



ORIGINAL RESEARCH PAPER

Analysis of mortality models with covariates missing at random

S. Shoaee*, R. Fathi

Department of Actuarial Science, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

ARTICLE INFO

Article History:

Received 23 April 2021

Revised 12 June 2021

Accepted 2 August 2021

Keywords:

Capital requirement

Full rank

Identifiability

Likelihood contribution

Missing at random

Mortality model

Parameter redundant

ABSTRACT

BACKGROUND AND OBJECTIVES: Demographic indicators such as mortality rates play a very important role in health, financial and pension policies. Therefore, the accuracy of mathematical models in estimating mortality rates is an important challenge. One of the tasks of actuaries is to construct a suitable mortality model for the available data so that these mortality models can calculate mortality for different ages and longevity, as well as the different information available to individuals on retirement plans. Missing data is a problem that may be faced by actuaries when they are analyzing the real data. Missing data can occur for a variety of reasons, such as unanswered or censored. The presence of missing data can pose a threat to the accuracy of the data analysis results. The purpose of this study is to model the mortality in a retirement plan. In this regard, it is assumed that data are available at the individual level, including date of birth, date of joining the retirement plan, date of completion of the observation, and reason for discontinuation (usually death or right censoring). Information on covariate variables such as gender, benefits or size of pension, demographic geography or health status will also be available. More precisely, this study aims to model the mortality in a retirement plan based on missing data and access to information from various covariate variables, to carefully analyze the structure of different models, to estimate and finally to investigate the financial implications for different mortality experiences containing missing data.

METHODS: Demographic indicators such as mortality rates play a very important role in health, financial and pension policies. Therefore, the accuracy of mathematical models in estimating mortality rates is an important challenge. One of the tasks of actuaries is to construct a suitable mortality model for the available data so that these mortality models can calculate mortality for different ages and longevity, as well as the different information available to individuals on retirement plans. Missing data is a problem that may be faced by actuaries when they are analyzing the real data. Missing data can occur for a variety of reasons, such as unanswered or censored. The presence of missing data can pose a threat to the accuracy of the data analysis results. The purpose of this study is to model the mortality in a retirement plan. In this regard, it is assumed that data are available at the individual level, including date of birth, date of joining the retirement plan, date of completion of the observation, and reason for discontinuation (usually death or right censoring). Information on covariate variables such as gender, benefits or size of pension, demographic geography or health status will also be available. More precisely, this study aims to model the mortality in a retirement plan based on missing data and access to information from various covariate variables, to carefully analyze the structure of different models, to estimate and finally to investigate the financial implications for different mortality experiences containing missing data.

FINDINGS: First, complete data are analyzed and modeled for observations of members of a retirement plan, which includes survival time and ancillary variables for each individual. Estimation of parameters is obtained using the maximum likelihood method. However, when the data is missing, it is not easy to estimate the parameters with the maximum likelihood method. In this case, the model parameters are estimated by the maximum likelihood method which are calculated using the proposed algorithm; then, statistical indicators such as identifiability of parameters are calculated to evaluate the performance of the proposed structure and algorithm. Furthermore, the financial effects, in particular the annuity factors, and the mis-estimation risk capital requirements for the mortality experience which includes the maximum covariates variables are calculated and compared with the individual segments when the data are missing. In addition, it can be seen that when the two statistical variables are not observed together, the model is not identifiable according to the data.

CONCLUSION: It was found that if the data are missing, the statistical model is not always identifiable using the maximum likelihood, and data combination from two or more experiments can avoid identifiable barriers. The methods proposed in this paper can be useful for actuaries when calculating financial committees based on annuity factors. These methods may combine different datasets with equal or similar mortality experiences, increase sample size, and reduce parameter risk, thus, reducing capital requirements. Socio-economic variables such as the level of benefits and geographical characteristics of the population are given more attention if the interest rate is low.

*Corresponding Author:

Email: sh_shoaee@sbu.ac.ir

Phone: +9821 29902912

ORCID: [0000-0003-4755-0117](https://orcid.org/0000-0003-4755-0117)

DOI: [10.22056/ijir.2021.03.02](https://doi.org/10.22056/ijir.2021.03.02)





مقاله علمی

تجزیه و تحلیل مدل‌های مرگ‌ومیر با متغیرهای کمکی دارای گمشدگی تصادفی

شیرین شعاعی^{*}، ریحانه فتحی

گروه بیم‌سنجی، دانشکده علوم ریاضی، دانشگاه شهید بهشتی، تهران، ایران

اطلاعات مقاله

تاریخ‌های مقاله:

تاریخ دریافت: ۳ اردیبهشت ۱۴۰۰

تاریخ داور: ۲۲ خرداد ۱۴۰۰

تاریخ پذیرش: ۱۱ مرداد ۱۴۰۰

کلمات کلیدی:

الزام سرمایه

پارامتر افزونه

رتبه کامل

سهم در دستمای

شناسایی پذیری

گمشدگی تصادفی

مدل مرگ‌ومیر

چکیده:

پیشینه و اهداف: این پژوهش با هدف مدل‌بندی مرگ‌ومیر در یک طرح بازنشستگی بر اساس داده‌های گمشده و دسترسی به اطلاعات مختلف از متغیرهای کمکی، تجزیه و تحلیل دقیق ساختار مدل‌های مختلف، برآوردیابی و در نهایت بررسی تأثیرات مالی برای تجربه‌های مختلف مرگ‌ومیر و حاوی داده‌های گمشده انجام شده است.

روش‌شناسی: این مقاله با یک طرح بازنشستگی سروکار دارد که در آن طول عمر آتی هر فرد با مدل‌های بقای پارامتری با ترکیب متغیرهای کمکی که ممکن است برای برخی از افراد گمشده باشند، مدل‌سازی شده است. پارامترها با روش ماکسیمم درست‌نمایی برآورد شده و الگوریتمی پیشنهاد گردیده که بتواند وظیفه برآوردیابی را به بهترین شکل ممکن انجام دهد.

یافته‌ها: نتایج نشان داد در صورتی که داده‌ها گمشده باشند، مدل آماری همیشه با استفاده از ماکسیمم درست‌نمایی شناسایی‌پذیر نیست و تلفیق داده‌های حاصل از دو یا چند تجربه می‌تواند از موانع شناسایی‌پذیری جلوگیری نماید.

نتیجه‌گیری: روش‌های پیشنهادی این مقاله هنگام محاسبه کمیت‌های مالی مورد علاقه براساس عامل‌های مستمری، برای بیم‌سنجان می‌تواند مفید باشد. این روش‌ها ممکن است مجموعه داده‌های مختلف با تجربه مرگ‌ومیر برابر یا مشابه را با هم ترکیب کنند، اندازه نمونه را افزایش دهند و ریسک پارامتر را کاهش دهند، بنابراین منجر به کاهش الزام سرمایه شوند. متغیرهای اقتصادی-اجتماعی از جمله سطح مزایا و مشخصات جغرافیایی جمعیتی در صورت پایین بودن نرخ بهره بیشتر مورد توجه قرار می‌گیرند.

^{*}نویسنده مسئول:

ایمیل: sh_shoae@sbu.ac.ir

تلفن: ۰۲۹۹۰۲۹۱۲ ۹۸۲۱+

ORCID: 0000-0003-4755-0117

DOI: 10.22056/ijir.2021.03.02

توجه: مدت زمان بحث و انتقاد برای این مقاله تا ۱ اکتبر ۲۰۲۱ در وب‌سایت IJIR در «نمایش مقاله» باز می‌باشد

است خطای برآورد حاصل شود، که می‌توان با مدل‌سازی از یک جمعیت بزرگ‌تر آن را کاهش داد. به‌همین علت معمولاً از طرح‌های بازنشستگی ترکیبی که همگن باقی بمانند، استفاده می‌شود. این امر باعث افزایش اندازه نمونه، کاهش ریسک پارامتر و در نتیجه، کاهش الزام سرمایه می‌گردد.

در این پژوهش بعد از بیان مقدمه، در ادامه، مبانی نظری و پیشینه پژوهش به اختصار مرور شده است. بخش بعدی شامل نمادگذاری مورد نیاز برای توزیع احتمال متغیرهای کمکی، توصیف تابع درستمایی (Capital Requirement) و توسعه آن برای زمانی که داده‌ها گمشده هستند، است. همچنین در ابتدای این بخش، به تحلیل و مدل‌بندی داده‌های کامل برای مشاهدات اعضای یک طرح بازنشستگی که شامل زمان بقا و متغیرهای کمکی برای هر فرد است، پرداخته شده است. همچنین در این بخش براساس مشاهدات کامل از داده‌های صندوق بازنشستگی، ابتدا مجموعه داده‌هایی حاوی داده‌های گمشده با حذف متغیرهای کمکی برای برخی واحدها شبیه‌سازی شده‌اند. سپس مسئله برآورد، شرایط ریاضی برای شناسایی‌پذیری (Identifiability) برآورد پارامترها و الگوریتم برازش توصیف شده است. به‌علاوه مشاهده می‌شود در حالتی که دو متغیر آماری توأم مشاهده نمی‌شوند، مدل با توجه به داده‌ها شناسایی‌پذیر نیست. برای غلبه بر عدم شناسایی‌پذیری، نمونه بسیار کوچکی در نظر گرفته شده که متغیرهای کمکی آن کاملاً مشاهده شده است. در پایان درباره تأثیر مالی ترکیب مجموعه داده‌های مختلف برای برآورد شدت مرگ‌ومیر مطالعه شده و در نهایت نتیجه‌گیری ارائه گردیده است.

مبانی نظری پژوهش

در بررسی بسیاری از مجموعه داده‌ها، با داده‌های گمشده روبه‌رو می‌شویم. گمشدگی داده‌ها می‌تواند به‌علل مختلفی مانند پاسخ داده‌نشده یا سانسور شده به‌وجود آید. در مجموعه داده‌ها در موارد پاسخ داده‌نشده، برخی از اطلاعات هر فرد موجود نیست. موارد سانسور شده از جایی به بعد تمامی اطلاعات درباره فرد موجود نیست که می‌تواند به‌دلیل فوت یا هر علت دیگری فرد از مشاهده خارج شود. وجود داده گمشده می‌تواند منجر به تهدیدی برای صحت نتایج حاصل از تحلیل داده‌ها شود. در نظر نگرفتن داده گمشده یا استفاده از روش‌های نامناسب برخورد با این داده‌ها، به‌عنوان مثال جایگذاری داده‌های گمشده با میانگین می‌تواند به مشکلاتی همچون آماسیدگی (Inflation) در نقطه‌ای مانند میانگین منجر شود. بسیاری از محققین روش‌هایی را برای مقابله با مشکلات داده گمشده ارائه کرده‌اند. اکثر این روش‌ها وابسته به نوع سازوکار گمشدگی داده‌ها هستند و برحسب آن می‌توان روش تحلیل مناسبی را انتخاب نمود. در این راستا، Rubin (1976) سیستم طبقه‌بندی را برای مشاهدات داده گمشده معرفی نمود که به‌طور گسترده مورد استفاده قرار می‌گیرد. به‌طور معمول سه مکانیسم گمشدگی شامل گمشدگی تصادفی (Missing at Random (MAR))، گمشدگی کاملاً تصادفی

شاخص‌های جمعیتی همواره نقش مهمی در تصمیم‌گیری‌های جمعیتی داشته‌اند و نرخ مرگ‌ومیر مهم‌ترین عامل تغییر الگوی جمعیت به‌شمار می‌آید. داشتن اطلاعات درست از آینده جمعیت برای برآورد دقیق جمعیت سالخورده به‌منظور برنامه‌ریزی‌های دقیق مالی جهت تأمین آینده افراد، ساخت جداول آتی عمر مطابق با محتمل‌ترین الگوی مرگ‌ومیر و قیمت‌گذاری صحیح سالانه‌های عمر، تنها بخشی از کاربردهای پیش‌بینی مقادیر آتی نرخ مرگ‌ومیر در جمعیت‌شناسی، بیمه و آمار است.

مدل‌بندی نرخ مرگ‌ومیر پیشینه‌ای طولانی دارد و جداول مرگ‌ومیر برای بیم‌سازان، یکی از متداول‌ترین و قدیمی‌ترین روش‌های بررسی میزان مرگ‌ومیر و ابزار اصلی توصیف الگوی مرگ‌ومیر تلقی می‌شوند. رویکرد یک جدول مرگ‌ومیر بر این اساس است که آزمایش گذشته بدون هیچ دگرگونی در آینده تکرار شود. لذا تلاش‌های بسیاری برای دستیابی به مدلی که بتواند این جداول را تبیین کند، صورت گرفته است. یافتن مدل ریاضی برای احتمال مرگ‌ومیر از نظر کاربردی مفیدتر است. روشن است که استفاده از تابعی که پارامتر کمی دارد و از نظر هندسی دارای منحنی همواری است به‌مراتب آسان‌تر از کار با جدول عمری است که پارامترهای فراوانی دارد و در موارد بسیاری نیز احتیاج به درون‌یابی، برون‌یابی، هموارسازی و تعدیل‌های متعدد آماری دیگر، در جهت یافتن احتمال‌های فوت دارد که در محاسبه حق‌بیمه فنی استفاده می‌شوند. لذا بیم‌سازها علاقه‌مند هستند که مدل مرگ‌ومیری را ارائه دهند که بتواند برای سنین و طول عمرهای متفاوت و همچنین اطلاعات متفاوتی که برای افراد در طرح‌های بازنشستگی موجود است، شدت مرگ‌ومیر را محاسبه نماید.

بنابراین، مطالعه حاضر با هدف مدل‌بندی مرگ‌ومیر در یک طرح بازنشستگی انجام شده است. می‌توان فرض نمود که داده‌ها در سطح فردی شامل تاریخ تولد، تاریخ پیوستن به طرح، تاریخ اتمام مشاهده و دلیل قطع مشاهده (معمولاً مرگ یا سانسور از راست) در دسترس هستند. همچنین سایر اطلاعات به‌صورت متغیرهای کمکی در دسترس خواهند بود. از جمله متغیرهای کمکی، جنسیت است که معمولاً شناخته شده است. متغیرهای کمکی دیگری که شامل اطلاعات غنی‌تری هستند، از قبیل مزایا یا اندازه حقوق بازنشستگی، مشخصات جغرافیایی جمعیتی یا وضعیت سلامتی نیز در دسترس هستند.

در طرح‌های بازنشستگی اطلاعات متفاوتی برای هر فرد موجود است و وابستگی مرگ‌ومیر به این اطلاعات به‌خصوص زمانی که ریسک روی زیرمجموعه‌هایی از اعضا متمرکز است، بسیار اهمیت دارد. در بررسی داده‌ها در طرح‌های بازنشستگی با مواردی مواجه می‌شویم که برخی از اطلاعات در مورد هر فرد موجود نیست. در تحلیل مدل‌های مرگ‌ومیر با حضور مقادیر گمشده (Missing Values) می‌توان از روش ماکسیمم درست‌نمایی (Maximum Likelihood) استفاده نمود. در این‌گونه حالت‌ها اغلب هنگام برازش مدل ممکن

اگر در یک مدل آماری، داده‌ها گمشده وجود داشته باشد، ممکن است پارامتر افزونه حاصل شود و روی شناسایی پذیری اثر بگذارد. (Catchpole and Morgan 1997) در قضیه ۴ مقاله خود نشان دادند اگر یک مدل پارامتر افزونه باشد، به صورت محلی شناسایی پذیر نیست. مدل‌های آماری که پارامتر افزونه نیستند، به صورت رتبه کامل تعریف می‌شوند. اگر یک مدل رتبه کامل باشد، مقادیر ویژه ماتریس اطلاع تجربی ارزیابی شده توسط برآورد ماکسیمم درستنمایی (MLE) Maximum likelihood Estimation)) همگی منفی و از صفر دور خواهند بود. برای تشخیص رتبه کامل بودن به ماتریس ژاکوبین نیاز است. اگر رتبه نمادین این ماتریس با تعداد پارامترهای مدل (تعداد ردیف‌های ماتریس ژاکوبین Jacobian Matrix) برابر بود، مدل رتبه کامل (Full Rank) است. اما اگر رتبه نمادین ماتریس ژاکوبین کمتر از تعداد پارامترهای مدل حاصل شود، مدل پارامتر افزونه (Parameter Redundant) و در نتیجه شناسایی پذیر نیست. پارامتر افزونه به حالتی گفته می‌شود که بتوان مدل آماری مدنظر را با بردار پارامتری کوچک‌تری نیز تعریف نمود. بنابراین مدل دارای تعداد زیادی پارامتر (بیش از تعداد مورد نیاز) است و در نتیجه شناسایی پذیر نیست.

در طرح‌های بازنشستگی اطلاعات متفاوتی برای هر فرد موجود است و وابستگی مرگ‌ومیر به این اطلاعات به خصوص زمانی که ریسک روی زیرمجموعه‌هایی از اعضا متمرکز است، بسیار اهمیت دارد. در مدل‌بندی داده‌ها در طرح‌های بازنشستگی با مواردی مواجه می‌شویم که برخی از اطلاعات در مورد هر فرد موجود نیست. یکی از رویکردهای موجود در برآورد پارامترهای مدل‌های مرگ‌ومیر با حضور مقادیر گمشده استفاده از روش ماکسیمم درستنمایی است. اما در اغلب موارد به دلیل عدم دسترسی به برخی از داده‌ها در برآورد پارامترهای مدل خطای برآورد حاصل شود، که می‌توان آن‌ها را با مدل‌سازی از یک جمعیت بزرگ‌تر کاهش داد. به همین علت معمولاً از طرح‌های بازنشستگی ترکیبی که همگن باقی بمانند، استفاده می‌شود. این امر باعث افزایش اندازه نمونه، کاهش ریسک پارامتر و در نتیجه، کاهش الزام سرمایه می‌گردد.

روش‌شناسی پژوهش

مشاهدات و تابع درستنمایی

در این بخش، ساختار تابع درستنمایی توأم برای برآورد بردار پارامترهای مدل در حالت مشاهدات کامل ارائه و سپس برای مواردی که داده گمشده وجود دارد، توسعه شده است.

مشاهدات کامل

در تجزیه و تحلیل رگرسیونی، متغیرهای کمکی اغلب قطعی در نظر گرفته می‌شوند، در این بخش آن‌ها را به عنوان متغیر تصادفی مدل‌بندی می‌کنیم. مجموعه متغیرهای کمکی تصادفی یک فرد با بردار تصادفی P بعدی Z نشان می‌دهیم. مقدار تحقق یافته Z شدت مرگ‌ومیر فرد $\mu_x(z; \tau)$ را تعیین می‌کند. تابع توزیع و تابع چگالی

(Missing Completely at Random (MCAR)) و گمشدگی غیر تصادفی (Missing not at Random (MNAR)) وجود دارد. در این راستا، بردار پاسخ برای فرد i به صورت $y_i = (y_{i1}, \dots, y_{ij})$ است. برای هر متغیر پاسخ برداری از متغیرهای نشانگر به صورت $m_{ij} = (m_{i1}, \dots, m_{ij})$ که در آن $m_{ij} = 0$ اگر y_{ij} مشاهده شود و $m_{ij} = 1$ وقتی y_{ij} مشاهده نشود وجود دارد. بردار مشاهدات برای هر فرد به صورت y_i^{obs} (مقادیر مشاهده شده) و y_i^{mis} (مقادیر گمشده) تفکیک می‌شود.

در گمشدگی تصادفی داده‌ها به طور تصادفی گمشده هستند. بدین مفهوم که احتمال گمشدن داده‌ها در یک متغیر Y به برخی از متغیرهای اندازه‌گیری شده دیگر وابسته، اما به مقدار گمشده Y وابسته نیست. به عبارت دیگر بین داده‌های گمشده در Y و مقادیر Y پس از جدا کردن سایر متغیرها رابطه‌ای وجود ندارد.

امروزه بسیاری از محققین روش‌هایی را برای مقابله با مشکل داده گمشده ارائه کرده‌اند که اغلب آنها وابسته به نوع ساز و کار گمشدگی داده‌ها هستند. از جمله می‌توان به روش‌های Wilks (1932) و Lord (1955) اشاره کرد که بر توزیع نرمال تمرکز کرده‌اند، در حالی که در زمینه داده‌های بقا، Schluchter and Jackson (1989) یک مدل لوگ خطی برای تجزیه و تحلیل داده‌های بقا سانسور شده همراه با عوامل قطعی در نظر گرفتند (Herring and Ibrahim 2001) بر مدل خطرات متناسب کاکس (Cox Proportional Hazards Model) تمرکز کرده‌اند. در این دو مطالعه الگوی کلی داده‌های گمشده مطرح و پارامترها با استفاده از الگوریتم EM (Expectation Maximization) برآورد شده‌اند. Little and An (2004) یک روش مبتنی بر احتمال قوی MAR را برای داده‌های چند متغیره با مقادیر گمشده بر اساس رگرسیون روی اسپلاین‌های نمره‌های تمایل ارائه دادند. Tsiatis (2006) برای داده‌های گمشده و سانسور شده یک روش با مدل‌های نیمه پارامتری معرفی نمود. Chen et al. (2014) یک مدل‌بندی برای داده‌های طولی و بقا بر روی داده‌های گمشده و سانسور شده از چپ معرفی نمودند که بر پایه مدل کاکس، مدل آمیخته خطی و مدل‌های بیزی بود (Xu et al. 2017) از معادلات برآورد ساز مورد انتظار برای تولید یک نظریه کلی استفاده نمودند. شایان ذکر است که رویکرد آن‌ها در شرایطی که متغیرهای کمکی و برآمدها پیوسته یا گسسته هستند قابل استفاده است، اما برای برآورد واریانس از روش Bootstrap بایستی استفاده نمود. این روش نیمه پارامتری، یک طرح با داده‌های مقطعی و متغیرهای مستقل از زمان را در نظر می‌گیرد. در نهایت (Ungolo et al. 2019) به برآورد پارامترهای مدل برای داده‌های بقا بر اساس داده‌های گمشده به وسیله اعمال یک توزیع آمیخته برای متغیرهای کمکی و روش بهینه‌سازی مقید پرداختند. یکی از ویژگی‌های لازم برای کفایت مدل آماری، به خصوص زمانی که داده‌ها حاوی مقادیر گمشده‌اند، شناسایی پذیری است. در صورت عدم شناسایی پذیری، می‌توان ادعا نمود که مدل آماری رتبه کامل نبوده و مدل مناسبی برای داده‌ها نیست. یک مدل رتبه کامل مدلی است که پارامتر افزونه ندارد. به طور خاص

$$L_i(\tau, \zeta) = S_{x_i}(t_i, z_i; \tau) f_Z(z_i; \zeta), \quad (2)$$

که $S_{x_i}(t_i, z_i; \tau)$ تابع بقا و بیانگر این است که فرد x ساله حداقل t سال زنده مانده است و داریم:

$$S_{x_i}(t_i, z_i; \tau) = \exp\left\{-\int_0^{t_i} \mu_{x_i+s}(z_i; \tau) ds\right\}, \quad (3)$$

که μ_{x_i+s} بیانگر شدت مرگ‌ومیر فرد i ام دارای سن $x+S$ است. با تلفیق این دو رابطه و ارتباط شدت مرگ‌ومیر، تابع احتمال و بقا، سهم مشارکت در تابع درستنمایی فرد i ام به صورت زیر است:

$$L_i(\tau, \zeta) = \exp\left\{-\int_0^{t_i} \mu_{x_i+s}(z_i; \tau) ds\right\} \mu_{x_i+t_i}(z_i; \tau)^{d_i} f_Z(z_i; \zeta). \quad (4)$$

اگر عناصر Z برای تمام افراد مشاهده شود، تابع درستنمایی به یک تابع از τ و یک تابع از ζ تبدیل می‌شود:

$$L(\tau, \zeta) = \prod_i \underbrace{\exp\left\{-\int_0^{t_i} \mu_{x_i+s}(z_i; \tau) ds\right\} \mu_{x_i+t_i}(z_i; \tau)^{d_i}}_{L^{\tau}(\tau)} \prod_i \underbrace{f_Z(z_i; \zeta)}_{L^{\zeta}(\zeta)}. \quad (5)$$

در این حالت τ می‌تواند به صورت مستقل از ζ برآورد و توزیع متغیرهای کمکی در تحلیل رگرسیونی نادیده گرفته می‌شود. در بخش بعد مشاهده می‌شود که اگر بعضی از مشاهدات Z_i گمشده باشند، این کار امکان‌پذیر نیست و پارامترهای τ و ζ باید توأمآ برآورد شوند.

مشاهدات گمشده

در بررسی بسیاری از مجموعه داده‌ها، با داده‌های گمشده روبه‌رو می‌شویم. گمشدگی داده‌ها می‌تواند به علل مختلفی مانند پاسخ داده‌نشده یا سانسور شده به وجود آید. حال فرض کنید بعضی از مولفه‌های متغیرهای کمکی Z برای برخی از افراد تحت مطالعه گمشده هستند. لذا بردار p بعدی $Z_i = (Z_{i,p}, \dots, Z_{i,1})$ به دو بردار Z_i^{obs} و Z_i^{mis} تقسیم می‌شود. برای هر فرد تابع چگالی بردار متغیرهای کمکی $Z_i = (Z_i^{obs}, Z_i^{mis})$ به صورت زیر تجزیه می‌شود:

$$f_{Z_i}(z_i) = f_{Z_i^{obs}}(z_i^{obs}; \zeta) f_{Z_i^{mis}|Z_i^{obs}}(z_i^{mis} | z_i^{obs}; \zeta). \quad (6)$$

متغیرهای کمکی دارای گمشدگی تصادفی هستند، لذا عبارت $f_{Z_i^{mis}|Z_i^{obs}}(z_i^{mis} | z_i^{obs}; \zeta)$ را می‌توان نادیده گرفت. مشابه رابطه (4)، میزان سهم درستنمایی برای هر فرد، بر پایه طول عمر مشاهده شده T ، سن و مجموعه متغیرهای کمکی مشاهده شده Z^{obs} به صورت زیر است:

$$L_i(\tau, \zeta) = \int_{S_{z_i^{mis}}} \exp\left\{-\int_0^{t_i} \mu_{x_i+s}(z_i^{obs}, y; \tau) ds\right\} \mu_{x_i+t_i}^{d_i}(z_i^{obs}, y; \tau) f_Z(z_i^{obs}, y; \zeta) dy, \quad (7)$$

یا احتمال Z ، به ترتیب با $F_z(z; \zeta)$ و $f_z(z; \zeta)$ نشان داده می‌شود که ζ یک بردار پارامتری ناشناخته است. فرض کنید n تعداد افراد مشاهده شده در یک دوره زمانی محدود است. نمادهای زیر را تعریف می‌کنیم:

- ✓ x_i سن فرد i ام در آغاز دوره مشاهده است،
- ✓ t_i مقدار تحقق یافته متغیر تصادفی T_i ، بیانگر زمان زندگی شده فرد i ام در طول دوره است،
- ✓ d_i شمارنده‌ای است که مقدار تحقق یافته متغیر تصادفی D_i را نشان می‌دهد؛ اگر فرد i ام فوت کرده باشد مقدار آن یک و اگر بر اثر سانسور از مشاهده خارج شده باشد مقدار آن صفر است،
- ✓ z_i مقدار تحقق یافته متغیر تصادفی Z_i ، که متغیرهای کمکی فرد i ام را مشخص می‌کند.

در طول یک دوره مشاهدات متناهی برای آن دسته از افرادی که فوت می‌کنند، T_i بیانگر طول عمر باقی مانده برای سن x_i است. برای افراد زنده (سانسور شده از راست)، تمام آن چیزی که می‌دانیم این است که طول عمر باقی مانده آن‌ها بیشتر از T_i خواهد بود.

فرض می‌کنیم متغیرهای کمکی Z_1, \dots, Z_n مستقل و هم توزیع (i.i.d.) (Independent and Identically Distributed (i.i.d.) و روی مقادیر تحقق یافته خود شرطی هستند. به علاوه سنین مشاهده شده x_1, \dots, x_n و طول عمرهای مشاهده شده T_1, \dots, T_n به صورت i.i.d. هستند. در نهایت فرض می‌شود که سانسورها به صورت ناآگاهی بخش هستند. یعنی می‌توان طول عمرها را به عنوان متغیر تصادفی در نظر و توزیع آن‌را مستقل از زمان سانسور، برآورد نمود، (Macdonald et al., 2018).

در ادامه فرض می‌کنیم مشاهدات به صورت کامل در دسترس هستند، به طوری که هر بردار متغیر کمکی z_i به طور کامل مشاهده شده است. اگر سهم فرد i ام در تابع درستنمایی را با L_i نمایش دهیم، بنابراین تابع درستنمایی کل به صورت $L = \prod L_i$ خواهد بود. برای افرادی که فوت می‌شوند ($d_i = 1$) سهم تابع درستنمایی به وسیله ضرب چگالی شرطی T_i روی x_i و z_i مربوطه در تابع چگالی یا احتمال z_i به دست می‌آید که به صورت زیر محاسبه می‌گردد:

$$L_i(\tau, \zeta) = f_{x_i}(t_i | z_i; \tau) f_Z(z_i; \zeta), \quad (1)$$

که در آن (τ, ζ) بردار پارامترهای نامعلوم هستند و در رابطه فوق چگالی توأم (T, Z) به عنوان مدل کامل (Full Model) (برای سن x) تعریف می‌شود، و همچنین به چگالی T شرطی شده روی Z مدل جزئی (Partial Model) گفته می‌شود. عبارت $f_{x_i}(t_i | z_i; \tau)$ به این معنی است که فرد x ساله بعد از سن $x+t$ ، دیگر زنده نیست.

برای افرادی که طول عمرهای مشاهده شده آن‌ها سانسور شده از راست است ($d_i = 0$)، زمان دقیق فوت مشاهده نشده است. لذا سهم تابع درستنمایی برای این افراد به صورت زیر محاسبه می‌شود:

دو ایتیم آخر، بردار متغیرهای کمکی را تشکیل می‌دهند. در این مجموعه داده مزایا بر حسب پوند است و می‌توان با آن به‌عنوان متغیر پیوسته برخورد نمود، اما به چند دلیل آن‌را به‌صورت متغیر گسسته لحاظ می‌نماییم. اول: *Macdonald et al. (2018)* نشان دادند که استفاده مستقیم آن در مدل مناسب نیست. دوم: اعمال متغیر پیوسته به‌عنوان یک متغیر طبقه‌بندی شده، از لحاظ امساک در رویکرد مدل‌بندی راحت‌تر است. سوم: *Madrigal et al. (2011)* مشاهده کردند که مقدار پایین حقوق بازنشستگی می‌تواند یک شاخص گمراه‌کننده برای ثروتمندی فرد بوده که می‌تواند به این دلیل رخ دهد که یک کارمند (یا مستخدم، عضو) برای خدمات طولانی مدت حقوق پایینی داشته یا برای خدمات کوتاه مدت حقوق بالایی داشته است.

تجزیه و تحلیل مقدماتی داده‌ها

در این بخش ویژگی‌های مجموعه داده کامل را توصیف می‌کنیم. ابتدا نرخ خام مرگومیر D_x / E_x در هر سال واحد از سن محاسبه و در شکل ۱ نمایش داده شده است. نرخ‌های مرگومیر به‌طور جداگانه برای هر دو سطح مزایا (نمودار سمت چپ) و سه گروه مشخصات جغرافیایی جمعیتی (نمودار سمت راست) رسم شده است. در این نمودار D_x بیانگر تعداد افرادی است که بین سن‌های دقیق x و $x+1$ فوت شده‌اند و E_x (در معرض ریسک) بیانگر مجموع تمام زمان‌هایی است که تمام افراد بین این دو سن زنده هستند. براساس شکل ۱ سه نتیجه مهم حاصل می‌شود:

- ✓ تقریباً یک رابطه لگ-خطی بین سن و نرخ‌های مرگومیر وجود دارد.
- ✓ بین مزایای بالاتر و مرگومیر پایین‌تر ارتباط وجود دارد.
- ✓ بین مشخصات جغرافیایی جمعیتی دارای کم‌ترین محرومیت، با مرگومیر پایین‌تر ارتباط وجود دارد.

انتخاب مدل پارامتری

برای مدل‌بندی و به‌منظور ساده نگه داشتن مدل، از اختلاف کوچک بین نرخ مرگومیر برای گروه‌های مختلف در سنین بالا

که روی فضای وضعیت $S_{z_{mis}}$ (State Space) از Z_i^{mis} انتگرال گرفته می‌شود. اگر مولفه‌های Z_i^{mis} دارای فضای وضعیت گسسته باشند، انتگرال روی $S_{z_{mis}}$ تبدیل به مجموع خواهد شد. در حالت داده‌های گمشده، تابع درستنمایی کل را نمی‌توان مشابه رابطه (۵) به دو تابع جداگانه از τ و γ عامل‌بندی نمود. بنابراین نمی‌توان τ را به‌طور مستقل از γ برآورد کرد و بایستی آن‌ها را به‌صورت توأم برپایه طول عمر مشاهده‌شده و متغیرهای کمکی برآورد نماییم. همچنین زمانی‌که داده‌ها گمشده و پارامترهای τ و γ هر دو نامعلوم هستند، شناسایی‌پذیری مدل کامل را نمی‌توان تصدیق نمود.

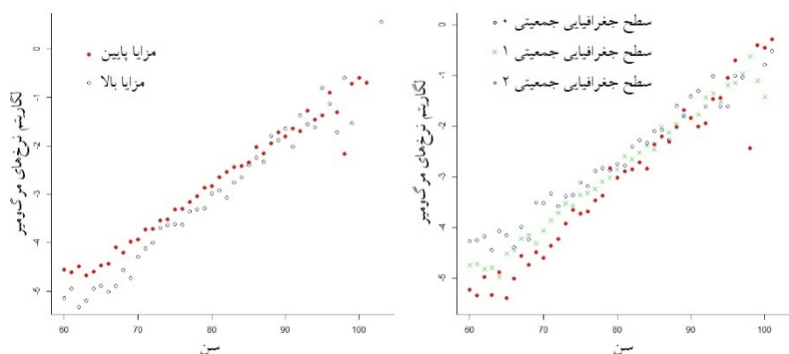
نتایج و بحث

تحلیل و مدل‌بندی بر اساس داده‌های کامل

هدف این بخش مدل‌بندی داده‌ها برای مشاهدات اعضای یک صندوق بازنشستگی است. برای هر عضو صندوق، دو متغیر کمکی وجود دارد. این مجموعه داده کامل نامیده می‌شوند.

مجموعه داده‌های کامل

- مجموعه داده‌های کامل دارای مشخصه‌های زیر است:
- ✓ ۱۸۷۴۱ مستمری در حال پرداخت است،
 - ✓ مجموع سال‌های در معرض ریسک افراد برابر $۱۷۲۶۰ \frac{۱}{۴}$ سال است،
 - ✓ تعداد افراد فوت شده ۴۹۵۶ نفر است،
 - ✓ دوره مشاهده از ۱۰ نوامبر ۱۹۹۲ تا ۳۱ دسامبر ۲۰۰۹ است،
 - ✓ مقدار مزایای سالانه برای هر فرد مشاهده‌شده را با B_i نشان می‌دهیم که یک متغیر کمکی با دو سطح مزایای بالا و مزایای پایین است،
 - ✓ مشخصات جغرافیایی جمعیتی برای هر فرد برپایه سیستم طبقه‌بندی موزاییک، را با C_i نمایش می‌دهیم و شامل سه سطح ۰ و ۱ و ۲ است. سطح ۰ نماینده محروم‌ترین مناطق و سطح ۲ مناطق با کم‌ترین محرومیت است.



شکل ۱. نمودار خام نرخ مرگومیر D_x / E_x بر اساس سن

جدول ۱. ساختار ۴ مدل بر اساس نیروی مرگومیر

مدل	ساختار مدل
M_0	$\exp\{\alpha + \beta x_i\};$
M_1	$\exp\{\alpha + \beta x_i + \gamma \mathbb{1}_{[b_i=High]}\};$
M_2	$\exp\{\alpha + \beta x_i + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]}\};$
M_3	$\exp\{\alpha + \beta x_i + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]}\}.$

این مدل‌ها بر اساس نیروی مرگومیر در جدول زیر ارائه شده است: در **جدول ۱**، مدل M_0 یک مدل گومپرتز بدون متغیر کمکی، مدل‌های M_1 و M_2 تنها دارای یک متغیر کمکی (به ترتیب مقدار مزایا و مشخصات جغرافیایی جمعیتی) و مدل M_3 هر دو متغیر کمکی را دارد. برای مجموعه داده کامل، تابع درستنمایی به صورت رابطه (۵) است. لذا تنها برآورد τ نیاز است، پس کافی است فقط عامل $L^r(\tau)$ را در نظر گرفته و $L^r(\zeta)$ را نادیده بگیریم. به عنوان مثال رابطه $L^r(\tau)$ برای مدل M_3 به صورت زیر است:

$$L^r(\tau) = \prod_i \exp \left\{ - \int_0^{t_i} \exp \left\{ \alpha + \beta(x_i + s) + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]} \right\} ds \right\} \times \exp \left\{ \alpha + \beta(x_i + t_i) + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]} \right\}^{d_i}, \quad (11)$$

لذا لگاریتم درستنمایی به صورت زیر حاصل می‌شود:

$$l(\tau) = - \sum_i \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp \left(\alpha + \beta x_i + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]} \right) \right\} + \sum_{d_i=1} \left(\alpha + \beta(x_i + t_i) + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]} \right). \quad (12)$$

تابع درستنمایی برای سایر مدل‌ها به طریق مشابه حاصل می‌گردد. برای بررسی نیکویی برازش مدل‌ها، معیار اطلاع بیزی (Bayesian Information Criterion) (BIC) را محاسبه و آن‌ها با یکدیگر مقایسه می‌کنیم. این نتایج برای مجموعه داده‌های کامل در **جدول ۲** ارائه شده است.

بر اساس نتایج **جدول ۲**، مدل‌های حاوی متغیرهای کمکی B و C (به صورت جداگانه یا توأم)، برازش مدل را بهبود می‌بخشند. کمترین مقدار BIC برای مدل M_3 است که هر دو متغیر کمکی را دارد. بنابراین M_3 مدل بهتری برای داده‌های موجود است. برای بررسی کیفیت برازش مدل‌ها، باقی‌مانده انحراف پواسون به صورت رابطه زیر محاسبه می‌شود:

$$r_i = \text{sign}(d_i - \Lambda_i) \sqrt{2 \left[d_i \log \frac{d_i}{\Lambda_i} - (d_i - \Lambda_i) \right]}, \quad (13)$$

چشم‌پوشی و از مدل Gompertz با جداسازی گروه‌های خاص استفاده می‌کنیم که در آن نیروی مرگومیر برای فرد i ام از رابطه زیر حاصل می‌شود:

$$\mu(b_i, c_i; \tau) = \exp \left\{ \alpha + \beta x_i + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]} \right\}, \quad (8)$$

که $\tau = (\alpha, \beta, \gamma, \delta_1, \delta_2)$ بردار پارامترهای نامعلوم، پارامتر β تأثیر سن، γ تأثیر مزایای بالا و δ_1 و δ_2 تأثیر مشخصات جغرافیایی جمعیتی مختلف است. تابع شدت مرگومیر ارائه شده در رابطه ۸ متناظر با مدل بقای $S_{x_i}(t_i | b_i, c_i; \tau)$ است که از رابطه زیر محاسبه می‌گردد:

$$S_{x_i}(t_i | b_i, c_i; \tau) = \exp \left\{ - \int_0^{t_i} \mu_{x+s} ds \right\} = \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp \left(\alpha + \beta x_i + \gamma \mathbb{1}_{[b_i=High]} + \delta_1 \mathbb{1}_{[c_i=1]} + \delta_2 \mathbb{1}_{[c_i=2]} \right) \right\}. \quad (9)$$

تابع چگالی برای طول عمر باقی‌مانده به صورت زیر حاصل می‌شود:

$$f_{x_i}(t_i | b_i, c_i; \tau) = S_{x_i}(t_i | b_i, c_i; \tau) \mu_{x_i+t_i}(b_i, c_i; \tau). \quad (10)$$

طبق رابطه (۸) پارامترهای δ_1 ، γ و δ_2 بیانگر تفاوت نرخ مرگومیر بین گروه‌های متفاوت است. مرگومیر پایه $\exp(\alpha + \beta x)$ در مدل، بیانگر نرخ مرگومیر برای فرد x ساله که دارای مزایا پایین و در مناطق جغرافیایی جمعیتی با محرومیت بیشتر زندگی می‌کند، است.

نتایج تجربی برای مجموعه داده‌های کامل

در ادامه به بررسی بهترین برازش مدل در رابطه ۸ بر اساس مدل‌های آشیانه‌ای با ترکیب‌های مناسب از پارامترهای δ_1 ، γ و δ_2 ، می‌پردازیم. در این راستا ۴ مدل مختلف حاصل می‌شود. ساختار

جدول ۲. برآورد پارامترها، لگاریتم درستنمایی، BIC و برآورد خطای استاندارد برآورد پارامترها.

مدل	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$L^T(\tau)$	BIC
M_0	-11.58	0.11	-	-	-	-۲۰۵۹۲/۶۱	۴۱۲۰۴/۹۰
$\hat{\sigma}(\hat{\theta}) \times 10^3$	130.99	1.67	-	-	-	-	-
M_1	-11.54	0.11	-۰/۲۸	-	-	-۲۰۵۶۱/۸۸	۴۱۱۵۳/۲۷
$\hat{\sigma}(\hat{\theta}) \times 10^3$	131.51	1.68	۱۳/۶۳	-	-	-	-
M_2	-11.42	0.11	-	-۰/۲۴	-۰/۴۸	-۲۰۵۲۸/۹۳	۴۱۰۹۷/۲۱
$\hat{\sigma}(\hat{\theta}) \times 10^3$	132.97	1.68	-	۳۳/۵۷	۴۳/۰۴	-	-
M_3	-11.41	0.11	-۰/۲۱	-۰/۲۲	-۰/۴۳	-۲۰۵۱۲/۶۹	۴۱۰۷۴/۵۷
$\hat{\sigma}(\hat{\theta}) \times 10^3$	133.25	1.69	۳۶/۸۵	۳۳/۷۵	۴۳/۸۹	-	-

C را مشاهده کردیم (B را حذف کردیم). این دو زیرمجموعه را به عنوان دو طرح بازنشستگی جداگانه در نظر می‌گیریم که علاقه‌مند به مدل‌بندی آن‌ها به صورت توأم هستیم.

از آنجائی که مجموعه داده ترکیبی توسط یک منبع تولید شده (مجموعه داده کامل)، لذا هیچ ناهمگنی بین داده‌ها وجود ندارد و می‌توان یک قانون مرگ‌ومیر یکسان برای اعضای P_1 و P_2 در نظر گرفت.

به دلیل این که متغیرهای گمشده رسته‌ای هستند، انتگرال در رابطه ۷ به یک مجموع متناهی روی فضای نمونه مشاهدات گمشده تبدیل می‌شود. برای هر فرد در P_1 که مشخصات جغرافیایی جمعیتی مشاهده نشده است، سهم تابع درستنمایی به صورت زیر است:

$$L_i(\tau, \zeta) \propto \sum_{c \in \{0,1,2\}} S_{x_i}(t_i | b_i, c; \tau) \mu_{x_i+t_i}^{d_i}(b_i, c; \tau) f_{B,C}(b_i, c; \zeta). \quad (15)$$

و برای اعضای P_2 که متغیر مزایا مشاهده نشده است، سهم تابع درستنمایی به فرم زیر است:

$$L_i(\tau, \zeta) \propto \sum_{b \in \{Low, High\}} S_{x_i}(t_i | b, c_i; \tau) \mu_{x_i+t_i}^{d_i}(b, c_i; \tau) f_{B,C}(b, c_i; \zeta). \quad (16)$$

بنابراین تابع درستنمایی به صورت زیر است:

$$L(\tau, \zeta | t, x, b_{obs}, c_{obs}, d) \propto \prod_{\{i \in P1\}} L_i \prod_{\{i \in P2\}} L_i. \quad (17)$$

هدف برآورد توأم پارامتر τ از تابع مرگ‌ومیر، پارامتر ζ از تابع احتمال توأم برای سطوح مزایا و مشخصات جغرافیایی جمعیتی است. یک روش منطقی ساختن یک توزیع توأم مناسب برای B و C است.

شناسایی‌پذیری

حال دو شرط ریاضی که برای شناسایی‌پذیری بردار پارامتری $\theta = (\tau, \zeta)$ لازم است را ارائه می‌کنیم. فرض می‌کنیم که هیچ‌کدام از متغیرهای کمکی B و C برای یک فرد مشاهده نشده‌اند. سهم درستنمایی توسط توزیع آمیخته متناهی که اجزای آن تمام نتایج ممکن (B,C) است، ارائه می‌شود. براساس رابطه (۷) تابع چگالی

d_i که مجموع تعداد مرگ‌ومیرها در سن i ام و $\lambda_{i,j}$ است. پارامتر توزیع پواسون λ برای مدل M_1 و $i = 1, \dots, 24$ و $j = 1, \dots, 18741$ از رابطه زیر محاسبه می‌شود:

$$\lambda_{i,j} = \left(\frac{\exp(\beta \min((60+i-x_{i,j}), (la_{i,j}-x_{i,j}))) - 1}{\beta} \right) \times \exp(\alpha + \beta x_{i,j} + \gamma 1_{[b_{i,j}=High]}), \quad (14)$$

که $\Lambda_i = \sum_{j=1}^{18741} \lambda_{i,j}$ است. باقی‌مانده انحراف پواسون برای همه مدل‌ها در شکل ۲ رسم شده است. مشاهده می‌شود که باقی‌مانده‌ها در اطراف صفر متمرکز و از مدل خاصی تبعیت نمی‌کنند.

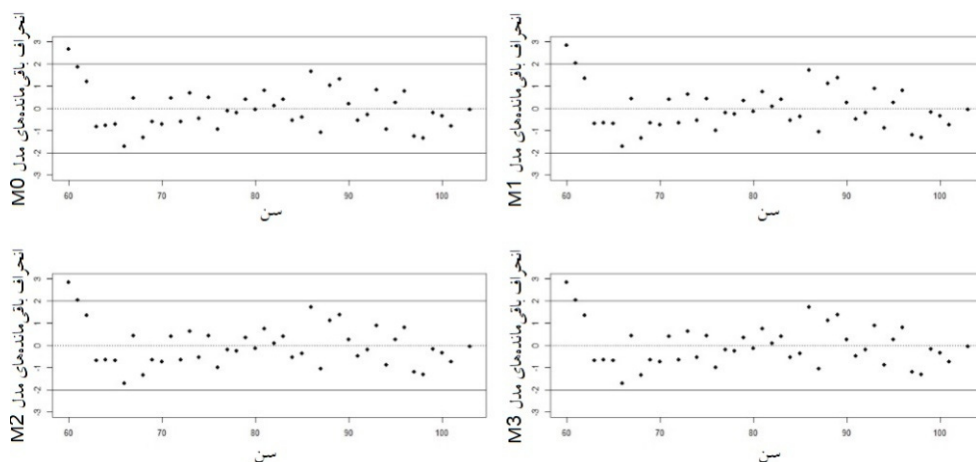
با توجه به مقادیر ویژه ماتریس هسین در برآورد پارامترها که همگی منفی و از صفر دور هستند، لذا تمامی مدل‌ها شناسایی‌پذیر خواهند بود. همچنین با توجه به مجموعه داده‌های کامل، توزیع توأم متغیرهای کمکی B و C به وسیله فراوانی نسبی آن‌ها که در جدول ۳ گزارش شده، برآورد شده است. این مقادیر همان مقادیر بردار پارامتری ζ هستند.

تحلیل و مدل‌بندی بر اساس مشاهدات گمشده

در این بخش با استفاده از حذف برخی از اطلاعات افراد، داده گمشده را شبیه‌سازی می‌کنیم. برای این منظور مجموعه داده کامل را به طور تصادفی به چند بخش تقسیم و یکی از متغیرهای کمکی یا هر دو آن‌ها را در هر زیرمجموعه داده کامل حذف می‌کنیم.

تولید مجموعه داده حاوی مشاهدات گمشده

برای تولید مجموعه داده‌های حاوی مشاهدات گمشده، مجموعه داده کامل را که شامل n فرد است، به صورت تصادفی به دو زیرمجموعه تقسیم و آن‌ها را با P_1 و P_2 نمایش می‌دهیم که هر کدام به ترتیب شامل $n_1 = 9370$ و $n_2 = 9371$ عضو هستند ($n_1 + n_2 = n$). فرض می‌کنیم برای هر فرد در P_1 تنها متغیر کمکی B را مشاهده کردیم (C را حذف کردیم) و برای هر فرد در P_2 تنها متغیر کمکی



شکل ۲. نمودار باقی مانده انحراف پواسون در برابر سن برای هر ۴ مدل

جدول ۳. فراوانی نسبی سطح مزایا و مشخصات جغرافیایی جمعیتی برای طرح بازنشستگی

سطح مزایا (B)	مشخصات جغرافیایی جمعیتی (C)		
	۲	۱	۰
پایین	۰/۷۵	۰/۴۳	۰/۱۹
بالا	۰/۲۵	۰/۱۳	۰/۰۳
	۱	۰/۲۲	۰/۲۲

آمیخته در رابطه ۱۸ شناسایی پذیر است.

زمانی که داده گمشده وجود دارد، شناسایی پذیری مدل را از روش‌های دیگری نیز می‌توان بررسی نمود. هنگام برآورد پارامترهای مدل ممکن است برآوردهایی به دست آیند که دارای واریانس‌های بسیار بزرگ و یا بسیار کوچک و نزدیک به صفر هستند و یا ممکن است برآوردها در فضای پارامتری موردنظر، تعریف نشوند، که می‌توان نتیجه گرفت مدل شناسایی پذیر نیست. همچنین شناسایی پذیری را می‌توان از طریق مقادیر ویژه ماتریس هسین (Hessian Matrix) (حاصل شده در مرحله برآوردیابی) نیز بررسی نمود. در صورتی که مقادیر ویژه (Eigenvalues)، همگی منفی و از صفر دور به دست آیند، نتیجه می‌شود مدل آماری شناسایی پذیر است.

الگوریتم برازش

زمانی که داده‌ها گمشده هستند، به حداکثر رساندن تابع درست‌نمایی ممکن است به آسانی صورت نگیرد. مسئله کلی محاسباتی این است که چگونه محدودیت‌ها روی بردار پارامتری ζ را با توجه به توزیع توام B و C ، کنترل نماییم (عناصر ζ بین صفر و یک محدود شده‌اند و مجموع آن‌ها برابر یک است). به همین دلایل هر بهینه‌سازی مقید معمولی ممکن است شامل عدم پایداری و استواری شود.

الگوریتم EM که ابتدا توسط (Dempster et al., 1997) ارائه

احتمال آمیخته با ۶ مولفه، را می‌توان به صورت زیر نوشت:

$$f_x(t; \tau, \zeta) = \zeta_1 f_x(t | b = Low, c = 0; \tau) + \zeta_2 f_x(t | b = High, c = 0; \tau) + \zeta_3 f_x(t | b = Low, c = 1; \tau) + \zeta_4 f_x(t | b = High, c = 1; \tau) + \zeta_5 f_x(t | b = Low, c = 2; \tau) + \zeta_6 f_x(t | b = High, c = 2; \tau). \quad (18)$$

شروط زیر برای شناسایی پذیری مدل آمیخته متناهی لازم است (McLachlan and Peel, 2000):

۱. برای سنین مختلف با طول عمرهای متفاوت، مقادیر متفاوت (B,C) باید تابع چگالی احتمال متفاوتی را نتیجه دهند. یعنی برای $h \neq k$ باید داشته باشیم:

$$f_x(t | (b, c) = h; \tau) \neq f_x(t | (b, c) = k; \tau). \quad (19)$$

۲. مقادیر ζ_j برای $j = 1, \dots, 6$ بایستی به صورت $0 < \zeta_j < 1$ باشد.

شایان ذکر است که شرط‌ها لازم هستند، اما کافی نیستند (Titterton et al., 1985). اگر دو شرط بالا برقرار شد، مدل

۴. مقادیر $\hat{\tau}$ و $\hat{\pi}$ را به وسیله

$$(\hat{\tau}, \hat{\pi}) = \arg \max_{(\tau, \pi)} \log L(\tau, ILR^{-1}(\pi)). \quad (25)$$

۵. با استفاده از طرح تکراری نیوتن رافسون با مقادیر اولیه $\hat{\tau}^{(0)}$ و $\hat{\pi}^{(0)}$ محاسبه نمایید.

۶. تبدیل ILR معکوس را برای محاسبه $\hat{\zeta}$ از روی $\hat{\pi}$ استفاده نمایید.

$$\hat{\zeta} = ILR^{-1}(\hat{\pi}). \quad (26)$$

از آنجائی که جواب رابطه درستنمایی می تواند از انتخاب مقادیرهای اولیه، تأثیر بگیرد. در پیاده سازی، چندین مقدار اولیه در نظر گرفته شد و MLE ای انتخاب شد که بیشترین مقدار را برای تابع لگاریتم درستنمایی نتیجه می داد.

نتایج تجربی برای مجموعه مشاهدات گمشده

برای تحلیل خصوصیات برآوردگر $(\hat{\tau}, \hat{\zeta})$ ، ابتدا مجموعه داده کامل را به صورت تصادفی به دو بخش مطابق ساختار بخش ۴-۲ تقسیم می کنیم. بدین ترتیب دو مجموعه داده مصنوعی P_1 و P_2 ایجاد می گردد و سپس آن ها را به صورت مصنوعی ترکیب می نماییم که برای برآورد $\theta = (\tau, \zeta)$ استفاده می شوند. برآوردگر بر پایه P_1 و P_2 با $\hat{\theta}_1$ نمایش داده می شود. با توجه به تابع چگالی رابطه ۱۸، لگاریتم درستنمایی برای مجموعه P_1 به صورت زیر است:

$$l_i(\tau, \zeta) \propto (\zeta_1 + \zeta_2) \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma 1_{[b=High]}) \right\} \\ \times \exp \left\{ \alpha + \beta(x_i + t_i) + \gamma 1_{[b=High]} \right\}^{d_i} \quad (27) \\ + (\zeta_3 + \zeta_4) \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma 1_{[b=High]} + \delta_1) \right\} \\ \times \exp \left\{ \alpha + \beta(x_i + t_i) + \gamma 1_{[b=High]} + \delta_1 \right\}^{d_i}$$

لگاریتم درستنمایی برای مجموعه P_2 به صورت زیر است:

$$l_i(\tau, \zeta) \propto (\zeta_1 + \zeta_3 + \zeta_5) \\ \times \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \delta_1 1_{[c=1]} + \delta_2 1_{[c=2]}) \right\} \\ \times \exp \left\{ \alpha + \beta(x_i + t_i) + \delta_1 1_{[c=1]} + \delta_2 1_{[c=2]} \right\}^{d_i} \quad (28) \\ + (\zeta_2 + \zeta_4 + \zeta_6) \\ \times \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma + \delta_1 1_{[c=1]} + \delta_2 1_{[c=2]}) \right\} \\ \times \exp \left\{ \alpha + \beta(x_i + t_i) + \gamma + \delta_1 1_{[c=1]} + \delta_2 1_{[c=2]} \right\}^{d_i}$$

شد، برآورد را برای برخی مسائل آسان می نمود. در ادامه یک اصلاح جدید برای تبدیل مسئله بهینه سازی مقید به بهینه سازی بدون قید که منجر به همگرایی سریع تر نسبت به الگوریتم EM و نیز استواری برآوردها نسبت به مقادیر اولیه متفاوت می شود، ارائه می شود. برای تبدیل مسئله بهینه سازی مقید به یک مسئله بدون قید، از تبدیل نسبت لگاریتمی Isometric Log-Ratio Transform (ILR) که توسط Egozcue et al. (2003) ارائه شد، استفاده می شود. این تبدیل حاوی نگاشت حفظ فاصله از k بعدی ساده به فضای \mathbb{R}^{k-1} بعدی است و بیشتر در تجزیه و تحلیل داده های ترکیبی برای پارامترسازی مجدد بردار $\zeta = (\zeta_1, \dots, \zeta_k)$ استفاده می شود.

$$\pi = ILR(\zeta) = \Psi^T \log(\zeta) \quad (20)$$

که π یک بردار $k-1$ بعدی و Ψ یک ماتریس با ابعاد $(k \times (k-1))$ است که ستون های $(\psi_1, \dots, \psi_{k-1})$ پایه های یکمتمم را برای ابرصفحه از \mathbb{R}^k متعامد به بردار k واحدی ۱ نشان را می دهند که به صورت زیر تعریف می شود:

$$\psi_i = \sqrt{\frac{i}{i+1}} \left[\underbrace{1, \dots, 1}_i, -1, \underbrace{0, \dots, 0}_{k-i-1} \right]. \quad (21)$$

برای به دست آوردن ζ برای هر π موجود، معکوس تبدیل ILR در رابطه ۲۰ به صورت زیر است:

$$\zeta = ILR^{-1}(\pi) = \left[\frac{\exp(\psi_1 \pi)}{\sum_{i=1}^k \exp(\psi_i \pi)}, \dots, \frac{\exp(\psi_k \pi)}{\sum_{i=1}^k \exp(\psi_i \pi)} \right]. \quad (22)$$

حال از تبدیل ILR برای تغییر مسئله بهینه سازی مقید به یک مسئله بدون قید، استفاده می کنیم. لذا الگوریتم بهینه سازی پیشنهادی شامل گام های زیر می شود:

۱. مقادیر اولیه ای برای بردارهای پارامتری $\hat{\tau}^{(0)}$ و $\hat{\zeta}^{(0)}$ در نظر بگیرید.
۲. براساس تبدیل ILR مقدار $\hat{\pi}^{(0)}$ را از روی $\hat{\zeta}^{(0)}$ محاسبه نمایید.

$$\hat{\pi}^{(0)} = ILR(\hat{\zeta}^{(0)}). \quad (23)$$

۳. در تابع لگاریتم درستنمایی J_j را با مولفه های فردی زیر جایگذاری نمایید.

$$\frac{\exp(\psi_j \pi)}{\sum_{i=1}^k \exp(\psi_i \pi)}, \quad j = 1, \dots, k \quad (24)$$

جدول ۴. برآورد و انحراف استاندارد پارامترها در داده کامل و گمشده با سناریو تقسیم‌بندی مختلف

برآورد پارامتر	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\zeta}_5$
$\hat{\theta}$	-۱۱/۴۱	۰/۱۱	-۰/۲۱	-۰/۲۲	-۰/۴۳	۰/۱۹	۰/۰۳	۰/۴۳	۰/۱۳	۰/۱۳
$\hat{\theta}_1$	-۱۱/۴۱	۰/۱۱	-۰/۰۸	-۰/۲۷	-۰/۳۷	۰/۲۱	۰/۰۰	۰/۵۴	۰/۰۲	۰/۰۰
$\hat{\theta}$	-۱۱/۴۲	۰/۱۱	-۰/۱۳	-۰/۲۳	-۰/۳۴	۰/۲۲	۰/۰۰	۰/۵۲	۰/۰۴	۰/۰۱
$\sigma_{PAR} \times 10^2$	۱۷/۴۳	۰/۲۱	۱۰۱/۲۴	۳۳/۲۲	۱۱۴/۹۴	۳/۰۹	۰/۸۷	۱۷/۷۴	۱۷/۴۹	۱۷/۰۸

جدول ۵. مقادیر برآورد $\hat{\phi}$ برای مجموعه داده کامل

مشخصات جغرافیایی	۰	۱	۲
جمعیتی (C)			
سطح مزایا (B)			
پایین	۱	۰/۸۰	۰/۶۵
بالا	۰/۸۱	۰/۶۵	۰/۵۳

محاسبه شده است. مشاهده می‌شود که دو مقدار بزرگ‌تر مقادیر ویژه ماتریس هسین بسیار نزدیک به صفر هستند که نشان دهنده این است که ماتریس اطلاع برآورد شده، رتبه کامل نیست و مدل برای داده‌های موجود، دارای پارامتر افزونه است. علاوه بر این ممکن است یک مسئله در ریسک مدل موجود باشد. ما یک مدل گومپرتز برای داده‌ها استفاده کرده‌ایم که ممکن است نتواند تابع مرگ‌ومیر درستی را برای این تجربه مرگ‌ومیر منعکس کند و یا ممکن است که اختلافات مرگ‌ومیر به دلیل ویژگی‌های مختلف اقتصادی و اجتماعی به‌طور متناسب بر روی لگاریتم خطر تأثیر نگذارد.

دسترسی به اطلاعات بیشتر

مسئله شناسایی‌پذیری زمانی رخ می‌دهد که دو متغیر کمکی که هیچ‌گاه به‌صورت مشترک مشاهده نشده‌اند، مانع از استنباط بیشتر در مورد تابع مرگ‌ومیر می‌شود. اکنون برای بررسی تأثیر دسترسی به مجموعه داده بیشتر حالت‌های مختلفی را مدنظر قرار می‌دهیم. مجموعه داده P_3 اندازه نمونه کوچکتری به تعداد n_3 دارد که نماینده طرح بازنشستگی است که هیچ اطلاعاتی از دو متغیر کمکی، گمشده نیستند. برای نشان دادن استواری نتایج، مجموعه داده P_4 با اندازه نمونه بزرگتر به تعداد n_4 در نظر گرفته می‌شود که نماینده طرح بازنشستگی است که در آن C و هر دو گمشده هستند.

حال چهار طرح بازنشستگی P_1, P_2, P_3, P_4 داریم که از طریق تقسیم‌بندی مجموعه داده‌های کامل به‌صورت تصادفی و حذف عوامل مقتضی به‌دست آورده می‌شوند. در این رویکرد، ۵ درصد به‌طور کامل مشاهده شده‌اند (P_3)، ۲۰ درصد فقط B مشاهده شده (P_1)، ۲۰ درصد فقط C مشاهده شده (P_2) و در ۵۵ درصد باقی‌مانده هر دو B و C مشاهده نشده‌اند (P_4). از این رو $n_1 = n_2 = 3748$ ، $n_3 = 937$ و $n_4 = 10308$ است. این نحوه تقسیم‌بندی تصادفی ۱۰۰۰ بار تکرار می‌شود. بنابراین سهم تابع در دستمایه برای هر فرد در P_3 (بدون مشاهده گمشده) به‌صورت زیر است:

ابتدا تنها یک‌بار سناریو تقسیم‌بندی تصادفی به P_1 و P_2 را انجام می‌دهیم. برآورد پارامترها در این سناریو را با $(\hat{\theta}_1)$ نمایش می‌دهیم. سناریوی دیگر ۱۰۰۰ بار تقسیم کردن مجموعه داده کامل به‌صورت تصادفی است و میانگین برآورد پارامترها $(\hat{\theta} = \frac{1}{1000} \sum_k \theta_k)$ و همچنین انحراف استاندارد پارامترها $(\sigma_{PAR} = \sqrt{\frac{1}{1000} \sum_k (\theta_k - \hat{\theta})^2})$ محاسبه می‌نماییم. برآورد پارامترها برای مجموعه داده کامل را با $(\hat{\theta})$ نمایش می‌دهیم. کلیه نتایج در جدول ۴ ارائه شده است.

برای بررسی دلایل برآورد ضعیف به‌دست آمده از ترکیب P_1 و P_2 ، به بررسی شناسایی‌پذیری پارامترها می‌پردازیم. برای این منظور، براساس تابع مرگ‌ومیر موجود تابع $\hat{\phi}$ را برای مقادیر مختلف (B, C) و با توجه به سطح پایه و سن X به صورت زیر تعریف می‌کنیم:

$$\hat{\phi} = \exp \left[\gamma 1_{[b=High]} + \delta_1 1_{[c=c_1]} + \delta_2 1_{[c=c_2]} \right]$$

جدول ۵ مقادیر $\hat{\phi}$ را برای مقادیر مختلف (B, C) برای مجموعه داده کامل، نشان می‌دهد.

در جدول ۵ مشاهده می‌شود که فردی با مشخصات جغرافیایی جمعیتی «صفر» و سطح مزایای بالا، با فردی که سطح مزایا پایین دارد و مشخصات جغرافیایی جمعیتی او «یک» است، مرگ‌ومیر بسیار مشابهی (در سن‌های یکسان) دارند. این موضوع برای افرادی که مزایای بالا و مشخصات جغرافیایی جمعیتی آن‌ها یک است، با افرادی که دارای مزایای پایین و مشخصات جغرافیایی جمعیتی «دو» هستند، نیز برقرار است. بنابراین شرایط شناسایی‌پذیری، برقرار نمی‌باشد.

همچنین شناسایی‌پذیری پارامترها در مدل مطرح‌شده با داده‌های موجود، می‌تواند از طریق بررسی مقادیر ویژه ماتریس هسین تحلیل شود. مقادیر ویژه برای تمام ۱۰۰۰ مجموعه داده شبیه‌سازی شده

جدول ۶) مقایسه می‌کنیم. خطاهای استاندارد با استفاده از منفی معکوس ماتریس اطلاع تجربی برآورد شده‌اند. پروسه برآورد در بخش ۵، شامل (τ, ζ) است. از این رو برآورد خطای استاندارد برای (τ, ζ) نیازمند این است که از روش دلنا برای توابع برداری استفاده شود. این روش تنها برای برآورد خطاهای استاندارد ζ استفاده می‌شود. تابع مورد استفاده، همان رابطه ۲۲ است. بنابراین مشتق تابع مد نظر نسبت به π ، یک برداری k بعدی به صورت زیر است:

$$\zeta'_j = \frac{\psi_j \exp(\psi_j \pi) \sum_{i=1}^k \exp(\psi_i \pi) - \exp(\psi_j \pi) \sum_{i=1}^k \psi_i \exp(\psi_i \pi)}{\left(\sum_{i=1}^k \exp(\psi_i \pi) \right)^2} \quad (۳۳)$$

که $j = 1, \dots, k$ است. بنابراین، واریانس آن به طور تقریبی از رابطه $var(\zeta) = \zeta' \hat{\Omega} \zeta$ حاصل می‌شود، که در آن Σ ، ماتریس واریانس کوواریانس است که از طریق منفی معکوس ماتریس هسین به دست آمده در برآورد بردار پارامتری π حاصل می‌شود. در نهایت با جذر گرفتن از رابطه فوق، خطای استاندارد مربوط به بردار پارامتری ζ حاصل می‌گردد.

همان‌طور که از مقادیر $\hat{\theta}_k$ و $\bar{\theta}$ مشاهده می‌شود برآورد پارامترها با برآوردهای به دست آمده براساس مجموعه داده کامل، مطابقت دارند. اما، چون بخش‌هایی از مجموعه داده کامل از استنباط خارج می‌شوند، خطای استاندارد برآورده شده پارامترها $\hat{\sigma}(\hat{\theta})$ به دلیل کوچک‌تر بودن اندازه نمونه افزایش می‌یابد.

افزایش خطای استاندارد برآورد شده، برای پارامترهای δ_1 ، γ و δ_2 به این دلیل است که تفاوت‌های مرگومیر بر پایه مزایا و مشخصات جغرافیایی جمعیتی، وابسته به متغیرهایی است که گاهی اوقات گمشده هستند. برای پارامترهای α و β این امر تأثیر کمتری دارد. به این دلیل که متغیر کمکی سن برای تمام افراد مشاهده شده است و خط مبنا برای همه افراد در چهار مجموعه داده طرح‌های بازنشستگی شبیه‌سازی شده، مشترک است. این تفاوت‌ها بیشتر هنگام مقایسه خطای استاندارد برآورد شده $\hat{\delta}_1$ ، $\hat{\gamma}$ و $\hat{\delta}_2$ وقتی که از هر چهار طرح بازنشستگی $(P_1 - P_4)$ با برآورد فقط بر اساس طرح‌های بازنشستگی $P_1 - P_3$ استفاده می‌شود، تأیید می‌شوند.

برای بررسی شناسایی‌پذیری مدل کامل با ۴ مجموعه داده $P_1 - P_4$ ، مشابه قبل می‌توان مقادیر ویژه ماتریس هسین حاصل از برآورد هر چهار مجموعه داده، را محاسبه نمود. مشاهده می‌شود که در تمام ۱۰۰۰ طرح تصادفی داده گمشده، مقادیر ویژه همگی منفی و دور از صفر هستند. همچنین نزدیک‌ترین مقدار ویژه به صفر در تمام شبیه‌سازی‌ها برابر با $-۹/۹۳$ است. بنابراین مدل بررسی شده، شناسایی‌پذیر است.

بررسی تأثیر مالی در استنباط آمار براساس مشاهدات گمشده

بررسی نرخ مرگومیر و عامل مستمری

فرض کنید متغیر تصادفی Y_x نشان دهنده ارزش فعلی جریان‌های نقدی در یک صندوق بازنشستگی برای عضو x ساله

$$L_i(\tau, \zeta) \propto S_{x_i}(t_i | b_i, c_i; \tau) \mu_{x_i+t_i}^{d_i}(b_i, c_i; \tau) f_{B,C}(b_i, c_i; \zeta). \quad (۲۹)$$

لذا:

$$L_i(\tau, \zeta) \propto (\zeta_1 + \zeta_2 + \zeta_3 + \zeta_4 + \zeta_5 + \zeta_6) \times \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma 1_{[b_i=High]} + \delta_1 1_{[c_i=1]} + \delta_2 1_{[c_i=2]}) \right\} \times \exp \left\{ \alpha + \beta(x_i + t_i) + \gamma 1_{[b_i=High]} + \delta_1 1_{[c_i=1]} + \delta_2 1_{[c_i=2]} \right\}^{d_i}. \quad (۳۰)$$

و سهم تابع درستنمایی برای افراد در P_4 (هر دو B و C گمشده) به صورت زیر است:

$$L_i(\tau, \zeta) \propto \sum_{b \in \{Low, High\}} \sum_{c \in \{0,1,2\}} S_{x_i}(t_i | b, c; \tau) \mu_{x_i+t_i}^{d_i}(b, c; \tau) f_{B,C}(b, c; \zeta). \quad (۳۱)$$

در نتیجه داریم:

$$L_i(\tau, \zeta) \propto \zeta_1 \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i) \right\} \exp \{ \alpha + \beta(x_i + t_i) \}^{d_i} \quad (۳۲) \\ + \zeta_2 \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma) \right\} \exp \{ \alpha + \beta(x_i + t_i) + \gamma \}^{d_i} \\ + \zeta_3 \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \delta_1) \right\} \exp \{ \alpha + \beta(x_i + t_i) + \delta_1 \}^{d_i} \\ + \zeta_4 \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma + \delta_1) \right\} \exp \{ \alpha + \beta(x_i + t_i) + \gamma + \delta_1 \}^{d_i} \\ + \zeta_5 \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \delta_2) \right\} \exp \{ \alpha + \beta(x_i + t_i) + \delta_2 \}^{d_i} \\ + \zeta_6 \exp \left\{ - \left(\frac{\exp(\beta t_i) - 1}{\beta} \right) \exp(\alpha + \beta x_i + \gamma + \delta_2) \right\} \exp \{ \alpha + \beta(x_i + t_i) + \gamma + \delta_2 \}^{d_i}.$$

در ادامه، مقادیر برآورد پارامترها و خطاهای استاندارد به دست آمده از مجموعه داده‌های مختلف را با هم مقایسه می‌کنیم. برآوردها بر اساس ترکیب چهار مجموعه داده $(P_1 - P_4)$ مشخص شده، براساس ترکیب P_1 ، P_2 و P_3 $(P_1 - P_3)$ مشخص شده، یعنی به غیر از مجموعه داده بزرگتر که هر دو B و C مشاهده نشده‌اند؛ و تنها برای P_3 (بدون داده گمشده ولی اندازه نمونه خیلی کوچک) را در نظر می‌گیریم.

برآورد پارامترها در مجموعه داده کامل $(\hat{\theta})$ ، برآورد پارامترها در یک بار تقسیم مجموعه داده کامل به صورت تصادفی $(\hat{\theta}_1)$ برآورد خطاهای استاندارد آن‌ها $(\hat{\sigma}(\hat{\theta}_1) \times 10^3)$ ، میانگین برآورد پارامترها به وسیله ۱۰۰۰ بار تقسیم کردن مجموعه داده‌های کامل به صورت تصادفی $(\hat{\theta} = \frac{1}{1000} \sum_k \hat{\theta}_k)$ ، میانگین برآورد خطای استاندارد و انحراف استاندارد پارامترها به ترتیب $(\hat{\sigma}(\hat{\theta}) = \frac{1}{999} \sum_k \hat{\sigma}(\hat{\theta}_k) \times 1000)$ و $\sigma_{PAR} = \sqrt{\frac{1}{999} \sum_k (\hat{\theta}_k - \hat{\theta})^2} \times 1000$ نمایش داده شده است.

این برآوردها را با نتایج حاصل از مجموعه داده کامل (خط اول

جدول ۶. برآورد پارامترها و برآورد خطای استاندارد مربوطه در سناریوهای مختلف

مجموعه داده	برآورد	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\zeta}_5$
کامل	$\hat{\theta}$	-۱۱/۴۱	۰/۱۱	-۰/۲۱	-۰/۲۲	-۰/۴۳	۰/۱۹	۰/۰۳	۰/۴۳	۰/۱۳	۰/۱۳
	$\hat{\theta}_1$	-۱۱/۴۶	۰/۱۱	-۰/۰۸	-۰/۱۶	-۰/۳۹	۰/۱۹	۰/۰۲	۰/۴۳	۰/۱۴	۰/۱۳
	$\hat{\sigma}(\hat{\theta}_1) \times 10^3$	۱۳۶/۲۶	۱/۷۱	۶۷/۹۶	۶۳/۱۶	۸۴/۷۱	۶/۶۹	۴/۰۸	۹/۴۴	۷/۹۴	۷/۹۰
	$\hat{\bar{\theta}}$	-۱۱/۴۵	۰/۱۱	-۰/۱۸	-۰/۱۹	-۰/۳۶	۰/۱۹	۰/۰۳	۰/۴۳	۰/۱۳	۰/۱۳
	$\hat{\bar{\sigma}}(\hat{\theta}) \times 10^3$	۱۳۶/۴۲	۱/۷۲	۶۸/۴۱	۶۱/۶۵	۸۲/۱۶	۶/۹۴	۴/۶۸	۹/۴۴	۷/۹۴	۷/۶۶
$P_1 - P_4$	$\sigma_{PAR} \times 10^3$	۲۵/۴۹	۰/۲۹	۵۰/۹۷	۴۸/۵۸	۶۳/۲۹	۶/۱۴	۴/۴۸	۸/۵۶	۷/۳۰	۷/۴۷
	$\hat{\theta}_1$	-۱۱/۵۲	۰/۱۱	-۰/۱۰	-۰/۱۹	-۰/۴۷	۰/۱۹	۰/۰۲	۰/۴۳	۰/۱۴	۰/۱۳
	$\hat{\sigma}(\hat{\theta}_1) \times 10^3$	۱۹۹/۲۱	۲/۵۲	۷۷/۰۹	۶۹/۰۱	۹۴/۲۹	۶/۶۹	۴/۰۹	۹/۴۴	۷/۹۵	۷/۹۱
	$\hat{\bar{\theta}}$	-۱۱/۴۳	۰/۱۱	-۰/۲۰	-۰/۲۱	-۰/۴۱	۰/۱۹	۰/۰۳	۰/۴۳	۰/۱۳	۰/۱۳
	$\hat{\bar{\sigma}}(\hat{\theta}) \times 10^3$	۲۰۰/۶۱	۲/۵۴	۷۴/۰۵	۶۶/۴۶	۸۸/۵۵	۶/۹۵	۴/۶۹	۹/۴۵	۷/۹۵	۷/۶۷
$P_1 - P_3$	$\sigma_{PAR} \times 10^3$	۱۵۴/۳۲	۱/۹۷	۵۹/۷۳	۵۴/۹۷	۷۱/۳۹	۶/۴۲	۴/۵۷	۸/۴۵	۷/۳۸	۷/۳۷
	$\hat{\theta}_1$	-۱۱/۵۴	۰/۱۱	-۰/۱۶	-۰/۱۳	-۰/۶۶	۰/۱۹	۰/۰۲	۰/۴۵	۰/۱۴	۰/۱۱
	$\hat{\sigma}(\hat{\theta}_1) \times 10^3$	۶۱۱/۵۲	۷/۷۴	۱۷۳/۳۷	۱۵۰/۹۲	۲۲۷/۵۲	۱۲/۹۰	۴/۳۶	۱۶/۲۵	۱۱/۴۴	۱۰/۳۵
	$\hat{\bar{\theta}}$	-۱۱/۵۱	۰/۱۱	-۰/۲۱	-۰/۲۲	-۰/۴۴	۰/۱۹	۰/۰۳	۰/۴۳	۰/۱۳	۰/۱۳
	$\hat{\bar{\sigma}}(\hat{\theta}) \times 10^3$	۶۰۴/۹۷	۷/۶۶	۱۶۶/۵۷	۱۵۲/۲۵	۱۹۹/۰۵	۱۲/۸۰	۵/۱۴	۱۶/۱۵	۱۱/۱۰	۱۰/۹۸
P_3	$\sigma_{PAR} \times 10^3$	۶۳۴/۰۷	۸/۱۳	۱۵۲/۹۵	۱۴۶/۰۲	۱۹۰/۲۱	۱۲/۶۹	۵/۰۱	۱۵/۹۴	۱۰/۸۶	۱۰/۷۰

و ۷۱ سال انجام شده است. اولین سن، همان سن معمول بازنشستگی و دومین سن به طور تقریبی، میانگین سن افراد زنده در مجموعه داده کامل در زمان اتمام دوره مشاهده است.

نتایج جدول نشان می‌دهد که برای سن‌های انتخابی و دو نرخ بهره محاسبه شده، بین عامل‌های مستمری به دست آمده از مدل مبتنی بر متغیرهای کمکی اقتصادی-اجتماعی و مقادیر به دست آمده از مدل M_0 تفاوت‌های قابل توجهی وجود دارد. این اختلافات برای نرخ بهره پایین‌تر، بیشتر است که بیانگر این است که وقتی نرخ بهره پایین است، مرگومیر عامل خطر مهم‌تری برای مستمری است. برای بررسی تأثیر مالی استفاده از مدل مرگومیر معرفی شده، می‌توان الزام سرمایه را بررسی نمود که نتیجه آن وابسته به عامل مستمری و در نتیجه مدل مرگومیر است.

ریسک برآورد نادرست الزام سرمایه

هدف این بخش نحوه کاهش الزام سرمایه براساس طرح بازنشستگی موجود، و ترکیب طرح‌های بازنشستگی با دیگر با هدف تخمین پارامترهای نرخ مرگومیر است.

با استفاده از رویکرد Richards (2016) برای محاسبه SCR (The Solvency Capital Requirement) ، بر اساس

است که در طول عمر باقی مانده خود آن را دریافت می‌کند. اگر این جریان نقدی به میزان یک واحد هر ساله به طور پیوسته پرداخت شود، مقدار مورد انتظار برای Y_x ، عامل مستمری نامیده شده که با \bar{a}_x نمایش و به صورت زیر محاسبه می‌شود (Dickson et al., 2009):

$$\bar{a}_x = \mathbb{E}(Y_x) = \int_0^{120-x} \exp(-rt) S_x(t) dt. \quad (34)$$

به دلیل گمشدگی داده‌ها، \hat{t} به \hat{t} وابسته است، توزیع نمونه‌گیری به صورت $\hat{\theta} = (\hat{t}, \hat{\zeta})$ مجانبی دارای توزیع نرمال چندمتغیره (Multivariate Normal Distribution) با میانگین θ و ماتریس واریانس کوواریانس Σ است، که از طریق منفی معکوس ماتریس اطلاع فیشر (Fisher Information Matrix) برآورد می‌شود.

برای بررسی نرخ مرگومیر و عامل مستمری، مقادیر عامل مستمری و درصد تغییر آن در مقایسه با عامل مستمری در مدل M_0 محاسبه و نتایج در جدول ۷ ارائه شده است. برآورد پارامترها با مشاهدات هر چهار طرح $P_1 - P_4$ در نظر گرفته می‌شوند. تجزیه و تحلیل عددی برای دو نرخ بهره ۱ و ۳ درصد و دو فرد به سن‌های ۶۵

جدول ۷. مقادیر مستمری، درصد تغییرات بر اساس مدل M_0 و نرخ‌های بهره مختلف

سن	مجموعه داده	تقسیم‌بندی	$r = 1\%$		$r = 3\%$	
			\bar{a}_x	درصد تغییر	\bar{a}_x	درصد تغییر
۶۵	$P_1 - P_4$	M_0	۱۶/۷۷	-	۱۳/۷۰	-
		$b = L, c = 0$	۱۵/۶۱	-۶/۹۳	۱۲/۹۰	-۵/۸۱
		$b = H, c = 0$	۱۶/۱۱	-۳/۹۳	۱۳/۲۵	-۳/۲۶
		$b = L, c = 1$	۱۶/۵۷	-۱/۲۱	۱۳/۵۶	-۰/۹۸
		$b = H, c = 1$	۱۷/۰۸	۱/۸۴	۱۳/۹۱	۱/۵۴
		$b = L, c = 2$	۱۷/۹۵	۷/۰۱	۱۴/۴۹	۵/۷۷
		$b = H, c = 2$	۱۸/۴۷	۱۰/۱۰	۱۴/۶۱	۶/۶۶
۷۱	$P_1 - P_4$	M_0	۱۲/۹۸	-	۱۱/۰۲	-
		$b = L, c = 0$	۱۱/۹۰	-۸/۳۳	۱۰/۲۲	-۷/۲۸
		$b = H, c = 0$	۱۲/۳۶	-۴/۷۸	۱۰/۵۷	-۴/۱۵
		$b = L, c = 1$	۱۲/۷۸	-۱/۵۴	۱۰/۸۸	-۱/۳۲
		$b = H, c = 1$	۱۳/۲۶	۲/۱۰	۱۱/۲۲	۱/۸۳
		$b = L, c = 2$	۱۴/۰۷	۸/۳۴	۱۱/۸۱	۷/۱۶
		$b = H, c = 2$	۱۴/۵۶	۱۲/۱۱	۱۱/۹۴	۸/۲۹

جدول ۸. الزام سرمایه برای طرح بازنشستگی P_1 و P_3 بر اساس دو نرخ بهره فرضی و طرح‌های مورد استفاده برای برآورد نرخ مرگومیر

نرخ بهره و الزام سرمایه					
نرخ بازنشستگی P_1					
نرخ بهره		$P_1 - P_3$		$P_1 - P_4$	
۳٪	۱٪	۳٪	۱٪	۳٪	۱٪
۲٪/۲۷	۲٪/۸۲	۱٪/۶۲	۱٪/۹۵	۱٪/۰۹	۱٪/۳۰
نرخ بازنشستگی P_3					
نرخ بهره		$P_1 - P_3$		$P_1 - P_4$	
۳٪	۱٪	۳٪	۱٪	۳٪	۱٪
۴٪/۶۱	۵٪/۷۹	۱٪/۵۴	۱٪/۸۸	۱٪/۰۳	۱٪/۲۶

که در این رابطه، w_i نشان‌دهنده مبلغ مستمری شخص i ام است. بنابراین ریسک برآورد نادرست الزام سرمایه را به‌عنوان مثال برای طرح بازنشستگی P_1 به سه روش محاسبه می‌کنیم:

✓ پارامتر تابع مرگومیر τ فقط با استفاده از داده‌های موجود در P_1 برآورد می‌شود. به‌منظور ساده نگه داشتن کارها و جلوگیری از مسئله شناسایی‌پذیری به‌دلیل نبود C ، تابع مرگومیر مانند مدل M_1 است که پارامترهای آن را می‌توان با حداکثر درست‌نمایی تخمین زد.

✓ تمام پارامترهای تابع مرگومیر با استفاده از ترکیب طرح‌های $P_1 - P_3$ ، از طریق رویکرد داده‌های گمشده، برآورد می‌شوند.
 ✓ تمام پارامترهای تابع مرگومیر با استفاده از ترکیب طرح‌های $P_1 - P_4$ برآورد می‌شوند.

این کار را برای طرح بازنشستگی P_3 تکرار می‌کنیم که در این طرح هر دو متغیر کمکی موجود است. الزامات سرمایه برای ریسک برآورد نادرست براساس این دو رویکرد و دو نرخ بهره، در جدول ۸ نشان داده شده‌است.

مجموعه بزرگی از مقادیر پارامتر جایگزین، ارزیابی‌های مکرری از کل پرتفوی مستمری انجام می‌دهیم. مقدار پرتفوی به‌عنوان تابعی از پارامتر θ با $P(\theta)$ نشان داده می‌شود. الزام سرمایه ناشی از ریسک پارامترها با برآورد نادرست، به شرح زیر محاسبه می‌شود:

$$\left(\frac{\text{چندک } 99/5 \text{ تابع } P(\theta)}{\text{میانگین } P(\theta)} - 1 \right) \times 100. \quad (35)$$

برای یافتن برآورد صدک و میانگین $P(\theta)$ ، مقدار ۱۰۰۰۰۰ شبیه‌سازی شده برای θ محاسبه می‌شود. مقادیر شبیه‌سازی شده را با θ' نمایش داده و k امین مقدار شبیه‌سازی شده برای پرتفوی $(P(\theta'))$ به‌صورت زیر محاسبه می‌شود:

$$P(\theta'_k) = \sum_{i=1}^n w_i \bar{a}_{x_i}(\theta'_k). \quad (36)$$

مشارکت نویسندگان

نویسنده اول نظارت بر کل فرآیند مقاله نویسی و بررسی کیفیت تحلیل داده‌ها و اجراء اصلاح و جمع‌بندی، مشاوره و نظارت بر روش‌شناسی تحقیق را داشته است. نویسنده دوم نیز جمع‌آوری مطالب مرتبط و تدوین مدل، مروری بر ادبیات پژوهش، روش‌شناسی و متدولوژی-نگارش اولیه، تحلیل و پیاده‌سازی را انجام داده است.

تشکر و قدردانی

نویسندگان مقاله از پیشنهادات مفید سردبیر و داوران محترم نشریه پژوهشنامه بیمه که در بهبود نتایج این پژوهش کمک شایانی کردند، تشکر می‌نمایند.

تعارض منافع

نویسندگان اعلام می‌دارند که در مورد انتشار این مقاله تضاد منافع وجود ندارد. علاوه بر این، موضوعات اخلاقی شامل سرقت ادبی، رضایت آگاهانه، سوءرفتار، جعل داده‌ها، انتشار و ارسال مجدد و مکرر توسط نویسندگان رعایت شده است.

دسترسی آزاد

کپی‌رایت نویسنده(ها) ©2021: این مقاله تحت مجوز بین‌المللی Creative Commons Attribution 4.0 اجازه استفاده، اشتراک‌گذاری، اقتباس، توزیع و تکثیر را در هر رسانه یا قالبی مشروط به درج نحوه دقیق دسترسی به مجوز CC منوط به ذکر تغییرات احتمالی بر روی مقاله می‌باشد. لذا به استناد مجوز مذکور، درج هرگونه تغییرات در تصاویر، منابع و ارجاعات یا سایر مطالب از اشخاص ثالث در این مقاله باید در این مجوز گنجانده شود، مگر اینکه در راستای اعتبار مقاله به اشکال دیگری مشخص شده باشد. در صورت عدم درج مطالب مذکور و یا استفاده فراتر از مجوز فوق، نویسنده ملزم به دریافت مجوز حق نسخه‌برداری از شخص ثالث می‌باشد.

به منظور مشاهده مجوز بین‌المللی Creative Commons Attribution 4.0 به آدرس زیر مراجعه گردد:
<http://creativecommons.org/licenses/by/4.0>

یادداشت ناشر

ناشر نشریه پژوهشنامه بیمه با توجه به مرزهای حقوقی در نقشه‌های منتشرشده بی‌طرف باقی می‌ماند.

منابع

Catchpole, E.A.; Morgan, B.J.T., (1997). Detecting parameter redundancy. *Biometrika*, 84(1): 187–196 (9 Pages).
Chen, Q.; May, R.C.; Ibrahim, J.G.; Chu, H.; Cole, S.R., (2014). Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates. *Stat.*

از نتایج این جدول مشاهده می‌شود که در نظر گرفتن سایر مجموعه داده‌ها برای برآورد نرخ مرگ‌ومیر، می‌تواند الزام سرمایه را کاهش دهد. در نظر گرفتن مشاهدات برای وقتی که در آن‌ها بعضی از متغیرهای کمکی یا تمامی آن‌ها گمشده هستند، اندازه نمونه را افزایش داده و سبب کاهش عدم حتمیت پارامتر می‌شود. این نتایج نشان می‌دهند که در نظر گرفتن نمونه‌هایی با مشاهدات گمشده، می‌تواند تغییرپذیری برآورد نمونه‌گیری را کاهش دهد و به نوبه خود سبب کاهش الزام سرمایه شود. و یکی از راه‌های به دست آوردن چنین مشاهدات اضافی، ترکیب داده‌های تجربی از طرح‌های مختلف بازنشتگی است با این فرض که اعضای آن طرح‌ها از قانون مرگ‌ومیر یکسانی پیروی کنند.

جمع بندی و پیشنهادها

در این پژوهش به ترکیب داده‌های حاصل از طرح‌های بازنشتگی مختلف، با فرض اینکه قانون احتمال یکسانی برای طول عمر آتی دارند، اما اطلاعات متغیر کمکی متفاوت است، پرداخته شد. اگر متغیرهای کمکی حتی در یک تجربه مرگ‌ومیر گمشده باشد، می‌توان از همین روش‌ها استفاده کرد.

همچنین به دست آوردن داده‌های کامل برای زیرمجموعه کوچکی از اعضا به ما این امکان را می‌دهد داده‌های حاصل از دو یا چند تجربه را با هم تلفیق و از موانع شناسایی‌پذیری جلوگیری کنیم. این نتیجه از لحاظ عملی اهمیت دارد، زیرا به دست آوردن اطلاعات کامل برای همه اعضای صندوق‌های بازنشتگی مختلف ممکن است دشوار یا گران باشد. روش‌های مطرح‌شده می‌تواند برای محاسبه کمیت‌های مالی مورد علاقه براساس عامل‌های مستمری، برای بیم‌سنگ‌ها مفید باشد. این روش مجموعه داده‌های مختلف با تجربه مرگ‌ومیر برابر یا مشابه را با هم ترکیب، اندازه نمونه را افزایش و ریسک پارامتر را کاهش می‌دهد، بنابراین منجر به کاهش الزام سرمایه می‌شوند. متغیرهای اقتصادی-اجتماعی از جمله سطح مزایا و مشخصات جغرافیایی جمعیتی در صورت پایین بودن نرخ بهره بیشتر مورد توجه قرار می‌گیرند.

بررسی این‌که آیا می‌توان بدون کنار گذاشتن داده‌های مربوطه در وضعیت پارامتر افزونه، استنباط آماری کرد، مساله مورد بررسی قرار نگرفته است و این مسئله می‌تواند در چارچوب استنباط بیزی بهتر حل شود، زیرا هنگامی که ماتریس اطلاع فشر معین مثبت اکید نیست، برآوردگر ماکسیمم درست‌نمایی، توزیع نرمال مجانبی ندارد (Watanabe and Opper, 2010). همچنین یکی از مفروضات مدل‌بندی این بود که اختلافات مرگ‌ومیر بین گروه‌های اقتصادی-اجتماعی با افزایش سن تغییر نمی‌کند. لذا تأثیر و تجزیه و تحلیل یک عبارت اثر متقابل می‌تواند در بررسی‌های آتی مدنظر قرار گیرد.

Med., 33(26): 4560–4576 (16 Pages).
Dempster, A.P.; Laird, N.M.; Rubin D.B., (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.*, 39(1): 1–38 (37 Pages).
Dickson, D.; Hardy, M.; Waters, H., (2009). *Actuarial math-*

- ematics for life contingent risks. international series on actuarial science. Cambridge University Press.
- Egozcue, J.J.; Pawlowsky-Glahn, V.; Mateu-Figueras, G.; Barceló-Vidal, C., (2003). Isometric logratio transformations for compositional data analysis. *Math. Geol.*, 35(3): 279–300 **(21 Pages)**.
- Gompertz, B., (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. Royal Soc. London*, 115(14): 513–583 **(70 Pages)**.
- Herring, A.H.; Ibrahim, J.G., (2001). Likelihood-Based methods for missing covariates in the cox proportional hazards model. *J. Am. Stat. Assoc.*, 96(453): 292–302 **(10 Pages)**.
- Komijani, A.; Koosheshi, M.; Niakan, L., (2013). Iran mortality rates using Lee-Carter model: Estimation and forecasting. *Iran. J. Insur. Res.*, 2(4): 1-25 **(24 Pages)**. [In Persian]
- Lin, X.S.; Liu, X., (2007). Markov aging process and Phase-Type law of mortality. *North Am. Actuarial J.*, 11(4): 92–109 **(17 Pages)**.
- Little, R.; An, H., (2004). Robust Likelihood-Based analysis of multivariate data with missing values. *Stat. Sin.*, 14(3): 949–968 **(19 Pages)**.
- Lord, F.M., (1955). Estimation of parameters from incomplete data. *J. Am. Stat. Assoc.*, 50(271): 870–876 **(6 Pages)**.
- Madrigal, A.M.; Matthews, F.E.; Patel, D.; Gaches, A.; Baxter, S., (2011). What longevity predictors should be allowed for when valuing pension scheme liabilities? *Br. Actuarial J.*, 16(1): 1–38 **(37 Pages)**.
- Macdonald, A.S., Richards, S.J.; Currie, I.D., (2018). Modelling mortality with actuarial applications. International Series on Actuarial Science. Cambridge University Press.
- Mahdavi, G.; Daghighi Asli, A.; Lotfi, N., (2011). Application of a mortality model with several risk factors in the termination of life insurance contracts (Case study: An insurance company). *Insur. Res. J. (Insur. Ind.)*, 26(3): 1-27 **(26 Pages)**. [In Persian]
- McLachlan, G.; Peel, D., (2000). Finite mixture models. Wiley series in probability and statistics, New York.
- Richards, S.J., (2016). Mis-Estimation risk: Measurement and impact. *Br. Actuarial J.*, 21(3): 429–457 **(28 Pages)**.
- Rubin, D.B., (1976). Inference and missing data. *Biometrika*, 63(3): 581–592 **(31 Pages)**.
- Schluchter, M.D.; Jackson, K.L., (1989). Log-Linear analysis of censored survival data with partially observed covariates. *J. Am. Stat. Assoc.*, 84(405): 42–52 **(10 Pages)**.
- Shojaee Azar, Z.; Hassan Zadeh, A., (2013). The use of phase-type distributions in mortality modeling. *Iran. J. Insur. Res.*, 3(1): 105-126 **(21 Pages)**. [In Persian]
- Titterton, D.M.; Smith, A.F.M.; Makov, U.E., (1985). Statistical analysis of finite mixture distributions. New York: Wiley.
- Tsiatis, A., (2006). Semiparametric theory and missing data. Springer Science and Business Media.
- Ungolo, F.; Christiansen, M.C.; Kleinow, T.; MacDonald, A.S., (2019). Survival analysis of pension scheme mortality when data are missing. *Scand. Actuarial J.*, 6(10): 523–547 **(24 Pages)**.
- Watanabe, S.; Opper, M., (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.*, 11(12): 3571-3594 **(23 Pages)**.
- Wilks, S.S., (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Ann. Math. Stat.*, 3(3): 163–195 **(32 Pages)**.
- Xu, Y.; Kim, J.K.; Li, Y., (2017). Semiparametric estimation for measurement error models with validation data. *Can. J. Stat.*, 45(2): 185–201 **(16 Pages)**.
- Yashin, A., (2001). Mortality models incorporating theoretical concepts of ageing. In *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective*. Dordrecht: Springer Netherlands.
- Zokaie, M.; Maghsoodi, M., (2010). Reconstruction of frailty-based mortality models using Gompertz distribution generalization. *Insur. Res. (Insur. Ind.)*, 25(4): 59-85 **(26 Pages)**. [In Persian]

AUTHOR(S) BIOSKETCHES	معرفی نویسندگان
<p>Shirin Shoaei, Associate Professor of Actuarial Science, Faculty of Mathematics, Shahrood University of Technology, Shahrood, Iran.</p> <ul style="list-style-type: none"> ▪ Email: sh_shoaei@sbu.ac.ir ▪ ORCID: 0000-0003-4755-0117 ▪ Homepage: http://scimet.sbu.ac.ir/Shirin_Shoaei 	
<p>Reyhane Fathi, Assistant Professor of Actuarial Science, Faculty of Mathematics, Shahrood University of Technology, Shahrood, Iran.</p> <ul style="list-style-type: none"> ▪ Email: reyhane.fathi1@gmail.com ▪ ORCID: 0009-0004-6350-4879 ▪ Homepage: https://mathsci.sbu.ac.ir 	

HOW TO CITE THIS ARTICLE	QR Code
<p>Shoaei, S.; Fathi, R., (2021). Analysis of mortality models with covariates missing at random. <i>Iran. J. Insur. Res.</i>, 10(3): 169-184.</p> <p>DOI: 10.22056/ijir.2021.03.02</p> <p>URL: https://ijir.irc.ac.ir/article_134709.html?lang=en</p>	