دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد ریاضی مالی

رتبه‌بندی ریسک های بیضوی چند متغیره بر اساس شاخص‌های جینی و انتخاب مدل بر مبنای منحنی های تمرکز و لورنز

نگارنده: مجید جهانی

استادان راهنما

دکتر علیرضا خدّامی
دکتر محمد میرباقری جم

بهمن ۱۳۹۹

تقدیم به سه وجود مقدس:
آنان که ناتوان شدند تا ما به توانایی برسیم...
موهایشان سپید شد تا ماروسفید شویم...
و عاشقانه سوختند تا گرمابخش وجود ما و روشنگر راهمان باشند...
پدرانمان
مادرانمان
استادانمان

سپاس‌گزاری...

بدین وسیله از زحمات و تلاش‌های بی‌دریغ اساتید محترم جناب آقای دکتر علیرضا خدّامی و جناب آقای دکتر محمد میرباقری جم و خانواده عزیزم صمیمانه سپاسگزاری می‌نمایم و همچنین از سایر همکاران و دوستانی که هر کدام به نحوی در تهیه این مجموعه با این جانب همکاری داشته اند تشکر نموده و موفقیت همه آنها را از خداوند متعال خواهانم.

مجید جهانی
بهمن ۱۳۹۹

تعهد نامه

اینجانب **مجید جهانی** دانشجوی کارشناسی ارشد رشته **ریاضی مالی دانشکده علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان **رتبه‌بندی ریسک های بیضوی چند متغیره بر اساس شاخص‌های جینی و انتخاب مدل بر مبنای منحنی های تمرکز و لورنز**، تحت راهنمایی **دکتر علیرضا خدّامی** و **دکتر محمد میرباقری جم** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام " دانشگاه صنعتی شاهرود " یا " Shahrood University of Technology " به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

مجید جهانی
بهمن ۱۳۹۹

مالکیت نتایج و حق نشر

چکیده

شاخص جینی، یکی از ابزارهای شناخته شده در اقتصاد است، که معمولا برای اندازه‌گیری عدم تساوی درآمد به کار گرفته می شود. در بیمه، این شاخص و تعمیم‌یافته آن برای مقایسه میزان ریسک سبد ها، رتبه بندی قراردادهای بیمه اتکایی و جمع‌بندی امتیازات بیمه (نسبیت‌ها) به کار می‌روند. در این پایان‌نامه، چندین ترتیب تصادفی بین شاخص‌های جینی ریسک‌های بیضوی چند متغیره با توزیع‌های حاشیه‌ای یکسان، اما ساختارهای وابستگی متفاوت ارائه می‌شود. در این پایان‌نامه با مقایسه شاخص‌های جینی (مقادیر ریسک برآورد شده به صورت تجربی)، یک تفسیر نظری برای این پدیده آماری ارائه می‌شود. به علاوه، مطالعاتی که بر روی مسئله تمرکز مرکزی توزیع‌های بیضوی انجام شده‌اند، تقویت می شود و ترتیب $pd-1$ پیشنهاد شده توسط Shaked و Tong در سال ۱۹۸۵، تعمیم می‌یابد.
به منظور تعیین حق بیمه مناسب، بایستی از ضوابط نیکویی برازش آماری به هنگام انتخاب یک حق بیمه مفروض استفاده کرد. در این پایان‌نامه نشان داده می شود که منحنی های تمرکز و منحنی های لورنز، ابزارهای موثری در اختیار بیمه گران قرار می دهند تا تعیین کنند که آیا حق بیمه مناسب است، یا اینکه آن را با گزینه‌های رقیب مقایسه کنند. ایده اصلی، مقایسه درآمد حق بیمه برای زیرپرتفوی هایی که ریسک های کم را گردآوری کرده اند (که حق بیمه پایین نامیده می شوند) با حق بیمه صحیح یا هم ارز آن، با زیان های واقعی است. مثال‌های عددی بر اساس داده های فرضی و داده های حقیقی، سودمندی رویکرد پیشنهادی را نشان می‌دهند.
در نهایت با استفاده از مدل‌های خطی تعمیم‌یافته و منحنی‌های تمرکز و لورنز عوامل اثرگذار بر شاخص کل بورس اوراق بهادار تهران را مورد بررسی قرار می‌دهیم، و مدلی که پیش‌بینی بهتری از شاخص بورس را ارائه می‌دهد، انتخاب می‌کنیم.

کلمات کلیدی: توزیع بیضوی، ساختار وابستگی، هم‌یکنوایی، ترتیب تصادفی معمولی، ترتیب محدب صعودی، ترتیب فوق مدولی، ترتیب pd-1، قیمت‌گذاری، رده‌بندی ریسک، منحنی تمرکز، منحنی لورنز، مدل خطی تعمیم‌یافته(GLM)، بورس اوراق بهادار تهران، انتخاب مدل.

فهرست مطالب

فهرست تصاویر ف

فهرست جداول ق

پیشگفتار ش

۱ پیش‌نیازها و تعاریف مقدماتی ۱

۱.۱ مقدمه ۱

۲.۱ تعاریف مقدماتی ۱

۳.۱ تابع مفصل ۱۰

۴.۱ رگرسیون خطی ۱۳

۵.۱ مفاهیم مربوط به حق بیمه‌های خالص و کاری ۱۴

۱.۵.۱ فرضیات ۱۴

۲.۵.۱ نمادگذاری ۱۶

۳.۵.۱ ترتیب محدب ۱۶

۴.۵.۱ تعاریف منحنی‌های عملکرد ۱۷

۲ رتبه‌بندی شاخص‌های جینی و انتخاب مدل بر اساس منحنی‌های تمرکز و لورنز ۲۱

۱.۲ مقدمه ۲۱

۲.۲ رتبه‌بندی شاخص‌های جینی براساس ترتیب $\leq_{icx}$ ۲۱

۳.۲ رتبه‌بندی شاخص‌های جینی بر اساس ترتیب $\leq_{st}$. ۲۵

۱.۳.۲ حالت تبادلی شرطی ۲۵

۲.۳.۲ شبه ـ تسلط بین ماتریس های کوواریانس ۲۸

۳.۳.۲ مقایسه با هم یکنوایی ۳۱

۴.۲ بحث درباره مرتبه $pd-1$ ۳۲

۵.۲ منحنی های عملکرد ۳۴

۱.۵.۲ برآورد منحنی‌های تمرکز و لورنز ۳۴

۲.۵.۲ از حق بیمه تا مرتبه ها ۳۶
۳.۵.۲ ویژگی‌های منحنی لورنز و تمرکز ۳۶
۴.۵.۲ اندازه‌گیری نیکویی برازش ۳۹
۶.۲ مقایسه عملکرد های دو پیش‌بین ۴۰
۱.۶.۲ منحنی های تمرکز و لورنز ۴۰
۲.۶.۲ منحنی‌های تمرکز و لورنز انتگرال‌دار ۴۳
۷.۲ برخی معیار های سودمند بیمه ۴۴
۱.۷.۲ شاخص جینی ۴۴
۲.۷.۲ معیار ABC (Area Between two Curves) ۴۵
۸.۲ مثال‌های عددی ۴۷
۱.۸.۲ فرضیات ۴۷
۲.۸.۲ تغییرپذیری ۴۸
۳.۸.۲ وابستگی ۴۹
۴.۸.۲ توزیع ۵۰
۵.۸.۲ مفصل های متقاطع ۵۱
۶.۸.۲ اثر مفصل با وابستگی غیر رگرسیونی ۵۲
۹.۲ مورد پژوهی ۵۴
۱.۹.۲ فراوانی ۵۴
۲.۹.۲ فراوانی و شدت ۵۸
۱۰.۲ نتیجه گیری ۵۹

۳ اعتبار سنجی تحقیق **۶۱**
۱.۳ مقدمه ۶۱
۲.۳ تحلیل وضعیت شاخص بورس طی دوره تحقیق ۶۲
۱.۲.۳ دوره تحقیق و داده‌های مربوطه ۶۲
۲.۲.۳ قیمت‌گذاری بیش از حد و کمتر از حد سهام ۶۳
۳.۲.۳ مقایسه بازده موثر بورس با سایر بازارهای موازی در دوره تحقیق . ۶۴
۳.۳ شناسایی عوامل موثر بر شاخص بورس اوراق بهادار تهران ۶۵
۱.۳.۳ عوامل موثر بر شاخص بورس در مطالعات تجربی انجام یافته . . ۶۶
۲.۳.۳ سنجش ارتباط شاخص کل بورس (TSE) با عوامل تعیین کننده . ۶۷
۴.۳ مدل‌سازی وپیش‌بینی شاخص کل بورس اوراق بهادار تهران ۷۰
۱.۴.۳ مدل‌ها و رویکردهای پیش‌بینی شاخص کل بورس ۷۰
۲.۴.۳ مدل‌های GLM در پیش‌بینی شاخص کل بورس ۷۲
۳.۴.۳ انتخاب مدل پیش‌بینی شاخص کل بر مبنای معیار ABC ۷۴

۵.۳ نتیجه گیری ۷۶
۶.۳ خلاصه فصل سوم ۷۶

مراجع **۷۷**

آ ضمیمه **۸۳**
آ.۱ جدول‌های سنجش ارتباط متغیرهای تحقیق ۸۳

واژه‌نامه فارسی به انگلیسی **۸۵**

واژه‌نامه انگلیسی به فارسی **۸۹**

فهرست تصاویر

۱.۲ منحنی‌های جهت‌دار ∂C_+، ∂C_-. ۲۲
۲.۲ منحنی‌های جهت‌دار $\partial C'_+$، ∂C_-. ۲۶
۳.۲ منحنی لورنز و منحنی‌های تمرکز مختلف برای واریانس‌های مختلف در توزیع مقدار انتظاری شرطی. ۴۹
۴.۲ منحنی لورنز و منحنی‌های تمرکز مختلف برای پارامترهای تاو متفاوت مفصل. ۵۰
۵.۲ منحنی لورنز و منحنی‌های تمرکز مختلف برای توزیع‌های مختلف از مقدار انتظاری شرطی. ۵۱
۶.۲ منحنی لورنز و دو منحنی تمرکز برای مفصل‌های مختلف. ۵۲
۷.۲ منحنی لورنز و منحنی‌های تمرکز برای مفصل با وابستگی غیرگرسیونی. . ۵۳
۸.۲ خطاهای برآورد درون نمونه و بیرون نمونه برای مدل‌های مدنظر. . . ۵۵
۹.۲ $\widehat{CC}$ برای مدل‌های مورد مطالعه. ۵۷
۱۰.۲ مقادیر ABC و ICC برآورد شده برای مدل‌های مدنظر. ۵۷
۱۱.۲ مقادیر ABC و ICC برآورد شده برای مدل‌های مدنظر. ۵۹
۱۲.۲ منحنی‌های تمرکز برای مدل‌های شدت فراوانی و توئیدی(Tweedie). . ۵۹

۱.۳ مقایسه TSE* و TSE ۶۴
۲.۳ میانگین بازده روزانه داده‌ها. ۶۵
۳.۳ منحنی تمرکز عوامل تعیین کننده شاخص بورس. ۷۰
۴.۳ منحنی تمرکز نمونه مدل‌های جدول ۸.۳ . . . . . ۷۴
۵.۳ منحنی تمرکز و لورنز مدل‌های برتر. ۷۵

فهرست جداول

۱.۱ مثال همبستگی رتبه‌ای اسپیرمن ۴

۱.۲ نتایج بدست آمده از نمودار۸.۲ . . . . . ۴۸
۲.۲ نتایج بدست آمده از نمودار۴.۲ . . . . . ۴۹
۳.۲ نتایج بدست آمده از نمودار ۵.۲ . . . . . ۵۱
۴.۲ نتایج بدست آمده از نمودار ۶.۲ . . . . . ۵۲
۵.۲ نتایج بدست آمده از نمودار ۷.۲ . . . . . ۵۳
۶.۲ نتایج بدست آمده از نمودار۱۱.۲ . . . . . ۵۸

۱.۳ جدول مشخصات متغیرها ۶۵
۲.۳ مقادیر ضریب همبستگی پیرسون برای متغیرها ۶۸
۳.۳ مقادیر ضریب رتبه‌ای اسپیرمن برای متغیرها ۶۹
۴.۳ مقادیر ضریب هماهنگی تاو کندال برای متغیرها ۶۹
۵.۳ مقایسه معیار ضریب منحنی تمرکز متغیرها با TSE ۶۹
۶.۳ خلاصه‌ای از رویکردهای مورد استفاده و پیش‌بینی شاخص بورس در مطالعات پیشین ۷۲
۷.۳ نمونه‌ای از مدل‌های GLM برآورد شده ۷۳
۸.۳ مقادیر منحنی تمرکز ، لورنز و ABC نمونه مدل‌های جدول ۷.۳ . . . . . ۷۳
۹.۳ مقادیر منحنی تمرکز، لورنز و ABC مدل‌های برتر ۷۵

آ.۱ مقایسه ضریب همبستگی پیرسون متغیرها ۸۳
آ.۲ مقایسه ضریب همبستگی اسپیرمن متغیرها ۸۴
آ.۳ مقایسه ضریب هماهنگی تاو کندال متغیرها ۸۴

پیشگفتار

بیش از صد سال پیش، کورادو جینی[1] یک شاخص برای اندازه‌گیری تمرکز یا عدم تساوی درآمدها معرفی کرد [۳۶]. این شاخص بعدها به شاخص جینی[2] معروف شد و در حوزه‌های متعدد مانند اقتصاد، بیمه، دانش مالی و آمار کاربرد گسترده‌ای پیدا کرد. به عنوان مثال، در بیمه و آمار، این شاخص برای مقایسه توزیع ریسک و توزیع قیمت مورد استفاده گرفته است [۳۱].

این مقایسه ها معمولاً مبتنی بر امتیازات بیمه نسبت به هزینه بیمه هستند که به «نسبیت» معروف هستند و به عدم توافق احتمالی بین توزیع ریسک و توزیع قیمت می‌پردازند. بعد از رتبه بندی ریسک و قیمت بر اساس نسبیت، به منحنی لورنز مرتب شده می‌رسیم که می توان آن را با استفاده از شاخص جینی، جمع‌بندی کرد. جالب است که منحنی لورنز و شاخص جینی که از طریق نسبیت تعریف می شوند، می توانند هر نوع مجموعه انتخابی را بررسی کنند، به اندازه گیری سود بالقوه کمک کنند و به عنوان ابزارهای مفیدی در مدل های پیش بین مورد استفاده قرار گیرند. برای اطلاعات بیشتر می‌توان به مرجع [۳۱] مراجعه نمود.

به علاوه، منحنی لورنز و ترتیب لورنز که مفاهیمی بسیار نزدیک به شاخص جینی هستند، توسط Denuit و همکارانش برای رتبه‌بندی قراردادهای بیمه اتکایی مورد استفاده قرار گرفته‌اند [۲۴].

کاربردهای آماری دیگر در بیمه بر این حقیقت تاکید دارند که شاخص جینی، یک $L-$آماره‌است، که خصوصیات نظری آن کاملاً تعیین شده است و بنابراین می توان از آن برای ایجاد ابزارهای استنباطی آماری بهره گرفت. به عنوان مثال، در مقالات [۴۵]، [۴۴] و [۱۱]، چندین آزمون فرض[3] برای مقایسه میزان ریسک سبدهای بیمه با استفاده از شاخص جینی مورد استفاده قرار گرفته است. Samanthi و همکارانش، یک مطالعه شبیه‌سازی مبسوط انجام دادند، که در آن انواع مختلف وابستگی بین سبدها را در نظر گرفتند و دریافتند زمانی که سبدها، همبستگی مثبت بیشتری داشته باشند، تابع توان آزمون فرض مربوطه، افزایش می‌یابد [۶۱]. در این پایان‌نامه مقایسه شاخص‌های جینی (مقادیر ریسک برآورد شده به صورت تجربی) همراه با توضیح نظری برای این پدیده آماری خواهد بود.

[1] Corrado Gini
[2] Gini index
[3] Hypothesis test

همانگونه که توسط Samanthi و همکارانش شرح داده شده است، تابع توان آزمون فرض مدنظر، یک رویداد احتمالی است که شاخص جینی $\frac{1}{n^2}\sum_{1\leq i,j\leq n}|X_i - X_j|$ دارد که در آن، متغیرهای تصادفی $X_1, X_2, \ldots, X_n$ مقادیر ریسک تجربی را نشان می‌دهند که از مشاهدات n سبد بیمه برآورد شده است. همچنین بردار تصادفی $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ از یک توزیع نرمال چندمتغیره مجانبی تبعیت می‌کند [۶۱]. برای جزییات بیشتر در ارتباط با طراحی آزمون فرض، خواننده می‌تواند به مراجع [۶۱، ۱۱] مراجعه نماید.

برای شرح یکنوایی تابع توان آزمون از نظر قدرت وابستگی، حدس زیر ارائه می‌شود.

حدس: فرض کنید $(X_1, X_2, \ldots, X_n)$ دارای توزیع نرمال چندمتغیره با میانگین $\mathbf{0}$ و ماتریس کواریانس $\boldsymbol{\Sigma}$ باشد. به عبارت دیگر فرض کنید $(X_1, \ldots, X_n) \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$. در این‌صورت، زمانی که ماتریس کوواریانس $\boldsymbol{\Sigma}$ به صورت مولفه‌ای افزایش می‌یابد اما مولفه‌های قطری آن بدون تغییر باقی می‌مانند، شاخص جینی $\frac{1}{n^2}\sum_{1\leq i,j\leq n}|X_i - X_j|$ بر اساس ترتیب تصادفی معمولی کاهش می‌یابد (برای مشاهده تعریف، به بخش (۲.۱) رجوع کنید).

اساساً، هدف از حدس فوق، رتبه بندی شاخص‌های جینی ریسک‌های نرمال چندمتغیره با حاشیه‌های یکسان اما قدرت وابستگی متفاوت است. اثبات حدس، کار چالش‌برانگیزی است. در این پایان‌نامه، اثبات این حدس به صورت جزئی انجام می‌شود و همچنین نتیجه مربوطه به توزیع‌های بیضوی تعمیم داده می‌شود. اما برخی مسائل همچنان باز هستند. مقایسه شاخص‌های جینی ریسک‌های بیضوی چندمتغیره علاوه بر اینکه در کاربردهای آماری حائز اهمیت است، به خودی خود جالب توجه است. اساساً حدس بیان می‌کند که با افزایش $\boldsymbol{\Sigma}$ برای هر $t \geq 0$ ، احتمال $\mathbb{P}\{\frac{1}{n^2}\sum_{1\leq i,j\leq n}|X_i - X_j| \leq t\}$ افزایش می‌یابد. بنابراین، مطالعه حدس در حوزه مسئله تمرکز مرکزی توزیع‌های بیضوی به صورت زیر فرمول‌بندی می‌شود.

به ازای $C \in B(\mathbb{R}^n)$، احتمال زیر

$$\mathbb{P}_{\boldsymbol{\Sigma}}(C) = \mathbb{P}\{(X_1, \ldots, X_n) \in C\} \qquad (۱)$$

مشروط بر اینکه، $(X_1, \ldots, X_n)$ از توزیع بیضوی با میانگین $\mathbf{0}$ و ماتریس پراکندگی $\boldsymbol{\Sigma}$ تبعیت کند، چگونه با تغییر $\boldsymbol{\Sigma}$ تغییر می‌کند؟ این مسئله که برای اولین بار در مرجع [۶۶] مطالعه شده است، بیان می‌دارد که اگر $(X_1, \ldots, X_n)$ از توزیع نرمال چندمتغیره با میانگین $\mathbf{0}$ و ماتریس پراکندگی $\boldsymbol{\Sigma}$ تبعیت کند، همچنان که $\boldsymbol{\Sigma}$ به طور مولفه‌ای افزایش می‌یابد ولی مولفه‌های قطری آن بدون تغییر باقی بماند، آنگاه $\mathbb{P}_{\boldsymbol{\Sigma}}(C)$ برای هر مجموعه مستطیلی پایینی C، افزایش می‌یابد. مقالاتی، این مطالعه را به توزیع‌های بیضوی بسط دادند به شکلی که نواحی‌ای از شکل‌های مختلف را در نظر گرفتند، مثلاً مجموعه‌های مستطیلی بالایی، مستطیل‌ها و نواحی متقارن مرکزی و محدب. کسانی که به این موضوع علاقه‌مند هستند می‌توانند به مقالات [۱۷]، [۴۳] [۲۷] و [۹] مراجعه کنند. در همه این مطالعات، فرض‌های خاصی درباره ساختار ماتریس کواریانس در نظر گرفته شده است. نتایج بدست آمده در این پایان‌نامه، از این لحاظ به غنای مطالعات این حوزه می‌افزاید، که انتخاب مبسوط‌تری روی مجموعه C انجام می‌دهد.

به علاوه، اهمیت دیگر آن در مقایسه شاخص جینی است. این روش‌ها را می‌توان برای تعمیم ترتیب $pd-1$ به کار برد. ترتیب $pd-1$ در [۶۵] پیشنهاد شده است و برای مقایسه قدرت وابستگی بردارهای تصادفی تبادلی به کار رفته است. Chang در مرجع [۱۴] این مفهوم را از منظر تعمیم تصادفی گسترش داده است و کاربردهایی در حوزه تحقیقات عملیاتی برای آن یافته است [۱۴]. علاقه‌مندان می‌توانند برای مشاهده جمع‌بندی مطالعات انجام شده روی ترتیب $pd-1$ ، به فصل ۹ کتاب [۶۳] مراجعه کنند. در مقالات موجود، کاربرد ترتیب $pd-1$ بسیار محدود است، زیرا مقایسه تنها بر روی بردارهای تصادفی تبادلی انجام می‌شود. در این پایان‌نامه، ترتیب $pd-1$ به بردارهای تصادفی غیرتبادلی تعمیم داده می‌شود.

امروزه، بیمه گران به کمک ابزارهای آماری پیشرفته می‌توانند پروفایل ریسک بیمه‌گذاران را به صورت دقیق ارزیابی کنند. حق بیمه فنی مبتنی بر هزینه (بر خلاف حق بیمه تجاری مبتنی بر تقاضا که در واقع به بیمه‌گذاران پرداخت می‌شود)، باید ریسک هر بیمه‌گذار را مطابق با مشخصات فردی خودش، به صورت صحیح ارزیابی کند. اگر داده‌ها به گروه‌هایی تقسیم‌بندی شوند که مشخصات خاص خود را دارند، بیمه گران با رده‌های ریسک جمعیتی پراکنده مواجه می‌شوند، به گونه‌ای که میانگین‌گیری ساده مورد تردید قرار می‌گیرد و باید مدل‌های رگرسیونی تولید شوند. مدل‌های رگرسیونی، متغیر پاسخ را بر اساس تابع ویژگی‌ها (یا متغیرهای توصیفی[1]) و پارامترها پیش‌بینی می‌کنند. آنها از طریق ایجاد ارتباط بین پروفایل‌های ریسک مختلف، به مسائل کاملاً بخش‌بندی شده‌ای می‌پردازند که حاصل مقادیر عظیم اطلاعات موجود درباره بیمه‌گذاران است، که اکنون در اختیار بیمه‌گر قرار دارد. مدل‌های قیمت‌گذاری آماری معمولاً به گونه‌ای کالیبره می‌شوند که معیار نیکویی برازش بهینه شود (در اغلب موارد، انحراف یا درست‌نمایی لگاریتمی). ضوابط نیکویی برازش[2] بر اساس تبعیت از داده‌های مشاهده شده پیشین هستند. آنها حاوی خطاهای درون نمونه و خطاهای بیرون نمونه هستند که مورد دوم، به عنوان ضابطه عملکرد پیش‌بین نیز در نظر گرفته می‌شود. با این وجود، بیمه گران باید علاوه بر معیارهای آماری محض، معیارهای مرتبط با بیمه نیز داشته باشند به ویژه وقتی مدل‌های قیمت‌گذاری مختلف را مقایسه می‌کنند. در کاربردهای بیمه، باید ضوابط اقتصادی بیشتری مورد استفاده قرار گیرند تا بتوان تصمیم گرفت که آیا یک مدل ارزش پیاده‌سازی دارد یا خیر. یک مدل بسیار دقیق و در نتیجه‌گران‌قیمت که در مقایسه با مدل ساده موجود، بهره کمتری دارد، مطمئناً به فراموشی سپرده خواهد شد. مفهوم بالابری که توسط میرز و کامینگز پیشنهاد شده است، در این زمینه سودمند بوده است [۵۳]. به طور خلاصه، بالابری، توانایی مدل برای جلوگیری از انتخاب مغایر را اندازه‌گیری می‌کند. به عبارت دقیق‌تر، توانایی مدل برای انتساب یک آهنگ آماری مناسب به هر بیمه شده و در نتیجه کمینه کردن احتمال از دست رفتن تجارت توسط رقبا با استفاده از فهرست‌های قیمت کوچکتر را مشخص می‌کند. ارزیابی عملکرد حق بیمه انتخابی بر اساس دو ویژگی زیر است،

[1] Explanatory variable
[2] Goodness of fit

- تغییرپذیری مقادیر حق بیمه حاصله زیرا تغییرپذیری‌های بزرگتر حق بیمه، بالابری بیشتری را القا می‌کنند.
- توانایی درآمد حق بیمه برای انطباق با مورد صحیح به منظور افزایش پروفایل‌های ریسک.

اولی را می‌توان به کمک ترتیب محدب فرمول‌بندی کرد، که به کمک منحنی‌های لورنز مشخصه‌یابی می‌شود. این مفهوم معمولاً در احتمالات کاربردی برای مقایسه تغییرپذیری ذاتی توزیع‌های احتمال (ورای انحراف معیار) به کار می‌رود.
مورد دوم، به وسیله منحنی‌های تمرکز ارزیابی می‌شود. با در نظر گرفتن زیرپرتفوی که درصد مشخصی از بیمه‌نامه‌ها با کمترین میزان حق بیمه را گردآوری کرده است (یعنی آن دسته از بیمه‌گذارانی که به علت پروفایل ریسک پایین، در خطر هدف‌گیری توسط رقبا قرار می‌گیرند)، منحنی غلظت، سهم درآمد کل حق بیمه را متناظر با این گروه از قراردادها ارزیابی می‌کند و آن را با زیان‌های متناظر مقایسه می‌کند. بیمه‌گر می‌تواند به کمک مکان نسبی گراف‌های لورنز و منحنی‌های تمرکز، عملکرد حق بیمه مدنظر را به دقت ارزیابی کند. مقاله [۲۳] بر اساس مطالعات پیشین است که توسط گوریروکس ([۳۸]، همچنین [۳۹] را ببینید) و فریز و همکاران ([۳۱]،[۳۲]) انجام شده است. این نویسندگان، منحنی های عملکرد را ابتدا در امتیازدهی اعتباری و سپس برای مدل‌سازی زیان بیمه عمومی تعریف کردند. در مقایسه با کارهای پیشین، معیارهای جدیدی (مانند ABC و ICC که در بخش‌های بعدی توصیف می‌شوند) وجود دارند که آنها از نظر ویژگی‌های روش‌شناسی با استفاده از مفاهیم وابستگی و ترتیب تصادفی توسعه داده می‌شوند.

فصل ۱

پیش‌نیازها و تعاریف مقدماتی

۱.۱ مقدمه

دراین فصل به بیان مفاهیم و تعاریفی که در سرتاسر این پایان نامه به آن‌ها نیاز داریم، خواهیم پرداخت.

۲.۱ تعاریف مقدماتی

در سراسر این پایان‌نامه، از حروف برجسته برای نمایش بردار یا ماتریس تصادفی استفاده می‌کنیم. به عنوان مثال، $\mathbf{X} = (X_1, \ldots, X_n)$ یک بردار سطری است و $\mathbf{\Sigma} = (\sigma_{ij})_{n\times n}$ یک ماتریس $n \times n$ است. به طور خاص، نماد $\mathbf{0}$ بیانگر بردار سطری‌ای است که تمامی درایه‌های آن برابر با $\circ$ باشند و $\mathbf{1}_{n\times n}$ بیانگر یک ماتریس $n \times n$ است که تمامی درایه های آن برابر با ۱ باشند. نامساوی بین بردارها یا ماتریس ها، نامساوی مولفه ای را نشان می دهد. به عنوان مثال، $(x_1, \ldots, x_n) \leq (y_1, \ldots, y_n)$ بیانگر این است که برای هر $i = 1, \ldots, n$ ، رابطه $x_i \leq y_i$ برقرار است.

تعریف ۱.۲.۱. فرض کنید Ω یک مجموعه ناتهی و $F \subseteq 2^\Omega$ (2^Ω مجموعه توانی Ω یا مجموعه تمام زیرمجموعه‌های Ω است). در این‌صورت F را یک σـ میدان یا σـ جبر گوییم هر گاه،

- $\emptyset \in F$
- اگر $A \in F$ آنگاه $A^C \in F$
- اگر $A_1, A_2, A_3, \cdots \in F$ آنگاه $\bigcup_{i=1}^{\infty} A_i \in F$.

هر گاه F یک σـ میدان روی Ω باشد آنگاه به دوتایی (Ω, F) یک فضای اندازه‌پذیر گوییم.

نکته: اگر $\Omega = \mathbb{R}$ آنگاه σـ جبر تولید شده توسط تمام بازه‌های باز را σـ جبر بورل می‌نامیم و آن را با نماد $B(\mathbb{R})$ نمایش می‌دهیم.

تعریف ۲.۲.۱. فرض کنید (Ω, F) یک فضای اندازه‌پذیر باشد و $X : \Omega \to \mathbb{R}$ یک تابع در نظر گرفته شود به طوری که به ازای هر $B \in B(\mathbb{R})$ داشته باشیم، $X^{-1}(B) \in F$، در این‌صورت X را F اندازه‌پذیر می‌نامیم و به X یک متغیر تصادفی می‌گوییم.

تعریف ۳.۲.۱. فرض کنید Ω یک مجموعه ناتهی و F یک σـ میدان روی Ω باشد. هم‌چنین فرض کنید $\mathbb{P} : F \to [0, 1]$ یک تابع با خواص زیر در نظر گرفته شود.

- $\mathbb{P}(\Omega) = 1$
- به ازای هر دنباله از اعضای دو به دو مجزای F یعنی به ازای هر دنباله مانند $\{A_i\}_{i=1}^{\infty} \subseteq F$ با شرط $i \neq j \Rightarrow A_i \cap A_j = \emptyset$،

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

در این‌صورت $\mathbb{P}$ را اندازه احتمال می‌نامیم و به سه تایی $(\Omega, F, \mathbb{P})$ فضای احتمال گفته می‌شود.

تعریف ۴.۲.۱. فرض کنید $(\Omega, F, \mathbb{P})$ یک فضای احتمال باشد و $X, Y : \Omega \to \mathbb{R}$ دو متغیر تصادفی باشند در این‌صورت گوییم X, Y قریب به یقین با هم برابرند و می‌نویسیم $X = Y a.s.$ هرگاه،

$$\mathbb{P}(\{\omega \in \Omega | X(\omega) \neq Y(\omega)\}) = 0$$

تعریف ۵.۲.۱. • فرض کنید $(\Omega, F, \mathbb{P})$ یک فضای احتمال باشد و $X = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ یک متغیر تصادفی ساده نامنفی با نمایش استاندارد لحاظ شود. در این‌صورت انتگرال X نسبت به اندازه احتمال $\mathbb{P}$ به صورت، $\int_\Omega X \mathbf{d}\mathbb{P} = \sum_{i=1}^{n} \alpha_i \mathbb{P}(A_i)$ تعریف می‌شود.

- اگر $X : \Omega \to [0, \infty)$ یک متغیر تصادفی نامنفی باشد در این‌صورت انتگرال X نسبت به $\mathbb{P}$ به صورت زیر تعریف می‌شود،
  $\int_\Omega X \mathbf{d}\mathbb{P} = \sup \left\{ \int_\Omega \varphi \mathbf{d}\mathbb{P} | 0 \leq \varphi \leq X \text{که} \varphi \text{متغیر تصادفی ساده‌ای است} \right\}$
  منظور از $0 \leq \varphi \leq X$ یعنی به ازای هر $\omega \in \Omega$، $0 \leq \varphi(\omega) \leq X(\omega)$.

- اگر $X : \Omega \to \mathbb{R}$ یک متغیر تصادفی باشد به شرط آنکه حداقل یکی از دو انتگرال $\int_\Omega X^- \mathbf{d}\mathbb{P}$ یا $\int_\Omega X^+ \mathbf{d}\mathbb{P}$ متناهی باشد آنگاه تعریف می‌کنیم، $\int_\Omega X \mathbf{d}\mathbb{P} = \int_\Omega X^+ \mathbf{d}\mathbb{P} - \int_\Omega X^- \mathbf{d}\mathbb{P}$.

تعریف ۶.۲.۱. اگر $(\Omega, F, \mathbb{P})$ یک فضای احتمال و $X : \Omega \to \mathbb{R}$ یک متغیر تصادفی انتگرال پذیر باشد، آنگاه امید ریاضی و واریانس متغیر تصادفی X به صورت زیر تعریف می‌شود.

$$E(X) = \int_\Omega X \mathbf{d}\mathbb{P}$$
$$Var(X) = E\big((X - E(X))^۲\big)$$

به سادگی می‌توان نشان داد که $Var(X) = E(X^۲) - (E(X))^۲$.

تعریف ۷.۲.۱. فرض کنید $(\Omega, F, \mathbb{P})$ یک فضای احتمال و X، Y دو متغیر تصادفی باشند، در این‌صورت X، Y را هم توزیع می‌نامیم و می‌نویسیم $X \overset{d}{=} Y$ هر گاه،

$$\forall B \in B(\mathbb{R}), \mathbb{P}(X^{-۱}(B)) = \mathbb{P}(Y^{-۱}(B))$$

تعریف ۸.۲.۱. فرض کنید $(\Omega, F, \mathbb{P})$ یک فضای احتمال و $X_۱, X_۲, \ldots, X_n$، n متغیر تصادفی باشند آنگاه $\mathbf{X} = (X_۱, \ldots, X_n) : \Omega \to \mathbb{R}^n$ با ضابطه $\mathbf{X}(\omega) = (X_۱(\omega), \ldots, X_n(\omega))$ یک بردار تصادفی نامیده می‌شود.

تعریف ۹.۲.۱. اگر $\mathbf{X} = (X_۱, \ldots, X_n)$ یک بردار تصادفی از متغیرهای تصادفی انتگرال‌پذیر باشد آنگاه امید $\mathbf{X}$ با نماد $\mu_{\mathbf{X}}$ یا $E(\mathbf{X})$ نمایش داده می‌شود و عبارت‌است از

$$\mu_{\mathbf{X}} = E(\mathbf{X}) = (E(X_۱), \ldots, E(X_n))$$

تعریف ۱۰.۲.۱. برای متغیرهای تصادفی XوY که امید ریاضی آنها $E(X) = \mu$ و $E(Y) = \nu$ هستند کوواریانس به صورت زیر تعریف می‌شود،

$$Cov(X,Y) = E\big((X - E(X))(Y - E(Y))\big) = E((X - \mu)(Y - \nu)) = E(XY) - E(X)E(Y).$$

با توجه به این که اگر XوY مستقل باشند، رابطه $E(XY) = E(X)E(Y)$ برقرار است، بنابراین کوواریانس آنها صفر خواهد بود. واضح است که $Cov(X,X) = Var(X)$.

تعریف ۱۱.۲.۱. اگر XوY دو متغیر تصادفی باشند، آنگاه ضریب همبستگی(ضریب همبستگی پیرسون) آنها با نماد $\rho(X,Y)$ نمایش داده می‌شود و آن عبارت است از

$$\rho(X,Y) = \frac{Cov(X,Y)}{(Var(X)Var(Y))^{\frac{۱}{۲}}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

که در آنσ_Xوσ_Y به ترتیب انحراف معیار XوY هستند. به سادگی می‌توان نشان داد که، $\rho(X,Y) = \rho(Y,X)$.

ضریب همبستگی یکی از مشهورترین شیوه‌های اندازه‌گیری وابستگی بین دو متغیر تصادفی است.

در مثال زیر شیوه محاسبه رتبه n مقدار از یک متغیر تصادفی را بیان می‌کنیم.

مثال ۱.۲.۱. فرض کنید $x_۱ = ۰/۸$، $x_۲ = ۱/۲$، $x_۳ = ۱/۲$، $x_۴ = ۲/۳$ و $x_۵ = ۱۸$ مقادیر مربوط به یک متغیر تصادفی مانند X باشند که به‌صورت صعودی مرتب شده‌اند. مرتبه‌های آنها که به ترتیب با $r_{x_۱}, r_{x_۲}, r_{x_۳}, r_{x_۴}, r_{x_۵}$ نمایش داده می‌شوند، به‌صورت بیان شده در جدول ۱.۱ قابل محاسبه است.

جدول ۱.۱: مثال همبستگی رتبه‌ای اسپیرمن

مقدار متغیر x_i	موقعیت x_i در ترتیب صعودی	رتبه x_i
۰/۸	۱	۱
۱/۲	۲	$\frac{۲+۳}{۲} = ۲/۵$
۱/۲	۳	$\frac{۲+۳}{۲} = ۲/۵$
۲/۳	۴	۴
۱۸	۵	۵

تعریف ۱۲.۲.۱. اگر $x_۱, x_۲, \ldots, x_n$، n مقدار از متغیر تصادفی X باشند و $r_{x_۱}, r_{x_۲}, \ldots, r_{x_n}$ به ترتیب رتبه‌های آنها در نظر گرفته شوند و به‌طور مشابه $y_۱, y_۲, \ldots, y_n$، n مقدار از متغیر تصادفی Y با رتبه‌های $r_{y_۱}, r_{y_۲}, \ldots, r_{y_n}$ لحاظ شوند، آنگاه ضریب همبستگی رتبه‌ای اسپیرمن که به صورت r_s نشان داده می‌شود طبق رابطه زیر تعریف می‌شود.

$$r_s(X,Y) = \frac{Cov(r_x, r_y)}{\sigma_{r_x}\sigma_{r_y}} = \frac{\sum_i (r_{x_i} - \overline{r_x})(r_{y_i} - \overline{r_y})}{(\sum_i (r_{x_i} - \overline{r_x})^۲ \sum_i (r_{y_i} - \overline{r_y})^۲)^{\frac{۱}{۲}}}$$

تعریف ۱۳.۲.۱. فرض کنید زوج‌های $(x_۱, y_۱), (x_۲, y_۲), \ldots, (x_n, y_n)$ مشاهدات متغیرهای X و Y را تشکیل دهند. زوج (x_i, y_i) و (x_j, y_j) را «هماهنگ»[1] می‌گویند اگر $x_j \geq x_i$ آنگاه $y_j \geq y_i$. به بیان دیگر اگر داده‌های این زوج‌ها را براساس مولفه اول یا دوم مرتب کنیم، دارای رتبه‌های یکسانی خواهند بود. در حالت عکس این زوج‌ها را «ناهماهنگ»[2] می‌نامند.

حال براساس تعریف هماهنگ و ناهماهنگ برای زوج‌ها، اگر تعداد زوج‌های هماهنگ را با $|Con|$ و تعداد زوج‌های ناهماهنگ را نیز با $|Dis|$ نشان دهیم، ضریب هماهنگی تاو کندال به صورت زیر محاسبه می‌شود.

$$\tau = \frac{|Con| - |Dis|}{\frac{n(n-۱)}{۲}}$$

[1] Concordant
[2] Discordant

تعریف ۱۴.۲.۱. فرض کنید $(\Omega, F, \mathbb{P})$ یک فضای احتمال و $\mathbf{X} = (X_1, \ldots, X_n)$ یک بردار تصادفی باشد به طوری‌که به ازای هر $1 \leq i \leq n$ داشته باشیم $E(X_i) < \infty$ و $Var(X_i) < \infty$. در این‌صورت ماتریس کوواریانس $\mathbf{X}$ که با $\mathbf{\Sigma}$ نمایش داده می‌شود یک ماتریس $n \times n$ است که درایه iام و ستون jام آن عبارت است از $Cov(X_i, X_j)$. به عبارت دیگر،

$$\mathbf{\Sigma} = \left[Cov(X_i, X_j)\right]_{n\times n}$$

همچنین می‌توان گفت،

$$\mathbf{\Sigma} = Var(\mathbf{X}) = Cov(\mathbf{X}, \mathbf{X}) = E((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T) = E(\mathbf{X}\mathbf{X}^T) - \mu_{\mathbf{X}}\mu_{\mathbf{X}}{}^T$$

تعریف ۱۵.۲.۱. فرض کنید $\mathbf{X} = (X_1, \ldots, X_n)$ یک بردار تصادفی دارای توزیع نرمال با میانگین $\circ$ و ماتریس کوواریانس $\mathbf{\Sigma}$ باشد. در این‌صورت می‌نویسیم، $\mathbf{X} \sim MVN(\circ, \mathbf{\Sigma})$.

تعریف ۱۶.۲.۱. فرض کنید $(\Omega, F, \mathbb{P})$ یک فضای احتمال و $X : \Omega \to \mathbb{R}$ یک متغیر تصادفی در نظر گرفته شود در این‌صورت،

$$\begin{aligned}
&\varphi_X : \mathbb{R} \to \mathbb{C}\\
&\varphi_X(t) = E(e^{itX})\\
&\quad = \int_{\mathbb{R}} e^{itx}\mathrm{d}F_X(x)\\
&\quad = \int_{\mathbb{R}} e^{itx} f_X(x)\mathrm{d}x
\end{aligned}$$

تابع مشخصه متغیر تصادفی X تعریف می‌شود.

تابع مشخصه یک روش دیگری برای توصیف یک متغیر تصادفی است. همانند تابع توزیع تجمعی $F_X(x) = E(\chi_{\{X\leq x\}})$ که به طور کامل رفتار و ویژگی‌های متغیر تصادفی X را تعیین می‌کند، تابع مشخصه $\varphi_X(t) = E(e^{itX})$ به طور کامل رفتار و ویژگی‌های توزیع احتمال متغیر X را مشخص می‌نماید. این دو روش برای شناسایی معادل هستند، به این معنی که شناسایی یکی از آن‌ها منجر به پیدا کردن دیگری خواهد بود.

تعریف ۱۷.۲.۱. فرض کنید $\mathbf{X}=(X_1, \ldots, X_n)$ یک بردار تصادفی در $\mathbb{R}^n$ باشد، در این‌صورت برای هر $\mathbf{t}=(t_1, \ldots, t_n)$، تابع با ضابطه

$$\begin{aligned}
&\varphi_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{C}\\
&\varphi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}\mathbf{X}'})
\end{aligned}$$

که در آن $\mathbf{X}'$ ترانهاده بردار $\mathbf{X}$ است، تابع مشخصه بردار تصادفی $\mathbf{X}$ نامیده می‌شود.

نکته: هرگاه $\mathbf{X}=(X_1, \ldots, X_n)$ یک بردار تصادفی باشد و $\varphi_{\mathbf{X}}$ تابع مشخصه $\mathbf{X}$ در نظر گرفته

شود و $\mathbf{t}=(t_1,\dots,t_n)$ آنگاه،

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{t}\mathbf{X}'}(1)$$
$$\varphi_{\mathbf{X}}((t,\dots,t)) = \varphi_{X_1+X_2+\cdots+X_n}(t)$$
$$\varphi_{\mathbf{X}}((t,0,\dots,0)) = \varphi_{X_1}(t)$$

تعریف ۱.۲.۱۸. فرض کنید $\mathbf{X}$ یک ماتریس تصادفی $m \times n$ باشد. در این‌صورت،

$$\varphi_{\mathbf{X}} : \mathbb{R}^{m\times n} \to \mathbb{C}$$
$$\varphi_{\mathbf{X}}(\mathbf{t}) = E(e^{itr(\mathbf{t}'\mathbf{X})})$$

تابع مشخصه ماتریس تصادفی $\mathbf{X}$ نامیده می‌شود. که در آن $tr(\mathbf{t}'\mathbf{X})$ اثر ماتریس مربعی $t'\mathbf{X}$ می‌باشد.

تعریف ۱.۲.۱۹. ماتریس $M_{n\times n}$ را معین مثبت می‌نامیم هر گاه،

$$\forall \mathbf{X} \in \mathbb{R}^n - \{0\}, \mathbf{X}M\mathbf{X}' > 0$$

دقت کنید که هر $\mathbf{X} \in \mathbb{R}^n$ ماتریس $1 \times n$ به فرم $\mathbf{X} = (x_1, x_2, \dots, x_n)$ است.

تعریف ۱.۲.۲۰. ماتریس $M_{n\times n}$ را نیمه معین مثبت می‌نامیم هرگاه،

$$\forall \mathbf{X} \in \mathbb{R}^n, \mathbf{X}M\mathbf{X}' \geq 0$$

ماتریس نیمه معین مثبت را معین نامنفی هم می‌نامیم.

تعریف ۱.۲.۲۱. بردار تصادفی n بعدی $\mathbf{X}$ دارای توزیع بیضوی[1] است، هرگاه تابع مشخصه آن به‌صورت زیر باشد.

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}\mathbf{X}'}) = e^{i\mathbf{t}\mu'}\psi(\mathbf{t}\Sigma\mathbf{t}') \qquad (1.1)$$

که در آن $\mu \in \mathbb{R}^{\mathbf{n}}$، $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ یک ماتریس نیمه‌معین[2] مثبت و $\psi : [0,\infty) \to \mathbb{R}$ مولد مشخصه توزیع بیضوی است. μ که همان میانگین X در نظر گرفته می‌شود بردار مکان و $\mathbf{\Sigma}$ ماتریس کوواریانس(پراکندگی) نامیده می‌شوند.
هرگاه $\mathbf{X}$ دارای توزیع بیضوی با پارامترهای $\mathbf{\Sigma}, \psi$ و μ باشد، آنگاه می‌نویسیم $\mathbf{X} \sim EC_n(\mu, \mathbf{\Sigma}, \psi)$.

نکته: هرگاه $\mathbf{X} \sim EC_n(\mu, \mathbf{\Sigma}, \psi)$ و $\psi : [0,\infty) \to \mathbb{R}$ باضابطه $\psi(Z) = e^{-\frac{z}{2}}$ اختیار شود آنگاه $\mathbf{X} \sim N_n(\mu, \mathbf{\Sigma})$. یعنی $\mathbf{X}$ دارای توزیع نرمال با پارامترهای μ و $\mathbf{\Sigma}$ خواهد بود.

[1] Elliptical distribution
[2] Semidefinite matrix

تعریف ۲۲.۲.۱. فرض کنید

$$\mathbf{\Psi_n} = \{\psi : [0,\infty) \to \mathbb{R} | \text{با ضابطه } \mathbf{t} \to \psi(\|\mathbf{t}\|^2) \text{ یک تابع مشخصه} n \text{بعدی باشد}\} \text{تابع} \mathbb{R}^n \to \mathbb{R}$$

به وضوح $\mathbf{\Psi}_n \supseteq \cdots \supseteq \mathbf{\Psi}_3 \supseteq \mathbf{\Psi}_2 \supseteq \mathbf{\Psi}_1$. اعضای $\mathbf{\Psi}_n$ توابع مشخصه nبعدی تولید می‌کنند.
توابعی چون $\psi : [0,\infty) \to \mathbb{R}$ که توسط آنها، توابع مشخصه با بعدهای دلخواه تولید شوند را با نماد $\mathbf{\Psi}_\infty$ نمایش می‌دهیم.

McNeil و همکارانش (۲۰۰۵) خاطر نشان ساختند که در کل، مولدهای مشخصه را تنها می توان در ابعاد خاص مورد استفاده قرار داد[۵۲]. در این پایان‌نامه، ما بر رده خاصی از مولدها و توزیع‌های بیضوی ناشی از این رده متمرکز می‌شویم. خانواده توزیع بیضوی ناشی از Ψ_∞ ، بسیاری از توزیع‌های مهم از جمله توزیع نرمال چندمتغیره و توزیع t چندمتغیره را در بر می‌گیرد.

تعریف ۲۳.۲.۱. بردار تصادفی $\mathbf{X}=(X_1,\ldots,X_n)$ را در نظر بگیرید. شاخص جینی آن به صورت $\frac{1}{n^2}\sum_{1\le i,j\le n}|X_i - X_j|$ تعریف می شود. شاخص جینی، میزان پراکندگی مولفه های بردار تصادفی را اندازه گیری می‌کند. به عنوان مثال، اگر همه مولفه ها برابر باشند، آنگاه شاخص جینی صفر است . برای سادگی، از نمادگذاری زیر استفاده می کنیم،

$$G(\mathbf{X}) = \sum_{1\le i,j\le n} |X_i - X_j| \qquad (۲.۱)$$

$G(\mathbf{X})$ شاخص جینی مقیاس شده است و همان متغیر تصادفی‌ای است که ما می خواهیم در سراسر این پایان‌نامه، مطالعه کنیم. به آسانی می توان دید که $G(\mathbf{X})$ را می توان بر حسب آماره ترتیبی[1] به صورت زیر بازنویسی کرد،

$$G(\mathbf{X}) = \sum_{i=1}^{n} (4i - 2n - 2) X_{(i)} \qquad (۳.۱)$$

که در آن، $X_{(i)}$، i امین، مولفه بزرگ $\{X_1,\ldots,X_n\}$ را نشان می دهد.
برای مقایسه شاخص های جینی، برخی از تعاریف ترتیب تصادفی را ارائه می کنیم.

تعریف ۲۴.۲.۱. فرض کنید X و Y دو متغیر تصادفی باشند. X را در ترتیب تصادفی معمولی[2] کوچکتر از Y می گوییم (که به صورت $X \le_{st} Y$ نشان داده می شود)، اگر برای همه $t \in R$ ها داشته باشیم، $\mathbb{P}\{X > t\} \le \mathbb{P}\{Y > t\}$.
X را در ترتیب محدب صعودی[3] ، کوچکتر از Y می گوییم (که به صورت $X \le_{icx} Y$ نشان داده می شود)، اگر برای هر تابع محدب صعودی u داشته باشیم، $E\big(u(X)\big) \le E\big(u(Y)\big)$ مشروط بر اینکه امیدهای ریاضی موجود باشد.

[1] Order statistics
[2] stochastic order
[3] Usual stochastic order

در این پایان‌نامه، مشخصات زیر نیز برای ترتیب تصادفی معمولی بیان شده است.

گزاره ۱.۲.۱. فرض کنید X و Y دو متغیر تصادفی باشند. $X \leq_{st} Y$ است اگر و تنها اگر $E(u(X)) \leq E(u(Y))$ برای هر تابع صعودی u برقرار باشد، مشروط به اینکه امید های ریاضی مربوطه موجود باشند.

به علاوه، به منظور مقایسه بردارهای تصادفی، مفهوم ترتیب فوق‌مدولی[1] موردنیاز است.

تعریف ۲۵.۲.۱. تابع $f : \mathbb{R}^n \to \mathbb{R}$ را فوق‌مدولی می‌گوییم اگر برای هر $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ رابطه زیر برقرار باشد.

$$f(\mathbf{x}) + f(\mathbf{y}) \leq f(\mathbf{x} \wedge \mathbf{y}) + f(\mathbf{x} \vee \mathbf{y})$$

که در آن، عملگرهای $\wedge$ و $\vee$، به ترتیب کمینه و بیشینه مولفه‌ای هستند، یعنی

$$(x_1, \ldots, x_n) \wedge (y_1, \ldots, y_n) = \big(min(x_1, y_1), \ldots, min(x_n, y_n)\big)$$
$$(x_1, \ldots, x_n) \vee (y_1, \ldots, y_n) = \big(max(x_1, y_1), \ldots, max(x_n, y_n)\big)$$

بردار تصادفی $\mathbf{X}$ را در ترتیب فوق‌مدولی، کوچکتر از $\mathbf{Y}$ می‌گوییم (که به صورت $\mathbf{X} \leq_{sm} \mathbf{Y}$ نمایش داده می‌شود)، اگر برای هر تابع فوق‌مدولی f داشته باشیم $E(f(\mathbf{X})) \leq E(f(\mathbf{Y}))$ مشروط بر اینکه امیدهای ریاضی مربوطه موجود باشد.

به آسانی می‌توان ثابت کرد که اگر $\mathbf{X} \leq_{sm} \mathbf{Y}$ و $\mathbf{X} \geq_{sm} \mathbf{Y}$ برقرار باشند، آنگاه $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ برقرار است، که در آن، $\stackrel{d}{=}$ بیانگر «تساوی در توزیع» است. مطابق با مقاله [۴۸]، ادعای (۱)، نتیجه زیر حاصل می شود.

گزاره ۲.۲.۱. تابع $f : \mathbb{R}^n \to \mathbb{R}$ فوق‌مدولی است، اگر و تنها اگر $f(x_1, \ldots, x_n)$ به عنوان تابعی از (x_i, x_j) برای هر x_k ثابت فوق‌مدولی باشد، که در آن برای هر $1 \leq i \leq j \leq n$ داشته باشیم $k \neq i, j$.

برای بررسی اثر وابستگی آماره آزمون به آزمون فرض پیشنهاد شده توسط [۱۱]، کافی است که توزیع نرمال چندمتغیره را مطالعه کنیم، زیرا توزیع مجانبی، مقادیر ریسک تجربی است. با این وجود، تنها از روی علاقه، می‌خواهیم مطالعه خود را به توزیع‌های بیضوی بسط دهیم. بنابراین، ابتدا برخی مفاهیم پایه راجع به توزیع های بیضوی را ارائه می‌کنیم. تعریف و مشخصات زیر برای توزیع بیضوی از مقاله [۵۲] گرفته شده است.

برای بحث بیشتر راجع به این خانواده، به فصل سوم [۵۲] مراجعه کنید. به علاوه، یک ویژگی مفید این خانواده این است که یک نمایش تصادفی بر حسب توزیع نرمال چندمتغیره دارد که به صورت گزاره ۳.۲.۱ بیان شده است. گزاره ۳.۲.۱، اساساً ترکیبی از قضیه ۳ – ۲۵ و تعریف ۳ – ۲۶ در [۵۲] است بنابراین، اثبات آنها ارائه نشده است.

[1] Supermodular

گزاره ۳.۲.۱. بردار تصادفی $\mathbf{X} \sim EC_n(\mu, \mathbf{\Sigma}, \psi)$ با $\psi \in \mathbf{\Psi}_\infty$ موجود است، اگر و تنها اگر بردار تصادفی $\mathbf{Z}$ و متغیر تصادفی R به گونه ای موجود باشند که رابطه زیر برآورده شود،

$$\mathbf{X} \overset{d}{=} \mu + R\mathbf{Z} \qquad (۴.۱)$$

که در آن $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{\Sigma})$ و $R \geq 0$ یک متغیر تصادفی مستقل از $\mathbf{Z}$ است.

گزاره ۳.۲.۱ یک رابطه مهم بین توزیع های بیضوی و نرمال چندمتغیره را نشان می‌دهد. با این نحوه نمایش، می‌توان بسیاری از خصوصیات توزیع نرمال چندمتغیری را به توزیع بیضوی تعمیم داد. در بخش‌های بعدی چند مثال ارائه می شود.
این پایان‌نامه، راجع به ساختار وابستگی است. هم‌یکنوایی ، یک وابستگی مثبت کامل است و کاربردهای مهمی در علوم آماری و دانش مالی دارد. Dhaene و همکارانش (۲۰۰۲) یک مطالعه مقایسه‌ای روی مفهوم هم‌یکنوایی و کاربردهای آن انجام دادند [۲۵]. در ذیل، تعاریف آنها و چندین مشخصه هم‌ارز هم‌یکنوایی ارائه می‌شوند.

تعریف ۲۶.۲.۱. مجموعه $A \subset \mathbb{R}^n$ را هم‌یکنوا می‌نامیم اگر برای هر $\mathbf{x}, \mathbf{y} \in A$، رابطه $\mathbf{x} \leq \mathbf{y}$ یا $\mathbf{y} \leq \mathbf{x}$ برقرار باشند. مسلم است که یک مجموعه هم‌یکنوا [1] است اگر و تنها اگر به صورت کلی مرتب باشد.

تعریف ۲۷.۲.۱. برای یک بردار تصادفی، تکیه‌گاه آن به صورت زیر تعریف می‌شود،

$$supp(\mathbf{X}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbb{P}\{\mathbf{X} \in B(x, r)\} > 0, \forall r > 0\}, \qquad (۵.۱)$$

که در آن، $B(\mathbf{x}, r)$ گوی‌هایی با مرکز x و شعاع r است.

تعریف ۲۸.۲.۱. بردار تصادفی X هم‌یکنوا است، اگر تکیه‌گاه آن هم‌یکنوا باشد.

Dhaene و همکارانش (۲۰۰۲) چندین مشخصه معروف هم‌یکنوایی را توسعه داده اند [۲۵]. به عنوان مثال، عبارات ذیل هم‌ارز هستند.

- $\mathbf{X} = (X_1, \ldots, X_n)$ هم‌یکنوا است.
- یک متغیر تصادفی Z و توابع صعودی $f_1, \ldots, f_n$ وجود دارند به گونه‌ای که، رابطه $(X_1, \ldots, X_n) \overset{d}{=} (f_1(Z), \ldots, f_n(Z))$ برقرار است.
- برای همه $(x_1, \ldots, x_n) \in \mathbb{R}^n$ داریم،
$$\mathbb{P}\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = \min\{\mathbb{P}\{X_1 \leq x_1\}, \ldots, \mathbb{P}\{X_n \leq x_n\}\}$$

به علاوه، قضیه ۵ در مقاله [۲۵] یک مشخصه هم‌یکنوایی را برای توزیع نرمال چندمتغیری $\mathbf{X} = (X_1, \ldots, X_n) \sim MVN(\mu, \mathbf{\Sigma})$ از طریق ماتریس کوواریانس آن ارائه می‌کند. به طور خاص،

[1] Comonotonicity

هم‌یکنوا است اگر و تنها اگر $corr(X_i, X_j) = ۱$ برای همه i, j ها برقرار باشد (یعنی $rank(\mathbf{\Sigma}) = ۱$). به علاوه، اگر همه توزیع‌های حاشیه‌ای، واریانس مشابه ۱ داشته باشند، آنگاه هم‌یکنوایی $\mathbf{X}$ هم ارز با $\mathbf{\Sigma} = ۱_{n\times n}$ است.

در واقع، این مشخصه را می توان به توزیع های بیضوی ناشی از Ψ_∞ بسط داد. در حالت خاص، توزیع بیضوی با $\psi \in \Psi_\infty$ ، هم یکنواست اگر و تنها اگر ماتریس پراکندگی، رتبه ۱ داشته باشد. McNeil و همکارانش (۲۰۰۵)، این حقیقت را برای توزیع t چندمتغیره بیان داشتند [۵۲] (در فصل ۵). در ذیل، این مشخصه به صورت رسمی بیان می شود و اثبات آن در حالت عمومی داده می شود.

گزاره ۴.۲.۱. فرض کنید $\mathbf{X} \sim EC_n(\mu, \mathbf{\Sigma}, \psi)$ با $\psi \in \Psi_\infty$ برقرار هستند. $\mathbf{X}$ هم‌یکنواست اگر و تنها اگر $rank(\mathbf{\Sigma}) = ۱$ باشد.

برهان. با توجه به رابطه (۱.۲.۲)، $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{\Sigma})$ و $R \geq ۰$ مستقل از $\mathbf{Z}$ وجود دارد به گونه ای که رابطه $\mathbf{X} \overset{d}{=} \mu + R\mathbf{Z}$ برقرار است. بخش «اگر». فرض کنید $rank(\mathbf{\Sigma}) = ۱$ باشد، آنگاه $corr(Z_i, Z_j) = ۱$ برای همه $jوi$ ها برقرار است. بنابراین، $Z \sim N(۰, ۱)$ وجود دارد به گونه‌ای که $Z_i = a_i Z$ با $a_i \geq ۰$ برای همه $i = ۱, \ldots, n$ ها برقرار باشد. بلافاصله نتیجه می‌شود که $\mathbf{X}$ با توجه به نمایش تصادفی ۲.۲.۲ هم‌یکنوا است و مشخصا تابعی هم‌یکنوا بدست می‌آید.
بخش «تنها اگر». فرض کنید $\mathbf{X}$ هم یکنوا باشد. $\mathbf{y}, \mathbf{z} \in supp(\mathbf{Z})$ را در نظر بگیرید. از آنجا که R مستقل از $\mathbf{Z}$ است، بنابراین برای هر $۰ < r \in supp(\mathbf{Z})$ ، $\mu + r\mathbf{y}, \mu + r\mathbf{z} \in supp(\mathbf{X})$. از هم یکنوایی $\mathbf{X}$ نتیجه می‌گیریم که $\mu + r\mathbf{y} \leq \mu + r\mathbf{z}$ یا $\mu + r\mathbf{y} \geq \mu + r\mathbf{z}$ که دلالت بر رابطه $\mathbf{y} \leq \mathbf{z}$ یا $\mathbf{y} \geq \mathbf{z}$ دارد. بنابراین، نتیجه می‌گیریم که $\mathbf{Z}$ هم‌یکنوا است و در نتیجه $rank(\mathbf{\Sigma}) = ۱$ است. □

۳.۱ تابع مفصل

می‌توان گفت تاریخچه مفصل توسط فرشه در مرجع [۳۰] آغاز شد. او مسئله زیر را مطالعه کرد، که در اینجا به صورت ۲ بعدی بیان شده است. اگر توابع توزیع $F_۱$ و $F_۲$ مربوط به دو متغیر تصادفی $X_۱$ و $X_۲$ که روی فضای احتمال یکسان $(\Omega, F, \mathbb{P})$ تعریف شده اند باشد، آنگاه در مورد مجموعه $\Gamma(F_۱, F_۲)$ متشکل از توابع توزیع توام $X_۱$ ، $X_۲$ که توابع حاشیه‌ای آنها $F_۱$ ، $F_۲$ باشد، چه می‌توان گفت؟.
مجموعه $\Gamma(F_۱, F_۲)$ کلاس فرشه $F_۱$ و $F_۲$ نامیده می‌شود. به‌عنوان مثال اگر $X_۱$ و $X_۲$ مستقل از هم باشند و $F_۱$ و $F_۲$ به ترتیب توابع توزیع آنها در نظر گرفته شوند، آنگاه $F : \mathbb{R}^۲ \to [۰, ۱]$ با ضابطه $F(x_۱, x_۲) = F_۱(x_۱)F_۲(x_۲)$ تابع توزیع توام آنهاست، لذا $F \in \Gamma(F_۱, F_۲)$، این را نشان می‌دهد که $\Gamma(X_۱, X_۲) \neq \emptyset$.

مطالعات جزئی در مورد مسئله فوق توسط افرادی انجام شد، اما در سال ۱۹۵۹ اسکلار [1] عمیق‌ترین نتیجه در ارتباط با این مسئله را با معرفی مفهومی به نام مفصل بدست آورد و در این خصوص قضیه‌ای اثبات نمود که به نام خودش معروف شد.

تعریف ۱.۳.۱. متغیر تصادفی X دارای توزیع یکنواخت روی بازه $[a,b]$ است، هرگاه تابع چگالی احتمال آن به‌صورت زیر تعریف شود.

$$f = \frac{۱}{b-a}\chi_{[a,b]}$$

به عبارت دیگر

$$f(t) = \begin{cases} \frac{۱}{b-a} & a < t < b \\ ۰ & \text{در غیر اینصورت} \end{cases}$$

همچنین هرگاه X دارای توزیع یکنواخت روی $[a,b]$ باشد می‌نویسیم $X \sim \mathcal{U}([a,b])$.

تعریف ۲.۳.۱. فرض کنید $n \geq ۲$ یک عدد طبیعی باشد. در این‌صورت یک مفصل $n-$بعدی (یا بطور خلاصه یک $n-$مفصل) یک تابع توزیع $n-$متغیره روی $[۰,۱]^n$ مانند C است، که توابع توزیع حاشیه‌ای آن یکنواخت استاندارد روی $[۰,۱]$ هستند.
بنابراین هر $n-$مفصل می‌تواند به یک بردار تصادفی مانند $\mathbf{U} = (U_۱, U_۲, \ldots, U_n)$ مربوط شود به‌طوریکه به ازای هر $۱ \leq i \leq n$، $U_i \sim \mathcal{U}([۰,۱])$ و $\mathbf{U} \sim C$.
همچنین هر بردار تصادفی که مولفه‌های آن روی $[۰,۱]$ دارای توزیع یکنواخت باشند، متناظر با یک مفصل می‌باشد.

قضیه ۱.۳.۱. تابع $C : [۰,۱]^n \to [۰,۱]$ یک مفصل $n-$بعدی است، اگر و فقط اگر موارد زیر برقرار باشند.

- برای هر $۱ \leq j \leq n$، $C(۱, \ldots, ۱, u_j, ۱, \ldots, ۱) = u_j$
- به ازای هر $u, v \in I^n$ که $u \leq v$ ، $C(u) \leq C(v)$
- هرگاه به ازای هر $۱ \leq i \leq n$ داشته باشیم $a_i \leq b_i$ آنگاه

$$\sum_{i_۱=۱}^{۲} \cdots \sum_{i_n=۱}^{۲} (-۱)^{i_۱+\cdots+i_n} C(u_{۱,i_۱}, \ldots, u_{n,i_n}) \geq ۰$$

که در آن برای $۱ \leq j \leq ۲$، $a_j = u_{j,۱}, b_j = u_{j,۲}$.

مثال ۱.۳.۱. • اگر $C : [۰,۱]^n \to [۰,۱]$، با ضابطه $C(x_۱, x_۲, \ldots, x_n) = x_۱ x_۲ \ldots x_n$ متناظر با بردار تصادفی $\mathbf{U} = (U_۱, U_۲, \ldots, U_n)$ باشد که مولفه‌های آن مستقل از هم و دارای توزیع یکنواخت روی $[۰,۱]$ باشند، آنگاه C را مفصل مستقل می‌نامند.

[1] Sklar

- اگر $C : [۰,۱]^n \to [۰,۱]$، با ضابطه $C(x_۱,\dots,x_n) = \min\{x_۱,\dots,x_n\}$ متناظر با بردار تصادفی $\mathbf{U} = (U_۱, U_۲, \dots, U_n)$ باشد که مولفه‌هایش دارای توزیع یکنواخت روی $[۰,۱]$ و $U_۱ = U_۲ = \cdots = U_n a.s.$ آنگاه C را مفصل هم‌یکنوا می‌نامند.

- اگر $C : [۰,۱]^۲ \to [۰,۱]$، با ضابطه $C(x_۱,x_۲) = \max\{x_۱ + x_۲ - ۱, ۰\}$ متناظر با بردار تصادفی $\mathbf{U} = (U_۱, U_۲)$ باشد، که مولفه‌های آن دارای توزیع یکنواخت روی $[۰,۱]$ هستند و $U_۱ = ۱ - U_۲ a.s.$ آنگاه C را مفصل عدم یکنوایی می‌نامند.

قضیه ۱.۲.۳. اسکلار:[1] تابع توزیع تجمعی $n-$بعدی F با حاشیه‌های $F_۱, F_۲, \dots, F_n$ را در نظر بگیرید. در این‌صورت مفصل $C : [۰,۱]^n \to [۰,۱]$ چنان موجود است، که برای هر $۱ \le i \le n$، $x_i \in (-\infty, \infty)$

$$F(x_۱,\dots,x_n) = C(F_۱(x_۱),\dots,F_n(x_n)).$$

هر گاه برای هر $۱ \le i \le n$، F_i پیوسته باشد، آنگاه مفصل C منحصر به‌فرد است.
هرگاه F_iها پیوسته نباشند، آنگاه C روی $Ran(F_۱) \times \cdots \times Ran(F_n)$ به‌طور منحصر به‌فرد تعریف می‌شود.

تعریف ۱.۳.۳. یک مولد ارشمیدسی[2] تابعی نزولی و پیوسته مانند $\psi : [۰,\infty] \to [۰,۱]$ است، که در شرایط زیر صدق کند.

- $\psi(۰) = ۱$
- $\lim_{t\to\infty} \psi(t) = ۰$
- روی $[۰, \inf\{t|\psi(t) = ۰\})$ اکیدا نزولی باشد.

تعریف ۱.۴.۳. یک مفصل $n-$ متغیره C را ارشمیدسی[3] می‌نامیم، هرگاه برای هر $\mathbf{U} \in [۰,۱]^n$ داشته باشیم.

$$C(\mathbf{U}) = \psi(\psi^{-۱}(u_۱) + \psi^{-۱}(u_۲) + \cdots + \psi^{-۱}(u_n))$$

که در آن $\psi[۰,\infty] \to [۰,۱]$ تابع مولد ارشمیدسی است.
همچنین معکوس آن به صورت زیر است.

$$\psi^{-۱}(t) = \begin{cases} \psi^{-۱}(t) & ۰ \le t \le \psi(۰) \\ ۰ & \psi(۰) \le t \le \infty \end{cases} \qquad (۱.۱)$$

[1] Sklar's Theorem
[2] Archimedean generator
[3] Archimedean copula

تعریف ۵.۳.۱. تابع مفصل کلایتون[1] برای دو متغیر به شکل زیر تعریف می‌شود.

$$C_\theta^{clayton}(u,v) = \max\{(u^{-\theta}+v^{-\theta}-۱)^{-\frac{۱}{\theta}}, ۰\}$$

و تابع مولد ارشمیدسی این خانواده از مفصل‌ها توسط

$$\psi_\theta(t) = \big(\max\{۱+\theta t, ۰\}\big)^{-\frac{۱}{\theta}}$$

بیان می‌شود، که در آن $\theta \in [-۱, +\infty]$ خواهد بود. تابع مفصل کلایتون دارای توزیع نامتقارن است، به‌طوریکه در این تابع وابستگی به دنباله منفی از وابستگی به دنباله مثبت بیشتر است.

نکته: هر گاه $\theta > ۰$ آنگاه مفصل کلایتون به‌صورت زیر تبدیل می‌شود.

$$C_\theta^{clayton}(u,v) = (u^{-\theta}+v^{-\theta}-۱)^{-\frac{۱}{\theta}}$$

تعریف ۶.۳.۱. تابع مفصل $n-$بعدی فرانک[2] به صورت زیر تعریف می‌شود.

$$C_\theta^{Frank}(u) = -\frac{۱}{\theta}\log\big(۱+\frac{\Pi_۱^n(e^{-\theta u_i}-۱)}{(e^{-\theta}-۱)^{n-۱}}\big)$$

که در آن $\theta > ۰$ است. وقتی $\theta \to ۰^+$ آنگاه مفصل فرانک به مفصل مستقل تبدیل می‌شود. در حالت $n = ۲$ پارامتر θ می‌تواند برای $\theta < ۰$ تعریف شود. تابع مولد ارشمیدسی این خانواده از مفصل‌ها توسط $\psi_\theta(t) = -\frac{۱}{\theta}\log\big(۱-(۱-e^{-\theta})e^{-t}\big)$ تعریف می‌شود.

تعریف ۷.۳.۱. تابع مفصل دومتغیره گامبل[3] به صورت زیر تعریف می‌شود.

$$C_\theta^{Gu}(u,v) = e^{-\left((-\ln u)^\theta+(-\ln v)^\theta\right)^{\frac{۱}{\theta}}}$$

که در آن $\theta \in [۱, \infty)$. در حالت $\theta = ۱$ مفصل گامبل به مفصل مستقل تبدیل می‌شود. در حالتی که $\theta \to \infty$ مفصل گامبل به مفصل هم‌یکنوا تبدیل خواهد شد.
تابع مولد ارشمیدسی این خانواده از مفصل‌ها توسط $\psi(t) = e^{-t^{\frac{۱}{\theta}}}$ تعریف می‌شود.

۴.۱ رگرسیون خطی

واژه رگرسیون به معنی بازگشت است و نشان می‌دهد که مقدار یک متغیر به متغیر دیگری بر می‌گردد. روابطی که بتوان با آن میزان متغیری را با توجه به میزان دو یا چند متغیر دیگر پیدا کرد مرتبط به رگرسیون چندگانه خواهد بود.
در رگرسیون به‌دنبال برآورد رابطه‌ای ریاضی و تحلیل آن هستیم، به‌طوری که با آن بتوان کمیت

[1] Clayton copula
[2] Frank copula
[3] Gumbel copula

متغیری مجهول را با استفاده از متغیر یا متغیرهای معلوم تعیین کنیم. سپس در همبستگی به دنبال تعیین نوع رابطه و میزان ارتباطی هستیم، که متغیرها را به هم ربط می‌دهد.
به عنوان مثال اگر X، Y دو متغیر تصادفی باشند، به‌طوری که $Y = a + bX$ در این‌صورت،

- اگر $b > 0$ آنگاه X، Y دارای رابطه مستقیم هستند، یعنی با افزایش یکی دیگری نیز افزایش می‌یابد و با کاهش یکی دیگری نیز کاهش می‌یابد، لذا رابطه بین X، Y خطی مستقیم است.
- اگر $b < 0$ در این‌صورت رابطه بین دو متغیر معکوس است. بدین معنی که با افزایش یکی دیگری کاهش و با کاهش یکی دیگری افزایش می‌یابد. به‌عبارت دیگر رابطه بین X، Y خطی معکوس است.
- اگر $b = 0$ نشان‌دهنده آن است، که دو متغیر X، Y رابطه خطی نداشته و مستقل هستند.

۵.۱ مفاهیم مربوط به حق بیمه‌های خالص و کاری

۱.۵.۱ فرضیات

پاسخ Y و مجموعه ویژگی‌های $X_1, \ldots, X_p$ که تحت بردار $\mathbf{X}$ گردآوری شده‌اند را در نظر بگیرید. ساختار وابستگی درون بردار تصادفی $(Y, X_1, \ldots, X_p)$ برای استخراج اطلاعات موجود در $\mathbf{X}$ درباره Y مورد استفاده قرار می‌گیرد. هدف از قیمت‌گذاری آماری، ارزیابی حق بیمه خالص با بیشترین دقت ممکن است. این بدین معنی است که این هدف، مقدار امید ریاضی شرطی $\mu(\mathbf{X}) = E(Y|\mathbf{X})$ از پاسخ Y (تعداد ادعا یا مقدار ادعا) بر اساس اطلاعات موجود $\mathbf{X}$ است. از اینجا به بعد، $\mu(\mathbf{X})$ را به عنوان حق بیمه صحیح (خالص) در نظر می‌گیریم. توجه داشته باشید که تابع $\mathbf{x} \mapsto \mu(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ معمولاً برای بیمه‌گر ناشناخته است و ممکن است رفتار پیچیده‌ای بر حسب $\mathbf{x}$ نشان دهد. به این دلیل است که این تابع را با حق بیمه (کاری یا واقعی) $\mathbf{x} \mapsto \pi(\mathbf{x})$ تقریب می‌زنند، که در مقایسه با تابع رگرسیون مجهول $\mathbf{x} \mapsto \mu(\mathbf{x})$ ساختار نسبتاً ساده‌ای دارد. این بدین معنی است، که شایستگی ابزار قیمت‌گذاری داده شده را می‌توان با استفاده از زوج $(\mu(\mathbf{X}), \pi(\mathbf{X}))$ ارزیابی کرد، به گونه‌ای که اگر هزارها ویژگی در $\mathbf{X}$ وجود داشته باشد، باز هم به حالت دومتغیری برمی‌گردیم. اندازه‌گیری بالابری، یک مولفه مهم دیگر در اعتبارسنجی مدل است. زمانی که یک مدل پیش‌بین ساخته شد، باید عملکرد آن در پیش‌بینی حق بیمه $\mu(\mathbf{X})$ صحیح بر اساس ویژگی‌های موجود در $\mathbf{X}$ تعیین شود. توجه کنید که پاسخ Y خود نقش مستقیمی در تعیین حق بیمه ایفا نمی‌کند (ورای تعریف $\mu(\mathbf{X})$ و کالیبراسیون مدل رگرسیون پیشنهادی که قیمت π را نشان می‌دهد)، زیرا وقتی از خروجی‌های $Y - \mu(\mathbf{X})$ روی یک پرتفوی به اندازه کافی بزرگ میانگین‌گیری می‌شود، آنها همدیگر را حذف می‌کنند (این

لازمه بیمه است). حق بیمه $\pi(\mathbf{X})$ باید تا حد ممکن با حق بیمه صحیح $\mu(\mathbf{X})$ نزدیک باشد. خواننده می‌تواند برای کسب اطلاعات بیشتر راجع به هدف مهم نرخ‌گذاری به [۵۳] مراجعه کند، که زیان‌های واقعی Y را پیش‌بینی نمی‌کند، بلکه برآوردهای دقیقی از $\mu(\mathbf{X})$ بدست می‌آید که مشاهده نشده‌اند.

نکته: در بسیاری از مدل‌هایی که در مطالعات بیمه به کار می‌روند، ویژگی‌های $\mathbf{X}$ ترکیب می‌شوند تا امتیاز S را شکل دهند (یعنی تابع مقدار حقیقی با ویژگی‌های $(X_1, \ldots, X_p)$ و حق بیمه π یک تابع یکنوا (یعنی صعودی) از S است. طبق قرارداد $\pi(\mathbf{X}) = h(S)$ برای یک تابع صعودی h برقرار است.

برخی مدل‌ها، امتیازات متعددی دارند، که هر کدام یک جنبه خاص از هزینه منتقل شده به بیمه‌گر را نشان می‌دهد. به عنوان مثال در مورد مدل‌های رگرسیون افزوده صفر که در آنها، جرم احتمال در صفر و هزینه انتظاری در زمان پر کردن ادعانامه، هر دو با یک امتیاز ویژه مدل‌سازی می‌شوند. از آنجا که این امتیازات همگی توابعی از ویژگی‌ای موجود هستند، در این پایان‌نامه، نماد $\pi(\mathbf{X})$ را تا حد ممکن، در حالت عمومی حفظ می‌کنیم.

برای راحتی تفسیر، فرض کنید که حق بیمه $\pi(\mathbf{X})$ مدنظر و همچنین مقادیر انتظاری شرطی $\mu(\mathbf{X})$ متغیرهای تصادفی پیوسته‌ای هستند، که توابع چگالی احتمال را می‌پذیرند. زمانی که حداقل یک ویژگی پیوسته در اطلاعات موجود $\mathbf{X}$ وجود داشته باشد و تابع π یک تابع صعودی پیوسته از امتیاز حقیقی S باشد، این مسئله صادق است. با این وجود، ممکن است پیش‌بینی‌های مبتنی بر ویژگی های گسسته و همچنین پیش‌بین های ثابت تکه ای مانند یک درخت تنها را، در نظر نگیرند. در واقع، $\pi(\mathbf{X})$ تنها تعداد محدودی ورودی می پذیرد. از آنجا که امروزه، قیمت گذاری آماری مبتنی بر مدل‌های پیچیده‌تر است (به عنوان مثال، درخت‌هایی که با جنگل‌های تصادفی ترکیب می‌شوند)، این فرض پیوستگی، عمومیت رویکرد را خدشه دار نمی‌کند. به علاوه، فرض می‌کنیم که پیش‌بین به صورت میانگین صحیح است، یعنی داریم،

$$E\big(\pi(\mathbf{X})\big) = E\big(\mu(\mathbf{X})\big) = E(Y) \tag{۱.۱}$$

همچنین فرض می‌کنیم که هر دو پاسخ Y و پیش‌بین π غیرمنفی باشند و $E(Y) < \infty$ برقرار باشد. این فرضیات در سراسر پایان‌نامه حفظ شده‌اند. توجه داشته باشید که این پایان‌نامه به صورت کامل به حق بیمه‌های فنی اختصاص دارد. این امر سبب می‌شود که فرض پیوستگی $\pi(\mathbf{X})$ کاملاً معقول باشد، زیرا امروزه بیمه گران به ویژگی‌های بسیار بیشتری دسترسی دارند و برای اینکه بتوانند پروفایل ریسک بیمه‌گذاران را به دقت ارزیابی کنند، به ابزارهای آماری پیشرفته متوسل می‌شوند. با این وجود، حق بیمه فنی باید بتواند ریسک هر بیمه‌گذار را مطابق با مشخصات فردی به دقت ارزیابی کند و هدف حق بیمه تجاری پرداخت شده به بیمه‌گذار، بهینه‌سازی سود بیمه‌گر است، که باید با قوانین اعمال شده توسط مقامات ناظر محلی مطابقت داشته باشد و تحت محدودیت‌های IT و محدودیت‌های کانال‌های حراج و دپارتمان‌های بازاریابی است. ساده‌سازی قیمت فنی که به قیمت تجاری می‌انجامد، منجر

به طرح قیمت‌گذاری گسسته می‌شود. به طور خاص، ویژگی پیوسته مانند سن بیمه‌گزار معمولاً در فهرست قیمت تجاری رده‌بندی می‌شود در حالیکه در حق بیمه فنی به کمک مدل‌های افزوده تعمیم‌یافته، به عنوان متغیر پیوسته در نظر گرفته می‌شود. بنابراین در عمل، نتایج بدست آمده در ذیل را نمی‌توان در فهرست قیمت تجاری قرار داد (که بیشترین احتمال برای گسسته بودن را دارند) در حالیکه باید در فهرست قیمت فنی باشد (که با بیشترین احتمال پیوسته هستند). اکنون، حتی اگر نتایج فنی بدست آمده در بخش‌های ۵.۲ و ۶.۲ برای حق‌بیمه‌های پیوسته باشند، معیارهای کلیدی «منحنی‌های تمرکز انتگرال گیری شده» و «مساحت بین منحنی‌ها» که در بخش‌های ۲.۶.۲ و ۷.۲ معرفی می‌شوند، همچنان برای حق بیمه‌های گسسته معنادار هستند. خوانندگان علاقه‌مند می‌توانند برای مشاهده توضیحات شفاف راجع به ساده‌سازی احتمالی در فهرست قیمت فنی برای رسیدن به فهرست قیمت تجاری با استفاده از ضرایب درجه‌بندی طبقه‌ای، به مقاله [۴۱] مراجعه کنند. بخش ۲ـ۴ در مقاله [۷۲] تنظیمات موردنیاز برای توزیع‌های گسسته را بیان می‌کند. با این وجود توجه داشته باشید که زمانی که فهرست قیمت خیلی ساده باشد، تعداد محدودی رده ریسک با مواجهه کافی ایجاد می‌شود که می‌توان تجارب را در آن سطح جمع‌آوری کرد و با استفاده از گراف‌های واقعی ساده در مقابل گراف‌های انتظاری، با مقادیر مورد انتظار مقایسه کرد.

۲.۵.۱ نمادگذاری

تابع توزیع [۱] $\pi(\mathbf{X})$ به صورت زیر تعریف می‌شود،

$$F_\pi(t) = \mathbb{P}[\pi(\mathbf{X}) \leq t], t \geq 0$$

که f_π تابع چگالی احتمال متناظر است و داریم،

$$F_\pi(t) = \int_0^t f_\pi(s)\, \mathrm{ds}\, , t \geq 0$$

و F_π^{-1} تابع چندک مربوطه (یا دارایی در خطر) است که به صورت وارون تعمیم‌یافته F_π تعریف می‌شود یعنی داریم،

$$F_\pi^{-1}(\alpha) = \inf\{t | F_\pi(t) \geq \alpha\} \text{ برای سطح احتمال } \alpha$$

برای سطح احتمال فرض پیوستگی ما تضمین می کند که اتحاد $F_\pi(F_\pi^{-1}(\alpha)) = \alpha$ برای همه سطوح احتمال برقرار باشد.

۳.۵.۱ ترتیب محدب

مشخصاً هر چه $\pi(\mathbf{X})$ پراکنده‌تر باشد، حاوی اطلاعات بیشتری راجع به حق بیمه واقعی است. پیش‌بین ثابت $\pi(\mathbf{X}) = E(Y)$ که کمترین پراکندگی را دارد، هیچ اطلاعاتی راجع به ریسک

[1] Distribution function

نسبی بیمه نامه‌های مختلف به همراه ندارد. بنابراین، به نظر می‌رسد که مقایسه تغییرپذیری آن در مسئله مورد مطالعه، حائز اهمیت باشد. ترتیب محدب[1] ، یک ابزار احتمالاتی موثر برای ارزیابی پخش تصادفی متغیرها فراتر از شاخص‌های ساده مانند انحراف معیار است. از مطالعه دنویت و همکارانش ([۱۹] ، بخش ۳ـ۴) می‌دانیم که وقتی دو متغیر تصادفی Z و T داریم، می‌گوییم T در ترتیب محدب کوچکتر از Z است (با $T \preceq_{cx} Z$ نشان داده می‌شود) اگر

$$E\big(g(T)\big) \leq E\big(g(Z)\big) \tag{۲.۱}$$

برای همه توابع محدب g برقرار باشد، مشروط بر اینکه امید ریاضی موجود باشد.
یکی از روش‌های مشخصه‌یابی مهم ترتیب محدب، ایجاد فضای احتمال یکسان به کمک امید ریاضی شرطی است. به ویژه، متغیرهای تصادفی Z و T رابطه $T \preceq_{cx} Z$ را برآورده می‌سازند اگر و تنها اگر دو متغیر تصادفی $\tilde{F}$ و $\tilde{Z}$ وجود داشته باشند و روی یک فضای احتمال تعریف شده باشند و $\tilde{F}$ به صورت T توزیع شود و $\tilde{Z}$ به صورت Z توزیع شود و $\{\tilde{F}, \tilde{Z}\}$ یک مارتینگل باشد یعنی $E[\tilde{Z}|\tilde{F}] = \tilde{F}$ به صورت تقریباً قطعی برقرار باشد. مستقیماً از آن نتیجه می‌شود که افزایش تعداد ویژگی‌ها زمانی سودمند است که داشته باشیم،
$\mathbf{X}_1 \subseteq \mathbf{X}_2 \Rightarrow \mu(\mathbf{X}_1) \preceq_{cx} \mu(\mathbf{X}_2)$.
وقتی از $\mathbf{X}_1$ به اطلاعات غنی‌تر $\mathbf{X}_2$ برویم، حق بیمه‌های پخش شده (μ) بیشتری تولید می‌شود.

۴.۵.۱ تعاریف منحنی‌های عملکرد

بر اساس مطالعه Frees و همکارانش ([۳۱]، [۳۲])، معیار عملکرد پایه پیش‌بین را روی منحنی های تمرکز و لورنز آن پیشنهاد می کنیم که تعریف آنها در ادامه متن یادآوری می‌شود.

تعریف ۱.۵.۱. منحنی تمرکز[2] حق بیمه واقعی $\mu(\mathbf{X})$ نسبت به حق بیمه کاری π بر اساس اطلاعات موجود در بردار $\mathbf{X}$ را می‌توان به صورت زیر تعریف کرد.

$$\alpha \mapsto CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha] = \frac{E\big(\mu(\mathbf{X}) I[\pi(\mathbf{X}) \leq F_{\pi}^{-1}(\alpha)]\big)}{E\big(\mu(\mathbf{X})\big)}$$

در اینجا بر خلاف مطالعه فریز و همکارانش ([۳۱]، [۳۲]) که نسبت دو حق بیمه را درنظر گرفتند، خود حق بیمه $\pi(\mathbf{X})$ در تعریف منحنی تمرکز وارد می‌شود. بنابراین معیارهای عملکرد پیش‌بین $\pi(\mathbf{X})$ برای پاسخ دوتایی $Y \in \{0, 1\}$ که توسط جوریروکس و جاسیاک ([۳۸]،[۳۹]) پیشنهاد شده بود، توسعه می‌یابد. برای مطالعه جامع خصوصیات منحنی غلظت، خواننده علاقه‌مند را به مطالعه [۷۲] ارجاع می‌دهیم.

[1] Convex order
[2] concentration curve

اکنون معنای شهودی منحنی غلظت $\pi(\mathbf{X})$ را شرح می‌دهیم. طبق تعریف ۱.۵.۱ می‌بینیم که $CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha]$ سهم درآمد کلی از حق بیمه واقعی را در زیر پرتفوی $\pi(\mathbf{X}) \leq F_{\pi}^{-1}(\alpha)$ نشان می‌دهد یعنی متناظر با $100\alpha\%$ قراردادها با کمترین حق بیمه π است. از آنجا که این بیمه نامه‌ها، پروفایل ریسک پایینی دارند، در معرض خطر ترک پرتفوی و جذب توسط رقیب هستند. بنابراین، مهم است که روی این گروه از بیمه‌گزاران، غلو نکنیم. در نتیجه، اهمیت منحنی تمرکز برای ارزیابی شایستگی حق بیمه مد نظر π است.
منحنی تمرکز به تنهایی برای ارزیابی عملکرد π کافی نیست. دلیل آن را در مقایسه زیر می توان دریافت. گورریروکس و جاسیاک ([۳۸]) به امتیازدهی اعتباری یعنی انتخاب اولیه متقاضیان بر اساس تمایل به بازپرداخت وام علاقه‌مند بودند. بنابراین، مقدار واقعی π اهمیتی نداشت، تنها رتبه ای که القا می کردند و آستانه ای که پذیرش یا رد تقاضانامه را تعیین می کرد، اهمیت داشت. با این وجود، در قیمت گذاری، ما باید مقادیر حق بیمه $\pi(\mathbf{X})$ را همراه با درآمد حق بیمه واقعی بر اساس $\mu(\mathbf{X})$ در نظر بگیریم. به این دلیل است که از منحنی لورنز پیش‌بین نیز به همراه منحنی تمرکز پاسخ نسبت به پیش‌بین استفاده می‌کنیم.

تعریف ۲.۵.۱. منحنی لورنز (LC) مربوط به پیش‌بین $\pi(\mathbf{X})$ به صورت زیر تعریف می‌شود،

$$\begin{aligned}\alpha \mapsto LC[\pi(\mathbf{X}); \alpha] &= CC[\pi(\mathbf{X}), \pi(\mathbf{X}); \alpha] \\ &= \frac{E\big(\pi(\mathbf{X}) I[\pi(\mathbf{X}) \leq F_{\pi}^{-1}(\alpha)]\big)}{E\big(\pi(\mathbf{X})\big)}\end{aligned}$$

بنابراین، منحنی لورنز طبق تعریف، وابستگی کامل به پراکندگی (یا تغییرپذیری) دارد. بر اساس مطالعه دنویت و همکارانش ([۱۹] ویژگی ۳ – ۴ – ۴۱)، می‌دانیم که افزایش پیش‌بین $\pi(\mathbf{X})$ در ترتیب محدب، منحنی لورنز را به سمت پایین حرکت می‌دهد.

نکته: اگر یک بیمه‌گر به حق بیمه واقعی $\mu(\mathbf{X})$ بر اساس اطلاعات موجود در $\mathbf{X}$ دسترسی داشته باشد، آنگاه نیازی نیست که CC را از LC متمایز کند. این بدین دلیل است که اگر $\pi(\mathbf{X}) = \mu(\mathbf{X})$ باشد، داریم،

$$LC[\pi(\mathbf{X}); \alpha] = CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha]$$

که برای همه سطوح احتمال α برقرار است.

به عبارت دیگر، دو منحنی عملکرد، به منحنی لورنز $\mu(\mathbf{X})$ کاهش می‌یابند. این بدین معنی است که زیرپرتفوی متناظر با $\pi(\mathbf{X}) \leq F_{\pi}^{-1}(\alpha)$ در تعادل است، زیرا حق بیمه با امید ریاضی شرطی میانگین مطابقت دارد. اختلاف بالای بین دو منحنی عملکرد نشان می‌دهد که پیش‌بین مدنظر، حق بیمه فنی واقعی را به شکل ضعیفی تقریب می‌زند.
بدین دلیل است که بسیاری از مطالعات تجربی تنها از منحنی لورنز $\pi(\mathbf{X})$ برای ارزیابی عملکرد مدل پیش‌بین بهره می گیرند. آنها پیش‌بین π را با امید ریاضی شرطی اشتباه می گیرند در

حالیکه بسیار محتمل است که $\pi(\mathbf{X})$ تنها $\mu(\mathbf{X})$ را تقریب بزند در صورتی که در واقعیت با آن تفاوت دارد. به علت این اختلاف، ما باید دوباره به زوج منحنی‌های $CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha]$ و $LC[\pi(\mathbf{X}); \alpha]$ متوسل شویم تا عملکرد مدل قیمت‌گذاری را ارزیابی کنیم.

فصل ۲

رتبه‌بندی شاخص‌های جینی و انتخاب مدل بر اساس منحنی‌های تمرکز و لورنز

۱.۲ مقدمه

در این فصل ضمن رتبه‌بندی ضرایب جینی بر اساس ترتیب محدب به انتخاب مدل بر مبنای منحنی‌های تمرکز و لورنز می‌پردازیم.

۲.۲ رتبه‌بندی شاخص‌های جینی براساس ترتیب $\leq_{icx}$

لم ۱.۲.۲. فرض کنید $G(\mathbf{x}) = G(x_1, \ldots, x_n)$ به صورت ۲۴.۲.۱ تعریف شود یعنی داشته باشیم، $G(\mathbf{x}) = \sum_{1 \leq i,j \leq n} |x_i - x_j|$ آنگاه $-G(\mathbf{x})$ فوق‌مدولی است.

برهان. اثبات لم ۱.۲.۲ مطابق با گزاره ۲.۲.۱ کافی است نشان دهیم

$$-G_{ij}(x_i, x_j) = -G(x_1, \ldots, x_n)$$

به عنوان یک تابع دومتغیره (x_i, x_j) برای هر x_k ثابت، $k \neq i, j$ و هر $1 \leq i \leq j \leq n$ فوق‌مدولی است. به طور ویژه، می‌خواهیم نشان دهیم که برای هر $1 \leq i \leq j \leq n$ و هر $x_i \leq y_i$ ، $x_j \leq y_j$ داریم،

$$G_{ij}(x_i, x_j) + G_{ij}(y_i, y_j) \leq G_{ij}(y_i, x_j) + G_{ij}(x_i, y_j)$$

توجه داشته باشید که

$$G_{ij}(x_i, x_j) = 2|x_i - x_j| + 2\sum_{k \neq i,j} \left(|x_i - z_k| + |x_j - z_k|\right) + \sum_{k,l:\{k,l\}\cap\{i,j\}=\varnothing} |z_k - z_l|$$

که هم‌ارز با بیان زیر است،

$$2|x_i - x_j| + 2|y_i - y_j| \leq 2|y_i - x_j| + 2|x_i - y_j|$$

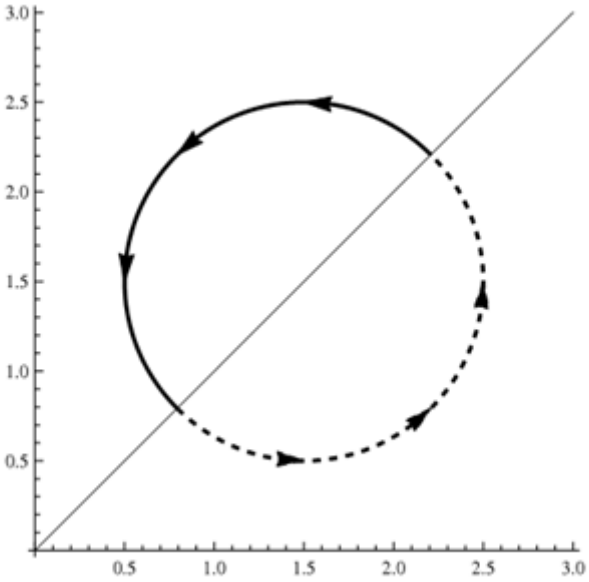
که اثبات آن آسان است. □

لم ۲.۲.۲. فرض کنید $G : \mathbb{R}^3 \to \mathbb{R}$ به صورت ۲۴.۲.۱ تعریف شود یعنی داشته باشیم، $G(x_1, x_2, x_3) = 2(|x_1 - x_2| + |x_2 - x_3| + |x_3 - x_1|)$. آنگاه $-u(G(x_1, x_2, x_3))$ برای هر u محدب صعودی، فوق‌مدولی است.

برهان. اثبات لم ۲.۲.۲ توجه داشته باشید که $u \circ G$ ناوردای جایگشتی است و کافی است نشان دهیم،

$$u\big(G(x_1, x_2, x_3)\big) + u\big(G(y_1, y_2, y_3)\big) \leq u\big(G(y_1, x_2, x_3)\big) + u\big(G(x_1, y_2, y_3)\big)$$

که برای هر $(x_1, x_2, x_3) \leq (y_1, y_2, y_3)$ برقرار است.



شکل ۱.۲: منحنی‌های جهت‌دار ∂C_+، ∂C_-.

مطابق با لم ۱.۲.۲ داریم، $G(x_1,x_2,x_3)+G(y_1,y_2,y_3)\leq G(y_1,x_2,x_3)+G(x_1,y_2,y_3)$.
یادآوری می کنیم که برای یک تابع محدب صعودی برای هر a,b,c,d داریم،
$u(a)+u(b)\leq u(c)+u(d)$ به گونه ای که $a+b\leq c+d$ و $\max\{a,b\}\leq\max\{c,d\}$ است. کافی است نشان دهیم،

$$\max\{G(x_1,x_2,x_3),G(y_1,y_2,y_3)\}\leq\max\{G(y_1,x_2,x_3),G(x_1,y_2,y_3)\} \qquad (۱.۲)$$

برای هر $(x_1,x_2,x_3)\leq(y_1,y_2,y_3)$ برقرار است.
توجه کنید $G(x_1,x_2,x_3)=۴\big(\max\{x_1,x_2,x_3\}-\min\{x_1,x_2,x_3\}\big)$ برقرار است.
اگر $x_1>min\{x_2,x_3\}$ باشد، آنگاه داریم، $y_1\geq x_1>\min\{x_2,x_3\}$ و بنابراین

$$\begin{aligned}G(x_1,x_2,x_3)&=۴\big(\max\{x_1,x_2,x_3\}-\min\{x_1,x_2,x_3\}\big)\\&\leq ۴\big(\max\{y_1,x_2,x_3\}-\min\{y_1,x_2,x_3\}\big)=G\big(y_1,x_2,x_3\big)\end{aligned}$$

در غیر این‌صورت اگر $x_1\leq\min\{x_2,x_3\}$ باشد، آنگاه $x_1\leq\min\{y_2,y_3\}$ و در نتیجه داریم،

$$\begin{aligned}G(x_1,x_2,x_3)=&۴\big(\max\{x_1,x_2,x_3\}-\min\{x_1,x_2,x_3\}\big)\\=&۴\big(\max\{x_2,x_3\}-x_1\big)\\\leq&۴\big(\max\{y_2,y_3\}-x_1\big)\\=&۴\big(\max\{x_1,y_2,y_3\}-\min\{x_1,y_2,y_3\}\big)\\=&G\big(x_1,y_2,y_3\big)\end{aligned}$$

بنابراین، نتیجه‌گیری می‌کنیم که، $G(x_1,x_2,x_3)\leq\max\{(y_1,x_2,x_3),G(x_1,y_2,y_3)\}$ برقرار است. همچنین می‌توان نشان داد که $G(y_1,y_2,y_3)\leq\max\{G(y_1,x_2,x_3),G(x_1,y_2,y_3)\}$ نیز برقرار است.

□

در سال ۲۰۰۱ یک شرط لازم و کافی برای ترتیب فوق مدولی بین توزیع‌های نرمال چندمتغیره توسط Muller در مرجع [۵۵] توسعه داده شد که در لم ۳.۲.۲ آن را بیان می‌کنیم.

لم ۳.۲.۲. فرض کنید

$$\mathbf{X}=(X_1,\ldots,X_n)\sim MVN(\mu_1,\mathbf{\Sigma}_1)$$

و

$$\mathbf{Y}=(Y_1,\ldots,Y_n)\sim MVN(\mu_2,\mathbf{\Sigma}_2)$$

برقرار باشند. آنگاه $\mathbf{X}\leq_{sm}\mathbf{Y}$ است، اگر وتنها اگر $\mu_1=\mu_2$ و $\mathbf{\Sigma}_1\leq\mathbf{\Sigma}_2$ برقرار باشند و همه مولفه‌های قطری برابر باشند.

بر اساس [۱۰] (نتیجه ۲ – ۳)، و همچنین [۵۶] یک شرط لازم و کافی برای توزیع های بیضوی بدست می آوریم.

گزاره ۱.۲.۲. فرض کنید $\mathbf{X} \sim EC_n(\mu_1, \Sigma_1, \psi_1)$ و $\mathbf{Y} \sim EC_n(\mu_2, \Sigma_2, \psi_2)$ با $\psi \in \Psi_\infty$ برقرار باشند. آنگاه $\mathbf{X} \leq_{sm} \mathbf{Y}$ است اگر و تنها اگر $\mu_1 = \mu_2$ و $\Sigma_1 \leq \Sigma_2$ برقرار و همه مولفه های قطری برابر باشند.

با ترکیب گزاره ۱.۲.۲ و لم ۲.۲.۲ ، نتیجه زیر را بدست می آوریم.

گزاره ۲.۲.۲. فرض کنید $\mathbf{X} = (X_1, X_2, X_3) \sim EC_3(\mathbf{0}, \Sigma_1, \psi)$ و $\mathbf{Y} = (Y_1, Y_2, Y_3) \sim EC_3(\mathbf{0}, \Sigma_2, \psi)$ با $\psi \in \Psi_\infty$ برقرار باشند. اگر $\Sigma_1 \leq \Sigma_2$ باشد و همه مولفه های قطری برابر باشند، آنگاه داریم، $G(\mathbf{X}) \geq_{icx} G(\mathbf{Y})$.

همچنین با ترکیب گزاره ۱.۲.۲ و لم ۱.۲.۲ ، نتیجه زیر را برای ریسک های دارای ابعاد بالا بدست می آوریم.

گزاره ۳.۲.۲. فرض کنید $\mathbf{X} \sim EC_n(\mathbf{0}, \Sigma_1, \psi)$ و $\mathbf{Y} \sim EC_n(\mathbf{0}, \Sigma_2, \psi)$ با $\psi \in \Psi_\infty$ برقرار باشند. اگر $\Sigma_1 \leq \Sigma_2$ باشد و همه مولفه های قطری برابر باشند، آنگاه داریم، $E\big(G(\mathbf{X})\big) \geq E\big(G(\mathbf{Y})\big)$ مشروط بر اینکه امیدهای ریاضی مربوطه وجود داشته باشند.

هدف نهایی، نشان دادن این مطلب است که وقتی که ماتریس پراکندگی برای ریسک بیضوی چندمتغیره $\mathbf{X}$ افزایش می‌یابد اما مولفه‌های قطری ثابت می‌مانند، $G(\mathbf{X})$ درترتیب تصادفی معمولی کاهش می‌یابد. اکنون گزاره ۲.۲.۲ یک گام مهم در حالت سه بعدی را در راستای این هدف تکمیل می‌کند. با این وجود، مباحث مربوط به اثبات گزاره ۲.۲.۲ را نمی‌توان به ابعاد بالاتر تعمیم داد. مشکل اصلی این است که در مورد ابعاد بالاتر، تابع ترکیبی $u(G(x))$، همانند حالت سه بعدی برای هر تابع صعودی u، لزوماً سه‌بعدی نیست، یعنی لم ۲.۲.۲ را نمی‌توان به ابعاد بالاتر تعمیم داد. مثال زیر، این نکته را نشان می‌دهد. بنابراین، خاطر نشان می‌کنیم که رتبه‌بندی $G(\mathbf{X})$ در ترتیب محدب صعودی برای ریسک در ابعاد بالاتر، هنوز یک مسئله حل نشده باقی مانده است.

مثال ۱.۲.۲. $n = ۴$ را در نظر بگیرید. فرض کنید $u(x) = x^۲$. آنگاه u در $x \geq ۰$، محدب صعودی است. باید نشان دهیم $-u(G(x_1, x_2, x_3, x_4))$ فوق‌مدولی است. برای این کار، فرض کنید $\mathbf{x} = (۳,۱,۰,۰)$ و $\mathbf{y} = (۲,۲,۰,۰)$ را داشته باشیم. آنگاه $\mathbf{x} \wedge \mathbf{y} = (۲,۱,۰,۰)$ و $\mathbf{x} \vee \mathbf{y} = (۳,۲,۰,۰)$ برقرار هستند. محاسبات مستقیم نشان می‌دهند که $u(G(\mathbf{x})) = ۲۰^۲ = ۴۰۰$ ، $u(G(\mathbf{y})) = ۱۶^۲ = ۲۵۶$ ، $u(G(\mathbf{x} \wedge \mathbf{y})) = ۱۴^۲ = ۱۹۶$ و $u(G(\mathbf{x} \vee \mathbf{y})) = ۲۲^۲ = ۴۸۴$ برقرار هستند که رابطه $-u(G(\mathbf{x})) - u(G(\mathbf{y})) \leq -u(G(\mathbf{x} \wedge \mathbf{y})) - u(G(\mathbf{x} \vee \mathbf{y}))$ را برآورده نمی‌سازد.

۳.۲ رتبه‌بندی شاخص‌های جینی بر اساس ترتیب $\leq_{st}$.

در این بخش، هم‌یکنوایی $G(\mathbf{X})$ را در ترتیب تصادفی معمولی با اعمال فرض‌های خاص بر ماتریس پراکندگی مورد بحث و بررسی قرار می‌دهیم.

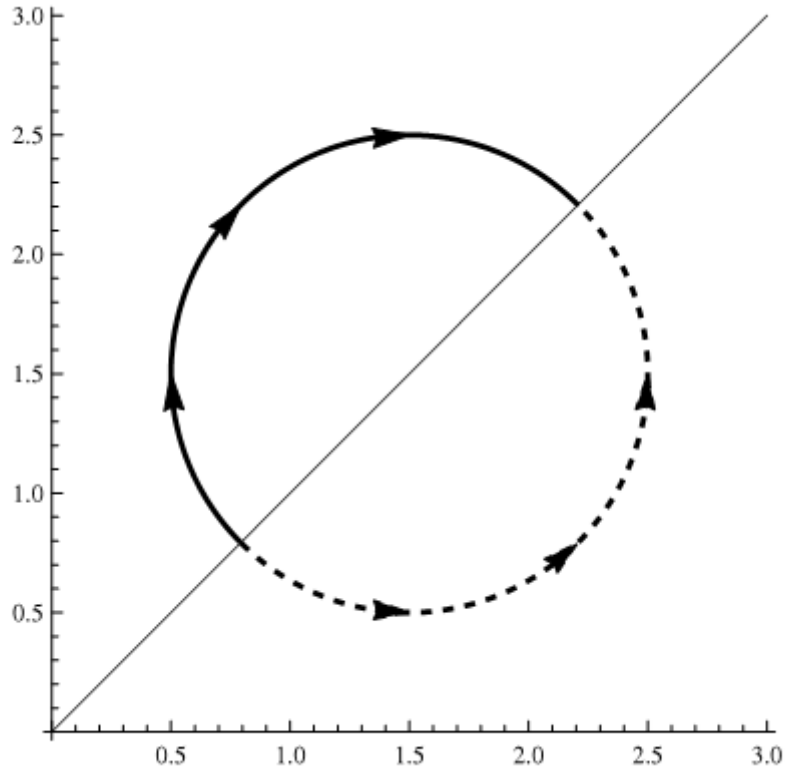
۱.۳.۲ حالت تبادلی شرطی

گزاره ۱.۳.۲. فرض کنید $\mathbf{X} = (X_1, \ldots, X_n)$ از یک توزیع نرمال چندمتغیری با میانگین $\mathbf{0}$ و ماتریس کوواریانس معین مثبت $\mathbf{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{n\times n}$ تبعیت می‌کند. اگر $\sigma_{1k} = \sigma_{2k}$ برای همه $k = 3, \ldots, n$ برقرار و $\sigma_{11} = \sigma_{22}$ باشد، آنگاه $\mathbb{P}\{G(\mathbf{X}) \leq t\}$ در برای هر $t \geq 0$ صعودی است.

به منظور اثبات گزاره ۱.۳.۲ ، از لم زیر بهره می‌گیریم.

لم ۱.۳.۲. فرض کنید (X, Y) یک بردار تصادفی نرمال دومتغیره تبادلی با ضریب همبستگی ρ داشته باشد. همچنین فرض کنید $C \subset \mathbb{R}^2$ هر مجموعه محدبی باشد که در آن، $\{(x, y)|(y, x) \in C\} = C$ برقرار است. بنابراین، $\mathbb{P}\{(X, Y) \in C\}$ در ρ صعودی است.

برهان. برای سادگی، فرض می‌کنیم C کراندار است. حالت بدون کران را می‌توان با مبحث حد تقریب زد. مرز C با جهت مثبت (پادساعتگرد) را با ∂C نشان می‌دهیم. بنابراین، ∂C به علت محدب بودن C به صورت تکه‌ای هموار است. به علاوه، داریم، $\partial C_+ = \{(x, y) \in \partial C | y \geq x\}$ و $\partial C_- = \{(x, y) \in \partial C | y \leq x\}$ ، بنابراین $\partial C = \partial C_+ \cup \partial C_-$. فرض کنید $\partial C'_+$ همان ∂C_+ باشد اما در جهت مخالف (ساعتگرد). آنگاه $\partial C'_+$ و ∂C_- بازتاب همدیگر نسبت به خط $y = x$ هستند. شکل‌های ۱ و ۲ تصویری (نه نمایش دقیق) از منحنی‌های جهت‌دار هستند.



شکل ۲.۲: منحنی‌های جهت‌دار $\partial C'_+$، ∂C_-.

بدون از دست دادن عمومیت، فرض می‌کنیم $E(X) = E(Y) = ۰$ و $Var(X) = Var(Y) =$ ۱. بنابراین، تابع چگالی (X, Y) به صورت $f(x,y) = \frac{۱}{۲\pi\sqrt{۱-\rho^۲}} \exp\left\{-\frac{x^۲+y^۲-۲\rho xy}{۲(۱-\rho^۲)}\right\}$ است. اتحاد پلاکت Plackett در مرجع [۵۹] بیان می‌کند که رابطه $\frac{\partial}{\partial\rho}f(x,y) = \frac{\partial^۲}{\partial x\partial y}f(x,y)$ برقرار است. مطابق با قضیه فوبینی داریم،

$$\begin{aligned}
\frac{\partial}{\partial\rho}\mathbb{P}\{(X,Y)\in C\} &= \frac{\partial}{\partial\rho}\int_c f(x,y)\mathbf{d}x\mathbf{d}y = \int_c \frac{\partial}{\partial\rho}f(x,y)\mathbf{d}x\mathbf{d}y \\
&= \int_c \frac{\partial^۲}{\partial x\partial y}f(x,y)\mathbf{d}x\mathbf{d}y \overset{*}{=} \oint_{\partial C}\frac{\partial}{\partial y}f(x,y)\mathbf{d}y \\
&= \left(\int_{\partial C_+}+\int_{\partial C_-}\right)\frac{\partial}{\partial y}f(x,y)\mathbf{d}y \\
&= \left(-\int_{\partial C'_+}+\int_{\partial C_-}\right)\frac{\partial}{\partial y}f(x,y)\mathbf{d}y
\end{aligned} \tag{۱.۲}$$

که در آن، تساوی $(*)$ از قضیه گرین بدست آمده است.

توجه کنید که $\frac{\partial}{\partial y}f(x,y)=\frac{\rho x-y}{1-\rho^2}f(x,y)$ برقرار است. با استفاده از (۱.۲) داریم،

$$\begin{aligned}\frac{\partial}{\partial\rho}\mathbb{P}\{(X,Y)\in C\} &= -\int_{\partial C'_+}\frac{\rho x-y}{1-\rho^2}f(x,y)\mathbf{d}y+\int_{\partial C_-}\frac{\rho x-y}{1-\rho^2}f(x,y)\mathbf{d}y\\ &= -\int_{\partial C'_+}f(x,y)dy\times\int_{\partial C'_+}\frac{\rho x-y}{1-\rho^2}\frac{f(x,y)}{\int_{\partial C'_+}f(x,y)\mathbf{d}y}\mathbf{d}y\\ &+\int_{\partial C_-}f(x,y)dy\times\int_{\partial C_-}\frac{\rho x-y}{1-\rho^2}\frac{f(x,y)}{\int_{\partial C_-}f(x,y)\mathbf{d}y}\mathbf{d}y\\ &= -\int_{\partial C'_+}f(x,y)\mathbf{d}y\times E\left(\frac{\rho X-Y}{1-\rho^2}\middle|(X,Y)\in\partial C'_+\right)\\ &+\int_{\partial C_-}f(x,y)\mathbf{d}y\times E\left(\frac{\rho X-Y}{1-\rho^2}\middle|(X,Y)\in\partial C_-\right)\end{aligned}$$

به علت تقارن بین ∂C_+ و $\partial C'_-$ و تبادل‌پذیری (X,Y) (یا $f(x,y)$)، داریم،
$\int_{\partial C'_+}f(x,y)\mathbf{d}y=\int_{\partial C_-}f(x,y)\mathbf{d}y$ و

$$\begin{aligned}E\left(\frac{\rho X-Y}{1-\rho^2}\middle|(X,Y)\in\ \partial C'_+\right) &= E\left(\frac{\rho Y-X}{1-\rho^2}\middle|(Y,X)\in\ \partial C'_+\right)\\ &= E\left(\frac{\rho Y-X}{1-\rho^2}\middle|(X,Y)\in\partial C_-\right)\end{aligned}$$

بنابراین،

$$\begin{aligned}\frac{\partial}{\partial\rho}\mathbb{P}\{(X,Y)\in C\} &= \int_{\partial C_-}f(x,y)\mathbf{d}y\\ &\times\left(-E\left(\frac{\rho X-Y}{1-\rho^2}\middle|(X,Y)\in\ \partial C'_+\right)+E\left(\frac{\rho X-Y}{1-\rho^2}\middle|(X,Y)\in\ \partial C_-\right)\right)\\ &= \int_{\partial C_-}f(x,y)\mathbf{d}y\\ &\times\left(-E\left(\frac{\rho Y-X}{1-\rho^2}\middle|(X,Y)\in\ \partial C_-\right)+E\left(\frac{\rho X-Y}{1-\rho^2}\middle|(X,Y)\in\ \partial C_-\right)\right)\\ &= \int_{\partial C_-}f(x,y)\mathbf{d}y\times E\left(\frac{(1+\rho)(X-Y)}{1-\rho^2}\middle|(X,Y)\in\ \partial C_-\right)\geq 0\end{aligned}$$

نامساوی آخر به این دلیل برقرار است که $(X,Y)\in\partial C_-$ دلالت بر $X\geq Y$ دارد. بلافاصله نتیجه می شود که $\mathbb{P}\{(X,Y)\in C\}$ در ρ صعودی است. □

برهان. **اثبات گزاره ۱.۳.۲** از مبحث شرط‌گذاری استفاده می‌کنیم. مطابق با ویژگی توزیع نرمال چندمتغیره، می‌دانیم که مشروط به $\{X_3=x_3,\ldots,X_n=x_n\}$، (X_1,X_2) یک توزیع نرمال دومتغیره تبادلی با کوواریانس $\partial^*_{12}=\partial_{12}-s$ دارد که در آن، s با مولفه های دیگر ماتریس کوواریانس تعیین می‌شود. برای هر $x_3,\ldots,x_n$ ثابت و $t\geq 0$ ، داریم،
$C=\{(x_1,x_2)|G(x_1,x_2,x_3,\ldots,x_n)\leq t\}$. توجه داشته باشید که $C\subset R^2$ یک چندوجهی

محدب است و نسبت به خط $x_۱ = x_۲$ متقارن است. مطابق با لم ۱.۳.۲،
$\mathbb{P}\{G(\mathbf{X}) \le t|X_۳ = x_۳, \ldots, X_n = x_n\} = \mathbb{P}\{(X_۱, X_۲) \in C|X_۳ = x_۳, \ldots, X_n = x_n\}$ در $\sigma^*_{۱۲}$ و در نتیجه ، در $\sigma_{۱۲}$ صعودی است. بنابراین، $\mathbb{P}\{G(\mathbf{X}) \le t\} = E\big(\mathbb{P}\{G(\mathbf{X}) \le t|X_۳, \ldots, X_n\}\big)$ نیز در $\sigma_{۱۲}$ صعودی است. □

گزاره ۱.۳.۲ پیشنهاد می‌کند که اگر یک ساختار تبادلی شرطی را بر بردار تصادفی نرمال چند متغیره $\mathbf{X}$ اعمال کنیم، یعنی $(X_۱, X_۲)$ روی مولفه‌های باقیمانده شرطی تبادلی باشد، آنگاه $G(\mathbf{X})$ در $Cov(X_۱, X_۲)$ که یک مولفه از ماتریس کوواریانس است، نزولی تصادفی می باشد. در ذیل، یک نتیجه مشابه با استفاده از گزاره ۳.۲.۱ برای توزیع بیضوی بدست می‌آید.

گزاره ۲.۳.۲. فرض کنید $\mathbf{X} \sim EC_n(\mathbf{0}, \mathbf{\Sigma}, \psi)$ با $\mathbf{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{n\times n}$ و $\psi \in \Psi_\infty$ برقرار باشند. اگر برای همه $k = ۳, \ldots, n$ داشته باشیم $\sigma_{۱k} = \sigma_{۲k}$ و $\sigma_{۱۱} = \sigma_{۲۲}$ باشد، آنگاه $\mathbb{P}\{G(\mathbf{X}) \le t\}$ برای هر $t \ge ۰$ در σ_{12} صعودی است.

برهان. فرض کنید $\mathbf{X}' \sim EC_n(\mathbf{0}, \mathbf{\Sigma}', \psi)$ با $\mathbf{\Sigma}' = (\sigma'_{ij}) \in \mathbb{R}^{n\times n}$ برقرار باشند که در آنها برای همه $۱ \le i \le j \le n$ و $(i, j) \ne (۱, ۲)$ داریم، $\sigma'_{ij} = \sigma_{ij}$ و $\sigma'_{12} \ge \sigma_{12}$. مطابق با گزاره ۳.۲.۱، $\mathbf{Y} \sim MVN(\mathbf{0}, \mathbf{\Sigma})$، $\mathbf{Y}' \sim MVN(\mathbf{0}, \mathbf{\Sigma}')$ و متغیر تصادفی $R \ge ۰$ مستقل از $\mathbf{Y}$ و $\mathbf{Y}'$ وجود دارد به گونه ای که روابط $\mathbf{X} \overset{d}{=} R\mathbf{Y}$ و $\mathbf{X}' \overset{d}{=} R\mathbf{Y}'$ برقرار هستند. توجه کنید که برای هر $r > ۰$ معین، $r\mathbf{Y}$ و $r\mathbf{Y}'$ از توزیع های نرمال چندمتغیره تبعیت می‌کنند و ماتریس های کوواریانس شرط گزاره ۱.۳.۲ را برآورده می‌سازند. بنابراین، ما داریم، $\mathbb{P}\{G(r\mathbf{Y}) \le t\} \le \mathbb{P}\{G(r\mathbf{Y}') \le t\}$. در نتیجه

$$\begin{aligned}\mathbb{P}\{G(\mathbf{x})\} &= \mathbb{P}\{G(R\mathbf{Y}) \le t\} = E\big(\mathbb{P}\{G(R\mathbf{Y}) \le t\}|R\big)\\ &\le E\big(\mathbb{P}\{G(R\mathbf{Y}') \le t\}|R\big) = \mathbb{P}\{G(R\mathbf{Y}') \le t\} = \mathbb{P}\{G(\mathbf{X}') \le t\}.\end{aligned}$$

□

۲.۳.۲ شبه ـ تسلط بین ماتریس های کوواریانس

ابتدا، قضیه ۴ـ۱ بیان شده توسط Eaton and Perlman در مرجع [۲۷] را در ذیل بیان می‌کنیم [۲۷].

لم ۲.۳.۲. فرض کنید $\mathbf{X}_i \sim MVN(\mathbf{0}, \mathbf{\Sigma}_i)$ برای $i = ۱, ۲$ برقرار باشد. اگر $\mathbf{\Sigma}_۲ - \mathbf{\Sigma}_۱$ نیمه‌معین مثبت باشد، آنگاه $\mathbb{P}\{\mathbf{X}_۲ \in C\} \le \mathbb{P}\{\mathbf{X}_۱ \in C\}$ برای هر مجموعه متقارن مرکزی و محدب C (یعنی $C = -C$) برقرار است.

می‌گوییم ماتریس P بر ماتریس دیگر Q غالب است، اگر $P - Q$ نیمه‌معین مثبت باشد. در این حالت، تعریف ۱.۳.۲ را شبه تسلط می نامیم. لم ۲.۳.۲ بیانگر این است که ماتریس

کوواریانس درجه تمرکز مرکزی یک توزیع نرمال چندمتغیره را تعیین می‌کند. به ویژه، هر چه ماتریس کوواریانس کوچکتر باشد، بردار تصادفی نرمال در یک ناحیه متقارن مرکزی و محدب، متمرکزتر است.

گزاره ۳.۳.۲. فرض کنید $\mathbf{X} \sim MVN(\mathbf{0}, \mathbf{\Sigma}_X)$ و $\mathbf{Y} \sim MVN(\mathbf{0}, \mathbf{\Sigma}_Y)$ برقرار باشند. اگر $a \in \mathbb{R}$ وجود داشته باشد به گونه‌ای که،

$$a\mathbf{1}_{n\times n} + \mathbf{\Sigma}_X - \mathbf{\Sigma}_Y \qquad (۲.۲)$$

نیمه معین مثبت باشد، آنگاه $G(\mathbf{X}) \geq_{st} G(\mathbf{Y})$ برقرار است.

برهان. **اثبات گزاره ۳.۳.۲** رابطه (۳.۱) را یادآوری می کنیم،

$$G(\mathbf{X}) = \sum_{k=1}^{n}(4k - 2n - 2)X_{(k)} \underset{=}{\Delta} \sum_{k=1}^{n} c_k X_{(k)}$$

که در آن، $c_k = 4k - 2n - 2$. توجه داشته باشید که $\{c_k, k = 1, 2, \ldots, n\}$ یک دنباله صعودی است و با بازآرایی نامساوی، داریم،

$$G(\mathbf{X}) = \max_{\pi\in P}\left\{\sum_{k=1}^{n} c_k X_{\pi(k)}\right\} = \max_{\pi\in P}\left\{\sum_{k=1}^{n} c_{\pi(k)} X_k\right\}$$

که در آن، P مجموعه همه جایگشت های $(1, 2, \ldots, n)$ را نشان می دهد. فرض کنید $\mathbf{C} \in \mathbb{R}^{n!\times n}$ ماتریسی باشد که با همه جایگشت های مختلف $(c_1, \ldots, c_n)$ تولید شده است. در نتیجه، $G(\mathbf{X})$ و $G(\mathbf{Y})$ به ترتیب بزرگترین آمار ترتیبی بردارهای تصادفی $\mathbf{CX}^T$ و $\mathbf{CY}^T$ هستند. از سوی دیگر، از آنجا که $\mathbf{X}$ و $\mathbf{Y}$ از توزیع‌های نرمال چندمتغیره تبعیت می‌کنند، در نتیجه $\mathbf{CX}^T$ و $\mathbf{CY}^T$ نیز همین‌گونه هستند. به ویژه داریم،
$\mathbf{CX}^T \sim MVN(\mathbf{0}, \mathbf{C\Sigma}_X\mathbf{C}^T)$ و $\mathbf{CY}^T \sim MVN(\mathbf{0}, \mathbf{C\Sigma}_Y\mathbf{C}^T)$.

توجه داشته باشید که، $\mathbf{C1}_{n\times n}\mathbf{C}^T = \mathbf{0}$ است زیرا $\sum_{k=1}^{n} c_k = 0$ است، از مقایسه ماتریس های کواریانس $\mathbf{CX}^T$ و $\mathbf{CY}^T$ نتیجه می‌گیریم،

$$\begin{aligned}\mathbf{C\Sigma}_X\mathbf{C}^T - \mathbf{C\Sigma}_Y\mathbf{C}^T =& \mathbf{C\Sigma}_X\mathbf{C}^T - \mathbf{C\Sigma}_Y\mathbf{C}^T + a\mathbf{C1}_{n\times n}\mathbf{C}^T \\ =& \mathbf{C}(\mathbf{\Sigma}_X - \mathbf{\Sigma}_Y + a\mathbf{1_{n\times n}})\mathbf{C}^T\end{aligned}$$

که نیمه‌معین مثبت است زیرا $\mathbf{\Sigma}_X - \mathbf{\Sigma}_Y + a\mathbf{1_{n\times n}}$ نیمه‌معین مثبت است. یادآوری می‌کنیم که $G(\mathbf{X}) = \max\{row_i\mathbf{CX}^T, i = 1, \ldots, n!\}$. از آنجا که $\{c_1, \ldots, c_n\} = \{-c_1, \ldots, -c_n\}$ است، بنابراین داریم، $\{row_i\mathbf{CX}^T, i = 1, \ldots, n!\} = \{-row_i\mathbf{CX}^T, i = 1, \ldots, n!\}$. بنابراین داریم،

$$\begin{aligned}\mathbb{P}\{G(\mathbf{X}) \leq t\} =& \mathbb{P}\{\mathbf{CX}^T \leq (t, \ldots, t)^T\} = \mathbb{P}\{-\mathbf{CX}^T \leq (t, \ldots, t)^T\} \\ =& \mathbb{P}\{(-t, \ldots, -t)^T \leq \mathbf{CX}^T \leq (t, \ldots, t)^T\} \\ \underset{=}{\Delta}& \mathbb{P}\{\mathbf{CX}^T \in Q_t\}\end{aligned}$$

که در آن، Q_t ابرمکعبی است که مرکز آن در مبدأ است و طول ضلع آن $2t$ است. مشخص است که Q_t محدب و متقارن مرکزی است. مطابق با لم ۲.۳.۲ داریم،

$$\mathbb{P}\{G(\mathbf{X}) \leq t\} = \mathbb{P}\{\mathbf{C}\mathbf{X}^T \in Q_t\} \leq \mathbb{P}\{\mathbf{C}\mathbf{Y}^T \in Q_t\} = \mathbb{P}\{G(\mathbf{Y}) \leq t\}$$

برای هر t که طبق تعریف ۲۴.۲.۱ ، دلالت بر $G(\mathbf{X}) \geq_{st} G(\mathbf{Y})$ دارد. □

در این پایان‌نامه، ما به مقایسه ساختار وابستگی بدون تغییر حاشیه‌ها می‌پردازیم. این بدین معنی است که ماتریس های پراکندگی ای که مقایسه می‌کنیم، مولفه های قطری مشابهی دارند. زمانی که تفاضل محاسبه می‌شود، مولفه های قطری صفر می‌شوند. در این حالت، ما انتظار نداریم که یک ماتریس پراکندگی بر دیگری غالب شود زیرا ماتریس تفاضل، نیمه معین مثبت نیست. بنابراین، شرط تسلط در لم ۲.۳.۲ به شرط «شبه» تسلط در ۱.۳.۲ تبدیل می شود که با این شرایط متناسب است.

مثال ۱.۳.۲. مثال هایی که شرط ۱.۳.۲ را برآورده می سازند.

- همه مولفه های غیرقطری ماتریس کوواریانس، به یک میزان افزایش می‌یابند یعنی $\mathbf{\Sigma}_Y = \mathbf{\Sigma}_X + \sigma(\mathbf{1_{n\times n}} - \mathbf{I}_n)$ با $\sigma \geq 0$ برقرار است. این بیانگر حالتی است که در آن، $\mathbf{X}$ و $\mathbf{Y}$ هر دو قابل تبادل باشند. نتیجه گزاره ۳.۳.۲ برای $\mathbf{X}$ و $\mathbf{Y}$ تبادلی، با استفاده از رویکردهای دیگر اثبات شده است به عنوان مثال می‌توان به قضیه ۶ – ۲۵ در مقاله Tong مراجعه کرد [۶۷].

- مولفه‌های غیرقطری کوواریانس روی سطر و ستون kام، به یک میزان افزایش می‌یابند یعنی داریم، $\mathbf{\Sigma}_Y = \mathbf{\Sigma}_X + \sigma \sum_{j\neq k} (\mathbf{\Delta}_{jk} + \mathbf{\Delta}_{kj})$ و $\sigma > 0$ که در آن، $\mathbf{\Delta}_{kj}$ ماتریسی با عدد ۱ در موقعیت (k, j) و ۰ در موقعیت های دیگر را نشان می دهد. با استفاده از بحثی مشابه با گزاره ۲.۳.۲، به آسانی می‌توان نتیجه گزاره ۳.۳.۲ را به توزیع های بیضوی تعمیم داد. اثبات آن مشابه اثبات گزاره ۲.۳.۲، است و در نتیجه از آن صرف نظر می شود.

گزاره ۴.۳.۲. فرض کنید $\mathbf{X} \sim EC_n(\mathbf{0}, \mathbf{\Sigma}_X, \psi)$ و $\mathbf{Y} \sim EC_n(\mathbf{0}, \mathbf{\Sigma}_Y, \psi)$ با $\psi \in \Psi_\infty$ برقرار باشند. اگر یک $a \in \mathbb{R}$ وجود داشته باشد به گونه ای که $a\mathbf{1_{n\times n}} + \mathbf{\Sigma}_X - \mathbf{\Sigma}_Y$ نیمه معین مثبت باشد، آنگاه داریم، $G(\mathbf{X}) \geq_{st} G(\mathbf{Y})$.

۳.۳.۲ مقایسه با هم یکنوایی

در بالا، ما ترتیبی اتخاد کردیم که با تبعیت ماتریس های پراکندگی از ساختارهای خاص، شاخص‌های جینی ریسک های بیضوی چندمتغیره مرتب شوند. از سوی دیگر، سروکار با ماتریس های پراکندگی با ساختار عمومی کار دشواری است. در این بخش، اثبات می‌شود که هم یکنوایی، کوچکترین شاخص جینی تصادفی را در میان همه ریسک های بیضوی چندمتغیری با حاشیه های مشترک ایجاد می‌کند. مباحث هندسی مورد استفاده در این بخش، مبتنی بر قضیه $A۲$ در مرجع [۴۳] هستند، که ترتیب هماهنگی بین توزیع های بیضوی را مشخص می کنند.

گزاره ۵.۳.۲. فرض کنید $\mathbf{X}=(X_۱,X_۲,X_۳)$ و $\mathbf{Y}=(Y_۱,Y_۲,Y_۳)$ از توزیع های نرمال چندمتغیره با میانگین $\circ$ و توزیع های حاشیه ای مشترک تبعیت کنند. اگر $(Y_۱,Y_۲,Y_۳)$ هم‌یکنوا باشد، آنگاه $G(\mathbf{Y})\leq_{st}G(\mathbf{X})$ برقرار است.

برهان. واریانس‌های توزیع‌های حاشیه ای را با $\sigma_۱^۲\leq\sigma_۲^۲\leq\sigma_۳^۲$ نشان می‌دهیم.
$\{Z_۱,Z_۲,Z_۳\}\overset{i.i.d}{\sim}N(\circ,۱)$ وجود دارند به گونه‌ای که $(Y_۱,Y_۲,Y_۳)\overset{d}{=}(\sigma_۱Z_۱,\sigma_۲Z_۱,\sigma_۳Z_۱)$ باشد و

$$(X_۱,X_۲,X_۳)\overset{d}{=}\begin{pmatrix}\sigma_۱&\circ&\circ\\ \circ&\sigma_۲&\circ\\ \circ&\circ&\sigma_۳\end{pmatrix}\begin{pmatrix}l_{۱۱}&\circ&\circ\\ l_{۲۱}&l_{۲۲}&\circ\\ l_{۳۱}&l_{۳۲}&l_{۳۳}\end{pmatrix}\begin{pmatrix}Z_۱\\ Z_۲\\ Z_۳\end{pmatrix}$$

که در آن، $\mathbf{L}_X=\begin{pmatrix}l_{۱۱}&\circ&\circ\\ l_{۲۱}&l_{۲۲}&\circ\\ l_{۳۱}&l_{۳۲}&l_{۳۳}\end{pmatrix}$ تجزیه چولسکی ماتریس همبستگی $\Big(X_۱,X_۲,X_۳\Big)$ است و بدین معنی است که $l_{۱۱}=۱$ و $l_{۲۱}^۲+l_{۲۲}^۲=۱$ و $l_{۳۱}^۲+l_{۳۲}^۲+l_{۳۳}^۲=۱$ برقرار هستند. بنابراین، $\mathbb{P}\{G(\mathbf{X})\leq t\}=\mathbb{P}\{(Z_۱,Z_۲,Z_۳)\in R_X(t)\}$ و $\mathbb{P}\{G(\mathbf{Y})\leq t\}=\mathbb{P}\{(Z_۱,Z_۲,Z_۳)\in R_Y(t)\}$ برقرار است که در آن داریم،

$$\begin{aligned}R_Y(t)&=\{(z_۱,z_۲,z_۳)\mathbf{۴}(\sigma_۳-\sigma_۱)|z_۱|\leq t\}\\ R_X(t)&=\{(z_۱,z_۲,z_۳)\mathbf{۲}|\sigma_۱z_۱'-\sigma_۲z_۲'|+\\ &\mathbf{۲}|\sigma_۲z_۲'-\sigma_۳z_۳'|+\mathbf{۲}|\sigma_۳Z_۳'-\sigma_۱Z_۱'|\leq t\}\\ &=\{(z_۱,z_۲,z_۳):\mathbf{۴}|\sigma_۱z_۱'-\sigma_۲z_۲'|\leq t\}\\ &\cap\{(z_۱,z_۲,z_۳):\mathbf{۴}|\sigma_۲z_۲'-\sigma_۳z_۳'|\leq t\}\\ &\cap\{(z_۱,z_۲,z_۳):\mathbf{۴}|\sigma_۳z_۳'-\sigma_۱z_۱'|\leq t\}\end{aligned}$$

که در آن، $z_۱'=z_۱$ ، $z_۲'=l_{۲۱}z_۱+l_{۲۲}z_۲$ و $z_۳'=l_{۳۱}z_۱+l_{۳۲}z_۲+l_{۳۳}z_۳$ برقرار هستند. اکنون، دو ناحیه $R_Y(t)$ و $R_X^۱(t)=\{(z_۱,z_۲,z_۳):\mathbf{۴}|\sigma_۳z_۳'-\sigma_۱z_۱'|\leq t\}$ را مقایسه می‌کنیم. توجه کنید

که هر دو آنها، نواحی‌ای بین یک زوج صفحه موازی هستند. برای $R_Y(t)$ فاصله بین صفحات مرزی، $\frac{t}{۲(\sigma_۳-\sigma_۱)}$ است. برای $R_X^۱(t)$ ، فاصله بین صفحات مرزی، $\frac{t}{۲\sqrt{\sigma_۱^۲+\sigma_۳^۳-۲l_{۳۱}\sigma_۱\sigma_۳}} \leq \frac{t}{۲(\sigma_۳-\sigma_۱)}$ است. از آنجا که هم $R_Y(t)$ و هم $R_X^۱(t)$ مرکزی در مبدأ دارند، نتیجه می‌گیریم که $R_X^۱(t)$ و در نتیجه، $R_X(t)$ ، به عنوان زیرمجموعه $R_X^۱(t)$ را می‌توان از طریق تبدیلات چرخشی خاصی به درون $R_Y(t)$ منتقل کرد. از آنجا که $(Z_۱, Z_۲, Z_۳)$ توزیع نسبت به چرخش ناورداست، بلافاصله نتیجه می شود، $\mathbb{P}\{(Z_۱, Z_۲, Z_۳) \in R_X(t)\} \leq \mathbb{P}\{(Z_۱, Z_۲, Z_۳) \in R_Y(t)\}$ یعنی برای هر $t \geq ۰$ داریم، $\mathbb{P}\{G(\mathbf{X}) \leq t\} \leq \mathbb{P}\{G(\mathbf{Y}) \leq t\}$ که دلالت بر رابطه $G(\mathbf{X}) \geq_{st} G(\mathbf{Y})$ دارد.

□

مشابه با گزاره های ۲.۳.۲ و ۳.۳.۲ می‌توان گزاره ۵.۳.۲ را نیز به توزیع های بیضوی تعمیم داد. همانند قبل از اثبات آن صرف نظر می شود.

گزاره ۶.۳.۲. فرض کنید $\mathbf{X} = (X_۱, X_۲, X_۳)$ و $\mathbf{Y} = (Y_۱, Y_۲, Y_۳)$ از توزیع های بیضوی با بردار میانگین $\mathbf{0}$ ، توزیع های حاشیه ای مشترک و یک مولد مشترک $\psi \in \Psi_\infty$ تبعیت کنند. اگر $\mathbf{Y}$ هم یکنوا باشد، آنگاه $G(\mathbf{Y}) \leq_{st} G(\mathbf{X})$ برقرار است.

۴.۲ بحث درباره مرتبه $pd-۱$

روش مورد استفاده برای تحلیل شاخص جینی در بخش های بالا را می‌توان برای تعمیم ترتیب $pd-۱$ که توسط Shaked and Tong در مرجع [۶۵] پیشنهاد شده، به کار برد. ترتیب $pd-۱$، یک ترتیب جزئی است که درجه پراکندگی بردارهای تصادفی تبادلی را مقایسه می‌کند. ما تعریف ترتیب $pd-۱$ را به صورت زیر بیان می کنیم.

تعریف ۱.۴.۲. فرض کنید $\mathbf{X} = (X_۱, \ldots, X_n)$ و $\mathbf{Y} = (Y_۱, \ldots, Y_n)$ دو بردار تصادفی تبادلی با حاشیه‌های مشترک باشند. می نویسیم $\mathbf{X} \leq_{pd-۱} \mathbf{Y}$ اگر $|\sum_{i=۱}^n c_i Y_{(i)}| \geq_{st} |\sum_{i=۱}^n c_i X_{(i)}|$ برای هر $\sum_{i=۱}^n c_i = ۰$ برقرار باشد.

Shaked and Tong در سال ۱۹۸۵ ، چندین مثال ارائه کردند که این ترتیب را برآورده می کرد از جمله توزیع های نرمال چندمتغیره و توزیع های نمایی. برای کسب اطلاعات بیشتر راجع به این ترتیب و دیگر ترتیب های مربوطه، خواننده می‌تواند به مراجع [۶۵]،[۶۳] مراجعه کند.

توجه کنید که این ترتیب، محدود است از این نظر که تنها بر بردارهای تصادفی تبادلی اعمال می شود. به منظور اعمال آن بر بردارهای تصادفی غیرتبادلی، می‌توانیم از ترتیب ضعیف‌تر استفاده کنیم و این رتبه‌بندی را در میان بردارهای تصادفی بیضوی برقرار کنیم.

گزاره ۱.۴.۲. فرض کنید $\mathbf{X} = (X_۱, \ldots, X_n)$ و $\mathbf{Y} = (Y_۱, \ldots, Y_n)$ دو بردار تصادفی با حاشیه های مشترک باشند. می‌نویسیم $\mathbf{X} \leq_{wpd-۱} \mathbf{Y}$ اگر $|\sum_{i=۱}^n c_i Y_{(i)}| \geq_{st} |\sum_{i=۱}^n c_i X_{(i)}|$ برای همه

$c_1 \le \cdots \le c_n$ ها برقرار باشد به گونه ای که داشته باشیم، $\{c_1,\ldots,c_n\} = \{-c_1,\ldots,-c_n\}$. مشخصاً $\mathbf{X} \le_{pd-1} \mathbf{Y}$ دلالت بر $\mathbf{X} \le_{wpd-1} \mathbf{Y}$ دارد، زیرا $\{c_1,\ldots,c_n\} = \{-c_1,\ldots,-c_n\}$ دلالت بر $\sum_{i=1}^{n} c_i = 0$ دارد. از آنجا که ترتیب $pd-1$ ضعیف، برای بردار تصادفی عمومی تعریف می شود، قید $\{c_1,\ldots,c_n\} = \{-c_1,\ldots,-c_n\}$ به آن افزوده می شود تا عدم تقارن ناشی از کنار گذاشتن فرض تبادل را جبران کند.

گزاره ۲.۴.۲. فرض کنید $\mathbf{X} \sim EC_n(\mathbf{0}, \mathbf{\Sigma}_X, \psi)$ ، با $\psi \in \Psi_\infty$ برقرار باشد و $\mathbf{X}$ و $\mathbf{Y}$ توزیع های حاشیه ای مشترک داشته باشند. اگر یک $a \in \mathbb{R}$ وجود داشته باشد به گونه ای که $a\mathbf{1_{n\times n}} + \mathbf{\Sigma}_X - \mathbf{\Sigma}_Y$ نیمه معین مثبت باشد، آنگاه $\mathbf{X} \le_{wpd-1} \mathbf{Y}$ برقرار است.

برهان. توجه کنید که مباحث مطرح شده برای اثبات گزاره ۲.۳.۲ همچنان برای هر $c_1,\ldots,c_n$ برقرار هستند و شرایط تعریف ۱.۴.۲ را برآورده می سازند. به سادگی با ترکیب گزاره ۳.۳.۲ و ۴.۳.۲ ، به نتیجه دلخواه می رسیم. از جزئیات صرف نظر شده است. □

همانند گزاره ۶.۳.۲ ، می‌توان اثبات کرد که ریسک بیضوی هم یکنوا، کران پایین همه ریسک های بیضوی چندمتغیره با حاشیه مشترک بر اساس مرتبه $pd-1$ ضعیف است.

گزاره ۳.۴.۲. فرض کنید $\mathbf{X} = (X_1, X_2, X_3)$ و $\mathbf{Y} = (Y_1, Y_2, Y_3)$ ، از توزیع های بیضوی با حاشیه های مشترک و یک مولد مشترک $\psi \in \Psi_\infty$ تبعیت کنند. اگر $\mathbf{Y}$ هم یکنوا باشد، آنگاه $\mathbf{X} \le_{wpd-1} \mathbf{Y}$ برقرار است.

برهان. برای هر c_3 به گونه ای که $c_1 \le c_2 \le c_3$ برقرار باشد، $\{c_1, c_2, c_3\} = \{-c_1, -c_2, -c_3\}$ داریم $c_2 = 0$ و $c_3 = -c_1 \ge 0$. بنابراین، $\sum_{i=1}^{3} c_i X_{(i)} = c_3\big(X_{(3)} - X_{(1)}\big) = \frac{c_3}{4} G(\mathbf{X})$ برقرار است. مطابق با گزاره ۶.۳.۲ ، می دانیم که $G(\mathbf{X}) \ge_{st} G(\mathbf{Y})$ است و بنابراین، داریم $\frac{c_3}{4} G(\mathbf{X}) \ge_{st} \frac{c_3}{4} G(\mathbf{Y})$ که دلالت بر رابطه $\mathbf{X} \le_{wpd-1} \mathbf{Y}$ دارد. □

یک کاربرد جالب ترتیب $pd-1$ ضعیف، اندازه گیری فاصله بردار تصادفی از ساختار وابستگی هم‌یکنوا است. برای بردار تصادفی $\mathbf{X} = (X_1,\ldots,X_n)$ ، یک ضریب جدید به صورت $H(\mathbf{X}) = \sum_{i=1}^{[\frac{n+1}{2}]} \big(X_{(n+1-i)} - X_{(i)}\big)$ تعریف می کنیم. مشخصاً ضرایب آمار ترتیبی، شرایط تعریف ۱.۴.۲ را برآورده می سازند. به علاوه، دلالت بر رابطه $H(\mathbf{X}) \ge_{st} H(\mathbf{Y})$ دارد. از منظر هندسی، ساختار وابستگی هم‌یکنوا، با خطر $l = \{(t,\ldots,t), t \in \mathbb{R}\}$ نشان داده می‌شود و فاصله بین ساختار وابستگی $\mathbf{X}$ و ساختار وابستگی هم‌یکنوا را می‌توان با فاصله هندسی بین $(X_1,\ldots,X_n)$ و خط ۱ اندازه گیری کرد. اگر از نرم مینکوفسکی مرتبه ۱ استفاده کنیم، یعنی $d_1\big((x_1,\ldots,x_n),(y_1,\ldots,y_n)\big) = \sum_{i=1}^{n} |x_i - y_i|$ ، آنگاه فاصله بین $(X_1,\ldots,X_n)$ و خط l ، $d_1\big((X_1,\ldots,X_n),l\big) = \inf_t d_1\big((X_1,\ldots,X_n),(t,\ldots,t)\big) = \inf_t \sum_{i=1}^{n} |X_i - t|$ است. توجه

داشته باشید که،

$$\sum_{i=۱}^{n}|X_t-t|=\sum_{i=۱}^{n}|X_{(i)}-t|=\sum_{i=۱}^{[\frac{n+۱}{۲}]-۱}\Big(|X_{(i)}-t|+|X_{(n+۱-i)}-t|\Big)+\sum_{i=[\frac{n+۱}{۲}]}^{n+۱-[\frac{n+۱}{۲}]}\Big|X_{(i)}-t\Big|\geq\sum_{i=۱}^{[\frac{n+۱}{۲}]}\big(X_{(n+۱-i)}-X_{(i)}\big)=H(\mathbf{X})$$

تساوی در $t\in\left[X_{\left([\frac{n+۱}{۲}]\right)},X_{\left(n+۱-[\frac{n+۱}{۲}]\right)}\right]$ بدست می‌آید. بنابراین، داریم، $d_۱(\mathbf{X},l)=H(\mathbf{X})$.
در این مورد، شاخص $H(\mathbf{X})$ فاصله یک بردار تصادفی از ساختار وابستگی هم‌یکنوا یا درجه هم‌یکنوایی را نشان می دهد. مشهود است که هر چه $H(\mathbf{X})$ کوچکتر باشد، بردار تصادفی، درجه هم‌یکنوایی بالاتری دارد.
ساختار وابستگی هم‌یکنوایی، کاربردهای گسترده‌ای دارد. در دانش مالی و بیمه، در حالتی که ساختار وابستگی نامشخص است، هم‌یکنوایی، محافظه‌کارانه‌ترین فرض است. از سوی دیگر، هنوز هم ارزیابی شایستگی فرض هم‌یکنوایی بسیار مهم است، که با بررسی فاصله بین ساختار وابستگی مربوطه و هم‌یکنوایی انجام می‌شود. Dhaene و همکاران در سال ۲۰۱۲ رفتار جمعیت را در بازار بورس مطالعه کردند که اساسا، درجه هم‌یکنوایی سهام خاصی را نشان می‌دهد. آنها ضریب رفتار جمعیت را برای اندازه‌گیری درجه حرکت همزمان سهام پیشنهاد کردند که به صورت نسبت واریانس یک سبد و وایانس همان سبد با هم‌یکنوایی تعریف می‌شود. نتایج ارائه شده در این بخش نشان می‌دهند که ضریب $H(\mathbf{X})$ یا کمیت مرتبط با آن مانند $E(H(\mathbf{X}))$ همچنان به عنوان معیاری از درجه هم‌یکنوایی مورد استفاده است بنابراین، مطالعات Dhaene و همکارانش در سال ۲۰۱۲ را تکمیل می‌کند [۲۶].

۵.۲ منحنی های عملکرد

۱.۵.۲ برآورد منحنی‌های تمرکز و لورنز

حق بیمه واقعی در حقیقت مشاهده نمی‌شود و این مسئله، برآورد $CC[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha]$ را دشوار می سازد. دریافته شده است که اتحاد زیر برای هر سطح احتمال α بر قرار است،

$$CC[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha]=CC[Y,\pi(\mathbf{X});\alpha]=\frac{E(YI[\pi(\mathbf{X})\leq F_{\pi}^{-۱}(\alpha)])}{E(Y)}\qquad(۱.۲)$$

اتحاد (۱.۲) از قانون انتظاری کل زیر بدست آمده می آید،

$$E\Big(E(Y|\mathbf{X})I\big(\pi(\mathbf{X})\leq t\big)\Big)=E\Big(E\big(YI[\pi(\mathbf{X})\leq t]|\mathbf{X}\big)\Big)=E\big(YI[\pi(\mathbf{X})\leq t]\big)$$

بنابر این منحنی تمرکز را می‌توان به صورت زیان های کلی Y مربوط به زیرپرتفوی که سهم معین α بیمه نامه ها را با کمترین پیش بینی ها گردآوری می‌کند، تفسیر کرد. فرمول (۱.۲) نشان می دهد که می‌توان حق بیمه خالص $\mu(\mathbf{X})$ را در منحنی تمرکز با پاسخ Y جایگزین کرد. این ویژگی بسیار مهم است زیرا بیمه گر می‌تواند در ارزیابی عملکرد پیش بین مورد مطالعه ، حق بیمه مشاهده نشده صحیح $\mu(\mathbf{X})$ را با مقادیر پاسخی واقعی جایگزین کند. بنابراین، با فرض اینکه برای هر$i = ۱, \ldots, n$ نمونه های (Y_i, X_i) ، مستقل هستند و به صورت یکسان توزیع شده اند، منحنی تمرکز حق بیمه خالص را می‌توان به صورت زیر برآورد کرد،

$$\begin{aligned}\widehat{CC}[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha] &= \widehat{CC}[Y, \pi(\mathbf{X}); \alpha] \\ &= \frac{۱}{n\bar{Y}} \sum_{i|\hat{\pi}(X_i) \le \widehat{F}_{\pi}^{-۱}(\alpha)} Y_i \\ &= \frac{\sum_{i|\hat{\pi}(X_i) \le \widehat{F}_{\pi}^{-۱}(\alpha)} Y_i}{\sum_{i=۱}^{n} Y_i}.\end{aligned}$$

در اینجا $\hat{\pi}$ پیش‌بین برآورد شده است. به عنوان مثال، برای مدل حق بیمه GLM ،مقادیر پیش بینی شده مبتنی برر پارامتر های رگرسیون برآورد شده را نشان می دهد. $\widehat{F_{\pi}}$ برآورد تابع توزیع تجمعی پیش بینی‌های حاصله را نشان می دهد،

$$\widehat{F}_{\pi}(t) = \frac{۱}{n} \sum_{i=۱}^{n} I[\hat{\pi}(X_i) \le t]$$

همتای تجربی $\widehat{CC}$ برای منحنی تمرکز جمعیت CC را می‌توان به صورت زیر تفسیر کرد، نسبت زیان کل تولید شده با آن بیمه‌نامه هایی که پیش‌بین برآورد شده $\hat{\pi}$ آنها در زیر چندک تجربی در سطح α قرار دارد و زیان تجمعی در کل پرتفوی. این بدین معنی است که CC این زیان کلی زیرپرتفوی را به شکل نسبی و به صورت درصد زیان تجمعی در سطح کلی پرتفوی بیان می‌کند.

نسخه تجربی منحنی لورنز را می‌توان به صورت زیر بدست آورد،

$$\widehat{LC}\big(\pi(\mathbf{X}); \alpha\big) = \frac{\sum_{i|\hat{\pi}(\mathbf{X}_i) \le \widehat{F}_{\pi}^{-۱}(\alpha)} \hat{\pi}(\mathbf{X}_i)}{\sum_{i=۱}^{n} \hat{\pi}(X_i)}.$$

به عبارت دیگر، $\widehat{LC}$ درصد درآمد حق بیمه کل متناظر با ٪۱۰۰α حق بیمه های کوچکتر است، که با استفاده از پیش بین π محاسبه شده اند. اگر تعادل کلی $\sum_{i=۱}^{n} \hat{\pi}(\mathbf{X}_i) = \sum_{i=۱}^{n} Y_i$ برقرار باشد (که به عنوان مثال مادامی که توابع اتصال کانونی به کار روند،برای GLM ها برقرار است)، آنگاه $\widehat{LC}$ این نسبت را بر اساس زیان تجمعی در سطح پرتفوی کل بیان می‌کند.

۲.۵.۲ از حق بیمه تا مرتبه ها

به رابطه زیر توجه کنید،

$$\pi(\mathbf{X}) \leq F_{\pi}^{-1}(\alpha) \Leftrightarrow F_{\pi}\big(\pi(\mathbf{X})\big) \leq \alpha$$

این بدین معنی است که کافی است، رتبه‌بندی حاصل از پیش‌بین را در نظر بگیریم، یعنی می‌توانیم هر پیش‌بین $\pi(\mathbf{X})$ را با مرتبه متناظر زیر جایگزین کنیم،

$$\Pi = F_{\pi}\big(\pi(\mathbf{X})\big)$$

که از توزیع یکنواخت واحد تبعیت می‌کند. معنای شهودی Π به صورت زیر است. Π مرتبه یک بیمه‌گذار است وقتی همه قراردادها مطابق با بیمه های متناظرشان (به ترتیب صعودی) رتبه‌بندی شده باشند. در امتیاز گذاری اعتباری، تنها Π و آستانه پذیرش اهمیت دارند. در قیمت گذاری بیمه، مقادی واقعی $\pi(\mathbf{X})$ نیز مهم هستند و این اطلاعات را می‌توان با منحنی لورنز بدست آورد. ما می‌توانیم منحنی تمرکز را بر حسب رابطه وابستگی بین زیان ها و حق بیمه مدل‌سازی شده بیان کنیم. به ویژه می‌توانیم صورت کسری منحنی تمرکز را به صورت زیر بنویسیم.
اگر مفصل زوج (Y, Π) را با C نشان دهیم داریم،

$$\begin{aligned} E\big(YI[\pi(\mathbf{X}) \leq F_{\pi}^{-1}(\alpha)]\big) &= E\big(YI[\Pi \leq \alpha]\big) \\ &= \int_{0}^{\infty}\int_{0}^{1} yI[z \leq \alpha]\mathbf{d}F_{Y,\Pi}(y,z) \\ &= \int_{0}^{\infty}\int_{0}^{1} yI[z \leq \alpha]\mathbf{d}C\big(F_{y}(y), z\big) \\ &= \int_{0}^{1} F_{y}^{-1}(u).\acute{C}_{1}(u,a)\mathbf{d}u \end{aligned}$$

که در آن C'_1 مشتق جزیی C نسبت به آرگومان اول آن است. در واقع، C، مفصل زوج $(Y, \pi(\mathbf{X}))$ است.

۳.۵.۲ ویژگی‌های منحنی لورنز و تمرکز

اکنون برخی خصوصیات مهم منحنی‌های لورنز و تمرکز را به صورت خلاصه مرور می‌کنیم. یادآوری می‌کنیم که منحنی لورنز، خصوصیات منحنی تمرکز را به عنوان یک حالت خاص به ارث می‌برد.

یکنوایی

منحنی تمرکز مبتنی بر تابع $t \mapsto E\big(YI[\pi(\mathbf{X}) \leq t]\big)/E(Y)$ است. که در آن چندک های $\pi(\mathbf{X})$ ارزیابی شده است. این تابع،مشخصا غیرنزولی است و از $(\circ,\circ)$ شروع شده و به $(۱,۱)$ می رسد. بنابراین $\alpha \mapsto CC[Y,\pi(\mathbf{X});\alpha]$ غیر نزولی است و رابطه زیر را برآورد می سازد،

$$\lim_{\alpha\to\circ} CC[Y,\pi(\mathbf{X});\alpha] = \circ \text{ و } \lim_{\alpha\to۱} CC[Y,\pi(\mathbf{X});\alpha] = ۱$$

خط استقلال/تساوی

در حالت خاص که حق بیمه، هیچ اطلاعاتی راجع به پاسخ نمی‌دهد (چون Y و $\pi(\mathbf{X})$ دو به دو مستقل هستند)، منحنی تمرکز، یک خط ۴۵ درجه است که در مقالات به خط استقلال مشهور است. طبق قرارداد، اگر Y و $\pi(\mathbf{X})$ مستقل باشند، اتحاد زیر برقرار است،

$$CC[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha] = \frac{E(Y)P\big[\pi(\mathbf{X}) \leq F_{\pi}^{-۱}(\alpha)\big]}{E(Y)} = \alpha$$

اکنون، مکان منحنی تمرکز را نسبت به خط ۴۵ درجه مطالعه می‌کنیم. اگر π اطلاعات زیادی راجع به حق بیمه $\mu(\mathbf{X})$ به همراه داشته باشد، بدین معنی است که این متغیرهای تصادفی وابستگی قوی دارند و منحنی تمرکز باید دور از خط استقلال باشد. به علاوه، شکل منحنی تمرکز، به نوع رابطه بین $\mu(\mathbf{X})$ و $\pi(\mathbf{X})$ بستگی دارد. نتیجه بعدی نشان می‌دهد که تحت وابستگی مثبت ضعیف، هر منحنی تمرکز، زیر خط استقلال قرار می‌گیرد.

ویژگی ۲ـ۵ـ۱ اگر $\mu(\mathbf{X})$ مقدار انتظاری مثبت وابسته به $\pi(\mathbf{X})$ باشد. یعنی اگر نامساوی زیر برای همه t ها برقرار باشد،

$$E\big(\mu(\mathbf{X})\big) \geq E(\mu(\mathbf{X})|\pi(\mathbf{X}) \leq t)$$

آنگاه منحنی تمرکز زیر خط ۴۵ درجه قرار می‌گیرد یعنی برای همه سطوح احتمال α داریم،

$$CC[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha] \leq \alpha$$

برای اثبات کافی است بنویسیم،

$$\begin{aligned}\frac{E\big(\mu(\mathbf{X})I[\pi(\mathbf{X}) \leq t]\big)}{E(Y)} &= \frac{P[\pi(\mathbf{X}) \leq t]E(\mu(\mathbf{X})|\pi(\mathbf{X}) \leq t)}{E(Y)} \\ &\leq P[\pi(\mathbf{X}) \leq t]\end{aligned}$$

سپس با جایگزینی t با $F_{\pi}^{-۱}\alpha$ ، قضیه اثبات می‌شود.

با در نظر گرفتن $LC[\pi(\mathbf{X});\alpha]$، خط ۴۵ درجه رامی‌توان به یک حالت محدودکننده دیگر به نام خط تساوی منتسب کرد. فرض کنید که π ثابت باشد، یعنی داشته باشیم $\pi(\mathbf{X}) = E(Y)$

. این رابطه می‌تواند بر اساس این حقیقت باشد که هیچکدام از ویژگی‌های موجود در $\mathbf{X}$ به پاسخ Y ارتباط نمی‌یابند (در نتیجه با خط استقلال ارتباط ندارند). بنابراین داریم،

$$\pi(\mathbf{X}) = E(Y) \Rightarrow LC[\pi(\mathbf{X}); \alpha] = \alpha$$

به گونه‌ای که این مورد خاص، متناظر با خط ۴۵ درجه است. توجه داشته باشید که باید این مورد محدود کننده را با دقت به کار گیریم زیرا پیش‌بین ثابت یک متغیر تصادفی پیوسته نیست (زیرا با فرضیات فنی بیان شده در بخش ۵.۱ در تناقض است و برخی از فرمول‌های ارائه شده در این مقاله را نقض می‌کند).

تحدب

نتیجه بعدی، شرط وابستگی مثبت را بیان می‌کند که تحت این شرط، منحنی تمرکز محدب است. باز هم شکل منحنی به نوع رابطه موجود بین پاسخ و پیش‌بین بستگی دارد. برای اثبات می‌توانید به مرجع [۷۲] مراجعه کنید.

ویژگی ۲ـ۵ـ۳ منحنی تمرکز $\alpha \mapsto CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha]$ محدب است اگر و تنها اگر $\mu(\mathbf{X})$، رگرسیون مثبت باشد که وابسته به $\pi(\mathbf{X})$ است یعنی اگر تابع زیر

$$t \mapsto E(\mu(\mathbf{X})|\pi(\mathbf{X}) = t) \qquad (۲.۲)$$

غیرنزولی باشد.

اکنون به صورت مختصر در مورد شرط (۲.۲) صحبت می‌کنیم. در کل، شرط $\pi(\mathbf{X}) = t$ مقدار $\mu(\mathbf{X})$ را ثابت نگه نمی‌دارد. برای تحقق این امر، حق بیمه GLM خطی – لگاریتمی کلاسیکی زیر را در نظر بگیرید،

$$\pi(\mathbf{X}) = \exp\left(\beta_\circ + \sum_{j=1}^{p} \beta_j X_j\right)$$

در این مورد، $\pi(\mathbf{X}) = t \Leftrightarrow \sum_{j=1}^{p} \beta_j X_j = \ln t - \beta_\circ$ برقرار است، بنابراین شرط (۲.۲) به صورت زیر بازنویسی می‌شود،

$$E\Big(\mu(\mathbf{X})|\sum_{j=1}^{p} \beta_j X_j = \ln t - \beta_\circ\Big)$$

از آنجا که ویژگی‌های X_j را نه تنها از طریق ترکیب خطی $\sum_{j=1}^{p} \beta_j X_j$ بلکه به شکل دلخواه می‌توان در $\mu(\mathbf{X})$ وارد کرد، $\mu(\mathbf{X})$ تصادفی مشروط به امتیاز GLM باقی می‌ماند.

به پاسخ Y باز گردیم، تحدب منحنی تمرکز $\alpha \mapsto CC[Y, \pi(\mathbf{X}); \alpha]$ تضمین می‌کند که نموهای تابع زیر

$$CC[Y, \pi(\mathbf{X}); \alpha + \Delta] - CC[Y, \pi(\mathbf{X}); \alpha] = \frac{E\big(Y I[F_\pi^{-1}(\alpha) < \pi(X) \leq F_\pi^{-1}(\alpha + \Delta)]\big)}{E(Y)}$$

برای هر Δ مثبت در α غیرنزولی هستند، به گونه‌ای که $\alpha + \Delta \leq 1$ است. این بدین معنی است که، زیرپرتفوی‌های ایجاد شده، از طریق جداسازی یک سهم Δ از بیمه‌نامه‌ها با حق بیمه‌های محدود بین $F_\pi^{-1}(\alpha)$ و $F_\pi^{-1}(\alpha + \Delta)$ با افزایش α، سهم زیان را به صورت میانگین افزایش می‌دهند.

این ویژگی مرتبط با نمودارهای ترفیع است، که در مقاله Tevet در سال ۲۰۱۳ توصیف شده است [۶۸]. برای رسم این گراف‌ها، مجموعه داده باید بر اساس مقادیر $\pi(\mathbf{X})$ مرتب شود. آنگاه داده‌ها بر اساس چندک‌ها، به صورت رده‌هایی با جمعیت مساوی طبقه‌بندی می‌شوند. در هر دسته، میانگین زیان پیش‌بینی شده، به کمک π و هزینه زیان واقعی Y محاسبه می‌شود. آنگاه، هزینه‌های زیان پیش‌بینی شده میانگین و زیان واقعی میانگین برای هر رده به صورت گراف رسم می‌شوند. تحلیل‌گر باید برای ارزیابی استواری π بررسی کند، که آیا وقتی به سمت دسته‌های بالاتر می‌رویم، پاسخ واقعی به صورت یکنوا افزایش می‌یابد یا خیر (طبق تعریف، در مورد پیش‌بینی‌ها اینگونه است). ویژگی ۳.۵.۲، شرط لازم برای مشاهده روند افزایشی در این نمودار ترفیع را مشخص می‌کند.

از آنجا که وابستگی رگرسیون مثبت دلالت بر وابستگی مقدار انتظاری مثبت دارد (به عنوان مثال به [۷۰] مراجعه کنید)، شرط ویژگی ۳.۵.۲ تضمین می‌کند که منحنی تمرکز غیرنزولی و محدب است و از $(0, 0)$ شروع شده و در $(1, 1)$ پایان می‌یابد و گراف آن در زیر خط ۴۵ درجه قرار می‌گیرد. توجه کنید که هر منحنی لورنز بر اساس ویژگی ۳.۵.۲ محدب است.

۴.۵.۲ اندازه‌گیری نیکویی برازش

از اینجا به بعد، عملکرد پیش‌بین $\pi(\mathbf{X})$ را به کمک مکان‌های نسبی دو منحنی ارزیابی می‌کنیم.

$$\alpha \mapsto CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha] \text{ و } \alpha \mapsto LC[\pi(\mathbf{X}); \alpha]$$

عبارت اول، سهم حق بیمه‌های گردآوری شده از $100\alpha\%$ بیمه‌نامه‌ها از پرتفوی با کمترین مقادیر $\pi(\mathbf{X})$ را نشان می‌دهد. عبارت دوم، سهم متناظر حق بیمه واقعی را بدست می‌دهد که باید از این زیرپرتفوی گردآوری شود. از آنجا که درآمد انتظاری کل π و μ مطابق با کل زیان مورد انتظار تحت رابطه (۱.۱) است، دو نسبت به صورت مستقیم قابل مقایسه هستند. در واقع، ما می‌توانیم دو درصد را با همدیگر مقایسه کنیم. یکی درصدی که با $\pi(0)$ بدست می‌آید و دیگری متناظر با $\mu(0)$ است. ما به عنوان بیمه گر، دوست داریم که گراف CC تا حد امکان

به گراف LC نزدیک باشد. به عبارت دیگر، هر چه مساحت بین دو منحنی کوچکتر باشد، بهتر است.

۶.۲ مقایسه عملکرد های دو پیش‌بین

۱.۶.۲ منحنی های تمرکز و لورنز

فرض کنید که دو پیش‌بین π_1 و π_2 دارید. هر دو تلاش می‌کنند تا حق بیمه خالص مجهول $\mu(\mathbf{X})$ را پیش‌بینی کنند. روش‌های مختلفی برای بدست آوردن این پیش‌بین‌ها گزارش شده است که از میانگین براکت تجربی تا GLM کلاسیک و روش‌های پایه درختی یا شبکه‌های عصبی متغیر هستند.
توجه داشته باشید که این پیش‌بین‌ها ممکن است از نظر شکل تابعی (π_1 به جای π_2) ویا اطلاعاتی (X_1به جای X_2) که بر اساس آن عمل می‌کنند فرق داشته باشند. دو نکته مهم در زمان ارزیابی عملکرد دو پیش‌بین وجود دارد. اولین مورد، تغییرپذیری نسبی آنها یعنی توانایی آنها برای شناسایی پروفایل‌های ریسک مختلف است. مورد دوم، همبستگی آنها با $\mu(\mathbf{X})$ است، یعنی میزان اطلاعاتی که درباره حق بیمه خالص می‌دهند.
به نظر می‌رسد که ترتیب محدب، یک ابزار مناسب برای اندازه‌گیری درجه جابه‌جایی ناشی از حق بیمه منتخب π باشد. این ابزار احتمالاتی، در واقع، ناسازگاری بین ارزان‌ترین و پرهزینه‌ترین پروفایل‌های ریسک شناسایی شده در مدل را ارزیابی می‌کند. در این حالت به نظر می‌رسد که جایگزینی π با یک متغیر دیگر که مبتنی بر ترتیب محدب است، یک استراتژی نویدبخش باشد. با در نظر گرفتن حالت خاص $g(X) = x^2$ در رابطه (۲.۱) درمی‌یابیم که،

$$\pi_2(\mathbf{X}_2) \preceq_{cx} \pi_1(\mathbf{X}_1) \Rightarrow Var(\pi_2(\mathbf{X}_2)) \leq Var(\pi_1(\mathbf{X}_1)) \qquad (۱.۲)$$

این مسئله شرح می‌دهد که چرا $\preceq_{cx}$ یک ترتیب تغییرپذیر است، زیرا تنها بر متغیرهای تصادفی با مقدار انتظاری یکسان اعمال می‌شود و پراکندگی نسبی آنها را مقایسه می‌کند. ترتیب محدب یک مقایسه پیچیده‌تر است که تنها بر روی واریانس متمرکز نیست، با این حال رابطه (۱.۲) نشان می‌دهد که با این رویکرد مطابقت دارد. بنابراین می‌توانیم $\pi_2(\mathbf{X}_2) \preceq_{cx} \pi_1(\mathbf{X}_1)$ را تفسیر کنیم زیرا $\pi_1(\mathbf{X}_1)$ تغییرپذیری بیشتری نسبت به $\pi_2(\mathbf{X}_2)$ دارد و باید در نظر داشت که تغییرپذیری مدنظر، ورای مقایسه ساده انحراف معیارهاست.
تعریف بعدی، منحنی‌های عملکرد Gourieroux and Jasiak که سال ۲۰۰۷ در مرجع [۳۹] بیان شده است را به زیان‌های عمومی‌تر ورای پاسخ‌های دودویی توسعه می‌دهد [۳۹].

تعریف ۱.۶.۲. حق بیمه $\pi_1(\mathbf{X}_1)$ تمایزپذیرتر از $\pi_2(\mathbf{X}_2)$ است اگر و تنها اگر داشته باشیم،

$$\pi_2(\mathbf{X}_2) \preceq_{cx} \pi_1(\mathbf{X}_1) \Leftrightarrow LC[\pi_1(\mathbf{X}_1)] \leq LC[\pi_2(\mathbf{X}_2); \alpha]$$

برای همه α ها برقرار باشد و نامساوی

$$CC[\mu(\mathbf{X}), \pi_1(\mathbf{X}_1); \alpha] \leq CC[\mu(\mathbf{X}), \pi_2(\mathbf{X}_2); \alpha]$$

برای همه سطوح احتمال α وجود داشته باشد.

مرتب‌سازی منحنی‌های تمرکز بدین معنی است، که سهم درآمد حق بیمه کل متناظر با زیرپرتفوهایی که $100\alpha\%$ بیمه نامه‌ها را گردآوری کرده‌اند، با کمترین حق بیمه، برای π_1 کوچکتر از π_2 است. این شرط را می‌توان به صورت زیر بازنویسی کرد. توابع توزیع دو پیش‌بین π_1و π_2را به ترتیب با F_{π_1} و F_{π_2} نشان می‌دهیم. فرض می‌شود که هر دو F_{π_k} پیوسته و اکیداً صعودی هستند و $k = 1, 2$ است. امتیازات $\Pi_1 = F_{\pi_1}(\pi_1(\mathbf{X}_1))$ و $\Pi_2 = F_{\pi_k}(\pi_2(\mathbf{X}_2))$ را تعریف می‌کنیم که هر دو به صورت یکنواخت در بازه واحد $[0, 1]$ توزیع شده‌اند. بنابراین

$$E(Y|\pi_1(\mathbf{X}_1) \leq F^{-1}_{\pi_1}(\alpha)) \leq E(Y|\pi_2(\mathbf{X}_2) \leq F^{-1}_{\pi_2}(\alpha))$$
$$\Leftrightarrow E(Y|\Pi_1 \leq \alpha) \leq E(Y|\Pi_2 \leq \alpha)$$

بنابراین امید ریاضی Yباید مقدار مثبت‌تری باشد که به جای Π_2 وابسته به Π_1 است، زیرا کاهش مقدار امید ریاضی با توجه به $\Pi_k \leq \alpha$ در Π_1 بزرگتر از Π_2 است. به عنوان مثال این نوع شرط را می‌توان در مراجع [۵۴]، [۱۸]، [۶۴] مشاهده کرد.
نتیجه بعدی، یک شرط کافی برای مقایسه پیش‌بین‌های رقیب از نظر رابطه تحدب و مرتبه هماهنگی با پاسخ را نشان می‌دهد. از مطالعه Denuit و همکارانش در سال ۲۰۰۵ می‌دانیم که دو متغیر تصادفی هماهنگ هستند، اگر هر دو با همدیگر بزرگ یا با همدیگر کوچک شوند [۱۹]. مرتبه هماهنگی بیانگر این هستند که مقادیر کوچک و بزرگ معمولاً مرتبط با توزیعی هستند، که بر دیگری غالب است. اکنون دو زوج تصادفی (Z_1, Z_2) و (V_1, V_2) با توزیع‌های حاشیه‌ای یکسان را در نظر می‌گیریم، یعنی برای $k \in \{1, 2\}$ داریم،
$P[Z_k \leq t] = P[V_k \leq t] = F_k(t)$.

اگر

$$P[Z_1 \le t_1, Z_2 \le t_2] \le P[V_1 \le t_1, V_2 \le t_2] \qquad (۲.۲)$$

برای همه t_1 وt_2 ها برقرار باشد، یا هم ارز آن، اگر

$$P[Z_1 > t_1, Z_2 > t_2] > P[V_1 > t_1, V_2 > t_2] \qquad (۳.۲)$$

برای همه t_1 وt_2 ها برقرار باشد، آنگاه می‌گوییم (Z_1, Z_2)، هماهنگی کمتری نسبت به (V_1, V_2) دارند. این نکته از اینجا به بعد به صورت $(Z_1, Z_2) \preceq_{conc} (V_1, V_2)$ نشان داده می‌شود. معنای شهودی رتبه‌بندی نسبت به $\preceq_{conc}$ را می‌توان ازرابطه (۲.۲) دریافت. در واقع، $P[Z_1 \le t_1, Z_2 \le t_2]$ و $P[V_1 \le t_1, V_2 \le t_2]$ به صورت زیر خوانده می‌شوند، «Z_1و Z_2هر دو کوچک هستند» و «V_1 و V_2هر دو کوچک هستند». کوچک بدین معنی که Z_1 (V_1) کوچکتر از آستانه t_1 است و Z_2 (V_2) کوچکتر از آستانه t_2 است. بنابراین رابطه (۲.۲)بدین معنی است که زمانی که $(Z_1, Z_2) \preceq_{cx} (V_1, V_2)$ برقرار باشد، احتمال اینکه V_1 و V_2 هر دو کوچک باشند بیشتر از این احتمال متناظر است که Z_1و Z_2 هر دو کوچک باشند. همچنین از رابطه (۳.۲)دریافته می‌شود که $(Z_1, Z_2) \preceq_{conc} (V_1, V_2)$بدین معنی است که احتمال اینکه Z_1و Z_2هر دو بزرگ باشند، کمتر از احتمال متناظر برای V_1 و V_2است. این توضیح با مفهوم شهودی زیر متناظر است، « (V_1, V_2)وابستگی مثبت بیشتری نسبت به (Z_1, Z_2)دارند». به علاوه، ضریب همبستگی پواسون و همچنین ضرایب همبستگی[1] مرتبه کندال و اسپیرمن همگی با ترتیب هماهنگی مطابقت دارند. این عوامل، معنای شهودی $\preceq_{conc}$را به عنوان ابزاری برای مقایسه شدت وابستگی، تقویت می‌کنند. اکنون می‌توانیم ارتباط بین ترتیب هماهنگی و توان تشخیصی یک پیش‌بین را بیابیم.

ویژگی۲ـ۶ـ۱. اگر

$$\pi_2(\mathbf{X}_2) \preceq_{cx} \pi_1(\mathbf{X}_1) \text{ و } (Y, \Pi_2) \preceq_{cx} (Y, \Pi_1)$$

آنگاه پیش‌بین $\pi_1(\mathbf{X}_1)$ قدرت تشخیصی بیشتری برای پاسخ Y نسبت به پیش‌بین $\pi_2(\mathbf{X}_2)$ دارد.

برهان. نتیجه از رابطه زیر بدست می‌آید:

$$E(Y|\Pi_2 \le \alpha) - E(Y|\Pi_1 \le \alpha) = \frac{1}{\alpha} \int_0^\infty \big(P[Y \le y, \Pi_1 \le \alpha] - P[Y \le y, \Pi_2 \le \alpha]\big) \mathrm{d}y$$

که در واقع مثبت است، اگر (Y, Π_1) هماهنگ‌تر از (Y, Π_2) باشد، زیرا انتگرال برای همه مقادیر Y در این حالت غیرمنفی است.

□

[1] correlation coefficient

بنابراین، می‌ببنیم که $\pi_1(\mathbf{X}_1)$ برای پاسخ Y نسبت به $\pi_2(\mathbf{X}_2)$ دارد اگر $\pi_1(\mathbf{X}_1)$ همزمان تغییرپذیری بیشتر (از نظر ترتیب تحدب) و همبستگی بیشتری (از نظر وابستگی مقدار انتظاری مثبت یا ترتیب هماهنگی قوی‌تر) با پاسخ Y نسبت به $\pi_2(\mathbf{X}_2)$ داشته باشد.

۲.۶.۲ منحنی‌های تمرکز و لورنز انتگرال‌دار

مرجع پیشنهاد شده در تعریف ۱.۶.۲، تنها یک رتبه‌بندی جزئی را انجام می‌دهد. دو پیش‌بین ممکن است غیرقابل مقایسه باشند، زیرا منحنی‌های تمرکز یا لورنز نسبی آنها همدیگر را قطع می‌کنند. مثلاً یک پیش‌بین برای ریسک‌های پایین بهتر و برای ریسک‌های بالا بدتر است. در چنین حالتی، می‌توانیم مقایسه‌ای روی انتگرال منحنی‌های تمرکز داشته باشیم. یعنی می‌توان منحنی تمرکز انتگرال‌دار را به صورت زیر تعریف کرد،

$$\begin{aligned} ICC\big[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha\big] &= \int_0^\alpha CC\big[\mu(\mathbf{X}),\pi(\mathbf{X});\xi\big]d\xi \\ &= \int_0^\alpha \frac{E\big(\mu(\mathbf{X})I[\Pi\le\xi]\big)}{E(Y)}d\xi \\ &= \frac{E\big(\mu(\mathbf{X})(\alpha-\Pi)_+\big)}{E(Y)} \\ &= \frac{Cov\big(\mu(\mathbf{X}),(\alpha-\Pi)_+\big)}{E(Y)} + E\big((\alpha-\Pi)_+\big) \end{aligned}$$

جمله اول از همبستگی بین پاسخ و پیش‌بین بدست می‌آید در حالیکه جمله دوم تنها یک ثابت است. زیرا Π دارای توزیع یکنواخت واحد است.

$$E\big((\alpha-\Pi)_+\big) = \int_0^\alpha (\alpha-\xi)\mathrm{d}\xi = \frac{\alpha^2}{2}$$

منظور ما از ICC، انتگرال منحنی تمرکز در کل بازه $[0,1]$است، یعنی داریم،

$$\begin{aligned} ICC &= ICC\big[\mu(\mathbf{X}),\pi(\mathbf{X});1\big] \\ &= \frac{Cov\big(\mu(\mathbf{X}),1-\Pi\big)}{E(Y)} + \frac{1}{2} \\ &= \frac{1}{2} - \frac{Cov\big(\mu(\mathbf{X}),\Pi\big)}{E(Y)} \end{aligned}$$

باز هم چون داریم،

$$\begin{aligned} E\big((\alpha-\Pi)_+\big) &= E\Big(E\big((\alpha-\Pi)_+|\mathbf{X}\big)\Big) \\ &= E\big(\mu(\mathbf{X})(\alpha-\Pi)_+\big) \end{aligned}$$

می‌توانیم در تعریف منحنی تمرکز انتگرال‌دار، $\mu(\mathbf{X})$ را با Y جایگزین کنیم. این بدین معنی است که می‌توانیم از آن برای اندازه‌گیری عملکرد $\pi(\mathbf{X})$ در تولید حق بیمه خالص مجهول

$\mu(\mathbf{X})$ بهره بگیریم.
اکنون یک تفسیر شهودی برای ICC ارائه می‌کنیم. $۱۰۰\alpha\%$ بیمه نامه‌ها با کمترین مقدار π را به صورت $(\alpha - \Pi)_+ = ۰$ برای $\alpha \leq \Pi$ درنظر می‌گیریم. اکنون، ICC مبتنی بر کواریانس بین $\mu(X)$ و $(\alpha - \Pi)_+$ است. ایده این است که هر چه Πنسبت به αکوچکتر باشد (یعنی $(\alpha - \Pi)_+$ بزرگتر باشد)، حق بیمه واقعی کوچکتر خواهد بود. بنابراین، رابطه مثبت بین $\pi(X)$ و $\mu(X)$ به معنی کوواریانس منفی بین $\mu(X)$ و $(\alpha - \Pi)_+$ است. و هر چه جمله کوواریانس وارد شده در تجزیه ICC منفی‌تر باشد، حق بیمه منتخب متناظر بهتر است.
اگر برای منحنی لورنز نیز همین روند را اتخاذ کنیم، می‌توانیم منحنی لورنز انتگرال دار را به صورت زیر تعریف کنیم،

$$\begin{aligned} ICC\big[\pi(\mathbf{X}); \alpha\big] &= \int_{۰}^{\alpha} \frac{E\big(\pi(\mathbf{X}) I[\Pi \leq \xi]\big)}{E(\pi(\mathbf{X}))} \mathrm{d}\xi \\ &= \frac{E\big(\pi(\mathbf{X})(\alpha - \Pi)_+\big)}{E\big(\pi(\mathbf{X})\big)} \\ &= \frac{Cov\big(\pi(\mathbf{X}), (\alpha - \Pi)_+\big)}{E\big(\pi(\mathbf{X})\big)} + E\big((\alpha - \Pi)_+\big). \end{aligned}$$

۷.۲ برخی معیار های سودمند بیمه

۱.۷.۲ شاخص جینی

مطالعات تجربی متعددی، از شاخص‌های جینی مبتنی بر منحنی لورنز پیش‌بین بهره می‌برند. اکنون دلیل استفاده از این پیش‌بین را شرح می‌دهیم. توجه داشته باشید که بحث مشابهی را می‌توان در بخش ۲ – ۱ مرجع [۷۱] یافت. یادآوری می کنیم که اختلاف میانگین جینی، یک معیار احتمالی برای تغییرپذیری است، که برای متغیر تصادفی پیوسته غیرمنفی Z به صورت زیر تعریف می‌شود.

$$Gini[Z] = E\big(|Z_۱ - Z_۲|\big) = E\big(max\{Z_۱, Z_۲\}\big) - E\big(min\{Z_۱, Z_۲\}\big)$$

که در آن، $Z_۱$ و $Z_۲$ مستقل هستند و همانند Z توزیع شده‌اند. این رابطه، میانگین تفاضل مطلق بین دو مشاهده که مانند Z توزیع شده‌اند را نشان می‌دهد. این ضریب ارتباط نزدیکی با واریانس دارد و می‌توان هم‌ارز آن را به شکل زیر بیان کرد،

$$Var[Z] = \frac{۱}{۲} E\big((Z_۱ - Z_۲)^۲\big)$$

در اینجا، میانگین تفاضل جینی مبتنی بر تفاضل مطلق است، در حالیکه در واریانس از تفاضل مربعی استفاده می‌شود. اگر Z پیوسته باشد، آنگاه می‌توان نشان داد که،

$$Gini[Z] = ۴Cov(Z, F_Z(Z))$$

بنابراین، میانگین تفاضل جینی، ارتباط بین متغیر تصادفی و رتبه آن را اندازه‌گیری می‌کند. به عبارت دیگر، با در نظر گرفتن دنباله‌ای از مشاهدات که به ترتیب صعودی رتبه‌بندی شده‌اند، میانگین تفاضل جینی، رابطه بین مقدار واقعی Z و مکان آن در دنباله را نشان می‌دهد. مشخصاً هر چه تغییرپذیری Z بیشتر باشد، مقدار واقعی آن در رتبه بالا بیشتر است (یعنی در میان بزرگترین مشاهدات است). در حالیکه میانگین تفاضل جینی کوچکتر بیانگر این است که مشاهدات بیشتر حول مقدار مرکزی متمرکز شده‌اند.
این فرمول، میانگین تفاضل جینی را به منحنی لورنز مرتبط می‌سازد. انحراف مطلق پایین $Z_۱$ را به صورت $E((t-Z_۱)I[Z_۱ \le t])$ تعریف می‌کنیم و t را با $Z_۲$ جایگزین می‌کنیم، تا رابطه زیر بدست آید.

$$E\big((Z_۱ - Z_۲)I[Z_۱ \le Z_۲]\big) = \frac{۱}{۲}E\big(|Z_۱ - Z_۲|\big) = \frac{۱}{۲}Gini[Z] = ۲Cov\big(Z, F_Z(Z)\big)$$

اکنون، ناحیه بین خط یکانی و منحنی لورنز Zبه صورت $\frac{۱}{E(Z)}Cov(Z, F_Z(Z))$ است. بنابراین، هر چه میانگین تفاضل جینی بالاتر باشد، منحنی لورنز پیش‌بین فاصله بیشتری از خط ۴۵ درجه (یعنی از منحنی لورنز پیش‌بین غیرآگاهی‌بخش که پیوسته برابر با $E(Z)$ است) دارد. بدین دلیل است که حق بیمه‌های منتخب با میانگین تفاضل جینی بزرگتر مرجع هستند.
شاخص جینی، میانگین تفاضل جینی تقسیم بر دو برابر میانگین است. این شاخص به نسبت تمرکز معروف است و ناحیه بین خط ۴۵ درجه و منحنی لورنز واقعی را تقسیم بر ناحیه بین خط ۴۵ درجه و منحنی لورنزی که مقدار بیشینه این شاخص است، نشان می‌دهد.

۲.۷.۲ معیار ABC (Area Between two Curves)

میانگین تفاضل جینی، ناحیه بین خط ۴۵ درجه و منحنی لورنز را اندازه‌گیری می‌کند. همانگونه که پیشتر شرح دادیم، بهتر است هر دو منحنی را به صورت همزمان در نظر بگیریم، منحنی لورنز پیش‌بین و منحنی تمرکز پاسخ نسبت به پیش‌بین. بنابراین، ما به فاصله بین دو منحنی $CC[\mu(\mathbf{X}), \pi(\mathbf{X}); \alpha]$ و $LC[\pi(\mathbf{X}); \alpha]$ علاقه‌مند هستیم، با توجه به اینکه می‌دانیم در $\pi(\mathbf{X}) = \mu(\mathbf{X})$ تلاقی می‌کنند، همچنین ناحیه بین دو منحنی CC و LC شاخص بهتری برای عملکرد یک پیش‌بین معین است. ناحیه بین منحنی‌ها که به اختصار ABC نامیده می‌شود،

به صورت زیر بیان می‌شود،

$$\begin{aligned} ABC[\pi(\mathbf{X})] &= \int_{\circ}^{1} \big(CC[Y,\pi(\mathbf{X});\alpha] - LC[\pi(\mathbf{X});\alpha]\big)\mathbf{d}\alpha \\ &= \frac{1}{E(\pi(\mathbf{X}))}\int_{\circ}^{1}\big(E(YI[\Pi\le\alpha]) - E(\pi(\mathbf{X})I[\Pi\le\alpha])\big)\mathbf{d}\alpha \\ &= \frac{1}{E(\pi(\mathbf{X}))}\int_{\circ}^{1}\int_{\circ}^{\infty}\big(P(\pi(\mathbf{X})\le y,\Pi\le\alpha) - P(Y\le y,\Pi\le\alpha)\big)\mathbf{d}y\mathbf{d}\alpha \\ &= \frac{1}{E(\pi(\mathbf{X}))}\big(Cov(\pi(\mathbf{X}),\Pi) - Cov(y,\Pi)\big) \end{aligned} \tag{۱.۲}$$

که در آن، تفاضل بین میانگین شاخص جینی پیش‌بین $\pi(\mathbf{X})$ (تا ضریب ۴) و کوواریانس جینی پاسخ Y و پیش‌بین $\pi(\mathbf{X})$ بدست می‌آید. فرض کنید رابطه (۱.۲) را بتوانیم به صورت زیر بازنویسی کنیم،

$$ABC[\pi(\mathbf{X})] = \frac{1}{E\big(\pi(\mathbf{X})\big)}Cov\big(\pi(\mathbf{X}) - Y,\Pi\big)$$

بنابراین، اگر $\pi(\mathbf{X}) - Y$ را منفعت مرتبط با یک بیمه نامه در نظر بگیریم، آنگاه ABC را می‌توان متناسب با کوواریانس بین منفعت و مرتبه حق بیمه‌ها در نظر گرفت. این تفسیر مشابه تفسیر ارائه شده در مطالعه Frees و همکارانش در سال ۲۰۱۳ برای شاخص جینی است [۳۲].
معیارهایی مانند ICC و ABC پیشنهادی، شباهت‌هایی با قوانین امتیازدهی مناسبی که برای ارزیابی کیفیت پیش‌بینی‌های احتمالاتی به کار می‌روند (مرجع [۳۷] را ببینید) دارند. با این وجود، همانگونه که در بخش ۵.۱ بیان شد، هدف قیمت‌گذاری بیمه، برآورد امید ریاضی شرطی است، نه پیش‌بینی پاسخ در حالیکه هدف قوانین امتیازدهی، توزیع پیش‌بین است. شایان ذکر است که ارتباطاتی بین این دو مفهوم وجود دارد. به عنوان مثال، امتیاز احتمال رتبه‌بندی شده پیوسته (یا CRPS ، رابطه (۲۰) بیان شده در مرجع [۳۷] ببینید) همانند ABC، شامل میانگین تفاضل جینی است.

نکته ۲ـ۷ـ۱ توجه کنید که برای $(\mu(\mathbf{X}),\pi(\mathbf{X}))$ یکنوا، برای همه $\alpha\in(\circ,1)$ داریم، $CC[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha] = LC[\pi(\mathbf{X});\alpha]$ اگر و تنها اگر $\mu(\mathbf{X}) = \pi(\mathbf{X})$ باشد. در واقع، برای $(\mu(\mathbf{X}),\pi(\mathbf{X}))$ یکنوا می‌توانیم بنویسیم،

$$\mu(\mathbf{X}) = F_{\mu}^{-1}(U) \text{ و } \pi(\mathbf{X}) = F_{\pi}^{-1}(U)$$

که در آن، U به صورت یکنواخت روی بازه واحد $[\circ,1]$ توزیع شده است و F_{μ}^{-1} تابع چندک مربوط به تابع توزیع F_{μ} روی $\mu(\mathbf{X})$ است. بنابراین، از آنجا که به $E(\mu(\mathbf{X})) = E(\pi(\mathbf{X}))$ نیاز داریم، رابطه $CC[\mu(\mathbf{X}),\pi(\mathbf{X});\alpha] = LC[\pi(\mathbf{X});\alpha]$ را برای همه $\alpha\in(\circ,1)$ ها بدست می‌آوریم

اگر و تنها اگر داشته باشیم،

$$
\begin{aligned}
&E\big(\mu(\mathbf{X})I[\pi(\mathbf{X}) \le F_{\pi}^{-1}(\alpha)]\big) = E\big(\pi(\mathbf{X})I[\pi(\mathbf{X}) \le F_{\pi}^{-1}(\alpha)]\big) \\
&\Leftrightarrow E\big(F_{\mu}^{-1}(U)I[F_{\mu}^{-1}(U) \le F_{\pi}^{-1}(\alpha)]\big) = E\big(F_{\pi}^{-1}(U)I[F_{\pi}^{-1}(U) \le F_{\pi}^{-1}(\alpha)]\big) \\
&\Leftrightarrow \int_{\circ}^{\alpha} F_{\mu}^{-1}(u)\mathbf{d}u = \int_{\circ}^{\alpha} F_{\pi}^{-1}(u)\mathbf{d}u \qquad (۲.۲)
\end{aligned}
$$

در نتیجه، از آنجا که رابطه (۲.۲) باید برای همه $\alpha \in (\circ, ۱)$ برقرار باشد، نتیجه می‌گیریم که $F_{\mu}^{-1}(U) = F_{\pi}^{-1}(u)$ برای همه $u \in (\circ, ۱)$ برقرار است و بنابراین داریم، $\mu(\mathbf{X}) = \pi(\mathbf{X})$

۸.۲ مثال‌های عددی

۱.۸.۲ فرضیات

فرض کنید $\pi(\mathbf{X})$ از توزیع گاما با میانگین μ واریانس $\sigma^{۲}$ تبعیت می‌کند و در نتیجه با $Gam(\mu, \sigma^{۲})$ نشان داده می‌شود. ما $\mu = ۱$ را در نظر می‌گیریم. این پیش‌بین‌ها را می‌توان بر حسب $\preceq_{cx}$ نسبت به α مرتب‌سازی کرد.
اکنون دو توزیع را برای امید ریاضی شرطی $\mu(\mathbf{X})$ در نظر می‌گیریم.

- توزیع گاما با میانگین واحد
- توزیع لگ نرمال با میانگین واحد یعنی $\ln \mu(\mathbf{X})$ توزیع نرمالی با میانگین $\frac{-\sigma_Y^{۲}}{۲}$ و واریانس $\sigma_Y^{۲}$ دارد که از اینجا به بعد با $\ell Nor(\frac{-\sigma_Y^{۲}}{۲}, \sigma_Y)$ نشان داده می‌شود.

تعریف ۱.۱ در هر دو مورد برآورده می‌شود، زیرا داریم، $E(Y) = E(\mu(\mathbf{X})) = E(\pi(\mathbf{X})) = ۱$.
توجه کنید که پاسخ می‌تواند گسسته باشد (مانند تعداد ادعاهای در نظر گرفته شده در این موردپژوهی) زیرا فرض پیوستگی تنها مربوط به $\pi(\mathbf{X})$ و $\mu(\mathbf{X})$ است.
علاوه بر پارامترهای σ و σ_Y حاکم بر تغییر پذیری پیش‌بین و حق بیمه صحیح، ما ساختارهای وابستگی متفاوتی که این دو متغیر تصادفی را به همدیگر ربط می‌دهند و به وسیله مفاصل توصیف می‌شوند نیز در نظر می‌گیریم. ابتدا، فرض می‌کنیم که مفصل مرتبط کننده $\pi(\mathbf{X})$ به $\mu(\mathbf{X})$، $\preceq_{conc}$ یکنوا است (با پارامترش افزایش می‌یابد). به عنوان مثال این مسئله در مورد مفصل‌های فرانک و کلایتون که در این بخش در نظر گرفته شده‌اند، صادق است. از مطالعه Denuit و همکارانش در مرجع [۱۹] یادآوری می‌کنیم که مفصل کلایتون به صورت زیر داده می‌شود،

$$C_{\theta}(u, \nu) = (u^{-\theta} + \nu^{-\theta} - ۱)^{\frac{-۱}{\theta}}, \theta > \circ$$

همچنین مفصل فرانک به صورت زیر داده می‌شود،

$$C_\theta(u,\nu) = -\frac{۱}{\theta}\log\left(۱ + \frac{(e^{(-\theta u)} - ۱)(e^{(-\theta \nu)} - ۱)}{e^{(-\theta)} - ۱}\right), \theta \neq ۰$$

این دو مفصل، وابستگی مثبت را بیان می‌کنند (در مورد فرانک، برای مقادیر مثبت پارامتر) به گونه‌ای که نتایج بدست آمده در بخش‌های پیشین برقرار می‌مانند [۱۹].
پارامتر θ را می‌توان به صورت شدت وابستگی بین $\mu(\mathbf{X})$ و $\pi(\mathbf{X})$ تفسیر کرد. برای اینکه پارامتر وابستگی θ دل‌پذیرتر باشد، از ضریب همبستگی مرتبه کندال متناظر یا تاو کندال استفاده می‌کنیم. برای مفصل کلایتون، تاو کندال برابر با $\frac{\theta}{\theta+۲}$ است، در حالیکه برای مفصل فرانک، تاو کندال یک تابع صعودی از θ است (عبارت شامل تابع دبای).

۲.۸.۲ تغییرپذیری

مشخص شده است که اگر میانگین را ثابت نگهداریم، خانواده توزیع‌های گاما نسبت به واریانس با ترتیب $\preceq_{cx}$ مرتب می‌شود. اکنون سطوح واریانس مختلف را در نظر می‌گیریم و مفصل را برابر با کلایتون فرض می‌کنیم و پارامتر تاو کندال را برابر با ۰/۵ در نظر می‌گیریم. جدول ۱.۲، نتایج ارائه شده در شکل ۳.۲ را جمع‌بندی می‌کند.

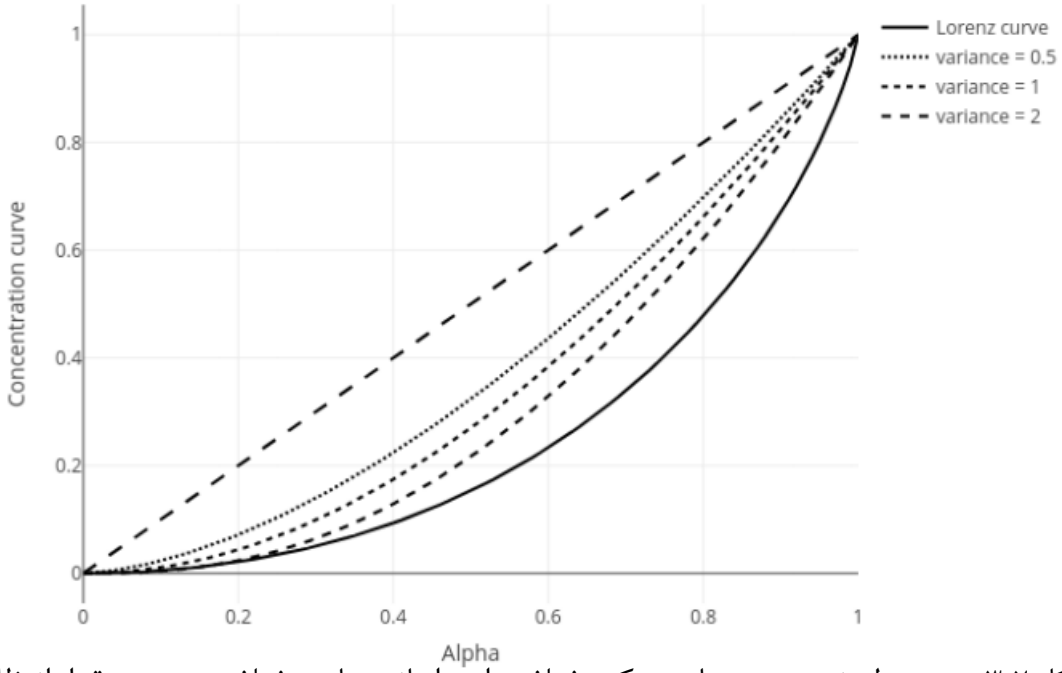
جدول ۱.۲: نتایج بدست آمده از نمودار۸.۲

Line type	$\pi(\mathbf{X})$	$\mu(\mathbf{X})$	Copula C	ABC
medium dash	$Gam(۱,۱)$	$Gam(۱,۲)$	$Clayton(\tau = ۰/۵)$	۶/۳۳٪
short dash	$Gam(۱,۱)$	$Gam(۱,۱)$	$Clayton(\tau = ۰/۵)$	۹/۶۶٪
dotted	$Gam(۱,۱)$	$Gam(۱,۰/۵)$	$Clayton(\tau = ۰/۵)$	۱۳/۰۸٪

می‌بینم که منحنی‌های تمرکز، در نتیجه ترتیب محدب میان توزیع‌های مختلف مقادیر امید ریاضی شرطی، متقاطع نیستند. به علاوه، هرچه واریانس $\mu(\mathbf{X})$ کوچکتر باشد، منحنی تمرکز از منحنی لورنز دورتر است، که این مسئله منجر به کاهش مقدار ABC با واریانس $\mu(\mathbf{X})$ می‌شود. بنابراین، این مثال بیانگر این واقعیت است که، زمانی که پیش‌بین‌هایی با توزیع یکسان داریم که عملکرد مشابه‌ای از نظر وابستگی به حق بیمه صحیح دارند، معیار ABC در مواردی مطلوب است، که حق بیمه صحیح بیشترین تغییر را داشته باشد (از لحاظ ترتیب محدب). همچنین برای یک حق بیمه صحیح داده شده و پیش‌بین‌هایی که از نظر وابستگی به حق بیمه صحیح به یک روش عمل می‌کنند، معیار ABCای مطلوب پیش‌بینی است، که تغییرپذیری کمتری از نظر ترتیب محدب داشته باشد.
موقعیت $\mu(\mathbf{x}) \preceq_{cx} \pi(\mathbf{X})$ می‌تواند به علت بیش‌برازش باشد. این مسئله به عنوان مثال، زمانی رخ می‌دهد که π روی نویز تصادفی انتگرال‌گیری کند. در واقع، فرض کنید که تنها q ویژگی اول

$X_1,\ldots,X_q,(q<p)$ اهمیت دارند و $X_{q+1},\ldots,X_p$ متغیرهای تصادفی مستقل از $X_1,\ldots,X_q$ با میانگین صفر هستند. بنابراین، امتیاز واقعی $\beta_0+\sum_{j=1}^{q}\beta_jX_j$ از نظر تحدب، به صورت $\beta_0+\sum_{j=1}^{p}\beta_jX_j$ است.

در مقابل، موقعیتی داریم که در آن، $\mu(\mathbf{x})\preceq_{cx}\pi(\mathbf{X})$ به علت برازش، کمتر از مقدار واقعی باشد، به عنوان مثال زمانی که داشته باشیم، $\mu(\mathbf{X})=E(Y|X_1,\ldots,X_q,X_{q+1},\ldots,X_p)$ و $\pi(\mathbf{X})=E(Y|X_1,\ldots,X_q)$. در واقع، همانگونه که در بخش ۳.۵.۱ بیان شده است، افزایش تعداد ویژگی‌ها منجر به تولید حق بیمه‌های پراکنده‌تر می‌شود.



شکل ۳.۲: منحنی لورنز و منحنی‌های تمرکز مختلف برای واریانس‌های مختلف در توزیع مقدار انتظاری شرطی.

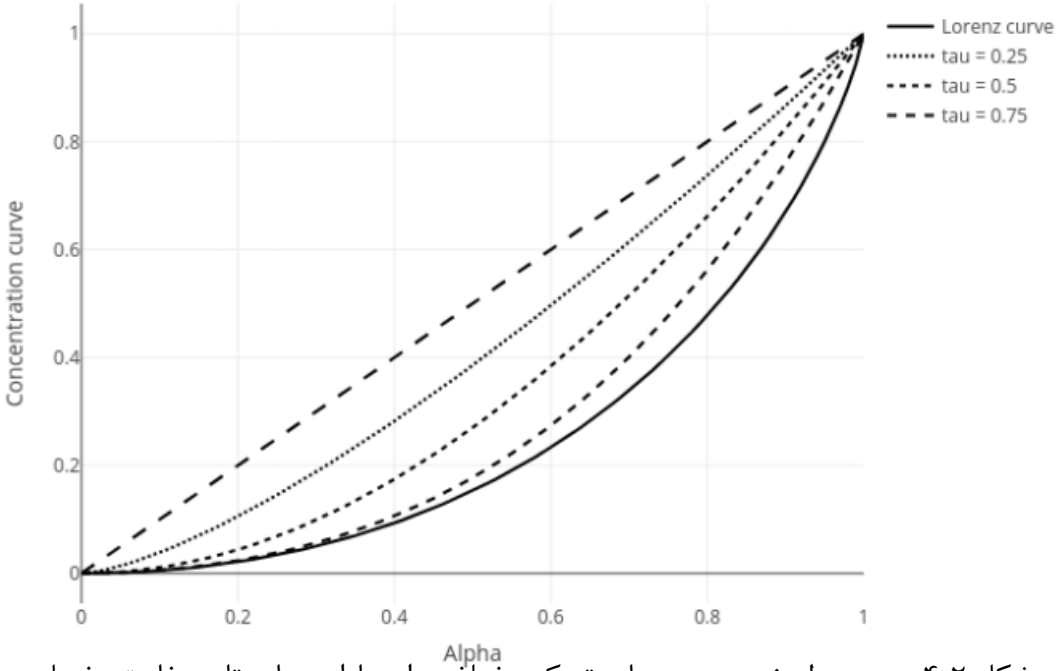
۳.۸.۲ وابستگی

اکنون اگر توزیع شرطی و توزیع پیش‌بین را ثابت نگهداریم، می‌توانیم اثر شدت وابستگی را با در نظر گرفتن مفاصل برای مقادیر پارامتری مختلف وابسته به ضرایب تاو کندال مختلف بدست آوریم. چیدمان شکل ۴.۲ در جدول ۲.۲ جمع‌بندی شده است.

جدول ۲.۲: نتایج بدست آمده از نمودار۴.۲

Line type	$\pi(\mathbf{X})$	$\mu(\mathbf{X})$	C	ABC
medium dash	$Gam(1,1)$	$Gam(1,1)$	$Clayton(\tau=0/75)$	۳/۴۶٪
short dash	$Gam(1,1)$	$Gam(1,1)$	$Clayton(\tau=0/50)$	۹/۶۶٪
dotted	$Gam(1,1)$	$Gam(1,1)$	$Clayton(\tau=0/25)$	۱۷/۰۴٪

مشابه با اثر واریانس می‌توانیم مشاهده کنیم که هر چه وابستگی ضعیف‌تر باشد، منحنی تمرکز از منحنی لورنز دورتر است. این مشاهده را می‌توان به صورت زیر شرح داد. افزایش تاو کندال منجر می‌شود وقتی دو مولفه زوج تصادفی $(\mu(\mathbf{X}), \pi(\mathbf{X}))$ از طریق مفصل کلایتون به همدیگر متصل شوند، این زوج از لحاظ $\preceq_{conc}$ بزرگتر باشد. آنگاه از ویژگی ۱.۶.۲ در می‌یابیم که منحنی تمرکز کاهش می‌یابد و به منحنی لورنز نزدیک‌تر می‌شود. افزایش تاو کندال بدین معنی است که $\pi(\mathbf{X})$ آگاهی‌بخشی بیشتری در رابطه با حق بیمه صحیح $\mu(\mathbf{X})$ خواهد داشت. همچنین زمانی که با اندازه‌گیری تاو کندال مشخص شود پیوستگی بیشتر شده است، مقادیر ABC کاهش می‌یابند. این رابطه در شکل ۴.۲ ترسیم شده است.



شکل ۴.۲: منحنی لورنز و منحنی‌های تمرکز مختلف برای پارامترهای تاو متفاوت مفصل.

۴.۸.۲ توزیع

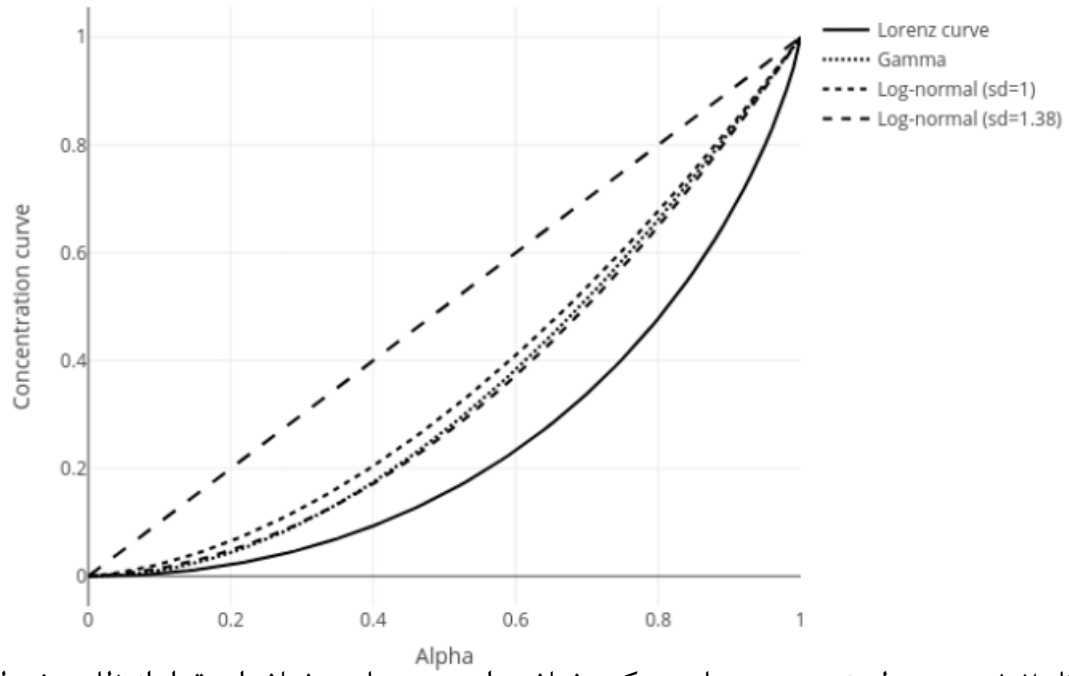
ما موردهایی را دیده‌ایم که در آنها، برای توزیع مرتب یا مفصل، ترتیب منحنی تمرکز حفظ می‌شود. با این وجود، این ترتیب در موارد دلخواه ممکن است حفظ نشود.

در شکل ۵.۲ می‌توانیم موقعیتی را ببینیم که در آن، پیش‌بین از توزیع گامای مرجع تبعیت می‌کند، در حالیکه مقادیر امید ریاضی شرطی از توزیع‌های لگ نرمال با پارامترهای متفاوت پیروی می‌کنند. ما مفصل را نیز به صورت کلایتون در نظر گرفته و تاو کندال را برابر با ۰/۵ قرار داده‌ایم. مقادیر متناظر در جدول ۳.۲ فهرست شده‌اند.

جدول ۳.۲: نتایج بدست آمده از نمودار ۵.۲

Line type	$\pi(\mathbf{X})$	$\mu(\mathbf{X})$	C	ABC
medium dash	$Gam(۱,۱)$	$\ell Nor(-\frac{(۱/۲۵\sqrt{\ln ۲})^۲}{۲}, ۱/۲۵\sqrt{\ln ۲})$	$Clayton(\tau = ۰/۵)$	۹/۲۰٪
short dash	$Gam(۱,۱)$	$\ell Nor(-\frac{\ln ۲}{۲}, \sqrt{\ln ۲})$	$Clayton(\tau = ۰/۵)$	۱۱/۵۸٪
dotted	$Gam(۱,۱)$	$Gam(۱,۱)$	$Clayton(\tau = ۰/۵)$	۹/۶۶٪

وقتی واریانس‌های دو توزیع برابر باشند، (با ۱) آنگاه منحنی تمرکز لگ نرمال (خط‌چین کوتاه) نسبت به گاما (نقطه‌چین) از منحنی لورنز دورتر می‌شود. در این دو مورد، وابستگی به $\pi(\mathbf{X})$ نیز یکسان است و مقدار ABC در مواردی مطلوب است، که توزیع‌های $\pi(\mathbf{X})$ و $\mu(\mathbf{X})$ مشابه باشند. اکنون وقتی σ_Y را با ضریب ۱/۲۵ افزایش می‌دهیم ، منحنی‌های تمرکز، در حدود نقطه ۰/۳۴ همدیگر را قطع می‌کنند. بنابراین، توزیع‌های مختلف منجر به انحناهای متفاوت در منحنی تمرکز می‌شوند و با تنظیم تغییرپذیری یا شدت وابستگی مفصل می‌توانیم این تقاطع را در α دلخواه تولید کنیم. اکنون متوجه می‌شویم که مقدار ABC، کمترین مقدار است که با توجه به بخش ۲.۸.۲ تعجب‌آور نیست زیرا در این حالت، واریانس $\mu(\mathbf{X})$ بزرگترین مقدار را دارد.



شکل ۵.۲: منحنی لورنز و منحنی‌های تمرکز مختلف برای توزیع‌های مختلف از مقدار انتظاری شرطی.

۵.۸.۲ مفصل های متقاطع

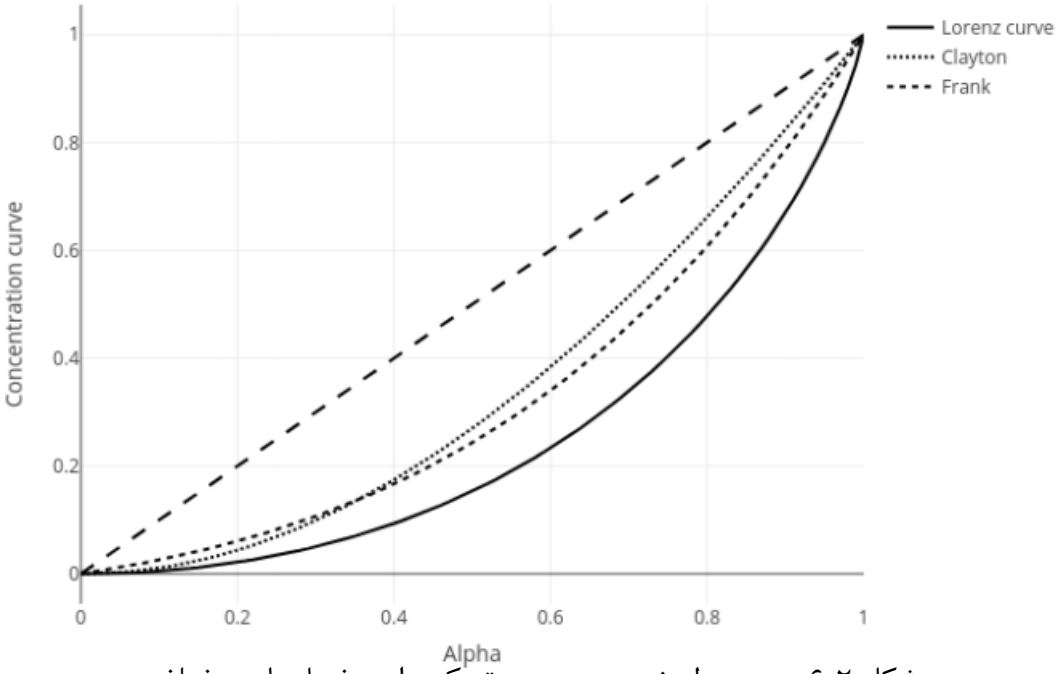
می‌توانیم همانند مثال قبل، از مفاصل مختلف برای بدست آوردن منحنی‌های تمرکز متقاطع استفاده کنیم. اکنون، یک مفصل کلایتون $C_۱$ و یک مفصل فرانک $C_۲$ را همانند مثال ۲ ـ ۳ ارائه شده در مرجع [۲۰] در نظر می‌گیریم. در این حالت، می‌دانیم که یک تابع f وجود دارد

که $C_1(u,\nu) - C_2(u,\nu) \leq 0$ در آن برقرار است، اگر $\nu \leq f(u)$ باشد و $C_1(u,\nu) - C_2(u,\nu) \geq 0$ در آن برقرار است، اگر $\nu \geq f(u)$ باشد. بنابراین این دو مفصل مطابق با ترتیب هماهنگ، مرتب‌سازی نشده‌اند. در شکل ۶.۲ یک توزیع پیش‌بین و مقدار امید ریاضی شرطی اما با مفاصل متفاوت کلایتون و فرانک را می‌بینیم که تاو کندال برابر با ۰/۵ دارند. این چیدمان در جدول زیر جمع‌بندی شده است.

جدول ۴.۲: نتایج بدست آمده از نمودار ۶.۲

Line type	$\pi(\mathbf{X})$	$\mu(\mathbf{X})$	C	**ABC**
short dash	$Gam(1,1)$	$Gam(1,1)$	$Frank(\tau = 0/5)$	۷/۷۹٪
dotted	$Gam(1,1)$	$Gam(1,1)$	$Clayton(\tau = 0/5)$	۹/۶۶٪

منحنی‌ها همدیگر را (حدود نقطه ۰/۳۵) قطع می‌کنند که نتیجه مشهودی است، زیرا مفصل کلایتون به چارک پایین بیش از مفصل فرانک وابستگی دارد و اگر وابستگی کل در هر دو برابر باشد، در ربع بالا، عکس این قضیه برقرار است. هر چه وابستگی قوی‌تر باشد، منحنی تمرکز به منحنی لورنز نزدیک‌تر است. بنابراین مفصل کلایتون در مقادیر کوچک به منحنی لورنز نزدیک‌تر و در مقادیر بزرگ دورتر است.



شکل ۶.۲: منحنی لورنز و دو منحنی تمرکز برای مفصل‌های مختلف.

۶.۸.۲ اثر مفصل با وابستگی غیر رگرسیونی

اکنون مفصلی را در نظر می‌گیریم که وابستگی چارک مثبت نشان نمی‌دهد، بلکه تنها وابستگی مورد انتظار مثبت نشان می‌دهد. در این بخش مطابق با مطالعه Egozcue و همکارانش در

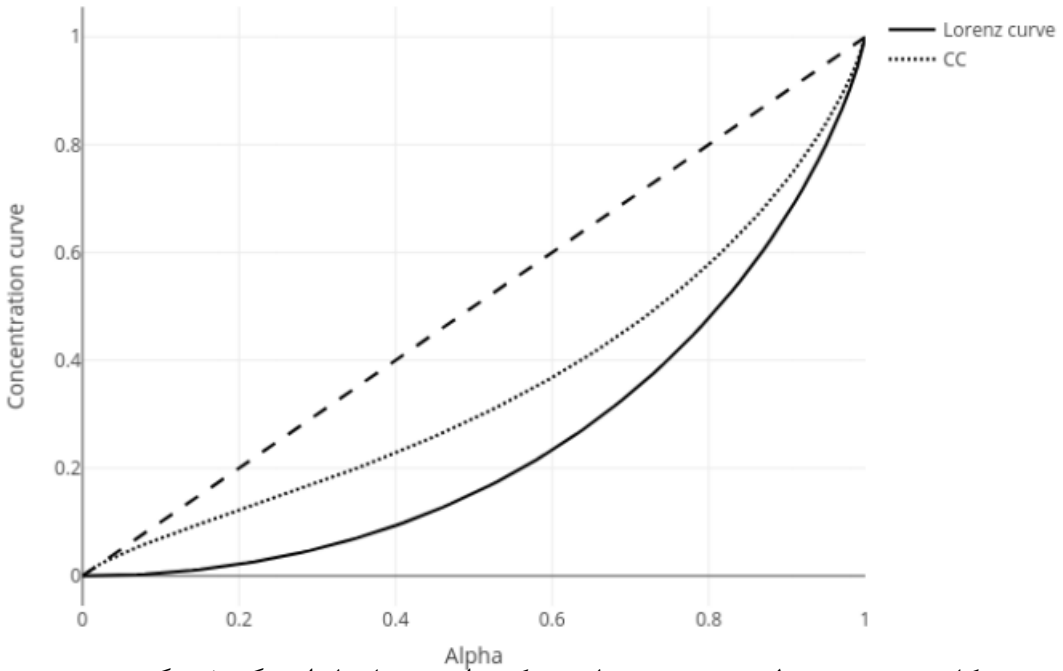
مرجع [۲۸] پیش می‌رویم و دو مفصل را که بیانگر وابستگی چارک با علائم مخالف هستند، مخلوط می‌کنیم. به عنوان مثال، با در نظر گرفتن کران‌های بالا و پایین مفاصل فرشه – هوفدینگ می‌توانیم از ترکیب زیر استفاده کنیم،

$$C(u,\nu) = (۱-\theta)\min\{u,\nu\} + \theta\max\{۰, u+\nu-۱\} \qquad (۱.۲)$$

که مطابق با مثال ۲ – ۱ در مطالعه Denuit and Mesfioui و در مرجع [۲۱] است. از مطالعه Egozcue و همکاران می‌دانیم که این مخلوط، وابستگی مورد انتظار مثبت را بیان می‌کند، اگر و تنها اگر $\theta \leq \frac{۱}{۲}$ باشد. همچنین، مفصل کران پایین فرشه – هوفدینگ را می‌توان با یک مفصل دیگر جایگزین کرد که وابستگی چارک منفی نشان دهد (مانند مفصل فارلی – گامبل – مورگنسترن [۱] یا (FGM) که پارامتر وابستگی منفی دارد، به مرجع [۲۸] مراجعه کنید). در شکل ۷.۲ ما می‌توانیم تایید کنیم، که مفصل ترکیبی بالا یک منحنی تمرکز غیرمحدب ایجاد می‌کند. چیدمان در نظر گرفته شده برای این مطالعه عددی به‌صورت زیر است.

جدول ۵.۲: نتایج بدست آمده از نمودار ۷.۲

Line type	$\pi(\mathbf{X})$	$\mu(\mathbf{X})$	C	**ABC**
dotted	$Gam(۱,۱)$	$Gam(۱,۱)$	$(۵/۱)with\theta = ۰/۸$	۱۰٪



شکل ۷.۲: منحنی لورنز و منحنی‌های تمرکز برای مفصل با وابستگی غیرگرسیونی.

[1] Farlie-Gumbel-Morgenstern

۹.۲ مورد پژوهی

۱.۹.۲ فراوانی

ما در اینجا یک مجموعه داده متناظر با پرتفوی بیمه مسئولیت شخص ثالث موتور در فرانسه را در نظر می‌گیریم که در بسته CASdatasets در R موجود است. به ویژه به مجموعه داده‌های freMTPL2freq توجه داریم که شامل ۶۷۸۰۱۳ مشاهده از ادعاها (پاسخ Y) همراه با نه متغیر توصیفی ($\mathbf{X} = (X_1, \ldots, X_9)$) است. متغیرهای توصیفی، مشخصات متفاوتی را مد نظر قرار می‌دهند،

- بیمه نامه، مواجهه
- بیمه‌گزار، سن، تمرکز ساکنان در خانه، شهر، منطقه، ناحیه بونوس – مالوس
- ماشین، قدرت، سن، برند، نوع سوخت

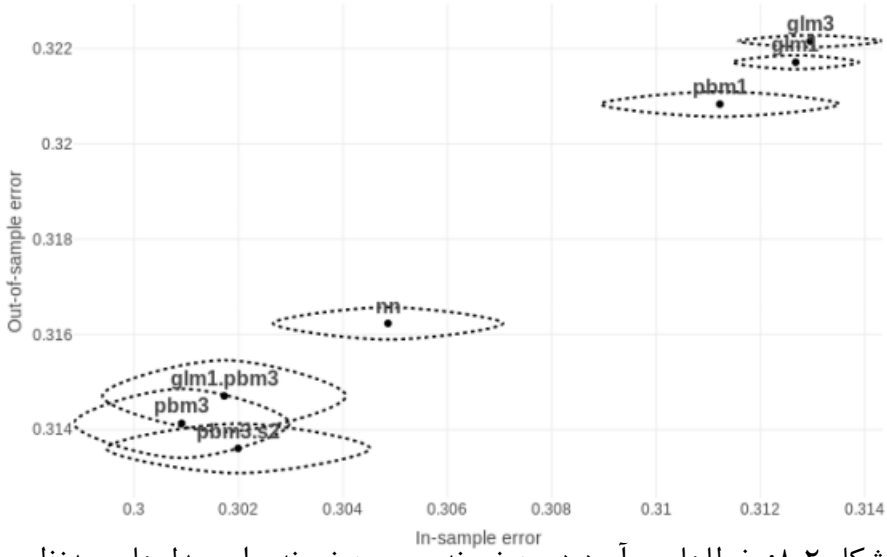
این مجموعه داده ها اخیراً در مطالعه Noll و همکارانش در سال ۲۰۱۸ با تکنیک‌های یادگیری ماشینی و آماری مختلف بررسی شده‌اند .خواننده می‌تواند برای مشاهده توصیف دقیق داده‌ها و جزئیات اجرای مدل به آن مورد پژوهی مراجعه کند [۵۸]. نول و همکارانش (۲۰۱۸) مدل‌های مختلف را بر اساس خطاهای درون نمونه و بیرون نمونه مقایسه کردند. تعریف دقیق خطاهای درون نمونه و بیرون نمونه در مطالعه نول و همکارانش (۲۰۱۸) با روابط (۲ – ۲) و (۲ – ۳) داده شده است [۵۸]. در این بخش، ما برخی از این مدل‌ها (که بر اساس تابع کاهش انحراف پواسون ساخته شده‌اند) را با استفاده از معیار نیکویی برازش [1] که در این پایان نامه معرفی شده است، مقایسه می‌کنیم. به طور ویژه، ما در این بخش، مدل‌های زیر را برای پیش‌بین $\pi_k(X_k)$ بر اساس مطالعه نول و همکارانش (۲۰۱۸) در نظر می‌گیریم،

- glm1 ــ پواسون GLM با تابع پیوند لگاریتمی و همه متغیرهای توصیفی
- glm3 ــ همان glm1 اما بدون متغیر های مساحت و ناحیه
- Pbm1 ــ درخت SBS تقویت شده (قطعه‌های دوتایی استانداردسازی شده) (عمق = ۱، تکرار =۳۰).
- Pbm3 ــ درخت SBS تقویت شده (عمق = ۳، تکرار = ۵۰).
- pbs3.s2 ــ درخت SBS تقویت شده (عمق = ۳، تکرار = ۵۰، انقباض = ۰/۵).
- glm1-pbm3 ــ درخت SBS تقویت شده که از برازش glm1 شروع می‌شود (عمق = ۳، تکرار ۵۰).

[1] goodness of fit

- nn ــ شبکه عصبی کم‌عمق (۲۰ نورون با یک لایه پنهان).

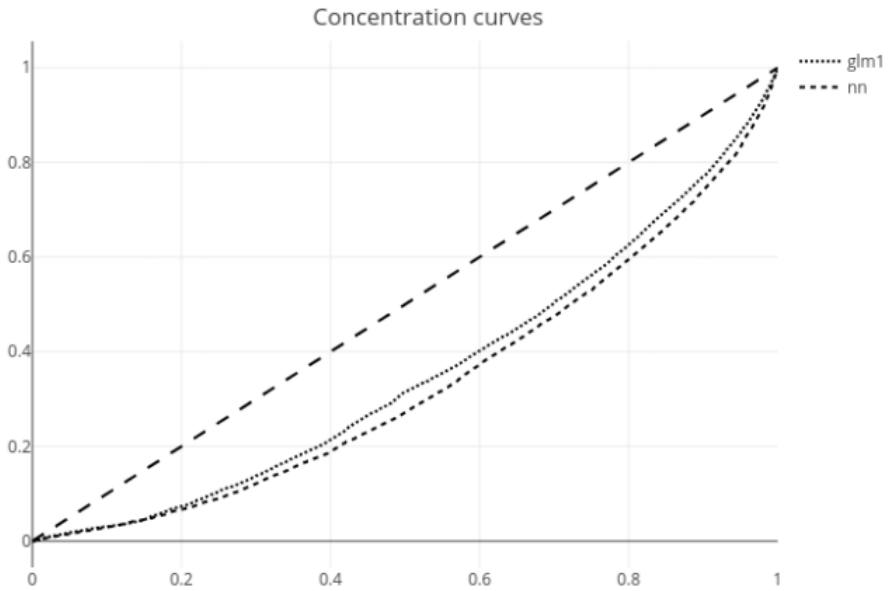
مجموعه داده‌ها به صورت یک نمونه آموزشی با ۶۱۰۰۰۰ مشاهده و یک نمونه آزمون با مشاهدات باقیمانده تقسیم‌بندی شده‌اند. شکل ۸.۲ خطاهای داخل و خارج نمونه برای مدل‌های مورد مطالعه را در بازه‌های اطمینان ۹۵٪ خودگردان‌سازی شده نشان می‌دهد. کران‌ها برای خطاهای درون و بیرون نمونه به صورت مجزا بدست آمده‌اند، بنابراین تنها فواصل عمودی و افقی معنادار هستند. به طور خاص، شکل بیضوی از طریق هموارسازی اسپلاینی نقاط (خطای درون نمونه، خطای بیرون نمونه) = {(کمتر، مشاهده شده)، (مشاهده شده، بیشتر)، (بیشتر، مشاهده شده)، (مشاهده شده، کمتر)} بدست آمده است. در کل، خطای درون نمونه و خطای بیرون نمونه مدل‌ها را به استثنای مدل درخت تقویت شده (pbm3) و نسخه کوتاه شده آن، به یک روش طبقه‌بندی می‌کنند. در مدل‌های آخر، وارد کردن یک ضریب انقباض، خطای درون نمونه را افزایش می‌دهد در حالیکه خطای بیرون نمونه را کاهش می‌دهد. این مسئله تعجب‌آور نیست زیرا هدف از وارد کردن ضریب انقباض، عدم مواجهه با بیش‌برازش است. همچنین لازم به ذکر است که مدل GLM تقویت شده (glm1.pbm3) نسبت به مدل GLM اصلی (glm1) بسیار ارتقا یافته است. با این وجود، بهتر از درخت SBS تقویت شده (pbm3) عمل نمی‌کند. این مشاهده آخر نشان می‌دهد که شکل ساختاری ثابت که مدل GLM بر فراوانی ادعای موردانتظار تحمیل می‌کند، هیچ بینش توصیفی مکملی در مقایسه با درخت SBS تقویت‌شده ایجاد نمی‌کند. در نهایت، مدل بهینه نسبت به معیار خطای خارج از نمونه، مدل درخت تقویت شده با یک ضریب انقباض است (pbm3.s2).



شکل ۸.۲: خطاهای برآورد درون نمونه و بیرون نمونه برای مدل‌های مدنظر.

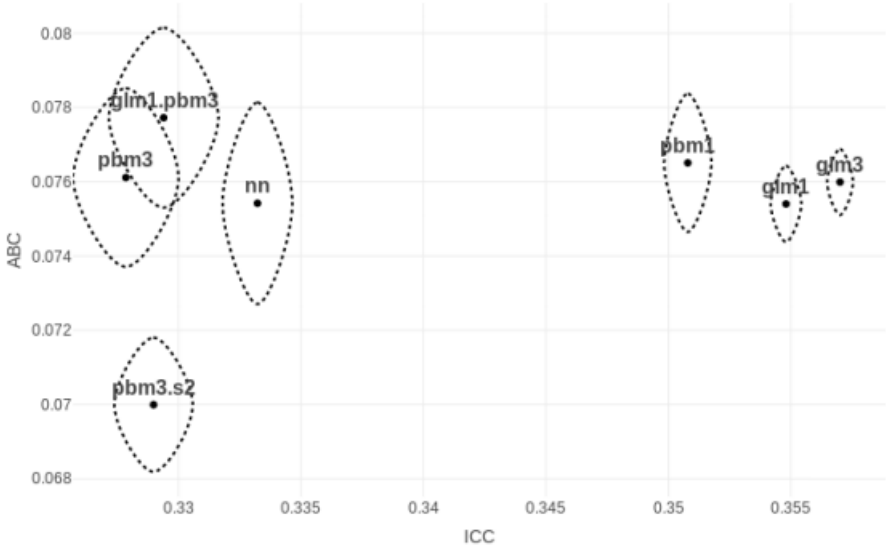
همه مدل‌ها به استثنای مدل مبتنی بر (pbm3) را می‌توان با توجه به بازه‌های اطمینان خودگردان‌سازی شده، به خوبی جداسازی کرد. به نظر می‌رسد که تغییرپذیری نتایج در مدل‌های تقویت شده نسبت به GLM ها یا مدل شبکه عصبی (برای خطای خارج از نمونه)

بیشتر باشد. اکنون به معیار نیکویی برازش که در این پایان نامه معرفی شد بازمی‌گردیم. در ادامه، از نسخه‌های تجربی منحنی تمرکز $\widehat{CC}$ و منحنی لورنز $\widehat{LC}$ که در نمونه آزمون محاسبه شده‌اند، برای بدست آورد مقادیر ABC و ICC استفاده می‌کنیم. در شرایطی که تعداد مشاهدات کافی نباشد، می‌توان به جای آن از یک نسخه هموار شده از منحنی تمرکز تجربی $\widehat{CC}$ استفاده کرد. در اینجا، اندازه نمونه آزمون کافی بر اساس $\widehat{CC}$ سنجیده می‌شود، که در شکل ۹.۲ برای دو مدل مدنظر نشان داده شده است. مدل‌های باقیمانده به یکی از این دو مدل نزدیک هستند و مشخصاً دو گروه منحنی را شکل می‌دهند. باید توجه داشت که این دو گروه برای α بالاتر از ۰/۱۵ ایجاد می‌شود. گروه منحنی‌های بالاتر مربوط به مدل‌های glm1، glm3 و pbm1 هستند که بدترین مدل‌ها از نظر خطاهای خارج از نمونه می‌باشند. مقادیر ABC و ICC در شکل ۱۰.۲ نمایش داده شده‌اند و بازه‌های اطمینان خودگردان‌سازی شده نیز مشخص شده‌اند. باید توجه داشت که معیار ICC، مدل‌ها را به غیر از مدل‌های pbm3 و pbm3.s2 به شکل معیارهای خطای خارج از نمونه در نظر می‌گیرد. در حالیکه هر دو معیار متفقاً بیان می‌دارند که این دو مدل آخر، بهترین مدل‌ها هستند اما مدل pbm3.s2 از نظر خطای خارج از نمونه بهتر از pbm3 عمل می‌کند در حالیکه در مورد ICC قضیه بالعکس است. در مورد مقادیر ABC می‌توان مشاهده کرد که مدلی با ICC پایین، یا ABC بزرگ دارد یا ABC کوچک. به عنوان مثال، glm1.pbm3 که یکی از بهترین مدل‌ها از نظر ICC است، بالاترین ABC را دارد در حالیکه pbm3.s2 هم ICC و هم ABC پایین دارد. اگر glm1.pbm3 و pbm3.s2 را مقایسه کنیم که طبق ICC درجه‌های ترفیع برابر دارند، می‌توان دریافت که معیار ABC مطلوب pbm3.s2 است، زیرا تغییرپذیری کمتری نسبت به glm1.pbm3 دارد که این نتیجه همراستا با نتیجه بخش ۲.۸.۲ نیز می‌باشد. به همین صورت، در حالیکه pbm3 و pbm3.s2، ICC مشابه دارند، pbm3.s2 از نظر ABC بهتر از pbm3 عمل می‌کند و تغییرپذیری کمتری نسبت به آن دارد (زیرا هر دو مدل تعداد درخت‌های یکسانی دارند در حالیکه pbm3.s2 یک پارامتر انقباض دارد). در نهایت، مدل بهینه از نظر ABC، pbm3.s2 است.



شکل ۹.۲: $\widehat{CC}$ برای مدل‌های مورد مطالعه.

در انتهای این مثال، در شکل ۱۱.۲، ICC و ABC را به صورت توابعی از α نشان دادیم (یعنی انتگرال‌گیری روی بازه $[0, \alpha]$ به جای بازه کلی $[0, 1]$). ما تنها منحنی‌های این دو مدل را ارائه می‌کنیم، زیرا منحنی‌های دیگر، بسیار شبیه هستند و نمی‌خواهیم که تصویر مبهم شود. به منظور متمایزسازی دقیق‌تر منحنی‌ها، باید روی نواحی خاص (α) متمرکز شویم. با این وجود، حتی در اینجا هم اگر به pbm1 و glm1.pbm3 نگاه کنیم می‌بینیم که مورد آخر، ICC کمتری دارد در حالیکه مقادیر ABC حدود ۹۱ درصد چندک را قطع می‌کنند. بنابراین از این آستانه، pbm1 از نظر معیار ABC بهتر از glm1.pbm3 عمل می‌کند.



شکل ۱۰.۲: مقادیر ABC و ICC برآورد شده برای مدل‌های مدنظر.

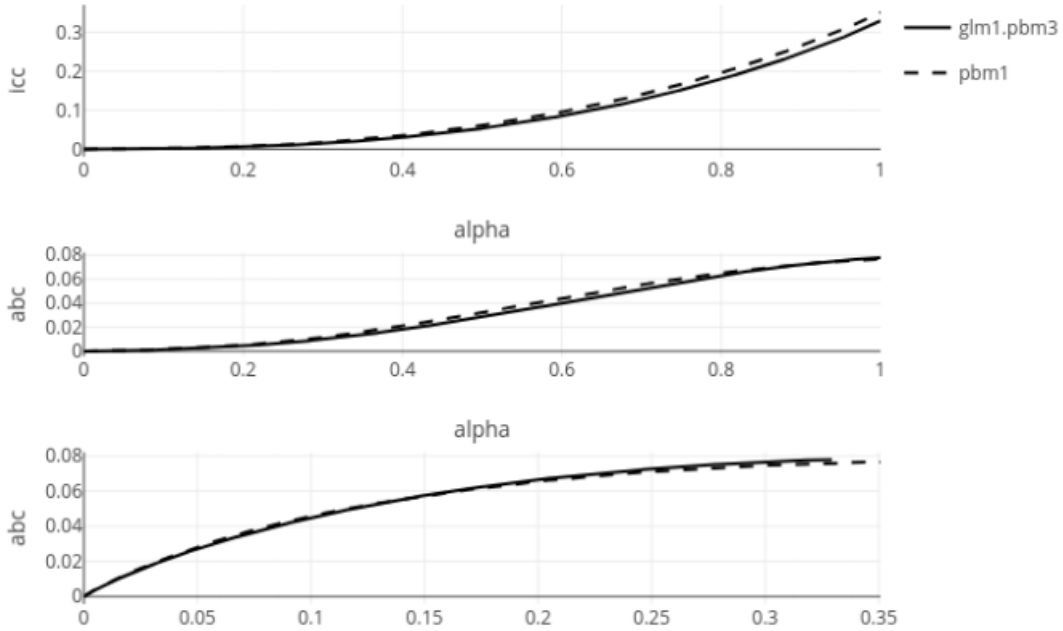
۲.۹.۲ فراوانی و شدت

در بخش پیشین، فراوانی ادعا را مورد بحث قرار دادیم. در این بخش، می‌خواهیم تحلیل شدت را انجام دهیم و از آن به عنوان یک نمونه استفاده کنیم، تا نشان دهیم که مفاهیم وارد شده در مدل‌های ترکیبی چه عملکردی دارند. به طور خاص، ما از freMTPL2sev بهره می‌گیریم که حاوی ۲۶۶۳۹ ادعای تجمیع شده است، که متناظر با فراوانی می‌باشد که پیشتر توصیف کردیم. ما از مدل glm3 به عنوان مدل فراوانی بهره می‌گیریم و همان مجموعه متغیرهای کمکی را با توزیع گاما و با اتصال لگاریتمی برای مدل شدت در نظر می‌گیریم. بدین صورت یک مدل ترکیبی (فراوانی – شدت) برای حق بیمه خالص بدست می‌آوریم. همچنین، یک مدل GLM توئیدی را برای حق بیمه خالص در نظر می‌گیریم. ما از توان ۱/۶ (که با جستجوی شبکه‌ای ساده بدست آمده است) برای در نظر گرفتن جرم ذاتی در صفر استفاده می‌کنیم.
در شکل ۱۲.۲، منحنی‌های تمرکز را برای هر دو مدل نشان داده‌ایم. می‌توان مشاهده کرد که دو منحنی در دو سوم بازه خود متفاوت هستند. به طور خاص، مدل توئیدی ادعاهای بالاتر را به بیمه نامه‌های کم ریسک‌تر از مدل ترکیبی نسبت می‌دهد در حالیکه ادعاهای پایین را به پروفایل‌های ریسک میانگین مرتبط می‌داند. این روند می‌تواند به تنظیم‌کنندگان قیمت کمک کند تا با مدل‌های حق بیمه خالص با هدف قطعه‌بندی یا استراتژی بازاریابی انطباق یابند.
معیارهای متناظر برای دو مدل به صورت زیر هستند.

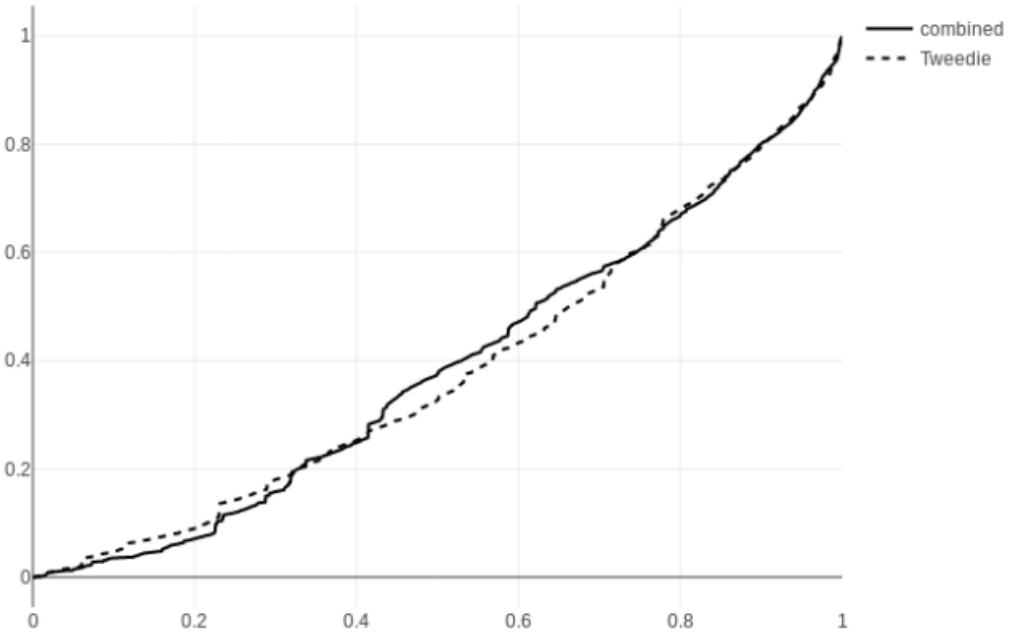
جدول ۶.۲: نتایج بدست آمده از نمودار۱۱.۲

Line type	model	ICC	ABC
Solid	Combined	۰/۳۸۹۶	۰/۱۳۳۶
dashed	Tweedie	۰/۳۸۱۶	۰/۰۴۴۶

اکنون باید توجه داشت که ما در میان مشاهدات مجموعه داده، که مربوط به مجموعه آزمون بوده‌اند (که در واقع برای بدست آوردن منحنی‌ها و مقادیر به کار گرفته شده‌اند)، بالاترین ادعا را حذف کرده‌ایم. بالاترین مشاهده، جهش بزرگی (تقریباً ۵۰ درصد) در منحنی تمرکز و در نقاطی ایجاد می‌کند که به صورت پروفایل‌های ریسک برآورد شده متناظر برچسب خورده‌اند.



شکل ۱۱.۲: مقادیر ABC و ICC برآورد شده برای مدل‌های مدنظر.



شکل ۱۲.۲: منحنی‌های تمرکز برای مدل‌های شدت فراوانی و توئیدی(Tweedie).

۱۰.۲ نتیجه گیری

مطالعات Brazauskas و همکارانش در سال ۲۰۰۷ و Samanthi و همکارانش ، انگیزه‌ای برای مقایسه شاخص‌های جینی ریسک‌های بیضوی چندمتغیره با حاشیه‌های مشترک اما ماتریس‌های پراکندگی متفاوت بود. در این پایان‌نامه، ابتدا به صورت کلی، ترتیب محدب صعودی بین شاخص‌های جینی ریسک‌های بیضوی چندمتغیره ایجاد شد. آنگاه، زمانی که ماتریس های

پراکندگی، از ساختارهای خاص تبعیت می‌کردند، شاخص‌های جینی با ترتیب تصادفی معمولی مرتب شدند. به علاوه، نشان داده شد که در میان همه ساختارهای وابستگی، هم‌یکنوایی، کوچکترین شاخص جینی را از دیدگاه ترتیب تصادفی معمولی تولید می‌کند. مقایسه شاخص‌های جینی، غیر از اینکه در کاربردهای آماری سودمند است، به خودی خود جالب توجه است. اول اینکه، به غنای مطالعاتی که درباره تمرکز بردارهای تصادفی بیضوی روی نواحی متقارن مرکزی محدب انجام می‌شوند، کمک می‌کند. دوم اینکه، مفهوم ترتیب $pd-1$ را تعمیم می‌دهد و منجر به کاربردهای بیشتر در تحقیقات عملی می‌شود.

از سوی دیگر، این پایان‌نامه، مسائل بازی را مطرح می‌کند. به عنوان مثال، شاخص‌های جینی ریسک‌های بیضوی چندمتغیره را از دیدگاه ترتیب تصادفی معمولی، تا چه اندازه می‌توان مرتب کرد؟ آیا نتیجه برای ریسک‌های با ابعاد بالا و توزیع بیضوی عمومی همچنان برقرار است؟

در این پایان‌نامه معیارهای جدیدی برای ارزیابی عملکرد یک مدل پیش‌بین ارائه شد. یعنی ABC و ICC مبتنی بر گراف‌های منحنی‌های تمرکز و لورنز و نسخه‌های انتگرال‌گیری شده آنها. این شاخص‌ها، درجه جابه‌جایی ناشی از پیش‌بین مدنظر را به دقت کمیت‌سنجی می‌کنند و یک تفسیر شهودی ارائه می‌کنند که در کاربردهای بیمه سودمند است. سودمندی این معیارها از نظر اقدامات بیمه‌ای به کمک مثال‌های شبیه‌سازی شده و همچنین یک موردپژوهی درباره تعداد ادعاهایی که در مجموعه داده بیمه موتور فرانسه مشاهده شده است، اثبات گردید. معیارهای مربوط به زوج‌های $(Y,\pi(\mathbf{X}))$ در ارزیابی عملکرد یک مدل قیمت‌گذاری معین بسیار سودمند هستند. در بسیاری از ضوابط، در کنار شاخص‌های جینی از احتمالات هماهنگی مانند تاو کندال استفاده می‌شود. خوانندگان علاقه‌مند می‌توانند برای مشاهده پاسخ‌های دودویی به مطالعه Denuit و همکاران در مرجع [۲۲] مراجعه کنند. با این وجود، در این رویکرد، نمی‌توانیم ارتباط $(\mu(\mathbf{X}),\pi(\mathbf{X}))$ را ارزیابی کنیم و Y تنها یک نسخه پرهیاهو از $\mu(\mathbf{X})$ است.

البته، در موقعیت‌های محتمل، هیچ عملکرد بهتری نسبت به رقبای خود ندارد. بنابراین، نشانگرهای ABC و ICC، جعبه‌ابزار آماری را تقویت کرده و ضوابط جدیدی در اختیار تحلیل‌گر قرار می‌دهند تا بتواند عملکرد یک پیش‌بین معین را ارزیابی کند.

فصل ۳

اعتبار سنجی تحقیق

۱.۳ مقدمه

در اقتصاد، بازارهای مالی نقش اساسی در انتقال منابع مالی از بخش غیرمولد به بخش تولید دارند. لذا این بازارها در ایجاد اشتغال، رشد اقتصادی، ثبات متغیرهای پولی و در نهایت رفاه جامعه می‌توانند موثر باشند. تامین مالی شرکت‌ها، تخصیص بهینه منابع، افزایش نقدشوندگی دارایی‌ها، افزایش شفافیت در اقتصاد از مهمترین کارکردهای بازارهای مالی است. از بین کارکردهای ذکر شده تامین مالی شرکت‌ها بیشترین اثر را بر رشد اقتصادی دارد. یکی از بزرگترین اهداف شرکت‌ها و بنگاه‌های مالی در کشورهای توسعه یافته و در حال توسعه، تامین مالی است.

در بازار سرمایه عوامل تأثیرگذار سیاسی و اقتصادی بسیار زیادی وجود دارند که بر روی وضعیت این بازار نقش تعیین کننده‌ای دارند و تصمیمات اقتصادی فعالان بازار سرمایه را تحت الشعاع قرار می‌دهند. این عوامل تا اندازه‌ای برای سهامداران مهم تلقی می‌شود، که عدم توجه برخی از سرمایه‌گذاران به این عوامل باعث متضرر شدن آن‌ها می‌شود. از سرمایه‌گذاران خرد تا شرکت‌های سرمایه‌گذاری دوست دارند بدانند که روند آتی بازار به چه شکل خواهد بود. آیا بازار رونق می‌گیرد یا خیر؟ شناسایی عوامل تعیین‌کننده شاخص کل بورس و پیش‌بینی دقیق روند آتی شاخص، کمک شایانی بر اتخاذ تصمیمات مالی فعالان بازار خواهد داشت.

در این فصل به مدل‌سازی و پیش‌بینی شاخص کل بورس اوراق بهادار تهران به روش GLM

و همچنین سنجش اثر عوامل تعیین کننده شاخص کل بورس به عنوان بازار تامین کننده منابع مالی برای بنگاه‌های تولیدی می‌پردازیم.

۲.۳ تحلیل وضعیت شاخص بورس طی دوره تحقیق

۱.۲.۳ دوره تحقیق و داده‌های مربوطه

تمامی متغیرها به‌صورت روزانه از فاصله زمانی ۲۶ـ۱۲ـ۱۳۹۴ تا ۰۱ـ۰۷ـ۱۳۹۹ می‌باشد که شامل ۱۶۵۴ مشاهده است. به منظور پیش‌بینی و انتخاب مدل این داده‌ها به دو دسته تقسیم شده‌اند، دسته اول که شامل ۹۰ درصد از مشاهدات است، در جهت ساخت مدل پیش‌بینی و دسته دوم که ۱۰ درصد از مشاهدات را شامل می‌شود جهت ارزیابی مدل پیش‌بینی است. تمامی داده‌های این دو گروه به صورت تصادفی انتخاب شده است.
داده‌های مورد استفاده در این فصل از سایت بانک مرکزی (www.cbi.ir) و شبکه اطلاع رسانی طلا و ارز (www.tgju.ir) استخراج شده است.

شاخص کل بورس اوراق بهادار تهران(TSE)

شاخص کل بورس نشان‌دهنده میانگین بازدهی سرمایه‌گذاران در بورس اوراق بهادار است، که با نام TEPIX شناخته می‌شود. نحوه محاسبه شاخص کل به صورت زیر است.

$$\text{شاخص کل} = \frac{\text{ارزش جاری بازار سهام در زمان محاسبه}}{\text{ارزش جاری بازار سهام در تاریخ مبدا}} \times ۱۰۰$$

قیمت اونس طلای جهانی (Gold)

انس یک واحد وزن برای طلا می‌باشد که در جهان متداول است، مانند مثقال که آن هم یک واحد وزن است و در ایران متداول می‌باشد. هر انس طلا حدودا ۳۱ گرم طلا با خلوص بسیار بالا یعنی ۹.۹۹ یا همان عیار ۲۴ می‌باشد. قیمت ارزی طلا و سکه در ایران از قیمت اونس طلای جهانی تاثیر می‌پذیرد.

قیمت سکه (Coin)

خرید و فروش سکه به صورت عمده و خرد، در ایران انجام می‌شود. با توجه به کاربردهای سکه، مصرف و داد و ستد سکه طلای بهار آزادی، از تواتر، تعدد و تکرارپذیری بالایی برخوردار است، به نحوی که حتی در مواقع اوج مصرف و داد و ستد (اعیاد رسمی و مذهبی) حجم نقدینگی در گردش این کالا، از برخی از کالاهای اساسی نیز بالاتر است. بنابراین قیمت سکه را به عنوان یک عامل تاثیرگذار در اقتصاد و همچنین شاخص بورس مورد بررسی قرار می‌دهیم.

قیمت دلار (Dollar)

دلار سالهای زیادی است که به عنوان ارز شناخته شده در بازارهای جهانی و اقتصاد بین‌الملل در حال معامله و داد و ستد است. از این جهت حدود ۴۰ سال پیش شاخصی به نام ایندکس دلار در آمریکا تعیین شد و ارزش دلار را در مقابل ۶ ارز جهانی و معتبر دیگر می‌سنجید. ایندکس دلار همچنان در بورس آمریکا در حال معامله است و تغییرات دلار را نسبت به سبد ارزی ذکر شده می‌سنجد.

قیمت یورو (Euro)

یکای پول کشورهای منطقه یورو است که به ۱۰۰ سنت تقسیم می‌شود. استفاده از یورو در اول ژانویه سال ۲۰۰۲ رسماً در کشورهای آلمان، فرانسه، بلژیک، اتریش، اسپانیا، پرتغال، یونان، ایتالیا، ایرلند، لوکزامبورگ، هلند و فنلاند آغاز شد و اکنون در ۱۹ کشور از ۲۷ عضو اتحادیه اروپا به‌طور رسمی در جریان است. از جمله عوامل دیگر تاثیرگذار در شاخص کل بورس اوراق بهادار می‌توان به تغیرات قیمت یورو اشاره کرد.

قیمت نفت اوپک (Oil)

۱۲ سپتامبر ۱۹۶۰ کشورهای صادرکنندهٔ نفت با هدف محافظت از منافع خود اقدام به تأسیس سازمان واحدی کردند که به اختصار «اوپک» نامیده شد. سازمان صادر کننده نفت خام با نام اختصاری اوپک یک کارتل بین‌المللی نفتی است که متشکل از کشورهای الجزایر، ایران، عراق، کویت، لیبی، نیجریه، قطر، عربستان سعودی، امارات متحده عربی، اکوادور، آنگولا، ونزوئلا و کنگو است. مقر بین‌المللی اوپک از بدو تأسیس در سال ۱۳۳۹ در ژنو بود و در سال ۱۳۴۴ به شهر وین در کشور اتریش انتقال یافت. قیمت نفت اوپک را به عنوان یک عامل تاثیرگذار بر شاخص در این بازه زمانی مورد بررسی قرار می‌دهیم.

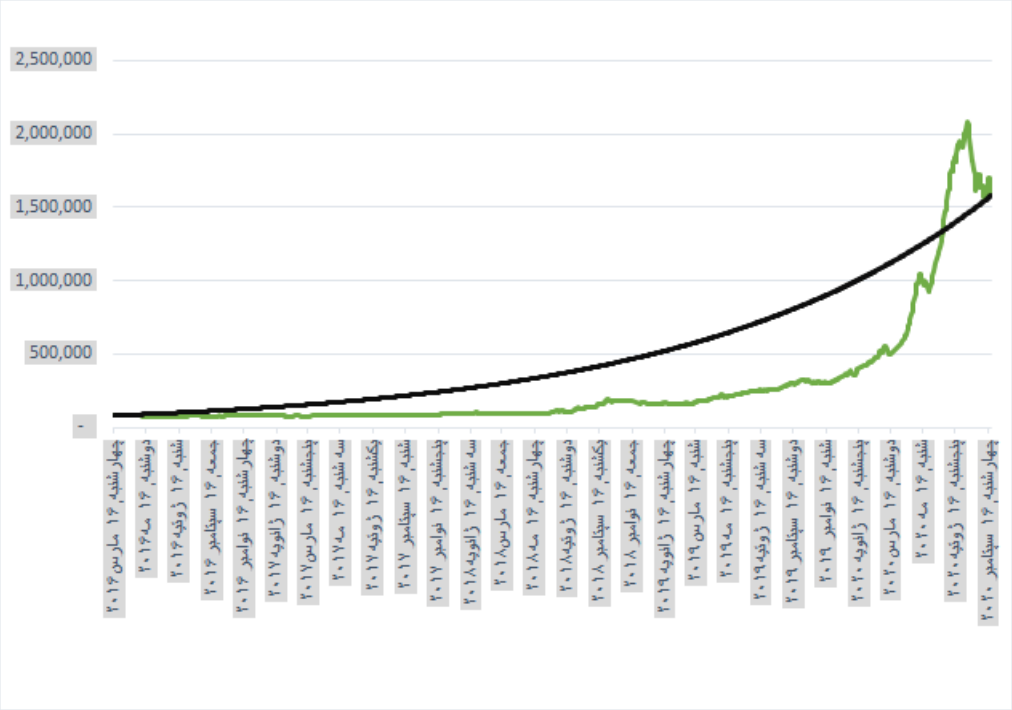
سطح عمومی قیمت‌ها(CPI)

شاخص‌های کالاها و خدمات مصرفی (CPI) از انواع شاخص‌های قیمت است، که نشان دهنده تغییرات قیمت کالاها و خدماتی است، که توسط خانوارها دریك دوره زمانی مورد مصرف قرار می‌گیرد. قیمت‌های مورد استفاده در این شاخص، قیمت‌های خرده فروشی کالاها و خدمات می باشد.

۲.۲.۳ قیمت‌گذاری بیش از حد و کمتر از حد سهام

آنچه باعث تعیین قیمت و ارزش واقعی یک دارایی می‌گردد، میزان عرضه و تقاضایی است، که توسط افراد و شرکت ها برای آن دارایی با توجه به اطلاعاتی که در اختیار دارند صورت

می‌پذیرد. فرض کنید r^* نرخ تنزیل ارزش زمانی پول باشد اگر P_0 ابتدای دوره در بورس سرمایه‌گذاری شود انتظار داریم که پس از گذشت T دوره حداقل ارزش سرمایه‌گذاری به سطح $P_T = P_0(1+r^*)^T$ برسد. در صورتی که ارزش سرمایه کمتر از سطح مذکور باشد آنگاه می‌گوییم که سرمایه در بورس کمتر از حد ارزش‌گذاری شده است و اگر ارزش سرمایه بیشتر از سطح مذکور باشد می‌گوییم دارایی و سرمایه بیشتر از حد ارزش‌گذاری شده است یه عبارتی زمانی ارزش‌گذاری دارایی ما بیش از حد (کمتر از حد) است که، ارزش بازار آن بالاتر (پایین‌تر) از ارزش واقعی آن باشد. به عنوان مثال اگر ارزش سهام بیش از ارزش محاسبه شده بر اساس رابطه $P_T = P_0(1+r^*)^T$ باشد، آنگاه سهام ارزش‌گذاری بیش از حد شده است. اگر ارزش سهام کمتر از ارزش محاسبه شده بر اساس این رابطه باشد، آنگاه سهام ارزش‌گذاری کمتر از حد شده است.



شکل ۱.۳: مقایسه TSE* و TSE

۳.۲.۳ مقایسه بازده موثر بورس با سایر بازارهای موازی در دوره تحقیق

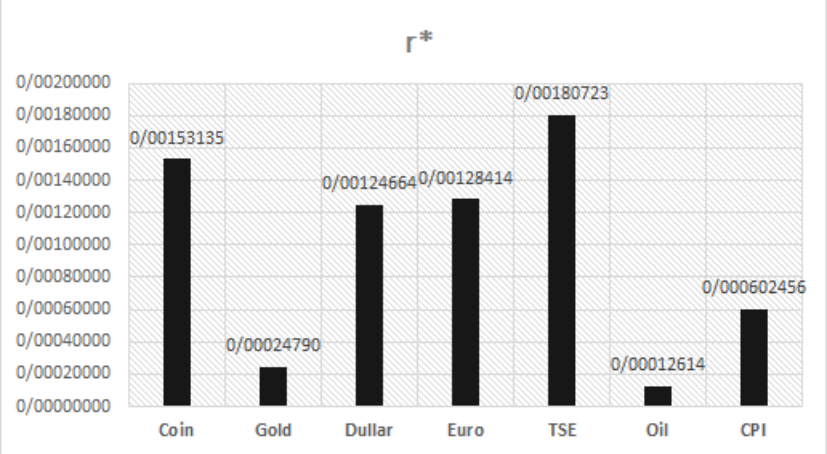
همانطور که در جدول ۱.۳ مشخص است، شاخص کل بورس اوراق بهادار تهران در بازه زمانی مورد نظر بیش از ۱۹ برابر رشد داشته است. بعد از شاخص کل بیشترین بازده متعلق به سکه طلا و کمترین میزان رشد نیز مربوط به قیمت نفت اوپک بوده است.

می‌دانیم r* به معنی میانگین بازده روزانه سرمایه‌گذاری در دارایی مربوطه است. با توجه به جدول ۱.۳ بیشترین نرخ بهره در این بازه بعد از شاخص کل بورس (TSE) متعلق به Coin،

Euro و Dollar است. همچنین کمترین نرخ بازده روزانه نیز متعلق به Oil می‌باشد. در سطر سوم از جدول ۱.۳ OPD [1] به معنای تعداد روزهایی است که قیمت هر کدام از متغیرها بیشتر از ارزش واقعی خود بوده اند، که در این بین قیمت نفت اوپک (Oil) ، ۸۴۵ روز بیشتر از ارزش واقعی خود بوده و همچنین قیمت سکه (Coin) تنها ۲۰ روز بیشتر از ارزش واقعی خود معامله شده است. در جدول ۱.۳، UPD [2] بیانگر تعداد روزهایی است که هر کدام از متغیرها کمتر از ارزش واقعی خود معامله شده‌اند.

جدول ۱.۳: جدول مشخصات متغیرها

	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
r*	۰/۰۰۱۵۳	۰/۰۰۰۲۴	۰/۰۰۱۲۴	۰/۰۰۱۲۸	۰/۰۰۱۸۰	۰/۰۰۰۱۲	۰/۰۰۰۶۰
Kt/K۰	۱۲/۵۰۷۵	۱/۵۰۵۷	۷/۸۲۱۸	۸/۳۲۰۸	۱۹/۷۰۸۴	۱/۲۳۱۵	۲/۶۱۳۵
OPD	۲۰	۱۱۴	۲۳	۸۶	۶۳	۸۴۵	۲۴
UPD	۱۱۴۹	۱۱۲۹	۱۲۲۰	۱۲۰۰	۹۷۴	۸۰	۱۲۶۴
NA	۴۸۱	۴۰۷	۴۰۷	۳۶۴	۶۱۳	۷۲۴	۳۶۱



شکل ۲.۳: میانگین بازده روزانه داده‌ها.

۳.۳ شناسایی عوامل موثر بر شاخص بورس اوراق بهادار تهران

عوامل موثر بر شاخص بورس اوراق بهادار متعدد است. احساسات سرمایه‌گذاران، بازدهی بازارهای جایگزین، چشم انداز آتی اقتصاد کشور، نرخ تورم، قیمت نفت، نرخ ارز، شرایط

[1] Over Pricing Day
[2] Upper Pricing Day

سیاسی و اقتصادی کشور و حتی بلایای طبیعی و دست‌ساز بر بارده بازار سهام و شاخص سهام موثر است در ادامه عوامل موثر بر شاخص بورس بر اساس مطالعات تجربی بیان می‌شود.

۱.۳.۳ عوامل موثر بر شاخص بورس در مطالعات تجربی انجام یافته

تقوی، محمدی و برزنده در سال ۱۳۷۶ متغیرهای اقتصادی که بر شاخص قیمت سهام بورس اوراق بهادار تهران اثرگذار بود را مورد بررسی قرار دادند. آنها اثرگذاری قیمت ارز، مسکن و خودرو را به عنوان عوامل مهم و تاثیر گذار بر شاخص قیمت اوراق بهادار تهران با استفاده از مدل خودرگرسیونی مورد بررسی قرار دادند، که نتیجه آن حکایت از وجود یک رابطه معنادار بین نرخ ارز و قیمت وسایل نقلیه با شاخص قیمت سهام دارد [۴].

نجارزاده در سال ۱۳۸۸ به بررسی اثر نرخ ارز و تورم بر شاخص قیمت سهام با استفاده از مدل خودرگرسیونی، تجزیه واریانس وتوابع واکنش تکانه می‌پردازد. دوره زمانی این مطالعه مربوط به فروردین ۱۳۸۲ تااسفند ۱۳۸۵ است. نتایج این بررسی نشان می‌دهد در میان‌مدت و کوتاه‌مدت نوسانات قیمت ارز باعث افزایش قیمت سهام و در بلندمدت باعث کاهش قیمت در بورس اوراق بهادار می‌شود [۸].

علیرضا بادکوبه‌ای در سال ۱۳۷۴ اثر تورم بر قیمت سهام شرکت های پذیرفته شده در بورس را مورد مطالعه و بررسی قرار داد، که بر اساس این مطالعه نرخ تورم و قیمت سهام یک رابطه مستقیم با هم دارند، به‌عبارت دیگر افزایش نرخ تورم افزایش قیمت کل سهام را درپی خواهد داشت [۱].

کریم زاده در سال ۱۳۸۵ به بررسی رابطه شاخص قیمت سهام بورس تهران با متغیرهای پولی کلان به کمک روش هم‌جمعی پرداخت. در این مطالعه با استفاده از نظریه پرتفولیو و نظریه اساسی فیشر رابطه شاخص قیمت سهام بورس تهران با متغیرهای کلان پولی بررسی شد. متغیرهای استفاده شده در این پژوهش نقدینگی، نرخ سود بانکی، قیمت سهام بورس، و نرخ ارز بوده و از روش ARDL به‌منظور برآورد مدل استفاده شده است. نتایج حاصل از این تحقیق نشان از وجود یک بردار هم‌جمعی بین متغیرهای پولی کلان و شاخص قیمت سهام دارد. همچنین در بلند مدت نتایج نشان از تاثیر مثبت بین نقدینگی و شاخص قیمت و تاثیر منفی بین نرخ ارز و سود بانکی با شاخص قیمت سهام دارد[۷].

چاتزی انتونیو و همکاران [1] در مقاله‌ای با عنوان واکنش بورس سهام به شوک های سیاست‌های مالی و پولی، رابطه‌ای پویا بین سیاست های مالی و پولی و عملکرد بورس سهام وارز را با استفاده از چهار چوب SVAR مورد بررسی قرار دادند. آنها از داده های هر سه ماه یکبار از سال ۱۹۹۱ تا ۲۰۱۰ در سه کشور، یعنی آلمان، بریتانیا و آمریکا استفاده کردند. متغیر های تحت کنترل شاخص فعالیت اقتصادی جهانی، شاخص قیمت کالاهای مصرفی، مخارج دولتی(به عنوان یک شاخصی از وضعیت سیاست مالی)، M1 (به عنوان یک شاخصی از عرضه پول)،

[1] Chatziantoniou, I. Duffy, D. Filis, G.

نرخ بازار بین بانکی سه ماهه و شاخص های بورس سهام و ارز برای این سه کشوراست، که برای آلمان DAX ۳۰، برای بریتانیا Share All FTSE و برای امریکا Jones Dow می باشد. که در نتیجه این مطالعه، آنها شواهدی را یافتند که نشان می داد ، سیاست های پولی ومالی هر دو بر بورس سهام وارز جه به‌صورت مستقیم و چه غیر مستیم تاثیر می گذارد. مهمتر اینکه، ما شواهدی را یافتیم که نشان می‌دهد تعامل بین این دو سیاست در توصیف توسعه‌ی بورس سهام وارز بسیار مهم است [۱۵].

کانگ و همکاران [1] در مقاله ای با عنوان اثر تغییر زمانی شوک های بازار نفت بر بازار سهام، تغییرات در واریانس شوک‌های ساختاری ، در بازار نفت خام در طول زمان و انتقال شوک‌های بازار نفت به بازار سهام آمریکا در طول زمان را بررسی کردند. اگر شوک‌های قیمتی نفت در زمان تغییر کند، اثرات تغییر را بر اقتصاد واقعی، از طریق مصرف کننده و رفتار شرکت داریم. آنگاه باید تغییری در اثرات قابل مشاهدهٔ شوک‌های قیمتی نفت بر بازار سهام داشته باشیم. که در نتیجه این مطالعه، شوک‌های ساختاری نفت ناشی از بازار جهانی نفت خام، ۷.۲۵ درصد تغییرات بلندمدت در بازده‌های واقعی سهام آمریکا را به خود اختصاص داده است. در مدل پارامتر متغیر با زمان VAR، شوک‌های تقاضای خاص بازار نفت، شوک‌های تقاضای کلی جهانی و شوک‌های عرضه نفت به ترتیب ۱.۹ درصد، ۳.۸ درصد و ۳.۸ درصد، از تغییرات بلندمدت بازده‌های واقعی سهام را در آمریکا تشکیل می‌دهند [۴۷].

۲.۳.۳ سنجش ارتباط شاخص کل بورس (TSE) با عوامل تعیین کننده

معیارهای متعددی برای سنجش ارتباط دو متغیر وجود دارد. مهمترین این معیارها عبارت‌اند از:

ضریب همبستگی اسپیرمن، ضریب همبستگی پیرسون، ضریب هماهنگی تاو کندال، معیار ضریب منحنی تمرکز. که در ادامه به بررسی هر کدام از این معیار ها و سنجش ارتباط شاخص کل بورس با عوامل تعیین کننده می‌پردازیم.

ضریب همبستگی(پیرسون)

یکی از مشهورترین شیوه‌های اندازه‌گیری وابستگی بین دو متغیر کمی، محاسبه ضریب همبستگی (پیرسون) است. این شاخص توسط «کارل پیرسون» آماردان انگلیسی در سال ۱۹۰۰ طی مقاله‌ای معرفی شد. می‌دانیم هر چه مقدار ضریب همبستگی به ۱ یا ۱ـ نزدیک شود، وجود رابطه خطی بین دو متغیر بیشتر می‌شود، چنانچه ضریب همبستگی مثبت باشد، رابطه بین دو متغیر را مستقیم و اگر منفی باشد، رابطه بین دو متغیر معکوس خواهد بود.اگر دو متغیر مستقل باشند، ضریب همبستگی پیرسون برابر با صفر خواهد بود. البته عکس این

[1] Kang, W. Ratti, R. Yoon, K.

موضوع صحیح نیست. یعنی ممکن است ضریب همبستگی پیرسون برای دو متغیر برابر با صفر باشد در حالیکه آن دو متغیر مستقل نیستند.
همانطور که در جدول ۲.۳ مشخص است، بیشترین مقدار ضریب همبستگی(پیرسون) بین شاخص کل بورس (TSE) با سطح عمومی قیمت‌ها (CPI) است پس یک همبستگی قوی بین این دو متغیر وجود دارد. لذا رابطه بین این دو متغیر خطی است. از طرفی چون این مقدار مثبت است، پس رابطه بین این دو متغیر مستقیم است. همچنین مشاهده می‌شود که متغیرهای قیمت سکه طلا (Coin)، اونس جهانی طلا (Gold)، قیمت دلار (Dollar)و قیمت یورو (Euro) نسبت به شاخص کل بورس اوراق بهادار تهران (TSE) همبستگی مثبت و نزدیک به یک دارند که نشان از وجود رابطه خطی مستقیم بین تک تک این متغیرها با TSE است.
همانطور که در جدول ۲.۳ مشاهده می‌شود، ضریب همبستگی بین قیمت نفت اوپک(Oil) با شاخص کل بورس اوراق بهادار تهران (TSE) منفی است و عددی است نزدیک به صفر، پس همبستگی خیلی ضعیفی بین این دو متغیر وجود دارد و مقدار منفی آن ناشی از رابطه معکوس بین این متغیرها است.

جدول ۲.۳: مقادیر ضریب همبستگی پیرسون برای متغیرها

	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
TSE	۰/۹۱۷۸	۰/۹۲۱۱	۰/۹۰۹۱	۰/۸۷۶۰	۱	ـ۰/۳۳۴۹	۰/۹۴۳۰

ضریب همبستگی رتبه‌ای اسپیرمن

با توجه به اینکه ضریب همبستگی پیرسون براساس میانگین و واریانس محاسبه می‌شود، ممکن است در مقابل داده‌های دورافتاده، منحرف شده و میزان همبستگی را به درستی نشان ندهد. در چنین مواقعی از ضریب همبستگی رتبه‌ای اسپیرمن استفاده می‌شود. همبستگی رتبه‌ای اسپیرمن به مانند ضریب پیرسون، نشان می‌دهد تمایل یک متغیر به پیروی کردن از مقدارهای متغیر دیگر چقدر است.
در این ضریب همبستگی به جای محاسبه روی مقدارها، از رتبه‌ها استفاده می‌شود. به همین دلیل به آن ضریب همبستگی رتبه‌ای می‌گویند. مشخص است که در ضریب همبستگی رتبه‌ای اسپیرمن، اساس رتبه‌ها هستند، نه خود مقدارها. همچنین ضریب همبستگی اسپیرمن شدت رابطه خطی را اندازه‌گیری نمی‌کند. به این معنی که ممکن است ضریب همبستگی رتبه‌ای اسپیرمن برابر ۱ باشد در حالی که رابطه خطی بین دو متغیر وجود نداشته باشد. در جدول ۳.۳ می‌توانیم به این موضوع دست پیدا کنیم که دو متغیر (CPI) و (TSE) از همدیگر پیروی می‌کنند، یعنی اگر سطح عمومی قیمت‌ها (CPI) افزایش پیدا کند شاخص کل بورس اوراق بهادار تهران (TSE) هم افزایش پیدا خواهد کرد. همچنین یک همبستگی قوی بین Coin، Dollar و Euro با TSE وجود دارد که به معنی پیروی کردن این متغیرها از هم است. متغیر دیگر

یعنی قیمت نفت اوپک (Oil) یک هم‌بستگی بسیار ضعیفی با TSE دارد.

جدول ۳.۳: مقادیر ضریب رتبه‌ای اسپیرمن برای متغیرها

	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
TSE	۰/۹۵۷۶	۰/۷۰۵۶	۰/۹۵۲۳	۰/۹۴۷۶	۱	۰/۱۱۶۴	۰/۹۷۸۵

ضریب هماهنگی تاو کندال

ضریب هماهنگی تاو کندال نیز مانند ضریب همبستگی اسپیرمن، به جای مقدار از ترتیب مقدارها برای اندازه‌گیری میزان وابستگی استفاده می‌کند. اگر همه زوج‌ها با هم هماهنگ باشند مقدار ضریب هماهنگی کندال برابر است با ۱، اگر همه زوج‌ها ناهماهنگ باشند ضریب هماهنگی کندال برابر است با ۱ـ، اگر X و Y مستقل باشند، انتظار داریم که ضریب هماهنگی کندال نیز برابر با ۰ باشد.

درجدول (۴.۳) ضریب هماهنگی متغیرهای Coin، Dollar، Euro و CPI با شاخص کل بورس اوراق بهادار (TSE) نشان از هماهنگی اغلب زوج‌ها دارد، در این بین ضریب هماهنگی کندال قیمت نفت اوپک (Oil) با شاخص کل بورس اوراق بهادار تهران (TSE) نشان از این امر دارد که اغلب زوج ها ناهماهنگ هستند.

جدول ۴.۳: مقادیر ضریب هماهنگی تاو کندال برای متغیرها

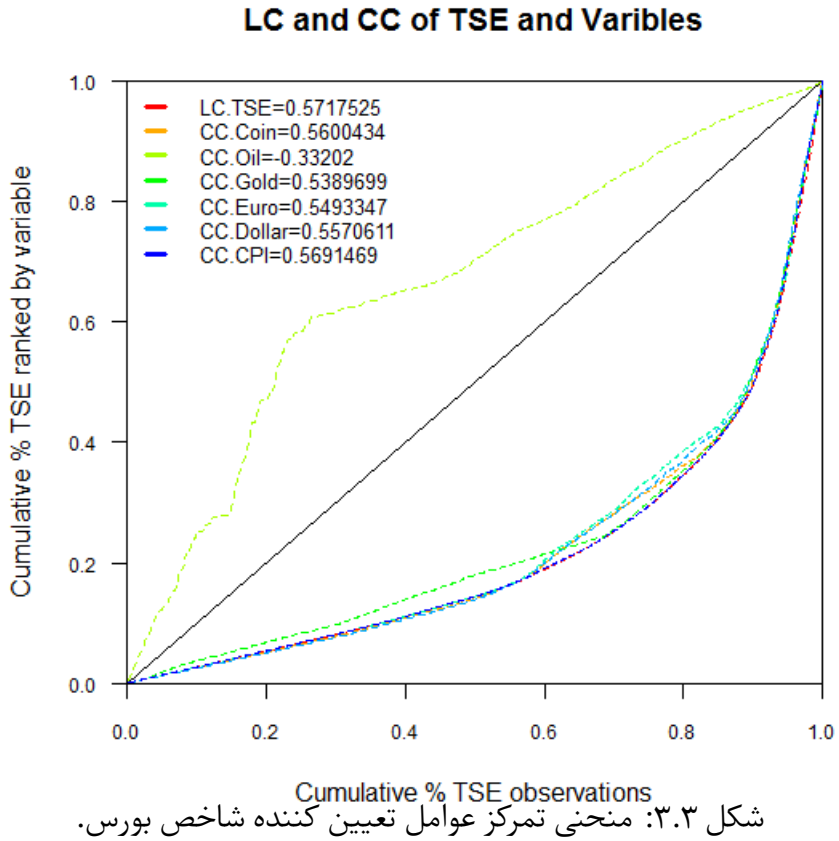
	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
TSE	۰/۸۲۲۶	۰/۵۴۵۰	۰/۸۱۳۵	۰/۸۰۵۶	۱/۰۰۰	۰/۱۱۷۷	۰/۸۹۸۹

معیار ضریب منحنی تمرکز(CC)

با توجه به جدول ۵.۳ مشاهده می‌شود که قیمت نفت اوپک (Oil) یک رابطه منفی با شاخص کل بورس دارد و مابقی عوامل یک رابطه مستقیم با شاخص کل دارند. در بین این عوامل قیمت طلا (Coin)، (CPI) ارتباط قوی‌تری با شاخص کل بورس دارند. همچنین منحنی لورنز TSE در شکل ۳.۳ با LC نشان داده شده است.

جدول ۵.۳: مقایسه معیار ضریب منحنی تمرکز متغیرها با TSE

	Coin	Gold	Dollar	Euro	Oil	CPI
TSE	۰/۵۶۰۰۴۳	۰/۵۳۸۹۶۹	۰/۵۵۷۰۶۱	۰/۵۴۹۳۳۴	ـ۰/۳۳۲۰۲	۰/۵۶۹۱۴۶



شکل ۳.۳: منحنی تمرکز عوامل تعیین کننده شاخص بورس.

۴.۳ مدل‌سازی وپیش‌بینی شاخص کل بورس اوراق بهادار تهران

۱.۴.۳ مدل‌ها و رویکردهای پیش‌بینی شاخص کل بورس

جهت پیش‌بینی شاخص کل بورس اوراق بهادار رویکردها و مدل‌های متفاوتی وجود دارد. در ابتدا برخی از این مدل‌ها خلاصه بیان می‌شود سپس مدل و رویکرد مورد استفاده در تحقیق (GLM) به تفسیر ارائه می‌شود. در نهایت در جدول ۶.۳ پژوهشگرانی که از این مدل ها در مطالعات تجربی خود استفاده کردند را معرفی می‌کنیم.

ماشین بردار پشتیبانی (Support Vector Machines)

این روش از جملهٔ روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان داده‌است. مبنای کاری دسته‌بندی کنندهٔ SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد [۱۲].

یادگیری عمیق (Deep Learning)

یادگیری عمیق، رده‌ای از الگوریتم‌های یادگیری ماشین است که از چندین لایه برای استخراج ویژگی‌های سطح بالا از ورودی خام استفاده می‌کنند. به بیانی دیگر، رده‌ای از تکنیک‌های یادگیری ماشین که از چندین لایه‌ی پردازش اطلاعات و به‌ویژه اطلاعات غیرخطی بهره می‌برد، تا عملیات تبدیل یا استخراج ویژگی نظارت‌شده یا نظارت‌نشده را عموماً با هدف تحلیل یا بازشناخت الگو، کلاس‌بندی یا خوشه‌بندی کند [۲۹، ۶۹].

شبکه عصبی مصنوعی (Artificial Neural Networks)

یک شبکه عصبی مصنوعی، از سه لایهٔ ورودی، خروجی و پردازش تشکیل می‌شود. هر لایه شامل گروهی از سلول‌های عصبی (نورون) است که عموماً با کلیهٔ نورون‌های لایه‌های دیگر در ارتباط هستند، مگر این که کاربر ارتباط بین نورون‌ها را محدود کند؛ ولی نورون‌های هر لایه با سایر نورون‌های همان لایه، ارتباطی ندارند. نورون کوچک‌ترین واحد پردازشگر اطلاعات است که اساس عملکرد شبکه‌های عصبی را تشکیل می‌دهد. یک شبکهٔ عصبی مجموعه‌ای از نورون‌هاست که با قرار گرفتن در لایه‌های مختلف، معماری خاصی را بر مبنای ارتباطات بین نورون‌ها در لایه‌های مختلف تشکیل می‌دهند [۱۶، ۴۲].

منطق فازی (Fuzzy Logic)

منطق فازی شکلی از منطق‌های چند ارزشی بوده که در آن ارزش منطقی متغیرها می‌تواند هر عدد حقیقی بین ۰ و ۱ و خود آن‌ها باشد. این منطق به منظور به‌کارگیری مفهوم درستی جزئی، به‌کارگیری می‌شود، به طوری که میزان درستی می‌تواند هر مقداری بین کاملاً درست و کاملاً غلط باشد [۳۵].

الگوریتم ژنتیک (Genetic Algorithms)

تکنیک جستجو در علم رایانه برای یافتن راه‌حل تقریبی برای بهینه‌سازی مدل و همچنین یافتن راه حل ریاضی و مسائل جستجو است. الگوریتم ژنتیک نوع خاصی از الگوریتم‌های تکاملی است که از تکنیک‌های زیست‌شناسی فرگشتی مانند وراثت، جهش زیست‌شناسی و اصول انتخابی داروین برای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو استفاده می‌شود. الگوریتم‌های ژنتیک اغلب گزینه خوبی برای تکنیک‌های پیش‌بینی بر مبنای رگرسیون هستند [۴۹].

شبکه‌های بیزی (Bayesian Networks)

شبکه‌های بیزی یک گراف جهت‌دار غیرمدور است، که مجموعه‌ای از متغیرهای تصادفی و نحوه ارتباط مستقل آن‌ها را نشان می‌دهد. به عنوان نمونه یک شبکه بیزی می‌تواند نشان دهنده ارتباط بین بیماری‌ها و علائم آن‌ها باشد [۵۰].

درخت‌های تصمیم (Decision Trees)

یک ابزار برای پشتیبانی از تصمیم است که از درخت‌ها برای انتخاب مدل استفاده می‌کند. درخت تصمیم به‌طور معمول در تحقیق‌ها و عملیات‌های مختلف استفاده می‌شود. به‌طور خاص در آنالیز تصمیم، برای مشخص کردن استراتژی که با بیشترین احتمال به هدف برسد بکار می‌رود. استفاده دیگر درختان تصمیم، توصیف محاسبات احتمال شرطی است [۱۳].

جدول ۶.۳: خلاصه‌ای از رویکردهای مورد استفاده و پیش‌بینی شاخص بورس در مطالعات پیشین

نویسنده مقالات	SVM	DL	ANN	FL	B	DT	GA
Bustos et .al (۲۰۱۸)	X		X				
Chakraborty et .al (۲۰۱۸)	X					X	
Coyne et al . (۲۰۱۸)			X				
Fischer and Krauss (۲۰۱۸)		X					
Hu et .al (۲۰۱۸)			X				
Malagrino et .al (۲۰۱۸)					X		
Ren et .al (۲۰۱۸)	X						
Wang،Xu، et .al (۲۰۱۸)	X	X	X				
Ghanavati et .al (۲۰۱۶)				X			
Leito et .al (۲۰۱۷)							X

۲.۴.۳ مدل‌های GLM در پیش‌بینی شاخص کل بورس

مدل خطی تعمیم‌یافته

مدل خطی تعمیم‌یافته [1] ، یک مدل پارامتری بوده و بسط مدل‌های خطی می‌باشد. در مدل خطی تعمیم‌یافته فرمول ارایه می‌گردد و رابطه بین متغیرهای تبیینی و پاسخ به وسیله پارامتر برآورد شده رگرسیونی به اضافه فواصل اطمینان سنجش می‌شود. مدل‌های خطی تعمیم‌یافته برای مواقعی که مشاهدات بطور نرمال توزیع نیافته‌اند و زمانی که سایر روش‌های رگرسیون

[1] Generalized Linear Model

مناسب نمی‌باشد، ابداع شدند. این مدل، در بین روش‌های مدل‌سازی دارای عملکرد خوبی است.
با توجه به این که تعداد متغیرها ۶ مورد هستند، تمام مدل های خطی تعمیم یافته ۶۳ حالت می‌باشد. نمونه‌هایی از مدل های برآورد شده در جدول ۷.۳ قرار گرفته است.

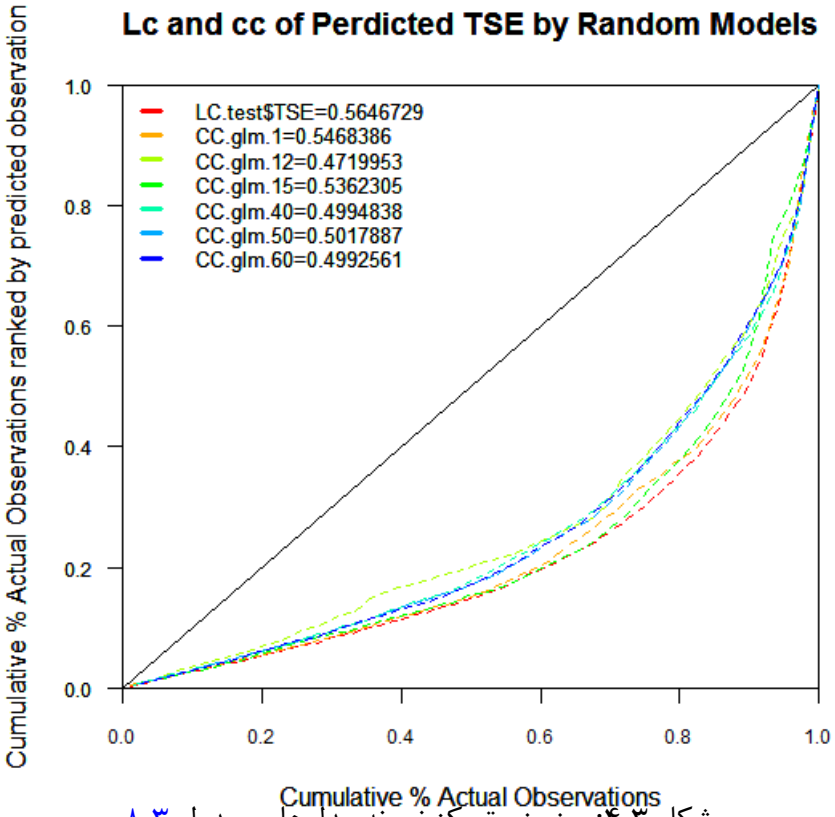
جدول ۷.۳: نمونه‌ای از مدل‌های GLM برآورد شده

glm.۱	TSE~Coin
glm.۱۲	TSE~Oil+Gold
glm.۴۰	TSE~Gold+Dollar+CPI
glm.۱۵	TSE~Oil+CPI
glm.۵۰	TSE~Oil+Gold+Euro+CPI
glm.۶۰	TSE~Coin+Oil+Gold+Euro+Dollar+CPI

همانطور که در جدول ۸.۳ مشاهده می‌شود، مقادیر CC، LC و ABC هر کدام از مدل‌های مربوطه قرار داده شده است. ABC ها از تفاضل LC و CC بدست می‌آیند، که هر چه این مقدار کمتر باشد مدل ما پیش‌بینی دقیق تری انجام می‌دهد.

جدول ۸.۳: مقادیر منحنی تمرکز ، لورنز و ABC نمونه مدل‌های جدول ۷.۳

	glm.۶۰	**glm۵۰**	**glm.۴۰**	**glm.۱۵**	**glm.۱۲**	**glm.۱**
LC	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷
CC	۰/۴۹۹۲۵۶۱	۰/۵۰۱۷۸۸۷	۰/۴۹۹۴۸۳۸	۰/۵۳۶۲۳۰۵	۰/۴۷۱۹۹۵۳	۰/۵۴۶۸۳۸
ABC	۰/۰۶۵۴۱۶۷۵	۰/۰۶۲۸۸۴۱۹	۰/۰۶۵۱۸۹۰۸	۰/۰۲۸۴۴۲۴۲	۰/۰۹۲۶۷۷۵۴	۰/۰۱۷۸۳۴۳۲



شکل ۴.۳: منحنی تمرکز نمونه مدل‌های جدول ۸.۳

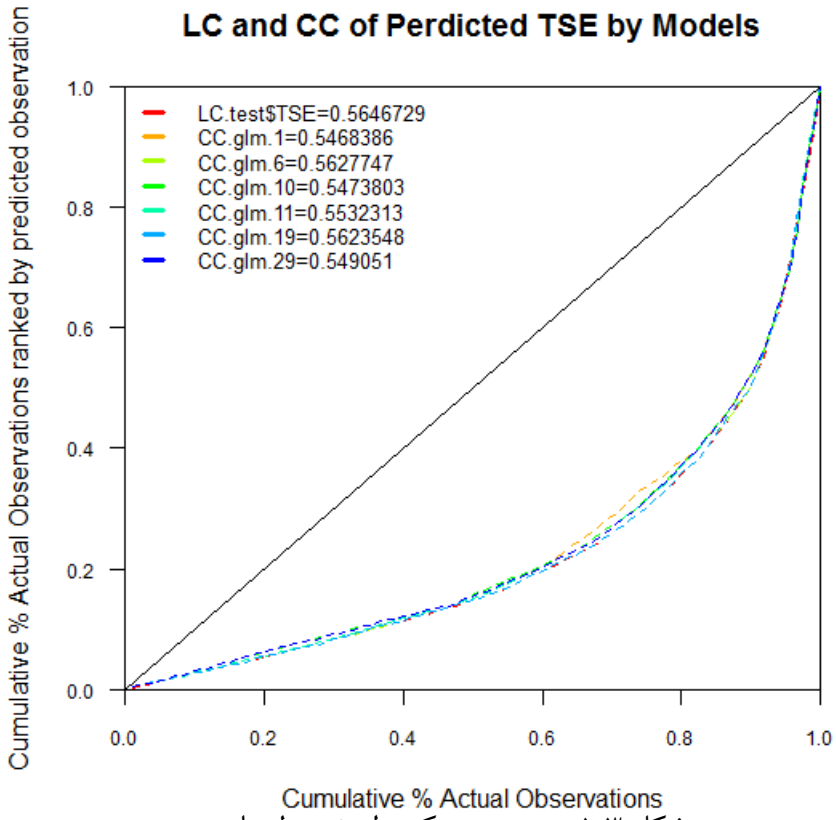
۳.۴.۳ انتخاب مدل پیش‌بینی شاخص کل بر مبنای معیار ABC

از بین تمامی مدل‌های بدست آمده، مدل‌های زیر بر مبنای معیار ABC پیش‌بینی بهتری از شاخص بورس ارائه کردند که به ترتیب عبارت‌اند از:

- glm.۶=TSE~CPI
- glm.۱۹=TSE~Euro+CPI
- glm.۱۱=TSE~Coin+CPI
- glm.۲۹=TSE~Coin+Euro+CPI
- glm.۱۰=TSE~Coin+Euro
- glm.۱=TSE~Coin

در انتخاب مدل‌های ذکر شده ابتدا منحنی لورنز و تمرکز هر کدام از مدل ها را رسم کردیم. می‌دانیم با توجه به۲.۷.۲ مدلی دقیق‌تر است، که فاصله بین منحنی‌های آن مقدار کمتری باشد. بنابراین مدل‌های بالا در بین تمامی مدل ها کمترین مقدار بین منحنی‌ها (ABC) را دارا بودند. در جدول۹.۳ مقادیر بین منحنی‌های هر کدام از مدل‌ها آورده شده است. حال از

بین مدل‌های ذکر شده با توجه به جدول۹.۳ مدل glm.6 که کمترین مقدار ABCرا دارد، بهترین مدل در بین تمامی مدل‌های برآورد شده است. در شکل ۵.۳ منحنی های تمرکز مربوط به مدل ها و همچنین منحنی لورنز مربوط به شاخص کل بورس رسم شده است. همانطور که مشخص است تمامی این مدل ها نسبت به سایر مدل ها انطباق بیشتری بر منحنی لورنز شاخص کل بورس دارد که همین امر نشان از انتخاب درست مدل ها دارد. در این شکل نیز glm.6 بیشترین انطباق را بر منحنی لورنز شاخص کل بورس دارد بنابراین نسب به مدل‌های جدول ۹.۳ پیش‌بینی دقیق‌تری از شاخص بورس ارائه می‌دهد.



شکل ۵.۳: منحنی تمرکز و لورنز مدل‌های برتر.

جدول ۹.۳: مقادیر منحنی تمرکز، لورنز و ABC مدل‌های برتر

	glm.۲۹	glm۱۹	glm.۱۱	glm.۱۰	glm.۶	glm.۱
LC	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷	۰/۵۶۴۶۷
CC	۰/۵۴۹۰۵	۰/۵۶۲۳۵۴	۰/۵۵۳۲۳۱	۰/۵۴۷۳۸۰	۰/۵۶۲۷۷۴	۰/۵۴۶۸۳۸
ABC	۰/۰۱۵۶۲۱	۰/۰۰۲۳۱۸	۰/۰۱۱۴۴۱	۰/۰۱۷۲۹۲	۰/۰۰۱۸۹۸	۰/۰۱۷۸۳۴

۵.۳ نتیجه گیری

در این فصل ابتدا بحث قیمت‌گذاری سهام در بازار ارائه شد، اگر ارزش سهام بیش از ارزش محاسبه شده بر اساس رابطه $P_T = P_0(1+r*)^T$ باشد، آنگاه سهام ارزش‌گذاری بیش از حد شده است. دلایل متعددی برای ارزش‌گذاری بیش از حد وجود دارد از جمله عدم تقارن اطلاعات بین سهامداران، رفتارهای هیجانی و... .
سنجش نسبی هر عامل یا سنجش اندازه اثر عوامل تعیین کننده شاخص کل بورس برای تصمیمات مالی و اتخاذ سیاست‌های درست در بازار اهمیت خاصی دارد. لذا در این تحقیق اندازه اثر با معیارهای چهارگانه (پیرسون، اسپیرمن، ...) سنجیده شده است.
در این پژوهش همچنین جهت سادگی بحث از روش GLM که یک روش خطی است، برای مدلسازی و پیش‌بینی شاخص بورس استفاده شده است. تعداد ۶۳ مدل خطی مختلف در مدل‌سازی و پیش‌بینی استفاده شده است، از بین این مدل‌ها ۶ مدل برتر که پیش‌بینی بهتری را طبق معیار ABC می‌دادند انتخاب کردیم. نتایج حاصل از این معیار در جدول ۹.۳ ارائه شده است. به طور کلی در این پژوهش ما از منحنی تمرکز و منحنی لورنز که حالت خاصی از منحنی تمرکز است، برای سنجش اثر عوامل تعیین کننده شاخص و انتخاب مدل برتر پیش‌بینی شاخص استفاده کردیم.

۶.۳ خلاصه فصل سوم

در این فصل ابتدا در قسمت ۱.۲.۳ به بیان متغیرها و داده‌های مورد استفاده در این تحقیق پرداختیم، پس از آن بازدهی بازار بورس را با سایر بازارهای موازی بررسی نمودیم. نتایج نشان داد بازار بورس ارزش‌گذاری بیش از حد شده است.
در بخش ۳.۳ ابتدا عوامل موثر بر شاخص بورس در مطالعات پیشین بررسی شد و سپس اثر این عوامل را با معیارهای متعدد (پیرسون، تاوکندال، اسپیرمن و ضریت منحنی تمرکز) بر شاخص کل سنجیدیم. در قسمت ۱.۳.۳ ارتباط شاخص کل بورس با متغیرها را به کمک معیار منحنی تمرکز(CC) و سایر معیار ها مورد بررسی قرار داریم. در ادامه کار مدل‌های GLM تعریف کردیم و با استفاده از آن تمامی مدل‌های ممکن را برآورد نمودیم و بر اساس مدل‌های برآورد شده شاخص بورس پیش‌بینی شد. در قسمت ۴.۳ از بین تمامی مدل‌ها بهترین آن ها را انتخاب کرده، و در نهایت در ۳.۴.۳ با استفاده از معیار ABC بهترین مدل را انتخاب کردیم.

مراجع

[۱] بادکوبه‌ای هزاره، علیرضا. (۱۳۷۴)، پایان‌نامه ارشد: ''به بررسی و مطالعه اثر تورم بر قیمت سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران''، دانشگاه تهران.

[۲] بیانی، عذرا. (۱۳۹۱)، پایان‌نامه ارشد: ''تحلیل تاثیر متغیرهاي منتخب اقتصاد کلان بر نوسان‌پذیري شاخص قیمت سهام بورس اوراق بهادار تهران''، دانشکده علوم اداري و اقتصاد، دانشگاه اصفهان.

[۳] پاکدین امیری، ع. پاکدین امیری، م. پاکدین امیری، م. (۱۳۸۸). ''ارائه مدل پیش‌بیني شاخص کل قیمت سهام با رویکرد شبکه‌هاي عصبي''، دوفصلنامه علمی‌ـ‌پژوهشی، ۸۳ـ۱۰۸.

[۴] تقوی، مهدی. محمدی، تیمور. برزنده، محمد. (۱۳۷۶). ''بررسی متغیرهای اقتصادی اثر گذار بر شاخص قیمت سهام بورس اوراق بهادار تهران، مجله برنامه و بودجه، شماره ۴۰و۴۱.

[۵] جعفریان، زینب. کارگر، منصوره. (۱۳۹۶). ''مدل‌سازی پراکنش گونه‌های گیاهی حفاظتی و باارزش در منطقه‌ی توریستی پلور با استفاده از مدل خطی تعمیم‌یافته و مدل جمعی تعمیم‌یافته''، جغرافیا و توسعه شماره ۴۶، ۱۱۷ـ۱۳۲.

[۶] رسد، حمیدرضا. (۱۳۸۸)، پایان‌نامه ارشد: ''بررسی اثر تورم بر روند شاخص‌های بورس تهران''، دانشکده علوم اجتماعی، دانشگاه رازی.

[۷] کریم‌زاده، مصطفی.(۱۳۸۵). ''بررس رابطه بلندمدت شاخص قیمت سهام بورس با متغیرهای کلان پولی با استفاده از روش هم‌جمعی در اقتصاد ایران''. فصلنامه پژوهش‌های اقتصادی ایران، ۲۶، ۴۱ـ۵۴.

[۸] نجارزاده، رضا و دیگران.(۱۳۸۰). ''بررسی نوسانات شوک‌های ارزی و قیمتی بر شاخص قیمت سهام بورس اوراق بهادار تهران با استفاده از رهیافت خودرگرسیون برداری''، فصلنامه پژوهش‌های اقتصادی. شماره اول.

[9] Anderson, T.W., 1996. ”Some inequalities for symmetric convex sets with applications”. **Ann. Statist**. 24 (2), 753–762.

[10] Block, H.W., Sampson, A.R., 1988. ”Conditionally ordered distributions”. **J. Multivari-ate Anal**. 27 (1), 91–104.

[11] Brazauskas, V., Jones, B.L., Puri, M.L., Zitikis, R., 2007. ”Nested L-statistics and their use in comparing the riskiness of portfolios”. **Scand. Actuar. J.** (3), 162–179.

[12] Bustos, O., Pomares, A., Gonzalez, E. (2018). ”A comparison between SVM and multilayer perceptron in predicting an emerging financial market: Colombian stock market”. **In 2017 Congreso internacional de innovacion y tendencias en in genieria, CONIITI 2017 - conference proceedings**: vol. 2018-January (pp. 1–6).

[13] Chakraborty, P., Pria, U., Rony, M., Majumdar, M. (2018).” Predicting stock movement using sentiment analysis of Twitter feed”. **In 2017 6th International conference on informatics, electronics and vision and 2017 7th international symposium in computational medical and health technology, ICIEV-ISCMHT 2017**, vol. 2018-January (pp. 1–6).

[14] Chang, C.S., 1992. ”A new ordering for stochastic majorization: Theory and applications”. **Adv. Appl. Probab**. 604–634.

[15] Chatziantoniou, I., Duffy, D., Filis, G., 2013. ”Stock market response to monetary and fiscal policy shocks: Multi-country evidence”. **Economic Modelling** 30, 754-769.

[16] Coyne, S., Madiraju, P., Coelho, J. (2018). ”Forecasting stock prices using social media analysis”. **In Proceedings**, pp. 1031–1038.

[17] Das Gupta, S., Eaton, M.L., Olkin, I., Perlman, M.D., Savage, L.J., Sobel, M., 1972. ”Inequalities on the probability content of convex regions for elliptically contoured distributions”. **In: Sixth Berkeley Symposium on Probability and Statistics, II**, pp. 241–265.

[18] Denuit, M. (2010). ”Positive dependence of signals”. **Journal of Applied Probability** 47, 893-897.

[19] Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005), ”Actuarial Theory for Dependent Risks: Measures, Orders and Models”. Wiley, New York.

[20] Denuit, M., Mesfioui, M. (2013). ”A sufficient condition of crossing type for the bivariate orthant convex order”. **Statistics and Probability Letters**, 83, 157-162.

[21] Denuit, M., Mesfioui, M. (2017). ”Preserving the Rothschild-Stiglitz type increase in risk with background risk: A characterization”. **Insurance: Mathematics and Economics**, 72, 1-5.

[22] Denuit, M., Mesfioui, M., Trufin, J. (2018). ”Bounds on concordance based validation statistics in regression models for binary responses”. **Methodology and Computing in Applied Probability**, in press.

[23] Denuit, M., Sznajder, D., Trufin, J., (2019). ”Model selection based on Lorenz and concentration curves, Gini indicesand convex order” **Insurance: Mathematics and Economics**.

[24] Denuit, M., Vermandele, C., 1999. ”Lorenz and excess wealth orders, with applications in reinsurance theory”. **Scand. Actuar. J.** (2), 170–185.

[25] Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R., Vyncke, D., 2002. ”The concept of comonotonicity in actuarial science and finance: theory”. **Insurance Math. Econom**, 31 (1), 3–33.

[26] Dhaene, J., Linders, D., Schoutens, W., Vyncke, D., 2012. ”The herd behavior index: a new measure for the implied degree of co-movement in stock markets”. **Insurance Math. Econom**, 50 (3), 357–370.

[27] Eaton,M.L., Perlman,M.D., 1991.”Multivariate probability inequalities: convolution theorems, composition theorems, and concentration inequalities”. **In: Lecture Notes-Monograph Series**, pp. 104–122.

[28] Egozcue, M., Garcia, L.-F., Wong, W.-K., Zitikis, R. (2011). ”Gruss type bounds for covariances and the notion of quadrant dependence in expectation”. **Central European Journal of Mathematics**, 9, 1288-1297.

[29] Fischer, T., Krauss, C. (2018). ”Deep learning with long short-term memory networks for financial market predictions”. **European Journal of Operational Research**, 270 (2), 654–669.

[30] Frechet, M., (1951). ”Sur les tableaux de correlation dont les marges sont donnees”. **Ann. Univ. Lyon.Sect. A.**, (3)14, 53–77.

[31] Frees, E.W., Meyers, G., Cummings, A.D. (2011). ”Summarizing insurance scores using a Gini index”. **Journal of the American Statistical Association**, 106, 1085-1098.

[32] Frees, E.W., Meyers, G., Cummings, A.D. (2013). ”Insurance ratemaking and a Gini index”. **Journal of Risk and Insurance**, 81, 335-366.

[33] Frees, E.W., Meyers, G., Cummings, A.D., 2014. ”Insurance ratemaking and a Gini index”. **J. Risk Insurance**, 81 (2), 335–366.

[34] Frees, E.W.,Meyers, G., Cummings, A.D., 2011. ”Summarizing insurance scores using a Gini index”. **J. Amer. Statist. Assoc**, 106 (495), 1085–1098.

[35] Ghanavati, M., Wong, R., Chen, F., Wang, Y., Fong, S. (2016). ”A generic service framework for stock market prediction”. **In Proceedings**, (pp. 283–290).

[36] Gini, C., 1936. ”On the measure of concentration with special reference to income and statistics”. General Series, vol. 208. Colorado College Publication, pp. 73–79.

[37] Gneiting, T., Raftery, A.E. (2007). ”Strictly proper scoring rules, prediction, and estimation”. **Journal of the American Statistical Association**, 102, 359-378.

[38] Gourieroux, C. (1992). ”Courbes de performance, de s´election et de discrimination”. **Annales d´ Economie et de Statistique**, 28, 107-123.

[39] Gourieroux, C., Jasiak, J. (2011). ”The Econometrics of Individual Risk: Credit, Insurance, and Marketing”. **Princeton University Press**.

[40] Haugh, m., (2016). ”An Introduction to Copulas”. Quantitative Risk Management.

[41] Henckaerts, R., Antonio, K., Clijsters, M., Verbelen, R. (2018). ”A data driven binning strategy for the construction of insurance tariff classes”. **Scandinavian Actuarial Journal**, 2018(8), 681-705.

[42] Hu, H., Tang, L., Zhang, S., Wang, H. (2018). ”Predicting the direction of stock markets using optimized neural networks with Google trends”. Neuro computing .

[43] Joe, H., 1990. ”Multivariate concordance”. **J. Multivariate Anal**, 35, 12–30.

[44] Jones, B.L., Puri, M.L., Zitikis, R., 2006. ”Testing hypotheses about the equality of several risk measure values with applications in insurance”. **Insurance Math. Econom**, 38 (2), 253–270.

[45] Jones, B.L., Zitikis, R., 2005. ”Testing for the order of risk measures: an application of L-statistics in actuarial science”. Metron LXIII (2), 193–211.

[46] Kaas, R., Hesselager, O. (1995). ”Ordering claim size distributions and mixed Poisson probabilities”. **Insurance: Mathematics and Economics**, 17, 193-201.

[47] Kanga, W., Rattib, R.A., Yoon, K.H., 2015 ”Time-varying effect of oil market shocks on the stock market”. **Journal of Banking and Finance**.

[48] Kemperman, J.H.B., 1977. ”On the FKG-inequality for measures on a partially ordered space”. **Indag. Math**, 80 (4), 313–331.

[49] Leito, J., Neves, R. F., Horta, N. (2016). ”Combining rules between pips and sax to identify patterns in financial markets”. **Expert Systems with Applications**, 65 , 242–254.

[50] Malagrino, L., Roman, N., Monteiro, A. (2018). ”Forecasting stock market index daily direction: A Bayesian network approach”. **Expert Systems with Applications**, 105 , 11–22.

[51] Marshall, A.W., Olkin, I., Arnold, B., 2010. ”Inequalities: Theory of Majorization and its Applications”. **Springer Science and Business Media**.

[52] McNeil, A.J., Frey, R., Embrechts, P., 2005. ”**Quantitative Risk Management: Concepts,Techniques and Tools**”. Princeton University Press.

[53] Meyers, G., Cummings, A. D. (2009). ”“Goodness of Fit” vs. “Goodness of Lift””. **Actuarial Review**, 36, 16-17.

[54] Muliere, P., Petrone, S. (1992). ”Generalized Lorenz curve and monotone dependence orderings”. Metron 50, 19-38.

[55] Müller, A., 2001. ”Stochastic ordering of multivariate normal distributions”. **Ann. Inst.Statist. Math**. 53 (3), 567–575.

[56] Müller, A., Scarsini, M., 2000. ”Some remarks on the supermodular order”. **J. Multivariate Anal**, 73, 107–119.

[57] Müller, A., Stoyan, D., 2002. ”**Comparison Methods for Stochastic Models and Risks**”.Wiley.

[58] Noll, A., Salzmann, R., W¨uthrich, M. (2018). ”Case study: French motor third-party liability claims”. Available at SSRN: https://ssrn.com/abstract=3164764

[59] Plackett, R.L., 1954. ”A reduction formula for normal multivariate integrals”. **Biometrika** 41 (3–4), 351–360.

[60] Ren, R., Wu, D., Liu, T. (2018). ”Forecasting stock market movement direction using sentiment analysis and support vector machine”. **IEEE Systems Journal**.

[61] Samanthi, R.G.M., Wei, W., Brazauskas, V., 2015. ”Comparing the riskiness of dependent portfolios via nested L-statistics”. **submitted for publication**.

[62] Samanthi, R.G.M., Wei, W., Brazauskas, V., 2016. ”Ordering Gini indexes of multivariate elliptical risks”. **Insurance: Mathematics and Economics** 68, 84–91.

[63] Shaked, M., Shanthikumar, J.G., 2007. ”Stochastic Orders”. **Springer, New York**.

[64] Shaked, M., Sordo, M. A., Suarez-Llorens, A. (2012). ”Global dependence stochastic orders”. **Methodology and Computing in Applied Probability** 14, 617- 648.

[65] Shaked, M., Tong, Y.L., 1985. ”Some partial orderings of exchangeable randomvariables by positive dependence”. **J. Multivariate Anal**. 17, 333–349.

[66] Slepain, D., 1962. ”The one-sided barrier problemfor Gaussian noise”. **Bell Syst. Tech.J**. 41, 463– 501.

[67] Tong, Y.L., 1990. ”The Multivariate Normal Distributions”. **Springer, New York**.

[68] Tevet, D. (2013). ”Exploring model lift: Is your model worth implementing”. **Actuarial Review** 40, 10-13.

[69] Wang, Q., Xu, W., Zheng, H. (2018). ”Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles”. **Neuro computing**, 299 , 51–61.

[70] Wright, R. (1987). ”Expectation dependence of random variables, with an application in portfolio theory”. **Theory and Decision** 22, 111-124.

[71] Yitzhaki, S. (2003). ”Gini’s mean difference: a superior measure of variability for non-normal distributions”. Metron LXI(2), 285-316.

[72] Yitzhaki, S., Schechtman, E. (2013). ”The Gini Methodology: A Primer on Statistical Methodology”. **Springer**.

پیوست آ

ضمیمه

آ.۱ جدول‌های سنجش ارتباط متغیرهای تحقیق

جدول آ.۱: مقایسه ضریب همبستگی پیرسون متغیرها

	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
Coin	۱	۰/۷۵۰۳	۰/۹۸۴۶	۰/۹۹۳۲	۰/۹۱۷۸	ـ۰/۰۶۵۳	۰/۹۷۰۸
Gold	۰/۷۵۰۳	۱	۰/۷۱۱۹	۰/۶۸۳۲	۰/۹۲۱۱	ـ۰/۴۶۸۶	۰/۸۱۵۸
Dollar	۰/۹۸۴۶	۰/۷۱۱۹	۱	۰/۹۷۵۷	۰/۹۰۹۱	ـ۰/۰۹۹۹	۰/۹۸۳۴
Euro	۰/۹۹۳۲	۰/۶۸۳۲	۰/۹۷۵۷	۱	۰/۸۷۶۰	۰/۰۳۷۲	۰/۹۵۲۲
TSE	۰/۹۱۷۸	۰/۹۲۱۱	۰/۹۰۹۱	۰/۸۷۶۰	۱	ـ۰/۳۳۴۹	۰/۹۴۳۰
Oil	ـ۰/۰۶۵۳	ـ۰/۴۶۸۶	ـ۰/۰۹۹۹	۰/۰۳۷۲	ـ۰/۳۳۴۹	۱	ـ۰/۱۲۱۱
CPI	۰/۹۷۰۸	۰/۸۱۵۸	۰/۹۸۳۴	۰/۹۵۲۲	۰/۹۴۳۰	ـ۰/۱۲۱۱	۱

جدول آ.۲: مقایسه ضریب همبستگی اسپیرمن متغیرها

	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
Coin	۱	۰/۶۵۹۶	۰/۹۷۹۴	۰/۹۸۷۰	۰/۹۵۷۶	۰/۱۴۸۷۸	۰/۹۷۴۷
Gold	۰/۶۵۹۶	۱	۰/۶۰۵۴	۰/۵۷۴۲	۰/۷۰۵۶	-۰/۳۵۲۶۸	۰/۷۰۷۳
Dullar	۰/۹۷۹۴	۰/۶۰۵۴	۱	۰/۹۶۹۹	۰/۹۵۲۳	۰/۰۰۸۶۱	۰/۹۶۷۸
Euro	۰/۹۸۷۰	۰/۵۷۴۲	۰/۹۶۹۹	۱	۰/۹۴۷۶	۰/۲۲۵۳	۰/۹۵۸۸
TSE	۰/۹۵۷۶	۰/۷۰۵۶	۰/۹۵۲۳	۰/۹۴۷۶	۱	۰/۱۱۶۴	۰/۹۷۸۵
Oil	۰/۱۴۸۷	-۰/۳۵۲۶	۰/۰۰۸۶	۰/۲۲۵۳	۰/۱۱۶۴	۱	۰/۱۵۷۶
CPI	۰/۹۷۴۷	۰/۷۰۷۳	۰/۹۶۷۸	۰/۹۵۸۸	۰/۹۷۸۵	۰/۱۵۷۶	۱

جدول آ.۳: مقایسه ضریب هماهنگی تاو کندال متغیرها

	Coin	Gold	Dollar	Euro	TSE	Oil	CPI
Coin	۱	۰/۴۹۵۸	۰/۸۸۹۰	۰/۹۱۳۰	۰/۸۲۲۶	۰/۱۸۳۲	۰/۸۹۰۷
Gold	۰/۴۹۵۸	۱	۰/۴۳۴۲	۰/۴۲۰۷	۰/۵۴۵۰	-۰/۲۵۰۵	۰/۵۴۷۶
Dullar	۰/۸۸۹۰	۰/۴۳۴۲	۱	۰/۸۷۲۵	۰/۸۱۳۵	۰/۰۶۴۱	۰/۸۶۹۳
Euro	۰/۹۱۳۰	۰/۴۲۰۷	۰/۸۷۲۵	۱	۰/۸۰۵۶	۰/۲۳۸۲	۰/۸۵۴۱
TSE	۰/۸۲۲۶	۰/۵۴۵۰	۰/۸۱۳۵	۰/۸۰۵۶	۱	۰/۱۱۷۷	۰/۸۹۸۹
Oil	۰/۱۸۳۲	-۰/۲۵۰۵	۰/۰۶۴۱	۰/۲۳۸۲	۰/۱۱۷۷	۱	۰/۱۵۹۸
CPI	۰/۸۹۰۷	۰/۵۴۷۶	۰/۸۶۹۳	۰/۸۵۴۱	۰/۸۹۸۹	۰/۱۵۹۸	۱

واژه‌نامه فارسی به انگلیسی

آزمون فرض . . . . . Hypothesis test
اعتبارسنجی مدل . . . . . Model validation
اقتصاد . . . . . Economics
امید ریاضی . . . . . Expected value
انحراف معیار . . . . . Standard deviation
بازار سرمایه . . . . . Stock market
بردار تصادفی . . . . . Random vector
بردار تصادفی تعویض‌پذیر . . . . . Exchangeable random vector
بردار مکان . . . . . Location vector
بیمه . . . . . Insurance
بیمه اتکایی . . . . . Reinsurance
بیمه عمومی . . . . . General insurance
بیمه‌گران . . . . . Actuaries
بیمه‌گذاران . . . . . Policyholders
پیش‌بینی کنندگان . . . . . Predictors
تابع توان . . . . . Power function
تابع چگالی احتمال . . . . . Probability density function
تابع چندک . . . . . Quantil function
تابع رگرسیون . . . . . Regression function
تابع فوق‌مدولی . . . . . Supermodular function
تابع محدب صعودی . . . . . Increasing convex function
تابع مفصل . . . . . Copula function
تابع مشخصه . . . . . Characteristic function
ترتیب تصادفی معمولی . . . . . Usual stochastic order
ترتیب فوق‌مدولی . . . . . Supermodular order
ترتیب هماهنگی(توافقی) . . . . . Concordance order

ترتیب محدب ..... Convex order
ترتیب محدب صعودی ..... Increasing convex order
لیفت(بالابری) ..... Lift
تکیه‌گاه ..... Support
توزیع تراز بیضوی ..... Elliptically contoured distribution
توزیع بیضوی ..... Elliptical distribution
توزیع حاشیه‌ای ..... Marginal distribution
توزیع حاشیه‌ای مشترک ..... Common marginal distribution
توزیع حاشیه‌ای یکسان ..... Same marginal distribution
توزیع‌های گسسته ..... Discrete distributions
توزیع نرمال چند متغیره ..... Multivariate normal distribution
حق بیمه ..... Premium
مقادیر حق‌بیمه ..... Premium amounts
جنگل‌های تصادفی ..... Random forests
حق بیمه خالص ..... Pure premium
حق بیمه خالص نامعین ..... Unknown pure premium
حق بیمه فنی ..... Technical premium
رده‌بندی ریسک ..... Risk classification
ریسک بیضوی ..... Elliptical risk
ریسک تجربی ..... Empirical risk
ریسک بیضوی چند متغیره ..... Multivariate elliptical risk
ساختار تبادلی ..... Exchangeable structure
ساختار تبادلی شرطی ..... Conditional exchangeable structure
ساختار وابستگی ..... Dependence structure
سنجه ریسک ..... Risk measure
شبکه‌های عصبی ..... Neural networks
شبیه‌سازی مونت کارلو ..... Monte carlo simulation
شاخص جینی ..... Gini index
قیمت‌گذاری ..... Pricing
ماتریس پراکندگی ..... Dispersion matrix
ماتریس کوواریانس ..... Covariance matrix
ماتریس نیمه معین مثبت ..... Positive semidefinite matrix
مالی ..... Finance
متغیر تصادفی ..... Random variable

مفصل ارشمیدسی ..... Archimedean copula
مفصل فرانک ..... Frank copula
مفصل کلایتون ..... Clayton copula
مفصل گامبل ..... Gumbel copula
مدل افزوده تعمیم‌یافته ..... Generalized additive model
مدل پیش‌بینی شده ..... Predictive model
مدل رگرسیونی ..... Regression model
مدل قیمت گذاری بیمه‌ای ..... Actuarial pricing model
منحنی تمرکز ..... Concentration curve
منحنی تمرکز انتگرال‌دار ..... Integrated concentration curve
منحنی لورنز ..... Lorenz curve
مولد ارشمیدسی ..... Archimedean generator
مولد مشترک ..... Common generator
مولد مشخصه ..... Characteristic generator
نسبیت ..... Relativities
نیکویی برازش ..... Goodness-of-fit
نیمه‌معین مثبت ..... Positive semidefinite
واحد بازاریابی ..... Marketing department
هم‌یکنوایی ..... Comonotonicity
یکنوایی ..... Monotonicity

واژه‌نامه انگلیسی به فارسی

بیمه‌گران ..... Actuaries
مدل های قیمت گذاری بیمه‌ای ..... Actuarial pricing models
مفصل ارشمیدسی ..... Archimedean copula
مولد ارشمیدسی ..... Archimedean generator
تابع مشخصه ..... Characteristic function
مولد های مشخصه ..... Characteristic generators
مفصل کلیتون ..... Clayton copula
مولد مشترک ..... Common generator
حاشیه مشترک ..... Common marginal
توزیع حاشیه‌ای مشترک ..... Common marginal distribution
هم‌یکنوایی ..... Comonotonicity
منحنی‌های تمرکز ..... Concentration curves
ترتیب هماهنگی ..... Concordance order
ساختار تبادلی شرطی ..... Conditional exchangeable structure
ترتیب محدب ..... Convex order
تابع مفصل ..... Copula function
ماتریس کوواریانس ..... Covariance matrix
ساختار وابستگی ..... Dependence structure
توزیع‌های گسسته ..... Discrete distributions
ماتریس پراکندگی ..... Dispersion matrix
اقتصاد ..... Economics
توزیع بیضوی ..... Elliptical distribution
ریسک بیضوی ..... Elliptical risk
توزیع تراز بیضوی ..... Elliptically contoured distribution
ریسک تجربی ..... Empirical risk
ساختار تبادلی ..... Exchangeable structure

بردار تصادفی تعویض‌پذیر ..... Exchangeable random vector
امید ریاضی ..... Expected value
مالی ..... Finance
مفصل فرانک ..... Frank copula
بیمه عمومی ..... General insurance
مدل افزوده تعمیم‌یافته ..... Generalized additive model
شاخص جینی ..... Gini index
نیکویی برازش ..... Goodness-of-fit
مفصل گامبل ..... Gumbel copula
آزمون فرض ..... Hypothesis test
تابع محدب صعودی ..... Increasing convex function
ترتیب محدب صعودی ..... Increasing convex order
بیمه ..... Insurance
منحنی تمرکز انتگرال‌دار ..... Integrated concentration curve
بالابری ..... Lift
بردار مکان ..... Location vector
منحنی لورنز ..... Lorenz curve
توزیع حاشیه‌ای ..... Marginal distribution
واحد بازاریابی ..... Marketing department
اعتبارسنجی مدل ..... Model validation
یکنوایی ..... Monotonicity
شبیه‌سازی مونت کارلو ..... Monte carlo simulations
ریسک بیضوی چند متغیره ..... Multivariate elliptical risk
توزیع نرمال چندمتغیره ..... Multivariate normal distribution
شبکه‌های عصبی ..... Neural networks
بیمه‌گران ..... Policyholders
نیمه‌معین مثبت ..... Positive semidefinite
تابع توان ..... Power function
مدل پیش‌بینی شده ..... Predictive model
پیش‌بینی کنندگان ..... Predictors
حق‌بیمه ..... Premium
مقادیر حق‌بیمه ..... Premium amounts
قیمت‌گذاری ..... Pricing
تابع چگالی احتمال ..... Probability density function

حق بیمه خالص ..... Pure premium
تابع چندک ..... Quantile function
جنگل‌های تصادفی ..... Random forests
متغیر تصادفی ..... Random variable
بردار تصادفی ..... Random vector
بیمه اتکایی ..... Reinsurance
تابع رگرسیون ..... Regression function
مدل رگرسیونی ..... Regression model
نسبیت ..... Relativities
رده‌بندی ریسک ..... Risk classification
سنجه ریسک ..... Risk measure
حاشیه یکسان ..... Same marginal
ماتریس نیمه معین ..... Semidefinite matrix
انحراف معیار ..... Standard deviation
ترتیب تصادفی ..... Stochastic order
بازار سرمایه ..... Stock market
تکیه‌گاه ..... Support
فوق‌مدولی ..... Supermodular
تابع فوق‌مدولی ..... Supermodular function
ترتیب فوق‌مدولی ..... Supermodular order
حق بیمه فنی ..... Technical premium
حق بیمه خالص نامعین ..... Unknown pure premium
ترتیب تصادفی معمولی ..... Usual stochastic order

Abstract

Gini index is a well-known tool in economics that is often used for measuring income inequality. In insurance, the index and its modifications have been used to compare the riskiness of portfolios, to order reinsurance contracts, and to summarize insurance scores(relativities). In this thesis, wee stablish several stochastic orders between the Gini indexes of multivariate elliptical risks with the same marginals but different dependence structures. The comparison of the Gini indexes (of empirically estimated risk measures) presented in this thesis provides a theoretical explanation to this statistical phenomenon. Moreover, it enriches the studies of the problem of central concentration of elliptical distributions and generalizes the pd-1 order proposed by Shaked and Tong(1985)

In order to determine an appropriate amount of premium, statistical goodness-of-fit criteria must be supplemented with actuarial ones when assessing performance of a given candidate pure premium. In this thesis, concentration curves and Lorenz curves are shown to provide actuaries with effective tools to evaluate whether a premium is appropriate or to compare competing alternatives. The idea is to compare the premium income for sub-portfolios gathering low risks(identified as low by means of the premiums under consideration) to the true one, or equivalently, to the actual losses. Numerical illustrations performed on hypothetical data and real ones demonstrate the usefulness of the proposed approach.

Finally, using Generalized linear models with the concentration and Lorenz curves, we examined the factors affecting the Tehran Stock Exchange (TSE) index, and selected the model that provided a better forecast of the TSE index.

Keyword: Elliptical distribution, Dependence structure, Comonotonicity, Usual stochastic order, Increasing convex order, Supermodular order, Pd-1 order, Pricing, Risk classification, Concentration curve, Lorenz curve, Generalized Linear Model (GLM), Tehran stock exchange(TSE), Model selection.

Faculty of Mathematical Sciences

MSc Thesis in Financial Mathematics

Ordering Gini indexes of multivariate elliptical risks and model selection based on Lorenz and concentration curves

By: Majid Jahani

Supervisors:

Dr. Ali Reza Khoddami
Dr. Mohammad Mirbagherijam

February 2021