DEPARTMENT OF STATISTICAL & ACTUARIAL SCIENCES

Master's Day

July 28, 2023

19th Annual Master's Day

Contents

Schedule of Events

Location: Western Science Centre Room 240

8:45 a.m. Student and Faculty - Coffee and Mingle

9:10 a.m. Opening Remarks, Dr. Jiandong Ren (Graduate Chair)

9:15 a.m. Session 1

Dr. Ricardas Zitikis
1) **Zhiwen Dong (MSc FM)**

Dr. Ricardas Zitikis
2) **Azar Eftekhari Targhi (MSc FM)**

Dr. Ricardas Zitikis
3) **Karzan Saeedi (MSc FM)**

Dr. Serge Provost/Dr. Rogemar Mamon
4) **Zhuangqing (Skyler) Li (MSc FM)**

10:15 a.m. Break and set up for Session II

10:30 a.m. Session II

Dr. Hyukjun Gweon
5) **Ruimin (Jasmine) Gao (MSc SS)**

Dr. Hyukjun Gweon
6) **Jingyi Liu (MSc SS)**

Dr. Hyukjun Gweon/Dr. Reg Kulperger/ Dr. Hao Yu
7) **Hongjun Sim (MSc SS)**

Dr. Simon Bonner
8) **Yize Yuan (MSc SS)**

11:30 a.m. Light Lunch

12:30 p.m. Session III

Dr. Marcos Escobar-Anel
9) **Ruoyan Gao (MSc FM)**

Dr. Marcos Escobar-Anel
10) **Xinkai Zhuang (MSc FM)**

Dr. Marcos Escobar-Anel
11) **Jueyi Shao (MSc FM)**

Dr. Marcos Escobar-Anel
12) **Yifan You (MSc FM)**

13:30 p.m. Break and set up for Session IV

13:45 p.m. Session IV

Dr. Lars Stentoft
13) **Kiera Sobolewski (MSc FM)**

Dr. Katsu Goda/Dr. Jiandong Ren
14) **Peiyu Huang (MSc AS)**

Dr. Matt Davison/Dr. Douglas Woolford
15) **Daniel Guerrero Santaren (MSc.T SS)**

14:30 p.m. Closing Remarks

Titles and Abstracts

Zhiwen Dong, Financial Modelling, supervised by Dr. Ricardas Zitikis

Title: A Monte Carlo Based Method for Estimating the Present Value of Annuities with Interest Rates Driven by Stochastic Processes

Annuities are mostly valued with constant interest rates, and in this case nice "academic" formulas are available for calculating the present value of annuities. However, in the real world, the interest rates are not fixed and may even change frequently. An algorithm for calculating the present value of annuities with stochastic interest-rate process is therefore needed. In our study, we introduce and explore a Monte Carlo based method for estimating the present value of annuities. The method can be applied to tackle very complicated models and, when needed, can be further generalized and improved from various perspectives, including accuracy.

Azar Eftekhari Targhi, Financial Modelling, supervised by Dr. Ricardas Zitikis

Title: An Application and Assessment of Multivariate GARCH Estimation of CoVaR for Systemic Risk Measurement

In this study, we delve into the idea of systemic risk by closely looking at a key tool called the Conditional Value-at-Risk (CoVaR). This tool is extremely useful in understanding the shared risk within the financial system. We thoroughly explore the tool, especially the use of a specific model (bivariate GARCH) to deal with changing correlations over time and for investigating severe financial distress events. Building on this, we follow the footsteps of researchers Giulio Girardi and A. Tolga Ergun who also tweaked the CoVaR for a better understanding of financial distress. In this study we focus on learning and applying these complex techniques on new data. The data are comprised of recent information from Canadian financial institutions, with the goal of understanding shared risk. This study is not just a learning exercise of contemporary and highly useful techniques but also offers insights into how to build stronger risk management strategies and boost the overall stability of financial systems.

Karzan Saeedi, Statistics, Financial Modelling, supervised by Dr. Ricardas Zitikis

Title: The Euler-based Capital Allocation Rule and its Compatibility with RORAC

This research studies capital allocations in financial institutions, focusing specifically on the Euler method. Our goal is to understand the Euler method and offer a guide to help financial businesses to use this method effectively.

When choosing an allocation method, several factors must be considered. These include the method's ability to fully allocate, its compatibility with performance measurements such as return on risk-adjusted capital calculations, and its contribution to diversification. Additionally, the method should be practical and easy to implement.

First, we start by looking closely at the basics of risk, risk management, and quantitative risk management. We also study the Basel Accords, which are agreements that set standards for banks. This gives us a strong understanding of the subject. Next, we examine the Return on Risk-Adjusted Capital (RORAC) and then the Euler method from a mathematical perspective. We investigate properties of risk measures under Euler's method. We then show that the Euler method works well with RORAC and the diversification index.

In the final part of our study, we use the Euler Capital Allocation method for a portfolio with risky positions, analyzing it with risk measures such as the Value at Risk, Expected Shortfall, and Standard Deviation. Based on a numerical example, we discuss and compare the results. We find that the Euler method meets all the criteria, including RORAC and diversification, and is also relatively simple to implement. As a result, we suggest using the Euler method in real-world applications.

Zhuangqing (Skyler) Li, Financial Modelling, supervised by Dr. Serge Provost, Dr. Rogemar Mamon

Title: Lending Club Data Analysis Based on Machine Learning Algorithm

This project aims to develop a predictive model for loan default using a dataset sourced from the Lending Club. The study employs Principal Component Analysis (PCA) to reduce the dimensionality of the dataset and utilizes three popular machine learning algorithms, namely XGBoost, logistic regression, and random forest, to construct predictive models. The objectives consist of evaluating the performance of these models and identifying the most effective one for predicting loan default based on various borrower attributes and credit-related factors. The analysis reveals that the loan status is particularly sensitive to three key variables: the borrower's FICO score range at loan origination, the total bank card credit limit, and the total credit balance excluding mortgage. These variables have a significant impact on the likelihood of loan default. The findings of this research ought to foster an improved appreciation of certain machine learning techniques in connection with the prediction of loan default. The selected model could assist financial institutions and lending platforms in making informed decisions regarding loan approval and risk assessment, ultimately aiding in reducing default rates and improving lending practices

Ruimin (Jasmine) Gao, Statistics, supervised by Dr. Hyukjun Gweon

Title: Label Powerset Classifier Chains for Multilabel Classification

In multi-label classification, each instance may be associated with multiple labels simultaneously. Multi-label data are typically modeled by some combinations of binary or multi-class models. One of the most well-known approaches for multi-label classification is Classifier Chain, in which label dependencies are addressed by a chain of sequential binary classifiers. In this project, we expand the idea of Classifier Chain and use a chain of segregated multi-class classifiers. The performance of the proposed method is examined under different parameter settings. We also compared the proposed method with other commonly used methods using some benchmark datasets.

Jingyi Liu, Statistics, supervised by Dr. Hyukjun Gweon, Dr. Hao Yu

Title: Investigating the Impact of Labeling Errors on Active Learning Performance in Time Series Classification

Active learning is a machine learning framework that can enhance prediction while minimizing the number of labelled instances required for model training. An active learning process interacts with an oracle (e.g., a human expert) to request the labelling of informative unlabeled instances. However, real-world applications often involve human experts prone to make labelling errors, affecting the predictive model's classification accuracy. This project aims to examine how labeling errors in active learning affect classification performance. Under different mislabeling scenarios, experimental results obtained from the ECG5000 and Wafer datasets demonstrate that a) the classification process is less sensitive to small labeling errors, b) when labeling errors are highly frequent, the performance of active learning deteriorates which is similar to that of passive learning approach, and c) the performance degradation due to labeling errors depends on the robustness of the classifier to noisy instance.

Hongjun Sim, Statistics, supervised by Dr. Hyukjun Gweon, Dr. Reg Kulperger, Dr Hao Yu

Title: Assessing Probabilistic Classifier Calibration: A Local Alternative Approach with Standardized Reliability Diagram

In supervised learning, a reliability diagram can be used to assess the quality of these probability estimates. It visualizes the level of calibration of the probability outcomes at a given probability interval based on data binning. However, determining the reliability of probability estimates remains challenging. A standardized reliability diagram can be employed to address this, providing an immediate visual interpretation of a probabilistic classifier. An approach to choosing the number of bins for the consistent assessment of the standardized reliability diagram involves finding an optimal number of bins, each with equal size, so that the power of each bin is close to our desired power under constant alternatives. However, the desired power may be underachievable or over-achievable if the variance in a bin is too large or too small, respectively. In this study, to overcome this limitation, we introduce a local alternative to study the power more finely. Using variable bin sizes based on the local alternative setup, we can achieve uniform desired power across all bins by adjusting the bin sizes so that the deviations from alternative to null are uniformly close to the same distances. A simulation study was conducted to assess the performance of the proposed local alternative setup across different model.

Yize Yuan, Statistics, supervised by Dr. Simon Bonner

Title: Accounting for sampling error in modelling the effect of the abundance of an invasive species of cattail on muskrat occupancy

Motivated by a study of the relationship between the abundance of specific species of cattail and muskrat occupancy, this paper addresses the problem of fitting logistic regression models when a covariate measuring a proportion is observed with error due to finite sampling. We aim to develop Bayesian methods to fit these models. We propose a two-component framework that first models the Cattails Sampling and then predicts Muskrat Presence. In the Cattails Sampling component, the binomial distribution is used to model the proportion of cattails of a specific species at each site. The Muskrat Presence component utilizes a logistic regression model to predict muskrat presence in grid cells based on the proportion of cattails of the species of interest. We fit the models in the Bayesian framework using a latent variable to connect the two components. We evaluate the new model's performance through a simulation study and compare it with estimation based on the model using the fixed covariate with measurement error. The model fit with the observed proportion exhibits higher bias in estimating the true parameter values. The posterior standard deviations of the model fit with observed proportion were generally lower than those of the model including the true proportion as a latent variable, indicating higher precision in parameter estimation. However, the coverage probability sometimes fell well below the nominal value, depending on the true parameter values. This shows that the proposed Bayesian approach, incorporating a latent variable that connects the Cattails Sampling and Muskrat Presence components to account for measurement error, improves parameter estimation compared to the model fit using the fixed covariate with measurement error. Our developed Bayesian framework enables parameter estimation, effectively handles measurement errors, and provides reliable predictions for logistic regression models with sampling errors of the covariate proportion. We believe these methods are applicable not only in ecology but also in many other fields, including health and medicine (medical diagnosis) and social science (election polling.

Ruoyun Gao, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: ESG Rating and Performance Measures

ESG acronym stands for environment, social and governance. The ESG rating system has been an increasingly discussed topic as it provides investors, stakeholders, and consumers with a non-financial perspective about a company. However, while ESG rating system offers a measurement of companies' long-term sustainability in various perspectives and is used as a tool by many people for making investment decisions and corporate assessments, the relationship between a company's ESG score and its return remains to be studied. In this project, our goal is to explore how ESG scores and companies' stock returns jointly behave.

Using RepRisk ESG ratings, we choose a total of 15 companies from different ESG rating levels: A, B, and C. Each level contains five companies. We evaluate the companies' risk-return performances using four performance measures: Sharpe Ratio, Omega Ratio, Sortino Ratio, and Treynor Ratio. Finally, our results indicate that companies with higher ESG ratings do not necessarily exhibit better performance compared to those with lower ESG ratings. Specifically, the impact of ESG rating on companies' risk-return performances is found to be minimal. Additionally, we analyze the pattern of RepRisk ESG ratings over time using time series models to provide further insights into their behavior confirming autoregressive with no GARCH residuals.

Xinkai Zhuang, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: Comparative Analysis of Characteristics and Prepayment Modeling for Mortgage-Backed Securities: A Focus on NHA MBS in Canada

Mortgage-backed securities (MBS) are debt obligations that represent claims to the cash flows from pools of mortgage loans, most commonly on residential property. While previous studies have primarily focused on MBS in the United States, limited research has been conducted on Canadian MBS (NHA MBS). To address this gap, this project compares mortgage characteristics in the United States and Canada, highlighting the unique term and distinct prepayment behavior of Canadian mortgages. Addressing the risk associated with MBS prepayment behavior, we develop a more suitable piecewise linear model for estimating prepayment rates specifically for NHA MBS, utilizing historical data. Additionally, this project presents a pricing methodology for MBS, akin to bond pricing, considering interest rate term structures and credit spreads. The proposed approach is specifically tailored to the Canadian market and offers practical implications for MBS valuation.

Jueyi Shao, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: Tackling Non-modellable Risk Factors with Variations of Linear Regressions

This paper addresses the challenge of quantifying non-modellable risk factors (NMRFs) in regulatory capital requirements for financial institutions, a process critical to robust financial risk management. Traditionally, NMRFs have been approximated with proxy variables, but this practice is no longer deemed acceptable by current banking regulations according to the fundamental reviews of the trading book. In this work, we propose a methodology to decompose NMRFs using multivariate regression and adding a covariance/correlation term in to the objective values of the optimization process, hence reducing the bias in capital requirement estimations.

Our study focuses on the stock returns of Apple, Google, and Amazon as NMRFs, utilizing the return of the SP500 index as the modellable risk factor. The proposed approach modifies the traditional multivariate regression objective function to minimize the residuals' covariance, thereby leading to representation of NMRF in tune with regulations.

To validate our methodology, we experiment with eight different models, each with the aim to capture the systematic, predictable variation in the NMRFs, referred to as the 'Modellable Component' of the risk. The ultimate goal is to minimize the residual correlation across different NMRFs, thus preventing the overestimation of the residual component in our decomposition, which could otherwise lead to inaccurate capital requirement assessments. This research contributes a potentially more accurate and regulation compliant approach to the complex task of capital requirement estimation, mitigating the risk of under- or over-estimation resulting from nonmodellable risk factors.

Yifan You, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: Non-Modelable Risk Factors and Residual Correlation

This study investigates the relationship between non-modelable risk factors (NMRF) and modelable risk factors (MRF) with the primary objective of reducing residual correlations among relevant NMRFs. For practicing purposes, the NMRFs considered are the implied volatilities of Apple (APPL), Amazon (AMZN), and Google (GOOGL), while the $S\&P$ 500's implied volatility serves as the modelable risk factor (MRF). Three distinct modeling methods are utilized for achieving the objectives: linear regression, quantile regression (median quantile regression), and random forest.

In the contexts of linear regression and quantile regression, this project assumes four general settings: linear, polynomial, logarithmic, and exponential functions, all linear in the parameters. On each setting, any reduction in the correlation of residuals is reported. The random forest approach involves dividing the dataset into multiple groups to enhance the modeling process. Group selection is determined based on the date and the attributes of the risk factors, such as Moneyness and Maturity time. Within the random forest framework, the four functions mentioned above are considered for each regression tree at each specific group.

From the current results, none of the three methods and four functionals successfully achieved the goal of significantly and consistently reducing the correlation in the residuals across. The lack of desirable outcomes may be attributed to the limited availability of independent (MRF) variables. Future research should explore other MRF, particularly in the groups where one MRF is not enough.

Kiera Sobolewski, Financial Modelling, supervised by Dr. Lars Stentoft

Title: Finance and Data Analytics Internship at SE Health

Currently, I am completing my internship as a Finance and Data Analytics intern at SE Health. SE Health is a not-for-profit social enterprise that provides health care services across many areas of the health system including home care, long-term care and acute care. In this presentation, I will go over projects I have been working on, why they are important and how they contribute to the company. Projects include ramp time, looking at number of beds vs. number of admissions, patient segmentation and patient population statistics effects on service leve.

Peiyu Huang, Actuarial Science, supervised by Dr. Katsu Goda, Dr. Jiangdon Ren

Title: Pricing the Catastrophe risk bonds for household seismic losses in Victoria, BC

Catastrophe events such as earthquakes, hurricanes, floods, can cause huge losses to insurance companies. An important tool is to transfer part of the risks to financial markets by issuing catastrophe risk bonds (CAT bonds). The CAT bonds can provide a means for investors to diversify their investment portfolio because their returns are uncorrelated with broader financial markets. In this project, based on the simulated losses for a portfolio of residential wooden houses in Victoria, BC, we calculate the price of CAT bond with different maturity and triggering points. The prices are compared with regular bond to highlight the effect of the catastrophe risk.

Daniel Guerrero Santaren, Statistical Science, supervised by Dr. Matt Davison, Dr. Douglas Woolford

Title: Exploring the Use of Real Options in the Wildfire Contex

Wildfires are a pressing concern in Canada, with an estimated annual burned area that, historically, is roughly equivalent to the area of India. While most wildfires are effectively managed, the potential impact of catastrophic events or the secondary impacts of fires, such as smoke, cannot be ignored. Given the uncertainty surrounding wildfire spread, the adoption of robust decision-making frameworks becomes crucial. This presentation focuses on the practical application of the real options approach to wildfire evacuations, recognizing the uncertainties involved and emphasizing the value of gathering information before making critical decisions. Specifically, we explore the use of a "wait-and-see" strategy in the context of real options for decision-makers considering if and when to evacuate. This presentation will provide an overview of the work done so far and discuss ongoing research directions.