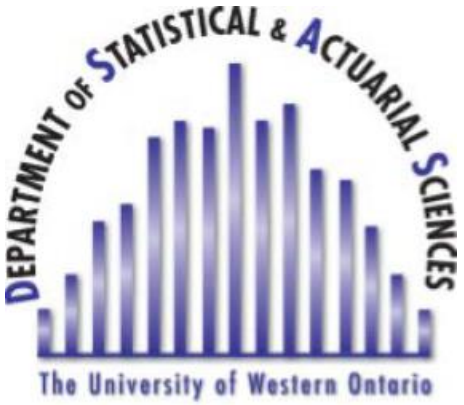
DEPARTMENT OF STATISTICAL & ACTUARIAL SCIENCES

Master's Day

July 30, 2020



16th Annual Master's Day

Contents

Abstracts 4

Poster Session One: Dr. Marcos Escobar-Anel 4

Yizhou Cai, Financial Modelling, supervised by Dr. Hristo Sendov 4

Madhusoodan Gunasingam, Financial Modelling, supervised by Dr. Hristo Sendov 4

Reza Mosayebi Kulluje, Financial Modelling, supervised by Dr. Ricardas Zitikis 5

Shengying Yang, Financial Modelling, supervised by Dr. Mark Reesor 5

Sahab Zandi, Financial Modelling, supervised by Dr. Cristian Bravo Roman 6

Junsen Zhang, Financial Modelling, supervised by Dr. Matt Davison 6

Xiaoyu Zhang, Financial Modelling, supervised by Dr. Matt Davison 7

Dongfang Zhao, Financial Modelling, supervised by Dr. Ricardas Zitikis 7

Xiyuan Zheng, Financial Modelling, supervised by Dr. Rogemar Mamon 8

Poster Session Two: Dr Ricardas Zitikis 9

Jonathan Jones, Financial Modelling, supervised by Dr. Rogemar Mamon 9

Javad Kazemian, Financial Modelling, supervised by Dr. Marcos Escobar-Anel 9

Xinxin Li, Financial Modelling, supervised by Dr. Lars Stentoft 10

Junjun Liu, Financial Modelling, supervised by Dr. Lars Stentoft 10

Yuhong Peng, Financial Modelling, supervised by Dr. Marcos Escobar-Anel 11

Amirreza Seddighi, Financial Modelling, supervised by Dr. Cristian Bravo Roman 11

Li Zhang, Financial Modelling, supervised by Dr. Marcos Escobar-Anel 11

Poster Session Three: Dr. Xiaoming Liu 12

Yunfei Du, Actuarial Science, supervised by Dr. Shu Li 12

Wei Li Fan, Financial Modelling, supervised by Dr. Rogemar Mamon 12

Dechen Gao, Actuarial Science, supervised by Dr. Kristina Sendova 13

Yue Gu, Financial Modelling, supervised by Dr. Ricardas Zitikis 13

Yuwei Huang, Actuarial Science, supervised by Dr. Kristina Sendova 14

Duo Xu, Actuarial Science, supervised by Dr. Shu Li 14

Poster Session Four: Dr. Wenqing He 15

Shiyu He, Statistical Science, supervised by Dr. Doug Woolford and Dr. Reg Kulperger 15

Xingfu Zhang, Statistical Science, supervised by Dr. Serge Provost and Dr. Hyukjun Gweon 15

Manjia Zhao, Statistical Science, supervised by Dr. Doug Woolford 16

Poster Session Five: Dr. Simon Bonner 17

Gansen Deng, Statistical Science – Biostats, supervised by Dr. Wenqing He 17

Fatemeh Gholizadeh, Statistical Science, supervised by Dr. Camila de Souza 17

Chufan Wu, Statistical Science, supervised by Dr. Hao Yu and Dr. Hyukjun Gweon 18

Chengqian Xian, Statistical Science, supervised by Dr. Camila de Souza 18

Schedule of Events

9:00 a.m. Student and Faculty Zoom Log-in Now Open – *Master's Day – Open/Close*

9:15 a.m. Opening Remarks, Marcos Escobar-Anel (Graduate Chair)

9:30 a.m. Proceed to 2nd Zoom Link for Master's Day Presentations

Poster Session One: Dr. Marcos Escobar-Anel
Zoom Link

- Yizhou Cai
- Madhusoodan Gunasingam
- Reza Mosayebi Kullije
- Shengying Yang
- Sahab Zandi
- Junsen Zhang
- Xiaoyu Zhang
- Dongfang Zhao
- Xiyuan Zheng

Poster Session Two: Dr. Ricardas Zitikis
Zoom Link

- Jonathan Jones
- Javad Kazemian
- Xinxin Li
- Junjun Liu
- Yuhong Peng
- Amirreza Seddighi
- Li Zhang

Poster Session Three: Dr. Xiaoming Liu
Zoom Link

- Yunfei Du
- Dechen Gao
- Wei Li Fan
- Yue Gu
- Yuwei Huang
- Duo Xu

Poster Session Four: Dr. Wenqing He
Zoom Link

- Shiyu He
- Xingfu Zhang
- Manjia Zhao

Poster Session Five: Dr. Simon Bonner
Zoom Link

- Gansen Deng
- Fatemeh Gholizadeh
- Chufan Wu
- Chengqian Xian

12:00 p.m. Completion of Master's Day Sessions

12:30 p.m. Rejoin *Master's Day – Open/Close* Zoom for closing remarks and Awards, Graduate Chair

Abstracts

Poster Session One: Dr. Marcos Escobar-Anel

Zoom Link

Yizhou Cai, Financial Modelling, supervised by Dr. Hristo Sendov

Title: Optimal Gambling Systems and Application"

Abstract: In our report, we introduce a general system which can explain various forms of gambling systems. We derive an alternative form of the arbitrage theorem for the system which holds in greater generality. Defining gamble strategy and growth rate, we prove an important theorem regarding the expected growth rate by stating series of propositions. Afterward, we use coin toss as an example to see how the system and theory works in specific cases.

Madhusoodan Gunasingam, Financial Modelling, supervised by Dr. Hristo Sendov

Title: Optimal Gambling Systems and Application"

Abstract: In our report, we introduce a general system which can explain various forms of gambling systems. We derive an alternative form of the arbitrage theorem for the system which holds in greater generality. Defining gamble strategy and growth rate, we prove an important theorem regarding the expected growth rate by stating series of propositions. Afterward, we use coin toss as an example to see how the system and theory works in specific cases.

Reza Mosayebi Kulluje, Financial Modelling, supervised by Dr. Ricardas Zitikis

Title: Empirical Assessment of GARCH and Deep Learning Option Pricing Methods

Abstract: Derivative pricing including option pricing is a challenging field in financial modeling. Since the introduction of Generalized Autoregressive Conditional Heteroscedastic (GARCH) process, GARCH models have been widely used in the option pricing because of their ability in modeling time-varying volatility and volatility clustering. Through the years, they have become more and more sophisticated by incorporating different aspects present in the dynamics of the market.

In this project, a comparison has been made on the out-of-sample empirical performance of three well recognized GARCH models that incorporate volatility asymmetry differently (the model by Glosten-Jagannathan-Runkle (GJR-GARCH), the model by Engle and Ng (NGARCH) and the model proposed by Heston and Nandi). Model parameters were estimated using Maximum Likelihood Estimation using daily return series of SP500. European option valuations were done by incorporating Duan-Simonato modification in Monte Carlo simulations using risk-neutralized parameters. To reduce the computations variance, a control variate technique was also used.

Another option pricing approach widely under attention is the use of Neural Networks. Since early 1990s, they have been used as a nonparametric method for option pricing. In this work, a data driven approach is used by simulating Black-Scholes option prices based on random generated parameters and feeding them to neural network models to be trained. Neural network models were trained with different hyper-parameters for finding the combination that worked better. This approach was repeated with the real market data. An out-of-sample test was done on the trained Deep Neural Network (DNN) using an unseen market data. The results showed that the trained DNN outperforms the Black-Scholes formula considerably, and that there is a great capability in Deep Neural Networks to be trained sophisticatedly and be used extensively in option pricing.

Keywords: Option pricing, GARCH, Empirical performance, GJR-GARCH, NGARCH, Heston-Nandi, European options, Monte Carlo, EMS, Deep Learning, Neural Networks

---

Shengying Yang, Financial Modelling, supervised by Dr. Mark Reesor

Title: Incorporating an Epidemic (SIR) Model into the Vasicek Portfolio Credit Risk Model

Abstract: Commercial banks are highly "pressured" due to spikes in credit losses amid COVID-19. A regular portfolio credit risk model is not enough for making an accurate approximation of portfolio loss under an epidemical situation. The Vasicek portfolio credit risk model and the SIR (S: susceptible, I: infectious, R: recovery) model are well-known and wildly used in the financial and epidemiological fields, respectively. However, few previous kinds of research on the Vasicek model involved epidemics. This project aims to transform the SIR model to be part of the risk model. Therefore, under the existence of an epidemic, the model would be able to update the credit loss distribution according to the real-time situation. We tried three different methods: 1) add the SIR model as an additional systematic factor to the Vasicek model; 2) have the SIR model affect the account level correlation between default drivers (eg., give more weight to the systematic factor as the epidemic increases); and 3) combine both of these methods. Monte Carlo simulations and numerical methods are employed to investigate the properties of each proposed model. We look at and compare the changes in loss distributions and some calculations, such as the average of portfolio losses, value-at-risk(VaR), and tail VaR, when the parameter values in the SIR model changes. The results indicate that the mixture model seems to have a better performance. However, real data is required for further evaluation and validation.

Sahab Zandi, Financial Modelling, supervised by Dr. Cristian Bravo Roman

Title: Effects of Edge-Coloured Networks on Classification Models for Single-Family Loan-Level Dataset

Abstract: Freddie Mac is a public enterprise which covers the secondary market for mortgages in US. This corporation buys mortgages on the secondary market, pools them, and sells them as a mortgage-backed security to investors on the open market. In this work, the single-family loan-level dataset from Freddie Mac is studied for a period of 10 years from 2009 to 2018. Two classification models are developed for loan default risk prediction, and important variables are identified. Zip codes and loan providers serve as two key variables for construction of the edge-coloured network. A sophisticated social network analytics method is applied to propagate influence from prior defaulters throughout the network to produce Personalized PageRank (PR) scores. It is demonstrated that adding these scores to the original data can help improve the performance when measured in terms of AUC. The effect is more pronounced for the Logistic Regression model than a Stochastic Gradient Boosting (XGBoost) model, which might be due to sample effects. In future research, the impact across all customers (with no sampling) will be studied.

Junsen Zhang, Financial Modelling, supervised by Dr. Matt Davison

Title: Performance Share Units as Long-Term Incentive Plans: The Effects of a Change to the Total Value of the Performance Share Units

Abstract: Long-term incentive plans are important for companies as it is the large component of executive pay for most companies. In Canada, 95% of companies use performance share units (PSUs) as their long-term incentives. PSUs are shares awarded at a future date to executives. This thesis models a high-dimensional PSUs, which the total PSUs vested are determined by the rank of the total shareholder returns among multiple comparators – the resulting structure resembles a multi-dimensional stock option. We consider a particular 6 comparator model based on the Celestica long-term incentive plan contract. For researching the impact of number of comparators, we also consider the 2 and 3 comparators models. Using the analytic solution of the geometric Brownian motion (GBM) stochastic differential equation (SDE) and applying Cholesky Decomposition theorem, the present value of the PSUs is simulated through the Monte Carlo method. Our main model requires a 6-dimensional Brownian motion SDE – with 6 volatility and 15 correlation parameters – in a risk-neutral measure with just a single risk neutral drift. We propose this method to observe which parameters have the biggest impact to the present value of the total value, which is the multiplication of the total PSUs vested and share price at maturity, in our studied high-dimensional PSUs. The results suggest that the total values are mainly affected by the number of comparators, the volatilities of stock price returns and the correlations between comparator stock price returns. Finally, we interpret the results and discuss how a company might achieve their strategic objectives and maximize shareholders value through the contract design of performance share units.

Keyword: Performance Share Units, Long-Term Incentive Plans, Multi-dimensional Stock Options, Geometric Brownian Motion, Stochastic Calculus, Monte Carlo Simulation, Cholesky Decomposition

Xiaoyu Zhang, Financial Modelling, supervised by Dr. Matt Davison

Title: Quantitative Methodologies to detect Money Laundering

Abstract: Since the 1970s, money laundering has increasingly become a major public hazard to the global economy. Money laundering is the illegal process of making "dirty" money appear legitimate by concealing the identity, source and destination. According to IMF (International Monetary Fund), the estimated amount of money laundered globally accounts for 3% to 5% of global GDP every year. The large sums of money laundered every year post a threat to the global economy and its security. Most researchers focus on detecting which is the first step in anti-money laundering, since a successful detection system can identify the trend or pattern of suspicious behavior and detect the introduction of new anti-money laundering patterns in time. The aim of this study is to review the research in the field of anti-money laundering and introduce the application of several quantitative methodologies commonly used. The existing quantitative analysis methods based on machine learning described in the literature are summarized and compared. Additionally, I present some results from a rich anti-money laundering dataset obtained from Kaggle and I show exploratory descriptive analytics including those obtained from K-means clustering and tree classifiers.

Dongfang Zhao, Financial Modelling, supervised by Dr. Ricardas Zitikis

Title: A Statistical Analysis of the 2020 U.S. Stock Market Fall Using Historical Financial Crisis Period Data and Time Series Techniques

Abstract: The stock market has encountered a sharp decline at the first half of 2020. To predict the stock price under the risk stock market as accurately as possible is important for investors and researchers. Time series analysis is used to develop several stock price prediction models such as the autoregressive integrated moving average (ARIMA) model and generalized autoregressive conditional heteroskedasticity (GARCH) model. This project presents specific time series analysis process of constructing predictive models using historical financial crisis period data. Historical Financial Crisis Period data and 2020 stock price data are obtained from Yahoo Finance, and S&P 500 index (^GSPC) is used for analyzing predictive models. ARIMA model reveals that the 2020 stock market will decrease after June. For ARIMA with GARCH model predicts that the stock return will not change.

Keywords: Time Series Analysis, Stock Price Prediction.

Xiyuan Zheng, Financial Modelling, supervised by Dr. Rogemar Mamon

Title: A financial cure for the pandemic? Dissecting the World Bank's emergency financing facility

Abstract: The pandemic bond issued by the World Bank (WB) in 2017 is a financial innovation, enabling the transfer of pandemic risk in the underdeveloped countries to the financial market. It covers various perils of diseases that could overwhelm the regional or global public-health systems and adversely impact the economy. If all of the six parametric triggers are activated - some of them are dependent on each other - the bond's principal and coupons will be used to finance coordinated, swift and resilient medical response to safeguard the well-being of the populace and to provide monetary support for affected businesses and households especially to those who are most vulnerable. However, this product is criticised for its onerous trigger requirements. In this research, the WB's pandemic bond-pricing framework, which requires inputs that are only partially available, is examined. From a rather unstructured COVID-19 data set, we underscore a major contribution of organising an information database customised to facilitate the pandemic-bond valuation. The pricing is analogous to that of a non-defaultable bond but with trigger-risk adjustments and recovery rate. A multivariate time series model in conjunction with the Monte-Carlo simulation is utilised to capture the triggers' underlying variables along with the trigger risk's quantification. We use a LIBOR-based discount factor and bootstrapping to address parameter estimation based on data that may have some reliability issue. Our findings show that the bond value is particularly sensitive to both the conditional probability of a pandemic happening as well as the unknown and random total number of deaths in covered areas. This is a joint work with R. Mamon

Keywords: Vector autoregression moving average model, catastrophe bond, trigger events, floating-interest rate

Poster Session Two: Dr Ricardas Zitikis

Zoom Link

Jonathan Jones, Financial Modelling, supervised by Dr. Rogemar Mamon

Title: The price tag of sustainability risk in fisheries: An environmental-derivatives approach

Abstract: In this study, sustainability is defined as prevention from ecological harm. We proposed a financial-market mechanism that could finance efforts for a resilient marine system. A stochastic version of the Gordon-Schaefer bioeconomic model is considered to capture the salient features of the fish biomass dynamics. We take into account various factors such as fish-harvesting effects, growth & decline of the fish population, fishing effort, catchability, and habitat's carrying capacity. Such a model is developed to support the valuation and risk management of environmental derivatives in fisheries. These over-the-counter market instruments function as hedging tools in the implementation of sustainability and conservation initiatives in the fishing industry. Our derivatives pricing framework is quite general in the sense that the stochastic interest rate process is also correlated with the evolution of the biomass level. Both European and American option-type products are valued with the latter type being priced via a least-square Monte-Carlo method. We perform sensitivity analyses, which revealed that the growth's intrinsic rate and volatility have the most significant impact on the biomass movement and option prices. Option values though are not substantially affected by the interest rate. The kernel density estimation and bootstrapping techniques are utilised to compute value-at-risk amounts. This research also underscores the intimate interplay between pricing and risk management. This is a joint work with R. Mamon.

Keywords: Fisheries, conservation, financial innovation, risk management, biomass modelling.

Javad Kazemian, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: Portfolio Optimization using Machine Learning Algorithms

Abstract: A simple one-layer Neural Network (NN) model with three hidden nodes is implemented here. The goal is for a defined contribution (DC) pension fund to allocate in risky assets such that a convenient terminal performance is accomplished. Optimal Neural Network control is determined for a portfolio of one risky and one risk-free asset using a target final wealth and a 30-year accumulation period. Asset prices are obtained directly from historical market returns, in addition to the parametric Monte Carlo model for return process. Five different strategies were examined for both historical and parametric datasets. The results showed that in early years, one can allocate most wealth in the risky asset. But to achieve higher final wealth, it is necessary to keep a higher proportion of risky asset for a longer period leading to higher risk exposures. In general, shifting from a lower to a higher final wealth demands a longer period of asset allocation in risky asset. However, for a low targeted wealth, the best strategy is to allocate a larger proportion of wealth in the risk-free account with a fixed interest rate.

Keywords: portfolio optimization; neural network; asset allocation; target wealth;

Xinxin Li, Financial Modelling, supervised by Dr. Lars Stentoft

Title: A study of default prediction model based on machine learning methods under COVID-19

Abstract: As the COVID-19 pandemic continues to evolve, there has been in a huge negative effect on the economy and the credit quality. Due to the decreased economic income caused by the virus disease, many borrowers can no longer meet the schedules for the interests and principal payments, causing the loan default. The aim of this project was to determine if default prediction models based on machine learning methods adapted to the crisis. Using various analytical models (Logistic regression, Random Forest algorithm and Quadratic Discriminant Analysis), this paper builds a prediction model from Lending Club's published data on loans issued on the first quarter of 2019 and 2020, respectively. The model's performance is detailed with respect to the accuracy, and the quality of the classifier is inspected through the Receiver Operator Characteristic (ROC) curve fitting, which is a graphical plot could evaluate the performances of binary classifies. The results show that: the prediction model could adapt to both status before and under COVID-19. In addition, Random Forest algorithm outperforms the Quadratic Discriminant Analysis and Logistic regression in predicting default rates.

Keywords: Loan default; Machine learning; Credit risk model

Junjun Liu, Financial Modelling, supervised by Dr. Lars Stentoft

Title: Estimating the Early Exercise Premium of American Put Option

Abstract: The Early Exercise Premium is interpreted as the price gap between American option and corresponding European Option. The Chicago Board Options Exchange simultaneously listed both of them on the S&P 100 index, allowing for a direct estimation of EEP. In this paper, we discover that the mean value of EEP of the put option is $ -0.5 for the sample from Jan 1, 2002, through Jan 1, 2010. Negative premiums may because that the market marker discouraged sellers of American options. Moreover, we find that the magnitude of EEP has a positive correlation with the maturity, dividend yield rate and implied volatility of stock price and negatively related to the moneyness and interest rate. Additionally, methods for pricing European and American options are discussed. There are lots of evidence that show the volatility is stochastic and the distribution of the returns has longer tails than a normal distribution. Therefore, we make comparisons between the option pricing model with constant volatility and stochastic volatility based on the early exercise premium. The Heston model (Heston, 1993) is also employed for pricing the option. The Euler Scheme is chosen for discretization of the Heston model. Model performance is measured in two ways. First, we show the model pricing error by model predicted price(PE), the root of the mean squared pricing error (RMSE) and the mean pricing error (MPE). Second, we assess the unbiasedness of models in predicting market EEP by using the regression method. We find that the SV model yields overall better results than CV model, especially for the long-maturity time and deep out of money option.

Yuhong Peng, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: Dynamic Programming of VaR constrained Portfolio

Abstract: Risk management quantitative measures like Value at Risk and Expected Shortfall have become essential tools to regulate the finance industry. However, dynamic optimization methods do not account for constraints on wealth. Kraft and Steffensen (2013) were first to apply dynamic programming for constrained portfolios, providing a methodology on how to construct a solution. They show how to set up Hamilton-Jacobi-Bellman equations for problems with constraints on wealth by involving financial derivatives and matching highly non-linear partial differential equations. In this project, I follow Kraft's methods for a VaR constrained portfolio, then Implementing their approach numerically under various market conditions and extreme scenarios using Matlab. The numerical solution of each setting is used to calculate the optimal initial unconstrained investment strategy and dynamic optimal strategies. To this end, their method gives valid strategies and we show the impact of various parameters, which comprehensively display the convenience of the approach.

Amirreza Seddighi, Financial Modelling, supervised by Dr. Cristian Bravo Roman

Title: Application of Multimodal Learning in Credit Risk Modeling Using Satellite Imagery

Abstract: Multimodal learning analytics obtains information from multiple sources and combines them in the form of a predictive model. Information sources may include audio, image, text, structured data, etc. and since such models aggregate information of different types, they can potentially have a higher performance compared to conventional models. The purpose of this work is to develop a mortgage credit risk model by combining structured data and satellite images through convolutional neural networks. When compared to a conventional machine learning model with structured data as the only input, one could quantify the contribution of images as the second source of information to the overall model performance. In this research, the Freddie Mac Single Family Loan-Level dataset has been studied where applicants' zip codes are used to detect their area of residence. With Logistic Regression and XG-Boosting algorithms as baseline models, the default rate in each area was then predicted using structured data. Next, satellite images corresponding to each area were collected and utilized in a deep learning model to predict the same variable. Both baseline models performed with AUC of 0.64 while the deep learning model had a lower performance with AUC of 0.58. The effectiveness of multimodal learning model where both structured and unstructured datasets are combined together is then to be assessed with regards to the baseline models.

Key words: Multimodal Learning, Satellite Imagery, Convolutional Neural Networks

---

Li Zhang, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Title: Implementation of Optimal Investments with Continuous Constraints

Abstract: In this report, we focus on implementing an optimal expected utility portfolio with a value-at-risk constraint. This is a significant problem in the banking and insurance sector which needs to be solved in order to control risks of the optimal portfolio and ensure that regulatory requirements can be fulfilled under the optimal portfolio. The optimal investment happens to be a barrier-type contingent claim on the unconstrained optimal investment. The problem is formulated as a maximization problem of expected utility of wealth. Some application results of this problem are shown in this paper.

Keywords: Optimal portfolio, Value-at-risk, continuous constraints

Poster Session Three: Dr. Xiaoming Liu

Zoom Link

Yunfei Du, Actuarial Science, supervised by Dr. Shu Li

Title: Pricing GMDB under a drawdown-dependent fee structure

Abstract: Variable Annuity (VA) is a type of insurance or annuity contract, whose value varies based on the performance of the underlying portfolio. VA is always combined with a range of guaranteed minimum benefits in the form of "GMxB", to cover death benefits, maturity value, income or accumulation value. Guaranteed minimum death benefit (GMDB) guarantees a minimum death benefit which provides protections when the underlying asset performs poorly. We consider a VA with GMDB rider under a special fee structure, which charges a level fee when the account value drops below a certain predetermined proportion of its running maximum. For traditional equity-linked products, the fee is always charged in proportion to the account value, so that higher fee is charged when the account value but the guarantee is out of money at that moment. It attracts our interest that how we can re-design a fee structure which can mitigate this misalignment. In our design, the fee is charged when the guarantee is close to being in the money. Under certain assumptions, we derive the valuation formula through change of measure and renewal arguments. Analytical and numerical results are also presented for illustration purpose.

Wei Li Fan, Financial Modelling, supervised by Dr. Rogemar Mamon

Title: Vasiček 'meets' Kermack & McKendrick: A stochastic SIR model for the valuation of a pandemic-bond instrument with a simple trigger event

Abstract: A catastrophe (CAT) bond is a mechanism that generates money for insurers and re-insurance companies when a natural disaster happens. In this research, a CAT bond that allows the issuer to receive funding from the bond is linked to a mainly pandemic-driven variable. A compartmentalised deterministic epidemic model was previously proposed by Kermack & McKendrick. We consider its stochastic analogue to describe the random number of individuals in three classes: susceptible (S), infected (I), and removed (R). Hence, in each compartment at time t, S(t), I(t) and R(t) represent the numbers of individuals satisfying a system of interacting stochastic differential equations. The CAT bond is triggered, and hence a payoff materialises for the holder, when the number of infected persons exceeds a pre-determined threshold. Under our generalised pricing framework, a Vasiček-based financial risk factor is correlated with the SIR dynamics. The probability of a pandemic occurrence during the bond's tenor from the time of issue to maturity is calculated via a Poisson process using actual data in the last 3 centuries. The result of Pederson and Cox (2000) was invoked to justify our choice of a pricing measure. Our sensitivity analyses ascertained that the SIR's disease transmission and recovery rates as well as the mean-reverting level of the Vasiček's interest rate have substantial effect on the bond price. Lastly, the proposed synthesised model was tested and validated on the Canadian COVID-19 data. This is a joint work with R. Mamon.

Keywords: Catastrophe bond; infectious-disease modelling; Poisson process; Vasiček model; sensitivity analysis; pandemic-trigger events.

Dechen Gao, Actuarial Science, supervised by Dr. Kristina Sendova

Title: Applications of the classical compound Poisson model with claim sizes following a compound distribution.

Abstract: In this paper we introduce a generalization of the classical compound Poisson Model with claim sizes following a compound distribution. As an application, we consider a model involving a Poisson-Geometric claim-counting process. We also provide some distributional quantities and properties of the Poisson-Geometric process, as well as some ruin-related quantities under the resulting Poisson-Geometric risk model.

Yue Gu, Financial Modelling, supervised by Dr. Ricardas Zitikis

Title: Assessing Financial Risks Associated with Health Insurance and Actual Health-related Costs

Abstract: In China, many people buy health insurance, in the hope that it will reduce healthcare expenditure. However, how this characterization of health insurance plays out in practice? How well health insurance protects people from illness-related financial risk? Yet there is no obvious evidence for that in the real-world health insurance always reduces financial risk. In this project we want to examine whether health insurance will reduce total health cost, based on the newest dataset from China Health and Nutrition Survey (CHNS), 2015.

We use Generalized Linear Models (GLM) and a Tobit model here to model the health cost data. Specifically, Tobit model has been widely assumed for modeling health cost data. However, Tobit does not work well here because of the excess of zeros. Although Poisson model is designed for count data, it is proved that this model is still available for continuous health data. It assumes that the mean and variance are equal. However, this restriction is violated in many applications because data are often over-dispersed. When the source of overdispersion is the excess of zeroes, the Zero-inflated Poisson regression model fits data better. In this paper, we conduct both Poisson regression and Zero-inflated Poisson regression to analysis effect factors of health cost.

As a result, contrary to the consensus, we find that health insurance is going to increase our cause out-of-pocket payments. In the end, this paper proposes relevant suggestions based on the analysis results.

Key words: Health Insurance; Tobit Model; Zero-inflated Poisson Regression

Yuwei Huang, Actuarial Science, supervised by Dr. Kristina Sendova

Title: Smoothing and studying of factors affecting mortality rates by province in China

Abstract: Endowment has been one of the most important issues for the Chinese government while there is little research about variability among provinces. In this project, after processing large amounts of missing data of deaths, mortality rates are smoothed by using a combination of a generalized additive model and a Gompertz model. The statistical significance of possible influencing factors (income, employment, crime, healthcare and geographical location) was tested through a generalized linear model. As a result, we have obtained a smoothed mortality table for both insureds of endowment of urban and of rural residents and urban employees, which are also classified by provinces. We found that employment and crime are powerful influencing factors for insureds of endowment of urban and of rural residents.

Duo Xu, Actuarial Science, supervised by Dr. Shu Li

Title: The fair valuation of ultimate drawdown insurance

Abstract: For the surplus process of a company, the term drawdown size is defined as the difference between running maximum and current surplus level, which provides the managers an intuitionistic variable to assess the surplus process, especially its downside risk. In a classic drawdown insurance, a claim merely occurs when the drawdown size exceeds a fixed barrier. However, such drop-in surplus might be caused by different exogenous and endogenous risk factors, and thus it is reasonable to allow for a grace period for a company to recover through its own ability of liquidation or restructure. Such an idea of grace period is defined as a Parisian drawdown time, which considers the amount of time spent under a soft barrier. The ultimate drawdown, combining the Parisian drawdown with a hard barrier, provides the policyholder the protection against the downside market behavioral in order to mitigate the risk of bankruptcy. In this project, we propose an insurance contract against the drawdown risk, where a claim is incurred when the ultimate drawdown occurs. We obtain the fair market premium under equivalence principle. We also examine the optimal termination strategy for the policyholder when a cancellation feature is allowed in the insurance contract.

Poster Session Four: Dr. Wenqing He

Zoom Link

Shiyu He, Statistical Science, supervised by Dr. Doug Woolford and Dr. Reg Kulperger

Title: Dirichlet Regression with Applications to Alberta Forest Fire Data

Abstract: Forest fire seasons start in the spring and end in the fall. During the fire season, temperatures rise and fuel moisture reduces, causing fires more likely to be ignited and spread rapidly. Many Canadian provinces, including Alberta, have experienced more extreme fire seasons in recent years. There is concern about to what extent characteristics of fire seasons may be changing due to climate change. Past research has modelled the start and end of the fire season separately for their trends over time. Those marginal approaches did not account for the joint distribution of the timing of the start and end, which may not be independent. The main objective of this project is to advance the methodology for modelling fire season data, using Dirichlet regression that permits the joint modelling of the timing of the start and end of the fire season. We show that compared to its marginal Beta regression, Dirichlet regression can model the joint distribution of three aspects of the fire seasonality: start, length and end while preserving their trends with year.

Xingfu Zhang, Statistical Science, supervised by Dr. Serge Provost and Dr. Hyukjun Gweon

Title: Threshold Moving Approaches for the Class Imbalance Problem and their Application to Multi-label Classification

Abstract: We propose simple threshold moving techniques for coping with the class imbalance problem. Decision thresholds are adjusted to match the class distribution of the training data to that of the predicted outcomes for unseen data. The proposed approaches are applied to multi-label classification wherein each instance may simultaneously belong to more than one label. Experimental results on two data sets confirm that the performance of the well-known binary relevance method can be improved upon when combined with the proposed threshold moving techniques.

Manjia Zhao, Statistical Science, supervised by Dr. Doug Woolford

Title: Factors Contributing to Detection, Response and Initial Attack Failures in the Northwest Fire Region of Ontario

Abstract: The Ontario Ministry of Natural Resources and Forestry (OMNRF) defines initial attack (IA) action as successful if fire is either in the condition of "Being Held" by 12:00 local time the day following its report or has a final size of 4.0 ha or less. We further decomposed the OMNRF fire management system into the following three mutually-exclusive systems: a detection system which aims to discover a fire at size 4.0 ha or less; a response system which aims to respond to fire, that were already successfully detected, before its size reach 4 ha; and the IA system itself with success as defined above. We examined the effects of potentially driving factors for detection failure, response failure and IA failure. This was done using logistic regression and lasso regression on 12,360 fires within the Northwest Region of Ontario from 1995 to 2019. The driving factors of IA failure were rate of spread, initial attack type, size of IA ground force, season, ISI, district fire load and latitude. IA failure probability also varied between districts. Detection failure probability was related to fires' coordinate, district fire load and specific cause of fire, and varied between fire management zones. Response probability was related to rate of spread, district fire load, FWI and longitude. Our work can be used as guidance for future fire management strategies, aiming at lowering detection, response and IA attack failure rate.

Poster Session Five: Dr. Simon Bonner

Zoom Link

Gansen Deng, Statistical Science – Biostats, supervised by Dr. Wenqing He

Title: Statistical Analyses for a Clinical Sepsis Study

Abstract: Sepsis is a life-threatening syndrome caused by the immune system working overtime to fight infection. A sepsis clinical study has been conducted to investigate the effect of sepsis in humans on MAIT cells in the blood by looking at their surface markers and their function (production of cytokines), and we provide statistical analysis to the team. The difference between different cohorts are compared using t-tests and Mann-Whitney U tests, and GEE (Generalized Estimating Equations) is utilized to identify variables with significant longitudinal trends. Correlation analysis is also involved to check if variables are significantly correlated with each other. The key conclusion is that MAIT cells decrease in abundance in septic patients comparing with healthy control, but surprisingly that their function progressively declines as the patient gets better. *MAIT % of total* and *CD14 HLADR* are found to be two important indicators of sepsis.

Fatemeh Gholizadeh, Statistical Science, supervised by Dr. Camila de Souza

Title: Clustering Brain Cells' Single-Cell RNA Sequencing Data

Abstract: Single-cell RNA sequencing (scRNA-seq) technology has attracted the attention of researchers in identification of sub-types of cells. However, no clustering method has been introduced as a benchmark method in such types of data, especially neuronal scRNA-seq. Therefore, we aimed to evaluate the performance of Seurat and SC3 clustering methods in brain cell's scRNA-seq data and investigate how comparable their findings are with a widely used method, called Backspin. For this purpose, we used published Backspin results obtained from analysing neuronal scRNA-seq, which were validated in laboratory, and compared those results with Seurat and SC3 using co-clustering plots, t-SNE plots and CH-index. The co-clustering showed that Seurat performed slightly better than SC3, while CH-index showed a slightly better performance for SC3. As the results based on the co-clustering plots and CH-index were very close for both methods, we concluded that both methods perform comparably. Both methods could not recover the clustering results obtained by Backspin and validated in laboratory.

Chufan Wu, Statistical Science, supervised by Dr. Hao Yu and Dr. Hyukjun Gweon

Title: Wind Turbine Performance Index for Early Anomaly Detection

Abstract: Wind turbine anomaly detection is an application of outlier detection for wind turbine time series data. Turbine performance observed in fault condition can be considered outliers. This project proposes a Performance Index (PI) for early detection of anomalous wind turbines.

Based on PI, we develop a distance-based normal behavior method to detect potential anomalous turbines. Using data in the training stage, the proposed method compares the distances between individual wind turbines in terms of PI and estimates specific thresholds for each turbine that identify the anticipated normal pattern of that turbine. Those thresholds are used to determine anomalous behavior of each turbine in the testing stage. The proposed algorithm is applied to multivariate time series data provided by EnBridge Inc. headquartered in Canada. Due to the structure of the wind turbine data, we preprocess the data by identifying and modifying pseudo outlier points observed due to extreme weather conditions rather than mechanical malfunctions.

Potential anomalous turbines and their corresponding timestamps detected by the proposed method provide targeted information for turbine maintenance.

Chengqian Xian, Statistical Science, supervised by Dr. Camila de Souza

Title: Predictive Modelling of Health Outcomes in Intensive Care Units in London

Abstract: The literature in Intensive Care Units (ICUs) data analysis focuses on predictions of length-of-stay (LOS) and mortality based on patient acuity scores such as APACHE, SOFA, etc. Unlike ICUs in other areas around the world, ICUs in Ontario collect two main intensive care Scoring Systems, a therapeutic acuity score called the "Multiple Organs Dysfunctional Score" (MODS) and a nursing workload score called the "Nine Equivalents Nursing Manpower Use Score" (NEMS). The dataset analyzed in this project contains patients' NEMS and MODS scores measured upon patient admission into the ICU as well as other characteristics (e.g., age, sex, admission diagnosis). This data was collected between January 1st, 2015 and December 31st, 2018 at two teaching hospital ICU's in Ontario, Canada. In this project, I developed logistic regression, random forests (RF) and neural network (NN) models for mortality (discharged or deceased) and LOS (short or long stay) predictions based on previous studies. Considering the effect of mortality outcome on LOS, I also combined mortality and LOS to create a new a response called LMClass (short discharged, short deceased or long and unknown outcome), and applied multinomial regression, RF and NN for its prediction. Five repetitions corresponding to five random starting points have been done in RF and NN for model optimization, and 5-fold cross-validation (CV) was also carried out for model stability checking. Results show that all models are stable based on 5-fold CV with area under the curve (AUC) values ranging from 0.790 to 0.791 in testing set for mortality prediction, AUC values from 0.695 to 0.706 for LOS prediction, accuracy from 0.631 to 0.638 for LMClass prediction. This study also demonstrates that MODS and NEMS as well as their components measured upon patient arrival significantly contribute to health outcome prediction in ICUs. For future work, survival analysis will be applied for LOS prediction and a new larger dataset including cases from several other ICUs from the Critical Care Information System in Ontario will be obtained for further analysis.