



ORIGINAL RESEARCH PAPER

Co-robust modeling of deferred loss storage data of body insurance and third party automobile insurance of an Iranian insurance company: a Bayesian method

M. Goudarzi<sup>1,\*</sup>, M. Zokaei<sup>2</sup>

<sup>1</sup> Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

<sup>2</sup> Department of Bimometry, School of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

ARTICLE INFO

Article History

Received: 21 April 2017

Revised: 31 May 2017

Accepted: 03 January 2018

Keywords

*Deferred Loss Storage; Scale-Mixed Distributions; Bayesian Inference; Case-by-Case Elimination; Colback-Liebler Divergence.*

ABSTRACT

Saving deferred losses is one of the most fundamental issues in general insurance. In this article, a generalized Bayesian method is used to model the storage of deferred loss data in the fields of body insurance and third party automobile insurance of an Iranian insurance company using Student's t and Pearson's 7th type bivariate distributions. When the data do not follow the assumption of normality, heavy-tailed distributions such as Student's t and Pearson type 7 lead to more robust inferences. These distributions belong to the category of mixed-scale normal distributions. The hierarchical structure of this category allows parameter estimation to be done easily using Markovian chain Monte Carlo methods in the Bayesian framework. For mean sampling distributions, three models of analysis of variance, analysis of covariance, and random walk are considered. In addition, to identify effective samples, a sensitivity analysis study has been conducted based on the Kolbeck-Leibler divergence in the models. The results show that the random walk model with the bivariate Student's t distribution has a better performance for damage payments.

\*Corresponding Author:

Email: [m\\_goudarzi@sbu.ac.ir](mailto:m_goudarzi@sbu.ac.ir)

DOI: [10.22056/ijir.2018.01.05](https://doi.org/10.22056/ijir.2018.01.05)



## مدل‌بندی توأم استوار داده‌های ذخیره‌سازی خسارات معوق رشته‌های بیمه بدنه و بیمه شخص ثالث اتومبیل یک شرکت بیمه ایرانی: یک روش بیزی

منیر گودرزی<sup>۱\*</sup>، محمد ذکایی<sup>۲</sup>

<sup>۱</sup>گروه آمار، دانشکده علوم ریاضی، دانشگاه شهید بهشتی، تهران، ایران  
<sup>۲</sup>گروه بیمه‌سنجی، دانشکده علوم ریاضی، دانشگاه شهید بهشتی، تهران، ایران

### چکیده:

ذخیره‌سازی خسارات معوق یکی از اساسی‌ترین مسائل در بیمه عمومی است. در این مقاله، یک روش بیزی تعمیم‌یافته برای مدل‌بندی توأم داده‌های ذخیره‌سازی خسارات معوق رشته‌های بیمه بدنه و بیمه شخص ثالث اتومبیل یک شرکت بیمه ایرانی با استفاده از توزیع‌های  $t$  ی استیودنت و پی‌یرسون نوع هفتم دومتغیره به کار گرفته می‌شود. هنگامی که داده‌ها از فرض نرمال بودن پیروی نمی‌کنند، توزیع‌های  $t$  ی استیودنت و پی‌یرسون نوع هفتم به استنباط‌های استوارتری منجر می‌شوند. این توزیع‌ها به رده توزیع‌های آمیخته-مقیاس نرمال تعلق دارند. ساختار سلسله‌مراتبی این رده سبب می‌شود که در چارچوب بیزی، برآورد پارامترها به‌سادگی با استفاده از روش‌های مونت کارلوی زنجیر مارکوفی انجام شود. برای میانگین توزیع‌های نمونه‌گیری، سه مدل آنالیز واریانس، آنالیز کوواریانس، و قدم‌زدن تصادفی در نظر گرفته می‌شود. به‌علاوه، برای شناسایی نمونه‌های مؤثر، یک مطالعه آنالیز حساسیت بر اساس واگرایی کولبک-لیبلر در مدل‌ها انجام شده است. نتایج نشان می‌دهد که مدل قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دومتغیره برای پرداخت‌های خسارت، عملکرد بهتری دارد.

### اطلاعات مقاله

تاریخ دریافت: ۰۱ اردیبهشت ۱۳۹۶  
تاریخ داوری: ۱۰ خرداد ۱۳۹۶  
تاریخ پذیرش: ۱۳ دی ۱۳۹۶

### کلمات کلیدی

ذخیره‌سازی خسارات معوق  
توزیع‌های آمیخته-مقیاس  
استنباط بیزی  
حذف موردی  
واگرایی کولبک-لیبلر

\*نویسنده مسئول:

ایمیل: [m\\_goudarzi@sbu.ac.ir](mailto:m_goudarzi@sbu.ac.ir)

DOI: 10.22056/ijir.2018.01.05

یک بیمه‌نامه قراردادی است بین دوطرف، بیمه‌گر (شرکت بیمه) و بیمه‌گذار. در این قرارداد بیمه‌گر در قبال مبلغی ثابت (حق بیمه) متعهد می‌شود که در صورت بروز خسارت برای موضوع قرارداد، خسارت وارده را جبران و یا وجه معینی را بپردازد. در صورتی که یک رویداد (خسارت) اتفاق افتد، بیمه‌گذار یک ادعا به بیمه‌گر ارائه می‌کند. وجهی را که بیمه‌گر باید بپردازد، مقدار خسارت می‌نامند. در طول یک سال مالی، شرکت بیمه نمی‌تواند برخی از خسارات را بلافاصله تسویه کند، بنابراین باید برای پرداخت چنین خساراتی، ذخیره‌ای تحت عنوان ذخیره خسارت معوق<sup>۱</sup> در نظر بگیرد، این امر به دلایل زیر است (Wüthrich and Merz, 2008):

۱. معمولاً در گزارش خسارتهای تأخیر وجود دارد (فاصله بین وقوع خسارت و گزارش به شرکت بیمه طولانی است)؛ بنابراین گزارش یک خسارت ممکن است چند سال طول بکشد. برخی از دلایل تأخیر در گزارش عبارتند از:
    - خسارت در روزهای پایانی اتفاق افتاده و بیمه‌گذار فرصت نکرده است به بیمه‌گر اعلام کند؛
    - بیمه‌گذار از تحت پوشش بیمه‌ای بودن خسارت اطمینان ندارد؛
    - بیمه‌گذار از تحقق خطر منجر به خسارت بی‌خبر است؛
    - بیمه‌گذار قادر نیست که به بیمه‌گر اطلاع بدهد.
  ۲. بعد از گزارش ممکن است چند سال طول بکشد تا یک خسارت به طور کامل تسویه شود.
  ۳. همچنین خسارتی که تسویه شده است (بسته شده است) ممکن است به هر دلیلی دوباره مطرح شود.
- بدهی بابت خسارت معوق که توسط هر یک از مدیریت‌های مختلف بیمه‌ای (باربری، آتش‌سوزی، اتومبیل، مهندسی، مسئولیت، و ...) در شرکت‌های بیمه برآورد و برای انعکاس در حساب‌های شرکت اعلام می‌شود، شامل موارد زیر است (بازنگری آیین‌نامه ذخایر فنی مؤسسات بیمه ارائه پیشنهاد‌های اصلاحی، ۱۳۹۵):

- خسارتهای واقع‌شده‌ای که گزارش نشده‌اند (IBNR)<sup>۲</sup>؛
  - خسارت گزارش‌شده‌ای که پرداخت نشده است (RBNS)<sup>۳</sup>؛
  - خسارتی که به طور کامل گزارش نشده است (IBNER)<sup>۴</sup>؛
- هزینه تسویه خسارت.

محاسبه ذخایر دارای اهمیت زیادی است زیرا کم‌برآوردی آن سبب می‌شود که شرکت بیمه نتواند تعهدات خود را انجام دهد و بیش‌برآوردی آن نیز سبب می‌شود که به صورت غیرلازم سرمایه اضافی نگه داشته شود. لذا پیش‌بینی خسارات معوق و ذخیره مناسب برای پرداخت چنین ادعاهایی نقش مهمی در توانایی مالی و استمرار کسب‌وکار شرکت‌های بیمه‌ای دارد؛ بنابراین پیش‌بینی مطالبات با روش‌های مناسب و برآورد خطای پیش‌بینی، هدف ذخیره‌سازی است.

بردار تصادفی  $\mathbf{X}_{ij} = (X_{ij}^{(1)}, X_{ij}^{(2)})' \in R^2$  را که در آن  $X_{ij}^{(l)}$  نشان‌دهنده میزان پرداخت‌های خسارات غیرتجمعی<sup>۵</sup> به‌وسیله یک شرکت بیمه برای سال مبدأ ۶ (سال یا دوره صدور بیمه‌نامه)  $i \in \{1, \dots, I\}$  و سال تأخیر<sup>۶</sup> (تعداد سالها یا فصلهایی که تا پرداخت خسارت، تأخیر به وجود می‌آید)  $j \in \{1, \dots, J\}$  در  $l \in \{1, 2\}$  امین رشته بیمه است، در نظر بگیرید. داده‌های مربوط به پرداخت‌های خسارت در هر رشته بیمه همان‌گونه که در جدول ۱ دیده می‌شود، معمولاً به صورت یک مثلث تأخیر<sup>۸</sup> نشان داده می‌شوند.

---

1. Outstanding Loss Reserves  
 2. Incurred But Not Reported  
 3. Reported But Not Settled  
 4. Incurred But Not Enough Reported  
 5. Incremental Paid Losses  
 6. Accident Year  
 7. Development Year  
 8. Run-Off Triangle

جدول ۱: مثلث تأخیر

آسال مبدأ	آسال تأخیر				
	۱	۲	...	z	J
۱					
۲					
⋮					
z-۱					
⋮					
۱					

(پرداختهای خسارت)  $X_{ij}^{(t)}$  مقادیر

که پیش‌بینی می‌شوند.  $X_{ij}^{(t)}$  مقادیر

مدلهای آماری متعددی برای پیش‌بینی میزان خسارات معوق در مثلث پایینی در جدول ۱ وجود دارد؛ برای مرور روشهای ذخیره‌سازی در یک رشته بیمه می‌توان به تیلور<sup>۱</sup> (۲۰۰۰)، انگلند و ورال<sup>۲</sup> (۲۰۰۲ و ۲۰۰۶)، و وترپچ و مرز<sup>۳</sup> (۲۰۰۸) مراجعه کرد. مدل‌بندی توأم مثلثهای تأخیر مرتبط با رشته‌های مختلف بیمه‌ای به منظور در نظر گرفتن ساختارهای وابستگی بین آنها در محاسبه ذخایر، یک مسئله جدید در نوشتگان ذخیره‌سازی تصادفی است، برای نمونه می‌توان براون<sup>۴</sup> (۲۰۰۴)، هس<sup>۵</sup> و همکاران (۲۰۰۶)، اشمیت<sup>۶</sup> (۲۰۰۶)، مرز و وترپچ (۲۰۰۹a و ۲۰۰۹b)، ژانگ<sup>۷</sup> (۲۰۱۰)، شی و فریز<sup>۸</sup> (۲۰۱۱)، سالزمن<sup>۹</sup> و وترپچ (۲۰۱۲)، ژانگ و همکاران (۲۰۱۲)، دی جونگ<sup>۱۰</sup> (۲۰۱۲)، شی و همکاران (۲۰۱۲)، هپ<sup>۱۱</sup> و وترپچ (۲۰۱۳)، مرز و همکاران (۲۰۱۳) و شی (۲۰۱۴) را نام برد. شی و همکاران (۲۰۱۲) یک مدل لگ نرمال بیزی را در پیش‌بینی خسارات معوق برای رشته‌های وابسته بیمه‌ای پیشنهاد کردند. در این مقاله، روش شی و همکاران (۲۰۱۲) را به حالتی که خسارات از توزیعهای  $t$  ی استیودنت و پی‌یرسون نوع هفتم<sup>۱۲</sup> دومتغیره پیروی می‌کنند، تعمیم می‌دهیم و بر این اساس و در یک چارچوب بیزی به مدل‌بندی توأم داده‌های ذخیره‌سازی خسارات بیمه بدنه و بیمه شخص ثالث اتومبیل یک شرکت بیمه ایرانی می‌پردازیم. توزیعهای  $t$  ی استیودنت و پی‌یرسون نوع هفتم، دُمهای سنگین‌تری از توزیع نرمال دارند و لذا اغلب به استنباطهای استوآتری<sup>۱۳</sup> منجر می‌شوند. همچنین برای تابع میانگین، مدل‌های آنالیز واریانس (ANOVA<sup>۱۴</sup>)، آنالیز کوواریانس (ANCOVA<sup>۱۵</sup>)، و قدم‌زدن تصادفی<sup>۱۶</sup> را در نظر می‌گیریم.

1. Taylor  
 2. England and Verrall  
 3. Wüthrich and Merz  
 4. Braun  
 5. Hess  
 6. Schmidt  
 7. Zhang  
 8. Shi and Frees  
 9. Salzmann  
 10. De Jong  
 11. Happ  
 12. Pearson Type VII  
 13. Robust  
 14. Analysis of Variance  
 15. Analysis of Covariance  
 16. Random Walk

بعد از برازش مدل، مطالعات حساسیت برای کشف مشاهدات مؤثر، یعنی مشاهداتی که حذف آنها منجر به تغییرات اساسی در برآوردهای پارامترها یا تابعی از پارامترها در یک تحلیل آماری می‌شود، مورد علاقه بسیاری از محققان است. یک روش شناخته‌شده برای شناسایی نمونه‌های مؤثر روش حذف موردی معرفی‌شده توسط کوک<sup>۱</sup> (۱۹۸۶) است. این روش با موفقیت برای مدل‌های آماری مختلفی به کار گرفته شده است. در چارچوب بیزی، چو<sup>۲</sup> و همکاران (۲۰۰۹) تشخیص‌های تأثیر حذف موردی برای مدل‌های بقا را از طریق روش بیزی توسعه دادند. روش تشخیص بیزی برای مدل‌بندی توأم ذخیره‌سازی خسارات معوق تاکنون در تحقیقات قبلی مورد توجه قرار نگرفته است؛ بنابراین یکی دیگر از اهداف این مقاله، مطالعه اندازه‌ای تشخیصی برای شناسایی نمونه‌های مؤثر بر اساس واگرایی کولبک-لیبلر<sup>۳</sup> (K-L) پیشنهادشده توسط چو و همکاران (۲۰۰۹) در مدل‌بندی توأم خسارات در مسئله ذخیره‌سازی بیمه‌های غیرعمر است. معیارهای بیزی به‌آسانی با نرم‌افزاری استاندارد بیزی مانند OpenBUGS محاسبه می‌شوند.

ادامه مقاله به صورت زیر سازمان‌دهی شده است: در بخش ۲، داده‌های خسارات رشته‌های بیمه بدنه و بیمه شخص ثالث اتومبیل یک شرکت بیمه ایرانی توصیف شده‌اند. در بخش ۳، مدل آماری و استنباط بیزی ارائه شده است. روش تشخیصی تأثیر نمونه‌ها به روش بیزی در بخش ۴ بررسی شده است. در بخش ۵ به تحلیل داده‌ها و بیان نتایج پرداخته و در پایان نتیجه‌گیری بیان شده است.

#### توصیف داده‌ها

داده‌های مثلث‌های تأخیر به‌کاررفته در این مقاله، پرداخت‌های خسارات و حق‌بیمه‌های عایدشده برای دو رشته بیمه بدنه و بیمه شخص ثالث اتومبیل در سالهای ۱۳۹۲-۱۳۹۵ یک شرکت بیمه ایرانی است، که به صورت فصلی ثبت شده‌اند. ضریب همبستگی پی‌یرسون بین میزان پرداخت‌های خسارت سلولهای متناظر در مثلث‌های تأخیر این دو رشته بیمه، ۰/۶۷ است که بیانگر وجود همبستگی بین آنهاست و با توجه به صفربودن پی-مقدار، همبستگی میان دو مثلث در سطح معنی‌داری ۰/۰۱ پذیرفته می‌شود. برای هر یک از دو رشته بیمه ( $l=1,2$ ),

پرداخت‌های خسارت را با تقسیم بر حق‌بیمه‌های عایدشده سال وقوع متناظر آن نرمال می‌کنیم، یعنی  $Y_{ij}^{(l)}$  را به صورت

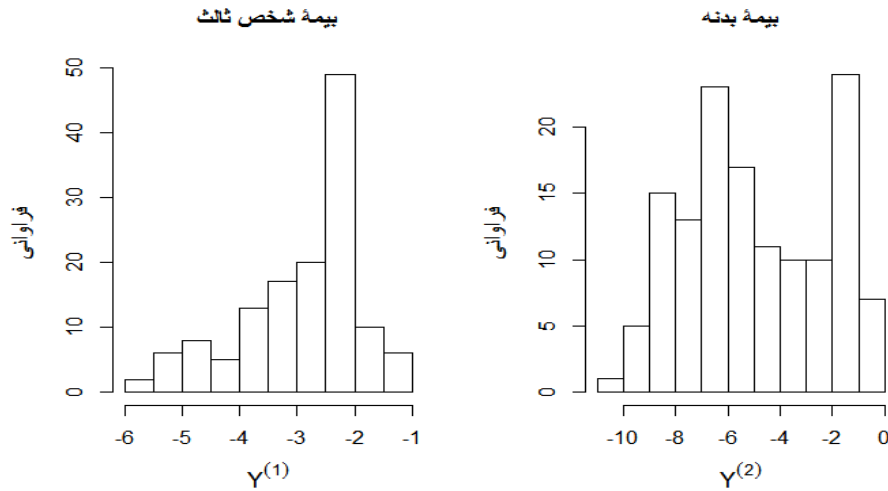
$$Y_{ij}^{(l)} = \log\left(\frac{X_{ij}^{(l)}}{P_i^{(l)}}\right), \quad i, j=1, \dots, 16,$$

تعریف می‌کنیم، که در آن  $P_i^{(l)}$  حق‌بیمه عایدشده مربوط به فصل صدور  $i$  ام و مثلث رشته  $l$  ام است (Côté and et al., 2016). مقادیر لگاریتم پرداخت‌های خسارات نرمال‌شده در دو مثلث تأخیر مربوط به رشته‌های بیمه بدنه و بیمه شخص ثالث اتومبیل در جدولهای ۳ و ۴ نشان داده شده‌اند. شکل‌های ۲ و ۳ به ترتیب نمودارهای بافت‌نگار و سریهای زمانی لگاریتم پرداخت‌های خسارات نرمال‌شده دو رشته بیمه را به صورت جداگانه نشان می‌دهد.

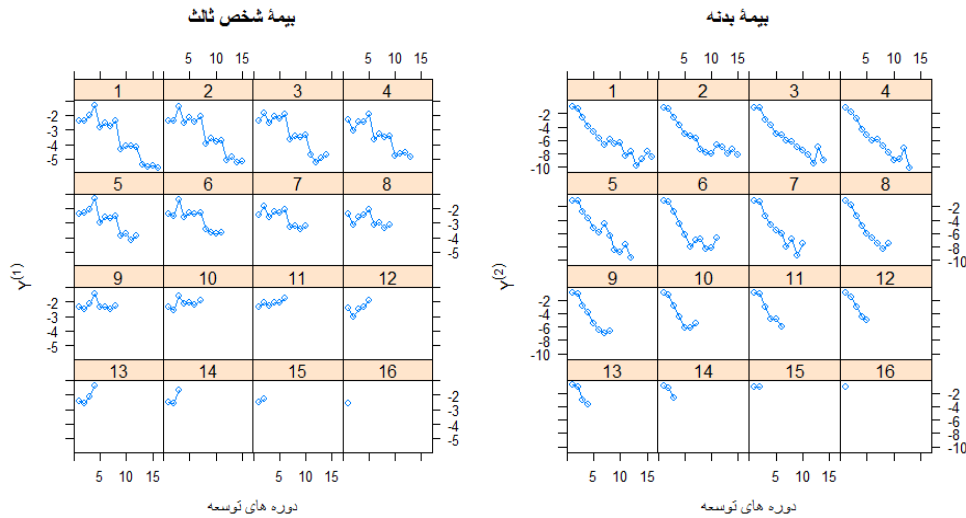
<sup>1</sup>. Cook

<sup>2</sup>. Cho

<sup>3</sup>. Kullback-Leibler Divergence



شکل ۲: نمودارهای بافت‌نگار لگاریتم پرداخت‌های خسارات نرمال‌شده دو مثلث تأخیر



شکل ۳: نمودارهای سری‌های زمانی لگاریتم پرداخت‌های خسارات نرمال‌شده دو مثلث تأخیر

مدل آماری و استنباط بیزی

توزیع پی‌یرسون نوع هفتم چندمتغیره

توزیع پی‌یرسون نوع هفتم چندمتغیره (Pearson, 1916)، به صورت

$$f(y; \mu, \Lambda, m) = \frac{\Gamma(m)}{\pi^{\frac{d}{2}} \Gamma(m - \frac{d}{2})} |\Lambda|^{-\frac{1}{2}} [1 + \Delta^2]^{-m},$$

تعریف می‌شود، که در آن  $\mu$  بردار میانگین،  $|\Lambda|$  نشان‌دهنده دترمینان ماتریس کوواریانس  $\Lambda$ ،  $\Delta^2 = (\mathbf{y} - \mu)^T \Lambda^{-1} (\mathbf{y} - \mu)$  یک فاصله ماهالانوبیس،  $m$  درجه آزادی و  $d$  بعد  $\mathbf{y}$  است. این توزیع به طور گسترده در مدل‌بندی بسیاری از پدیده‌ها مورد استفاده قرار می‌گیرد.

فرناندز و استیل<sup>۱</sup> (۲۰۰۰)، نشان دادند که توزیع پی‌یرسون نوع هفتم به رده توزیعهای آمیخته-مقیاس نرمال چندمتغیره تعلق دارد. بر این اساس، یک نمایش جایگزین سلسله مراتبی دو مرحله‌ای از توزیع  $\mathbf{y}$  به صورت

$$\mathbf{Y} | \Lambda = \lambda \sim N_L(\boldsymbol{\mu}, \lambda^{-1}\boldsymbol{\Sigma}), \Lambda \sim \text{gamma}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right),$$

است، که در آن  $N_L(\cdot, \cdot)$  نشان‌دهنده توزیع نرمال چندمتغیره،  $\Lambda$  یک متغیر تصادفی آمیزنده با تابع چگالی گاما و پارامترهای  $\mathbf{v} = (\nu_1, \nu_2)$  است. هنگامی که  $\nu_1 = \nu_2$  باشد، دارای توزیع  $\mathbf{t}$  ی استیودنت چندمتغیره با درجه آزادی  $\nu_1$  خواهد بود.

مدل بیزی استوار توأم داده‌های ذخیره‌سازی خسارات معوق

فرض می‌کنیم که  $Y_{ij}^{(l)}$  با توزیع پی‌یرسون نوع هفتم یا  $\mathbf{t}$  ی استیودنت دومتغیره مدل‌بندی شوند؛ بنابراین بردار تصادفی  $\mathbf{Y}_{ij}$  را بر اساس نمایش سلسله مراتبی به صورت

$$\mathbf{Y}_{ij} | \Lambda_{ij} = \lambda_{ij} \sim N_L(\boldsymbol{\mu}_{ij}, \lambda_{ij}^{-1}\boldsymbol{\Sigma}), \Lambda_{ij} \sim \text{gamma}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right), \quad (1)$$

می‌توان در نظر گرفت، که در آن  $\mathbf{y}_{ij}$ ،  $\boldsymbol{\mu}_{ij}$  و  $\boldsymbol{\Sigma}$  به صورت

$$\mathbf{y}_{ij} = \begin{bmatrix} y_{ij}^{(1)} \\ \vdots \\ y_{ij}^{(L)} \end{bmatrix}, \quad \boldsymbol{\mu}_{ij} = \begin{bmatrix} \mu_{ij}^{(1)} \\ \vdots \\ \mu_{ij}^{(L)} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^{(1,1)} & \dots & \sigma^{(1,L)} \\ \vdots & \ddots & \vdots \\ \sigma^{(L,1)} & \dots & \sigma^{(L,L)} \end{bmatrix}$$

تعریف می‌شوند. ماتریس کوواریانس  $\boldsymbol{\Sigma}$  بیانگر وابستگی دوبه‌دوی سولوها از یک مثلث و سلول متناظر از مثلث دیگر است. برای  $l$  امین مثلث تأخیر، مدل‌های زیر را برای میانگین در نظر می‌گیریم:

۱. مدل ANOVA: مدل ANOVA برای  $i \in \{1, \dots, I\}$  و  $j = 1, \dots, J - i + 1$ ، به صورت (Verral, 1991)

$$\mu_{ij}^{(l)} = \mu^{(l)} + \alpha_i^{(l)} + \beta_j^{(l)}, \quad (2)$$

است و در شرط‌های  $\sum_{i=1}^I \alpha_i^{(l)} = \sum_{j=1}^J \beta_j^{(l)} = 0$  صدق می‌کند. در رابطه (۲) اثرهای ثابت  $\alpha_i^{(l)}$  و  $\beta_j^{(l)}$  به ترتیب پارامترهای فصل‌های مبدأ و تأخیر را نشان می‌دهند.

۲. مدل ANCOVA: در مدل ANCOVA، فرض می‌کنیم که اثرهای فصل مبدأ خطی هستند، بنابراین تابع میانگین (۲) به صورت

$$\mu_{ij}^{(l)} = \mu^{(l)} + i\alpha^{(l)} + \beta_j^{(l)},$$

تغییر می‌کند و در شرط  $\sum_{j=1}^J \beta_j^{(l)} = 0$  صدق می‌کند. در این مدل فرض می‌کنیم اثرهای فصل مبدأ، تصادفی و اثرهای فصل تأخیر ثابت هستند.

۳. مدل قدم‌زدن تصادفی: با استفاده از رابطه (۲) و

$$\alpha_i^{(l)} = \alpha_{i-1}^{(l)} + h_i^{(l)}, \quad \beta_j^{(l)} = \beta_{j-1}^{(l)} + u_j^{(l)}$$

یک مدل قدم‌زدن تصادفی را برای تابع میانگین در نظر می‌گیریم که برای  $i = 2, \dots, I$  و  $j = 2, \dots, J - i + 1$  در شرط‌های  $\alpha_1^{(l)} = \beta_1^{(l)} = 0$  صدق می‌کنند. در این مدل فرض می‌کنیم اثرهای فصل مبدأ و تأخیر هر دو تصادفی هستند.

<sup>1</sup>. Fernandez and Steel

توزیع‌های پیشینی و پسینی

برای کامل کردن توصیف بیزی مدلها، مشخص کردن توزیع پیشینی برای همه پارامترهای نامعلوم ضروری است. به دلیل آنکه هیچ اطلاعاتی درباره پارامترهای مدلها پیشنهاد شده وجود ندارد، توزیعهای پیشینی مزدوج کم‌آگاهنده<sup>1</sup> نسبت داده می‌شود؛ بنابراین، بر اساس نوشتگان موضوع برای پارامترهای مدل ANOVA، پیشین‌های

$$\mathbf{a}_i = (\alpha_i^{(1)}, \dots, \alpha_i^{(L)})' \sim N_L(\mathbf{0}, \Sigma_{\mathbf{a}_i}), \quad \mathbf{\beta}_j = (\beta_j^{(1)}, \dots, \beta_j^{(L)})' \sim N_L(\mathbf{0}, \Sigma_{\mathbf{\beta}_j}),$$

$$\tau_{\Sigma} = \Sigma^{-1} \sim \text{Wishart}(\mathbf{R}, L),$$

را در نظر می‌گیریم، که در آن  $\Sigma_{\mathbf{\beta}_j} = \text{diag}(\sigma_{\beta_j^{(1)}}^2, \dots, \sigma_{\beta_j^{(L)}}^2)$  و  $\Sigma_{\mathbf{a}_i} = \text{diag}(\sigma_{\alpha_i^{(1)}}^2, \dots, \sigma_{\alpha_i^{(L)}}^2)$  یک ماتریس نیمه‌مثبت با ابعاد  $L \times L$  است. برای مدل ANCOVA پیشین‌های

$$\mathbf{a} = (\alpha^{(1)}, \dots, \alpha^{(L)})' \sim N_L(\mathbf{0}, \Sigma_{\mathbf{a}}), \quad \tau_{\alpha} = \sigma_{\alpha}^{-2} \sim \text{Gamma}(\kappa_1, \nu_1),$$

$$\mathbf{\beta}_j = (\beta_j^{(1)}, \dots, \beta_j^{(L)})' \sim N_L(\mathbf{0}, \Sigma_{\mathbf{\beta}_j}), \quad \tau_{\Sigma} = \Sigma^{-1} \sim \text{Wishart}(\mathbf{R}, L),$$

به کار گرفته می‌شود که در آن  $\Sigma_{\mathbf{a}} = \sigma_{\alpha}^2 \mathbf{I}_{L \times L}$  است، و برای مدل قدم‌زدن تصادفی، پیشین‌های

$$\tau_{\Sigma} = \Sigma^{-1} \sim \text{Wishart}(\mathbf{R}, L), \quad h_i^{(l)} \sim N(0, \sigma_h^2), \quad \tau_h = \sigma_h^{-2} \sim \text{Gamma}(\kappa_1, \nu_1),$$

$$u_i^{(l)} \sim N(0, \sigma_u^2), \quad \tau_u = \sigma_u^{-2} \sim \text{Gamma}(\kappa_2, \nu_2),$$

در نظر گرفته می‌شوند. توزیع پسینی توأم پارامترها با استفاده از قضیه بیز، با ترکیب کردن توزیع درست‌نمایی در رابطه (۱) و توزیعهای پیشینی به دست می‌آید. با توجه اینکه توزیعهای پسینی به دست آمده دارای شکل پیچیده‌ای هستند، برای به دست آوردن برآورد بیز پارامترهای مدلها می‌توان از روشهای MCMC مانند نمونه‌گیر گیبز<sup>۲</sup> و الگوریتم متروپولیس-هستینگ<sup>۳</sup> که با نمونه‌گیری مکرر از توزیع شرطی کامل هر پارامتر عمل می‌کنند، استفاده کرد. روشهای MCMC به آسانی با نرم‌افزارهای استاندارد بیزی مانند OpenBUGS محاسبه می‌شوند.

مقایسه مدل‌ها

برای مقایسه مدلها، معیارهای اطلاع انحراف<sup>۴</sup> (DIC) (Spiegelhalter et al., 2002) و لگاریتم شبه درست‌نمایی حاشیه‌ای<sup>۵</sup> (LPML)

را به کار برده‌ایم. فرض کنیم  $\mathbf{y}_{ij} = (y_{ij}^{(1)}, \dots, y_{ij}^{(L)})'$  و  $\boldsymbol{\theta}$  مجموعه همه پارامترهای مدل باشد. معیار انحراف<sup>۶</sup> را که به صورت

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y} | \boldsymbol{\theta}) = -2 \sum_{i=1}^I \sum_{j=1}^{J+1-i} \log f(\mathbf{y}_{ij} | \boldsymbol{\theta}),$$

تعریف می‌شود، در نظر بگیرید. پس  $\bar{D}(\boldsymbol{\theta}) = E(D(\boldsymbol{\theta}))$  معیاری در برازش مدل است و می‌تواند با استفاده از نمونه‌های MCMC، به صورت

$$\bar{D} = -2 \frac{1}{M} \sum_{i=1}^I \sum_{j=1}^{J+1-i} \sum_{m=1}^M \log f(\mathbf{y}_{ij} | \boldsymbol{\theta}^{(m)}),$$

<sup>1</sup>. Conjugate but Weakly Informative Prior

<sup>2</sup>. Gibbs Sampling

<sup>3</sup>. Metropolis Hastings Algorithm

<sup>4</sup>. Deviance Information Criterion

<sup>5</sup>. Log Pseudo Marginal Likelihood

<sup>6</sup>. Deviance



تقریب زده شود، که در آن  $\theta^{(m)}$ ،  $m$  امین تکرار زنجیر MCMC مدل و  $M$  تعداد تکرارهاست. با در نظر گرفتن اندازه انحراف، معیار بیزی DIC به صورت

$$DIC = \bar{D} + \hat{p}_D,$$

تعریف شده است، که در آن  $\hat{p}_D = \bar{D} - \hat{D}$  با  $\hat{D} = D\left(\frac{1}{M} \sum_{m=1}^M \theta^{(m)}\right)$  است. هر اندازه مقدار DIC کوچکتر باشد، مدل بهتر برازش داده شده است. آماره عرض پیشگوی شرطی<sup>۱</sup> (CPO)، معیار مشهور دیگری است که معمولاً برای مقایسه مدلها در متون بیزی به کار گرفته می-شود (برای توضیحات بیشتر درباره این آماره و کاربردهای آن به گلفند<sup>۲</sup> و همکاران (۱۹۹۲) مراجعه کنید). آماره CPO به صورت

$$CPO_{ij} = p(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}) = \int_{\theta \in \Theta} f(\mathbf{y}_{ij} | \theta) \pi(\theta | \mathbf{y}_{(-ij)}) d\theta = \left\{ \int_{\theta \in \Theta} \frac{\pi(\theta | \mathbf{y})}{f(\mathbf{y}_{ij} | \theta)} d\theta \right\}^{-1},$$

تعریف می-شود، که در آن  $\mathbf{y}_{(-ij)}$  نشان دهنده داده‌ها با حذف بردار  $\mathbf{y}_{ij}$  است و  $p(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)})$  توزیع پیشگوی یک مشاهده جدید به شرط  $\mathbf{y}_{(-ij)}$  است و  $\pi(\theta | \mathbf{y})$  توزیع پسینی  $\theta$  است.  $CPO_{ij}$  را می‌توان با استفاده از نمونه‌های MCMC به صورت

$$CPO_{ij} = \left\{ \frac{1}{M} \sum_{m=1}^M \frac{1}{f(\mathbf{y}_{ij} | \theta^{(m)})} \right\}^{-1},$$

تقریب زد (Chen et al., 2000) و بر اساس آن معیار LPML به صورت

$$LPML = \sum_{i=1}^I \sum_{j=1}^{J+1-i} \log(CPO_{ij}),$$

تعریف می‌شود. یک LMPL بزرگتر نشان دهنده عملکرد بهتر مدل است.

مباحث تشخیصی تأثیر موردی بیزی

مباحث تشخیصی تأثیر، یک گام مهم در تحلیل مدل‌های آماری است. فرض کنید  $L(\theta | \mathbf{y})$  و  $L(\theta | \mathbf{y}_{(-ij)})$  به ترتیب نشان دهنده تابع درست‌نمایی بر اساس داده‌های کامل و تابع درست‌نمایی بر اساس داده‌ها با حذف  $ij$  امین نمونه باشند. توزیعهای پسینی برای داده‌های کامل و داده‌ها با حذف  $ij$  امین نمونه می‌تواند به ترتیب به صورت  $\pi(\theta | \mathbf{y}) \propto L(\theta | \mathbf{y})\pi(\theta)$  و  $\pi(\theta | \mathbf{y}_{(-ij)}) \propto L(\theta | \mathbf{y}_{(-ij)})\pi(\theta)$  تعریف شوند، که در آن  $\pi(\theta)$  توزیع پیشینی  $\theta$  است. فرض کنید  $K(P, P_{(ij)})$  نشان دهنده فاصله کولبک-لیبار بین  $P$  و  $P_{(ij)}$  باشد، که در آن  $P$  نشان-دهنده  $\pi(\theta | \mathbf{y})$  و  $P_{(ij)}$  نشان دهنده  $\pi(\theta | \mathbf{y}_{(-ij)})$  است، لذا

$$K(P, P_{(ij)}) = \int \pi(\theta | \mathbf{y}) \log \left\{ \frac{\pi(\theta | \mathbf{y})}{\pi(\theta | \mathbf{y}_{(-ij)})} \right\} d\theta \\ = -\log[E_{\theta | \mathbf{y}}[\{f(\mathbf{y}_{ij} | \theta)\}^{-1}]]^{-1} + E_{\theta | \mathbf{y}}[\log\{f(\mathbf{y}_{ij} | \theta)\}],$$

بنابراین  $K(P, P_{(ij)})$  تأثیر حذف  $ij$  امین نمونه از داده‌های کامل روی توزیع پسینی توأم  $\theta$  را اندازه‌گیری می‌کند. برآورد مونت کارلوی  $K(P, P_{(ij)})$  به صورت

<sup>1</sup>. Conditional Predictive Ordinate

<sup>2</sup>. Gelfand

$$\hat{K}(P, P_{(ij)}) = - \log \left( \left\{ \frac{1}{M} \sum_{m=1}^M \frac{1}{f(\mathbf{y}_{ij} | \boldsymbol{\theta}^{(m)})} \right\}^{-1} \right) + \frac{1}{M} \sum_{m=1}^M \log[f(\mathbf{y}_{ij} | \boldsymbol{\theta}^{(m)})],$$

به دست می‌آید. یک مشاهده با مقدار کولبک-لیبلر بزرگ به‌عنوان یک مشاهده مؤثر در نظر گرفته می‌شود.

#### تحلیل داده‌ها

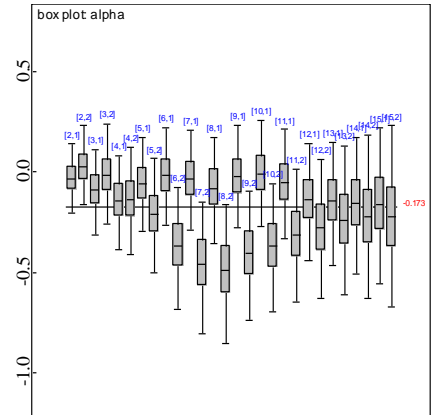
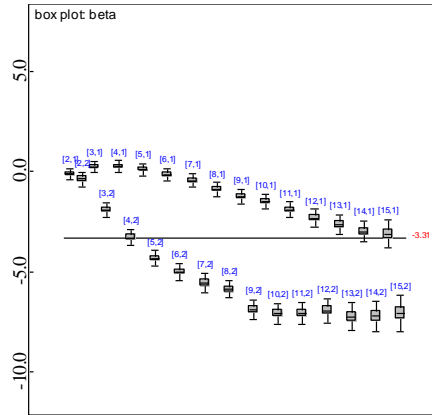
برای انجام مدل‌ها با استفاده از روش شبیه‌سازی بیزی، با به کار بردن نرم‌افزار OpenBUGS، نمونه‌گیر گیبز را برای ۲۵۰۰۰۰ تکرار اجرا کردیم. پس از حذف اولین ۵۰۰۰ تکرار به‌عنوان دوره داغیدن، برای کاهش همبستگی نمونه‌های شبیه‌سازی شده به صورت سیستماتیک هر ۴۰امین تکرار در زنجیر به کار گرفته شد. درنهایت، ۵۰۰۰ نمونه شبیه‌سازی شده خواهیم داشت. برای ارزیابی همگرایی زنجیر مارکف، نمودارهای روند، نمودارهای خودهمبستگی و خط‌های مونت کارلوی پارامترها مورد بررسی قرار گرفت. جدول ۲، مقایسه بین مدل‌های مختلف برازش داده شده را با استفاده از معیارهای انتخاب مدل بحث شده در بخش ۳-۴ نشان می‌دهد. به منظور ارزیابی توانایی پیش‌بینی یک سال جلوتر مدل‌ها، ۱۵ فصل اول برای برازش داده‌ها به کار گرفته می‌شود.

جدول ۲: مقایسه بین مدل‌های برازش داده شده

مدل میانگین	توزیع خسارت	DIC	LMPL
ANOVA	نرمال	۴۰۰/۹	-۲۰۵/۷۸
	پی‌یرسون نوع هفتم	۳۸۷/۴	-۱۹۸/۷۷
	ی استیودنت t	۳۸۱/۴	-۱۹۸/۵۷
ANCOVA	نرمال	۳۸۹/۴	-۲۰۱/۰۳
	پی‌یرسون نوع هفتم	۳۷۴	-۱۹۴/۷۲
	ی استیودنت t	۳۶۹/۸	-۱۹۵/۳۴
قدم‌زدن تصادفی	نرمال	۳۶۷/۶	-۱۸۶/۹۶
	پی‌یرسون نوع هفتم	۳۵۰	-۱۸۱/۳۹
	ی استیودنت t	۳۴۵/۵	-۱۷۹/۶۳

بر طبق معیارها، مدل قدم‌زدن تصادفی با توزیع t ی استیودنت دومتغیره برای لگاریتم پرداخت‌های خسارات نرمال شده، برازش بهتری دارد.

در این مدل برای بیمه شخص ثالث (مثلث ۱) و بیمه بدنه (مثلث ۲)، میانگین‌های پسینی (انحراف‌های استاندارد) میانگین‌های کل  $\mu^{(1)}$ ،  $\mu^{(2)}$  به ترتیب برابر با ۲/۳۲- (۰/۱۳) و ۰/۷۵- (۰/۱۷) است. شکل ۴، نمودار جعبه‌ای نمونه‌های پسینی اثر فصل‌های مبدأ در مدل قدم‌زدن تصادفی با توزیع t ی استیودنت دومتغیره را برای دو مثلث نشان می‌دهد. یک روند غیرثابت در بین برآورد میانگین اثر فصل‌های مبدأ در مثلث بیمه شخص ثالث و یک روند کاهشی تا ۸امین فصل و پس از آن یک روند افزایشی در بین برآورد میانگین اثر فصل‌های مبدأ در مثلث بیمه بدنه مشاهده می‌شود. شکل ۵، نمودار جعبه‌ای نمونه‌های پسینی اثر فصل‌های تأخیر در مدل قدم‌زدن تصادفی با توزیع t ی استیودنت دومتغیره برای دو مثلث را نشان می‌دهد. یک روند کاهشی در برآورد میانگین اثر فصل‌های تأخیر از اولین تا آخرین فصل تأخیر در هر دو مثلث مشاهده می‌شود.



شکل ۵: نمودار جعبه‌ای نمونه‌های پسینی اثر فصل‌های ی استیودنت t تأخیر برای مدل قدم‌زدن تصادفی با توزیع دومتغیره در دو مثلث

شکل ۴: نمودار جعبه‌ای نمونه‌های پسینی اثر فصل‌های ی استیودنت t مبدأ برای مدل قدم‌زدن تصادفی با توزیع دومتغیره در دو مثلث

میزان ادعاهای پیش‌بینی شده مثلث پایینی که در واقع هدف نهایی در مسئله ذخیره‌سازی است در هر دو رشته بیمه بر اساس مدل قدم‌زدن تصادفی با توزیع t ی استیودنت دومتغیره برای لگاریتم پرداخت‌های خسارات نرمال شده در جدولهای ۳ و ۴ نشان داده شده است.

جدول ۳: مثلث تأخیر میزان لگاریتم پرداخت‌های خسارات نرمال شده بیمه شخص ثالث (مثلث بالایی) و میزان لگاریتم پرداخت‌های خسارات نرمال شده پیش‌بینی شده بر اساس مدل قدم‌زدن تصادفی با توزیع t ی استیودنت دومتغیره (مثلث پایینی)

سال خسارت	فصل خسارت	فصل توسعه خسارت														
		۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
۱۳۹۲	بهار	-۲/۳۵	-۲/۳۹	-۱/۹۷	-۱/۳۴	-۲/۸۵	-۲/۵۲	-۲/۷۷	-۲/۳۶	-۴/۳۵	-۴/۰۸	-۴/۱۳	-۴/۱۹	-۵/۴	-۵/۵۸	-۵/۴۷
	تابستان	-۲/۴	-۲/۳۵	-۱/۴۳	-۲/۵۶	-۲/۱۸	-۲/۴۶	-۲/۰۵	-۳/۹۷	-۲/۶۲	-۳/۸	-۳/۷۱	-۵/۰۹	-۴/۸۵	-۵/۲۴	-۵/۴۶
	پاییز	-۲/۳۵	-۱/۸۲	-۲/۵۳	-۲/۱	-۲/۲۵	-۱/۹۳	-۳/۶۷	-۳/۴۷	-۳/۴۹	-۳/۳۸	-۴/۶۹	-۵/۲۱	-۴/۹۵	-۵/۴۳	-۵/۵۳
	زمستان	-۲/۳۲	-۳/۰۵	-۲/۴۵	-۲/۴۲	-۱/۹۳	-۳/۶۳	-۳/۲۶	-۳/۵۱	۳/۴۱	-۴/۸۲	-۴/۶۶	-۴/۵۹	-۵/۱	-۵/۴۸	-۵/۵۹
۱۳۹۳	بهار	-۲/۴	-۲/۳۲	-۲/۱	-۱/۲۸	-۲/۹۴	-۲/۵۷	-۲/۶۳	-۲/۵۳	-۲/۸۴	-۳/۷۵	-۴/۱۷	-۴/۷۱	-۵	-۵/۳۸	-۵/۵۷
	تابستان	-۲/۴	-۲/۴۹	-۱/۳۸	-۲/۶	-۲/۲۶	-۲/۳۷	-۲/۳۲	-۳/۴۱	-۲/۶۷	-۳/۷۲	۴/۲۴-	۴/۶۷-	۴/۹۸-	۵/۳۵-	۵/۴۶-
	پاییز	-۲/۴۴	-۱/۸۳	-۲/۵۹	-۲/۲	-۲/۲۷	-۲/۰۹	-۳/۲۴	-۳/۲۲	-۳/۳۹	۳/۸۵-	۴/۲۶-	۴/۶۸-	۴/۹۹-	۵/۳۶-	۵/۴۸-
	زمستان	-۲/۳۴	-۳/۱۴	-۲/۵۶	-۲/۴۱	-۲/۰۴	-۳/۱۴	-۳	-۳/۳۳	۳/۶۳-	۳/۸۹-	۴/۳۱-	۴/۷۳-	۵/۰۴-	۵/۴۱-	۵/۵۱-
۱۳۹۴	بهار	-۲/۳۲	-۲/۵۲	-۲/۱۳	-۱/۴۴	-۲/۳۶	-۲/۳۶	-۲/۴۹	۳/۲۲-	۳/۵۸-	۳/۸۳-	۴/۲۵-	۴/۶۶-	۴/۹۸-	۵/۳۵-	۵/۴۶-
	تابستان	-۲/۳۴	-۲/۶	-۱/۵۷	-۲/۱۶	-۲/۰۹	-۲/۲۴	۲/۷۷-	۳/۲۱-	۳/۵۷-	۳/۸۲-	۴/۲۳-	۴/۶۵-	۴/۹۷-	۵/۳۵-	۵/۴۴-
	پاییز	-۲/۳۶	-۲/۰۹	-۲/۳	-۲/۰۵	-۲/۰۵	۲/۵۱-	۲/۸۱-	۳/۲۵-	۳/۶-	۳/۸۶-	۴/۲۸-	۴/۷-	۵/۰۱-	۵/۳۹-	۵/۴۹-
	زمستان	-۲/۴۶	-۳/۰۲	-۲/۵۳	-۲/۳۲	۲/۳۵-	۲/۵۹-	۲/۹-	۳/۳۴-	۳/۷-	۳/۹۶-	۴/۳۶-	۴/۷۸-	۵/۰۹-	۵/۴۸-	۵/۵۸-
۱۳۹۵	بهار	-۲/۴	-۲/۶	-۲/۱۵	۲/۲-	۲/۳۷-	۲/۵۹-	۲/۹۱-	۳/۳۵-	۳/۷۱-	۳/۹۷-	۴/۳۷-	۴/۷۹-	۵/۱-	۵/۴۹-	۵/۵۹-
	تابستان	-۲/۵۱	-۲/۵۹	۲/۲۵-	۲/۲۱-	۲/۳۷-	۲/۶-	۲/۹۲-	۳/۳۵-	۳/۷۱-	۳/۹۷-	۴/۳۸-	۴/۸۱-	۵/۱۱-	۵/۴۹-	۵/۶-
	پاییز	-۲/۵	-۲/۵۸	-۲/۲۵	-۲/۲۳	۲/۳۶-	۲/۶۱-	۲/۹۲-	۳/۳۵-	۳/۷۱-	۳/۹۷-	۴/۳۹-	۴/۸۱-	۵/۱۲-	۵/۴۹-	۵/۵۹-

جدول ۴: مثلث تأخیر میزان لگاریتم پرداخت‌های خسارات نرمال شده بیمه بدنه (مثلث بالایی) و میزان لگاریتم پرداخت‌های خسارات نرمال شده پیش‌بینی شده بر اساس مدل قدم‌زدن تصادفی با توزیع  $t$ ی استیودنت دومتغیره (مثلث پایینی)

سال خسارت	فصل خسارت	فصل توسعه خسارت														
		۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
۱۳۹۲	بهار	-۱/۰۱	-۱/۲۸	-۲/۶۳	-۳/۹۸	-۴/۸۱	-۵/۶۵	-۶/۶۸	-۵/۸۴	-۶/۴۸	-۶/۴	-۸/۲۵	-۷/۶۶	-۹/۸	-۸/۸۳	-۷/۷۵
	تابستان	-۱/۰۹	-۱/۳۳	-۲/۵۶	-۳/۷۸	-۵/۰۵	-۵/۳۹	-۵/۶۸	-۷/۲۸	-۷/۸۷	-۸/۰۴	-۶/۶۸	-۶/۹۸	-۸/۰۴	-۷/۳۶	-۷/۸۱
	پاییز	-۱/۰۹	-۱/۱۶	-۲/۹	-۳/۶۸	-۵/۰۱	-۵/۱۸	-۶/۰۸	-۶/۲۵	-۶/۹۸	-۷/۵۸	-۸/۱۶	-۹/۴	-۶/۹۸	-۸/۰۴	-۷/۸۶
	زمستان	-۱/۱۶	-۱/۸۱	-۲/۸۵	-۴/۴۲	-۵/۲۲	-۶/۰۲	-۵/۹۳	-۶/۸۷	-۷/۸۱	-۸/۹۳	-۸/۸۵	-۷/۱۲	-۸/۱۵	-۸/۱۴	-۷/۹۸
۱۳۹۳	بهار	-۱/۰۶	-۱/۱۸	-۲/۷	-۳/۷۱	-۵/۱۹	-۵/۸۳	-۴/۶	-۶/۳۳	-۸/۴	-۸/۷۱	-۷/۶	-۷/۹۳	-۸/۲۱	-۸/۲۲	-۸/۰۵
	تابستان	-۱/۰۴	-۱/۳	-۲/۷	-۴/۵۲	-۶/۱۳	-۸/۰۴	-۶/۹۷	-۶/۸	-۸/۳۴	-۸/۱۱	-۸/۲۳	۸/۰۹	-۸/۳۹	-۸/۳۹	-۸/۲۲
	پاییز	-۱/۱۱	-۱/۳	-۳/۴۳	-۴/۶۸	-۵/۵۸	-۵/۹۴	-۷/۹۲	-۶/۷۹	-۹/۲۸	۸/۳۵	۸/۳۳	۸/۱۷	۸/۴۶	۸/۴۷	۸/۳۱
	زمستان	-۱/۱۷	-۱/۸۳	-۳/۳۶	-۴/۸۹	-۵/۹۹	-۶/۶۳	-۷/۴۲	-۸/۲۳	۸/۱۶	۸/۳۸	۸/۳۵	۸/۲	۸/۵	۸/۵۱	۸/۳۳
۱۳۹۴	بهار	-۰/۸۷	-۱/۱۵	-۲/۹۴	-۳/۸۲	-۵/۴۸	-۶/۴۵	-۶/۹۸	۷/۰۶	-۸/۱	۸/۲۹	۸/۲۷	۸/۱۱	۸/۴۳	۸/۴۳	۸/۲۵
	تابستان	-۰/۸۶	-۱/۲۹	-۲/۹۲	-۴/۵	-۶/۱	-۶/۱۸	۶/۶۹	۷/۰۱	۸/۰۵	۸/۲۵	۸/۲۴	۸/۰۶	۸/۳۷	۸/۳۸	۸/۲۱
	پاییز	-۰/۸۶	-۱/۰۵	-۳/۱۱	-۴/۷۹	-۴/۸۹	۶/۰۷	۶/۶۶	۶/۹۴	۷/۹۸	۸/۱۹	۸/۱۸	۸/۰۳	۸/۳۱	۸/۳۲	۸/۱۷
	زمستان	-۰/۹۶	-۱/۶۱	-۳	-۴/۵۸	۵/۴	۶/۰۶	۶/۶۱	۶/۹	۷/۹۵	۸/۱۷	۸/۱۵	۷/۹۸	۸/۲۹	۸/۲۹	۸/۱۲
۱۳۹۵	بهار	-۰/۷۲	-۱/۱۲	-۲/۹۹	۴/۲۸	۵/۳۷	۶/۰۱	۶/۵۷	۶/۸۸	۷/۹۲	۸/۱۳	۸/۱۲	۷/۹۴	۸/۲۶	۸/۲۶	۸/۰۸
	تابستان	-۰/۸۷	-۱/۱۶	۲/۹	۴/۲۵	۵/۳۳	۵/۹۸	۶/۵۶	۶/۸۵	۷/۸۹	۸/۱	۸/۰۹	۷/۹۳	۸/۲۱	۸/۲۳	۸/۰۶
	پاییز	-۰/۹۸	-۱/۳۶	-۲/۹	۴/۲۸	۵/۳۵	۶/۰۱	۶/۵۵	۶/۸۵	۷/۹۱	۸/۱	۸/۰۸	۷/۹۵	۸/۲۳	۸/۲۳	۸/۰۶

میانگین پسینی و انحراف استاندارد ذخایر دو مثلث بر اساس مدل‌های پیشنهاد شده و بر اساس مدل نردبان زنجیره‌ای<sup>۱</sup> در جدول ۵ گزارش شده است. برآورد ذخایر با استفاده از مدل نردبان زنجیره‌ای در دو مثلث به طور جداگانه و بدون در نظر گرفتن وابستگیها به دست آمده است. نتایج نشان می‌دهد که مدل نردبان زنجیره‌ای مقدار ذخایر را کمتر از روشهای پیشنهاد شده برآورد می‌کند و انحراف استاندارد برآوردهای به دست آمده بر اساس این مدل نیز از انحراف استاندارد برآوردهای به دست آمده بر اساس مدل‌های آنالیز کوواریانس و قدم‌زدن تصادفی بزرگتر است.

جدول ۵: برآوردهای ذخایر

بیمه بدنه		بیمه شخص ثالث		توزیع خسارت	مدل میانگین
انحراف استاندارد	میانگین	انحراف استاندارد	میانگین		
$۱۰۹ \times ۸۴۴/۹$	$۱۰۹ \times ۱۰۳۲$	$۱۰۹ \times ۶۳۴۴$	$۱۰۱۰ \times ۳۵۳۲$	نرمال	ANOVA
$۱۰۹ \times ۹۷۵/۹$	$۱۰۹ \times ۱۰۱۰$	$۱۰۹ \times ۶۴۲۸$	$۱۰۱۰ \times ۳۵۸۷$	پی‌یرسون نوع هفتم	
$۱۰۹ \times ۸۵۳/۹$	$۱۰۹ \times ۹۹۲/۴$	$۱۰۹ \times ۶۹۷۹$	$۱۰۱۰ \times ۳۶۰۴$	ی استیودنت t	
$۱۰۹ \times ۲۸۴/۸$	$۱۰۹ \times ۶۰۴/۵$	$۱۰۹ \times ۴۴۴۷$	$۱۰۱۰ \times ۳۵۷۵$	نرمال	ANCOVA
$۱۰۹ \times ۳۳۶/۷$	$۱۰۹ \times ۶۳۲/۳$	$۱۰۹ \times ۴۴۶۶$	$۱۰۱۰ \times ۳۵۸۸$	پی‌یرسون نوع هفتم	
$۱۰۹ \times ۳۱۰$	$۱۰۹ \times ۶۳۵/۳$	$۱۰۹ \times ۴۵۷۱$	$۱۰۱۰ \times ۳۵۹۹$	ی استیودنت t	
$۱۰۹ \times ۳۷۲/۱$	$۱۰۹ \times ۸۲۰/۵$	$۱۰۹ \times ۴۶۹۴$	$۱۰۱۰ \times ۳۴۹۳$	نرمال	قدم‌زدن تصادفی
$۱۰۹ \times ۳۶۳$	$۱۰۹ \times ۸۲۰/۱$	$۱۰۹ \times ۴۵۲۵$	$۱۰۱۰ \times ۳۵۱۹$	پی‌یرسون نوع هفتم	
$۱۰۹ \times ۴۸۵/۹$	$۱۰۹ \times ۸۲۸/۸$	$۱۰۹ \times ۴۷۶۸$	$۱۰۱۰ \times ۳۵۵۱$	ی استیودنت t	
$۱۰۹ \times ۹۴۷$	$۱۰۹ \times ۷۵۴$	$۱۰۹ \times ۵۰۲۶$	$۱۰۱۰ \times ۳۱۳۴$	مدل نردبان زنجیره‌ای	

<sup>1</sup>. Chain-Ladder

همان گونه که قبلاً بیان شد، علی‌رغم موجود بودن اطلاعات برای ۱۶ فصل، مدلها تنها به داده‌های مربوط به ۱۵ فصل اول برازش داده شد و خسارات پرداخت شده در فصل ۱۶م پیش‌بینی و برای این مقادیر، میانگین اریبی نسبی<sup>۱</sup>، با استفاده از فرمول

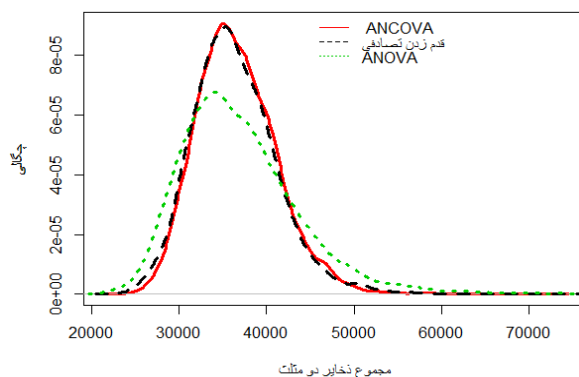
$$\text{Rel.Bias} = \frac{1}{NM} \sum_{m=1}^M \sum_A (\hat{y}_{ij}^{(i)}[m] - y_{ij}^{(i)}[\text{obs.}]) / y_{ij}^{(i)}[\text{obs.}],$$

برای هر دو مثلث بر اساس مدل قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دوم‌تغیره محاسبه شده است، که در رابطه فوق  $y_{ij}^{(i)}[\text{obs.}]$  و  $\hat{y}_{ij}^{(i)}[m]$  به ترتیب مقدار مشاهده شده و پیش‌بینی شده در  $m$  امین تکرار زنجیر MCMC در مثلث  $l$  ام و  $A$  مجموعه اندیسهای مربوط به خسارات پرداخت شده در فصل ۱۶م است که در داده‌های مورد مطالعه برای هر مثلث  $N = 14$  است. همچنین این معیار برای پیش‌بینی‌های خسارات پرداخت شده در فصل ۱۶م با استفاده از مدل نردبان زنجیره‌ای محاسبه شد.

جدول ۶: برآورد میانگین توان دوم خطای پیش‌بینی

Rel.Bias	مدل	نوع بیمه
-۰/۴۵	نردبان زنجیره‌ای	شخص ثالث
-۰/۳۷	ی استیودنت $t$ قدم‌زدن تصادفی با توزیع	
۱/۲۴	مدل نردبان زنجیره‌ای	بدنه
۱/۰۱	ی استیودنت $t$ قدم‌زدن تصادفی با توزیع	

جدول ۶ نشان می‌دهد که در هر دو مثلث عملکرد مدل قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دوم‌تغیره از مدل نردبان زنجیره‌ای بهتر بوده است. شکل ۶، نمودار توزیع پیش‌بینی پسینی مجموع ذخایر دو مثلث را بر اساس سه مدل ANOVA، ANCOVA و قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دوم‌تغیره برای لگاریتم پرداختهای خسارات نرمال شده با استفاده از نمونه‌های مونت کارلو نشان می‌دهد.

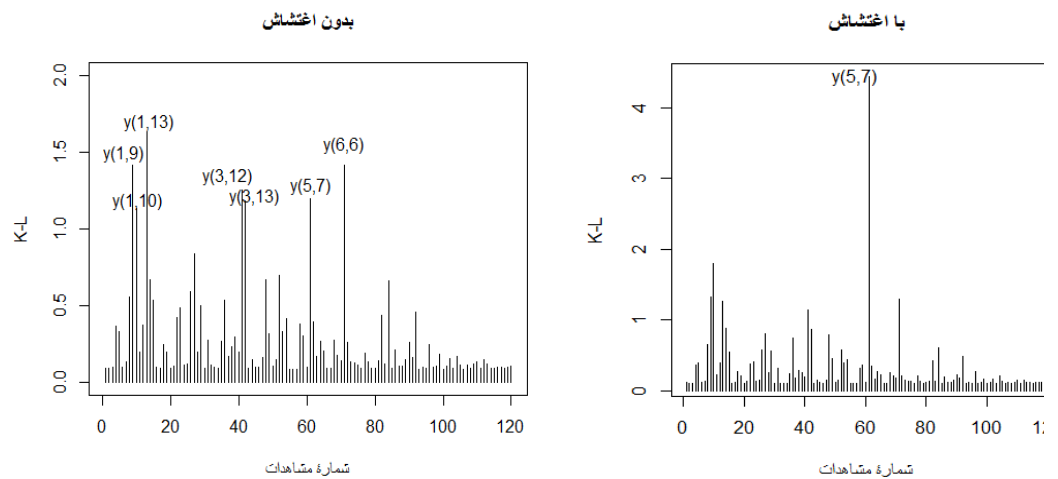


شکل ۶: نمودارهای توزیع پیش‌بینی پسینی مجموع ذخایر دو مثلث بر اساس توزیع  $t$  ی استیودنت

برای کشف نقاط مؤثر در داده‌ها، اندازه واگرایی کولبک-لیبلر که در بخش ۴ معرفی شد، بر اساس نمونه‌های پسینی پارامترهای مدل قدم-زدن تصادفی با توزیع  $t$  ی استیودنت دوم‌تغیره محاسبه شد. شکل ۸، نمودار اندازه واگرایی کولبک-لیبلر را برای همه نمونه‌های دوم‌تغیره دو رشته بیمه نشان می‌دهد. توجه به میزان پرداختهای خسارات در دوره‌های قبل و بعد در سطرها و ستونهای دو مثلث، اندازه‌های کولبک-لیبلر بالای نمونه‌های مؤثر شناسایی شده در شکل ۸ را در مقایسه با سایر نمونه‌ها بیان می‌کند. همچنین یک طرح اغتشاش در مشاهدات را

<sup>1</sup>. Relative Bias

با  $y_{57}^{(1)} + \delta = 2$  است، در نظر گرفتیم. همان‌گونه که در شکل ۸ مشاهده می‌شود، این مشاهده جدید به‌عنوان یک مشاهده پرت شناسایی شده است.



شکل ۷: اندازه واگرایی کولبک-لیبلر برای مدل قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دومتغیره

## نتایج و بحث

در این مقاله با استفاده از استنباط بیزی، سه نوع مدل ANOVA، ANCOVA، و قدم‌زدن تصادفی با توزیع‌های نرمال، توزیع پی‌یرسون نوع هفتم و  $t$  ی استیودنت دومتغیره برای لگاریتم پرداخت‌های خسارات نرمال‌شده را برای داده‌های ذخیره‌سازی خسارات دومتغیره بیمه بدنه و بیمه شخص ثالث اتومبیل یک شرکت بیمه ایرانی مورد بررسی قرار دادیم. نتایج نشان می‌دهد که مدل‌های با توزیع‌های دُم‌سنگین عملکرد بهتری نسبت به توزیع نرمال دارند.

همچنین با استفاده از مدل نردبان زنجیره‌ای در دو مثلث به طور جداگانه و بدون در نظر گرفتن وابستگیها، برآورد ذخایر را به دست آوردیم؛ نتایج نشان می‌دهد که مدل نردبان زنجیره‌ای مقدار ذخایر را کمتر از روش‌های پیشنهادشده برآورد می‌کند و انحراف استاندارد برآوردهای به‌دست‌آمده بر اساس این مدل نیز از انحراف استاندارد برآوردهای به‌دست‌آمده بر اساس مدل‌های آنالیز کوواریانس و قدم‌زدن تصادفی بزرگتر است. به منظور ارزیابی دقت پیش‌بینی مدل‌ها، میانگین اریبی نسبی مقادیر خسارات پرداخت‌شده در فصل ۱۶م، با دو مدل قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دومتغیره و مدل نردبان زنجیره‌ای محاسبه شد؛ نتایج نشان داد که در هر دو مثلث عملکرد مدل قدم‌زدن تصادفی با توزیع  $t$  ی استیودنت دومتغیره از مدل نردبان زنجیره‌ای بهتر بوده است؛ بنابراین در نظر گرفتن وابستگی‌های بین مثلث‌های تأخیر در محاسبه ذخایر می‌تواند به نتایج دقیق‌تری منجر شود. به‌علاوه، بر اساس واگرایی کولبک-لیبلر، نمونه‌های مؤثر بر توزیع‌های پسینی توأم شناسایی شدند.

بنابراین با توجه به اینکه بر اساس آیین‌نامه شماره ۵۸ بیمه مرکزی، روش محاسبه ذخایر خسارات معوق در شرکت‌های بیمه در ایران محدود و فقط با در نظر گرفتن یک مثلث تأخیر است، بر اساس نتایج این تحقیق پیشنهاد می‌شود که به منظور وارد کردن وابستگی‌های بین مثلث‌های تأخیر در محاسبه ذخایر از روش‌های بیزی استوار برای مدل‌بندی توأم خسارات معوق مربوط به رشته‌های مختلف بیمه‌ای استفاده شود. همچنین پیشنهاد می‌شود که پس از برآزش مدل‌ها، یک مطالعه آنالیز حساسیت بر اساس واگرایی کولبک-لیبلر به منظور شناسایی نمونه‌های مؤثر انجام شود. روش‌ها و معیارهای بیزی به‌آسانی با نرم‌افزاری استاندارد بیزی مانند

OpenBUGS محاسبه می‌شوند.

پژوهشکده بیمه، (۱۳۹۵). بازنگری آیین نامه ذخایر فنی مؤسسات بیمه و ارائه پیشنهادهاى اصلاحی.

- Braun, C., (2004). The prediction error of the chain ladder method applied to correlated run-off triangles. *ASTIN Bulletin*, 34 (2), pp. 399-424.
- Chen, M.; Shao, Q.; Ibrahim, J., (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- Cho, H.; Ibrahim J.G.; Sinha, D.; Zhu, H., (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, 65, pp.116-124.
- Cook, R., (1986). Assessment of local influence. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48, pp.133-169.
- Côté, M.P.; Genest, C.; Abdallah, A., (2016). Rank-based methods for modeling dependence between loss triangles. *European Actuarial Journal*, 6(2), pp. 377-408.
- De Jong, P., (2012). Modeling dependence between loss triangles using copula. *North American Actuarial Journal*, 16(1), pp. 74-86.
- England, P.D.; Verrall, R.J., (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(03), pp. 443-518.
- England, P.D.; Verrall, R.J., (2006). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, 1, pp. 221-270.
- Fernandez, C.; Steel, M.F., (2000). Bayesian regression analysis with scale mixtures of normals. *Econometric Theory*, 16(01), pp. 80-101.
- Gelfand, A.; Dey, D.; Chang, H., (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics*, 4, pp. 147-167.
- Happ, S.; Wüthrich, M.V., (2013). Paid-incurred chain reserving method with dependence modeling. *ASTIN Bulletin*, 43(1), pp. 1-20.
- Hess, K.; Schmidt, K.; Zocher, M., (2006). Multivariate loss prediction in the multivariate additive model. *Insurance: Mathematics and Economics*, 39(2), pp. 185-191.
- Merz, M.; Wüthrich, M.V., (2009a). Combining chain-ladder and additive loss reserving method for dependent lines of business. *Variance*, 3(2), pp. 270-291.
- Merz, M.; Wüthrich, M.V., (2009b). Prediction error of the multivariate additive loss reserving method for dependent lines of business. *Variance*, 3(1), pp. 131-151.
- Merz, M.; Wüthrich, M.V.; Hashorva, E., (2013). Dependence modeling in multivariate claims run-off triangles. *Annals of Actuarial Science*, 7(1), pp. 3-25.
- Pearson, K., (1916). *Mathematical contributions to the theory of evolution. XIX. Second supplement to a memoir on skew variation*. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 216, pp.429-457.
- Salzmann, R.; Wüthrich, M.V., (2012). Modeling accounting year dependence in run-off triangles. *European Actuarial Journal*, 2(2), pp. 227-242.
- Schmidt, K., (2006). Optimal and additive loss reserving for dependent lines of business. *Casualty Actuarial Society Forum*, pp. 319-351.
- Shi, P., (2014). A copula regression for modeling multivariate loss triangles and quantifying reserving variability. *ASTIN Bulletin*, 44(01), pp. 85-102.
- Shi, P.; Basu, S.; Meyers, G., (2012). A bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, 16(1), pp. 29-51.
- Shi, P.; Frees, E., (2011). Dependent loss reserving using copulas. *ASTIN Bulletin*, 41(2), pp. 449-486.

- Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Lindevan der, A., (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64, pp. 583–616.
- Taylor, G., (2000). *Loss Reserving: An Actuarial Perspective*. Dordrecht: Kluwer Academic Publishers.
- Verrall, R., (1991). Chain ladder and maximum likelihood. *Journal of the Institute of Actuaries*, 118, pp. 489–499.
- Wüthrich, M.V.; Merz, M., (2008). *Stochastic Claims Reserving in Insurance*. John Wiley Sons.
- Zhang, Y., (2010). A general multivariate chain ladder model. *Insurance: Mathematics and Economics*, 46, pp. 588-599.
- Zhang, Y.; Dukic, V.; Guszczka, J., (2012). A Bayesian non-linear model for forecasting insurance loss payments. *Journal of the Royal Statistical Society: Series A*, 175(2), pp. 637-656.