
# Insurance Data Science Conference 17 - 18 June 2024

Programme and Abstract Booklet

Scientific Committee

2024-05-14

## Contents

# Conference sponsors

## Gold sponsors

**Advisense** is a leading governance, risk and compliance powerhouse, offering best-in-class services and tech solutions to the European industry.

## Silver sponsors

**Mirai Solutions**: Smarter analytics - better decisions

**Swedbank**: Swedbank Försäkring.

**Swiss Re**: We make the world more resilient.

# Scientific Committee

| Academic | Industry |
|---|---|
| Arthur Charpentier (Université du Québec à Montréal) | Luca Baldassarre (Swiss Re, Zurich) |
| Boualem Djehiche (co-organiser, KTH Royal Institute of Technology) | Markus Gesmann (Insurance Capital Markets Research, London) |
| Montserrat Guillén (Universitat de Barcelona) | Erik Gustafsson (Advisense) |
| Mathias Lindholm (co-organiser, Stockholm University) | Mårten Marcus (Swedbank) |
| Filip Lindskog (co-organiser, Stockholm University) | Davide de March (Markel, London) |
| Mike Ludkovski (UC Santa Barbara) | Grainne McGuire (Optum) |
| Pietro Millossovich (Bayes Business School, City, University of London) | Wui Hua (John) Ng (Reinsurance Group of America) |
| Silvana Pesenti (University of Toronto) | Ronald Richman (Old Mutual Insure, South Africa) |
| Salvatore Scognamiglio (University of Naples) | Markus Senn (Partner Re, Zurich) |
| Andreas Tsanakas (Bayes Business School, City, University of London) | Jürg Schelldorfer (Swiss Re, Zurich) |
| Rui Zhu (Bayes Business School, City, University of London) | Giorgio Alfredo Spedicato (Unipol Group, Milan) |
| Johanna Ziegel (University of Bern) | |

## Keynotes

- **Katrien Antonio** (Professor, KU Leuven & University of Amsterdam)
- **Björn Dalemo** (CEO, Länsförsäkringar Alliance)
- **Mario V. Wüthrich** (Professor, ETH Zürich)

# Programme

## Venue

- Stockholm University, Department of Mathematics, Campus Albano, House 1, Stockholm

## 17 June 2024

**08:00 - 09:00 Registration**

**09:00 - 09:15 Room Hörsal 1: Opening remarks**

**09:15 - 10:15 Room Hörsal 1: Keynote 1 (Chair: Mathias Lindholm)**

- **Björn Dalemo** (Länsförsäkringar Alliance): Trust and stand by algorithmic decisions - a management perspective

**10:15 - 11:15 Regular Session 1**

**Room Hörsal 1: Generative AI in insurance 1 (Chair: Jürg Schelldorfer)**

- **Dylan Liew[a] and Malgorzata Smietanka[a,b]** ([a]Institute and Faculty of Actuaries Federated Learning Working Party, [b]UCL): Hush hush: Keeping neural network claims modelling private, secret, and distributed using federated learning
- **Yves-Cédric Bauwelinckx** (KU Leuven): Generating individual claims using generative adversarial networks
- **Louis DOUGE** (SwissRe): Gen AI: Developing and deploying a specialized underwriting AI assistant

**Room Lärosal 5: Machine learning & climate modelling (Chair: Filip Lindskog)**

- **Ronald Richman** (Old Mutual Insure, University of the Witwatersrand ): Claims modelling with climate data
- **Claudio Giancaterino** (Intesa San Paolo Vita): Boost Climate Risk Modelling with the help of Large Language Models
- **Efren Hernandez** (Management Solutions): Quantification of climate risk in Insurance

**11:15 - 11:40 Coffee break**

**11:40 - 12:40 Lightning Session 1**

**Room Hörsal 1: Stream 1 (Chair: Markus Gesmann)**

- **Alan Muro** (Swiss Re Corporate Solutions): Enabling business steering and decision-making through volatility modelling
- **Finn-Erik Langeggen** (Advisense): Monitoring and modelling ESG risks in insurance
- **Jakob Dambon** (Swiss Re): Advancements in embedded insurance pricing models for cargo insurance
- **Despoina Makariou** (University of St Gallen): A causal machine learning approach for estimating heterogeneous treatment effects in the primary catastrophe bond market
- **Daniel Knös** (Guy Carpenter): Cloudburst hazard analysis – a GIS approach
- **Rodrigo Targino** (Getulio Vargas Foundation): Challenges in actuarial learning for loss modeling of brazilian soybean crops

**Room Lärosal 5: Stream 2 (Chair: Rui Zhu)**

- **Jimmy Hollén** (The Swedish Financial Supervisory Authority): Machine learning and supervision
- **Erik Gustafsson** (Advisense): Challenges and opportunities when implementing AI/ML in pricing and reserving
- **Eivind Borg** (Analytika): Leveraging AI for DORA compliance assessment in insurance
- **Javier Calvo** (Management Solutions): The impact of artificial intelligence on financial institutions
- **Sven Haadem** (Aeda): Enabling advanced analytic capabilities by multiplying internal insurance data
- **Emil Bustos** (Research Institute of Industrial Economics): How well do firms recover from idiosyncratic shocks? Evidence from insurance claims

**12:40 - 13:40 Lunch**

**13:40 - 14:40 Regular Session 2**

**Room Hörsal 1: Machine learning & fairness (Chair: Mathias Lindholm)**

- **Arthur Charpentier** (UQAM, Montréal): Obtaining fair insurance premiums with multiple sensitive attributes
- **François Hu** (Milliman Paris): Ensuring algorithmic fairness in insurance pricing: A multi-class problem perspective
- **Tessa Steensgaard** (University of Copenhagen): Fair learning and testing for unfairness given protected features

**Room Lärosal 5: Machine learning & reserving 1 (Chair: Martin Bladt)**

- **Grainne McGuire** (Optum): Measuring loss reserving uncertainty with machine learning models
- **Emil Hofman** (University of Copenhagen / Alm Brand Group): A machine learning approach based on survival analysis for IBNR frequencies in non-life reserving
- **Oliver Lunding Sandqvist** (University of Copenhagen & PFA Pension): Estimation subject to reporting delays and incomplete event adjudication with applications to disability insurance

**14:40 - 15:40 Panel discussion (Room Hörsal 1)**

**Generative AI in insurance (Chair: Ronald Richman)**

**Panel**

- **Amelié Breitburd** (CNP Assurances)
- **Louis Douge** (Swiss Re)
- **Grainne McGuire** (Optum)
- **Per Jakobsson** (Finansinspektionen)

**15:40 - 16:00 Coffee break**

**16:00 - 17:00 Regular Session 3**

**Room Hörsal 1: Machine learning & pricing 1 (Chair: Mario V. Wüthrich)**

- **Łukasz Delong** (University of Warsaw): Isotonic regression for variance estimation and its role in mean estimation and model validation
- **Anne van der Scheer** (Perunum Actuarieel Advies): Credibility in network regression with sigma-hot encoding and weight balancing
- **Henning Zakrisson** (Stockholm University): A tree-based varying coefficient model

**Room Lärosal 5: Machine learning & life insurance (Chair: Salvatore Scognamiglio)**

- **Luca De Mori** (Bayes Business School): Mortality forecasting via multi-task neural networks
- **Gayani Thalagoda** (University of New South Wales): Variable annuity portfolio valuation with Shapley additive explanations

- **Francesco Ungolo** (University of New South Wales): A Dirichlet process mixture regression model for the analysis of competing risk events

**17:00 - 18:00 Keynote 2 (Room Hörsal 1) (Chair: Arthur Charpentier)**

- **Katrien Antonio** (KU Leuven & University of Amsterdam): TBA

**19:00 Conference dinner**

- **Stallmästaregården**, Stallmästaregården, Norrtull, 113 47 Stockholm

## 18 June 2024

**09:00 - 10:00 Room Hörsal 1: Keynote 3 (Chair: Filip Lindskog)**

- **Mario V. Wüthrich** (ETH Zürich): Experience rating

**10:00 - 11:00 Regular Session 4**

**Room Hörsal 1: Generative AI and underwriting (Chair: Erik Gustafsson)**

- **Benedikt Herwerth** (Swiss Re): LLM-Powered information extraction for claim documents
- **Jayant Apte** (Scor, Charlotte): Accelerated underwriting and underwriting with partial information
- **Per Wilhelmsson** (Länsförsäkringar): Unsupervised learning for efficient underwriting

**Room Lärosal 5: Machine learning & financial aspects of insurance (Chair: Boualem Djehiche)**

- **Patrick Cheridito** (ETH Zürich): Computing capital requirements with guarantees
- **Han Li** (The University of Melbourne): Augmenting hierarchical time series through clustering: Is there an optimal way for forecasting?
- **Salvatore Scognamiglio** (University of Naples Parthenope): Multiple yield curve modeling and forecasting using deep learning

**11:00 - 11:30 Coffe break**

**11:30 - 12:30 Regular Session 5**

**Room Hörsal 1: Machine learning & pricing 2 (Chair: Can Baysal)**

- **Paul Wilsens** (KU Leuven): Reducing the dimensionality and granularity in hierarchical categorical variables
- **Jan Küthe** (Akur8): Penalized regression - Between Credibility and GBMs
- **Mathias Lindholm** (Stockholm university): On duration effects in non-life insurance pricing

**Room Lärosal 5: Machine learning & reserving 2 (Chair: Grainne McGuire)**

- **Gabriele Pittarello** (University of Torino): Individual claims reserving using the Aalen-Johansen estimator
- **Martin Bladt** (University of Copenhagen): Bootstrapped multi-states and model uncertainty
- **Selim Gatti** (ETH Zürich): Modeling lower-truncated and right-censored insurance claims with an extension of the MBBEFD class

**12:30 - 13:25 Lunch**

**13:25 - 14:25 Regular Session 6**

**Room Lärosal 1: Machine learning in insurance 1 (Chair: TBA)**

- **Agathe Fernandes Machado** (UQAM, Montréal): Challenging calibration of insurance scoring algorithms
- **Charlotte Jamotton** (Université catholique de Louvain): Variational autoencoder for synthetic insurance data
- **Can Baysal** (Munich RE): Demand & elasticity modelling for P&C insurance pricing under various conditions

**Room Lärosal 5: Machine learning in insurance 2 (Chair: Boualem Djehiche)**

- **Meryem Yankol-Schalck** (IPAG Business School): Auto insurance fraud detection: Leveraging cost sensitive and insensitive algorithms for comprehensive analysis
- **Giovanni Rabitti** (Heriot-Watt University): Measuring unexplained variation in insurance data: a non-parametric approach based on global sensitivity indices

- **Mick Cooney** (Describe Data): Using random portfolios for managing and assessing insurance risks

**14:25 - 14:55 Lightning Session 2**

**Room Hörsal 1: Stream 1 (Chair: Mathias Lindholm)**

- **Yousra Cherkaoui** (CREST-ENSAE): Cyber risk modeling using a two-phase Hawkes process with external excitation
- **Emilio S. Guillén** (Bayes Business School, City, University of London): Generalized additive models and functional gradient boosting with geometrically designed (GeD) splines: Application to insurance data
- **Antoine Burg** (Université Paris Dauphine (CEREMADE) & SCOR): Leverage closed-form MLE for multivariate regression models: GLM-trees and actuarial applications

**Room Lärosal 5: Stream 2 (Chair: Markus Gesmann)**

- **Giulia Pucci** (KTH Royal Instituto of Technology): Network-based optimal control of pollution growth
- **Adam Bedwell-Smith** (Ki Insurance): Genetic algorithms for property binder portfolio optimisation
- **Soroush Amirhashchi** (Plannet Insurtech Hub): Modeling mortality rates: Nonparametric Bayesian inference with gaussian Cox processes

**14:55 - 15:10 Room Hörsal 1: Closing comments**

# Abstracts of contributed talks

## Genetic Algorithms for Property Binder Portfolio Optimisation

**Adam Bedwell-Smith, Ki Insurance (presenter)**

**Abstract**: The Lloyd's of London specialty insurance market is ripe for disruption. Global risk landscapes are changing at an unprecedented pace, and insurers must surely embrace algorithmic approaches to remain competitive. We present a novel application of modern portfolio theory within the context of specialty insurance. We first construct a deterministic loss simulation for our existing property binder portfolio, utilising the output of external catastrophe and in-house attritional loss models. The high-dimensionality of the portfolio and complex non-linear relationships captured within the simulation introduce a constraint-optimisation problem which is hard to solve using traditional methods. Hence, we reach for the class of biology-inspired stochastic optimisation techniques and implement genetic algorithms [1, 2] to optimise our portfolio for metrics including expected return and volatility. To the best of our knowledge, there has been no previous application of evolutionary algorithms for portfolio management within the insurance industry.

### References

1. Holland, J.H. (1975). Adaptation in natural and artificial systems. The U. of Michigan Press.
2. Alam, T., Qamar, S., Dixit, A., and Benaida M. (2020). Genetic Algorithm: Reviews, Implementations, and Applications.

## Challenging Calibration of Insurance Scoring Algorithms

**Agathe Fernandes Machado, UQAM, Montréal, Canada (presenter)**

**Arthur Charpentier, UQAM, Montréal, Canada**

**Emmanuel Flachaire, Aix-Marseille School of Economics, Marseille, France**

**Ewen Gallic, Aix-Marseille School of Economics, Marseille, France**

**François Hu, Université de Montréal, Montréal, Canada**

**Abstract**: The assessment of binary classifier performance traditionally centers on discriminative ability using metrics, such as accuracy. However, these metrics often disregard the model's inherent uncertainty, especially when dealing with sensitive decision-making domains, such as finance or healthcare. Given that model-predicted scores are commonly seen as event probabilities, calibration is crucial for accurate interpretation. For a classifier, calibration means that the estimated class probabilities are reflective of the true underlying probability of the sample. Following Van Calster et al. (2019), a (well-) calibrated model means that, "*among patients with an estimated risk of 20%, we expect 20 in 100 to have or to develop the event*".

Probabilities of occurrence find direct application in insurance, whether in determining mortality rates, assessing fraud risk, or predicting the likelihood of policy terminations. Here, we illustrate the evaluation of calibration for regression models (Logistic Regression) and ensemble methods within an insurance context. Employing commonly used recalibration techniques, such as local regression from Denuit et al. (2021), we aim to enhance the calibration of predicted scores. Furthermore, we extend the notion of calibration to address fairness considerations using group-wise calibration with respect to sensitive attributes, as presented in Baumann et al. (2023).

**Keywords**: Calibration, Binary Classification, Probabilities of Occurrence, Scores

### References

1. Baumann, J., & Loi, M. (2023). Fairness and Risk: An Ethical Argument for a Group Fairness Definition Insurers Can Use. *Philosophy & Technology*, 36(3), 45.
2. Denuit, M., Charpentier, A., & Trufin, J. (2021). Autocalibration and Tweedie-dominance for insurance pricing with machine learning. *Insurance: Mathematics and Economics*, 101, 485-497.
3. Machado, A. F., Charpentier, A., Flachaire, E., Gallic, E., & Hu, F. (2024). From Uncertainty to Precision: Enhancing Binary Classifier Performance through Calibration. arXiv:2402.07790
4. Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L., Steyerberg, E. W., & Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative Bossuyt Patrick Collins Gary S. Macaskill Petra McLernon David J. Moons Karel GM Steyerberg Ewout W. Van Calster Ben van Smeden Maarten Vickers Andrew J. (2019). Calibration: the Achilles heel of predictive analytics. *BMC medicine*, 17(1), 230.

## Enabling business steering and decision-making through volatility modelling

**Alan Muro, Swiss Re Corporate Solutions (presenter)**

**Jakob Dambon, Swiss Re**

**Abstract**: We introduce a new tool that assesses a company's Loss Ratio volatility and tail risk across various market scenarios and outward reinsurance options. The main goals are to support data-driven decision making around portfolio-strategy at Executive level, to allocate risk capital to granular underwriting portfolios for business planning and IFRS/pricing purposes, and to optimize outward reinsurance structures.

Our tool is based on Monte Carlo simulations at high portfolio granularity level, to which we apply correlations and complex reinsurance structures. In order to overcome computational challenges, we employ parallelization methods. We leverage historical data and exposure models to calibrate our loss models and correlations for both Manmade and Nat Cat perils. At the end, we analyze statistical volatility, tail risks, and Loss Ratio exceedance probabilities at different portfolio granularities. We compare these results across several market scenarios and reinsurance options to enable data-driven decision making. In order to share our results across the company, our tool and results are made available through an R package and RShiny app, which are hosted on the Posit Workbench ecosystem.

**Keywords**: Capital modelling, Risk modelling, Monte Carlo, Reinsurance, Posit Workbench.

## Credibility in network regression with sigma-hot encoding and weight balancing

**Anne van der Scheer, Perunum Actuarieel Advies**

**Abstract**: Actuarial credibility can be modeled by adding random effects to risk groups. In recent literature on neural network regression several extensions are proposed following the example of generalized linear mixed models (GLMMs).

We take a different approach by equalizing categorical and continuous input variables while keeping the original network. Standardization of continuous variables is generally considered crucial for the learning capacity of a network. Building on this, we introduce sigma-hot encoding as the alternative to one-hot encoding, using relative frequencies of observed categories to set input values. At the same time, we propose weight balancing for each categorical variable during training of the network. This enables an equal playing field for all input variables and ensures only minor deviating responses of relatively sparse categories.

With density estimation under maximum likelihood and constrain of the max norm of all weights, we illustrate sigma-hot encoding and weight balancing in two examples. In the first example we repeat a simulation study applied to GLMMNet and show that our method produces similar results. In the second example our simulation is showing promising results for a more complex disability insurance which does not adhere to the conditions of recent network models.

In conclusion, we believe sigma-hot encoding and weight balancing of categorical variables to solve actuarial credibility in an original Machine Learning way. This can broaden the application of credibility in an assumption free setting, and therefore increase the use of neural networks in the insurance industry.

**Keywords**: actuarial credibility, neural network, regression, sigma-hot encoding, weight balancing, density estimation

### References

1. Avanzi, B., Taylor, G., Wang, M., & Wong, B. (2023). Machine Learning with High-Cardinality Categorical Features in Actuarial Applications. *arXiv preprint arXiv:2301.12710*.
2. Chilinski, P., & Silva, R. (2020, August). Neural likelihoods via cumulative distribution functions. In *Conference on Uncertainty in Artificial Intelligence* (pp. 420-429). PMLR.
3. Richman, R., & Wüthrich, M. V. (2023). High-Cardinality Categorical Covariates in Network Regressions. *Available at SSRN 4549049*.

## Leverage closed-form MLE for multivariate regression models: GLM-trees and actuarial applications

**Antoine Burg, Université Paris Dauphine (CEREMADE), SCOR (presenter)**

**Christophe Dutang, Université Grenoble Alpes, CNRS, Grenoble INP, LJK (not attending)**

**Abstract**: The maximum likelihood estimator (MLE) remains the most frequently used method to estimate the parameters of generalized linear models (GLM). But even for distributions within the exponential family, MLEs are not always tractable and need to be computed with numerical methods like Newton-Rapshon-type algorithm. Alternative closed-form estimator have been found in case of categorical explanatory variables for univariate random variables of one-parameter exponential type. Extensions of these estimators in case of multivariate and/or multi-parameter exponential type distributions are proposed, and some of their properties are studied. Closed-form estimators are especially useful to fasten computational times, e.g. GLM-trees algorithm, which require a lot of successive GLMs fitting procedures. Use cases of such algorithms and estimators within standard actuarial modelling will be presented.

**Keywords**: Regression models, explicit estimators, GLM, categorical explanatory variables, multiparameters exponential distributions, GLM-trees; mortality, longevity, causes-of-death.

### References

1. Rusch, T., Zeileis, A. (2013). Gaining insight with recursive partitioning of generalized linear models. *Journal of Statistical Computation and Simulation* **83**(7), 1301–1315.
2. Brouste, A., Dutang, C., Rohmer, T. (2020), Closed-form maximum likelihood estimator for generalized linear models in the case of categorical explanatory variables: application to insurance lossmodeling. *Computational Statistics*, **35**, 689–724.
3. Dutang, C., Guibert, Q. (2022). An explicit split point procedure in model-based trees allowing for a quick fitting of GLM trees and GLM forests. *Statistics and Computing* **32**(1), 6.

## Obtaining Fair Insurance Premiums with Multiple Sensitive Attributes

**Arthur Charpentier, UQAM, Montréal, Canada (presenter)**

**Agathe Fernandes Machadoa, UQAM, Montréal, Canada**

**Suzie Grondin, ENSAE, Paris, France & UQAM, Montréal, Canada**

**François Hu, Université de Montréal, Montréal, Canada & Milliman Paris, France**

**Philipp Ratz, UQAM, Montréal, Canada**

**Abstract**: According to Kranzberg's law, "*technology is neither good nor bad, nor is it neutral*". In the case of discrimination, this is all the more true as the notion of intent no longer exists in law (indirect discrimination). Without bad intent, a predictive algorithm used in fraud analysis, to estimate compensation for bodily injury, or to set a rate, can perfectly well be racist, ageist, sexist, or all three at once. As claimed by Kearns et al. (2019), "*machine learning won't give you anything like gender neutrality 'for free' that you didn't explicitly ask for*". Here, we'll look at how to achieve (strict) neutrality in pricing algorithms, using post-processing approaches. In Charpentier et al. (2023), we discussed the use of Wasserstein barycenter to obtain fair price, well-balanced, and also that also satisfy statistical guarantees (smallest loss within the class of fair models). Unfortunately, they might struggle to provide a transparent and accurate path to achieve fairness when more than one sensitive variable, such as a combination of gender and race, is involved, for example. Here, we propose an interpretable approach, using some sequential mechanism, and we will illustrate the equipy python package on some insurance models.

**Keywords**: Discrimination, Fairness, Interpretable, Optimal Transport, Wassertein

### References

1. Charpentier, A., Hu, F., & Ratz, P. (2023). Mitigating Discrimination in Insurance with Wasserstein Barycenters. arXiv:2306.12912, *BIAS 2023*.
2. Charpentier, A. (2024). Insurance, biases, discrimination and fairness. Springer.
3. Roth, A., & Kearns, M. (2019). The Ethical Algorithm. Oxford University Press.
4. Hu, F., Ratz, P., & Charpentier, A. (2023). A sequentially fair mechanism for multiple sensitive attributes. arXiv:2309.06627, *AAAI 2024*.
5. Machado, A., Grondin, S., Hu, F., Ratz, P. & Charpentier, A. (2024). Sequential fairness using optimal transport. *In progress* equipy

## LLM-Powered Information Extraction for Claim Documents

**Benedikt Herwerth, Swiss Re (presenter)**

**Abstract**: Claims handling is an integral part of the insurance value chain, which often involves document-heavy workflows. Examples of documents that need to be processed include: First-notice-of-loss (FNOL) documents, description of the loss, communications (for example with brokers, insureds, or co-insurers), payment-related documents, and adjuster reports.

To address these points, Swiss Re started an initiative to enable large-scale processing of claims documents in commercial insurance. As part of this initiative, we developed a model that targets handling efficiency. The underlying business challenge is as follows. When Swiss Re receives a new claim notification through E-Mail, information needs to be identified from unstructured data. This workflow is currently done manually and involves assigning the claim to a policy, which is uniquely identified by its ID. The model facilitates this workflow by extracting the 'core part' of the policy ID from the notification and suggesting that to the human specialist. (A policy ID consists of multiple parts, where we refer to the main part of the policy as its 'core' as it remains the same in case of renewals.)

While this model solves a key business problem, it is not easily generalizable due to its heavy reliance on regular expressions. To extend its coverage, a proof-of-concept (PoC) has been instituted to put to the test new Large Language Models (LLMs) capabilities.

As part of the PoC, the Azure OpenAI GPT3.5 Turbo model was used (both 4k and 16k input context length variations) and additional techniques such as retrieval-augmented generation (RAG) using ChromaDB and hallucination guardrails were implemented. While text-level annotations were not available, we quantified the LLMs performance at document level where ground-truth data was available. Furthermore, key focus was spent on tracking/making results reproducible using MLflow and quantitatively measuring hallucinations rates.

The PoC showed some performance improvements and allowed to extend the previous model's capabilities to new entities. Particularly, it was possible to extract date of losses out of the box with a "fuzzy precision" of about 70% from a sample of more than 300 documents, where we compared the model's extraction against the ground truth using a fuzzy string comparison.

Because there are many other aspects to the underlying challenge, we argue that the LLM alone does not provide a catch-all solution to fully digitize the process. However, it can become a key component. When paired with human-in-the-loop approaches, it has the potential to yield novel efficiency benefits, for example by pre-populating fields for human specialists.

**Keywords**: Claims, Large Language Models, Generative AI, OpenAI, Information Extraction

### References

1. Azure OpenAI, https://azure.microsoft.com/en-us/products/ai-services/openai-service
2. ChromaDB, https://github.com/chroma-core/chroma
3. MLflow, https://mlflow.org/

## Demand & elasticity modelling for P&C insurance pricing under various conditions

**Can Baysal, Munich RE (presenter)**

**Abstract**: Demand and elasticity modelling play a crucial role in insurance pricing, offering insights into customer-price relationship, facilitating pricing decisions and portfolio steering. In this session, I will talk about various methodologies for data collection & test design, approaches to demand modelling & derivation of price elasticity, and practical use cases of demand models in insurance pricing.

The presentation encompasses the technical definition of price elasticity, a fundamental concept in demand modelling, and focuses on its varying implications for insurers, influenced by factors such as brand recognition and pricing strategies. It then delves into demand modelling methodologies, discussing both randomized price tests and approaches in the absence of such tests. It explores the structure of demand models, emphasizing the importance of test design and data collection for accurate analysis.

Furthermore, the study investigates the application of generalized linear models (GLM) and tree-based models in demand modelling, outlining their respective advantages and disadvantages and the disparities in derived elasticity.

In the final section of the presentation, practical use cases such as customer insight analysis, scenario testing, and price optimisation demonstrate the tangible benefits and real-world applications of demand and elasticity modelling in insurance pricing, providing guidance for enhancing pricing strategies and customer-centric approaches.

**Keywords**: Non-life insurance pricing, demand modelling, price elasticity, generalised linear models, gradient boosting models

### References

1. Serhat Guven, FCAS, MAAA, and Michael McPhail, FCAS, MAAA. “Beyond the Cost Model: Understanding Price Elasticity and Its Applications” CAS E-Forum 2. Casact.
2. Karl P. Murphy, Michael J. Brockman, and Peter K. W. Lee. “Using Generalized Linear Models to Build Dynamic Pricing Systems for Personal Lines Insurance” CAS Forum (Winter 2000): 107-139.

## Variational AutoEncoder for synthetic insurance data

**Charlotte Jamotton, Université catholique de Louvain (presenter)**

**Donatien Hainaut, Université catholique de Louvain**

**Abstract**: Previous research has demonstrated the successful use of generative models, particularly Variational AutoEncoders (VAEs) [2], in various domains such as image recognition [1], text classification [4], and recommender systems [3]. However, their application to insurance data, specifically heterogeneous insurance portfolios containing different types of covariates (i.e., (positive) numerical, binary, Multinomial, or ordinal) with a variety of marginal distributions (e.g., multimodal, long tail), remains unexplored. Our study aims at addressing challenges related to data availability in the insurance industry by generating synthetic insurance policies. In this article, we adapt the VAE architecture for unsupervised learning tasks in the actuarial field, including dimensionality reduction and synthetic data generation. We propose a VAE model with a quantile transformation of continuous data and a reconstruction loss that combines categorical cross-entropy and mean squared error, along with a KL divergence-based regularization term. We analyze our VAE's ability to reconstruct complex insurance data and generate synthetic insurance policies using a motor portfolio.

**Keywords**: autoencoder, variational inference, synthetic data generation, heterogeneous insurance data, dimension reduction.

### References

1. Baur, C., Wiestler, B., Albarqouni, S., & Navab, N. (2019). Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4 (pp.161-169). Springer International Publishing. doi:10.1007/978-3-030-11723-8.
2. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv:1312.6114.
3. Li, X., & She, J. (2017, August). Collaborative variational autoencoder for recommender systems. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 305-314. https://cseweb.ucsd.edu//classes/fa17/cse291-b/reading/p305-li.pdf.
4. Xu, W., Sun, H., Deng, C., & Tan, Y. (2017, February). Variational autoencoder for semi-supervised text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, No. 1. https://ojs.aaai.org/index.php/AAAI/article/view/10966.

## Boost Climate Risk Modelling with Large Language Models Data Augmentation

**Claudio Giorgio Giancaterino, Intesa San Paolo Vita (presenter)**

**Abstract**: Climate change poses a significant threat to global ecosystems and human societies. The resulting effects include extreme weather events, rising sea levels, and disruptions to ecosystems. Predicting the occurrence and impacts of these events is crucial for mitigating their consequences and protecting vulnerable communities.

The research has been focused on the flood event type from the US storm events database with data provided by the National Weather Service (NWS) and containing statistics on deaths, injuries and damage estimates.

By leveraging ChatGPT, the aim has been to retrieve features from narrative data related to climate events, and specifically, have been employed zero-shot text classification and embeddings techniques used as data augmentation.

To forecast deaths, injuries, and damages the performance of the baseline Naive method has been compared with several Machine Learning models including Generalized Linear Models (GLM), LightGBM, and Neural Networks.

**Keywords**: Climate Change, LigthGBM, GLM, Neural Networks, Naive Forecasting, LLMs, GenAI, ChatGPT

### References

1. P.J. Brockwell, R.A. Davis (2016). *Introduction to Time Series and Forecasting*, Springer
2. J. Birkmann, T. Welle (2015). *Assessing the risk of loss and damage: exposure, vulnerability and risk to climate-related hazards for different country classifications*, International Journal of Global Warming.
3. Dineva Snezhana, (2023). *Applying Artificial Intelligence (AI) for Mitigation Climate Change Consequences of the Natural Disasters*, SSRN

### Contact details

- Email: c.giancaterino@gmail.com
- Repository: https://github.com/claudio1975/Climate_Risk_Modelling_with_LLMs/
- Social media: https://www.linkedin.com/in/claudiods/

## Cloudburst hazard analysis – a GIS approach

**Daniel Knös, Model developer, Guy Carpenter**

**Abstract**: In recent years, the insurance industry in the Nordics have suffered from a number of costly events caused by cloudbursts. Heavy precipitation in urban areas in a short time exceeds the capacity of the groundwater system and leads to severe urban flooding. The requirements on spatial accuracy in traditional modelling makes this a complex peril to model because of the high computational costs. Here I present a simplified GIS approach based on the historical event from Gävle 2021 and claims data from the insurance industry, to quantify worst case scenarios from other locations. The historical event is moved to other cities, resulting in loss estimations given that the event would have occured over other, possibly more densely populated and exposed areas.

**Keywords**: Cloudburst, Climate, Natcat

## A causal machine learning approach for estimating heterogeneous treatment effects in the primary catastrophe bond market

**Despoina Makariou, University of St Gallen (presenter)**

**Pauline Barrieu, London School of Economics**

**Yining Chen, London School of Economics**

**Abstract**: We introduce a causal machine learning approach to predict treatment heterogeneity in the primary catastrophe bond market. We find that issuance timing affects catastrophe bond spreads, but this result varies according to several market and bond specific factors. Studying the issuance timing is important for optimising the cost of capital and ensuring the success of a catastrophe bond offering for sponsors and investors.

**Keywords**: Causal machine learning, heterogeneous treatment effects estimation, catastrophe bonds

### References

1. Susan Athey. Julie Tibshirani. Stefan Wager. "Generalized random forests." Ann. Statist. 47 (2) 1148 - 1178, April 2019. https://doi.org/10.1214/18-AOS1709
2. Makariou, D., Barrieu, P. and Chen, Y., 2021. A random forest based approach for predicting spreads in the primary catastrophe bond market. Insurance: Mathematics and Economics, 101, pp.140-162.

## Hush Hush: Keeping Neural Network Claims Modelling Private, Secret, and Distributed Using Federated Learning

**Dylan Liew, Institute and Faculty of Actuaries Federated Learning Working Party (presenter)**

**Malgorzata Smietanka, Institute and Faculty of Actuaries Federated Learning Working Party (presenter)**

**Scott Hand, Institute and Faculty of Actuaries Federated Learning Working Party**

**Harry Loh, Institute and Faculty of Actuaries Federated Learning Working Party**

**Michelle Chen, Institute and Faculty of Actuaries Federated Learning Working Party**

**Abstract**: Federated Learning is a new method of training Machine Learning models pioneered by Google in 2016 aimed for use on smartphones. This concept enables the direct training of machine learning models on users' devices, such as smartphones, eliminating the need to share or transfer potentially sensitive data to a centralized server. Unlike traditional machine learning methodologies, federated learning adopts a model where the algorithm is brought to the data, rather than transferring the data to the algorithm.

In this presentation, the Institute and Faculty of Actuaries (IFoA) Federated Learning Working Party (part of the IFoA Data Science Research Section) will illustrate how insurance companies can leverage this technique. Specifically, we will show how these companies can collaboratively develop a Neural Network model to predict claims frequency. This collaboration allows for the combination and utilization of their customer data without actually sharing or compromising any sensitive information.

We achieve this using the Flower package in Python along with PyTorch. We simulate a car insurance market's claims data with 10 companies in it using the freMTPL2freq dataset. We find that if all insurers are allowed to share their confidential data with each other they could collectively build a model that achieves an accuracy (Poisson Deviance Explained or "% PDE") of c.5% on an unseen sample. However if they are not allowed to share their customer data none of them can achieve more than c.3% PDE on the same unseen sample. If they use Federated Learning they could keep all of their customer data private, and build a model that achieves near the same accuracy as if their confidential data was shared, reaching c.5% on the same unseen sample.

**Keywords**: Federated Learning, Collaborative Modelling, Claims Frequency Prediction, Data Privacy, Deep Learning, Machine Learning, Flower, PyTorch

### References

1. Smietanka, M., Liew, D., Giorgio, C. (2021). All together now: modelling claims using federated learning. *The Actuary Magazine* **September 2021**, 22-26.

## Quantification of climate risk in Insurance

**Efren Hernandez, Management Solutions (presenter)**

**Abstract**: Quantifying climate risk presents an emerging and critical challenge for the insurance sector. This talk focuses on defining and applying climate scenarios, encompassing both physica and transition variables, to measure th financial impact of climate change on insurers. We explore advanced methodologies for estimating the financial effects derived from climate risk, highlighting the importance of a comprehensive understanding of risk scenarios and their potential impact on the financial future of insurance entities.

The relevance of developing robust predictive models that integrate climate and transition variables, thereby facilitating and accurate estimation of risks and their possible effects on the financial stability of insurers, is discussed. The presentation will also address the latest trends and challenges in measuring and managing climate risk, underscoring the crucial role of collaboration between insurers, regulators, and other relevant stakeholders to enhance climate risk quantification practices.

**Keywords**: Climate Risk, Climate Scenarios, Quantification, Climate Change

### References

1. Management Solutions (2020). Managing climate change risk
2. NGFS (2023). Climate Scenarios for Central Banks and Supervisors - Phase IV
3. IPCC (2023). Sixth Assessment Report
4. Battiston, S. (2022). Climate Policy Relevant Sectors

## Leveraging AI for DORA Compliance Assessment in Insurance

**Eivind Borg, Analytika (presenter)**

**Sven Haadem, phd., Analytika**

**Atle Melve, Analytika**

**Nito Tharmanathan, Analytika**

**Kjell Andreas Solbereg, Analytika**

**Abstract**: We exploit the capabilities of Artificial Intelligence (AI) to assess insurance policies for compliance with the Digital Operational Resilience Act (DORA), a new EU regulation aiming to enhance cybersecurity resilience in the financial sector. We address the limitations of generic Large Language Models (LLMs) for this task and propose a solution that combines human expertise and fine-tuned AI models.This project demonstrates the potential of AI for automating specific tasks within the DORA compliance process for the insurance industry. However, the success of such endeavors hinges on the combined expertise of AI engineers and compliance specialists, highlighting the importance of a collaborative approach.

Many companies struglee with DORA's complexity and reliance on diverse regulatory documents as it is a huge challenge for traditional compliance assessment methods. We identified an opportunity for AI due to the structured nature of both DORA regulations and insurance policies. However, initial experiments with generic LLMs yielded unreliable results, highlighting the need for a more nuanced approach. To overcome these challenges, we developed a multi-stage pipeline. First, we fine-tuned a large language model based on Mistral, specifically trained on relevant data. This model interacts with Langchain, a framework for building modular pipelines, and a set of custom agents, each tasked with verifying specific compliance aspects within a policy. The final assessment report is generated by aggregating the results from these agents. The pipeline is currently hosted on a combination of cloud and on-premise infrastructure, with the LLM residing on a dedicated GPU server.

Recognizing the limitations of AI, we integrated human expertise throughout the process. Compliance specialists played a crucial role in:

- Dividing DORA regulations into smaller, manageable segments for the LLM to analyze.
- Validating the LLM's findings to ensure accuracy and interpretability.

This human-in-the-loop approach ensures the reliability and trustworthiness of the AI-assisted assessments.

**Keywords**: Generativ Models, ML, LLM, Compliance

### References

1. Mistral 7B. Jiang et al. arXiv:2310.06825 [cs.CL]

## How Well do Firms Recover from Idiosyncratic Shocks? Evidence from Insurance Claims

**Emil Bustos, Research Institute of Industrial Economics (Presenter)**

**Oliver Engist, Stockholm School of Economics**

**Christian Thoman, Royal Institute of Technology**

**Abstract**: Firm-specific idiosyncratic supply shocks matter for understanding the evolution of firm growth as well as aggregate macroeconomic conditions (Barrot & Sauvagnat, 2016; Carvalho et al., 2021). Yet, little is known about how firms recover, primarily due to the scarcity of data on these idiosyncratic shocks and the difficulty in isolating the effects of supply or demand shocks.

We employ a novel dataset from a major Swedish insurance company covering 2008 to 2017. This dataset stands out for its detailed documentation of firm losses and for spanning a substantial portion of the Swedish corporate sector. We employ a stacked event study approach to estimate the effects of large fire losses, as these are largely unexpected.

We find no effect on physical capital, suggesting that firms can rebuild their lost assets. However, sales and employment decline and remain about 15%-20% lower in the medium term. Moreover, we find no effect on wages, which suggests that firms do not share these negative shocks with their workers, but rather reduce employment (Carlsson et al., 2016).

We then explore several potential channels. First, we show that group firms fare better than independent firms (Santioni et al., 2020). Secondly, we find that financing constraints do not explain these differences (Caggese et al., 2019). Finally, firms with low insurance coverage fare somewhat worse.

Taken together, these results suggest that real frictions rather than financial ones matter in explaining how firms recover from idiosyncratic productivity shocks in our setting.

**Keywords**: Idiosyncratic Shocks, Firm Growth, Insurance, Private Firms

### References

J.-N. Barrot and J. Sauvagnat. Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. *The Quarterly Journal of Economics*, 131(3):1543–1592, 2016.

A. Caggese, V. Cunat, and D. Metzger. Firing the Wrong Workers: Financing Constraints and Labor Misallocation. *Journal of Financial Economics*, 133(3):589–607, 2019.

M. Carlsson, J. Messina, and O. N. Skans. Wage Adjustment and Productivity Shocks. *The Economic Journal*, 126(595):1739–1773, 2016.

V. M. Carvalho, M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi. Supply Chain Disruptions: Evidence From the Great East Japan Earthquake. *The Quarterly Journal of Economics*, 136 (2):1255–1321, 2021.

R. Santioni, F. Schiantarelli, and P. E. Strahan. Internal Capital Markets in Times of Crisis: The Benefit of Group Affiliation. *Review of Finance*, 24(4):773–811, 2020.

## A machine learning approach based on survival analysis for IBNR frequencies in non-life reserving

**Emil Hofman, University of Copenhagen / Alm Brand Group (presenter)**

**Gabriele Pittarello, Sapienza, Università di Roma**

**Munir Hiabu, University of Copenhagen**

**Abstract**:We introduce new approaches for forecasting IBNR (Incurred But Not Reported) frequencies by leveraging individual claims data, which includes accident date, reporting delay, and possibly additional features for every reported claim. A key element of our proposal involves computing development factors, which may be influenced by both the accident date and other features. These development factors serve as the basis for predictions. While we assume close to continuous observations of accident date and reporting delay, the development factors can be expressed at any level of granularity, such as months, quarters, or year and predictions across different granularity levels exhibit coherence. The calculation of development factors relies on the estimation of a hazard function in reverse development time, and we present three distinct methods for estimating this function: the Cox proportional hazard model, a feed-forward neural network, and xgboost (eXtreme gradient boosting). In all three cases, estimation is based on the same partial likelihood that accommodates left truncation and ties in the data. While the first case is a semi-parametric model that assumes in parts a log linear structure, the two machine learning approaches only assume that the baseline and the other factors are multiplicatively separable. Through an extensive simulation study and real-world data application, our approach demonstrates promising results. This paper comes with an accompanying R-package, *ReSurv*, which can be accessed at https://github.com/edhofman/ReSurv.

**Keywords**: Reserving, Survival Analysis, IBNR Claims, Neural Networks, XGBoost, Cox Regression

### References

1. Hofman, Emil and Pittarello, Gabriele and Hiabu, Munir (2023) *A machine learning approach based on survival analysis for IBNR frequencies in non-life reserving*, arXiv preprint arXiv:2301.03858.
2. Hofman, Emil and Pittarello, Gabriele and Hiabu, Munir. *ReSurv: Toolbox for applying survival analysis for claims frequency reserving*, https://github.com/edhofman/ReSurv
3. Pittarello, Gabriele and Hiabu, Munir and Villegas, Andrés (2023). *Replicating and extending chain-ladder via an age-period-cohort structure on the claim development in a run-off triangle*, arXiv preprint arXiv:2301.03858.
4. Hiabu, Munir (2023). *On the relationship between classical chain ladder and granular reserving*, Scandinavian Actuarial Journal 2017, 708-729.

## Generalized Additive Models and Functional Gradient Boosting with Geometrically Designed (GeD) Splines: Application to Insurance Data

**Dimitrina S. Dimitrova, Bayes Business School, City, University of London**

**Emilio S. Guillén, Bayes Business School, City, University of London (presenter)**

**Vladimir K. Kaishev, Bayes Business School, City, University of London**

**Abstract**: Geometrically Designed Splines (GeDS), introduced by Kaishev, Dimitrova, Haberman, and Verrall (2016) and further developed by Dimitrova, Kaishev, Lattuada, and Verrall (2023), have been demonstrated to be an efficient tool for solving regression problems incorporating large data samples with one or two covariates. In this paper we extend further the GeD spline methodology by:

- Incorporating Generalized Additive Models (GAM), thereby making GeDS highly multivariate.
- Implementing Functional Gradient Boosting (FGB), which improves the construction of the underlying spline regression model.

This enhanced spline regression methodology has been implemented in the **R** package **GeDS**, which is available from https://cran.r-project.org/package=GeDS. We demonstrate its efficiency using large, multivariate insurance datasets.

**Keywords**: Variable-knot spline regression, Generalized Additive Models, Functional Gradient Boosting

### References

1. Kaishev VK, Dimitrova DS, Haberman S, Verrall RJ (2016). "Geometrically designed, variable knot regression splines." *Computational Statistics*, **31**(3), 1079–1105. ISSN 1613-9658. doi: 10.1007/s00180-015-0621-7. URL https://doi.org/10.1007/s00180-015-0621-7.
2. Dimitrova DS, Kaishev VK, Lattuada A, Verrall RJ (2023). "Geometrically designed variable knot splines in generalized (non-)linear models." *Applied Mathematics and Computation*, **436**. ISSN 0096-3003. doi: 10.1016/j.amc.2022.127493. © 2022. This manuscript version is made available under the CC-BY-NCND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/, URL https://www.sciencedirect.com/science/article/pii/S0096300322005677.
3. Dimitrova S.D., Guillen E.S., Kaishev V.K., Lattuada A., and Verrall R.J. (2024). *GeDS: Geometrically Designed Spline Regression*, R package version R package version 0.2.0, https://CRAN.R-project.org/package=GeDS.

## Challenges and opportunities when implementing AI/ML in pricing and reserving

**Erik Gustafsson, Advisense**

**Abstract**: Recent years' technical development in AI and ML has seen increased applications in the insurance sector, solving old problems of pricing and reserving with new methods and previously untapped data sources.

These advancements come with both challanges - e.g. unitended induced volatlity in more and more granular and precise reserving models - but also opportunities - e.g improved accuracy in pricing models due to access to more data and higher data quality. The talk will cover some reflections made having seen several examples of the application of modern ML algorithms from the insurance industry.

**Keywords**: AI, ML, Reserving, Pricing

## Monitoring and modelling ESG risks in insurance

**Finn-Erik Langeggen, Advisense**

**Abstract**: The Partnership for Carbon Accounting Financials (PCAF) is a pioneering industry-driven initiative focused on evaluating the greenhouse gas emissions associated with the financial industry.

In 2022, PCAF expanded its scope to include the insurance sector, introducing a structured framework for insurance companies to responsibly share the responsibility of their clients' emissions, guided by a standardized methodology.

Adoption of this framework presents several challenges, notably concerning the lack of comprehensive data, uncertainties on the quality of existing data and the necessity to establish and refine the models in this framework. This talk will consider how the methodology can be used for the insurance industry, including open insurance as a potential means to gather data, and how to encourage more companies to join the initiative.

**Keywords**: PCAF, ESG, open insurance

### References

1. PCAF's Insurance-Associated Emissions Working Group https://carbonaccountingfinancials.com/files/downloads/pcaf-standard-part-c-insurance-associated-emissions-nov-2022.pdf

## A Dirichlet Process Mixture regression model for the analysis of competing risk events

**Francesco Ungolo, University of New South Wales (presenter)**

**Edwin R. van den Heuvel, Eindhoven University of Technology**

**Abstract**: We present a regression modelling approach for the analysis of competing risk events. The joint distribution of the time to these events is flexibly characterized by a random effect which follows a discrete probability distribution drawn from a Dirichlet Process, explaining their variability. This entails an additional layer of flexibility of this joint model, whose inference is robust with respect to the misspecification of the distribution of the random effects. The model is analysed in a fully Bayesian setting, yielding a flexible Dirichlet Process Mixture model for the joint distribution of the time to events. An efficient MCMC sampler is developed for inference. The modelling approach is applied to the empirical analysis of the surrending risk in a US life insurance portfolio previously analysed by Milhaud and Dutang (2018). The approach yields an improved predictive performance of the surrending rates.

**Keywords**: Competing Risks, Survival Analysis, Dirichlet Processes, Bayesian analysis, Lapse risk, MCMC

### References

1. Ungolo, F., van den Heuvel, E.R. (2024). A Dirichlet Process Mixture regression model for the analysis of competing risk events. *Under review*
2. Milhaud, X., Dutang C. (2018). Lapse tables for lapse risk management in insurance: a competing risk approach. *European Actuarial Journal* **8**, 97-126

## Ensuring Algorithmic Fairness in Insurance Pricing: a Multi-Class Problem Perspective

**François Hu, Milliman Paris, France (presenter)**

**Christophe Denis, Sorbonne-Université, LPSM, France**

**Romuald Elie, Université Gustave Eiffel, LAMA, France**

**Mohamed Hebiri, Université Gustave Eiffel, LAMA, France**

**Abstract**: AI systems heavily rely on labeled data for decision-making, yet these datasets often harbor biases stemming from human decisions or societal inequities, which is particularly evident in insurance pricing due to varying regulatory frameworks across countries. For instance, in car insurance, some jurisdictions permit the consideration of personal factors like credit history, employment, or gender, while others prohibit their use due to concerns of unfair discrimination. However, even after removing these sensitive variables, biases persist. In response to the urgent need to mitigate bias in data-driven decision-making, algorithmic fairness has emerged as a critical ethical concern in predictive modeling. Despite its widespread application in insurance, there remains a lack of research addressing multi-class classification problems with fairness considerations, particularly within the insurance sector as multi-class tasks become more common. In this study, we extend both exact and approximate fairness definitions, focusing specifically on Demographic Parity, to the domain of multi-class classification. Our approach provides novel insights and methodologies based on Lagrangian methods directly relevant to the insurance industry. The proposed plug-in procedure is supported by rigorous theoretical guarantees, showcasing its capability to closely approximate the behavior of an optimal fair rule while maintaining effectiveness in predictive performance. Evaluation on synthetic and real insurance datasets demonstrates the efficacy of our approach in ensuring fairness.

**Keywords**: Algorithmic Fairness, Multi-class Classification, Insurance Pricing

### References

1. Denis, C., Elie, R., Hebiri, M., Hu, F. (2021). Fairness guarantee in multi-class classification. arXiv:2109.13642.
2. Elie, R., Hillairet, C., Hu, F., Juillard, M. (2021). Fair Active Learning: Solving the labeling problem in insurance. arXiv:2112.09466.
3. Hu, F., Ratz, P., Charpentier, A. (2023). Fairness in Multi-Task Learning via Wasserstein Barycenters. ECML-PKDD 2023.
4. Hu, F., Ratz, P., Charpentier, A. (2023). A sequentially fair mechanism for multiple sensitive attributes. arXiv:2309.06627, AAAI 2024.
5. Hu, F. (2022). Semi-supervised learning in insurance: fairness and active learning. Ph.D. Manuscript.

## Individual claims reserving using the Aalen-Johansen estimator

**Martin Bladt, University of Copenhagen**

**Gabriele Pittarello, University of Torino (presenter)**

**Abstract**: We propose an individual claims reserving model based on the conditional Aalen–Johansen estimator, as developed in [1]. In our approach, we formulate a multi- state problem, where the underlying variable is the individual claim size, rather than time. The states in this model represent development periods, and we estimate the cumulative density function of individual claim costs using the conditional Aalen–Johansen method as transition probabilities to an absorbing state. Our methodology reinterprets the concept of multi-state models and offers a strategy for modeling the complete curve of individual claim costs. To illustrate our approach, we apply our model to both simulated and real datasets. Having access to the entire dataset enables us to support the use of our approach by comparing the predicted total final cost with the actual amount, as well as evaluating it in terms of the continuously ranked probability score.

**Keywords**: multi-state models; conditional Aalen–Johansen; reserving; non-life insurance.

### References

1. Bladt, M., & Furrer, C. (2023). Conditional Aalen–Johansen estimation. arXiv preprint arXiv:2303.02119.

## Variable annuity portfolio valuation with Shapley additive explanations

**Gayani Thalagoda, University of New South Wales (presenter)**

**Katja Hanewald, University of New South Wales**

**Andrés Villegas, University of New South Wales**

**Jonathan Ziveyi, University of New South Wales**

**Abstract**: This study proposes a meta-modelling method that utilises Shapley additive explanations in selecting the representative sample in an explainable manner for valuing variable annuity portfolios. Selecting a representative sample with clear and explainable criteria is necessary when applying meta-modelling techniques to principle-based calculations. The proposed method involves (i) training a surrogate neural network with existing valuations for the same portfolio across multiple market conditions and (ii) decomposing the overall risk of a policy into clearly separated contributions from each risk driver using Shapley additive explanations. This decomposition results in an informative and explainable representation of the insurance policy data which can later be used for selecting the representative sample under a new market condition. Furthermore, by fine-tuning the surrogate neural network with the carefully chosen representative sample, the proposed method offers a systematic way to improve the neural network's estimation with limited data in the new market condition. The proposed method can assist users in explaining the reasoning behind selecting a policy as representative. Furthermore, the proposed method aligns with the U.S. National Association of Insurance Commissioners (NAIC)'s requirements for principle-based reserves for variable annuities, which necessitate representative policies to be selected in a manner that sufficiently reflects the impact of policy characteristics on the calculated risk measure. Our numerical analyses show that the proposed method outperforms conventional methods in the existing literature in prediction accuracy and goodness of fit.

**Keywords**: Variable annuities, Meta-modelling, Neural networks, Clustering, Shapley additive explanations

### References

1. Godin, F., Hamel, E., Gaillardetz, P., Ng, E. H. M. (2023). Risk allocation through shapley decompositions, with applications to variable annuities. *ASTIN Bulletin: The Journal of the IAA* **53/2**, 311–331.
2. Lin, X., Yang, S. (2020). Efficient dynamic hedging for large variable annuity portfolios with multiple underlying assets. *ASTIN Bulletin: The Journal of the IAA* **50/3**, 913–957.
3. Lundberg, S. M., Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30**, 4765-4774.
4. Gan, G. and Lin, X.S. (2017). Efficient Greek calculation of variable annuity portfolios for dynamic hedging: A two-level metamodeling approach. *North American Actuarial Journal* **21/2**, 161-177.
5. Gan, G. (2013). Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics* **53/3**, 795-801.

## Measuring unexplained variation in insurance data: a non-parametric approach based on global sensitivity indices

**Giovanni Rabitti, Heriot-Watt University**

**Abstract**: In many insurance datasets, it often happens that some risk factors are not included because they are unobservable or unreported. Then, it becomes essential to quantify the extent to which the observed explanatory risk factors provide relevant information on the quantity of interest. In previous works, this issue has been studied with a parametric approach using regression models.

In this work, we provide a non-parametric approach to quantify the importance of unobserved variables for the variation in the quantity of interest. We consider the total importance indices derived from sensitivity analysis literature, adapting them for the first time to address this particular problem. We discuss appealing theoretical properties of this importance measure in this context and propose a numerical estimation strategy.

To illustrate the versatility of our approach, we apply it to insurance datasets with different types of quantities of interest, including medical malpractice costs, cyber data breaches, country life expectancy, and conduct a comparative analysis of four car insurance datasets. In this latter case, we demonstrate that telematic variables explain a significant portion of the variance in policyholders' claim frequency.

**Keywords**: Global sensitivity analysis, unobserved risk factors, unexplained variance

### References

1. Hart, J., Gremaud, P.A. (2018). An approximation theoretic perspective of Sobol' indices with dependent variables. *International Journal for Uncertainty Quantification* **8(6)**, 483–493.
2. Honerkamp-Smith, G., Xu, R. (2016). Three measures of explained variation for correlated survival data under the proportional hazards mixed-effects model. *Statistics in Medicine* **35(23)**, 4153–4165.
3. Hössjer, O., Eriksson, B., Järnmalm, K., Ohlsson, E. (2009). Assessing Individual Unexplained Variation in Non-Life Insurance. *ASTIN Bulletin* **39 (1)**, 249–273.
4. Mase, M., Owen, A.B., Seiler, B.B (2024). Variable Importance Without Impossible Data. *Annual Review of Statistics and Its Application* **11(1)**, 1-27.

## Network-Based Optimal Control of Pollution Growth

**Giulia Pucci KTH Royal Instituto of Technology (presenter)**

**Fausto Gozzi Luiss University of Rome**

**Marta Leocata Luiss University of Rome**

**Abstract**: We are investigating pollution growth in time-space within a dynamic optimal control framework. We built an optimal control problem where a centralized social planner aims to find the optimal production policy over time and space to maximize welfare from consumption, taking into account a negative local pollution externality and the diffusive nature of pollution. We modelled the space as a network of interconnected geographic locations on a graph with state constraints, a choice motivated by the inherent graph-like structure that naturally characterises spatial data. This work aims to consider the spatial economic and ecological patterns generated by transboundary pollution on networks, where at each site (i.e., a node of the graph) some output is produced, consumed and locally invested. We are considering a linear production function to be maximised, depending on some capital input. We consider pollution mobility through a linear operator, which can be, for instance, the discrete Laplacian. We also included the possibility of investing in renewable energy as well. This will have a twofold effect: it increases the total production and influences pollution dynamics, even though with a lower impact compared to non-renewable production. Furthermore, in the perspective of transitioning towards cleaner energy, it makes sense to consider installation and maintenance costs. Several cases have been studied.

**Keywords**: Networks, Optimal Control, Pollution

### References

1. Boucekkine, R., Fabbri, G., Federico, S., & Gozzi, F. (2021). From firm to global-level pollution control: The case of transboundary pollution. European journal of operational research, 290(1), 331-345
2. Calvia, F. Gozzi, M. Leocata, G. I Papayiannis, A. Xepapadeas, A. N. Yannacopoulos. (2023) An optimal control problem with state constraints in a spatio-temporal economic growth model on networks, 2023 arXiv preprint arXiv:2304.11568.

## Measuring loss reserving uncertainty with machine learning models

**Grainne McGuire, Optum (presenter)**

**Greg Taylor, University of New South Wales**

**Abstract**:

Interest in the use of machine learning (ML) in reserving has been increasing in recent years, though its use in practice is not yet widespread. ML methods must tick a number of boxes before actuaries will feel comfortable deploying them – these include model stability, interpretability, ease of use, and the ability to estimate reserve uncertainty.

In this talk we will focus on this last topic and consider the following points:

- The components of loss reserving error
- How ML allows us to quantitatively estimate a greater proportion of the total loss reserving error than e.g. the Mack model or traditional bootstrapping approaches
- How to use bootstrapping with regularised regression models to obtain these estimates
- Issues faced in practice when bootstrapping ML models and how to deal with them.

Full R code for the analysis will be available on-line in the form of a detailed worked example on some real-life data.

**Keywords**: Bootstrapping, Lasso, Loss reserving, Uncertainty

### References

1. Taylor, G., McGuire, G. (2023). Model Error (or Ambiguity) and Its Estimation, with Particular Application to Loss Reserving. *Risks* **11(11)**, 185. https://doi.org/10.3390/risks11110185
2. McGuire, G. (2023) Model error via regularised regression - CAS monograph data. https://grainnemcguire.github.io/post/2023-05-04-model-error-example/

## Augmenting hierarchical time series through clustering: Is there an optimal way for forecasting?

**Han Li, The University of Melbourne (presenter)**

**Bohan Zhang, Beihang University**

**Anastasios Panagiotelis, The University of Sydney**

**Abstract**: Forecast reconciliation has attracted significant research interest in recent years, with most studies relying on pre-defined hierarchies constructed with time series metadata. With the goal of improving forecast accuracy in mind, we extend and contribute to the emerging research on the clustering-based reconciliation method by proposing a novel framework for hierarchy construction. This framework offers three approaches: cluster hierarchies, random hierarchies, and combination hierarchies. Utilizing the proposed approaches, we investigate the individual contributions of two primary factors, namely grouping and structure, to the performance of forecast reconciliation. Through a simulation study and experiments on two real-world datasets, we demonstrate the practical efficacy of different hierarchy construction approaches. Our findings provide new insights into the dynamics between grouping and structure, which lead to an improved understanding of forecast reconciliation.

**Keywords**: Forecast reconciliation; Hierarchical time series; Clustering; Forecast combination.

## A tree-based varying coefficient model

**Henning Zakrisson, Stockholm University (presenter)**

**Mathias Lindholm, Stockholm University**

**Abstract**: The paper introduces a tree-based varying coefficient model (VCM) where the varying coefficients are modelled using the cyclic gradient boosting machine (CGBM) from Delong et al. (2023). Modelling the coefficient functions using a CGBM allows for dimension-wise early stopping and feature importance scores. The dimension-wise early stopping not only reduces the risk of dimension-specific overfitting, but also reveals differences in model complexity across dimensions. The use of feature importance scores allows for simple feature selection and easy model interpretation. The model is evaluated on the same simulated and real data examples as those used in Richman and Wüthrich (2023), and the results show that it produces results in terms of out of sample loss that are comparable to those of their neural network-based VCM called LocalGLMnet.

**Keywords**: Generalised linear models, Multivariate gradient boosting, Feature selection, Interaction effects, Early stopping

### References

1. Zakrisson, H., Lindholm, M. (2024). A tree-based varying coefficient model. arXiv preprint arXiv:2401.05982
2. Delong, Ł., Lindholm M., and Zakrisson H. (2023). On cyclic gradient boosting. Available at SSRN 4352505.
3. Richman, R., and Wüthrich, M. (2023). LocalGLMnet: interpretable deep learning for tabular data. Scandinavian Actuarial Journal 2023.1: 71-95.

## Advancements in Embedded Insurance Pricing Models for Cargo Insurance

**Jakob Dambon, Swiss Re (presenter)**

**Tomas Hulan, Swiss Re**

**Samuel Kollar, Swiss Re**

**Gregoire Caro, Swiss Re**

**Abstract**: We present our ongoing research on cargo insurance pricing models, focusing on the enhancement of embedded insurance products by incorporating novel risk factors. In response to the dynamic landscape of cargo shipments by vessel, we have identified three main risk types originating from the vessel itself, the shipping route, and the seasonality. Our revised pricing model addresses these challenges comprehensively and aims to provide a more accurate and nuanced assessment of the policyholders' exposure.

The proposed model relies on an exhaustive analysis of historical data to derive three independent risk factors to ensure an evaluation of the overall risk. By leveraging the vessel specifications, we establish a detailed understanding of the vessels' risk profiles such as size, age, and incident history. Machine learning techniques are then employed to quantify the influence of these characteristics on the probability of insurable events. Similar analyses are used to assess the risk by shipping route and seasonality, respectively.

In summary, our approach involves a data-driven methodology, utilizing geo-data processing, advanced statistical techniques, and machine learning to the derive the risk factors. The resulting unified, dynamic pricing model combines the risk factors for the three types of risks. Key advantages of the proposed model include an increased granularity in risk evaluation leading to fairer and more tailored pricing structures, as well as a modular approach to account for evolving geo-political risks.

Patent Pending

**Keywords**: Cargo Insurance, Geo Data, Marine, Pricing, Vessel

### References

1. International Union of Marine Insurance (IUMI), (2023). *IUMI's 2023 analysis of the global marine insurance market* https://iumi.com/statistics/public-statistics
2. Pathak, K., Mahajan, S., Nordahl, H.A., Solum, A., (2024). *High waves, high claims: New study on container losses* https://www.gard.no/web/articles?documentId=36713383

## Penalized regression - Between Credibility and GBMs

**Jan Küthe, Actuarial Data Scientist at Akur8 (presenter)**

**Abstract**: Penalized regression is steadily becoming a mainstream application in ratemaking. There is momentum in the insurance space with innovation in research, software and production on penalized techniques that build and innovate the standard GLM models. We will display how Penalized regression can be related to GBMs, a powerful technique, yet entirely black box in nature. This will be helpful to get a better understanding of the advancing method landscape and should provide the tools to select the appropriate modelling framework.

Since Penalized regression can effectively tie standard Credibility practices and incorporate some of the benefits of GBMs, this presentation aims to contribute to the diffusion of these techniques as solid alternatives to standard GLMs for ratemaking.

Credibility and Penalized Regression, by Mattia Casotto, Marco Banterle, Guillaume Beraud-Sudreau

**Keywords**: penalized regression, GLM, credibility, modeling

### References

1. Casotto, M., Banterle M., Beraud-Sudreau, G.

## The Impact of Artificial Intelligence on Financial Institutions

**Javier Calvo, Management Solutions (presenter)**

**Abstract**: The advent of Artificial Intelligence (AI) is transforming the landscape of financial institutions, with a profound impact on the insurance sector. This presentation examines the multifaceted effects of AI adoption in insurers, from enhancing customer experience to improving risk assessment and management. We delve into the integration of AI technologies such as generative AI and machine learning models in insurance operations, illustrating their role in driving efficiency, accuracy, and innovation.

Moreover, we explore the challenges and ethical considerations surrounding data privacy, algorithmic bias, and the need for regulatory frameworks to govern AI usage in insurance. The EU AI Act serves as a pivotal regulation, aiming to safeguard ethical AI deployment by addressing these concerns. It establishes rigurous standards for transparency and accountability, thereby ensuring that AI advancements are leveraged innovatively while upholding a responsible approach.

Overall, the talk aims to provide a comprehensive overview of AI's transformative potential for insurers, highlighting best practices and future trends in leveraging AI to navigate the complexities of the modern risk landscape.

**Keywords**: Artificial Intelligence, Risk Management, Predictive Analytics

### References

1. Management Solutions (2023). Explainable artificial intelligence (XAI)
2. European Commission (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. EU AI Act (Draft)
3. European Insurance and Occupational Pensions Authority (EIOPA) (2023). Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector.
4. European Central Bank (2021). ECB Response to the European Commission's Consultation on the Artificial Intelligence Act.
5. The Geneva Association (2023). Regulation of Artificial Intelligence in Insurance: Balancing consumer protection and innovation.

## Accelerated underwriting and underwriting with partial information

**Jayant Apte, PhD, Data Analytics Solutions, Scor, Charlotte (presenter)**

**Antoine Ly, PhD, Data Analytics Solutions, Scor, Paris**

**Abstract**: Accelerated Underwriting (AUW) programs are popular in US life insurance market as a way for insurers to provide quicker and less intrusive customer journey to some customers when compared to traditional underwriting [1]. We consider the problem of designing a machine learning system to predict whether expensive/intrusive evidence can be waived for an applicant using features that can be derived from the inexpensive/non-intrusive evidence, in presence of multiple duration historical claims data. We show how this problem is fundamental to historical mortality data driven design of AUW models that does not regress to historical underwriting. Empirical results in the context of underwriting cost vs selected mortality tradeoff are presented based on NHANES data for several alternative solutions along with a novel labeling scheme that uses a pair of survival models.

**Keywords**: algorithmic underwriting, machine learning, accelerated underwriting, survival models

### References

1. Klein A., Li J. (2023). 2022 Accelerated Underwriting Practices Survey Report. *The Society of Actuaries Research Institute*, available online: link.
2. Chancel A., Bradier, L., Ly A., Ionescu R., Martin L., Sauce M., (2022) Applying Machine Learning to Life Insurance: some knowledge sharing to master it. *Arxiv preprint.* available online: link.

## Machine learning and supervision

**Jimmy Hollén, The Swedish Financial Supervisory Authority (presenter)**

**Farzad Ashouri, The Swedish Financial Supervisory Authority (presenter)**

**Jenny Önskog, The Swedish Financial Supervisory Authority**

**Abstract**: The Swedish Financial Supervisory Authority (SFSA) are supervising the financial institutions in Sweden. Given the ever increasing amount of data we have looked into different ways to improve and facilitate our supervisory work with help from different AI techniques. Within the Insurance division we have developed a machine learning model for analyzing qualitative structured data. The preferred model is a support vector machine algorithm evaluated with a repeated stratified 3-fold cross validation that was repeated 100 times. However, the main focus of the seminar will be a practitioners use of machine learning in supervision.

## Gen AI: Developing and deploying a specialized Underwriting AI Assistant

**Louis DOUGE, SwissRe (presenter)**

**Aneta RICHTEROVA, SwissRe**

**Jaouad MOUSSER, SwissRe**

**Robert SIMMEN, SwissRe**

**Abstract**: Swiss Re's Life Guide, a widely-used Underwriting Manual with over 23 million annual hits, aids Life & Health Underwriters by providing crucial guidance for assessing insurance applications. Recognizing the challenge of timely decision-making given the diverse profiles of applicants, we introduce an innovative add-on enabling human-like interaction, akin to a specialized assistant.

This solution, a first in integrating an advanced chat interface for underwriting decision-making, utilizes a Retrieval-Augmented Generation (RAG) system powered by GPT-4. It involves splitting content according to a text chunking strategy for precise information retrieval, followed by careful prompt construction to be fed to a Large Language Model (LLM) and finally generate responses. We detail the development process, from initial manual structuring and vector space embedding to deployment challenges, including specifics related to operating LLMs (text chunk management, cost and latency optimization, etc.). Our system also focuses on creating a trustworthy user experience despite the hallucinatory nature of LLMs.

Evaluation with Swiss Re Underwriters showed over 90% accuracy in retrieving and exposing relevant information, underscoring the system's impact on modernizing underwriting practices and addressing LLM limitations through effective evaluation and user-centered design.

**Keywords**: Retrieval-Augmented Generation, LLM, Underwriting

### References

1. Patrick Lewis, Ethan Perez et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *34th Conference on Neural Information Processing Systems*, (NeurIPS 2020).
2. Shahul Es, Jithin James et al. (2023). RAGAS: Automated Evaluation of Retrieval Augmented Generation. *ArXiv:2309.15217v1*.

## Mortality forecasting via multi-task neural networks

**Luca De Mori, Bayes Business School (presenter)**

**Abstract**: Neural networks have been being applied with a growing focus in recent years on the mortality forecasting field. We aim to implement a feed-forward type multi-task neural network to forecast mortality rates in a multi-population context. The multi-task neural networks' goal is to solve multiple tasks simultaneously by sharing some of the layers among the networks. As a dataset, we consider female and male populations of ten different developed countries from the Human Mortality Database. We also perform a preliminary cluster analysis on the countries considered to optimize the network structure. Then, we compare the forecast accuracy of the multi-task neural networks with traditional stochastic models and pre-existing single-task neural networks.

**Keywords**: Neural networks, mortality forecasting, clustering

### References

1. Richman, R., Wüthrich, M. (2021). A neural network extension of the Lee-Carter model to multiple populations. Annals of Actuarial Science.
2. Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey.

## Isotonic regression for variance estimation and its role in mean estimation and model validation

**Łukasz Delong, University of Warsaw, Faculty of Economic Sciences (presenter)**

**Mario Wüthrich, ETH Zürich, Department of Mathematics**

**Abstract**: We study isotonic regression which is a non-parametric rank-preserving regression technique. Under the assumption that the variance function of a response is monotone in its mean functional, we investigate a novel application of isotonic regression as an estimator of this variance function. Our proposal of variance estimation with isotonic regression is used in multiple classical regression problems focused on mean estimation and model validation. In a series of numerical examples, we (1) explore the power variance parameter of the variance function within Tweedie's family of distributions, (2) derive a semi-parametric bootstrap under heteroskedasticity, (3) provide a test for auto-calibration, (4) explore a quasi-likelihood approach to benefit from best-asymptotic estimation, (5) deal with several difficulties under lognormal assumptions. In all these problems we verify that the variance estimation with isotonic regression is essential for proper mean estimation and beneficial compared to traditional statistical techniques based on local polynomial smoothers.

**Keywords**: Isotonic regression, generalized linear model, Tweedie's family, power variance parameter, quasi-likelihood, lognormal model, auto-calibration, T-reliability diagram, bootstrap, best-asymptotic normal.

## Hush Hush: Keeping Neural Network Claims Modelling Private, Secret, and Distributed Using Federated Learning

**Dylan Liew, Institute and Faculty of Actuaries Federated Learning Working Party (presenter)**

**Malgorzata Smietanka, Institute and Faculty of Actuaries Federated Learning Working Party (presenter)**

**Scott Hand, Institute and Faculty of Actuaries Federated Learning Working Party**

**Harry Loh, Institute and Faculty of Actuaries Federated Learning Working Party**

**Michelle Chen, Institute and Faculty of Actuaries Federated Learning Working Party**

**Abstract**: Federated Learning is a new method of training Machine Learning models pioneered by Google in 2016 aimed for use on smartphones. This concept enables the direct training of machine learning models on users' devices, such as smartphones, eliminating the need to share or transfer potentially sensitive data to a centralized server. Unlike traditional machine learning methodologies, federated learning adopts a model where the algorithm is brought to the data, rather than transferring the data to the algorithm.

In this presentation, the Institute and Faculty of Actuaries (IFoA) Federated Learning Working Party (part of the IFoA Data Science Research Section) will illustrate how insurance companies can leverage this technique. Specifically, we will show how these companies can collaboratively develop a Neural Network model to predict claims frequency. This collaboration allows for the combination and utilization of their customer data without actually sharing or compromising any sensitive information.

We achieve this using the Flower package in Python along with PyTorch. We simulate a car insurance market's claims data with 10 companies in it using the freMTPL2freq dataset. We find that if all insurers are allowed to share their confidential data with each other they could collectively build a model that achieves an accuracy (Poisson Deviance Explained or "% PDE") of c.5% on an unseen sample. However if they are not allowed to share their customer data none of them can achieve more than c.3% PDE on the same unseen sample. If they use Federated Learning they could keep all of their customer data private, and build a model that achieves near the same accuracy as if their confidential data was shared, reaching c.5% on the same unseen sample.

**Keywords**: Federated Learning, Collaborative Modelling, Claims Frequency Prediction, Data Privacy, Deep Learning, Machine Learning, Flower, PyTorch

### References

1. Smietanka, M., Liew, D., Giorgio, C. (2021). All together now: modelling claims using federated learning. *The Actuary Magazine* **September 2021**, 22-26.

## Bootstrapped multi-states and model uncertainty

**Martin Bladt, University of Copenhagen (presenter)**

**Christian Furrer, University of Copenhagen**

**Abstract**: Aalen–Johansen estimation targets transition probabilities in multi-state Markov models subject to right-censoring. In particular, it belongs to the standard toolkit of actuaries and statisticians specializing in health and disability. The conditional Aalen-Johansen estimator is a recent kernel-based generalization that allows for the inclusion of covariates and allows for non-Markov behaviour. We establish consistency, normality, and validate a class of bootstrap procedures; here, the theory of empirical processes plays a central role. We also illustrate the practical implications, not least for assessing model uncertainty, and for quantifying the spread of the reserve in disability insurance.

**Keywords**: multi-states, model uncertainty, reserves, non-Markov models

### References

1. Bladt, M., & Furrer, C. Conditional Aalen–Johansen estimation. Preprint, University of Copenhagen. arXiv:2303.02119.

## On duration effects in non-life insurance pricing

**Mathias Lindholm, Stockholm university (presenter)**

**Taariq Nazar, Stockholm university**

**Abstract**: In this talk we discuss duration effects on the consistency of mean parameter and dispersion parameter estimators in exponential dispersion families (EDFs) that are the standard models used for non-life insurance pricing. Focus is on the standard generalised linear model assumptions where both the mean and variance, conditional on duration, are linear functions in terms of duration. We derive simple convergence results that highlight consequences when the linear conditional moment assumptions are not satisfied. These results illustrate that: (i) the resulting mean estimators always have a relevant. asymptotic interpretation in terms of the duration adjusted actuarially fair premium – a premium that only agrees with the standard actuarial premium using a duration equal to one, given that the expected value is linear in the duration; (ii) deviance based estimators of the dispersion parameter in an EDF should be avoided in favour of Pearson estimators; (iii) unless the linear moment assumptions are satisfied, consistency of dispersion and plug-in variance estimators can not be guaranteed and may result in spurious over-dispersion. The results provide explicit conditions on the underlying data generating process that will lead to spurious over-dispersion that can be used for model checking. This is illustrated based on real insurance data, where it is concluded that the linear moment assumptions are violated, which results in non-negligible spurious over-dispersion.

**Keywords**: Consistency, Duration, Exponential dispersion family, Generalised lin- ear model, Over-dispersion, Dispersion estimator

### References

1. Lindholm, M., Nazar, T. (2023) On Duration Effects in Non-Life Insurance Pricing. Available at *SSRN*: https://ssrn.com/abstract=4474908

## Auto Insurance Fraud Detection: Leveraging Cost Sensitive and Insensitive Algorithms for Comprehensive Analysis

**Meryem Yankol-Schalck, IPAG Business School (presenter)**

**Abstract**: As technology and the economy continue to grow, fraud has a significant negative impact on business and society, and insurance fraud remains an important issue, posing challenges in both detection and prevention. This article presents a direct cost-sensitive learning approach to improving traditional motor insurance fraud detection by using real-world data sets. In this approach, the results are obtained by using the information available at claim opening, FNOL. The dataset (FNOL) includes numeric, categorical and textual variables. Our study provides a comprehensive review of established econometric methods and machine learning techniques used in insurance fraud prevention.Various machine learning techniques are applied to automobile insurance fraud detection, including Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), Neural Network (NN), Extreme Gradient Boosting (XGB), Cost-Sensitive Decision Tree (CSDT), Cost-Sensitive Random Forest (CSRF), Cost-Sensitive Bayesian (CSBC) and Cost-Sensitive XGBoost (CS-XGB).However, empirical evidence suggests that machine learning classifiers outperform traditional econometric methods. It is important to understand that the success of machine learning techniques in detecting insurance fraud is highly dependent on the quality of the data and the specific context of the insurance industry. The results show that machine learning techniques perform better statistically and can also be more effective than standard approaches in reducing fraud-related costs. Furthermore, for most insurance companies, reducing the cost of the detection process is often a higher priority than reducing the classifier error rate. Our study confirms that Extreme Gradient Boosting (XGB) outperforms both cost-sensitive and cost-insensitive approaches based on performance measures.\While the inclusion of textual variables does not improve the statistical predictive performance of insurance fraud detection models, an in-depth study of agent text variables that highlight claim specificities is essential.

**Keywords**: Fraud detection; Automobile insurance; Cost Sensitive and Insensitive Algorithms, Natural language processing

### References

1. Banulescu-Radu, D., & Yankol-Schalck, M. (2023). Practical guideline to efficiently detect insurance fraud in the era of machine learning: A household insurance case. Journal of Risk and Insurance.
2. Yankol-Schalck, M. (2022). The value of cross-data set analysis for automobile insurance fraud detection. Research in International Business and Finance, 63, 101769.
3. Schalck, C., & Yankol-Schalck, M. (2021). Predicting French SME failures: new evidence from machine learning techniques. Applied Economics, 53(51), 5948-5963.

## Using Random Portfolios for Managing and Assessing Insurance Risks

**Mick Cooney, Describe Data**

**Abstract**: Random portfolios are a well established approach to assessing investment strategies. The concept is to randomly construct portfolios of risk within a given set of risk and size constraints. The performance of those randomised portfolios is compared to the actual portfolio to help discriminate returns derived from risk tolerance and those from risk selections. In this talk we discuss the use of this techniques for portfolios of insurance risks, and how the results can be used to for business planning, exposure management, risk optimisation and performance forecasting.

**Keywords**: Randomised Portfolios, Exposure Management, Business Planning

### References

1. Burns, Patrick (2009) Random Portfolios: Practice and Theory http://past.rinfinance.com/agenda/2009/randport_practice_theory_annotated.pdf

## Estimation subject to reporting delays and incomplete event adjudication with applications to disability insurance

**Kristian Buchardt, AP Pension**

**Christian Furrer, University of Copenhagen**

**Oliver Lunding Sandqvist, University of Copenhagen & PFA Pension (presenter)**

**Abstract**:

Estimating pricing and reserving models based on individual subject data can lead to more precise predictive models due to the use of more granular information. However, in order to use the most recent data and capture trends in a timely fashion, one has to adjust for the presence of incurred-but-not-reported (IBNR) and reported-but-not-settled (RBNS) claims as they lead to biased sampling. How to incorporate such sampling effects has recently been studied for Cox and Poisson models in Badescu et al. (2019) and Crevecouer et al. (2023), respectively, which both rely on the EM-algorithm for estimation.

In this talk, I will discuss how to accommodate IBNR and RBNS claims in the estimation of a general multistate model using a two-step estimation procedure and the asymptotics of the resulting estimator. The practical relevance is illustrated via a numerical study as well as via a disability insurance data application for a large Danish insurance portfolio that has been anonymized and slightly altered so as not to reveal any confidential information about the individual subjects or the insurance portfolio.

**Keywords**: Event history analysis, IBNR, RBNS, two-step M-estimation

### References

1. Badescu, A.L., Chen, T., Lin, X.S., Tang, D. (2019). A marked Cox model for the number of IBNR claims: Estimation and application. *ASTIN Bulletin: The Journal of the IAA* **49/3**, 709-739.
2. Crevecoeur, J., Antonio, K., Desmedt, S., Masquelein, A. (2023). Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims. *ASTIN Bulletin: The Journal of the IAA* **53/2**, 185-212.

## Computing capital requirements with guarantees

**Patrick Cheridito, ETH Zurich (presenter)**

**Moritz Weiss, ETH Zurich**

**Abstract**: We propose a deep learning method for computing capital requirements. It is based on a neural network approximation of the loss at the risk horizon together with a dual bound, which allows us to derive estimates of the accuracy of our numerical approximation.

**Keywords**: Capital requirements, VaR, expected shortfall, neural networks

## Reducing the dimensionality and granularity in hierarchical categorical variables

**Paul Wilsens, KU Leuven (presenter)**

**Katrien Antonio, KU Leuven, University of Amsterdam**

**Gerda Claeskens, KU Leuven**

**Abstract**: Hierarchical categorical variables often exhibit many levels (high granularity) and many classes within each level (high dimensionality). This may cause overfitting and estimation issues when including such covariates in a predictive model. In current literature, a hierarchical covariate is often incorporated via nested random effects. However, this does not facilitate the assumption of classes having the same effect on the response variable. In this paper, we propose a methodology to obtain a reduced representation of a hierarchical categorical variable. We show how entity embedding can be applied in a hierarchical setting. Subsequently, we propose a top-down clustering algorithm which leverages the information encoded in the embeddings to reduce both the within-level dimensionality as well as the overall granularity of the hierarchical categorical variable. In simulation experiments, we show that our methodology can effectively approximate the true underlying structure of a hierarchical covariate in terms of the effect on a response variable, and find that incorporating the reduced hierarchy improves model fit. We apply our methodology on a real dataset and find that the reduced hierarchy is an improvement over the original hierarchical structure and reduced structures proposed in the literature.

**Keywords**: hierarchical categorical variable, entity embedding, clustering, predictive modelling, machine learning

### References

1. Antonio, K. and Campo, B. D. (2023). On clustering levels of a hierarchical categorical risk factor. Annals of Actuarial Science.
2. Guo, C. and Berkhahn, F. (2016). Entity embeddings of categorical variables. arXiv preprint arXiv:1604.06737.
3. Carrizosa, E., Mortensen, L. H., Morales, D. R., and Sillero-Denamiel, M. R. (2022). The tree based linear regression model for hierarchical categorical variables. Expert Systems with Applications, 203:117423.

## Unsupervised Learning for Efficient Underwriting

**Elena Dalla Torre, Linköping University (co-presenter)**

**Per Wilhelmsson, Länsförsäkringar (co-presenter)**

**Abstract**: In the field of actuarial science, statistical methods have been extensively studied to estimate the risk of insurance. These methods are good at estimating the risk of typical insurance policies, as historical data is available. However, their performance can be poor on unique insurance policies, which require the manual assessment of an underwriter. A classification of insurance policies on a unique/typical scale would help insurance companies allocate manual resources more efficiently and validate the goodness of fit of the pricing models on unique objects. The aim of this thesis is to use outlier detection methods to identify unique non-life insurance policies. First, we use the principal component analysis as a baseline model to map insurance policies on the uniqueness scale. Then we see whether a better mapping can be achieved using autoencoders which can capture complex nonlinearities. The two methods are compared by performing a simulation study. The many categorical features present in insurance polcy data sets represent a challenge when applying outlier detection methods. In this thesis, we also explore different ways to derive informative numerical representations of categorical features. We compare the numerical representation obtained with the multiple correspondence analysis with the one obtained through joint embeddings learned as part of an autoencoder.

**Keywords**: Datadriven Underwriting, Outlier Detection, Autoencoders, Principal Component Analysis, Representation Learning

### References

1. Esbjörn Ohlsson and Björn Johansson. *Non-Life Insurance Pricing with Generalized Linear Models*. 01 2010.
2. Charu C. Aggarwal. *Outlier Analysis*. Springer, 2013.
3. Kishan G. Mehrotra, Chilukuri K. Mohan, and HuaMing Huang. *Anomaly Detection Principles and Algorithms*. Springer Publishing Company, Incorporated, 1st edition, 2017.
4. Christopher Blier-Wong, Jean-Thomas Baillargeon, Hélène Cossette, Luc Lamontagne, and Etienne Marceau. *Rethinking representations in p&c actuarial science with deep neural networks*, 2021
5. Delong Ł, Kozak A. *The use of autoencoders for training neural networks with mixed categorical and numerical features*. ASTIN Bulletin. 2023;53(2):213-232. doi:10.1017/asb.2023.15

## Challenges in Actuarial Learning for Loss Modeling of Brazilian Soybean Crops

**Rodrigo Targino, Getulio Vargas Foundation (presenter)**

**Abstract**: In Brazil, the agricultural sector plays a fundamental role, accounting for approximately 25% of the national Gross Domestic Product (GDP). Within this sector, soybean cultivation stands out as the most significant, with a forecast yield exceeding 150 million tons in the 2022/2023 harvest, which is equivalent to the combined production of all Organisation for Economic Co-operation and Development (OECD) nations. Nevertheless, soybean production is subject to considerable climatic vulnerabilities that not only affect the yield but also have a considerable impact on the formulation of crop insurance pricing. The consequences of drought and excessive rainfall in Brazil's South and Central-West regions severely affected the crop yield of 2021/2022, leading to a substantial rise in the loss ratio of insurer's portfolios. These climatic adversities led to a substantial rise in the loss ratio of insurers' portfolios. In this paper, we utilize actuarial learning models to predict losses in soybean crops, taking into account weather covariates, specifically rainfall, and temperature, given their relevance to crop productivity. Both generalized linear models and random forest have been benchmarked to assess their predictive performance within the two-stage frequency-severity framework. This framework comprises a model dedicated to claim occurrence and a second model focused on determining the severity of individual claims. We also compare the charged premium with the estimated premium by using standard insurance modeling techniques. Additionally, we carry out simulated stress tests to highlight the significance of weather-related variables in evaluating total losses during extreme events. Our assessment encompasses the estimation of losses across diverse scenarios. The implications of these results are significant for (re)insurance pricing, risk management, and solvency. They also bear substantial importance for formulating effective agricultural public policies.

## Claims modelling with climate data

**Ronald Richman (presenter), Old Mutual Insure, University of the Witwatersrand**

**Abstract**: Traditional short-term (or property and casualty) insurance pricing involves the use of statistical techniques such as Generalised Linear Models and policyholder specific rating factors to estimate the expected claims cost associated with a relevant risk profile. In recent years, the short-term insurance industry has been increasingly impacted by weather related events, which are expected to be exacerbated by both climate change and meteorological phenomena such as El Niño. Consideration of this risk has traditionally been captured in the actuarial pricing process indirectly through the use of geographical features such as area codes. In this talk, we will directly link an exposure dataset of buildings risk and associated claims experience to a high-resolution gridded precipitation dataset to investigate the predictive power of this feature on both an actual and forecasted basis. We establish a modelling framework using machine learning that allows for the estimation of both the frequency and severity of buildings claims given the combined dataset. We consider the relative importance of the added precipitation feature against traditionally used rating factors and the sensitivity of the underlying claims frequency and severity to changes in precipitation. Finally, by considering different precipitation scenarios, we show how the risks of excessive precipitation can be quantified, allowing for more accurate forecasts of financial performance to be made, and risk mitigation strategies to be investigated.

**Keywords**: Insurance pricing, climate change, natural perils

## Multiple Yield Curve Modeling and Forecasting using Deep Learning

**Ronald Richman, Old Mutual Insure and University of the Witwatersrand, Johannesburg, South Africa**

**Salvatore Scognamiglio, University of Naples Parthenope (presenter)**

**Abstract**: In this work, we develop deep learning models that simultaneously model and forecast the dynamics of the multiple yield curves, which could be related to different countries, credit qualities or liquidity characteristics; here, we focus on the first two of these. We aim to learn the dependence structure among the different yield curves induced by the globalization of financial markets and exploit it to produce more accurate predictions. By combining the self-attention mechanism [1] and nonparametric quantile regression, our model generates both point and interval forecasts of future yields. The architecture is designed to avoid quantile crossing issues affecting multiple quantile regression models. Numerical experiments conducted on two different datasets confirm that the proposed approach produces more accurate results with respect to some Nelson-Siegel type models [2]. Finally, we explore potential extensions and enhancements by incorporating deep ensemble methods [3] and transfer learning mechanisms.

**Keywords**: Deep Learning, Multiple Yield Curve modeling, Nelson-Siegel model, Attention Models, Transfer Learning, Value-at-Risk, Solvency II, IFRS 17.

### References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
2. Diebold, F. X., & Rudebusch, G. D. (2012). *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach.* Economics Books.
3. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30.

## Modeling lower-truncated and right-censored insurance claims with an extension of the MBBEFD class

**Selim Gatti, ETH Zurich (presenter)**

**Mario Wüthrich, ETH Zurich**

**Abstract**: In general insurance, claims are often lower-truncated and right-censored because insurance contracts may involve deductibles and maximal covers. Most classical statistical models are not (directly) suited to model lower-truncated and right-censored claims. A surprisingly flexible family of distributions that can cope with lower-truncated and right-censored claims is the class of MBBEFD distributions that originally has been introduced by Bernegger (1997) for reinsurance pricing, but which has not gained much attention outside the reinsurance literature. We derive properties of the class of MBBEFD distributions, and we extend it to a bigger family of distribution functions suitable for modeling lower-truncated and right-censored claims. Interestingly, in general insurance, we mainly rely on unimodal skewed densities, whereas the reinsurance literature typically proposes monotonically decreasing densities within the MBBEFD class.

**Keywords**: General insurance claims, lower-truncation, right-censoring, MBBEFD distribution, unimodal density, skewed density, normalized loss, exposure curve, Swiss Re exposure curve, Lloyd's exposure curve

### References

1. Bernegger, S. (1997). The Swiss Re exposure curves and the MBBEFD distribution class. *ASTIN Bulletin* **27/1**, 99-111.
2. Fung, T.C., Badescu, A.L., Lin, X.S. (2022). Fitting censored and truncated regression data using the mixture of experts models. *North American Actuarial Journal* **26/4**, 496-520.
3. Poudyal, C., Brazauskas, V. (2022). Robust estimation of loss models for truncated and censored severity data. *Variance* **15/2**.
4. Verbelen, R., Gong, L., Antonio, K., Badescu, A., Lin, S. (2015). Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. *ASTIN Bulletin* **45/3**, 729-758.
5. Wüthrich, M.V., Merz, M. (2023). *Statistical Foundations of Actuarial Learning and its Applications*, Springer Actuarial.

## Modeling Mortality Rates: Nonparametric Bayesian Inference with Gaussian Cox Processes

**Soroush Amirhashchi, Plannet Insurtech Hub (presenter)**

**Khaled Masoumifard, Plannet Insurtech Hub**

**Abstract**:

The modeling of mortality rates has been extensively explored from various perspectives. This research specifically focuses on mortality rates, utilizing the Cox Process as a foundational model.

To enhance our methodology, we employ a valuable technique that represents the intensity of the Cox Process through the transformation of a random realization from a Gaussian Process (GP), known as the Gaussian Cox Process (GCP).

The inference based on likelihood in these models leads to intractability due to integrals over an infinite-dimensional random function. To make the inference more manageable, we drew inspiration from techniques discussed in [1, 2] and applied the model to a mortality database sourced from the Human Mortality Database.

**Keywords**: Mortality Rate, Gaussian Cox Process, Bayesian Inference

### References

1. Adams, Ryan Prescott, Iain Murray, and David JC MacKay. "Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities." Proceedings of the 26th annual international conference on machine learning. 2009.
2. Lloyd, Chris, et al. "Variational inference for Gaussian process modulated Poisson processes." International Conference on Machine Learning. PMLR, 2015.

## Enabling Advanced Analytic Capabilities by Multiplying Internal Insurance Data

**Sven Haadem, phd. Aeda (presenter)**

**Are Flataker Viken, Aeda**

**Abstract**: The ever-growing field of data science hinges on the availability of high-quality data. However, many organisations struggle with limited data volume and poor data quality, hindering their ability to extract valuable insights. Aeda addresses this challenge by introducing a novel data sharing platform. This platform aggregates data from insurance companies and additional sources, implements data cleaning techniques and enrich and anonymize it before distributing it out amongst the companies. This provides everyone involved with a multiple on their data according to the quality and usefulness of the data. In addition we provide methods to monitor the quality and state of data and give detailed and dynamic reports from advanced ml and generativ models. The internally developed anonymization and distribution algorithm is based on techniques such as clustering, distribution estimation, optimal allocation, and data distrubution, quality, and sliding monitoring using ml.

**Keywords**: Data Sharing Platform, Data Acquisition, Data Cleaning, Data Quality, Data Analytics

### References

1. Lin, Jun-Lin & Cheng, Meng. (2008). An efficient clustering method for k-anonymization. 46-50. 10.1145/1379287.1379297.

## Fair learning and testing for unfairness given protected features

**Munir Eberhardt Hiabu, University of Copenhagen**

**Niklas Andreas Pfister, University of Copenhagen**

**Tessa Steensgaard, University of Copenhagen (presenter)**

**Abstract**: In the presence of a set of known protected features, we say that a predictor is fair if does not use, directly or indirectly, the effect of the protected features on the response. Within a causal framework, we show that this corresponds to a predictor that measures the average direct effect of the unprotected features on the response. This type of estimator is relevant in non-life insurance, where it is essential to remove discriminatory effects from prices.

In this talk, we accomplish two tasks:
(a) We propose to derive a fair predictor via the following steps. One first estimates the direct effect of the protected features using double machine learning. Subsequently, this estimator is used to remove the effect of the protected feature on the response variable. Finally, any learning algorithm can be applied on the adjusted data. We conduct a small sample study that shows promising results.
(b) In the second part, we introduce a measure of unfairness and develop a test that asymptotically has the right coverage. It takes any predictor as input and tests for unfairness on the original data.

**Keywords**: fairness, testing, causal model, double machine learning, non-life insurance

## Cyber risk modeling using a two-phase Hawkes process with external excitation

**Yousra Cherkaoui, CREST-ENSAE; (Presenter)**

**Alexandre Boumezoued, Milliman R&D**

**Caroline Hillairet, CREST-ENSAE**

**Abstract**:

With the growing digital transformation of the worldwide economy, cyber risk has become a major issue. As 1% of the world's GDP (around $1,000 billion) is allegedly lost to cybercrime every year, IT systems continue to get increasingly interconnected, making them vulnerable to accumulation phenomena that undermine the pooling mechanism of insurance. As highlighted in the literature, Hawkes processes appear to be suitable to capture contagion phenomena and clustering features of cyber events. This paper extends the standard Hawkes modeling of cyber risk frequency by adding external shocks, such as the publication of cyber vulnerabilities that are deemed to increase the likelihood of attacks in the short term. While the standard Hawkes model attributes all clustering phenomena to self-excitation, this paper introduces a model designed to capture external common factors that may explain part of the systemic pattern. This aims to provide a better quantification of contagion effects. We propose a Hawkes model with two kernels, one for the endogenous factor (the contagion from other cyber events) and one for the exogenous component (cyber vulnerability publications). We use parametric exponential specifications for both the internal and exogenous intensity kernels, and we compare different methods to tackle the inference problem based on public datasets containing features of cyber attacks found in the Hackmageddon database and cyber vulnerabilities from the Known Exploited Vulnerability database and the National Vulnerability Dataset. By refining the external excitation database selection, the degree of endogeneity of the model is nearly halved. We illustrate our model with simulations and discuss the impact of taking into account the external factor driven by vulnerabilities. Once an attack has occurred, response measures may be implemented to limit the effects of an attack. These measures include patching vulnerabilities and reducing the attack's contagion. We use an augmented version of the model by adding a second phase modeling a reduction in the contagion pattern from the remediation measures. Based on this model, we explore various scenarios and quantify the effect of mitigation measures of an insurance company that aims to mitigate the effects of a cyber pandemic in its insured portfolio. The paper is available here : https://arxiv.org/abs/2311.15701.

**Keywords**: Cyber risk, cyber insurance, Hawkes process, external excitation, exogenous factors, cyber vulnerabilities, cyber pandemic, reaction measures.

## Generating individual claims using Generative Adversarial Networks

**Yves-Cédric Bauwelinckx, KU Leuven (presenter)**

**Emmanuel Jordy Menvouta, KU Leuven**

**Jan Dhaene, KU Leuven**

**Tim Verdonck, University of Antwerp**

**Milan van den Heuvel, UGent**

**Abstract**: This paper addresses the challenge of generating synthetic data for individual claims reserving in the insurance sector, a critical area where the availability of large, detailed datasets is a limiting factor for research and development, especially for machine learning (ML) based models. Traditional simulation engines, while useful, rely heavily on assumptions and complex architectures, making them less adaptable and potentially less accurate for various data scenarios. In response, we propose a the use of a more data-driven approach utilizing generative adversarial networks (GANs), a methodology that has shown remarkable success in generating realistic imagery and has been increasingly applied to numerical and tabular data generation. Our approach leverages recent advancements in generative modeling to produce high-quality synthetic datasets. These datasets can closely mimic real-world data in structure and distribution, making them suitable for back-testing new claims reserving methods. Our methodology requires fewer assumptions, offers flexibility in handling different data types and distributions, and can be easily adapted to new datasets with varying variables and structures. We discuss the theoretical framework of GANs, describe our data structure and features, and present our findings, demonstrating the potential of our approach to significantly contribute to the field of actuarial science and claims reserving research.

**Keywords**: Synthetic data, Generative AI, Generative Adversarial Networks, Individual Claim Reserving

## Index of presenters