مجله مدل‌سازی پیشرفته ریاضی، دوره ۴، شماره ۲، پاییز و زمستان ۱۳۹۳

نتایج جدیدی درباره‌ی شناساپذیری مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی

الهام تبریزی، احسان بهرامی سامانی[۱] و ناصح جعفری

بخش آمار، دانشگاه شهید بهشتی

تاریخ دریافت: ۱۳۹۳/۷/۱۷ تاریخ پذیرش: ۱۳۹۴/۴/۱۰

چکیده: شناساپذیری یکی از ویژگی‌های لازم برای کفایت یک مدل آماری است. وقتی مدلی شناساپذیر نباشد، با هیچ اندازه‌ای از نمونه، نمی‌توان پارامتر حقیقی مدل را تعیین کرد. در این مقاله، مروری بر مفهوم مشهور شناساپذیری و ویژگی‌های آن شده است. به‌علاوه ازآنجایی‌که مشکل شناسا-ناپذیری در مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی بسیار رایج است، تمرکز اصلی ما بر روی این‌گونه از مدل‌ها بوده است. از سوی دیگر، معمولاً نرم‌افزارهای آماری، بعد از برازش مدل شناساناپذیر در آن‌ها، اشاره‌ای به این مسئله نکرده و خروجی‌های غیر معتبر ارائه می‌دهند. بنابراین یافتن راهی برای بررسی شناساپذیری مدل، قبل از برازش آن، خالی از فایده نخواهد بود. در این راستا، قضایای جدیدی در رابطه با شناساپذیری مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی بیان‌شده است. همچنین برای تشریح سودمندی قضایای مطرح‌شده، چند مطالعه‌ی شبیه‌سازی روی مدل‌های شناساناپذیر خطی تعمیم‌یافته و خطی تعمیم‌یافته با اثرهای تصادفی انجام‌شده و مشکل‌های حاصل از برازش آن‌ها موردبررسی قرارگرفته است.

واژه‌های کلیدی: شناساپذیری، مدل‌های آمیخته‌ی خطی، مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی.

رده‌بندی ریاضی (۲۰۱۰): ۶۲J۰۵، ۶۲J۱۲.

۱- مقدمه

در استنباط‌های آماری، یک مجموعه از داده‌ها، مشاهده‌هایی از یک آزمایش تصادفی تعریف‌شده روی فضای احتمال (Ω, F, P) است که اندازه‌ی احتمال P، جامعه نامیده می‌شود [۱]. هر مدل آماری، مجموعه‌ای از فرض‌ها را روی جامعه‌ی P اعمال می‌کند و بیان می‌دارد

۱- آدرس الکترونیکی نویسنده مسئول مقاله: ehsan_bahrami_samani@yahoo.com

که جامعه‌ی P به خانواده‌ای پارامتری مثل $\{P_{\theta}: \theta \in \Theta\}$ تعلق دارد. شناساپذیری، یک ویژگی در مور توزیع احتمال مشاهده‌ها و یا مدل است که در صورت برقرار نبودن این ویژگی، برآورد پارامترهای مدل و تحلیل آن بی‌معنی خواهد بود. این ویژگی زمانی برقرار است که نگاشت $\theta \to P_{\theta}$، نگاشتی یک‌به‌یک است. در واقع این خاصیت، یکتایی توزیع به‌عنوان تابعی از پارامترهای مدل را بیان می‌کند. از سوی دیگر، برآورد پارامترهای هر مدل شناساناپذیر، امری غیرمعقول بوده و در صورت برآورد کردن آن‌ها حتی با در دسترس داشتن تعداد نامتناهی از مشاهده‌ها (کل جامعه)، به مقدار حقیقی پارامتر مدل دست نخواهیم یافت.

بررسی شناساپذیری مدل با استفاده از تعریف اصلی آن و قبل از برازش مدل، توسط [۲] انجام‌شده است. یکی از شایع‌ترین حالت‌هایی که با مشکل شناساناپذیری مدل مواجه می‌شویم، هنگام برازش مدل‌های آمیخته‌ی خطی است. محققهای زیادی مانند [۳ و ۴] مدل‌های آمیخته‌ی خطی را موردمطالعه قرار داده‌اند.

شناساناپذیری مدل‌های آمیخته‌ی خطی وقتی‌که یکی از اثرهای تصادفی و خطاها باهم آمیخته شده‌اند نیز به‌صورت مختصر توسط [۵] بررسی‌شده است. همچنین در مدل‌های آمیخته‌ی خطی، می‌توان ساختارهای متفاوتی برای ماتریس کوواریانس پاسخ در نظر گرفت. این ماتریس کوواریانس، حاصل مجموع دو ماتریس کوواریانس مربوط به بردار اثرهای تصادفی و خطاها خواهد بود. ازاین‌رو، انتخاب ساختار کوواریانس مناسب برای بردار اثرهای تصادفی و خطاها امری اجتناب‌ناپذیر است. لیست جامعی از این ساختارها توسط [۶ و ۷] ارائه شده است. در این مدل‌ها، حتی اگر ماتریس کوواریانس پاسخ، بیش پارامتری شده نباشد (بیش پارامتری شدن ماتریس کوواریانس به این معنا است که تعداد پارامترهای نامعلوم موجود در این ماتریس بیش‌تر از تعداد عناصر قطری و بالا مثلثی آن شود)، آنگاه به دلیل تلفیقی بودن این ماتریس از دو ماتریس دیگر، ممکن است مشکل شناساناپذیری حاصل شود. در صورت برازش چنین مدل‌هایی، نرم‌افزارها، خروجی‌های قابل‌اطمینانی برای برآورد پارامترها نداشته و معمولاً اشاره‌ای به این مشکل نخواهند کرد.

لازم به ذکر است، مطالعات انجام‌شده توسط محققان علوم آمار کاربردی به بررسی شناساپذیری در مدل‌های آمیخته‌ی خطی خلاصه‌شده است درصورتی‌که در این مقاله به بررسی شناساپذیری در مدل‌های آمیخته‌ی خطی تعمیم‌یافته با اثرهای تصادفی پرداخته می‌شود که تاکنون موردتوجه محققان دیگر قرارگرفته نشده است. قضایای جدید ارائه‌شده در این مقاله، اولاً شناساپذیری مدل توأم خطی تعمیم‌یافته با اثرهای تصادفی را موکول به شناساپذیری مدل تکی کرده‌اند و ثانیاً شروط لازم برای شناساپذیری پارامترها در این مدل‌ها را ارائه می‌دهند. سابقاً شروط لازم شناساپذیری مدل‌های آمیخته خطی، توسط [۸] بررسی‌شده است اما موارد استفاده از آن‌ها، محدود به همین مدل‌هاست. بنابراین در این مقاله سعی بر یافتن ارتباطی بین

قضایای ارائه‌شده در [۸] با قضایای معادل آن‌ها در مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی شده است.

همچنین این مقاله به این صورت سازمان‌دهی گردیده است که در بخش دوم مروری کوتاه بر مفهوم شناساپذیری شده است و تعریف خاصی از شناساپذیری، برگرفته از تعریف اصلی آن، در مدل‌های خطی تعمیم‌یافته بیان‌شده است. در ادامه به چگونگی بررسی شناساپذیری مدل‌های خطی تعمیم‌یافته با استفاده از این تعریف، پرداخته‌شده است. در بخش سوم مروری بر مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی خواهد شد و با بیان قضایا و برهان آن‌ها در مورد شناساپذیری در مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی، راهکارهای جدیدی برای بررسی شناساپذیری در این مدل‌ها ارائه می‌شود. درنهایت در بخش چهارم، مشکل‌های حاصل از برازش مدل شناساناپذیر با استفاده از مطالعه‌های شبیه‌سازی، نشان داده می‌شود. مشکل‌های نرم‌افزاری ایجادشده در این مدل‌ها، علاوه بر مشکلاتی که در [۸] برای مدل‌های آمیخته‌ی خطی بیان‌شده است، موارد قابل تأمل دیگری مانند تغییر برآورد پارامترها با تغییر مقدار اولیه و یا حتی تعویض نرم‌افزار مورداستفاده را دارند. در این بخش به بررسی برخی از این مشکلات خواهیم پرداخت.

۲- شناساپذیری در مدل‌های خطی تعمیم‌یافته

۲-۱- مفهوم شناساپذیری

تعریف ۱. تعریف شناساپذیری بر اساس خانواده‌ی پارامتری

خانواده‌ی پارامتری $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ شناساپذیر است اگر و تنها اگر $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$ دلالت بر $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ بنماید [۱]. به عبارت بهتر، این خانواده‌ی پارامتری و به‌اصطلاح، $\boldsymbol{\theta}$ شناساپذیر است اگر نگاشت $\boldsymbol{\theta} \to P_{\boldsymbol{\theta}}$ نگاشتی یک‌به‌یک در نظر گرفته شود

تعریف ۲. تعریف شناساپذیری بر اساس خانواده‌ای از توزیع‌های آماری

مدل آماری (۱)، تعریف‌شده به‌واسطه‌ی خانواده‌ای از توزیع‌ها بر پایه‌ی بردار تصادفی $\mathbf{Y}$ که تحت بردار پارامتری $\boldsymbol{\theta}$ پارامتری شده است را در نظر بگیرید:

$$P = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \qquad (۱)$$

که در آن $P_{\boldsymbol{\theta}}$ توزیع بردار تصادفی $\mathbf{Y}$ است که به‌واسطه‌ی بردار پارامتری $\boldsymbol{\theta}$ پارامتری شده است. همچنین $\boldsymbol{\Theta}$ اشاره به فضای پارامتر دارد. این مدل روی فضای پارامتری $\boldsymbol{\Theta}$، شناساپذیر است اگر $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$، آن‌گاه $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ [۳]. تعریف شناساپذیری (۲)، تعریفی بسیار عمومی از

شناساپذیری به شمار می‌آید و در واقع، مفهوم شناساپذیری را بیان می‌کند. حال برای بررسی شناساپذیری پارامترهای یک مدل، اگر بتوان توابعی از پارامترهای مدل را چنان یافت که تابع چگالی احتمال مدل، به‌واسطه‌ی این توابع کاملاً تعیین شود، آن‌گاه می‌توان تعریف جدیدی برای شناساپذیری بر پایه‌ی تعریف (۳) بیان کرد. این روش برای ساده‌سازی بررسی شناساپذیری پارامترهای مدل، استفاده می‌شود.

۲-۲- چگونگی بررسی شناساپذیری مدل‌های خطی تعمیم‌یافته

مفهوم شناساپذیری با اندکی تغییرات در مدل‌های خطی تعمیم‌یافته نیز کاربرد دارد [۱۰]. در مدل‌های خطی تعمیم‌یافته نیز توزیع بردار پاسخ $\mathbf{Y}$ به‌واسطه‌ی $E(\mathbf{Y})$ و یا $E(\mathbf{Y})$ و پارامترهای دیگری مثل ϕ که به $E(\mathbf{Y})$ مرتبط نیستند، کاملاً تعیین می‌شود. هر مدل خطی تعمیم‌یافته به‌صورت زیر در نظر گرفته می‌شود:

$$E(\mathbf{Y}) = \mathbf{h}(\mathbf{X\beta}) \qquad (۲)$$

که در آن $\mathbf{h}(.)$ هر تابع یکنوا و مشتق‌پذیر است. معکوس این تابع، تابع ربط نامیده می‌شود. در مدل‌سازی‌ها، معمولاً ترجیح می‌دهند که از تابع‌های توزیع تجمعی به‌عنوان تابع $\mathbf{h}(.)$ استفاده کنند اما هر تابع دیگری که در این شرایط صدق کند، نیز می‌تواند به‌عنوان تابع $\mathbf{h}(.)$ مورداستفاده قرار گیرد [۱۱]. همچنین $\mathbf{Y} = (Y_1,...,Y_n)'$ بردار پاسخ، $\mathbf{X}_{n\times m}$ ماتریس طرح اثرهای ثابت و $\boldsymbol{\beta}$ بردار پارامتری اثرهای ثابت است. بنابراین می‌توان تعریف (۲) را به‌صورت زیر، برای مدل‌های خطی تعمیم‌یافته، تعمیم داد:

تعریف ۳. تعریف شناساپذیری در مدل‌های خطی تعمیم‌یافته

بردار پارامتری $\boldsymbol{\beta}$ شناساپذیر است اگر برای هر بردار $\boldsymbol{\beta}_1$ و $\boldsymbol{\beta}_2$، $\mathbf{h}(\mathbf{X\beta}_1) = \mathbf{h}(\mathbf{X\beta}_2)$ دلالت بر $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ کند. لازم به ذکر است اگر بردار $\boldsymbol{\beta}$ شناساپذیر باشد، آنگاه پارامتریدن $\mathbf{h}(\mathbf{X\beta})$ (مدل) شناساپذیر است. به‌علاوه $\mathbf{g}(\boldsymbol{\beta})$ ($\mathbf{g}(.)$ تابعی از R^p به R است که در آن p بعد بردار پارامتری $\boldsymbol{\beta}$ خواهد بود.) شناساپذیر است اگر $\mathbf{h}(\mathbf{X\beta}_1) = \mathbf{h}(\mathbf{X\beta}_2)$ دلالت بر $\mathbf{g}(\boldsymbol{\beta}_1) = \mathbf{g}(\boldsymbol{\beta}_2)$ کند. اگر نوعی پارامتریدن شناساپذیر نباشد امّا تابعی از پارامترها چنان وجود داشته باشد که شناساپذیر است، آنگاه گفته می‌شود این نوع پارامتریدن، شناساپذیر جزئی است.

نتیجه ۱. با توجه به این‌که در ویژگی‌های تابع برداری $\mathbf{h}(.)$، معکوس‌پذیر بودن آن نیز لحاظ شده است، از تعریف (۳) می‌توان نتیجه گرفت که بردار پارامتری $\boldsymbol{\beta}$ در مدل‌های خطی تعمیم‌یافته‌ی (۲) شناساپذیر است اگر ماتریس طرح $\mathbf{X}$، پر رتبه باشد.

قضیه ۱. تابع $\mathbf{g}(\boldsymbol{\beta})$ در مدل خطی تعمیم‌یافته‌ی (۲) شناساپذیر است اگر و تنها اگر $\mathbf{g}(\boldsymbol{\beta})$ تابعی از $\mathbf{h}(\mathbf{X}\boldsymbol{\beta})$ است [۱۰].

نتیجه ۲. تابع $\mathbf{g}(\boldsymbol{\beta})$ در مدل خطی تعمیم‌یافته‌ی (۲) شناساپذیر است اگر و تنها اگر $\mathbf{g}(\boldsymbol{\beta})$ تابعی از $\mathbf{X}\boldsymbol{\beta}$ است.

۲-۳- چند مثال در شناساپذیری مدل‌های خطی تعمیم‌یافته

در این قسمت به بررسی شناساپذیری در مدل‌هایی با پاسخ‌های گسسته‌ی دودویی همچون رگرسیون لوژستیک و پروبیت در دو حالت با وجود مفهوم متغیر پنهان و بدون وجود مفهوم متغیر پنهان پرداخته می‌شود و در صورت شناساناپذیری مدل راه‌کارهایی برای رفع مشکل بیان خواهد شد.

مدل (۲) را مدنظر قرار دهید. هرگاه متغیر پاسخ Y_i در این مدل، متغیر گسسته‌ی دودویی باشد، آنگاه مدل حاصل، رگرسیون دودویی خواهد بود. بنابراین در این مدل‌ها، $E[Y_i \mid \mathbf{x}_i] = h(\mathbf{x}_i'\boldsymbol{\beta})$ که در آن $\mathbf{x}_i$ بردار متغیرهای تبیینی مربوط به مشاهده iام و در واقع سطر iام از ماتریس طرح $\mathbf{X}$ است. به‌طور معمول، مؤلفه‌ی اول بردار $\mathbf{x}_i$، ۱ و مؤلفه‌ی اول بردار پارامتری $\boldsymbol{\beta}$، پارامتر عرض از مبدأ در نظر گرفته می‌شود. به‌علاوه $h(.)$ هر تابع پیوسته اکیداً صعودی و دو بار مشتق‌پذیر است. برای مطالعه بیشتر در مورد شرایط این تابع به [۳] مراجعه کنید.

۲-۳-۱- شناساپذیری در مدل رگرسیون لوژستیک

اگر در مدل $E[Y_i \mid \mathbf{x}_i] = h(\mathbf{x}_i'\boldsymbol{\beta})$، تابع $h(.)$، تابع توزیع تجمعی لوژستیک با پارامتر مکان صفر و پارامتر مقیاس یک برگزیده شود، آنگاه مدل حاصل، مدل رگرسیون لوژستیک نامیده می‌شود. از سوی دیگر، برای استفاده از تعریف (۳) و قضیه‌ی (۱) مناسب است اشاره‌ای به تابع برداری $h(.)$ بنماییم. توجه کنید که در این مدل:

$$\pi(\mathbf{x}_i) = E[Y_i \mid \mathbf{x}_i] = \frac{\exp\{\mathbf{x}_i'\boldsymbol{\beta}\}}{1+\exp\{\mathbf{x}_i'\boldsymbol{\beta}\}}, \qquad i = 1,\ldots,n.$$

بنابراین $h(s) = \frac{e^s}{1+e^s}$ و تابع برداری $\mathbf{h}(\mathbf{s})$ تابعی است که مؤلفه‌ی iام آن به‌صورت $h(s_i)$ است که در آن s_i مؤلفه‌ی iام بردار $\mathbf{s}$ خواهد بود. درنتیجه:

$$E[\mathbf{Y} \mid \mathbf{X}] = \mathbf{h}(\mathbf{X}\boldsymbol{\beta})$$

که در آن $\mathbf{Y}=(Y_1,...,Y_n)'$ و $\mathbf{h}(.)$ تابع برداری از R^n به R^n تعریف‌شده است. حال با استفاده از تعریف (۲) می‌توان شناساپذیری در مدل رگرسیون لوژستیک را به‌صورت زیر موردبررسی قرار داد:

ابتدا لگاریتم تابع درستنمایی بردار پاسخ $\mathbf{Y}$ به‌صورت زیر به دست می‌آید:

$$\ln f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\beta}) = \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^{n} \ln(1+e^{\mathbf{x}_i'\boldsymbol{\beta}}).$$

حال به‌راحتی می‌توان ثابت کرد که پر رتبه بودن ماتریس طرح $\mathbf{X}$ شرط لازم و کافی برای شناساپذیری بردار پارامتری $\boldsymbol{\beta}$ است. فرض کنید ماتریس طرح $\mathbf{X}$ پر رتبه نیست. در این صورت بردار $\mathbf{a}\neq 0$ چنان وجود دارد که

$$\mathbf{X}\mathbf{a} = \begin{bmatrix} \mathbf{x}_1'\mathbf{a} \\ \vdots \\ \mathbf{x}_n'\mathbf{a} \end{bmatrix} = \mathbf{0}$$

بنابراین به ازای هر بردار پارامتری $\boldsymbol{\beta}$، بردار پارامتری $\boldsymbol{\beta}^* = \boldsymbol{\beta}+\mathbf{a}$ چنان وجود دارد که به ازای هر $\mathbf{y}$ موجود در تکیه‌گاه، $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\beta}) = f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\beta}^*)$ درصورتی‌که $\boldsymbol{\beta}\neq\boldsymbol{\beta}^*$.

از طرف دیگر اگر بردار پارامتری $\boldsymbol{\beta}$ شناساپذیر در نظر گرفته نشود، آن‌گاه به ازای هر بردار پارامتری $\boldsymbol{\beta}_1$، بردار پارامتری $\boldsymbol{\beta}_2$ چنان وجود دارد که اولاً مخالف با $\boldsymbol{\beta}_1$ است و دوماً $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\beta}_1) = f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\beta}_2)$ و این یعنی $\mathbf{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)=\mathbf{0}$؛ بنابراین ماتریس طرح $\mathbf{X}$ پر رتبه نیست. همچنین با استفاده از تعریف (۳)، برای هر دو بردار $\boldsymbol{\beta}_1$ و $\boldsymbol{\beta}_2$ که $\mathbf{h}(\mathbf{X}\boldsymbol{\beta}_1) = \mathbf{h}(\mathbf{X}\boldsymbol{\beta}_2)$، نتیجه می‌شود که $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$ زیرا $\mathbf{x}_i'\boldsymbol{\beta}_1 = \mathbf{x}_i'\boldsymbol{\beta}_2 = \ln\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)}$؛ بنابراین بردار پارامتری $\boldsymbol{\beta}$ در مدل لوژستیک شناساپذیر است اگر و تنها اگر ماتریس طرح $\mathbf{X}$ پر رتبه در نظر گرفته شود.

۲-۳-۲- شناساپذیری در مدل رگرسیون پروبیت بر پایه متغیر پنهان

اگر در مدل $E[Y_i \mid \mathbf{x}_i] = h(\mathbf{x}_i'\boldsymbol{\beta})$ به‌جای تابع $h(.)$ از Φ که تابع توزیع تجمعی نرمال استاندارد است، استفاده شود، مدل پروبیت حاصل می‌شود. این تابع زمانی در تحلیل پاسخ‌های دودویی به کار می‌رود که قضاوت تحلیل‌گر بر اساس یک متغیر پنهان باشد.

بررسی شناساپذیری مدل رگرسیون پروبیت بر اساس تعریف (۲) به‌صورت زیر انجام می‌شود:

فرض کنید متغیر تصادفی گسسته دودویی Y_i بر پایه‌ی متغیر پنهان Y_i^*، به این صورت تعریف‌شده باشد:

$$Y_i = \begin{cases} ۱ & Y_i^* > ۰ \\ ۰ & Y_i^* \le ۰ \end{cases} \Rightarrow Y_i \sim \text{Ber}(P(Y_i^* > ۰)).$$

اگر برای Y_i^*، مدلی رگرسیونی به‌صورت زیر در نظر گرفته شود:

$$Y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \sim N(۰, \sigma^۲),$$

که در آن $\mathbf{x}_i'$ بردار متغیرهای تبیینی و $\boldsymbol{\beta}$ بردار پارامتری اثرهای ثابت است، آنگاه در این مدل بردار پارامتر یعنی $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma)'$ شناساپذیر نیست. برای بررسی این موضوع به [۱۲] مراجعه کنید. بررسی شناساپذیری مدل رگرسیون پروبیت بر اساس تعریف (۳) به‌صورت زیر انجام می‌شود: مدل رگرسیون پروبیت به‌صورت زیر بازنویسی می‌شود:

$$E[\mathbf{Y} \mid \mathbf{X}] = \mathbf{h}_{\sigma^۲}(\mathbf{X}\boldsymbol{\beta}),$$

که در آن تابع برداری $\mathbf{h}_{\sigma^۲}(.)$ تابعی است که هر مؤلفه‌ی آن، تابع توزیع تجمعی نرمال با میانگین صفر و واریانس $\sigma^۲$ است. از سوی دیگر تابع $\mathbf{h}(.)$ معرفی‌شده در مدل (۲) باید تابعی معلوم باشد؛ بنابراین مجدداً مدل به‌صورت زیر بازنویسی می‌شود:

$$E[\mathbf{Y} \mid \mathbf{X}] = \mathbf{h}(\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\Phi}(\mathbf{X}\frac{\boldsymbol{\beta}}{\sigma})$$

که در آن تابع برداری $\mathbf{h}(.)$، تابعی است که هر مؤلفه‌ی آن، تابع توزیع تجمعی نرمال استاندارد خواهد بود. حال طبق تعریف (۳) پر رتبه بودن ماتریس طرح $\mathbf{X}$ شرط لازم و کافی برای شناساپذیری بردار پارامتری $\boldsymbol{\beta}^* = \frac{\boldsymbol{\beta}}{\sigma}$ است. توجه کنید که توزیع بردار تصادفی $\mathbf{Y}$ متعلق به خانواده‌ی نمایی بوده و به‌واسطه‌ی دو گشتاور اوّل تعیین می‌شود و تحت همین شرط است که مجاز به استفاده از تعریف (۳) خواهیم بود. برای رفع این مشکل که $\boldsymbol{\beta}^*$ شناساپذیر است و نه $\boldsymbol{\beta}$ و σ (بنابراین نرم‌افزارها خروجی معقول را برای برآورد $\boldsymbol{\beta}^*$ نمایش خواهند داد) می‌توان قیدهای روی مدل اعمال کرد. برای مثال، قید $\sigma^۲ = ۱$ از متداول‌ترین این قیود است. لازم به ذکر است که اگر این رگرسیون بدون استفاده از متغیرهای پنهان برازش داده شود و در واقع از ابتدای مدل‌سازی، تابع $h(.)$ تابع توزیع نرمال استاندارد در نظر گرفته شود، آنگاه پر رتبه بودن ماتریس طرح $\mathbf{X}$ شرط لازم و کافی برای شناساپذیری مدل و به‌تبع بردار پارامتر $\boldsymbol{\beta}$ خواهد

بود. حال، اگر متغیر تصادفی گسسته دودویی Y_i بر پایه‌ی متغیر Y_i^* به‌صورت زیر تعریف شده باشد:

$$Y_i = \begin{cases} ۱ & Y_i^* > \theta \\ ۰ & Y_i^* \le \theta \end{cases}$$

که در آن θ پارامتری نامعلوم (پارامتر آستانه‌ای) است؛ بنابراین برای نمونه‌ای تصادفی به حجم n، مدل به‌صورت زیر است:

$$E[\mathbf{Y} \mid \mathbf{X}] = \mathbf{h}_{\theta,\sigma^۲}(\mathbf{X\beta}), \qquad (۴)$$

که در آن $\mathbf{h}_{\theta,\sigma^۲}(.)$ تابع برداری با مؤلفه‌های تابع توزیع نرمال با میانگین θ و واریانس $\sigma^۲$ است. مجدداً مدل فوق را برای رسیدن به تابع $\mathbf{h}(.)$ مطلوب و مستقل از پارامتر، بازنویسی می‌کنیم:

$$E[\mathbf{Y} \mid \mathbf{X}] = \mathbf{h}(\mathbf{X}(\frac{\boldsymbol{\beta}}{\sigma}) - \frac{\theta}{\sigma}). \qquad (۵)$$

برای استفاده از تعریف (۳)، مدل (۵) مطلوب نیست؛ بنابراین تعریف می‌کنیم $\mathbf{X}^* = [\mathbf{X} \ \ -\mathbf{1}_n]$ که در آن $\mathbf{1}_n$ بردار یک‌ها با طول n و $\boldsymbol{\beta}^* = \left(\frac{\boldsymbol{\beta}'}{\sigma}, \frac{\theta}{\sigma}\right)'$. مدل (۵) به‌صورت $E[\mathbf{Y} \mid \mathbf{X}] = \mathbf{h}(\mathbf{X}^*\boldsymbol{\beta}^*)$ بازنویسی می‌شود. بنابراین مانند قبل اولاً باید، قیدی برای مشکل وجود σ در مدل اعمال کرد و ثانیاً برای شناساپذیری $\boldsymbol{\beta}^*$ حتی بعد از اعمال قید روی σ، شرط پر رتبه بودن برای ماتریس طرح $\mathbf{X}^*$ لازم خواهد بود و اگر بردار پارامتری $\boldsymbol{\beta}$ شامل عرض از مبدأ $\beta_۰$ در نظر گرفته شود، آنگاه ستون اوّل ماتریس $\mathbf{X}$، $\mathbf{1}_n$ و بنابراین ماتریس طرح $\mathbf{X}^*$ پر رتبه نخواهد شد. راهی دیگر برای برطرف کردن این مشکل، برازش مدل بدون عرض از مبدأ $\beta_۰$ است؛ بنابراین در بردار $\boldsymbol{\beta}$، عرض از مبدأ $\beta_۰$ حذف خواهد شد.

۳- شناساپذیری در مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی

در این بخش ابتدا مروری بر مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی خواهیم کرد و در ادامه با ارائه‌ی چند قضیه‌ی جدید در مورد شناساپذیری مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی، راهکارهای آسان‌تری را برای بررسی شناساپذیری در این مدل‌ها بیان می‌کنیم.

۳-۱- مروری بر مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی

مدل آمیخته‌ی خطی زیر را در نظر بگیرید:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1,...,N \qquad (۶)$$

که در آن، $\mathbf{Y}_i$ بردار پاسخ $(n_i \times 1)$ مربوط به خوشه (فرد) iام، $\mathbf{X}_i$ ماتریس طرح اثرهای ثابت در خوشه iام، $\boldsymbol{\beta}$ بردار ضرایب اثرهای ثابت، $\mathbf{u}_i$ بردار اثرهای تصادفی و $\mathbf{Z}_i$ ماتریس طرح مربوط به اثرهای تصادفی در خوشه‌ی iام هستند. $\boldsymbol{\varepsilon}_i$ هم بردار خطای تصادفی خوشه i ام است. در این مدل $\mathbf{u}_i \sim (\mathbf{o},\Sigma_{\mathbf{u}})$ و $\boldsymbol{\varepsilon}_i \sim (\mathbf{o},\Sigma_{\varepsilon_i})$ هستند و $\mathbf{u}_1$ تا $\mathbf{u}_N$ و $\boldsymbol{\varepsilon}_1$ تا $\boldsymbol{\varepsilon}_N$ توأماً مستقل در نظر گرفته می‌شوند. به مدل (۶) به ازای همه iها مدل توأم و به ازای یک i ثابت، مدل تکی می‌گویند. بررسی شناساپذیری چنین مدلی تحت ساختارهای رایج ماتریس کواریانس توسط [۸] صورت گرفته است. در واقع همان‌گونه که بردار پاسخ پیوسته‌ی $\mathbf{Y}_i$ در مدل (۶)، می‌تواند توسط متغیرهای تبیینی و اثرهای تصادفی مدل شود و شناساپذیری پارامترهای آن موردبررسی قرار گیرد، $\mathbf{Y}_i = [Y_{i1},...,Y_{in_i}]'$ بردار پاسخ گسسته مربوط به فرد یا خوشه‌ی iام نیز می‌تواند بر اساس متغیرهای تبیینی و اثرهای تصادفی مدل شده و شناساپذیری پارامترهای آن بررسی شود. در این راستا فرض کنید $\mathbf{X}_i$، $\boldsymbol{\beta}$، $\mathbf{Z}_i$ و $\mathbf{u}_i$ بردارها و ماتریس‌های تعریف‌شده در رابطه‌ی (۶) هستند؛ بنابراین، هر مدل خطی تعمیم‌یافته‌ی اثرهای تصادفی با تابع ربط برداری $\mathbf{g}(.)$ به‌صورت زیر خواهد بود:

$$\mathbf{g}(E(\mathbf{Y}_i \mid \mathbf{u}_i, \mathbf{X}_i)) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i, \qquad i = 1,...,N. \qquad (۷)$$

مشخصه‌های عمومی مدل‌های خطی تعمیم‌یافته‌ی اثرهای تصادفی به‌صورت زیر است:

۱. متغیرهای پاسخ $Y_{i1},...,Y_{in_i}$ به‌شرط $\mathbf{u}_i$ دوبه‌دو مستقل هستند و توزیع آن‌ها عضوی از خانواده‌ی پراکندگی نمایی به‌صورت زیر است:

$$f_{Y_{ij}}(y_{ij} \mid \mathbf{u}_i) = \exp\left\{\frac{(y_{ij}\theta_{ij}) - b(\theta_{ij})}{a(\phi)} + c(y_{ij},\phi)\right\}$$

که در آن $a(.)$، $b(.)$ و $c(.)$ تابع‌های حقیقی مقدار، ϕ و θ_{ij} به ترتیب پارامتر پراکنش و پارامتر مکان هستند [۱۱].

۲. گشتاورهای شرطی $\mu_{ij} = E\left[Y_{ij}\middle|\mathbf{u}_i\right] = b'(\theta_{ij})$ و $v_{ij} = Var\left[Y_{ij}\middle|\mathbf{u}_i\right] = \frac{b''(\theta_{ij})}{\phi}$ که در آن $b'(.)$ و $b''(.)$ مشتق‌های مرتبه‌ی اوّل و دوم تابع $b(.)$ نسبت به θ_{ij} هستند در رابطه‌های زیر صدق می‌کنند:

$$\theta_{ij} = g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i, \quad v_{ij} = v(\mu_{ij})\phi,$$

که در آن $g(.)$ و $v(.)$ به ترتیب تابع ربط (مؤلفه‌های تابع ربط برداری $\mathbf{g}(.)$) و تابع واریانس هستند (برای مطالعه بیشتر در این زمینه به [۱۱ و ۱۳] مراجعه کنید).

۳. $\mathbf{u}_i$ دارای توزیع چندمتغیره مشخص است که معمولاً نرمال با بردار میانگین $\mathbf{o}$ و ماتریس کوواریانس $\Sigma_{\mathbf{u}}$، پارامتری شده به‌واسطه‌ی بردار پارامتری $\boldsymbol{\theta}_u$ برای آن رایج است.

از سوی دیگر برای بررسی شناساپذیری، به تابع چگالی توأم برای آزمودنی iام نیاز داریم؛ بنابراین طبق مشخصه‌ی ۱، این تابع به‌صورت زیر محاسبه می‌شود:

$$f_{\mathbf{Y}_i}(y_{i1},\ldots,y_{in_i};\boldsymbol{\beta},\phi,\boldsymbol{\theta}_u) = \int \prod_{j=1}^{n_i} f_{\mathbf{Y}_i}(y_{ij} \mid \mathbf{u}_i;\boldsymbol{\beta},\phi) f_{\mathbf{U}_i}(\mathbf{u}_i;\boldsymbol{\theta}_u)\, d\mathbf{u}_i \qquad (۸)$$

که در آن وابستگی به $\mathbf{x}_{ij}$ برای سادگی در نوشتار حذف‌شده است. در ضمن انتگرال فوق، یک انتگرال k گانه است.

۳-۲- چند نتیجه‌ی جدید درباره‌ی شناساپذیری مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی

در این زیر بخش، چند قضیه‌ی جدید در شناساپذیری مدل‌های خطی تعمیم‌یافته ارائه خواهد شد. با استفاده از این قضایا بررسی شناساپذیری در این مدل‌ها، ساده‌تر خواهد شد.

قضیه ۲. مدل توأم (۷) شناساپذیر است اگر و تنها اگر حداقل یکی از مدل‌های تکی آن شناساپذیر در نظر گرفته شود.

برهان. به پیوست مراجعه شود.

قضیه‌ی فوق بررسی شناساپذیری مدل توأم را به بررسی شناساپذیری مدل‌های تکی موکول می‌کند؛ بنابراین برای سهولت در نگارش، چون ما دائماً با یک i ثابت از مدل توأم، کار خواهیم کرد. زیروند i را از مدل (۷) برمی‌داریم و با مدل تکی کار می‌کنیم.

در ادامه قضیه‌ای مطرح خواهد شد که بیان آن مستلزم ارائه‌ی تعریفی برای شناساپذیری ماتریس کواریانس $\Sigma_{\mathbf{u}}$ بر پایه‌ی تعریف (۲) است. این تعریف توسط [۸] ارائه شده است.

قضیه ۳. پر رتبه بودن ماتریس طرح $\mathbf{X}$ و شناساپذیری ماتریس کواریانس $\Sigma_{\mathbf{u}}$ در مدل تکی (۷)، شرطی لازم برای شناساپذیری بردار پارامتری $\boldsymbol{\theta}$ است.

برهان. به پیوست مراجعه شود.

البته توجه به این نکته ضروری است که پارامتر پراکنش ϕ، یک در نظر گرفته شده است و توزیع به شکل توزیع طبیعی خانواده‌ی نمایی تبدیل شده است.

ملاحظه ۱. یک شرط لازم برای شناساپذیری پارامترهای مدل، یکتایی دو گشتاور اوّل به‌عنوان تابعی از پارامترهای مدل است [۳].

$$E_{\theta_1}(\mathbf{Y}) = E_{\theta_2}(\mathbf{Y}) \ , \ \ Co\,V_{\theta_1}(\mathbf{Y}) = CoV_{\theta_2}(\mathbf{Y}) \ \Rightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2.$$

توجه به این نکته ضروری است که این شرط لازم است و نه کافی. زمانی که توزیع متغیر پاسخ، به‌واسطه‌ی دو گشتاور اول کاملاً تعیین شود، این شرط لازم و کافی خواهد بود. به همین دلیل در مدل‌های رگرسیونی با فرض نرمالیتی، شرط ذکرشده، برای شناساپذیری پارامترهای مدل، لازم و کافی خواهد بود.

قضیه ۴. مدل (۷) را مدنظر قرار دهید. اگر به ازای هر بردار پارامتری $\boldsymbol{\theta}_1$ عضو مجموعه‌ی $\boldsymbol{\Theta}$، بردار پارامتری $\boldsymbol{\theta}_2$ در مجموعه‌ی $\boldsymbol{\Theta}$ و مخالف با $\boldsymbol{\theta}_1$ چنان وجود داشته باشد که تساوی

$$\left|\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\right| \mathbf{h}(\mathbf{X}\boldsymbol{\beta}_1 + \mathbf{Z}\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\mathbf{w}) = \left|\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\right| \mathbf{h}(\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{Z}\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\mathbf{w}) \qquad (۹)$$

به ازای هر $\mathbf{w}$ در R^n برقرار باشد، آن‌گاه مدل (۷) شناساناپذیر است.

برهان. به پیوست مراجعه شود.

نکته ۱. گاهی اوقات برخلاف مدل (۷)، بیش از یک مؤلفه‌ی تصادفی $(\mathbf{u})$ در مدل وجود دارد. نتایج بیان‌شده در این مقاله برای برخی از این گونه مدل‌ها نیز کاربرد دارند. برای مثال، مدل زیر با دو مؤلفه‌ی تصادفی را مدنظر قرار دهید:

$$\mathbf{g}(E(\mathbf{Y}_i \mid \mathbf{u}_{i1}, \mathbf{u}_{i2}, \mathbf{X}_i)) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_{i1} + \mathbf{Z}_i\mathbf{u}_{i2}, \qquad i = 1,...,N. \qquad (۱۰)$$

که در آن $\mathbf{u}_1 \sim (\mathbf{o}, \Sigma_{u1})$، $\mathbf{u}_2 \sim (\mathbf{o}, \Sigma_{u2})$. اگر ستون‌های ماتریس $\mathbf{Z}_1$ و ماتریس $\mathbf{Z}_2$ در مدل (۱۰) مستقل خطی باشند، آنگاه به‌راحتی مدل (۱۰) را به‌صورت زیر به مدل (۷) تبدیل کرده و از مطالب این مقاله استفاده می‌کنیم:

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2), \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \Sigma_u = \begin{bmatrix} \Sigma_{u1} & \Sigma_{u1,u2} \\ \Sigma_{u2,u1} & \Sigma_{u2} \end{bmatrix}.$$

اما در حالتی که ستون‌های ماتریس $\mathbf{Z}_1$ و ماتریس $\mathbf{Z}_2$ مستقل خطی نباشند، تنها در حالتی که $\mathbf{u}_1$ و $\mathbf{u}_2$ مستقل هستند، می‌توان از این نتایج استفاده کرد. در این حالت اگر فرض کنیم

$\boldsymbol{\varepsilon} = \mathbf{Z}_۴\mathbf{u}_۲$، آنگاه می‌توان از تمام قضایای [۸] که تحت ساختارهای رایج ماتریس کواریانس بیان‌شده است، برای بررسی شناساپذیری ماتریس کواریانس اثرهای تصادفی مدل استفاده کرد.

۴- چند مطالعه‌ی شبیه‌سازی

در این بخش، طی چند مطالعه شبیه‌سازی به بررسی اهمیت شناساپذیری پارامترهای مدل‌های خطی تعمیم‌یافته و مدل‌خطی تعمیم‌یافته با اثرهای تصادفی می‌پردازیم.

۴-۱- مطالعه‌ی شبیه‌سازی روی مدل‌های رگرسیون پروبیت بر پایه‌ی متغیر پنهان

در این بخش ابتدا، طی یک مطالعه‌ی شبیه‌سازی از مدل رگرسیون پروبیت بر پایه‌ی متغیر پنهان و بدون اثرهای تصادفی، به بررسی مشکل‌های حاصل‌شده از شناساناپذیری این مدل پرداخته و سپس قیدهای بیان‌شده برای شناساپذیری مدل را اعمال کرده و مجدداً به بررسی وضعیت مدل شناساشده و مقایسه‌ی آن با مدل قبل می‌پردازیم.

الف) مدل رگرسیون پروبیت بر پایه‌ی متغیر پنهان

مدل (۴)، حالتی که تنها یک متغیر تبیینی در مدل است، در نظر گرفته شود:

$$Y_i = \begin{cases} ۱ & Y_i^* > \theta \\ ۰ & Y_i^* \le \theta \end{cases}, \quad Y_i^* = \beta_۰ + \beta_۱ x_i + \varepsilon_i, \qquad i = ۱,\dots,n,$$

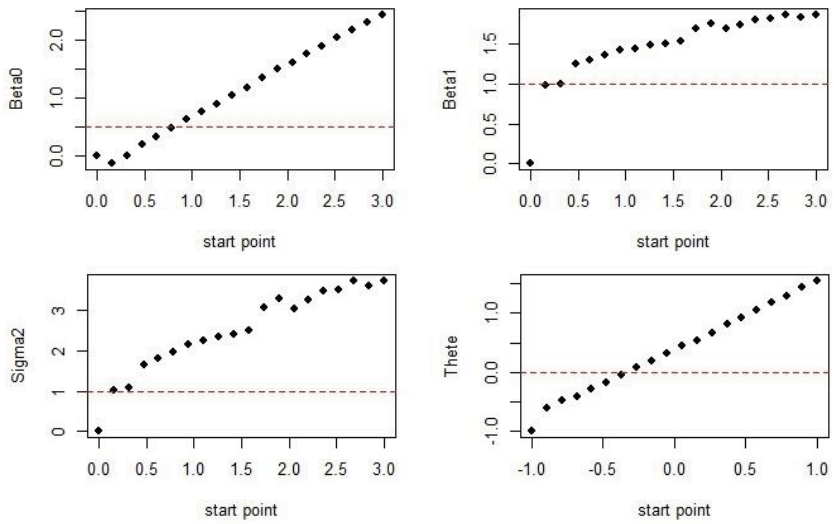
که در آن $\varepsilon_i \sim N(۰,\sigma^۲)$؛ بنابراین $Y_i \sim Ber(p(Y_i^* > \theta))$. بردار پارامتری $\boldsymbol{\beta}' = (\beta_۰,\beta_۱,\sigma^۲,\theta)$ شناساپذیر نبوده و تمام شش مشکل اشاره شده در بخش بحث و نتیجه‌گیری را خواهد داشت. به این منظور طی یک مطالعه‌ی شبیه‌سازی برای $n = ۱۰۰$، ۱۰۰۰ بار از مدل فوق شبیه‌سازی کرده و تعداد دفعاتی که الگوریتم عددی برای محاسبه‌ی پارامترها همگرا نشده را مد نظر قرار داده‌ایم. برای شبیه‌سازی مدل فوق $\boldsymbol{\beta}' = (۰/۵,۱,۱,۰)$ لحاظ شده است. در این مطالعه در هر ۱۰۰۰ دفعه شبیه‌سازی، الگوریتم همگرا شد و بنابراین نشانی از شناساناپذیری به‌واسطه‌ی همگرا نشدن، در این مدل یافت نمی‌شود. همچنین برای بررسی خطای استانداردهای محاسبه شده برای پارامترهای مدل، جدول (۱) تهیه شده است. همان‌گونه که در جدول (۱) ملاحظه می‌شود، تقریباً برای همه‌ی پارامترها به‌جز $\beta_۱$، چندک‌های ۵ درصد به بالای خطای استاندارد، مقدارهای بسیار بزرگ و غیرمعمولی نسبت به مقدار واقعی خود پارامتر دارند. برای پارامتر $\beta_۱$ چندک ٪۵۰ به بالا نامعقول و غیرواقعی است.

در هر مدل خطی با پاسخ پیوسته یا گسسته، برآورد پارامترهای مدل نباید به مقدار اولیه‌ای که به‌عنوان نقطه‌ی ابتدایی به تابع برآورد کننده‌ی پارامترها داده می‌شود، وابسته باشد. در واقع

الگوریتم‌های عددی محاسبه کننده‌ی برآورد پارامترها در مدل‌های بدون مشکل، وابسته به مقادیر اولیّه نیستند. به این منظور، یک‌بار از مدل (۴) شبیه‌سازی کرده و سپس طی ۲۰ دفعه تغییر مقدارهای اولیه، پارامترهای مدل را برآورد کردیم. در شکل (۱) سعی شده است تغییر برآورد پارامترها با تغییر مقدار اولیه در تابع پروبیت ۱، نوشته‌شده با استفاده از بسته‌ی کمینه‌سازی لگاریتم تابع درستنمایی، نمایش داده شود. این دستور، برآورد پارامترهای مدل به روش ماکسیمم درستنمایی را در خروجی ارائه می‌دهد. دامنه‌ی تغییرات پارامترها β_0، β_1، σ^2 و θ به ترتیب [۰،۳]، [۰،۳]، [۰،۳] و [۱،۱-] لحاظ شده است. شکل (۱) تغییر نامعقول برآورد پارامترها متناسب با تغییر مقدارهای اولیه در این مدل شناساناپذیر را نشان می‌دهد.

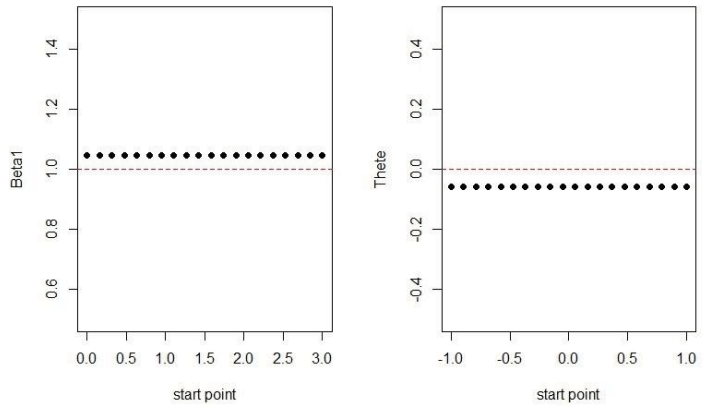
جدول ۱: چندک‌های خطای استاندارد برای پارامترهای مدل شناساناپذیر، به روش ماکسیمم درستنمایی طی هزار بار شبیه‌سازی

پارامتر	۰٪	۱٪	۵٪	۵۰٪	۶۰٪	۸۰٪	۹۰٪	۹۵٪	۹۹٪	۱۰۰٪
$\beta_0 = ۰/۵$	۲/۸۶	۶/۰۷	۱۵/۲۳	۴۲/۲۵	۵۱/۵۵	۸۱/۰۶	۱۱۷/۷۹	۱۸۳/۷۴	۳۹۵/۴۲	۵۷۶/۵۷
$\beta_1 = ۱$	۰/۱۲	۲/۴۳	۴/۹۱	۲۱/۸	۲۷/۷	۴۵/۴۳	۶۶/۰۷	۹۸/۱۵	۲۱۱/۶۶	۳۳۳/۴۴
$\sigma^2 = ۱$	۰/۰۳	۵/۲۷	۱۰/۶۷	۴۶/۱۹	۶۰/۹۶	۱۰۱/۳۹	۱۴۸/۱۶	۲۱۹/۷	۴۰۵/۷۷	۵۹۲/۸۲
$\theta = ۰$	۱/۷۱	۲/۸۸	۷/۳۸	۳۳/۹۸	۴۳/۸۰	۷۸/۶۱	۱۰۸/۸۲	۱۸۱/۱۲۳	۳۹۹/۴۶	۶۲۱/۲۹



شکل ۱: نمودار پراکنش مقدارهای اولیه تابع پروبیت ۱ در مقابل برآورد پارامترهای مدل شناساناپذیر

حال با اعمال قیدهای $\sigma^2 = 1$ و حذف β_0 از بردار پارامتری $\boldsymbol{\beta}$، مدل، شناساپذیر خواهد شد. همین مطالعه‌های شبیه‌سازی را برای مدل شناساشده انجام داده و نتایج در شکل (۲) و جدول (۲) نمایش داده شده‌اند. خطای استاندارد برای پارامترهای مدل شناساپذیر فوق، مقدارهای قابل‌قبول و معقولی اختیار می‌کنند. برآورد خطاهای استاندارد برای پارامتر β_1 در این مطالعه شبیه‌سازی متعلق به بازه‌ی (۰،۰/۳۹] و برای پارامتر θ در ٪۹۹ اوقات [۰،۰/۱۷) است.



شکل ۲: نمودار پراکنش مقدارهای اولیه تابع پروبیت ۱ در مقابل برآورد پارامترهای مدل شناساپذیر

جدول ۲: چندک‌های خطای استاندارد برای پارامترهای مدل شناساپذیر، به روش ماکسیمم درستنمایی طی هزار بار شبیه‌سازی

پارامتر	۰٪	۱٪	۵٪	۵۰٪	۶۰٪	۸۰٪	۹۰٪	۹۵٪	۹۹٪	۱۰۰٪
$\beta_1 = 1$	۰	۰/۱۵	۰/۱۶	۰/۱۹	۰/۲۰	۰/۲۲	۰/۲۴	۰/۲۵	۰/۳	۰/۳۹
$\theta = 0$	۰	۰/۱۲	۰/۱۳	۰/۱۴۶	۰/۱۴۸	۰/۱۵۳	۰/۱۵۸	۰/۱۶	۰/۱۷	۳/۲

با وجود شناساپذیر شدن مدل پروبیت بعد از اعمال قیدهای فوق، یکسان بودن برآورد پارامترها با وجود تغییر در مقدار اولیه ورودی تابع پروبیت ۱، امری طبیعی است که شکل (۲) نیز به تأیید آن می‌پردازد.

ب) مدل رگرسیون پروبیت بر پایه‌ی متغیر پنهان و با حضور اثر تصادفی

به مدل (۷) توجه کنید. مجدداً حالتی را که تنها یک متغیر تبیینی در مدل است، در نظر بگیرید:

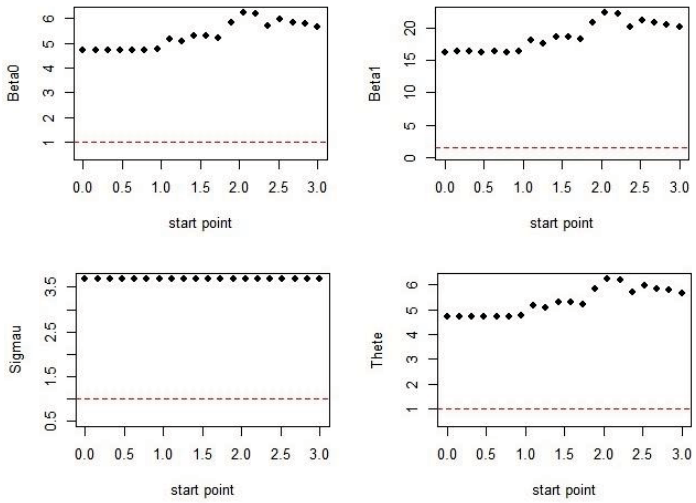
$$Y_{it}=\begin{cases}۱ & Y_{it}^{*}>\theta\\ ۰ & Y_{it}^{*}\le\theta\end{cases},\quad Y_{it}^{*}=\beta_{۰}+\beta_{۱}x_{it}+u_i+\varepsilon_{it},\qquad i=۱,...,n,\quad t=۱,...,T,$$

که در آن $\varepsilon_{it}\sim N(۰,۱)$، $u_i\sim N(۰,\sigma_u^{۲})$. همچنین u_iها و ε_{it}ها توأماً مستقل هستند. طبق قضیه‌ی (۳) بردار پارامتری $\boldsymbol{\beta}'=(\beta_{۰},\beta_{۱},\sigma_u,\theta)$ شناساپذیر نخواهد بود. برای بررسی مشکل‌های حاصل از برازش، مشابه بخش قبل، طی یک مطالعه‌ی شبیه‌سازی برای $n=۱۰۰$، ۱۰۰۰ بار از مدل فوق شبیه‌سازی کرده و تعداد دفعاتی که الگوریتم عددی برای محاسبه‌ی پارامترهای مدل، همگرا نشده است را مدنظر قرار داده‌ایم. در این مطالعه $\boldsymbol{\beta}'=(۱,۱/۵,۱,۱)$ لحاظ شده است. این بار در ۱۰۰ مورد از ۱۰۰۰ بار، الگوریتم همگرا نشد و در بقیه موارد موجب نمایش خروجی شده که به دلیل شناسا نبودن مدل معتبر نیست، ولیکن نرم‌افزار هیچ اشاره‌ای به معتبر نبودن خروجی نمی‌کند. جدول (۳) چندک‌های خطای استانداردهای محاسبه‌شده برای پارامترهای مدل را نشان می‌دهد. چندک‌های انحراف استاندارد پارامترها، بیش از حد معمول بزرگ برآورد شده‌اند. پارامترهای $\beta_{۰}$ و θ که دلیل ایجاد مشکل شناساناپذیری مدل هستند از چندک یک درصد به بعد مقادیر نامعقول اتخاذ کرده‌اند.

شکل (۳) تغییر نامعقول برآورد پارامترها با تغییر مقدارهای اولیه در این مدل شناساناپذیر را نمایش می‌دهد. مدل‌های شناساناپذیر تقریباً علائم مشابه دارند اما این‌که همواره این علامت‌ها در خروجی مشاهده شوند، قطعی نیست و همین مطلب ما را مشتاق به بررسی شناساپذیری مدل قبل از برازش آن می‌کند.

جدول ۳: چندک‌های خطای استاندارد برای پارامترهای مدل شناساناپذیر، به روش ماکسیمم درستنمایی طی ۱۰۰۰ بار شبیه‌سازی

پارامتر	۰٪	۱٪	۵٪	۵۰٪	۶۰٪	۸۰٪	۹۰٪	۹۵٪	۹۹٪	۱۰۰٪
$\beta_{۰}=۱$	۰/۱۷	۲۶/۵۰	۱۳۱/۸۴	۲۹۱۰/۸۶	۵۳۴۸/۸	۵۴۳۱۷	۱۱۶۷۱۴	۲۳۲۸۹۱	۳۲۵۸۳۳	۳۴۹۰۶۹
$\beta_{۱}=۱/۵$	۰/۶۸	۳/۴۵	۱۴/۵۰	۱۴۸۸۹	۲۸۰۰۲	۱۱۲۴۶۲	۳۳۳۹۵۶	۵۴۱۰۹۲	۷۰۶۸۰۱	۷۴۸۲۲۸
$\sigma_u^{۲}=۱$	۰/۶۸	۰/۹۱	۱/۸۱	۲۱۴۱۴	۴۵۳۸۳	۱۲۸۵۱۳	۴۶۱۴۳۶	۵۳۳۰۳۷	۵۹۰۳۱۸	۶۰۴۶۳۸
$\theta=۱$	۰/۱۷	۲۵/۸۱	۱۲۸/۳۶	۲۹۱۰	۵۳۴۸	۵۴۳۱۷	۱۱۶۷۱۴	۲۳۲۸۹۱	۳۲۵۸۳۳	۳۴۹۰۶۹



شکل ۳: نمودار پراکنش مقادیر اولیه تابع پروبیت ۲ در مقابل برآورد پارامترهای مدل شناساناپذیر

۵- بحث و نتیجه‌گیری

گاهی اوقات نرم‌افزارهای آماری در صورت برازش مدل‌های شناساناپذیر، پیام‌هایی مثل همگرا نبودن الگوریتم‌های عددی را گزارش می‌کنند. در ادامه برخی از مشکل‌های حاصل از برازش هر مدل شناساناپذیر ذکر می‌شود. **الف)** نمایش خطای استاندارد خیلی بزرگ و یا صفر بودن مقدار خطای استاندارد برای هر یک از پارامترهای مدل **ب)** ناتوان بودن نرم‌افزار در ساخت فاصله‌های اطمینان **ج)** فاصله‌های اطمینان ساخته‌شده غیرمعقول برای پارامترها. از سوی دیگر در برخی از اوقات، ممکن است هیچ‌کدام از مشکل‌های فوق در خروجی ظاهر نشوند و کاربر بدون اطلاع از این مسئله، از خروجی‌های غیر معتبر استفاده کرده و تفسیرهای نادرستی ارائه دهد. با این اوصاف، بررسی شناساپذیری هر مدل آماری، قبل از برازش آن، امری واجب خواهد بود. از سوی دیگر، اگر مدل خیلی پیچیده نباشد، آنگاه محاسبه‌ی مستقیم عناصر ماتریس کوواریانس پاسخ و استفاده از تعریف اصلی شناساپذیری، برای بررسی آن، ممکن است راهی پیش روی محقق قرار دهد. این کار توسط [۵] انجام‌شده است. بررسی مشکل‌های نرم‌افزاری حاصل از برازش مدل‌های شناساناپذیری توسط [۲ و ۶] صورت گرفته است. تمرکز اصلی [۶] بر روی مشکل‌های حاصل از شناساناپذیری در یکی از نرم افزارهای آماری بوده است، درصورتی‌که [۲] به مقایسه‌ی خروجی‌های حاصل از مدل‌های شناساناپذیر در چهار نرم‌افزار آماری پرداخته است.

در این مقاله، بر پایه‌ی تعریف اصلی شناساپذیری، قضایای دیگری که بررسی شناساپذیری مدل‌های خطی تعمیم‌یافته‌ی اثرهای تصادفی را ساده می‌کنند، بیان شد. موارد زیر برای

تحقیقات بیشتر پیشنهاد می‌شود: **الف)** ارائه شروط لازم و کافی در شناساپذیری پارامترهای ماتریس کوواریانس اثرهای تصادفی و اثرهای ثابت در مدل‌های خطی تعمیم‌یافته با اثرهای تصادفی: برای بررسی شروط کافی شناساپذیری در مدل‌های فوق، توجه به بحث زیر خالی از فایده نخواهد بود. در راستای بررسی کافی بودن شرط پر رتبه بودن ماتریس طرح و شناساپذیری ماتریس $\Sigma_{\mathbf{u}}$ در قضیه‌ی (۳)، باید به‌صورت زیر عمل کرد:

فرض می‌شود بردار پارامتری $\boldsymbol{\theta}$ شناساپذیر نیست و یا به عبارتی به ازای هر بردار پارامتری $\boldsymbol{\theta}_1$ که عضوی از مجموعه‌ی $\boldsymbol{\Theta}$ است، بردار پارامتری $\boldsymbol{\theta}_2$ در مجموعه‌ی $\boldsymbol{\Theta}$ چنان وجود دارد که اولاً مخالف با $\boldsymbol{\theta}_1$ است و دوماً $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_1)=f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_2)$. توجه کنید که $\{\boldsymbol{\theta}_1\neq\boldsymbol{\theta}_2\}$ معادل با $\{\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u1}=\boldsymbol{\theta}_{u2}\}$، $\{\boldsymbol{\beta}_1=\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u1}\neq\boldsymbol{\theta}_{u2}\}$ و یا $\{\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u1}\neq\boldsymbol{\theta}_{u2}\}$ است. در نتیجه:

$$\int\overbrace{\prod_{j=1}^{n}f_{Y_j}(y_j\mid\mathbf{u};\boldsymbol{\beta}_1)f_{\mathbf{u}}(\mathbf{u};\boldsymbol{\theta}_{u1})}^{K(u,\boldsymbol{\beta}_1,\boldsymbol{\theta}_{u1})}d\mathbf{u}=\int\overbrace{\prod_{j=1}^{n}f_{Y_j}(y_j\mid\mathbf{u};\boldsymbol{\beta}_2)f_{\mathbf{u}}(\mathbf{u};\boldsymbol{\theta}_{u2})}^{K(u,\boldsymbol{\beta}_1,\boldsymbol{\theta}_{u2})}d\mathbf{u}. \qquad (11)$$

از رابطه‌ی فوق، نتیجه می‌شود:

$$\int K(\mathbf{u},\boldsymbol{\beta}_1,\boldsymbol{\theta}_{u1})-K(\mathbf{u},\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u2})d\mathbf{u}=0$$

اگر تابع تحت انتگرال، به ازای هر بردار $\mathbf{u}$ ($\mathbf{u}$ عضوی از $\mathbf{R}^k$ است)، همواره مثبت یا همواره منفی در نظر گرفته شود (و یا به ازای هر مجموعه‌ی $\mathbf{B}$ عضو سیگما میدان برل $\mathbf{R}^k$، مقدار انتگرال فوق روی $\mathbf{B}$ برابر صفر باشد)، آنگاه این تابع تقریباً همه‌جا برابر صفر می‌شود و یا به عبارتی $K(\mathbf{u},\boldsymbol{\beta}_1,\boldsymbol{\theta}_{u1})=K(\mathbf{u},\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u2})$. حال سه حالت پیش می‌آید: آ) فرض کنید $\{\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u1}=\boldsymbol{\theta}_{u2}\}$. در نتیجه $f_{\mathbf{Y}}(\mathbf{y}\mid\mathbf{u};\boldsymbol{\beta}_1)=f_{\mathbf{Y}}(\mathbf{y}\mid\mathbf{u};\boldsymbol{\beta}_1)$. مجدداً یادآوری می‌کنیم که توزیع بردار تصادفی $\mathbf{Y}\mid\mathbf{u}$ عضوی از خانواده نمایی بوده و به‌واسطه‌ی دو گشتاور اوّل کاملاً تعیین می‌شود؛ بنابراین از (۷) داریم:

$$E_{\boldsymbol{\beta}_1}(\mathbf{Y}\mid\mathbf{u},\mathbf{X})=\mathbf{h}(\mathbf{X}\boldsymbol{\beta}_1+\mathbf{Z}\mathbf{u})=\mathbf{h}(\mathbf{X}\boldsymbol{\beta}_2+\mathbf{Z}\mathbf{u})=E_{\boldsymbol{\beta}_2}(\mathbf{Y}\mid\mathbf{u},\mathbf{X})$$

حال از معکوس‌پذیر بودن تابع $\mathbf{h}(.)$ نتیجه می‌شود که $\mathbf{X}(\boldsymbol{\beta}_2-\boldsymbol{\beta}_1)=\mathbf{0}$ و عبارت قبل نشان‌دهنده‌ی پر رتبه نبودن ماتریس طرح $\mathbf{X}$ است؛ اما این مسئله که تابع تحت انتگرال مذکور، تقریباً همه‌جا صفر می‌شود یا خیر، باز است. **ب)** فرض کنید که $\{\boldsymbol{\beta}_1=\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u1}\neq\boldsymbol{\theta}_{u2}\}$. در این صورت $f_{\mathbf{U}}(\mathbf{u};\boldsymbol{\theta}_{u1})=f_{\mathbf{U}}(\mathbf{u};\boldsymbol{\theta}_{u2})$ و این معادل با شناساناپذیری ماتریس کوواریانس اثرهای تصادفی است. **پ)** فرض کنید $\{\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2,\boldsymbol{\theta}_{u1}\neq\boldsymbol{\theta}_{u2}\}$. در این حالت از تبدیل $\mathbf{w}=\Sigma_u^{-\frac{1}{2}}\mathbf{u}$ استفاده می‌کنیم. بنابراین $\mathbf{w}$

دارای توزیع نرمال با بردار میانگین صفر و ماتریس کواریانس $\mathbf{I}$ (همانی) است. حال رابطه‌ی (۱۱) به‌صورت زیر بازنویسی می‌شود:

$$\int \left|\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\right| f_{\mathbf{Y}}(\mathbf{y} \mid \Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\mathbf{w}; \boldsymbol{\beta}_1, \boldsymbol{\theta}_{u1}) f_{\mathbf{w}}(\mathbf{w}) d\mathbf{w} = \int \left|\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\right| f_{\mathbf{Y}}(\mathbf{y} \mid \Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\mathbf{w}; \boldsymbol{\beta}_2, \boldsymbol{\theta}_{u2}) f_{\mathbf{w}}(\mathbf{w}) d\mathbf{w}. \quad (۱۲)$$

مجدداً اگر توابع تحت انتگرال باهم برابر شوند، یعنی:

$$\left|\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\right| f_{\mathbf{Y}}(\mathbf{y} \mid \Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\mathbf{w}; \boldsymbol{\beta}_1, \boldsymbol{\theta}_{u1}) = \left|\Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\right| f_{\mathbf{Y}}(\mathbf{y} \mid \Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\mathbf{w}; \boldsymbol{\beta}_2, \boldsymbol{\theta}_{u2}),$$

آنگاه شناساناپذیری هر دو بردار پارامتری $\boldsymbol{\beta}$ و $\boldsymbol{\theta}_u$ حاصل نخواهد شد؛ بنابراین کافی بودن شرایط قضیه‌ی (۳) برای بررسی شناساپذیری، مسئله‌ای باز است.

ب) بررسی شناساپذیری پارامترهای مدل‌های هم‌زمان آمیخته با پاسخ‌های گسسته و پیوسته تحت ساختارهای گوناگون ماتریس کوواریانس

پ) گاهی اوقات برخلاف مدل تکی (۶)، بیش از دو مؤلفه‌ی تصادفی $(\boldsymbol{\varepsilon}, \mathbf{u})$ در مدل وجود دارند. بررسی شناساپذیری در این مدل‌ها، مسئله‌ای باز است.

ت) بررسی شناساپذیری در مدل‌های خطی تعمیم‌یافته‌ی چند متغیره و مدل‌های خطی تعمیم‌یافته‌ی چند متغیره با اثرهای تصادفی

۶- پیوست

برهان قضیه (۲): ($\Leftarrow$) اگر مدل توأم شناساپذیر در نظر گرفته شود، آن‌گاه بدیهی است که حداقل یکی از مدل‌های تکی شناساپذیر است. در غیر این صورت اگر فرض کنیم همه مدل‌های تکی شناساپذیر نیستند، $(\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\theta}_u)')$ آن‌گاه $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ چنان وجود دارند که در هر مدل تکی، $f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\theta}_1) = f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\theta}_2)$؛ بنابراین بدیهی است که تابع چگالی توأم که حاصل‌ضرب توابع چگالی کناری است نیز به ازای این دو بردار پارامتری نابرابر، برابر خواهد شد.

($\Rightarrow$) حال اگر مدل تکی iام شناساپذیر در نظر گرفته شود، آن‌گاه $f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\theta}_1) = f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\theta}_2)$ دلالت بر $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ می‌کند. از آنجایی بردار پارامتر در مدل‌های تکی، وابسته به اندیس نیست، بنابراین مدل تکی jام نیز شناسا خواهد بود. پس مدل توأم شناساپذیر است(باید توجه شود که اگر مدل تکی iام شناساپذیر در نظر گرفته شود، به ازای **"هر"** $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ نتیجه می‌شود $(f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\theta}_1) \neq f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\theta}_2))$.

برهان قضیه (۳): آ) فرض کنید ماتریس $\mathbf{X}$ پر رتبه نیست، در این صورت بردار $\boldsymbol{\beta}^* \neq 0$ چنان وجود دارد که $\mathbf{X}\boldsymbol{\beta}^* = \mathbf{0}$ و یا به ازای هر j از ۱ تا n، $\mathbf{x}_j'\boldsymbol{\beta}^* = 0$، که در آن $\mathbf{x}_j'$، سطر jام ماتریس طرح $\mathbf{X}$ است. حال به تابع چگالی توأم (۸) توجه کنید. ازآنجایی‌که $Y_j \mid \mathbf{u}$ به ازای j از ۱ تا n، عضوی از خانواده‌ی پراکندگی نمایی است؛ بنابراین توزیع آن به‌واسطه‌ی دو گشتاور اوّل مشخص خواهد شد. با توجه به (۷) می‌دانیم $E(Y_j \mid \mathbf{u}, \mathbf{x}_j) = h(\mathbf{x}_j'\boldsymbol{\beta} + \mathbf{z}_j'\mathbf{u})$. پس اگر به ازای هر بردار پارامتری $\boldsymbol{\theta}_1$ قرار دهیم $\boldsymbol{\theta}_2 = ((\boldsymbol{\beta} + \boldsymbol{\beta}^*)', \boldsymbol{\theta}_u')'$ آنگاه $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_1) = f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_2)$.

ب) فرض کنید ماتریس Σ_u شناسا نیست؛ بنابراین به ازای هر بردار پارامتری $\boldsymbol{\theta}_{u1}$ عضو مجموعه‌ی $\boldsymbol{\Theta}$، بردار پارامتری $\boldsymbol{\theta}_{u2}$ در مجموعه‌ی $\boldsymbol{\Theta}_u$ و مخالف با $\boldsymbol{\theta}_{u1}$ چنان وجود دارد که $\Sigma(\boldsymbol{\theta}_{u1}) = \Sigma(\boldsymbol{\theta}_{u2})$. حال چون توزیع احتمال بردار اثرهای تصادفی $\mathbf{u}$ نرمال در نظر گرفته شده است، توزیع آن به‌واسطه‌ی دو گشتاور اوّل مشخص خواهد شد و بنابراین $f_{\mathbf{U}}(\mathbf{u};\boldsymbol{\theta}_{u1}) = f_{\mathbf{U}}(\mathbf{u};\boldsymbol{\theta}_{u2})$؛ بنابراین با توجه به (۸)، اگر به ازای هر بردار پارامتری $\boldsymbol{\theta}_1$ قرار دهیم $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}', \boldsymbol{\theta}_{u2}')'$ آنگاه $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_1) = f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_2)$.

پ) اگر ماتریس $\mathbf{X}$ پر رتبه و ماتریس کواریانس Σ_u شناساپذیر در نظر گرفته نشوند، آنگاه به‌راحتی با تعریف بردار پارامتری $\boldsymbol{\theta}_2 = ((\boldsymbol{\beta} + \boldsymbol{\beta}^*)', \boldsymbol{\theta}_{u2}')'$، شناساناپذیری مدل نتیجه می‌شود.

برهان قضیه (۴): فرض کنید به ازای هر بردار پارامتری $\boldsymbol{\theta}_1$ عضو مجموعه‌ی $\boldsymbol{\Theta}$، بردار پارامتری $\boldsymbol{\theta}_2$ در مجموعه‌ی $\boldsymbol{\Theta}$ و مخالف با $\boldsymbol{\theta}_1$ چنان وجود دارد که رابطه‌ی (۹) برقرار است. حال طرفین رابطه فوق را در $f_{\mathbf{W}}(\mathbf{w})d\mathbf{w}$ ($\mathbf{W}$ دارای توزیع نرمال با بردار میانگین صفر و ماتریس کواریانس همانی است) ضرب کرده و انتگرال می‌گیریم:

$$\int \left|\Sigma^{\frac{1}{2}}(\theta_{u1})\right| h(\mathbf{X}\boldsymbol{\beta}_1 + \mathbf{Z}\Sigma^{\frac{1}{2}}(\theta_{u1})\mathbf{w}) f_{\mathbf{w}}(\mathbf{w})d\mathbf{w} = \int \left|\Sigma^{\frac{1}{2}}(\theta_{u2})\right| h(\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{Z}\Sigma^{\frac{1}{2}}(\theta_{u2})\mathbf{w}) f_{\mathbf{w}}(\mathbf{w})d\mathbf{w}.$$

در عبارات سمت چپ و راست رابطه‌ی فوق، به ترتیب، تغییر متغیرهای $\mathbf{u} = \Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u1})\mathbf{w}$ و $\mathbf{u} = \Sigma^{\frac{1}{2}}(\boldsymbol{\theta}_{u2})\mathbf{w}$ را اعمال می‌کنیم تا به برابری زیر برسیم:

$$\int h(\mathbf{X}\boldsymbol{\beta}_1 + \mathbf{Z}\mathbf{u}) f_{\mathbf{U}}(\mathbf{u};\boldsymbol{\theta}_{u1})d\mathbf{u} = \int h(\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{Z}\mathbf{u}) f_{\mathbf{U}}(\mathbf{u};\boldsymbol{\theta}_{u2})d\mathbf{u}. \qquad (۱۳)$$

با توجه به رابطه‌ی (۷)، $E_{\boldsymbol{\theta}}(\mathbf{Y} \mid \mathbf{u}, \mathbf{X}) = h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$. لذا از برابری (۱۳) نتیجه می‌شود:

$$E_{\boldsymbol{\theta}_1}(\mathbf{Y}) = E(E_{\boldsymbol{\theta}_1}(\mathbf{Y} \mid \mathbf{u}, \mathbf{X})) = E(E_{\boldsymbol{\theta}_2}(\mathbf{Y} \mid \mathbf{u}, \mathbf{X})) = E_{\boldsymbol{\theta}_2}(\mathbf{Y})$$

با استفاده از ملاحظه (۱)، شناساناپذیری مدل (۷) ثابت می‌شود.

مراجع

[1] Shao, J. (2003). *Mathematical Statistics* (2nd ed.),. Springer.

[2] West, B., Welch, K.B. and Galecki, A. T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. NewYork: Chapman & Hall, CRC.

[3] Demidenko, E. (2004). *Mixed Models: Theory and Applications*. John Wiley & Sons.

[4] Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Model for Longitudinal Data*. Springer.

[5] Littell, R. C., Henry, P. R. and Ammerman, C. B. (1998). Statistical analysis of repeated measures data using sas procedures. *Journal of Animal Science*, **76**, 1216-1231.

[6] Pinheiro, J. and Bates, D. (2009). *Mixed-Effects Models in S and S-PLUS*. Springer.

[7] Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics: Simulation and Computation*, **22**(4), 1079-1106.

[8] Wang, W. (2013). Identifiability of linear mixed effects models. *Electronic Journal of Statistics*, **7**, 244–263.

[9] San Martin, E. and Quintana, E. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, **16**, 99–106.

[10] Christensen, R. (2011).1 Plane Answers to Complex Questions: *The Theory of Linear Models* (Fourth ed.). Springer.

[11] Agresti, A. (2002). Categorical Data Analysis, p. 133. John Wiley & Sons.

[۱۲] طهماسبی‌نژاد، ژاله (۱۳۹۰). *بررسی مدل‌های همبسته آمیخته با پاسخ‌های پیوسته و گسسته دودویی*. پایان‌نامه کارشناسی ارشد, دانشگاه علم و صنعت ایران، تهران.

[۱۳] قهرودی, زهرا رضایی (۱۳۸۷). *کاربردهای مدل‌های انتقالی برای تحلیل داده‌های طولی با پاسخ رسته‌ای با و بدون مقادیر گم‌شده*، پایان‌نامه دکترا، دانشگاه شهید بهشتی، تهران.

New Results on Identifiability of Generalized Linear Models with Random Effects

Elham Tabrizi, Ehsan Bahrami Samani and Naseh. Jafari

Department of Statistics, Shahid Beheshti University, Tehran, Iran

Abstract

Identifiablity is a necessary property for the adequacy of a statistical model. When a model is not identifiable, no amount of data cannot determine true parameter. In this article, well-known concept of identifiablity and its properties is reviewed. Moreover, since non-identifiablity problem in linear mixed effects models and generalized linear models with random effects is very common, our main focus is on these models. On the other hand, statistical software, after fitting non-identifiable models, don't usually indicate the problem and show invalid outputs. Consequently, it is useful to have a way to check model identifiability before fitting. In this regard, some new theorems to check identifiability in generalized linear models with random effects are presented. data from non-identifiable models are simulated and problems with model non-identifiablity are listed for showing advantages of the mentioned theorems.

Keywords: Identifiability, Linear Mixed Effects Models, Generalized Linear Random Effects Models.

Mathematics Subject Classification (2010): 62J12, 62J05.