر این به شرط اثرهای تصادفی بردار پاسخها در موقعیت‌های مختلف برای هر آزمودنی مستقل شرطی هستند. بردار اثرهای تصادفی $\mathbf{b}_i^{(۳)}$ برای وابستگی پارامتر آمیزنده در طول زمان در نظر گرفته می‌شوند.

همانند مدل (۱) برای مدل‌بندی توأم پاسخ‌های آمیخته، با وارد کردن اثرهای تصادفی همبسته به مدل (۴)، همبستگی بین \tilde{Y}_{it} و Y_{it} در نظر گرفته شده است. بنابراین با فرض $\mathbf{v}_i' = (\mathbf{b}_i^{(۲)'}, \mathbf{b}_i^{(۳)'})$ ، ماتریس واریانس-کوواریانس بردار \mathbf{v}_i به صورت $Cov(\mathbf{b}_i^{(۱)}, \mathbf{v}_i) = \mathbf{G}$ است و برای هر $j = ۱, ۲, ۳$ فرض می‌شود که $\mathbf{b}_i^{(j)} \sim MVN(\mathbf{0}, \Sigma_j)$. علاوه بر این، $(\mathbf{b}_i^{(۱)'}, \mathbf{v}_i')$ و \mathbf{v}_i توزیع‌های گاوسی با میانگین $\mathbf{0}$ و ماتریس کوواریانس $\Sigma_{\mathbf{b}_i^{(۱)}, \mathbf{v}_i}$ دارند که $\Sigma_{\mathbf{v}_i} = \begin{pmatrix} \Sigma_2 & \mathbf{0} \\ \mathbf{0} & \Sigma_3 \end{pmatrix}$ بردارهای پاسخ برای i امین آزمودنی، $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iT_i})'$ و $\tilde{\mathbf{Y}}_i = (\tilde{Y}_{i1}, \dots, \tilde{Y}_{iT_i})'$ به شرط \mathbf{v}_i و $\mathbf{b}_i^{(۱)}$ مستقل هستند. فرض کنید $\gamma = (\gamma^{(۱)'}, \gamma^{(۲)'})'$ اگر $T_{(۱)} \leq \dots \leq T_{(m)}$ شکل مرتب شده T_1, \dots, T_m باشد به طوری که برای هر $i = ۱, \dots, m$ ، $1 \leq T_{(i)} \leq T$ ، آنگاه $\beta^{(۱)} = (\beta_1^{(۱)'}, \dots, \beta_{T_{(۱)}}^{(۱)'}, \dots, \beta_{T_{(m)}}^{(j)'})'$ و $\gamma^{(j)} = (\gamma_1^{(j)'}, \dots, \gamma_{T_{(۱)}}^{(j)'}, \dots, \gamma_{T_{(m)}}^{(j)'})'$ همچنین $\lambda = (\lambda_1', \lambda_2', \lambda_3', \lambda_G')$ برداری شامل پارامترهای واریانس-کوواریانس مربوط به اثرهای تصادفی است. فرض کنید $\xi = (\beta', \gamma', \lambda', \sigma^2)'$ برداری مرکب از پارامترهای مدل باشد که متعلق به $\Omega \subseteq \mathbb{R}^d$ است (فضای اقلیدسی d بُعدی است که d تعداد همه پارامترهای مدل است).

تابع درست‌نمایی مدل (۴)، حاصل ضرب همه توابع درست‌نمایی تکی در نمونه به صورت

$$\begin{aligned} L(\xi) &= \prod_{i=1}^m \int_{\mathcal{D}_i} \prod_{t=1}^{T_i} f(y_{it}, \tilde{y}_{it}, r_{y_{it}}, r_{\tilde{y}_{it}} | \mathbf{x}_{it}, \mathbf{w}_{it}, \mathbf{m}_{\vee it}, \mathbf{m}_{\vee it}, \mathbf{b}_i^{(۱)}, \mathbf{v}_i) \\ &\quad \times \psi(\mathbf{b}_i^{(۱)}, \mathbf{v}_i) d\mathbf{b}_i^{(۱)} d\mathbf{v}_i \\ &= \prod_{i=1}^m \int_{\mathcal{D}_i} \prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{\vee it}, \mathbf{w}_{\vee it}, \mathbf{b}_i^{(۱)}, \mathbf{v}_i) \times f(\tilde{y}_{it} | \mathbf{x}_{\vee it}, \mathbf{w}_{\vee it}, \mathbf{x}_{\vee it}, \mathbf{w}_{\vee it}, \mathbf{b}_i^{(۱)}, \mathbf{v}_i) \\ &\quad \times f(r_{y_{it}} | \mathbf{m}_{\vee it}, y_{it}) \times f(r_{\tilde{y}_{it}} | \mathbf{m}_{\vee it}, \tilde{y}_{it}) \times \psi(\mathbf{b}_i^{(۱)}, \mathbf{v}_i) d\mathbf{b}_i^{(۱)} d\mathbf{v}_i \end{aligned} \quad (۵)$$

است، که در آن $\mathcal{D}_i = \{(\mathbf{b}_i^{(۱)}, \mathbf{v}_i); \mathbf{b}_i^{(۱)} \in \mathbb{R}^{d_1}, \mathbf{v}_i \in \mathbb{R}^{d_2}\}$ و $\psi(\cdot)$ تابع چگالی متناظر با توزیع

شیء تصادفی $(b_i^{(1)}, v_i)$ را نمایش می‌دهد. بدیهی است که $x_{it} = (x'_{1it}, x'_{2it}, x'_{3it})'$ (به طور مشابه برای w_{it}). چگالی شرطی $f(y_{it}|x_{1it}, w_{1it}, b_i^{(1)}, v_i)$ یک چگالی نرمال با میانگین μ_{it} و واریانس σ^2 و چگالی شرطی $f(\tilde{y}_{it}|x_{2it}, w_{2it}, x_{3it}, w_{3it}, b_i^{(1)}, v_i)$ به صورت

$$\begin{aligned} & [\pi_{it} + (1 - \pi_{it})(\eta_{k,it} - \eta_{k-1,it})]^{z_{it}} \\ & \times \left[(1 - \pi_{it}) \sum_{j \in \{0, \dots, c\} - k} [(\eta_{j,it} - \eta_{j-1,it}) I_{\{j\}}(\tilde{y}_{it})] \right]^{1-z_{it}} \end{aligned}$$

است، که در آن وقتی $\tilde{y}_{it} = k$ و z_{it} برابر با ۱، در غیر این صورت برابر با ۰ است.

تابع درستنمایی (۵) در حالات مختلفی، بسته به اینکه کدام پاسخ‌ها در کدام زمان‌ها شامل مقادیر گم شده هستند، متفاوت خواهد بود. در اینجا نیز انتگرال موجود در رابطه (۵) که نسبت به بردارهای اثرهای تصادفی است، با روش مونت کارلو تقریب زده می‌شود. برای برآورد پارامترهای مدل از روش بیشینه درستنمایی استفاده می‌شود. برای این منظور می‌توان از تابع `optim` و روش `Neder-Mead` یا تابع `nlminb` در نرم افزار R استفاده کرد. توضیحات مناسب در انتهای بخش ۲.۱ ارائه شده است.

شناساپذیری یکی از شرایط ضروری برای کفایت یک مدل آماری است. وانگ (۲۰۱۳) مسئله شناساپذیری در مدل‌های آمیخته خطی را مورد بررسی قرار داد. طبق قضایای اثبات شده توسط وانگ (۲۰۱۳) و نتایج به دست آمده، یکی از شروط کافی برای شناساپذیری پارامترهای کوواریانس، معلوم بودن ماتریس کوواریانس اثر تصادفی است. بدین منظور با دوباره پارامترسازی اثرهای تصادفی در مدل‌های مورد بررسی به مسئله شناساپذیری توجه شده است.

۳ مطالعات شبیه‌سازی

در این بخش عملکرد مدل (۱) و (۴) با تعدادی مطالعه شبیه‌سازی مورد ارزیابی قرار می‌گیرد. نمونه‌های شبیه‌سازی با چهار اندازه نمونه $m = 50, 75, 100, 500$ و برای طرح متعادل با $T_1 = \dots = T_m = 3$ مورد بررسی قرار می‌گیرند. فرض می‌شود $W_{1it} = W_{2it} = W_{3it}^* = 1$ و یک متغیر زمان-متغیر X_1 در نظر گرفته می‌شود که $X_{1i} = (X_{1i1}, X_{1i2}, X_{1i3})'$ ، $i = 1, \dots, m$ ، به طور مستقل از یک توزیع نرمال دومتغیره با میانگین $(0, 0, 0)'$ و ماتریس کوواریانس $\begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$

تولید می‌شود. تعداد تکرارهای شبیه‌سازی برای اندازه نمونه‌های $m = 100, 500$ برابر با ۱۰۰، و برای اندازه نمونه‌های $m = 50, 75$ برابر با ۱۰۰۰ خواهد بود. برای تقریب انتگرال موجود در تابع درست‌نمایی (بر حسب اثرهای تصادفی) با استفاده از روش مونت کارلو، تعداد ۵۰ نمونه مونت کارلو برای اثرهای تصادفی تولید می‌شوند. در همه مدل‌های مورد بررسی، اثرهای تصادفی توزیع نرمال استاندارد دارند و λ_1 و λ_2 مقادیر حقیقی مثبت هستند. با تبدیل $b_i = b_i / \sigma_b$ در انتگرال موجود در توابع درست‌نمایی اثرهای تصادفی بازمقیاس می‌شوند و به جای هر اثر تصادفی همچون b_i در مدل‌های توأم، λb_i قرار می‌گیرد. در واقع، ضریب λ همان پارامتر مربوط به واریانس اثر تصادفی هستند (σ_b). بنابراین اثرهای تصادفی مستقل از توزیع نرمال صفر و یک تولید می‌شوند. با این کار به مسئله شناساپذیری مدل‌ها نیز توجه شده است (وانگ، ۲۰۱۳).

۳.۱ شبیه‌سازی ۱

در مدل (۱) مجموعه \mathcal{A} یک بازه همچون (a, b) در نظر گرفته می‌شود. متغیر پیوسته از توزیع (a, b) -آماسیده نرمال در دو حالت که $g_{\mathcal{A}}$ یک توزیع یکنواخت و یک توزیع نرمال بریده دوگانه (DTN) است، و متغیر شمارشی نیز از توزیع دوجمله‌ای آماسیده در صفر تولید می‌شوند. بنابراین پاسخ‌های آمیخته پیوسته و شمارشی از چهار مدل توأم خاص تولید می‌شوند. در مدل (۱) زیرمدل‌های مربوط به نسبت‌های آمیزنده فارغ از اثر تصادفی خواهند بود و فرض می‌شود که $b_i^{(1)} = b_i^{(2)} = b_i^{(3)}$. $b_i^{(1)}$ و $b_i^{(2)}$ برای $m = 50, 75, 100, 500$ آزمودنی به طور مستقل از توزیع نرمال صفر و یک تولید می‌شوند.

در حالت اول Y_{it} از مدل

$$\begin{aligned}
 Y_{it} | b_i^{(1)}, \mu_{it}, \pi_{\mathcal{A}_{it}}, \sigma^2 &\sim (a, b)IN(\mu_{it}, \pi_{\mathcal{A}_{it}}, \sigma^2); \quad g_{\mathcal{A}_{it}} \sim U(a, b) \\
 \mu_{it} &= \beta_0^{(1)} + \beta_1^{(1)} X_{1it} + \lambda_1 b_i^{(1)}, \\
 \text{logit}(\pi_{\mathcal{A}_{it}}) &= \beta_0^{(2)} + \beta_1^{(2)} X_{1it},
 \end{aligned} \tag{۶}$$

N_{it} از مدل

$$\begin{aligned} N_{it}|b_i^{(1)}, \nu_{it}, \pi_{\cdot it} &\sim ZIB\left(\gamma, \frac{\nu_{it}}{1 + \nu_{it}}, \pi_{\cdot it}\right), \\ \log(\nu_{it}) &= \gamma_{\circ}^{(1)} + \gamma_1^{(1)} X_{\nu it} + \lambda_1 b_i^{(1)}, \\ \text{logit}(\pi_{\cdot it}) &= \gamma_{\circ}^{(2)} + \gamma_1^{(2)} X_{\nu it}. \end{aligned} \quad (7)$$

تولید می‌شود. در حالت دوم Y_{it} از مدل

$$\begin{aligned} Y_{it}|b_i^{(1)}, b_i^{(2)}, \theta_{it}^*, \mu_{it}, \pi_{A_{it}}, \sigma_*^{\vee}, \sigma^{\vee} &\sim (a, b)IN(\theta_{it}^*, \mu_{it}, \pi_{A_{it}}, \sigma_*^{\vee}, \sigma^{\vee}); g_{A_{it}} \sim DTN_{(a,b)} \\ \mu_{it} &= \beta_{\circ}^{(1)} + \beta_{\nu t}^{(1)} X_{\nu it} + \lambda_1 b_i^{(1)}, \\ \text{logit}(\theta_{it}^*) &= \beta_{\circ}^{(2)} + \beta_{\nu t}^{(2)} X_{\nu it} + \lambda_2 b_i^{(2)}, \\ \text{logit}(\pi_{A_{it}}) &= \beta_{\circ}^{(3)} + \beta_1^{(3)} X_{\nu it}, \end{aligned} \quad (8)$$

و N_{it} از مدل (7) تولید می‌شود. در مدل (8)، $\theta_{it} = (b - a) \times \theta_{it}^* + a$ ، $\theta_{it}^* < 1$ ، پارامتر مکان توزیع $DTN_{(a,b)}(\theta_{it}, \sigma_*^{\vee})$ است. با فرض امکان گمشدگی غیرتصادفی در پاسخ شمارشی در زمان اول، احتمال گمشدگی (سازوکار گمشدگی غیرقابل چشم‌پوشی) برای داده‌های ناکامل به صورت

$$Pr(R_{n_{i1}} = 1 | X_{\nu i1}, n_{i1}) = \frac{\exp(\alpha_{\nu 1}^{(2)} X_{\nu i1} + \alpha_{\nu 1}^{(1)} n_{i1})}{1 + \exp(\alpha_{\nu 1}^{(2)} X_{\nu i1} + \alpha_{\nu 1}^{(1)} n_{i1})}.$$

است. مقادیر پارامترها برای تولید داده از مدل‌های (6) و (8) در دو اندازه نمونه ۱۰۰ و ۵۰۰ عبارت از $(\gamma_{\circ}^{(1)}, \gamma_1^{(1)}, \gamma_{\circ}^{(2)}, \gamma_1^{(2)}) = (\beta_{\circ}^{(1)}, \beta_1^{(1)}, \beta_{\circ}^{(2)}, \beta_1^{(2)}, \beta_{\circ}^{(3)}, \beta_1^{(3)}) = (0.1, 1, -0.1, 1, 1, 1)$ و $(-0.1, -2, 1, 1)$ و $(\lambda_1, \lambda_2, \sigma^{\vee}, \sigma_*^{\vee}) = (1, 0.5, 1, 0.5)$ است. در دو اندازه نمونه ۵۰ و ۷۵، مقادیر واقعی پارامترهای مدل $(\beta_{\circ}^{(1)}, \beta_1^{(1)}, \beta_{\circ}^{(2)}, \beta_1^{(2)}, \beta_{\circ}^{(3)}, \beta_1^{(3)}) = (0.1, 0.5, 0.1, -0.5, 0.1, 2)$ و $(\lambda_1, \lambda_2, \sigma^{\vee}, \sigma_*^{\vee}) = (1, 0.5, 1, 0.5)$ هستند. مقادیر واقعی پارامترهای احتمال گمشدگی برای تولید داده‌های آمیخته با مقادیر گمشدگی غیرتصادفی در مدل‌های (6) و (8) به صورت $(\alpha_{\nu 1}^{(2)}, \alpha_{\nu 1}^{(1)}) = (0.5, 0.1)$ است. برای تولید مقادیر از مدل (6) و (8) مقدار (a, b) به ترتیب برابر با (۱، ۳) و (۱، ۴) در نظر گرفته می‌شود.

جدول ۰۱. نتایج شبیه‌سازی بخش ۳.۱، حالت اول ($UN - ZIB$) با ۱۰۰ تکرار شبیه‌سازی

پارامتر	مقدار واقعی	$m = 100$				$m = 500$			
		الف)		ب)		الف)		ب)	
		برآورد	SE	برآورد	SE	برآورد	SE	برآورد	SE
$\beta_0^{(1)}$	۰/۱	۰/۳۴۰	۰/۴۵۰	۰/۲۱۴	۰/۳۸۹	۰/۱۹۴	۰/۹۵	۰/۱۸۸	۰/۱۱۰
$\beta_1^{(1)}$	۱/۰	۱/۱۰۸	۰/۲۱۹	۱/۱۱۳	۰/۱۶۹	۰/۸۸۶	۰/۷۶	۰/۹۷۶	۰/۰۸۸
λ_1	۱/۰	۰/۷۳۵	۰/۱۶۰	۰/۸۰۱	۰/۲۰۱	۱/۰۳۸	۰/۷۲	۱/۱۰۴	۰/۰۶۱
σ^2	۱/۰	۰/۹۰۰	۰/۱۱۷	۱/۲۳۳	۰/۲۰۱	۱/۰۰۲	۰/۵۸	۱/۰۳۴	۰/۲۱۰
$\beta_0^{(2)}$	۱/۰	۱/۲۶۳	۰/۲۸۱	۱/۱۲۲	۰/۱۹۲	۰/۹۰۰	۰/۱۰۵	۰/۹۸۲	۰/۱۱۱
$\beta_1^{(2)}$	۱/۰	۱/۲۰۵	۰/۲۴۸	۱/۲۱۷	۰/۱۹۸	۱/۰۲۹	۰/۱۱۷	۱/۰۱۹	۰/۱۰۴
$\gamma_0^{(1)}$	-۰/۱	-۰/۱۱۰	۰/۱۹۷	-۰/۱۲۳	۰/۱۸۶	-۰/۱۳۱	۰/۹۱	-۰/۹۸	۰/۰۸۳
$\gamma_1^{(1)}$	-۲/۰	-۱/۶۸۱	۰/۲۸۶	-۱/۷۴۰	۰/۲۳۸	-۲/۰۲۷	۰/۸۸	-۲/۱۰۱	۰/۰۷۵
$\gamma_0^{(2)}$	۱/۰	۰/۵۵۲	۰/۴۰۳	۰/۶۸۹	۰/۳۹۱	۰/۶۶۸	۰/۸۹	۰/۷۲۱	۰/۰۶۷
$\gamma_1^{(2)}$	۱/۰	۱/۱۰۸	۰/۳۲۴	۱/۱۰۱	۰/۲۳۹	۱/۲۰۳	۰/۱۰۹	۱/۰۹۸	۰/۲۰۲
$\alpha_{11}^{(2)}$	۱/۰	۱/۳۹۱	۰/۴۷۱	۱/۲۴۷	۰/۳۹۷	۱/۰۹۷	۰/۲۱۷	۱/۰۸۴	۰/۲۰۱
$\alpha_{21}^{(2)}$	۱/۰	۱/۲۰۶	۰/۴۴۳	۱/۱۹۱	۰/۳۵۹	۱/۱۱۴	۰/۲۳۴	۱/۰۶۸	۰/۱۷۱

جدول ۰۲. نتایج شبیه‌سازی بخش ۳.۱، حالت دوم ($TN - ZIB$) با ۱۰۰ تکرار شبیه‌سازی

پارامتر	مقدار واقعی	$m = 100$				$m = 500$			
		الف)		ب)		الف)		ب)	
		برآورد	SE	برآورد	SE	برآورد	SE	برآورد	SE
$\beta_0^{(1)}$	۰/۱	۰/۳۷۸	۰/۳۱۶	۰/۳۲۴	۰/۲۸۹	۰/۱۹۷	۰/۱۲۹	۰/۲۰۹	۰/۱۳۳
$\beta_1^{(1)}$	۱/۰	۰/۹۷۳	۰/۱۹۵	۱/۲۱۴	۰/۱۶۶	۱/۰۸۸	۰/۹۱	۱/۰۷۷	۰/۰۴۹
λ_1	۱/۰	۱/۲۲۳	۰/۱۴۸	۱/۱۶۴	۰/۲۸۹	۱/۰۴۸	۰/۷۲	۱/۰۱۴	۰/۰۸۱
σ^2	۱/۰	۱/۰۷۶	۰/۱۳۸	۰/۹۱۴	۰/۲۵۱	۱/۰۴۶	۰/۵۷	۱/۰۲۸	۰/۰۶۶
$\beta_0^{(2)}$	-۰/۱	-۰/۱۳۹	۰/۰۹۹	-۰/۲۱۰	۰/۰۸۷	-۰/۱۱۸	۰/۴۱	-۰/۱۳۱	۰/۰۳۵
$\beta_1^{(2)}$	۱/۰	۱/۰۳۸	۰/۱۱۶	۱/۱۱۵	۰/۱۸۰	۱/۰۸۹	۰/۵۸	۱/۰۷۷	۰/۰۱۹
λ_2	۰/۵	۰/۶۳۳	۰/۱۲۲	۰/۷۱۴	۰/۳۱۱	۰/۴۹۴	۰/۵۲	۰/۵۱۲	۰/۰۴۱
σ_*^2	۰/۵	۰/۴۷۶	۰/۴۶	۰/۵۶۸	۰/۰۸۴	۰/۵۰۴	۰/۲۱	۰/۵۳۲	۰/۰۳۷
$\beta_0^{(3)}$	۱/۰	۱/۰۰۷	۰/۲۲۶	۱/۱۰۶	۰/۳۰۳	۱/۰۸۱	۰/۱۰۱	۱/۰۷۵	۰/۲۰۱
$\beta_1^{(3)}$	۱/۰	۱/۰۰۶	۰/۲۰۷	۰/۹۱۱	۰/۲۵۶	۱/۰۴۸	۰/۱۰۳	۰/۹۹۳	۰/۱۷۲
$\gamma_0^{(1)}$	-۰/۱	-۰/۴۵۸	۰/۲۶۰	-۰/۳۱۱	۰/۳۴۲	-۰/۱۵۰	۰/۱۱۲	-۰/۱۰۶	۰/۲۵۱
$\gamma_1^{(1)}$	-۲/۰	-۱/۹۴۳	۰/۲۶۴	-۲/۲۱۴	۰/۳۴۳	-۲/۰۸۹	۰/۱۱۷	-۲/۰۸۵	۰/۱۹۶
$\gamma_0^{(2)}$	۱/۰	۱/۰۹۶	۰/۱۷۳	۱/۱۲۸	۰/۱۸۸	۱/۰۵۲	۰/۸۱	۱/۰۱۴	۰/۰۳۶
$\gamma_1^{(2)}$	۱/۰	۰/۸۸۹	۰/۱۸۱	۰/۸۱۴	۰/۲۱۶	۱/۰۵۲	۰/۹۵	۰/۹۸۱	۰/۰۷۵
$\alpha_{11}^{(2)}$	۱/۰	۱/۱۳۵	۰/۲۰۰	۱/۲۰۱	۰/۱۷۸	۱/۰۰۱	۰/۹۵	۰/۹۹۵	۰/۰۹۰
$\alpha_{21}^{(2)}$	۱/۰	۰/۹۸۳	۰/۰۶۷	۰/۸۹۵	۰/۱۱۷	۰/۹۸۶	۰/۳۴	۰/۹۱۴	۰/۰۲۱

۳.۲ شبیه‌سازی ۲

در مدل (۴) متغیر پیوسته از توزیع نرمال و متغیر ترتیبی از توزیع چندجمله‌ای آماسیده در صفر تولید می‌شوند. بنابراین پاسخ‌های آمیخته پیوسته و ترتیبی از یک مدل توأم خاص تولید می‌شوند. در مدل (۴) نیز زیرمدل

جدول ۳. نتایج شبیه‌سازی بخش ۳.۱، حالت اول ($UN - ZIB$) با ۱۰۰۰۰ تکرار شبیه‌سازی

پارامتر	مقدار واقعی	$m = 75$				$m = 50$			
		(ب)		(الف)		(ب)		(الف)	
		برآورد	SE	برآورد	SE	برآورد	SE	برآورد	SE
$\beta_0^{(1)}$	۰٫۱	۰٫۱۳۹	۰٫۱۵۱	۰٫۲۰۱	۰٫۱۷۶	۰٫۱۲۱	۰٫۱۳۱	۰٫۱۶۶	
$\beta_1^{(1)}$	۰٫۵	۰٫۶۹۱	۰٫۱۱۶	۰٫۶۷۸	۰٫۱۳۱	۰٫۶۶۶	۰٫۱۳۱	۰٫۶۶۶	
λ_1	۱٫۰	۰٫۹۴۳	۰٫۱۳۶	۰٫۸۸۹	۰٫۱۹۹	۱٫۰۶۱	۰٫۱۹۹	۱٫۰۶۱	
σ^2	۱٫۰	۰٫۹۲۳	۰٫۱۵۹	۰٫۸۳۴	۰٫۱۵۱	۱٫۰۲۱	۰٫۱۵۱	۱٫۰۲۱	
$\beta_0^{(2)}$	۰٫۱	۰٫۲۱۳	۰٫۳۱۰	۰٫۳۳۴	۰٫۲۱۴	۰٫۱۴۷	۰٫۲۱۴	۰٫۱۴۷	
$\beta_1^{(2)}$	۱٫۰	۱٫۱۷۰	۰٫۱۲۸	۱٫۲۳۰	۰٫۲۰۱	۰٫۹۶۸	۰٫۲۰۱	۰٫۹۶۸	
$\gamma_0^{(1)}$	-۰٫۱	-۰٫۲۱۱	۰٫۱۸۳	-۰٫۱۳۶	۰٫۲۰۰	-۰٫۰۹۰	۰٫۲۰۰	-۰٫۰۹۰	
$\gamma_1^{(1)}$	-۲٫۰	-۱٫۷۸۸	۰٫۳۲۰	-۱٫۸۶۴	۰٫۲۲۳	-۲٫۰۵۹	۰٫۲۲۳	-۲٫۰۵۹	
$\gamma_0^{(2)}$	۱٫۰	۰٫۸۵۹	۰٫۲۰۳	۰٫۷۹۹	۰٫۳۰۰	۰٫۸۷۹	۰٫۳۰۰	۰٫۸۷۹	
$\gamma_1^{(2)}$	۱٫۰	۱٫۱۳۸	۰٫۲۳۹	۱٫۲۱۱	۰٫۲۳۱	۱٫۰۹۱	۰٫۲۳۱	۱٫۰۹۱	
$\alpha_{11}^{(2)}$	۱٫۰	۱٫۲۶۴	۰٫۳۸۷	۱٫۱۲۲	۰٫۲۶۸	۱٫۰۶۱	۰٫۲۶۸	۱٫۰۶۱	
$\alpha_{21}^{(2)}$	۱٫۰	۰٫۹۶۲	۰٫۲۲۹	۰٫۹۴۲	۰٫۱۸۸	۰٫۹۲۶	۰٫۱۸۸	۰٫۹۲۶	

جدول ۴. نتایج شبیه‌سازی بخش ۳.۱، حالت دوم ($TN - ZIB$) با ۱۰۰۰۰ تکرار شبیه‌سازی

پارامتر	مقدار واقعی	$m = 75$				$m = 50$			
		(ب)		(الف)		(ب)		(الف)	
		برآورد	SE	برآورد	SE	برآورد	SE	برآورد	SE
$\beta_0^{(1)}$	۰٫۱	۰٫۲۷۸	۰٫۲۹۹	۰٫۱۷۱	۰٫۲۴۳	۰٫۱۲۸	۰٫۲۴۳	۰٫۱۲۸	
$\beta_1^{(1)}$	۰٫۵	۰٫۶۹۷	۰٫۲۲۸	۰٫۶۱۴	۰٫۱۷۱	۰٫۵۸۰	۰٫۱۷۱	۰٫۵۸۰	
λ_1	۱٫۰	۱٫۱۲۱	۰٫۱۰۷	۱٫۰۸۹	۰٫۱۲۰	۱٫۰۳۳	۰٫۱۲۰	۱٫۰۳۳	
σ^2	۱٫۰	۱٫۱۳۰	۰٫۱۰۶	۱٫۱۲۱	۰٫۱۰۸	۱٫۰۲۰	۰٫۱۰۸	۱٫۰۲۰	
$\beta_0^{(2)}$	۰٫۱	۰٫۳۲۱	۰٫۳۶۶	۰٫۴۰۹	۰٫۲۸۶	۰٫۲۸۷	۰٫۲۸۶	۰٫۲۸۷	
$\beta_1^{(2)}$	-۰٫۵	-۰٫۸۲۹	۰٫۳۴۲	-۰٫۷۱۷	۰٫۲۸۹	-۰٫۶۸۶	۰٫۲۸۹	-۰٫۶۸۶	
λ_2	۰٫۵	۰٫۶۹۲	۰٫۲۱۰	۰٫۷۰۳	۰٫۲۲۹	۰٫۴۶۲	۰٫۲۲۹	۰٫۴۶۲	
σ_*^2	۰٫۵	۰٫۷۰۹	۰٫۱۳۶	۰٫۶۳۴	۰٫۲۰۹	۰٫۶۸۷	۰٫۲۰۹	۰٫۶۸۷	
$\beta_0^{(3)}$	۰٫۱	۰٫۳۰۱	۰٫۲۸۹	۰٫۲۷۲	۰٫۲۹۱	۰٫۱۷۹	۰٫۲۹۱	۰٫۱۷۹	
$\beta_1^{(3)}$	۲٫۰	۲٫۳۰۱	۰٫۱۵۶	۲٫۲۱۴	۰٫۱۲۵	۲٫۰۹۶	۰٫۱۲۵	۲٫۰۹۶	
$\gamma_0^{(1)}$	-۰٫۱	-۰٫۱۴۸	۰٫۲۴۳	-۰٫۲۰۱	۰٫۲۳۲	-۰٫۱۳۹	۰٫۲۳۲	-۰٫۱۳۹	
$\gamma_1^{(1)}$	-۲٫۰	-۱٫۸۵۲	۰٫۱۷۳	-۲٫۱۳۹	۰٫۱۵۰	-۲٫۰۳۹	۰٫۱۵۰	-۲٫۰۳۹	
$\gamma_0^{(2)}$	۱٫۰	۱٫۱۹۱	۰٫۱۵۵	۱٫۱۸۰	۰٫۱۵۲	۱٫۰۷۳	۰٫۱۵۲	۱٫۰۷۳	
$\gamma_1^{(2)}$	۱٫۰	۰٫۸۶۵	۰٫۲۱۴	۰٫۹۰۷	۰٫۲۲۱	۰٫۸۷۹	۰٫۲۲۱	۰٫۸۷۹	
$\alpha_{11}^{(2)}$	۱٫۰	۱٫۱۳۷	۰٫۲۹۱	۱٫۲۱۱	۰٫۲۵۵	۱٫۱۲۵	۰٫۲۵۵	۱٫۱۲۵	
$\alpha_{21}^{(2)}$	۱٫۰	۰٫۸۶۰	۰٫۱۱۲	۰٫۸۹۶	۰٫۱۰۰	۰٫۹۶۲	۰٫۱۰۰	۰٫۹۶۲	

مربوط به نسبت آمیزنده پاسخ ترتیبی فارغ از اثر تصادفی است و فرض می‌شود $b_i^{(2)} = b_i^{(1)}$. اثر تصادفی Y_{it} برای $m = 50, 75, 100, 500$ آزمودنی به طور مستقل از توزیع نرمال استاندارد تولید می‌شود. متغیر

از توزیع نرمال با میانگین $\mu_{it} = \beta_0 + \beta_1 X_{1it} + \lambda_1 b_i$ و واریانس σ^2 و \tilde{Y}_{it} از مدل (۴) با

$$\eta_{j,it} = \frac{\exp(\theta_j - (\gamma_1^{(1)} X_{1it} + \lambda_1 b_i^{(1)}))}{1 + \exp(\theta_j - (\gamma_1^{(1)} X_{1it} + \lambda_1 b_i^{(1)}))} \quad j = 0, 1 \quad (9)$$

و نسبت آمیزنده

$$\pi_{it} = \frac{\exp(\gamma_0^{(2)} + \gamma_1^{(2)} X_{1it})}{1 + \exp(\gamma_0^{(2)} + \gamma_1^{(2)} X_{1it})}$$

تولید می‌شود. فرض می‌شود امکان گم‌شدگی غیرتصادفی در پاسخ ترتیبی در زمان اول وجود دارد. احتمال گم‌شدگی برای داده‌های ناکامل به صورت زیر ا

$$Pr(R_{\tilde{Y}_{i1}} = 1 | X_{1i1}, \tilde{y}_{i1}) = \frac{\exp(\alpha_{11}^{(2)} X_{1i1} + \alpha_{21}^{(2)} \tilde{y}_{i1})}{1 + \exp(\alpha_{11}^{(2)} X_{1i1} + \alpha_{21}^{(2)} \tilde{y}_{i1})}$$

است. مقادیر واقعی پارامترها برای تولید داده در دو اندازه نمونه ۱۰۰ و ۵۰۰ عبارت است از $(\beta_0, \beta_1, \theta_1, \gamma_1^{(1)}, \gamma_0^{(2)}, \gamma_1^{(2)}, \lambda_1, \sigma^2) = (0.1, 1, 0.8, 0.5, 1, 1, 0.25, 1)$

و در دو اندازه نمونه ۵۰ و ۷۵ عبارت است از

$$(\beta_0, \beta_1, \theta_1, \gamma_1^{(1)}, \gamma_0^{(2)}, \gamma_1^{(2)}, \lambda_1, \sigma^2) = (0.1, 0.5, 0.8, 0.5, 1, 1, 0.25, 1)$$

مقادیر واقعی پارامترهای احتمال گم‌شدگی برای تولید داده‌های آمیخته با مقادیر گم‌شدگی غیرتصادفی به صورت $(\alpha_{11}^{(2)}, \alpha_{21}^{(2)}) = (0.5, 0.1)$ است.

جدول ۵. نتایج شبیه‌سازی بخش ۳.۲ با ۱۰۰ تکرار شبیه‌سازی

$m = 500$				$m = 100$				مقدار واقعی	پارامتر
(ب)		(الف)		(ب)		(الف)			
SE	برآورد	SE	برآورد	SE	برآورد	SE	برآورد		
۰/۱۸۱	۰/۱۱۳	۰/۰۹۵	۰/۱۹۴	۰/۳۶۸	۰/۳۱۱	۰/۴۵۰	۰/۳۴۰	β_0	
۰/۰۶۵	۰/۶۱۴	۰/۰۷۶	۰/۵۸۶	۰/۲۵۳	۰/۶۳۹	۰/۲۱۹	۰/۶۷۸	β_1	
۰/۰۷۹	۰/۲۴۲	۰/۰۷۲	۰/۲۳۸	۰/۱۸۹	۰/۲۱۴	۰/۱۶۰	۰/۲۳۵	λ_1	
۰/۰۸۷	۱/۰۷۹	۰/۰۵۸	۱/۰۰۲	۰/۱۱۹	۰/۸۳۱	۰/۱۱۷	۰/۹۰۰	σ^2	
۰/۰۹۹	۰/۸۰۶	۰/۱۱۷	۰/۸۱۹	۰/۱۹۱	۰/۸۲۸	۰/۲۴۸	۰/۸۱۵	θ_1	
۰/۱۰۰	۰/۵۱۱	۰/۰۸۸	۰/۵۲۷	۰/۱۹۶	۰/۶۲۷	۰/۲۸۶	۰/۶۸۱	$\gamma_1^{(1)}$	
۰/۱۰۹	۰/۷۵۱	۰/۰۸۹	۰/۶۶۸	۰/۳۱۲	۰/۷۱۴	۰/۴۰۳	۰/۵۵۲	$\gamma_0^{(2)}$	
۰/۱۲۱	۱/۱۸۰	۰/۱۰۹	۱/۲۰۳	۰/۲۹۲	۱/۱۶۲	۰/۳۲۴	۱/۱۰۸	$\gamma_1^{(2)}$	
۰/۰/۱۹۱	۰/۵۰۹	۰/۲۱۷	۰/۴۹۷	۰/۳۶۱	۰/۵۵۸	۰/۴۷۱	۰/۴۸۱	$\alpha_{11}^{(2)}$	
۰/۱۸۹	۰/۱۶۷	۰/۲۳۴	۰/۱۷۴	۰/۳۴۸	۰/۲۰۷	۰/۴۴۳	۰/۲۰۶	$\alpha_{21}^{(2)}$	

جدول ۶. نتایج شبیه‌سازی بخش ۳.۲ با ۱۰۰۰۰ تکرار شبیه‌سازی

پارامتر	مقدار واقعی	$m = 50$				$m = 75$			
		الف		ب		الف		ب	
		برآورد	SE	برآورد	SE	برآورد	SE	برآورد	SE
β_0	۰٫۱۰	۰٫۳۱۸	۰٫۳۶۷	۰٫۲۹۰	۰٫۳۷۹	۰٫۳۵۶	۰٫۲۹۱	۰٫۲۷۶	۰٫۳۱۶
β_1	۰٫۵۰	۰٫۶۶۵	۰٫۲۲۰	۰٫۶۴۱	۰٫۲۴۹	۰٫۵۹۰	۰٫۰۷۶	۰٫۶۰۴	۰٫۰۷۰
λ_1	۰٫۲۵	۰٫۳۲۲	۰٫۱۴۲	۰٫۲۸۷	۰٫۲۰۵	۰٫۲۲۸	۰٫۱۱۹	۰٫۲۵۷	۰٫۱۱۸
σ^2	۱٫۰۰	۱٫۱۴۰	۰٫۱۷۰	۱٫۲۰۴	۰٫۱۷۱	۱٫۰۹۶	۰٫۱۲۱	۰٫۹۲۵	۰٫۱۲۱
θ_1	۰٫۸۰	۰٫۹۰۲	۰٫۲۰۱	۰٫۸۹۹	۰٫۱۴۳	۰٫۹۳۶	۰٫۱۰۵	۰٫۹۲۲	۰٫۱۳۷
$\gamma_1^{(1)}$	۰٫۵۰	۰٫۶۶۱	۰٫۲۲۵	۰٫۶۱۲	۰٫۳۴۷	۰٫۵۸۹	۰٫۰۸۱	۰٫۵۸۳	۰٫۰۸۹
$\gamma_2^{(1)}$	۱٫۰۰	۰٫۶۰۳	۰٫۳۲۱	۰٫۶۱۰	۰٫۳۱۱	۰٫۶۵۲	۰٫۲۸۴	۰٫۵۴۴	۰٫۲۹۰
$\gamma_1^{(2)}$	۱٫۰۰	۱٫۱۲۹	۰٫۳۷۴	۱٫۳۸۰	۰٫۳۲۱	۱٫۲۰۹	۰٫۳۱۲	۱٫۲۰۲	۰٫۳۰۵
$\alpha_1^{(2)}$	۰٫۵۰	۰٫۶۳۱	۰٫۴۰۲	۰٫۶۴۹	۰٫۴۰۶	۰٫۵۱۳	۰٫۳۰۹	۰٫۵۲۰	۰٫۲۶۹
$\alpha_1^{(1)}$	۰٫۱۰	۰٫۲۶۱	۰٫۳۸۸	۰٫۲۳۱	۰٫۲۷۳	۰٫۲۱۰	۰٫۲۸۶	۰٫۲۱۵	۰٫۳۲۹

به منظور برازش مدل‌های مورد بررسی به نمونه‌های تولید شده، برآوردهای بیشینه درست‌نمایی پارامترهای مدل‌های توأم با بیشینه کردن تابع درست‌نمایی $L(\xi)$ با استفاده از تابع optim، در نرم‌افزار R به دست می‌آید. در حالت (الف) مقادیر اولیه برای این تابع همان مقادیر واقعی پارامترها هستند و در حالت (ب) مقادیر اولیه مقادیر تصادفی هستند. به عبارتی در حالت (ب)، در هر تکرار شبیه‌سازی با تولید اعداد تصادفی به تعداد پارامترهای مدل از توزیع یکنواخت پیوسته بر بازه صفر و یک، هر یک از مقادیر واقعی پارامترها با مقداری تصادفی جمع می‌شود و به‌عنوان مقدار اولیه در تابع optim استفاده می‌شوند. جداول ۱ تا ۶ نتایج مطالعات شبیه‌سازی را ارائه می‌دهند. ماتریس هسی مشاهده‌شده مربوط به لگاریتم تابع درست‌نمایی با استفاده از تابع fdHess در بسته nlme در نرم‌افزار R محاسبه می‌شود. سپس خطای استاندارد (SE) برآورده شده از ریشه دوم عناصر قطری ماتریس وارون هسی برای برآوردهای همه پارامترها به دست می‌آید. در هر دو حالت (الف) و (ب) که دو شیوه برای انتخاب مقادیر اولیه تابع optim در بیشینه کردن عددی تابع درست‌نمایی هستند، برآوردها به مقادیر واقعی پارامترها نزدیک هستند. این مسئله علاوه بر تأیید کارایی برآوردها، نشان می‌دهد که در این مطالعه نتایج شبیه‌سازی به انتخاب مقادیر اولیه برای تابع optim وابسته نیست. همچنین خطاهای استاندارد برای حجم نمونه بزرگتر مقدار کوچکتری دارند و این ویژگی سازگاری برآوردهای بیشینه درست‌نمایی و عملکرد مناسب مدل‌های مورد بررسی را نشان می‌دهد.

۴ کاربرد

در این بخش دو مجموعه داده واقعی تغییرات زندگی آمریکایی‌ها ACL و آزمون موفقیت فردی پیبادی PIAT بررسی می‌شوند. در مدل‌های بخش کاربردی همانند بخش شبیه‌سازی، $b_i^{(1)}$ توزیع نرمال استاندارد دارد. برای هر مدل معیار AIC و BIC نیز محاسبه می‌شود، که $AIC = -2 \log(L(\xi)) + 2d$ و $BIC = -2 \log(L(\xi)) + d(\log(m))$ ، که $d = \dim(\xi)$ و m تعداد آزمودنی‌ها است. مجدداً ماتریس هسی مشاهده‌شده برای لگاریتم درست‌نمایی با استفاده از تابع fdHess در بسته nlme در نرم‌افزار R محاسبه می‌شود. سپس خطاهای استاندارد برآوردشده برای همه برآوردهای پارامترها از ریشه‌های دوم عناصر قطری ماتریس وارون هسی محاسبه می‌شوند.

۴.۱ داده‌های ACL

سلسله مطالعه تغییرات زندگی آمریکایی‌ها (ACL) یک مطالعه طولی ملی طولانی‌مدت است که روی آمریکایی‌هایی متمرکز دارد که در میانه یا انتهای زندگی خود هستند. موج اول این داده‌ها در سال ۱۹۸۶ با یک آمارگیری رودرو در سطح کشور، از ۳۶۱۷ آمریکایی بزرگسال ۲۵ ساله و بالاتر، سیاه‌پوستان آمریکایی و افراد ۶۰ ساله و بالاتر با سهم نمونه‌ای دو برابر دیگران، آغاز شد. موج دوم شامل مصاحبه‌های مجدد رودرو در سال ۱۹۸۹ با کسانی که هنوز زنده بودند، است.

زیرنمونه‌ای تصادفی به حجم ۲۷۹ مربوط به موج‌های اول و دوم از این داده متغیر وزن و تعداد روزهای مصرف الکل در ماه قبل متغیرهای پاسخ هستند، که شریفیان و همکاران (۲۰۲۱) با در نظر گرفتن این دو متغیر پاسخ، یک مدل توأم برای پاسخ‌های آمیخته مجموعه آماسیده پیوسته و آماسیده در صفر دو جمله‌ای را به این زیرمجموعه تصادفی از داده‌ها برازش دادند. گمشدگی در دومین نقطه زمانی برای پاسخ شمارشی طبق یک سازوکار غیرتصادفی تولید می‌شود. حال این مجموعه داده با گمشدگی شبیه‌سازی شده مورد ارزیابی قرار می‌گیرد. در این داده‌ها تعداد زیادی از افراد هرگز الکل مصرف نکرده‌اند، و افرادی نیز وجود دارند که الکل مصرف می‌کنند اما در ماه قبل هیچ روزی الکل ننوشیده‌اند. بنابراین به نظر می‌رسد یک مدل ZIB یک مدل مناسب برای این پاسخ شمارشی باشد. زیرا تعداد زیاد صفرها و تولید شدن آن‌ها توسط دو منبع یکی تصادفی و دیگری غیرتصادفی باعث می‌شود چنین مدلی مناسب باشد. از نظر پزشکی، اضافه وزن به‌عنوان یکی از اثرات مصرف بالای الکل در نظر گرفته می‌شود (فرانچ و همکاران، ۲۰۱۰). بنابراین به دلیل ارتباط وزن و مصرف الکل، درصد بالای افراد غیرالکلی و همچنین مصرف کم الکل، انتظار می‌رود

جدول ۰۷. آمار توصیفی متغیر وزن و تعداد روزهای مصرف الکل در ماه گذشته (DD) برای زیرمجموعه داده‌های ACL در موج اول و دوم

متغیر پاسخ	آماره	موج اول	موج اول
تعداد روزهای مصرف الکل (DD)	میانگین	۴۸۸	۳۵۸
	چارک اول	۰/۰۰	۰/۰۰
	میانه	۰/۰۰	۰/۰۰
	چارک سوم	۳/۰۰	۲/۰۰
	انحراف استاندارد	۷/۰۴	۶/۶۱
	نما	۰/۰۰	۰/۰۰
	کمینه	۰/۰۰	۰/۰۰
	بیشینه	۳۱/۰۰	۳۱/۰۰
	درصد (DD = ۰)	۵۳/۴۱	۵۶/۹۹
	وزن	میانگین	۱۶۳/۰۰
چارک اول		۱۴۰/۰۰	۱۳۵/۰۰
میانه		۱۶۰/۰۰	۱۶۰/۰۰
چارک سوم		۱۸۵/۰۰	۱۸۵/۰۰
انحراف استاندارد		۳۶/۹۱	۳۷/۷۲
کمینه		۸/۰۰	۸/۰۰
بیشینه		۳۳۰/۰۰	۳۳۰/۰۰
برجستگی چولگی		۴/۳۹	۴/۲۴
		۰/۸۷	۰/۷۷

که آماسیدگی در یک محدوده خاص از متغیر وزن رخ دهد. ویژگی‌های توصیفی این دو پاسخ در جدول ۷ ارائه شده است.

با انجام یک آزمون ناپارامتری کولموگروف-اسمیرنوف و یک آزمون نسبت درستنمایی بوت‌استریپی فرض نرمال بودن متغیر وزن در دو زمان رد و فرض توزیع مجموعه آماسیده را تأیید می‌شود. همچنین با یک آزمون نسبت درستنمایی مناسب، فرض آماسیدگی در صفر برای پاسخ شمارشی را پذیرفته می‌شود. بازه (۱۲۰, ۲۰۰) به‌عنوان مقدار بازه (a, b) در نظر گرفته می‌شود. برای مشاهده جزئیات درباره انتخاب مقادیر a و b به شریفیان و همکاران (۲۰۲۱) مراجعه شود. رویکرد انتخاب اثر تصادفی دقیقاً مشابه با بخش شبیه‌سازی است. جنسیت (X_{1i}) ، سن (X_{2it}) ، میزان ساعت خواب در ۲۴ ساعت شبانه روز (X_{3it}) (HS) به‌عنوان متغیرهای کمکی در مدل متغیرهای پاسخ در نظر گرفته می‌شود. مدل توأم زیر

برای دو پاسخ آمیخته طولی به همراه سازوکار گم‌شدگی غیرقابل چشم‌پوشی در نظر گرفته می‌شود:

$$\begin{aligned}
 Y_{it}|b_i^{(1)}, \mu_{it}, \pi_{A_{it}}, \sigma^2 &\sim (120, 200)IN(\mu_{it}, \pi_{A_{it}}, \sigma^2), \quad g_A \sim U(120, 200), \\
 \mu_{it} &= \beta_0^{(1)} + \beta_1^{(1)}X_{1i} + \beta_2^{(1)}X_{rit} + \beta_3^{(1)}X_{rit} + \lambda_1 b_i^{(1)}, \\
 \text{logit}(\pi_{A_{it}}) &= \beta_0^{(2)} + \beta_1^{(2)}X_{1i} + \beta_2^{(2)}X_{rit}, \\
 N_{it}|b_i^{(1)}, \nu_{it}, \pi_{\circ it} &\sim ZIB(31, \frac{\nu_{it}}{1 + \nu_{it}}, \pi_{\circ it}), \\
 \log(\nu_{it}) &= \gamma_0^{(1)} + \gamma_1^{(1)}X_{1i} + \gamma_2^{(1)}X_{rit} + \gamma_3^{(1)}X_{rit} + \lambda_1 b_i^{(1)}, \\
 \text{logit}(\pi_{\circ it}) &= \gamma_0^{(2)} + \gamma_1^{(2)}X_{1i} + \gamma_2^{(2)}X_{rit}, \\
 \text{logit}(Pr(R_{n_{i2}} = 1|X_{1i}, n_{i2})) &= \alpha_{12}^{(2)}X_{1i} + \alpha_{22}^{(2)}n_{i2}. \quad (10)
 \end{aligned}$$

اگر در مدل (۱۰) ضریب $\alpha_{22}^{(2)}$ معنی‌دار شود، یعنی سازوکار گم‌شدگی MNAR و از این رو گم‌شدگی غیرقابل چشم‌پوشی، اگر فقط ضریب $\alpha_{12}^{(2)}$ معنی‌دار باشد، یعنی سازوکار گم‌شدگی MAR و اگر هیچ‌یک از این دو ضریب معنی‌دار نباشند یعنی سازوکار گم‌شدگی MCAR است. جدول ۸ نتایج تحلیل پاسخ‌های پیوسته و شمارشی با سازوکار گم‌شدگی غیرقابل چشم‌پوشی را نمایش می‌دهد. مشاهده می‌شود که اثر سن، جنسیت و میزان خواب در طول ۲۴ ساعت بر وزن، اثر سن بر نسبت آمیزنده وزن در بازه (۱۲۰، ۲۰۰)، و اثر سن و میزان خواب در طول ۲۴ ساعت بر میزان مصرف الکل، و اثر جنسیت بر نسبت آمیزنده پاسخ شمارشی در مقدار ۰ در سطح $\alpha = 0.05$ معنی‌دار است. در قسمت سازوکار گم‌شدگی، پارامتر $\alpha_{22}^{(2)}$ معنی‌دار است. به این ترتیب سازوکار گم‌شدگی از نوع MNAR و غیرقابل چشم‌پوشی است. ضریب متغیر کمکی در مدل احتمال گم‌شدگی معنی‌دار نیست. همچنین، پارامتر λ_1 ، پارامتر مربوط به همبستگی مدل، نیز در سطح $\alpha = 0.05$ معنی‌دار است.

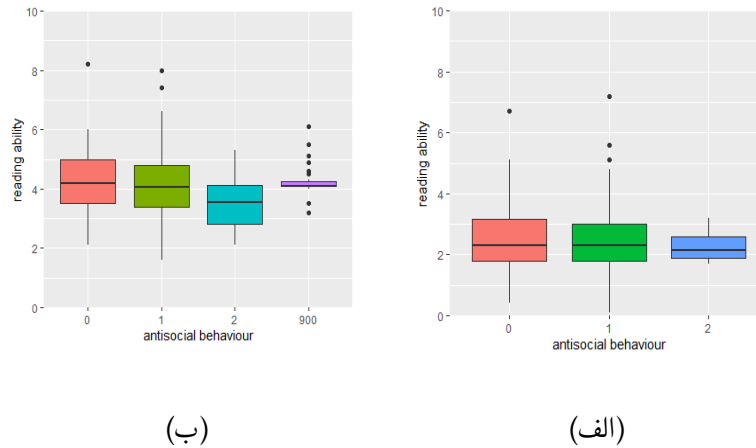
۴.۲ داده‌های PIAT

داده‌های PIAT که قبلاً توسط شریفیان و همکاران (۲۰۲۰) مورد تحلیل قرار گرفته است، در بین سال‌های ۱۹۸۶ و ۱۹۹۲ جمع‌آوری شده و شامل مصاحبه‌های دو سال یک بار از مادرها و فرزندهایشان است. هدف اصلی مطالعه، ارزیابی روابط بین پاسخ مادران بر رفتار ضد اجتماعی فرزندان و توانایی خواندن فرزندان است. این دو متغیر همبسته متغیرهای پاسخ آمیخته مورد بررسی خواهند بود به‌طوری

جدول ۸. نتایج برآورد مدل (۱^۰) برای داده‌های ACL وقتی که g, A چگالی توزیع $U(a, b)$ است (برآوردهای پارامترهایی که در سطح $\alpha = ۰/۰۵$ معنی‌دار هستند با علامت ستاره در توان مشخص شده‌اند).

بخش مرتبط به مدل	متغیر کمکی	پارامتر	برآورد	خطای استاندارد (SE)
پاسخ پیوسته	عرض از مبدأ	$\beta_0^{(1)}$	۲۴۴۵۹۰*	۱۳۲۵
	جنسیت	$\beta_1^{(1)}$	-۳۸۹۹۰*	۴۹۶۹
	سن	$\beta_2^{(1)}$	-۰۵۲۴*	۰/۱۵۸
	میزان خواب	$\beta_3^{(1)}$	-۳۲۷۶*	۱۳۸۴
	عرض از مبدأ	$\beta_0^{(2)}$	۱۹۴۴	۱۷۱۵
	جنسیت	$\beta_1^{(2)}$	۱۷۲۸	۲۰۵۱
سن	$\beta_2^{(2)}$	-۰۳۷۷*	۰/۱۳۶	
پاسخ شمارشی	عرض از مبدأ	$\gamma_0^{(1)}$	۱۵۵۳*	۰/۰۷۴
	جنسیت	$\gamma_1^{(1)}$	-۰۳۶۹	۰/۱۹۵
	سن	$\gamma_2^{(1)}$	-۰/۰۹۳*	$1.09e^{-2}$
	میزان خواب	$\gamma_3^{(1)}$	-۰/۰۸۵*	۰/۰۱۸
	عرض از مبدأ	$\gamma_0^{(2)}$	-۰۵۲۰*	$8.49e^{-5}$
	جنسیت	$\gamma_1^{(2)}$	۱۴۶۷*	$7.91e^{-2}$
سن	$\gamma_2^{(2)}$	-۱/۱۱۹	۹/۴۱۴	
سازوکار گم‌شدگی	سن	$\alpha_{12}^{(2)}$	۰/۶۱۴	۲/۰۲۲
	وزن	$\alpha_{22}^{(2)}$	۱۰/۴۸۴*	۰/۰۰۱
		λ_1	۳/۲۶۳*	۰/۱۵۶
		σ^2	۲۷۲۳۷۰۲*	۰/۲۱۰
	Log-likelihood		-۴۰۸۵۹۴/۵	
	AIC		۸۱۷۲۵۵/۰	
	BIC		۸۱۷۲۹۰/۳	

که پاسخ ترتیبی رفتار ضد اجتماعی فرزند (*anti*) و پاسخ پیوسته توانایی خواندن فرزند (*read*) هستند. زیر مقیاس‌های رفتار ضد اجتماعی شامل گزارش مادر بر شش قلم است که میزان رفتار ضد اجتماعی فرزند برای سه ماه قبل از مصاحبه را مورد ارزیابی قرار می‌دهد. این قلم‌ها عبارتند از فرزند ۱-تقلب کند یا دروغ بگوید، ۲-قلدری کند، یا ظالم باشد و یا با دیگران بدجنس باشد، ۳-وقتی بد رفتاری می‌کند از رفتارش پشیمان نشود، ۴-برای رسیدن به هدفش اشیا را بشکند و یا از روی عمد اشیا خود و دیگران را تخریب کند، ۵-در مدرسه از دستورات سرپیچی کند و ۶-با معلم خود مشکل داشته باشد. در اینجا، پاسخ ترتیبی بدین صورت رده‌بندی می‌شود: اگر فرزند هیچ یک از این شش قلم ضد اجتماعی را در مدت تعیین شده مرتکب نشود مقدار ۰، اگر بعضی موقع حداقل یکی از آنها را انجام دهد، مقدار ۱، و اگر اغلب حداقل یکی از آنها را انجام دهد، مقدار ۲ به او اختصاص می‌یابد. توانایی خواندن فرزندان نیز طبق آزمونی که از آنها گرفته می‌شوند، به صورت نمره‌ای از مقدار صفر تا ۸/۴ به دست می‌آید. ۴۰۵ فرزند به همراه مادرانشان



شکل ۱. نمودار جعبه‌ای توانایی خواندن برای سطوح مختلف رفتار ضد اجتماعی کودکان در الف- اولین دوره زمانی، ب- دومین دوره زمانی

در این مطالعه شرکت کرده‌اند. در دوره اول مطالعه هیچ یک از متغیرهای پاسخ دارای گم‌شدگی نیستند ولی در دوره دوم مطالعه شدت ناهنجاری دارای گم‌شدگی است (۸ درصد داده گم‌شده برای شدت ناهنجاری در دوره دوم). متغیرهای پیشگو شامل جنسیت فرزندان (X_1 ، دختر=۰ و پسر=۱)، سن فرزندان در زمان شروع مطالعه (X_2 ، از سن ۶ تا سن ۸)، سن مادرها در زمان شروع مطالعه (X_3 ، از سن ۲۱ تا سن ۲۹)، و زمان ($X_t = t, t = 1, 2$)، متناظر با اولین و دومین دوره زمانی) است که در تحلیل استفاده می‌شوند. شکل ۱ نمودار جعبه‌ای پاسخ‌های توانایی خواندن برای مقادیر مختلف از سطوح رفتارهای ضد اجتماعی را نشان می‌دهد. بر اساس این شکل، فرزندان با رفتارهای ضد اجتماعی کمتر دارای توانایی خواندن بیشتری نسبت به فرزندان با رفتارهای ضد اجتماعی بیشتر در تمام دوره‌های زمانی تعیین شده هستند. همچنین توانایی خواندن در طول زمان بیشتر می‌شود. آزمون کولموگروف-اسمیرنوف فرض نرمال بودن متغیر توانایی خواندن را تأیید می‌کند. همچنین فرض آماسیدگی پاسخ ترتیبی، رفتار ضد اجتماعی، در سطح دوم (مقدار ۱) در دو زمان با استفاده از یک آزمون نسبت درستنمایی مناسب، تأیید می‌شود ($P - value < 0.0001$).

مطلوب است که همزمان با بررسی تأثیر جنسیت کودک، سن کودک و سن مادر کودک، نوع گم‌شدگی در زمان دوم برای پاسخ *anti* نیز مشخص شود. بنابراین، مدل توأم زیر برای دو پاسخ آمیخته طولی به

جدول ۹. نتایج برآورد مدل نتایج برآورد مدل (۱۱) برای داده‌های PIAT

بخش مرتبط به مدل	متغیر کمکی	پارامتر	برآورد	خطای استاندارد (SE)
پاسخ پیوسته	عرض از مبدأ	β_0	۰٫۸۴۲*	۰٫۰۸۸
	جنسیت	β_1	-۰٫۰۲۱	۰٫۰۱۷
	سن کودک	β_2	۰٫۴۸۴*	۰٫۰۳۷
	سن مادر کودک زمان	β_3 β_4	۰٫۱۰۷* ۱٫۵۹۶	۰٫۰۳۰ ۰٫۰۵۴
پاسخ ترتیبی	عرض از مبدأ	θ_1	۲٫۴۰۶*	۰٫۲۹۶
	جنسیت	$\gamma_1^{(1)}$	-۰٫۸۳۶*	۰٫۲۴۶
	سن کودک	$\gamma_2^{(1)}$	۰٫۱۳۷*	۰٫۰۴۶
	سن مادر کودک زمان	$\gamma_3^{(1)}$ $\gamma_4^{(1)}$	-۰٫۰۸۲* ۰٫۱۹۸*	۰٫۰۱۶ ۰٫۰۷۰
سازوکار گم‌شدگی	عرض از مبدأ	$\gamma_0^{(2)}$	۰٫۴۴۷*	۰٫۱۲۸
	جنسیت	$\gamma_1^{(2)}$	-۰٫۰۸۸*	۰٫۰۳۷
	سن کودک	$\gamma_2^{(2)}$	-۰٫۲۲۵	۰٫۱۳۵
	سن مادر کودک رفتار ضد اجتماعی	$\alpha_{12}^{(2)}$ $\alpha_{22}^{(2)}$ $\alpha_{32}^{(2)}$ $\alpha_{42}^{(2)}$	۰٫۸۵۸* ۰٫۳۴۲* -۰٫۱۸۰ ۰٫۷۷۹*	۰٫۲۲۱ ۰٫۰۲۷ ۰٫۱۷۹ ۰٫۱۴۹
		λ_1	۰٫۴۶۶*	۰٫۰۵۲
		σ^2	۰٫۵۳۶*	۰٫۰۶۱
		Log-likelihood	-۴۸۳۹۴۸۴	
		AIC	۹۶۷۹۳۴۸	
		BIC	۹۶۸۰۱۰۹	

همراه سازوکار گم‌شدگی غیرقابل چشم‌پوشی در نظر گرفته می‌شود:

$$\begin{aligned}
 read_{it} &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \lambda_1 b_i^{(1)} + \epsilon_{it} \\
 \text{logit}(\eta_{0,it}) &= \theta_0 - (\gamma_1^{(1)} X_{1i} + \gamma_2^{(1)} X_{2i} + \gamma_3^{(1)} X_{3i} + \gamma_4^{(1)} X_{4i} + \lambda_1 b_i^{(1)}), \\
 \text{logit}(\eta_{1,it}) &= \theta_1 - (\gamma_1^{(1)} X_{1i} + \gamma_2^{(1)} X_{2i} + \gamma_3^{(1)} X_{3i} + \gamma_4^{(1)} X_{4i} + \lambda_1 b_i^{(1)}), \\
 \text{logit}(\pi_i) &= \gamma_0^{(2)} + \gamma_1^{(2)} X_{1i} + \gamma_2^{(2)} X_{2i}, \\
 \text{logit}(Pr(R_{\tilde{y}_{i2}} = 1 | X_i, \tilde{y}_{i2})) &= \alpha_{12}^{(2)} X_{1i} + \alpha_{22}^{(2)} X_{2i} + \alpha_{32}^{(2)} X_{3i} + \alpha_{42}^{(2)} \tilde{y}_{i2} \quad (11)
 \end{aligned}$$

که ϵ_{it} دارای توزیع نرمال با میانگین صفر و واریانس σ^2 است. اگر در مدل (۱۱) ضریب $\alpha_{42}^{(2)}$ معنی‌دار شود، یعنی سازوکار گم‌شدگی MNAR و از این رو گم‌شدگی غیرقابل چشم‌پوشی، اگر یکی از ضرایب دیگر معنی‌دار باشد، یعنی سازوکار گم‌شدگی MAR و اگر هیچ یک از این ضرایب معنی‌دار نباشند یعنی

سازوکار گم‌شدگی MCAR است.

جدول ۵ نتایج تحلیل پاسخ‌های پیوسته و ترتیبی با سازوکار گم‌شدگی غیرقابل چشم‌پوشی را نمایش می‌دهد. مشاهده می‌شود که اثر سن کودک، سن مادر و زمان بر توانایی خواندن کودک، اثر جنسیت کودک، سن کودک و زمان بر شدت رفتار ضداجتماعی، و اثر جنسیت بر نسبت آمیزنده پاسخ ترتیبی در رده میانی در سطح $\alpha = 0.05$ معنی‌دار است. در قسمت سازوکار گم‌شدگی، علاوه بر جنسیت و سن کودک، پارامتر $\alpha_{22}^{(2)}$ مربوط به توانایی خواندن کودک، نیز معنی‌دار است. به این ترتیب سازوکار گم‌شدگی از نوع MNAR و غیرقابل چشم‌پوشی است. همچنین، پارامتر λ_1 ، پارامتر مربوط به همبستگی مدل، نیز در سطح $\alpha = 0.05$ معنی‌دار است.

بحث و نتیجه‌گیری

در این مقاله سازوکار گم‌شدگی غیرقابل چشم‌پوشی در دو مدل توأم برای پاسخ‌های آمیخته مجموعه آماسیده پیوسته و آماسیده در صفر شمارشی و پاسخ‌های آمیخته پیوسته و k -آماسیده ترتیبی مورد بررسی قرار گرفت. در واقع، این دو مدل دو حالات خاص پاسخ‌های آمیخته با امکان گم‌شدگی غیرتصادفی را مورد بررسی قرار دادند. مدل‌ها در چند مطالعه شبیه‌سازی مورد بررسی قرار گرفتند و به دو مجموعه داده واقعی PIAT و ACL برازش داده شدند. معنی‌داری پارامترهای مربوط به سازوکار گم‌شدگی غیرتصادفی در داده‌های واقعی مشاهده شد. بنابراین در این دو مجموعه داده، سازوکار گم‌شدگی غیرقابل چشم‌پوشی است و مدل‌بندی سازوکار گم‌شدگی غیرتصادفی ضروری است. برآورد پارامترها در چنین مدل‌هایی با روش‌های دیگر و الگوریتم‌های مناسب برای دستیابی به برآوردهای بیشینه درست‌نمایی در پژوهش‌های آینده مورد بررسی قرار می‌گیرد.

تقدیر و تشکر

از داوران و ویراستار محترم مجله که با توصیه‌های ارزشمند خود سبب ارتقای این مقاله و ارائه بهتر آن شده‌اند کمال تشکر را داریم.

مراجع

اسماعیل‌زاده، م. و بهرامی سامانی، ا. (۱۳۹۸)، مدل‌های اثر تصادفی دو متغیره آماسیده برای پاسخ‌های آمیخته سری توانی نرمال، *مجله علوم آماری*، ۱۳، ۲۷۵-۳۰۰.

میری، س. م. و بهرامی سامانی، ا. (۱۳۹۱)، مدل مکانی عام برای پاسخ‌های همبسته پیوسته، ترتیبی و اسمی، *مجله علوم آماری*، ۶، ۹۹-۱۱۰.

Bahrami Samani, E., Ganjali, M. and Khodadadi, A. (2008), A Latent Variable Model for Mixed Continuous and Ordinal Responses, *Journal of Statistical Theory and Applications*, **7**, 337-349.

Bahrami Samani, E., and Ganjali, M. (2010), A Multivariate Latent Variable Mixed-Data from Continuous and Ordinal Responses with Possibility of Missing Responses, *Application and Applied Mathematics: An international Journal (AAM)*, **5** (10), 1364-1584.

Bahrami Samani, E., Ganjali, M. and Eftekhari, S. (2010), A Latent Variable Model for Mixed Continuous and Ordinal Responses with Nonignorable Missing Responses: Assessing the Local Influence via Covariance Structure, *Sankhya: The Indian Journal of Statistics*, **27** (B), 38-57.

Bahrami Samani, E., and Tahmasebinejad, Zh. (2011), Joint Modeling of Mixed Correlated Nominal, Ordinal and Continuous Responses. *Journal of Statistical Research*, **45**, 37-47.

Bahrami Samani, E., Ganjali, M., and Amirian, Y. (2012), Likelihood Estimation for Longitudinal Zero-Inflated Power Series Regression Models, *Journal of Applied Statistic*, **39**, 1965-1974.

Bahrami Samani, E. (2013), Local Influence in Bayesian Elliptically

Contoured-Ordinal Model for Mixed Data, *Application and Applied Mathematics: An international Journal (AAM)*, **8**, 391-403.

Bahrami Samani, E., and Ganjali, M. (2014), Mixed Correlated Bivariate Ordinal and Negative Binomial Longitudinal Responses with Nonignorable Missing Values, *Communications in Statistics-Theory and Methods*, **43**, 2059-2673.

De Leon, A. R., and Chough, K. C. (2013), *Analysis of Mixed Data, Methods & Applications*, Taylor & Francis.

French, M. T., Notron, E. C., and Fang, H. (2010), Alcohol Consumption and Body Weight, *Health Economics*, **19**, 814-832.

Kelley, M. E., and Anderson, S. J. (2008), Zero Inflation in Ordinal Data: Incorporating Susceptibility to Response Through the Use of a Mixture Model, *Statistics in Medicine*, **27**, 3674-88.

Lambert, D. (1992), Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, **34**, 1-14.

Little, R. J. and Rubin, D. (2002), *Statistical Analysis with Missing Data. Second Edition*, New York, Wiley.

Little, R. J., and Schluchter M. (1985), Maximum Likelihood Estimation for Mixed Continuous and Categorical Data with Missing Values, *Biometrika*, **72**, 497-512.

Mc Culloch, C. (2007), Joint Modeling of Mixed Outcome Type Using Latent Variables, *Statistical Methods in Medical Research*, **17**, 53-73.

Razie, F., Bahrami Samani, E., and Ganjali, M. (2017), Latent Variable Model for Mixed Correlated Power Series and Ordinal Longitudinal Re-

sponses with Nonignorable Missing Values, *Journal Communication in Statistics-Theory and Methods*, **46**, 5738-5753.

Razie, F., Bahrami Samani, E., and Ganjali, M. (2019), Analysis of Mixed Longitudinal (k,l)-Inflated Power Series, Ordinal and Continuous Responses with Sensitivity Analysis to Non-Ignorable Missing Mechanism, *Journal Communication in Statistics-Simulation and Computation*, 1-27. doi: 10.1080/03610918.2019.1601215

Sharifian, N., Bahrami Samani, E., and Ganjali, M. (2021), Joint Modeling for Longitudinal Set-Inflated Continuous and Count Responses, *Communications in Statistics-Theory and Methods*, **50**, 1134-1160.

Sharifian, N., Bahrami Samani, E., and Ganjali, M. (2020), Joint Model for Longitudinal Mixture of Normal and Zero-Inflated Power Series Correlated Responses, *Journal of Biopharmaceutical Statistics*, 1-24. doi:10.1080/10543406.2020.1814798

Teixeira-Pinto, A., and Normand, S. L. T. (2009), Correlate Continuous and Binary Outcomes: Issues and Applications, *Statistics in Medicine*, **28**, 1753-1773.

Wang, W. (2013), Identifiability of Linear Mixed Effects Models, *Electronic Journal of Statistics*, **7**, 244-263.

Yang, Y. and Kang, J.(2010), Joint Analysis of Mixed Poisson and Continuous Longitudinal Data with Nonignorable Missing Values, *Computational Statistics and Data Analysis*, **54**, 193-207.

Journal of Statistical Sciences, Autumn and Winter, 2021
Vol. 15, No. 2, pp 505-531
DOI: 10.29252/jss.15.2.505

Modeling of Continuous-Count and Continuous-Ordinal Mixed Longitudinal Bivariate Inflated Responses with Possibility of Non Random Missingness

Sharifian, N. and Bahrami Samani, E.
Department of Statistics, Shahid Beheshti University, Tehran , Iran.

Abstract: One of the most frequently encountered longitudinal studies issues is data with losing the appointments or getting censoring. In such cases, all of the subjects do not have the same set of observation times. The missingness in the analysis of longitudinal discrete and continuous mixed data is also common, and missing may occur in one or both responses. Failure to pay attention to the cause of the missing (the mechanism of the missingness) leads to unbiased estimates and inferences. Therefore, in this paper, we investigate the mechanism of nonignorable missing in set-inflated continuous and zero-inflation power series, as well as the continuous and k -inflated ordinal mixed responses. A full likelihood-based approach is used to obtain the maximum likelihood estimates of the parameters of the models. In order to assess the performance of the models, some simulation studies are performed. Two applications of our models are illustrated for the American's Changing Lives (ACL) survey, and the Peabody Individual Achievement Test (PIAT) data set.

Keywords: Longitudinal Mixed data, Inflation, Random effects, Missing values.

Mathematics Subject Classification (2010): 62J12.