

برآورد کوچک ناحیه‌ای بیز تجربی استوار با توزیع α -پایدار متقارن برای مولفه‌های خطا

شاهو زارعی

گروه آمار، دانشکده علوم پایه، دانشگاه کردستان

تاریخ دریافت: ۱۳۹۹/۰۲/۲۶ تاریخ پذیرش و انتشار: ۱۴۰۰/۰۸/۳۰

چکیده: در مدل سطح ناحیه یا فی-هریوت که پرکاربردترین مدل در برآورد کوچک ناحیه است، معمولاً فرض می‌شود که اثرهای تصادفی خاص هر ناحیه و خطاهای نمونه‌گیری دارای توزیع نرمال هستند. اما مقادیرهای خیلی بزرگ یا خیلی کوچک از مولفه‌های خطا باعث عملکرد ضعیف برآوردهای کوچک ناحیه‌ای بر اساس روش‌های برآورد کوچک ناحیه‌ای موجود می‌شوند. برای رفع این مشکل در این مقاله برای مولفه‌های خطا توزیع α -پایدار متقارن جایگزین توزیع نرمال می‌شود و با رهیافت بیز تجربی پارامترهای کوچک ناحیه‌ها برآورد می‌شوند. با استفاده از شبیه‌سازی و تحلیل داده‌های واقعی به مقایسه مدل جدید ارائه شده با روش بیز تجربی در مدل کلاسیک بر اساس نرمال بودن مولفه‌های خطا و روش‌های استوار بر اساس رهیافت بیز تجربی پرداخته و کارایی مدل پیشنهادی در برآورد دقیق پارامترهای کوچک ناحیه، به خصوص وقتی مولفه‌های خطا هر دو نرمال یا هر دو دارای توزیع دم‌بلند هستند، نشان داده می‌شود. **واژه‌های کلیدی:** برآورد کوچک ناحیه‌ای، مدل سطح ناحیه، بیز تجربی، توزیع α -پایدار.

۱ مقدمه

در سال‌های اخیر درخواست برای تعیین آمارهایی با دقت لازم برای برآورد پارامترهای توصیفی کوچک ناحیه به شدت افزایش یافته است. از این رو برآورد کوچک ناحیه^۱ (SAE) به یک موضوع جذاب و

آدرس الکترونیکی نویسنده مسئول مقاله: شاهو زارعی، sh.zarei@uok.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): 62F35, 60E07, 62D05

¹Small Area Estimation

پرطرفدار برای آمارشناسان تبدیل شده است. در برآورد کوچک ناحیه امکان کسب اطلاعات در سطوح کوچکتری از جامعه فراهم می‌شود. از نقطه نظر آماری اگر در طراحی یک آمارگیری، کوچکترین بخش جامعه آماری مورد توجه قرار گیرد، اما بخواهیم از این آمارگیری برآوردهایی در سطوح کوچکتر نیز به دست آوریم به این سطوح کوچکتر کوچک ناحیه گفته می‌شود (زارعی و همکاران، ۱۳۸۶). در کوچک ناحیه، برآوردهای مستقیم که فقط از اطلاعات خاص آن ناحیه استفاده می‌کنند، به علت اندازه کم نمونه معمولاً از دقت لازم برخوردار نیستند. از این رو با بکارگیری مدل‌ها که قابلیت لحاظ کردن اثرهای ناحیه و اطلاعات کمکی را در تحلیل‌ها دارند، می‌توان به برآوردهای مستقیم قدرت قرضی^۱ داده و دقت آن‌ها را افزایش داد. در این میان، مدل سطح ناحیه یا مدل فی-هریوت (فی و هریوت، ۱۹۷۹) به علت سادگی بیشتر و دسترسی آسان‌تر به اطلاعات کمکی سطح ناحیه و امکان بررسی فرضیات مدل به وسیله داده‌های نمونه‌ای، از اهمیت بیشتری برخوردار شده است. (مولینا و راتو، ۲۰۱۵).

در مدل سطح ناحیه برای راحتی و تفسیرپذیری بیشتر معمولاً فرض می‌شود هم اثرهای تصادفی سطح ناحیه و هم خطاهای نمونه‌گیری دارای توزیع نرمال هستند. این یک فرض محدود کننده بوده و در عمل مولفه‌های خطا (خطای مدل و خطای نمونه‌گیری) کمتر از توزیع نرمال پیروی می‌کنند. از طرف دیگر، وجود داده‌های دورافتاده باعث می‌شود پیش‌بینی دقیق پارامترهای کوچک ناحیه مشکل باشد و پذیرش فرض نرمال بودن برای مولفه‌های خطا نامعتبر گردد (تانگ و همکاران، ۲۰۱۸). معمولاً در روش‌های متداول به دست آوردن برآوردهای استوار کوچک ناحیه به بررسی وجود داده‌های دورافتاده فقط در اثرهای تصادفی سطح ناحیه پرداخته می‌شود. برای این کار دو دلیل وجود دارد (بل و هواهنگ، ۲۰۰۶). اول، این حالت توسعه طبیعی مدل فی-هریوت نرمال است و دوم اینکه قضیه حد مرکزی نرمال بودن تقریبی خطاهای نمونه‌گیری را نتیجه می‌دهد. اما همان‌طور که بل و هواهنگ (۲۰۰۶) بیان می‌کنند داده‌های دورافتاده می‌توانند در هر دو مولفه خطا رخ دهند و در عمل تفکیک اینکه داده دورافتاده مشاهده شده مربوط به کدام مولفه است دشوار و نشدنی می‌باشد. بنابراین در این مقاله به بررسی تاثیر داده‌های دورافتاده در مولفه‌های خطا بر برآوردهای میانگین‌های کوچک ناحیه‌ها با کاربست رهیافت بیز تجربی^۲ (EB) پرداخته می‌شود.

در سال‌های اخیر روش‌های استوار در حالت وجود داده‌های دورافتاده در برآوردهای مستقیم، بخش مهمی از مسائل مربوط به مدل سطح ناحیه را تشکیل می‌دهد. به عنوان مثال می‌توان به برآورد بیزی استوار در مدل فی-هریوت بر اساس توابع موثر (گوش و همکاران، ۲۰۰۸)، استفاده از تابع ψ -هوبر توسط

¹Borrow Strength

²Empirical Bayes

سینها و رائو (۲۰۰۹) و استفاده از واگرایی قدرت چگالی^۱ (باسو و همکاران، ۱۹۹۸) توسط سوگاساوا (۲۰۲۰) اشاره کرد.

ریتر (ریتر، ۲۰۱۵) داده‌های دورافتاده را به دو نوع خفیف^۲ و شدید^۳ تقسیم کرد. داده‌های دورافتاده خفیف از نمونه‌گیری از یک جامعه یا از یک مدل مفروض تولید می‌شوند. این داده‌ها نتیجه مسئله مشخصه‌سازی نادرست هستند. برای بررسی داده‌های دورافتاده خفیف یا می‌توان از مدل‌های استوار استفاده کرد و یا می‌توان آن‌ها را از مدل حذف و جداگانه بررسی کرد. داده‌های دورافتاده شدید غیرقابل پیش‌بینی و محاسبه هستند و با توزیع‌های معمول آماری مدل نمی‌شوند. اگر Q دامنه میان چارکی باشد، معمولاً داده‌ای که $k \geq 1/5$ از چارک اول کوچکتر یا از چارک سوم بزرگتر باشد، داده دورافتاده به حساب می‌آید. اگر $k \in [1/5, 3]$ باشد، داده دورافتاده خفیف و اگر $k \geq 3$ باشد، داده دورافتاده شدید است. برای مدل‌بندی داده‌های دورافتاده شدید می‌توان از توزیع‌های α -پایدار استفاده کرد (نولن، ۲۰۱۶). زارعی و همکاران (۲۰۲۱) برای اولین بار توزیع α -پایدار متقارن را برای اثرهای تصادفی سطح ناحیه بر اساس رهیافت بیز سلسه مراتبی به کار بردند. آن‌ها مدل حاصل را مدل α -پایدار فی-هریوت نامیدند. در بخش ۲ توزیع α -پایدار متقارن معرفی می‌شود. در بخش ۳ مدل فی-هریوت با مولفه‌های خطای α -پایدار متقارن معرفی می‌گردد. نحوه برآورد پارامترهای مدل پیشنهادی در بخش ۴ ارائه خواهد شد. در بخش ۵ با مطالعه‌های شبیه‌سازی این مدل ارزیابی و با مدل‌های مرتبط مقایسه می‌شود. در بخش ۶ مدل پیشنهادی برای توصیف داده‌های واقعی مورد استفاده قرار گرفته و نهایتاً بحث و نتیجه‌گیری ارائه می‌شود.

۲ توزیع α -پایدار متقارن

توزیع‌های پایدار یک کلاس غنی از توزیع‌های آماری هستند. این توزیع‌ها همزمان امکان بررسی دم‌کلفتی^۴ و چولگی در داده‌ها را فراهم می‌کنند (نولن، ۲۰۱۶). لوی (۱۹۲۵) در سال ۱۹۲۰ در جریان بررسی ویژگی‌های متغیرهای مستقل و هم توزیع، این توزیع‌ها را معرفی کرد. متغیر تصادفی یک متغیره پایدار (α -پایدار) عضو خانواده توزیع‌های چهار پارامتری است. شاخص پایداری^۵ یا شاخص دم^۶ $\alpha \in (0, 2]$ ، پارامتر چولگی $\beta \in [-1, 1]$ ، پارامتر مقیاس $\gamma > 0$ و پارامتر مکان $\delta \in \mathbb{R}$ چهار پارامتر این توزیع

¹Density Power Divergence

²Mild

³Gross

⁴Heavy Tail

⁵Stable Index

⁶Tail Index

هستند. در این مقاله متغیر تصادفی یک متغیره α -پایدار با $S(\alpha, \beta, \gamma, \delta)$ نمایش داده می‌شود. توزیع α -پایدار تعمیم توزیع نرمال است. پارامتر α رفتار دم توزیع را کنترل می‌کند و وقتی که $\alpha \rightarrow 0$ توزیع دم‌بلندتر است. برای $\alpha < 2$ واریانس توزیع وجود ندارد (نولن، ۲۰۱۶). این ویژگی باعث می‌شود که این توزیع برای بررسی مدل‌هایی با داده‌های دورافتاده شدید مناسب باشد. تابع چگالی متغیرهای تصادفی α -پایدار به جزء چند مورد خاص، صورت تحلیلی ندارند. بنابراین برای بررسی خواص این توزیع‌ها از تابع مشخصه^۱ استفاده می‌شود. متغیر تصادفی X دارای توزیع α -پایدار یک متغیره است اگر تابع مشخصه آن به صورت

$$\varphi_X(u) = \begin{cases} \exp(-\gamma^\alpha |u|^\alpha (1 - i\beta \tan(\frac{\pi\alpha}{\gamma}) \text{sign}(u)) + i\delta u), & \alpha \neq 1, \\ \exp(-\gamma |u| (1 + i\frac{\gamma}{\pi} \beta \text{sign}(u) \log |u|) + i\delta u), & \alpha = 1, \end{cases}$$

باشد، که در آن $\text{sign}(\cdot)$ تابع علامت است. لازم به ذکر است که اگر $\beta = 0$ باشد، توزیع را α -پایدار متقارن در اطراف δ می‌نامند. همچنین توزیع نرمال با میانگین δ و واریانس $2\gamma^2$ حالت خاصی از توزیع α -پایدار متقارن به صورت $S(2, 0, \gamma, \delta)$ است. برای بررسی کامل‌تر توزیع‌های α -پایدار می‌توان به سامورودنیتسکی و تاکو (۱۹۹۴) و نولن (۲۰۱۶) مراجعه کرد.

۳ مدل فی-هریوت با توزیع α -پایدار متقارن

در مدل فی-هریوت هدف، استنباط درباره تابع مناسبی از میانگین کوچک ناحیه i ام به صورت $\theta_i = g(\bar{Y}_i)$ است (مولینا و راثو، ۲۰۱۵). θ_i با بردار p -بعدی متغیرهای کمکی x_i به صورت

$$\theta_i = x_i^T \beta + b_i v_i, \quad i = 1, \dots, m, \quad (1)$$

ارتباط دارد، که در آن b_i مقدارهای معلوم و ثابت هستند و بدون اینکه از کلیت مساله کاسته شود فرض می‌شود مقدار آن‌ها ۱ است، $\beta = (\beta_1, \dots, \beta_p)^T$ برداری p -بعدی از ضرایب رگرسیونی، v_i اثر تصادفی خاص کوچک ناحیه i ام و m تعداد کوچک ناحیه‌ها در نمونه بوده و فرض می‌شود با تعداد کوچک ناحیه‌ها در جامعه برابر است. برای انجام استنباط درباره θ_i ها بر اساس مدل (۱) فرض می‌شود برآوردگرهای

¹Characteristic Function

مستقیم این پارامترها یعنی y_i ها در دسترس هستند. رابطه بین برآوردهای مستقیم و θ_i ها به صورت

$$y_i = \theta_i + e_i, \quad i = 1, \dots, m, \quad (2)$$

است، که در آن e_i خطای نمونه‌گیری است. با توجه به طرح نمونه‌گیری p این خطاها مستقل و دارای میانگین و واریانس

$$E_p(e_i|\theta_i) = 0, \quad V_p(e_i|\theta_i) = \phi_i, \quad (3)$$

هستند، که واریانس‌های ϕ_i معلوم فرض می‌شوند. از ترکیب (۱) و (۲) مدل فی-هریوت به صورت

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, m. \quad (4)$$

به دست می‌آید. همچنین فرض می‌شود که بردار خطاهای نمونه‌گیری $\mathbf{e} = (e_1, \dots, e_m)^T$ و بردار اثرهای تصادفی $\mathbf{v} = (v_1, \dots, v_m)^T$ از هم مستقل هستند. در مدل فی-هریوت فرض بر این است که v_i ها متغیرهایی تصادفی مستقل و هم توزیع از توزیع نرمال باشند. همان‌طور که اشاره شد این فرضیات در حضور داده‌های دور افتاده در مشاهدات معتبر نیستند و برآوردهای حاصل از کارایی لازم برخوردار نیستند. همچنین با توجه به رابطه $y_i - \mathbf{x}_i^T \boldsymbol{\beta} = v_i + e_i$ مشخص کردن اینکه داده دورافتاده ناشی از خطای مدل است یا خطای نمونه‌گیری بسیار دشوار است (بل و هواهنگ، ۲۰۰۶).

در این مقاله برای به دست آوردن مدلی استوار در حضور داده‌های دورافتاده ناشی از هر دو منبع خطا، توزیع α -پایدار متقارن جایگزین توزیع نرمال می‌شود. مدل حاصل، مدل فی-هریوت α -پایدار اصلاحی^۱ نامیده شده و با $R\alpha SMFH$ نمایش داده می‌شود. مزیت عملی استفاده از توزیع α -پایدار متقارن آن است که دم‌های این توزیع در مقایسه با توزیع نرمال بسیار کلفت‌تر است و می‌تواند برآوردهایی استوار در حضور داده‌های دورافتاده شدید را فراهم کند. همچنین توزیع نرمال را به عنوان حالت خاص پوشش دهد و اگر داده‌ها چولگی بیشتر از حد نرمال داشته باشند، با مدل جدید بهتر بررسی می‌شوند.

ساختار آماری مدل $R\alpha SMFH$ به صورت

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, m. \quad (5)$$

¹Modified

۴۶۸ برآورد کوچک ناحیه‌ای بیز تجربی استوار با توزیع α -پایدار

است، که در آن $v_i \stackrel{i.i.d}{\sim} S(\alpha, \circ, \gamma_1, \circ)$ و $e_i \stackrel{ind}{\sim} S(\alpha, \circ, \psi_i, \circ)$. همانند مدل فی-هریوت فرض می‌شود پارامترهای مقیاس ψ_i معلوم باشند. در این مدل نیز، هدف اصلی پیش‌بینی پارامترهای کوچک ناحیه $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i$ با بیشترین دقت است. بر طبق خاصیت مجموع دو متغیر تصادفی α -پایدار (سامورودنیتسکی و تاکو، ۱۹۹۴) توزیع حاشیه‌ای y_i به صورت

$$y_i \sim S(\alpha, \circ, (\gamma_1^\alpha + \psi_i^\alpha)^{\frac{1}{\alpha}}, \mathbf{x}_i^T \boldsymbol{\beta}).$$

است. چون تابع‌های چگالی توزیع‌های α -پایدار در حالت کلی دارای فرم بسته نیستند، برآورد کردن پارامترها در مدل‌های شامل این توزیع‌ها کار سختی است.

۴ برآورد بیز تجربی پارامترهای مدل RaSMFH

رویکرد بیز تجربی (EB) در برآورد کوچک ناحیه را در حالت کلی می‌توان به صورت زیر خلاصه کرد (رائو، ۲۰۰۳). فرض کنید \mathbf{y} بردار مشاهدات، $\boldsymbol{\lambda}$ بردار پارامترهای مدل و $\boldsymbol{\theta}$ بردار پارامترهای (تصادفی) کوچک ناحیه باشد. در این صورت چگالی پسینی $\pi(\boldsymbol{\theta} | \mathbf{y}, \boldsymbol{\lambda})$ را با استفاده از چگالی‌های شرطی $\pi(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\lambda}_1)$ و $\pi(\boldsymbol{\theta} | \boldsymbol{\lambda}_2)$ به دست می‌آورند، که در آن $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T)^T$ بردار پارامترهای مدل است. سپس پارامترهای $\boldsymbol{\lambda}$ از چگالی حاشیه‌ای $\pi(\mathbf{y} | \boldsymbol{\lambda})$ برآورد می‌شوند. در نهایت از چگالی برآورد شده $\pi(\boldsymbol{\theta} | \mathbf{y}, \hat{\boldsymbol{\lambda}})$ که در آن $\hat{\boldsymbol{\lambda}}$ برآوردگری برای $\boldsymbol{\lambda}$ است، برای انجام استنباط درباره $\boldsymbol{\theta}$ استفاده می‌شود. برای آشنایی با روش پیدا کردن برآوردها به روش بیشینه درست‌نمایی برای توزیع‌های α -پایدار می‌توان تیموری (۱۳۹۹) را مشاهده کرد.

با فرض معلوم بودن پارامترهای مدل (۵) همانند رویکرد EB برای مدل فی-هریوت با مولفه‌های خطای نرمال (رائو، ۲۰۰۳)، فرضیات این مدل را به صورت یک مدل سلسله مراتبی دو مرحله‌ای به صورت

$$y_i | \theta_i \stackrel{ind}{\sim} S(\alpha, \circ, \psi_i, \theta_i) \quad \theta_i \stackrel{ind}{\sim} S(\alpha, \circ, \gamma_1, \mathbf{x}_i^T \boldsymbol{\beta})$$

می‌توان نمایش داد. پس در مدل α -پایدار پیشنهادی (۵)، $\lambda_1 = \alpha$ و $\boldsymbol{\lambda}_2^T = (\alpha, \gamma_1, \boldsymbol{\beta})$ است.

بنابراین برآوردگر بیزی θ_i با توجه به تابع زیان درجه دوم به صورت

$$E(\theta_i|y_i, \alpha, \gamma_1, \beta) = \frac{\int \theta_i \pi(y_i|\theta_i, \alpha) \pi(\theta_i|\alpha, \gamma_1, \beta) d\theta_i}{\int \pi(y_i|\theta_i, \alpha) \pi(\theta_i|\alpha, \gamma_1, \beta) d\theta_i} \quad (۶)$$

است. چون توابع چگالی احتمال α -پایدار $\pi(\theta_i|\alpha, \gamma_1, \beta)$ و $\pi(y_i|\theta_i, \alpha)$ در حالت کلی دارای فرم بسته نیستند، برآوردگر (۶) را به روش تحلیلی نمی‌توان محاسبه کرد، برای محاسبه آن از انتگرال مونت کارلویی استفاده می‌شود. اگر یک نمونه به حجم N از توزیع $\pi(\theta_i|\alpha, \gamma_1, \beta)$ به صورت $\theta_i^{(1)}, \dots, \theta_i^{(N)}$ تولید شود، خواهیم داشت

$$E(\theta_i|y_i) \simeq \frac{\frac{\sum_{j=1}^N \theta_i^{(j)} \pi(y_i|\theta_i^{(j)}, \alpha)}{N}}{\frac{\sum_{j=1}^N \pi(y_i|\theta_i^{(j)}, \alpha)}{N}} = \frac{\sum_{j=1}^N \theta_i^{(j)} \pi(y_i|\theta_i^{(j)}, \alpha)}{\sum_{j=1}^N \pi(y_i|\theta_i^{(j)}, \alpha)} \quad (۷)$$

این برآوردگر نقطه‌ای به پارامترهای α ، γ_1 و β بستگی دارد. برای برآورد این پارامترها به روش بیشینه درست‌نمایی از توزیع حاشیه‌ای y_i ها یعنی $(\gamma_1^\alpha + \psi_i^\alpha)^{\frac{1}{\alpha}}, \mathbf{x}_i^T \beta$ استفاده خواهد شد. چون این چگالی از نوع α -پایدار است و فرم بسته‌ای برای آن در حالت کلی وجود ندارد، تابع درست‌نمایی نیز دارای فرم بسته نیست. از این رو برای یافتن برآورد روش بیشینه درست‌نمایی از تابع $nlminb$ در نرم افزار R و بسته $STABLE$ ، قابل دسترس در <http://www.robustanalysis.com>، استفاده می‌شود. لازم به ذکر است که در محاسبات انجام شده در بخش‌های بعد $N = 5000$ انتخاب شده است.

۵ ارزیابی مدل $R\alpha SMFH$

در این بخش بر اساس چند مطالعه شبیه‌سازی، مدل استوار پیشنهادی بر اساس رهیافت بیز تجربی که با $R\alpha SEB$ نمایش داده می‌شود، با روش بیز تجربی کلاسیک براساس توزیع نرمال (CNEB)، روش بیز تجربی استوار ارائه شده توسط سوگاساوا (۲۰۲۰) در دو حالت و وقتی که ۱ و ۵ درصد به میانگین توان دوم خطای برآوردگر بیزی اضافه شود، که به ترتیب با EB\RD PD و EB\RD PD نشان داده خواهند شد، مورد مقایسه قرار می‌گیرد. سوگاساوا (۲۰۲۰) نشان داد که برآوردگر بیزی θ_i تحت مدل (۴) را می‌توان از رابطه

$$\hat{\theta}_i^B = y_i + \phi_i \frac{\partial}{\partial y_i} \log f(y_i|\lambda),$$

۴۷۰ برآورد کوچک ناحیه‌ای بیز تجربی استوار با توزیع α -پایدار

نیز به دست آورد، که در آن $\hat{\theta}_i^B$ برآوردگر بیزی پارامتر θ_i است. بنابراین با جانشین کردن $\log f(y_i|\lambda)$ با یک جایگزین استوار، مانند تابع درست‌نمایی روش واگرایی قدرت چگالی می‌توان به برآوردگرهای استوارتری رسید. بر اساس رهیافت **باسو و همکاران (۱۹۹۸)** برای به دست آوردن برآوردگرهای استوار پارامترهای مدل λ در مدل (۴) می‌توان تابع

$$\frac{1}{\rho} \sum_{i=1}^m f_i(y_i, \lambda)^\rho - \frac{1}{1+\rho} \sum_{i=1}^m \int f_i(t, \lambda)^{\rho+1} dt,$$

را بر حسب λ بیشینه کرد، که در آن $f_i(\cdot)$ تابع چگالی احتمال y_i و پارامتر $\rho \in (0, 1)$ تعادل بین استواری مدل و کارایی مجانبی را کنترل می‌کند. استفاده از این روش علی‌رغم افزایش استواری برآوردگرهای بیزی θ_i ، معمولاً منجر به افزایش میانگین توان دوم خطای این برآوردگر می‌شود. بنابراین باید تعادلی بین میزان افزایش میانگین توان دوم خطا و استواری مدل پیدا کرد. **سگاساوا (۲۰۲۰)** نشان داد که اگر شاخص ρ طوری تعیین شود که ۱ یا ۵ درصد به میانگین توان دوم خطای برآوردگر بیزی معمولی اضافه شود، می‌توان به این تعادل رسید.

۵.۱ مطالعه شبیه‌سازی اول

اولین شبیه‌سازی با روشی مشابه با **سگاساوا (۲۰۲۰)**، **چاکرابورتی و همکاران (۲۰۱۸)** و **زارعی و همکاران (۲۰۲۱)** انجام می‌شود. تعداد کوچک ناحیه‌های m ، ۲۰، ۵۰ و ۱۰۰ و شاخص پایداری $\alpha/8$ ، $1/5$ و ۲ در نظر گرفته شده است. به ازای $\alpha = 2$ ، توزیع مورد استفاده نرمال است. در حالت $\alpha = 1/5$ امکان تولید هر دو نوع داده‌های دورافتاده وجود دارد، ولی برای $\alpha = 0/8$ داده‌های تولید شده اغلب از نوع دور افتاده خیلی شدید هستند. برای هر انتخاب m ، پنج سناریوی مختلف تولید داده بررسی می‌شود. در سناریوی اول v_i و e_i به ترتیب دارای توزیع نرمال استاندارد و نرمال با میانگین صفر و واریانس ψ_i منظور شده‌اند، در سناریوی دوم $(S(1/5, 0, \frac{1}{\sqrt{4}}, 0))$ و $v_i \sim S(1/5, 0, \frac{1}{\sqrt{4}}, 0)$ و $e_i \sim N(0, \psi_i)$ در سناریوی سوم $(S(1/5, 0, \frac{1}{\sqrt{4}}, 0))$ و $v_i \sim S(1/5, 0, \frac{1}{\sqrt{4}}, 0)$ و $e_i \sim N(0, \psi_i/2)$ در سناریوی چهارم $(N(0, 1))$ و $v_i \sim N(0, 1)$ و $e_i \sim S(1/5, 0, \sqrt{\psi_i/2}, 0)$ و سناریوی پنجم همان سناریوی دوم است، با این تفاوت که شاخص پایداری $\alpha = 0/8$ است. همچنین فرض می‌شود فقط یک متغیر مستقل با توزیع یکنواخت پیوسته استاندارد وجود دارد. بعلاوه ضرایب رگرسیونی هم به صورت $\beta_1 = 2$ و $\beta_0 = 0$ هستند. واریانس‌های نمونه‌گیری همانند **چاکرابورتی و همکاران (۲۰۱۸)** و **تانگ و همکاران (۲۰۱۸)**

از مجموعه $\{0, 5/10, 2, 1/5, 3, 2/5, 4, 3/5, 4, 4/5, 5\}$ انتخاب می‌شوند و هر عضو این مجموعه شانسی یکسان برای انتخاب به عنوان واریانس نمونه‌گیری را دارد.

در هر سناریو، بعد از تولید نمونه‌ای تصادفی به حجم m از توزیع‌های مولفه‌های خطا، با توجه به مقادیر ضرایب رگرسیون، برآوردهای مستقیم از مدل

$$y_i = \beta x_i + v_i + e_i, \quad i = 1, \dots, m,$$

محاسبه می‌شوند. بنابراین مقدار واقعی θ_i در هر تکرار شبیه‌سازی برابر $\beta x_i + v_i$ است. برای مقایسه روش‌های برآورد کوچک ناحیه‌ای مورد اشاره، از متوسط توان دوم‌های انحراف^۱ (ASD) و متوسط قدرمطلق انحراف^۲ (AAD) (تانگ و همکاران، ۲۰۱۸) به صورت

$$ASD = \frac{1}{m} \sum_{i=1}^m (\theta_i - \hat{\theta}_i)^2, \quad AAD = \frac{1}{m} \sum_{i=1}^m |\theta_i - \hat{\theta}_i|,$$

استفاده می‌شود، که در آن $\hat{\theta}_i$ برآورد میانگین کوچک ناحیه i ام یعنی θ_i است. هر سناریوی شبیه‌سازی ۵۰۰ بار تکرار می‌شود. در هر تکرار معیارهای ASD و AAD و تمام پارامترهای مدل محاسبه و برآورد می‌شوند. با استفاده از رابطه $\hat{\theta}_i = \hat{\beta}^T x_i + v_i$ پارامتر کوچک ناحیه‌ای i ام نیز برآورد می‌شود. سپس متوسط مقادیرهای به‌دست آمده بر اساس این ۵۰۰ تکرار محاسبه می‌شوند.

جدول ۱ نشان می‌دهد که در سناریوی ۱ می‌توان گفت افزایش تعداد ناحیه‌ها باعث افزایش دقت برآورد میانگین کوچک ناحیه‌ها شده است. نتایج به‌دست آمده در سناریوهای ۱، ۲، ۴ و ۵ یعنی وقتی که مولفه‌های خطا دارای توزیع نرمال، توزیع $1/5$ -پایدار، وقتی که توزیع خطای مدل نرمال استاندارد و خطای نمونه‌گیری دارای توزیع $1/5$ -پایدار و هنگامی که هر دو مولفه خطا دارای توزیع $0/9$ -پایدار هستند، به‌طور کلی نشان‌دهنده بهتر بودن روش پیشنهادی است.

در سناریوی اول که توزیع مولفه‌های خطا هر دو از خانواده نرمال است، بر اساس هر دو شاخص، به‌طور کلی روش پیشنهادی از روش CNEB نیز بهتر است، علت این امر آن است که توزیع نرمال حالت خاصی از توزیع پایدار است. با این خاصیت اضافه که توزیع پایدار برای داده‌هایی که آزمون نرمال بودن آن‌ها رد نمی‌شود، ولی اندکی دارای چولگی باشند، برازش بهتری فراهم می‌کند. همچنین چون در این

¹ Average Squared Deviation

² Average Absolute Deviation

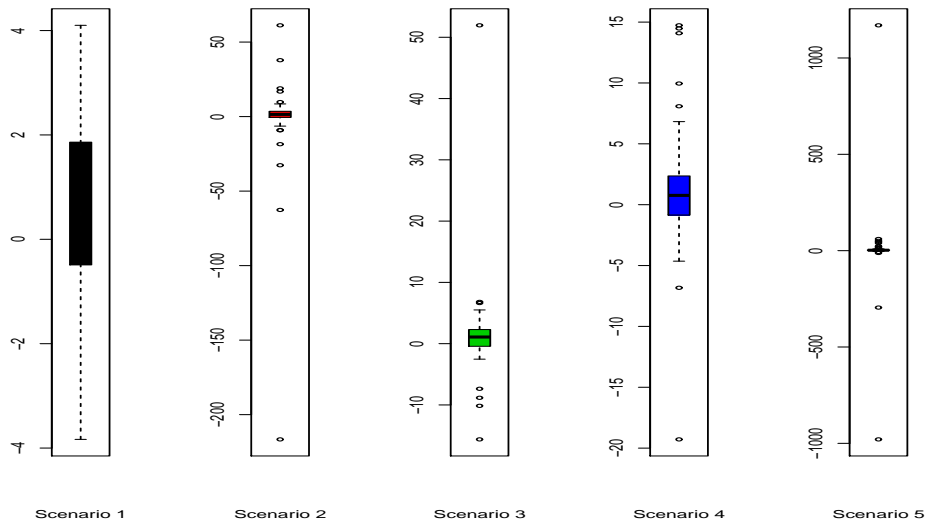
جدول ۱. متوسط ASD و AAD محاسبه شده بر اساس تکرارهای مختلف شبیه‌سازی در سناریوهای مختلف.

سناریو	معیار	برآوردگر			m			معیار	برآوردگر				
		۱۰۰	۵۰	۲۰	۱۰۰	۵۰	۲۰		۱۰۰	۵۰	۲۰		
۱	ASD	CNEB	۰/۷۲	۰/۸۰	۰/۹۱	CNEB	۰/۶۷	۰/۷۱	۰/۷۶	CNEB	۰/۶۷	۰/۷۱	۰/۷۶
		EB\RDPA	۰/۷۵	۰/۸۴	۰/۹۵	EB\RDPA	۰/۶۸	۰/۷۲	۰/۷۷	EB\RDPA	۰/۶۸	۰/۷۲	۰/۷۷
		EB\RDPA	۰/۷۳	۰/۸۲	۰/۹۴	EB\RDPA	۰/۶۸	۰/۷۱	۰/۷۶	EB\RDPA	۰/۶۸	۰/۷۱	۰/۷۶
		RaSEB	۰/۷۴	۰/۷۸	۰/۸۹	RaSEB	۰/۶۸	۰/۷۰	۰/۷۴	RaSEB	۰/۶۸	۰/۷۰	۰/۷۴
۲	ASD	CNEB	۴۷/۲۷	۵۵/۶۷	۲۷/۵۷	CNEB	۱/۷۲	۱/۵۳	۱/۶۳	CNEB	۱/۷۲	۱/۵۳	۱/۶۳
		EB\RDPA	۴۸/۶۶	۵۶/۳۹	۲۸/۲۱	EB\RDPA	۱/۶۲	۱/۴۸	۱/۶۰	EB\RDPA	۱/۶۲	۱/۴۸	۱/۶۰
		EB\RDPA	۴۸/۴۱	۵۶/۱۷	۲۸/۰۸	EB\RDPA	۱/۶۲	۱/۴۸	۱/۶۱	EB\RDPA	۱/۶۲	۱/۴۸	۱/۶۱
		RaSEB	۱۶/۹۳	۶/۴۸	۷/۲۸	RaSEB	۱/۱۲	۱/۰۶	۱/۲۳	RaSEB	۱/۱۲	۱/۰۶	۱/۲۳
۳	ASD	CNEB	۱/۶۳	۱/۴۳	۱/۶۵	CNEB	۰/۹۰	۰/۸۷	۰/۹۵	CNEB	۰/۹۰	۰/۸۷	۰/۹۵
		EB\RDPA	۱/۲۵	۱/۲۶	۱/۵۹	EB\RDPA	۰/۸۳	۰/۸۴	۰/۹۴	EB\RDPA	۰/۸۳	۰/۸۴	۰/۹۴
		EB\RDPA	۱/۲۹	۱/۳۰	۱/۶۴	EB\RDPA	۰/۸۴	۰/۸۵	۰/۹۵	EB\RDPA	۰/۸۴	۰/۸۵	۰/۹۵
		RaSEB	۳/۷۸	۳۷/۲۷	۳/۰۱	RaSEB	۰/۹۲	۱/۰۰	۰/۹۴	RaSEB	۰/۹۲	۱/۰۰	۰/۹۴
۴	ASD	CNEB	۳۱/۰۵	۵۸/۹۱	۴۳/۶۹	CNEB	۱/۳۸	۱/۵۵	۱/۶۱	CNEB	۱/۳۸	۱/۵۵	۱/۶۱
		EB\RDPA	۳۲/۲۹	۵۹/۸۹	۴۴/۲۸	EB\RDPA	۱/۳۱	۱/۵۰	۱/۵۹	EB\RDPA	۱/۳۱	۱/۵۰	۱/۵۹
		EB\RDPA	۳۱/۹۸	۵۹/۶۱	۴۴/۱۰	EB\RDPA	۱/۳۰	۱/۵۰	۱/۶۰	EB\RDPA	۱/۳۰	۱/۵۰	۱/۶۰
		RaSEB	۱/۵۱	۷/۷	۳/۳۸	RaSEB	۰/۷۹	۰/۸۸	۰/۹۴	RaSEB	۰/۷۹	۰/۸۸	۰/۹۴
۵	ASD	CNEB	۳۱۷۹۰۸۱	۵۱۷۰۹	۹۹۶۱۲	CNEB	۲۷/۲۶	۱۲/۹۴	۱۸/۶۰	CNEB	۲۷/۲۶	۱۲/۹۴	۱۸/۶۰
		EB\RDPA	۳۱۷۹۰۸۳	۵۱۷۱۱	۹۹۶۱۴	EB\RDPA	۲۷/۱۵	۱۲/۸۵	۱۸/۵۳	EB\RDPA	۲۷/۱۵	۱۲/۸۵	۱۸/۵۳
		EB\RDPA	۳۱۷۹۰۸۲	۵۱۷۱۰	۹۹۶۱۳	EB\RDPA	۲۷/۱۶	۱۲/۸۵	۱۸/۵۳	EB\RDPA	۲۷/۱۶	۱۲/۸۵	۱۸/۵۳
		RaSEB	۱۰۱۱۹	۴۸۹۴۲	۶۹۷۹	RaSEB	۶/۹۸	۸/۹۵	۷/۸۸	RaSEB	۶/۹۸	۸/۹۵	۷/۸۸

سناریو داده دور افتاده نداریم، مقادیرهای معیارهای ارزیابی کوچکتر محاسبه شده‌اند، که نشان‌دهنده دقت بهتر پیش‌بینی‌هاست.

در سناریوی دوم و پنجم مدل پیشنهادی از مدل‌های رقیب بسیار بهتر عمل کرده است. برای بررسی بیشتر در شکل ۱ نمودار جعبه‌ای برآوردهای مستقیم در سناریوهای مختلف رسم شده است. همان‌طور که از این شکل مشخص است، در سناریوی دوم و مخصوصاً پنجم، امکان وجود داده‌های دورافتاده شدید در مشاهدات خیلی بیشتر است. از آنجا که توزیع نرمال در رهیافت بیز تجربی کلاسیک و روش استوار معرفی شده توسط سوگاساوا (۲۰۲۰) نمی‌توانند این داده‌های دورافتاده شدید را تحت کنترل قرار بدهند، برآوردهای حاصل از واقعیت خیلی دور بوده و باعث ASD و ADD های خیلی بزرگ نسبت به روش RaSEB شده است. بنابراین به عنوان یک قاعده سرانگشتی برای استفاده از مدل پیشنهادی در عمل، می‌توان وجود داده‌های دورافتاده شدید در برآوردهای مستقیم را بررسی کرد.

در سناریوی سوم، فقط اثر ناحیه‌ها دارای توزیع دم کلفت ۱/۵-پایدار بوده و خطای نمونه‌گیری نرمال است. چون مدل پیشنهادی بر اساس توزیع α -پایدار با شاخص پایداری یکسان برای مولفه‌های خطاست، در این حالت RaSEB عملکرد ضعیف‌تری داشته است. در واقع در روش پیشنهادی، انحراف معیار بزرگ ASD، نشان می‌دهد که ASD در برخی از تکرارها بسیار بزرگ برآورد شده است و باعث شده به طور



شکل ۱. نمودار جعبه‌ای برآوردهای مستقیم در یکی از تکرارهای شبیه‌سازی برای سناریوهای مختلف. در سناریوی دوم و مخصوصاً پنجم امکان وجود داده‌های دورافتاده خیلی بزرگ (شدید) در مشاهدات بیشتر از سناریوهای سوم و چهارم است.

میانگین ASD این روش از ASD های روش‌های رقیب بزرگتر باشد. این نتیجه‌گیری بر اساس شاخص AAD نیز تقریباً یکسان است.

از نکات قابل توجه در نتایج این شبیه‌سازی بهتر بودن روش CNEB در سناریوهای ۲ و ۴ نسبت به روش‌های استوار EB\1RDPD و EB\2RDPD بر اساس شاخص ASD است. می‌توان گفت که اگر نشانه‌ای از دورافتاده بودن اثرهای تصادفی خاص هر ناحیه مشاهده نشد، بهتر است از روش‌های EB\1RDPD و EB\2RDPD استفاده نکرد.

برای اطمینان بیشتر به نتایج حاصل، حالتی که توزیع واقعی مولفه‌های خطا از خانواده توزیع‌های دم‌بلند متفاوت باشد، را برای سناریوی اصلی، یعنی سناریوی دوم هم بررسی می‌کنیم. فرض کنید اثرهای تصادفی خاص هر ناحیه دارای توزیع t -استیودنت با درجه آزادی ۳ باشند و این اثرها را از این توزیع تولید کنیم. علاوه بر این فرض کنید که e_i دارای توزیع کوشی با پارامتر مکان صفر و پارامتر مقیاس $\sqrt{\psi_i/2}$ بوده و ψ_i ها از توزیع گاما با پارامتر شکل ۴/۵ و پارامتر مقیاس ۲ تولید شوند. به‌عنوان نمونه در حالت $m = 50$ شبیه‌سازی را ۱۰۰ بار انجام داده مشاهده می‌کنیم که متوسط ASD های محاسبه شده در این حالت برای روش پیشنهادی ۵/۳۴ و برای روش CNEB و روش‌های EB\1RDPD و EB\2RDPD به

ترتیب برابر ۱۲۴/۵۲ ، ۱۲۵/۸۶ و ۱۲۵/۸۷ است و بهتر بودن روش پیشنهادی را مجدداً تأیید می‌کند.

برای بررسی بیشتر کارایی مدل پیشنهادی، توانایی آن در برآورد کردن پارامترها بررسی شده و متوسط برآورد پارامترها به همراه انحراف معیار مقادیر برآورد شده (داخل پرانتز) در تکرارهای مختلف شبیه‌سازی در جدول ۲ آمده است. در این جدول α_v و α_e به ترتیب شاخص‌های پایداری مربوط به خطای مدل و خطای نمونه‌گیری بوده و $\gamma = 2\gamma_1^2$ است. جدول ۲ نشان می‌دهد که روش $R\alpha SEB$ در برآورد شاخص‌های پایداری و ضرایب رگرسیونی بسیار خوب عمل می‌کند، ولی در برآورد پارامتر مقیاس نسبتاً خوب عمل می‌کند. همان‌طور که انتظار می‌رفت، وجود داده‌های دورافتاده بسیار شدید در سناریوی پنجم باعث به‌وجود آمدن انحراف معیارهای بزرگتری برای برآورد پارامترها شده و دقت برآوردها در این حالت نسبت به سناریوهای قبلی کمتر است.

جدول ۲. متوسط برآورد شده برای پارامترهای مدل در تکرارهای مختلف شبیه‌سازی.

m			مقدار واقعی پارامتر	سناریو
۱۰۰	۵۰	۲۰		
۱/۹۵ (۰/۰۸)	۱/۹۳ (۰/۱۴)	۱/۹۱ (۰/۱۷)	$\alpha_v = \alpha_e = 2$	۱
۱/۳۴ (۰/۱۵)	۱/۳۳ (۰/۲۲)	۱/۳۰ (۰/۳۷)	$\gamma = 1$	
-۰/۰۲ (۰/۳۷)	۰/۰۱ (۰/۵۳)	-۰/۰۴ (۰/۸۳)	$\beta_0 = 0$	
۲/۰۳ (۰/۷۲)	۱/۹۷ (۰/۹۵)	۲/۰۱ (۱/۵۶)	$\beta_1 = 2$	
۱/۵۳ (۰/۱۵)	۱/۵۶ (۰/۲۳)	۱/۵۳ (۰/۳۶)	$\alpha_v = \alpha_e = 1/5$	۲
۱/۴۱ (۰/۲۱)	۱/۳۶ (۰/۳۲)	۱/۲۹ (۰/۵۳)	$\gamma = 1$	
۰/۰۶ (۰/۴۴)	۰/۰۲ (۰/۵۲)	۰/۰۸ (۱/۰۶)	$\beta_0 = 0$	
۱/۸۹ (۰/۷۹)	۱/۹۸ (۰/۹۷)	۱/۹۲ (۱/۷۳)	$\beta_1 = 2$	
۱/۷۹ (۰/۱۵)	۱/۸۱ (۰/۱۷)	۱/۸۱ (۰/۲۶)	$\alpha_v = 2, \alpha_e = 1/5$	۳
۱/۴۰ (۰/۱۸)	۱/۳۸ (۰/۲۶)	۱/۳۱ (۰/۴۱)	$\gamma = 1$	
-۰/۰۵ (۰/۴۵)	-۰/۰۹ (۰/۶۳)	۰/۰۶ (۱/۰۱)	$\beta_0 = 0$	
۲/۰۳ (۰/۷۵)	۲/۱۲ (۱/۰۶)	۱/۷۴ (۱/۸۸)	$\beta_1 = 2$	
۱/۶۶ (۰/۱۶)	۱/۵۹ (۰/۲۴)	۱/۶۷ (۰/۳۱)	$\alpha_v = 1/5, \alpha_e = 2$	۴
۱/۳۵ (۰/۲۲)	۱/۲۸ (۰/۲۸)	۱/۳۳ (۰/۴۸)	$\gamma = 1$	
-۰/۰۳ (۰/۳۹)	-۰/۰۵ (۰/۵۶)	-۰/۰۸ (۱/۰۷)	$\beta_0 = 0$	
۲/۰۷ (۰/۶۸)	۱/۹۸ (۱/۰۲)	۲/۰۷ (۱/۸۶)	$\beta_1 = 2$	
۰/۸۸ (۰/۱۰)	۰/۹۰ (۰/۱۸)	۰/۹۵ (۰/۳۴)	$\alpha_v = \alpha_e = 0/8$	۵
۱/۶۴ (۰/۳۶)	۱/۷۵ (۰/۶۴)	۱/۷۲ (۱/۱۰)	$\gamma = 1$	
-۰/۰۱ (۰/۴۶)	-۰/۰۱ (۰/۹۸)	۰/۰۷ (۱/۶۱)	$\beta_0 = 0$	
۱/۹۰ (۰/۸۸)	۱/۹۱ (۱/۷۱)	۲/۱۶ (۲/۹۴)	$\beta_1 = 2$	

۵.۲ مطالعه شبیه‌سازی دوم

برای انجام دومین شبیه‌سازی از روشی مشابه با سوگاساوا (۲۰۲۰) استفاده می‌کنیم. هدف این شبیه‌سازی ارزیابی بیشتر کارایی مدل $R\alpha SEB$ و امکان مقایسه نتایج شبیه‌سازی با نتایج حاصل از تحلیل داده‌های واقعی است. تعداد کوچک ناحیه‌ها را $m = 40$ انتخاب می‌کنیم. فرض می‌شود این چهل کوچک ناحیه، در چهار حوزه بزرگ قرار داشته باشند. واریانس‌های نمونه‌گیری در این حوزه‌ها به ترتیب $0/2, 0/4, 0/8$ و 1 فرض می‌شوند. نحوه تولید داده مشابه اولین شبیه‌سازی است. در این حالت سناریوهای اول و دوم یعنی حالتی که مولفه‌های خطا هر دو نرمال یا هر دو دارای توزیع $1/4$ -پایدار باشند، مورد بررسی قرار می‌گیرند. در این حالت علاوه بر شاخص‌های ارزیابی ASD و AAD متوسط $RMSE$ برای کوچک ناحیه‌های داخل هر حوزه (ناحیه) بزرگ نیز محاسبه می‌شود. شاخص $RMSE$ برای کوچک ناحیه i ام به صورت

$$RMSE_i = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_i^{(r)} - \theta_i^{(r)})^2} \quad (8)$$

تعریف می‌شود، که در آن $\theta_i^{(r)}$ و $\hat{\theta}_i^{(r)}$ مقدار واقعی و مقدار برآورد شده پارامتر کوچک ناحیه i ام بوده و R تعداد دفعات انجام شبیه‌سازی و برابر 500 است. همانند سوگاساوا (۲۰۲۰) در هر ناحیه بزرگ، متوسط $RMSE$ را برای کوچک ناحیه‌های آن ناحیه بزرگ محاسبه می‌کنیم. همان‌طور که در جدول ۳ ملاحظه

جدول ۳. متوسط ASD و AAD و $RMSE$ محاسبه شده بر اساس تکرارهای مختلف.

$\alpha = 1/4$ سناریوی دوم				$\alpha = 2$ سناریوی اول				
$R\alpha SEB$	EB^*RDPD	EB^*RDPD	CNEB	$R\alpha SEB$	EB^*RDPD	EB^*RDPD	CNEB	
۲/۱۶	۱/۳۳	۱/۳۳	۱/۳۳	۰/۴۴	۰/۴۳	۰/۴۲	۰/۴۲	حوزه‌های ۱MD
۱/۶۹	۳/۱۳	۳/۱۴	۳/۱۳	۰/۵۸	۰/۵۶	۰/۵۶	۰/۵۶	۲MD
۱/۷۹	۴/۳۲	۴/۳۱	۴/۳۰	۰/۶۸	۰/۷۱	۰/۷۰	۰/۶۹	۳MD
۲/۰۴	۲/۲۲	۲/۲۱	۲/۲۳	۰/۷۷	۰/۸۶	۰/۸۵	۰/۸۳	۴MD
۵/۹۴	۲۱/۱۳	۲۱/۱۷	۲۱/۱۳	۰/۴۵	۰/۴۶	۰/۴۶	۰/۴۴	معیار ASD
۰/۷۹	۰/۸۶	۰/۸۶	۰/۸۷	۰/۴۸	۰/۴۸	۰/۴۸	۰/۴۸	AAD

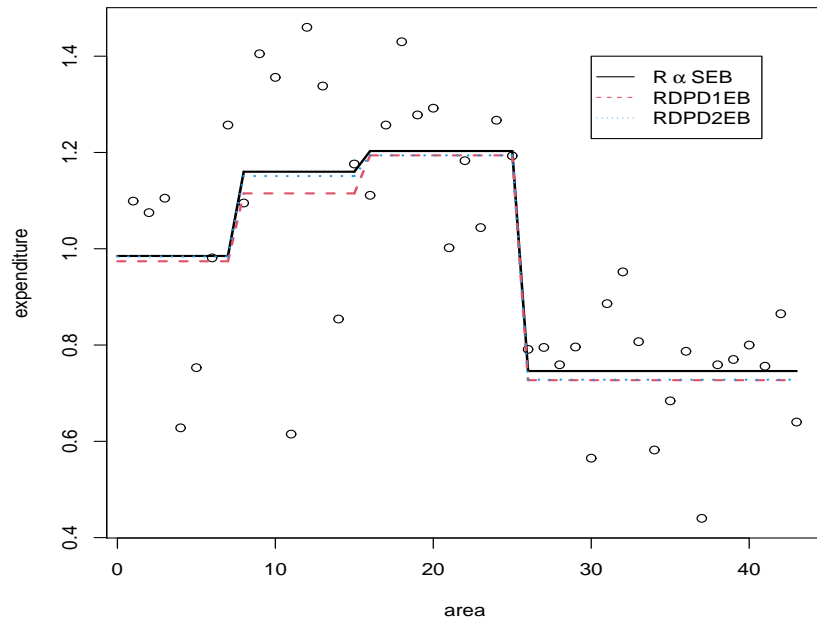
می‌شود، در سناریوی اول که مولفه‌های خطا از توزیع نرمال تولید می‌شوند، بررسی مقادیرهای متوسط $RMSE$ نشان می‌دهد که در حوزه اول و دوم که واریانس نمونه‌گیری کوچکتر است، عملکرد $R\alpha SEB$ از روش‌های رقیب اندکی ضعیفتر است. اما در حوزه‌های سوم و چهارم و با افزایش واریانس‌های نمونه‌گیری، عملکرد $R\alpha SEB$ از سایر روش‌ها بهتر شده است. در مجموع و با توجه به مقادیرهای به‌دست آمده برای

شاخص‌های ASD و ADD، می‌توان گفت عملکرد روش CNEB اندکی از روش‌های دیگر بهتر بوده و $R\alpha$ SEB در رتبه دوم قرار دارد. با توجه به نرمال بودن مولفه‌های خطا این نتیجه قابل انتظار است. اما در سناریوی دوم، به علت وجود داده‌های دورافتاده شدید در مشاهدات $R\alpha$ SEB (بجز درحوزه بزرگ اول که دارای کمترین واریانس نمونه‌گیری است) بهترین عملکرد را داشته است.

۶ تحلیل داده‌های واقعی

در این بخش برای بررسی بیشتر کارایی مدل پیشنهادی، آن را برای تحلیل داده‌هایی به نام «شیر» که از اداره آمار کار ایالات متحده آمده است، مورد استفاده قرار می‌گیرد. این داده‌ها قبلاً توسط **آرورا و لاهیری (۱۹۹۷)**، **یو و چپمن (۲۰۰۶)** و **سوگاساوا (۲۰۲۰)** تحلیل شده‌اند. به طور خاص ما نتایج به دست آمده را با نتایج **سوگاساوا (۲۰۲۰)** که یک روش EB استوار است، مقایسه می‌گیرد. در این داده‌ها مقدارهای برآورد شده متوسط مصرف شیر تازه برای سال ۱۹۸۹ که با y_i نمایش داده می‌شود، برای ۴۳ ناحیه با واریانس‌های نمونه‌گیری D_i در دسترس هستند. این داده‌ها در ۴ حوزه بزرگ به صورت $M_1 = \{1, \dots, 7\}$ ، $M_2 = \{8, \dots, 14\}$ ، $M_3 = \{15, \dots, 25\}$ و $M_4 = \{26, \dots, 43\}$ قرار دارند.

همانند **یو و چپمن (۲۰۰۶)** و **سوگاساوا (۲۰۲۰)** اگر i امین ناحیه متعلق به j امین حوزه بزرگ باشد ($j = 1, 2, 3, 4$)، آنگاه $\beta_j = \beta_j^i$ در نظر گرفته می‌شود. به عبارت دیگر در این داده‌ها متغیر کمکی نداریم. تحلیل مانده‌ها توسط **سوگاساوا (۲۰۲۰)** بیانگر وجود نقاط دور افتاده در حوزه‌های ۱ و ۲ است، بنابراین این داده‌ها برای بررسی هدف شبیه‌سازی مناسب هستند. ضرایب رگرسیونی برآورد شده بر اساس برازش مدل پیشنهادی به این داده‌ها برابر هستند با $\hat{\beta}_1 = 0/985$ ، $\hat{\beta}_2 = 1/160$ ، $\hat{\beta}_3 = 1/203$ و $\hat{\beta}_4 = 0/746$. چون مقدار واقعی متوسط میزان مصرف شیر تازه مشخص نیست، بنابراین برای مقایسه کردن روش پیشنهادی با روش‌های استوار ارائه شده توسط **سوگاساوا (۲۰۲۰)** از برآورد تقریبی متوسط واریانس پسینی برای بیز تجربی (**رائو، ۲۰۰۳**) یعنی $APV = \frac{1}{m} \sum_{i=1}^m \frac{Var(v_i)\psi_i}{Var(v_i)+\psi_i}$ استفاده می‌کنیم. در روش $R\alpha$ SEB مقدار $Var(v_i)$ را برابر با $\hat{\gamma}$ می‌گیریم. چون شاخص پایداری تقریباً ۲ برآورد شده و در این حالت توزیع پایدار همان توزیع نرمال است، این فرض درست است. مقدار متوسط این واریانس برای روش‌های $R\alpha$ SEB، EB\1RDPD و EB\2RDPD به ترتیب برابر است با ۰/۰۱۵۸، ۰/۰۲۰۶ و ۰/۰۲۰۷، که نشان دهنده بهتر بودن روش پیشنهادی است. در شکل ۲ برآورد خط‌های رگرسیونی برای سه روش مورد مقایسه در حوزه‌های مختلف رسم شده است.



شکل ۲. برآورد خط‌های رگرسیونی برای سه روش مورد مقایسه یعنی: $R\alpha SEB$ ، $EB \setminus RDPD$ و $EB \setminus RDPD$ در حوزه‌های مختلف.

جدول ۴. مقدارهای برآورد شده برای متوسط مصرف شیر تازه برای هر کوچک ناحیه بر اساس روش پیشنهادی $R\alpha SEB$.

شماره ناحیه	برآورد	شماره ناحیه	برآورد	شماره ناحیه	برآورد
۱	۱/۰۵۳	۱۶	۱/۱۷۰	۳۱	۰/۷۷۲
۲	۱/۰۶۳	۱۷	۱/۲۲۵	۳۲	۰/۷۹۴
۳	۱/۰۸۸	۱۸	۱/۲۶۷	۳۳	۰/۷۷۷
۴	۰/۷۱۰	۱۹	۱/۲۳۱	۳۴	۰/۶۲۴
۵	۰/۸۱۲	۲۰	۱/۲۳۰	۳۵	۰/۷۱۴
۶	۰/۹۸۳	۲۱	۱/۱۱۸	۳۶	۰/۷۶۳
۷	۱/۱۱۹	۲۲	۱/۱۹۸	۳۷	۰/۵۶۲
۸	۱/۱۰۴	۲۳	۱/۱۴۴	۳۸	۰/۷۵۲
۹	۱/۳۴۷	۲۴	۱/۲۲۱	۳۹	۰/۷۵۹
۱۰	۱/۳۰۲	۲۵	۱/۱۹۹	۴۰	۰/۷۷۳
۱۱	۰/۶۷۵	۲۶	۰/۷۶۸	۴۱	۰/۷۵۲
۱۲	۱/۳۶۳	۲۷	۰/۷۶۹	۴۲	۰/۸۰۲
۱۳	۱/۳۰۲	۲۸	۰/۷۴۸	۴۳	۰/۷۰۱
۱۴	۰/۹۱۲	۲۹	۰/۷۷۳		
۱۵	۱/۱۹۳	۳۰	۰/۶۳۵		

بیشترین تفاوت بین ضرایب رگرسیونی مربوط به حوزه دوم است، که می‌تواند ناشی از پراکندگی بیشتر مشاهدات در این حوزه و وجود داده دورافتاده باشد، که مدل پیشنهادی با توجه به نتایج شبیه‌سازی دوم بهتر این وضعیت را کنترل می‌کند. همچنین جدول ۴ مقادیرهای متوسط مصرف برآورد شده برای شیر تازه در هر کوچک ناحیه بر اساس روش $R\alpha SEB$ آمده است.

بحث و نتیجه‌گیری

در این مقاله برای اولین بار رهیافت بیز تجربی در برآورد کوچک ناحیه بر اساس توزیع α -پایدار متقارن برای مولفه‌های خطا معرفی و روش برآورد پارامترها بر اساس روش‌های عددی بیان شد. تحلیل‌های انجام شده بر اساس مطالعه‌های شبیه‌سازی و تحلیل داده‌های واقعی برتری مدل پیشنهادی را بر مدل‌های مورد مقایسه نشان داد. از آنجا که توزیع α -پایدار متقارن تعمیم توزیع نرمال است و همزمان دم کلفتی و چولگی موجود در داده‌ها را می‌تواند مدل کند، مدل پیشنهادی حالت‌هایی که توزیع مولفه‌ها نرمال هستند (یا حتی اندکی از توزیع نرمال تخطی دارند) را هم به خوبی پوشش می‌دهد. مدل α -پایدار استوار فی-هریوت پیشنهاد شده به علت برآورد پارامترها به روش‌های عددی در مقایسه با مدل‌های مورد مقایسه، اندکی زمان بیشتری برای اجرا لازم دارد. علاقمندان با درخواست مستقیم به نویسنده مقاله می‌توانند کدهای مورد استفاده در شبیه‌سازی‌ها را در اختیار قرار گیرند.

به عنوان اهداف آینده می‌توان مدل را به حالتی که خطاهای مدل وابسته هستند یا حالتی که واریانس‌های نمونه‌گیری مجهول هستند، توسعه داد. همچنین می‌توان از رهیافت بیز سلسله مراتبی به جای بیز تجربی در مدل پیشنهادی استفاده کرد.

تقدیر و تشکر

نویسندگان مقاله از پیشنهادها و نظرهای ارزشمند داوران محترم که باعث اصلاحات سازنده در محتوا و افزایش کیفیت مقاله شده است، کمال تشکر و قدردانی را دارند. این مقاله با شماره قرارداد ۱۸۶۳/۴ تحت حمایت معاونت محترم پژوهشی دانشگاه کردستان بوده است.

مراجع

- تیموری، م. (۱۳۹۹)، برآوردگر ماکسیمم درست‌نمایی توزیع آلفا-پایدار، مجله علوم آماری، ۱۴، ۹۴-۷۳.
- رائو، جی. ان. کی. (۲۰۰۳)، روش‌های برآورد کوچک ناحیه‌ای، ترجمه محمدرضا مشکانی، تهران: پژوهشکده آمار.
- زارعی، ش.، گرامی، ع. و جعفری، خ. م. (۱۳۸۶)، مقایسه برآورد ناحیه کوچک متوسط درآمد خانوار در برخی استان‌های کشور با روش بیز سلسله مراتبی، پژوهش‌های آماری ایران، ۴، ۹۰-۷۱.
- Arora, V. and Lahiri, P. (1997), On the Superiority of the Bayesian Method over the BLUP in Small Area Estimation Problems, *Statistica Sinica*, 1053-1063.
- Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998), Robust and Efficient Estimation by Minimizing a Density Power Divergence, *Biometrika*, **85**, 549-559.
- Bell, W. R. and Huang, E. T. (2006), Using the t -Distribution to Deal With Outliers in Small Area Estimation, *In Proceedings of Statistics Canada Symposium*.
- Chakraborty, A., Datta, G. S. and Mandal, A. (2018), Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model, *International Statistical Review*, **87**, 158-176.
- Fay, M. P. and Herriot, R. A. (1979), Estimates of Income for Small Places: an Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74**, 269-277.
- Ghosh, M., Maiti, T. and Roy, A. (2008), Influence Functions and Robust Bayes and Empirical Bayes Small Area Estimation, *Biometrika*, **95**, 573-585.

- Lévy, P. (1925), *Calcul des Probabilités*, Gauthier-Villars, Paris.
- Molina, I. and Rao, J. N. K. (2015), *Small Area Estimation*, John Wiley & Sons.
- Nolan, J. P. (2016), Stable Distributions: Models for Heavy-Tailed Data, *Birkhauser*, Boston, Unfinished Manuscript, Chapter 1 Online at academic2.american.edu/jpnolan.
- Ritter, G. (2015), Robust Cluster Analysis and Variable Selection, *Vol. 137 of Chapman & Hall/CRC Monographs on Statistics & Applied Probability*, CRC Press.
- Samorodnitsky, G. and Taqqu, M. S. (1994), *Stable non-Gaussian Random Processes*, Chapman and Hall, New York.
- Sinha, S. K. and Rao, J. N. K. (2009), Robust Small Area Estimation, *The Canadian Journal of Statistics*, **37**, 381-399.
- Sugasawa, S. (2020), Robust Empirical Bayes Small Area Estimation with Density Power Divergence, *Biometrika*, **107**, 467-480.
- Tang, X., Ghosh, M., Ha, N. S. and Sedransk, J. (2018), Modeling Random Effects Using Global-Local Shrinkage Priors in Small Area Estimation, *Journal of the American Statistical Association*, **113**, 1476-1489.
- You, Y. and Chapman, B. (2006), Small Area Estimation Using Area level Models and Estimated Sampling Variances, *Survey Methodology*, **32**, 97-103.
- Zarei, S., Arima, S. and Giovanna, J. L. (2021), A New Robust Bayesian Small Area Estimation Via α -Stable Model for Estimating the Proportion of Athletic Students in California, *Biometrical Journal (In press)*.

Journal of Statistical Sciences, Autumn and Winter, 2021
Vol. 15, No. 2, pp 463-480
DOI: 10.29252/jss.15.2.463

Robust Empirical Bayes Small Area Estimation with Symmetric α -Stable Distribution for Error Components

Zarei, S.

Department of Statistics, University of Kurdistan, Sanandaj , Iran.

Abstract: The most widely used model in small area estimation is the area level or the Fay-Herriot model. In this model, it is typically assumed that both the area level random effects (model errors) and the sampling errors have a Gaussian distribution. However, considerable variations in error components (model errors and sampling errors) can cause poor performance in small area estimation. In this paper, to overcome this problem, the symmetric α -stable distribution is used to deal with outliers in the error components. The model parameters are estimated with the empirical Bayes method. The performance of the proposed model is investigated in different simulation scenarios and compared with the existing classic and robust empirical Bayes methods. The proposed model can improve estimation results, in particular when both error components are normal or have heavy-tailed distribution.

Keywords: Small Area Estimation, Area Level Model, Empirical Bayes, α -Stable Distribution.

Mathematics Subject Classification (2010): 62F35, 60E07, 62D05.