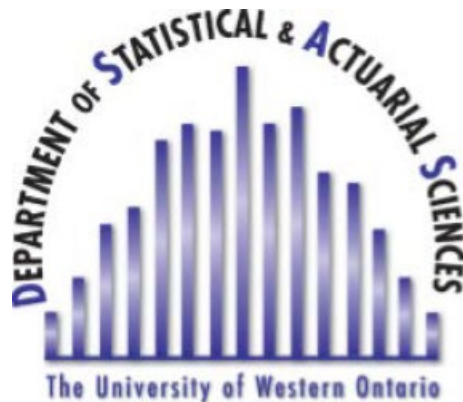


# The 20<sup>th</sup> M.Sc. Day

## DEPARTMENT OF STATISTICAL & ACTUARIAL SCIENCES

July 26, 2024



## Table of Contents

Schedule of Events .....	2
Titles and Abstracts.....	3
Yuxue Zhang, Actuarial Science, supervised by Dr. Shu Li .....	3
Bernard Emmanuel Bartels, Statistics, supervised by Dr. Kristina Sendova.....	3
Parinaz Zarei, Financial Modelling, supervised by Dr. Kristina Sendova.....	3
Aleksandr Chen, Financial Modelling, supervised by Dr. Ricardas Zitikis .....	4
Shumin Chen, Financial Modelling, supervised by Dr. Cristián Bravo Román .....	4
Wen-Zhan Cao, Financial Modelling, supervised by Dr. Marcos Escobar-Anel.....	4
Ziyuan Fang, Financial Modelling, supervised by Dr. Marcos Escobar-Anel .....	5
Hongchen Liu, Financial Modelling, supervised by Dr. Marcos Escobar-Anel .....	5
Sahand Seyfi, Financial Modelling, supervised by Dr. Marcos Escobar-Anel .....	6
Xiangshi Li, Statistics, supervised by Dr. Wenqing He .....	6
Jiachen Pan, Statistics, supervised by Dr. Hyukjun Gweon .....	7
Lin He, Statistics, supervised by Dr. Hyukjun Gweon and Dr. Shu Li.....	7
Aryan Rezanezhad, Statistics, supervised by Dr. Grace Yi.....	8
Zihan Zhang, Statistics, supervised by Dr. Grace Yi.....	8
Kasiwatte Kankanamge Dinithi De Silva, Statistics, supervised by Dr. Douglas Woolford.....	8
Luc John-Paul Benoit, Statistics - Hazards, Risks and Resilience, supervised by Dr. Douglas Woolford .....	9
Jessica Dobbin, Statistics - Hazards, Risks and Resilience, supervised by Dr. Douglas Woolford ..	9

## Schedule of Events

### Location: Western Science Centre Room 240

Student and Faculty - Coffee and Mingle	8:40 AM
<b>Opening Remarks, Dr. Kristina Sendova (Department Chair)</b>	<b>8:55 AM</b>
<b>Session I</b> (Chair: Dr. Ricardas Zitikis)	<b>9:00 AM</b>
1) Yuxue Zhang (MSc AS - Dr. Shu Li)	9:00 AM
2) Bernard Emmanuel Bartels (MSc ST - Dr. Kristina Sendova)	9:15 AM
3) Parinaz Zarei (MSc FM - Dr. Kristina Sendova)	9:30 AM
4) Aleksandr Chen (MSc FM - Dr. Ricardas Zitikis)	9:45 AM
5) Shumin Chen (MSc FM - Dr. Cristián Bravo Román)	10:00 AM
<b>Break and set up for Session II</b>	<b>10:15 AM</b>
<b>Session II</b> (Chair: Dr. Lars Stentoft)	<b>10:45 AM</b>
6) Wen-Zhan Cao (MSc FM - Dr. Marcos Escobar-Anel)	10:45 AM
7) Ziyuan Fang (MSc FM - Dr. Marcos Escobar-Anel)	11:00 AM
8) Hongchen Liu (MSc FM - Dr. Marcos Escobar-Anel)	11:15 AM
9) Sahand Seyfi (MSc FM - Dr. Marcos Escobar-Anel)	11:30 AM
<b>Light lunch</b>	<b>11:45 AM</b>
<b>Session III</b> (Chair: Dr. Shu Li)	<b>12:45 PM</b>
10) Xiangshi Li (MSc ST - Dr. Wenqing He)	12:45 PM
11) Jiachen Pan (MSc ST - Dr. Hyukjun Gweon)	1:00 PM
12) Lin He (MSc ST - Dr. Hyukjun Gweon and Dr. Shu Li)	1:15 PM
13) Aryan Rezanezhad (MSc ST - Dr. Grace Yi)	1:30 PM
<b>Break and set up for Session IV</b>	<b>1:45 PM</b>
<b>Session IV</b> (Chair: Dr. Jiandong Ren)	<b>2:15 PM</b>
14) Zihan Zhang (MSc ST - Dr. Grace Yi)	2:15 PM
15) Kasiwatte Kankanamge Dinithi De Silva (MSc-T ST - Dr. Douglas Woolford)	2:30 PM
16) Luc John-Paul Benoit (MSc-T ST-HR - Dr. Douglas Woolford)	2:45 PM
17) Jessica Dobbin (MSc-T ST-HR - Dr. Douglas Woolford)	3:00 PM
<b>Closing Remarks and Photo Taking</b>	<b>3:15 PM</b>

---

## Titles and Abstracts

---

**Yuxue Zhang, Actuarial Science, supervised by Dr. Shu Li**

*Equity-Linked Insurance in Financial Planning Under CRRA Utility*

Segregated funds, or the equity-linked insurance, provide guaranteed payments for policyholder's funds, and are considered as an option in financial planning. In this project, we investigate the performance of the segregated funds comparing to the mutual funds, under the constant relative risk aversion (CRRA) utility function. We extend the results in Huang, Milevsky and Wang (2008), and analyze the optimal allocation problem for family investing in segregated funds, insurance plans, and riskless assets, with a target function defined over the family's planning horizon covering both before and after retirement. Our numerical results show that the utility function on overall consumption, under the portfolio with segregated funds, could be higher than that with mutual funds. This finding suggests the benefits of using the equity-linked insurance as a vehicle for retirement planning.

---

**Bernard Emmanuel Bartels, Statistics, supervised by Dr. Kristina Sendova**

*The Impact of Marital Status on Longevity: A Comparative Analysis of Joint-Life and Single-Life Mortality Rates*

Joint-life tables provide insights into the mortality dynamics of individuals living in a couple, typically within the context of marriage or common-law. This project analyzes the Long Administrative Database (LAD) provided by Statistics Canada. We focus on data grouped by gender and marital status to investigate patterns in mortality rates of married or common-law individuals and compare with single individuals. The LAD comprises of a 20% sample of the Canadian population collected from yearly income tax returns.

Many previous studies consistently show that married individuals tend to have lower mortality rates compared to their single counterparts. The findings of this project contribute to a deeper comprehension of the reasons behind longevity patterns within marital contexts. Also, we aim to provide life tables to ensure accurately priced insurance policies for married and single individuals.

---

**Parinaz Zarei, Financial Modelling, supervised by Dr. Kristina Sendova**

*Analysis of the Financial Wellness of Canadian Households Using Regional Data*

This project involves merging and analyzing three comprehensive datasets that were collected up to December 31, 2023. These datasets originate from Statistics Canada and have been provided to us by an industrial partner. The merged dataset enables us to examine certain aspects of the economic behaviour and financial wellness of Canadian households.

Key variables include before-tax household income, after-tax household income, various forms of expenditures, net worth, assets, and different types of debt. The analysis employs multiple

regression and step-wise selection methods to identify significant predictors. Correlation analysis reveals the relationships among various financial variables, providing insights into household financial stability and economic behaviors.

---

## Aleksandr Chen, Financial Modelling, supervised by Dr. Ricardas Zitikis

### *Operational Effectiveness Internship at the IESO*

Currently, I am completing my internship as a member of the Operational Effectiveness and Assessments team at the Independent Electricity System Operator (IESO). The IESO is a not-for-profit organization, one of seven independent system operators in North America, which is responsible for operating the electricity market and managing the power system in real-time in the province of Ontario. In this presentation, I will go over the principles of the real-time energy market, how bids and offers are settled, and how to ensure that the electricity supply always meets the demand. I will also talk about some projects I have been working on, including the projects that aim to contribute to the Operating Reserve Market in Ontario.

---

## Shumin Chen, Financial Modelling, supervised by Dr. Cristián Bravo Román

### *Impact of Bond Market Shocks on Capital Ratios: Insights from a FAVAR Study*

The recent collapses of Credit Suisse and Silicon Valley Bank have heightened concerns about global banking stability. This study examines the effects of financial shocks on the stability of Global Systemically Important Banks and Domestic Systemically Important Banks in the US, Canada, and Europe by analyzing capital ratios. We focus on the Tier 1 Capital Ratio and Total Capital Ratio, which are crucial components of the Basel Standards. Using a Factor-Augmented Vector AutoRegression (FAVAR) model, we assess how fluctuations in the FTSE Canada All Corporate Bond Index (tracked by the XCB.TO ETF) impact banks' capital ratios. The results reveal varying responses to financial shocks across different banks. Impulse response functions indicate an immediate drop in both capital ratios following a shock, with diverse recovery patterns. This behavior was evident during crises such as the COVID-19 pandemic. Some banks' capital ratios stagnate near regulatory minimums, while others fall below, signaling significant risk during bond market-related financial stress episodes.

---

## Wen-Zhan Cao, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

### *An Excursion into The World of CBOE Indices - Some Examples and Properties*

The VIX index, also known as the fear gauge, is a volatility measure of the S&P500 index launched by the Chicago Board Options Exchange (CBOE) in the early 1990's. The popularity of this VIX index has increased substantially in the past two decades. Notably, this innovative index is just one example of the hundreds developed by CBOE. In the first part of this project, we collect, describe, classify, and motivate a subset of such indices. In the second part, we select 11 representative

indices to gain a better understanding of their statistical properties, and their financial functionality, worth.

In particular, we focus on the 5 VIXs capturing S&P500's volatility term structure, the VIX of two commodities (i.e., Oil and Gold), and four new interesting indices: VXTLT (i.e., the VIX of the long-term bond ETF), VVIX (i.e., volatility of the VIX), SKEW (i.e., steep level of the implied volatility curve of S&P500) and COR1M (i.e., correlation among S&P500' stocks). We check for auto-correlation, stationarity, Hurst index, heteroscedasticity, and the distributions of their innovations. By studying these statistical properties of the chosen indices, we hope to gain better insights into how to read the numbers and how the market would read them. Moreover, as some of the indices are relatively new and have not been fully exploited, we search for uses of the indices in industry and imagine potential application.

---

## Ziyuan Fang, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

### *Applications of Regression Random Forest to Asset Pricing*

This project explores the application of regression and tree-based machine learning models to asset pricing. Using daily data of excess returns from well-known assets like Apple (AAPL), Bank of America (BAC), Boeing (BA), and sector-specific portfolios HiTec and Manuf, along with market risk premium, we investigate the goodness of fit of various regression techniques between the assets and the market. We start with linear regression, documenting fitting limitations, particularly on the tails. Constant regression tree uses constant fitting for asset price and market, while random forest integrates multiple constant regression trees. However, the results of these two models are not much better than the benchmark, the Capital Asset Pricing model (i.e., linear regression or CAPM). We further explore model trees (i.e., linear regression trees) and local linear forests, which incorporate linear models within tree structures for refined analysis. Our results indicate that forest-based methods with linear, rather than constant structures, outperform traditional asset pricing models (i.e., linear regression models), offering valuable insights into asset pricing. This underscores the potential of machine learning in financial modeling.

---

## Hongchen Liu, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

### *ESG Ratings - A Comparison of Sectors at The Marginal and Dependence Levels*

ESG rating integrates environmental, social, and governance considerations into the evaluation of corporations, prioritizing companies' sustainability alongside profitability. Although originated in the 1970s, it is only recently that ESG criteria have become a pivotal factor in guiding sustainable investment decisions. This project explores the patterns of ESG ratings among companies within and across different sectors, identifying sectors with significantly higher or lower ESG ratings, examining correlations per sector, and how these ratings have evolved over time.

Using ESG ratings provided by RepRisk, we analyze data from 1,766 companies across ten sectors from 2010 to 2020, including Financial Services, Oil & Gas, Banks, Retail, Pharmaceuticals & Biotechnology, Personal & Household Goods, Food & Beverages, Construction & Materials, Mining,

and Software & Computer Services. We apply Welch's and Kruskal-Wallis hypothesis tests to determine if there are significant differences in various statistics capturing sector ratings. Additionally, we examine which sectors are more susceptible to ESG-related risks and analyze trends in ESG ratings of different sectors over time. Last but not the least, we examine the distribution of rating correlations within sectors to identify overarching patterns in ESG performance.

---

## Sahand Seyfi, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

### *Ensuring and Optimizing Ontario's Electricity for Reliability and Cost-Effectiveness*

The Independent Electricity System Operator (IESO) manages Ontario's electricity system by overseeing critical functions such as Control Room Operations to monitor performance, investigate issues, and enforce operational rules. They administer auctions for electricity trading and manage programs aimed at market improvement and ensuring adequate power supply. Additionally, the IESO supports energy projects, integrates diverse energy sources, and develops strategies for maintaining sufficient electricity supply.

My specific tasks include keeps tracking of resources performance in every Daily Operating Reserve Analysis (ORA), records number of consecutive failures to assess emergency power availability, preparing and revised the tool for Unit Readiness Reports (URR) to schedule power plant maintenance, and applying Locational Marginal Pricing (LMP) and Henry Hub Pricing to adjust electricity prices based on grid capacity and market conditions. These efforts ensure that Ontario's electricity remains reliable and cost-effective.

Additionally, I contribute to regulation assessment and frequency response services that balance resources with load and maintain the scheduled Interconnection frequency at 60 Hz. This involves committing qualified resources to adjust output or demand using Automatic Generation Control (AGC) to follow moment-by-moment changes in Area Control Error (ACE). The requirements for these services, set according to North American Electric Reliability Corporation (NERC) criteria, vary by hour and season to address net load variability within a 5-minute dispatch interval. I support studies on co-optimization processes to minimize the total cost of energy, operating reserves, and regulation services, with regulation service costs totaling \$31.0 million in 2022.

---

## Xiangshi Li, Statistics, supervised by Dr. Wenqing He

### *The Effect of Measurement Error in Covariate on Classifications: A Simulation Study*

Classification methods are widely used in statistical learning and real-world data analysis. However, due to inaccuracies in measurement instruments and human factors, measurement errors in covariates are inevitable. The impact of these errors on the performance of classifiers needs to be thoroughly investigated. This report conducts a comprehensive simulation study using various classifiers based on different data generation setups where measurement errors exist in covariates. The performances of these classifiers are evaluated and compared to provide a deeper understanding of how measurement errors affect classification accuracy. Our simulation results

show that when the magnitude of the measurement errors gets stronger, the accuracy of the classifiers decreases. Therefore, in classification problems, the measurement error issues need to be addressed.

---

## Jiachen Pan, Statistics, supervised by Dr. Hyukjun Gweon

### *Active Learning with Noisy Labels in Multi-Class Classification*

Multi-label classification is used in a wide range of applications. Like other classification problems, successful modeling for multi-label data requires a high-quality labeled dataset, which is often difficult to obtain. In this project, we consider practical situations where labeled data is scarce but unlabeled data is large. An effective data mining approach is active learning that iteratively updates the training data by adding the examples most informative for the current model to learn. We examine the use of active learning in multi-label classification where mislabeling is probable. We use a committee-based approach for identifying potential labelling errors. Our main findings show that relabeling instances in the presence of mislabeling helps reduce the Hamming loss but has no significant effect on the 0/1 loss. In addition, we find that the strategy of removing suspicious labels is ineffective for both evaluation metrics. Regarding active learning strategies, our results show that the example-label-based approach is more effective than the example-based approach in managing labelling errors in a multi-label classification task.

---

## Lin He, Statistics, supervised by Dr. Hyukjun Gweon and Dr. Shu Li

### *Random Forest Credibility Model*

Credibility theory is a cornerstone in actuarial science for insurance experience rating. The traditional Buhlmann and Buhlmann-Straub models only rely on the past claim experience for the credibility premium calculation. Incorporating covariate variables help to improve the prediction performance of credibility premiums. However, many existing methods lack flexibility in capturing nonlinear and interaction effects of covariates due to their reliance on pre-specified parametric regression forms. As shown in Diao and Weng (2019), their Regression Tree Credibility (RTC) model utilizes machine learning techniques to enhance premium prediction accuracy within credibility theory. The RTC model eliminates the need for a preliminary analysis of the relationship between individual net premiums and covariates. Their proposed algorithm autonomously identifies influential covariates and optimal cutting points to form data space partitions, leading to effective premium prediction rules.

In this study, we extend the RTC model and propose a Random Forest Credibility (RFC) model, aiming to improve the predictive performance through ensemble approaches. The superiority of the RFC model in terms of prediction accuracy is demonstrated through simulation studies. This study highlights the potential of machine learning techniques to revolutionize actuarial methodologies, providing more accurate and reliable premium predictions.



---

## Aryan Rezanezhad, Statistics, supervised by Dr. Grace Yi

### *ASIMEX: An R Package for Estimation of Average Treatment Effect with mismeasured covariates and misclassified outcomes*

In regression contexts, it is well documented that ignoring measurement error often yields biased inference results. In recent years, there is increasing research on causal inference in the presence of measurement error. Addressing misclassified binary outcome variables, Shu and Yi, (2019) proposed methods to correct for misclassification effects using inverse probability weighted estimation of average treatment effects. To facilitate the practical application of these methods, we implement correction methods proposed by Shu and Yi, (2019) and develop the R package ASIMEX for general users. This package corrects for measurement error in covariates and misclassification in outcome to ensure to provide valid estimation of average treatment effects. The ASIMEX function employs logistic regression for treatment assignment and outcome modeling, supplemented by bootstrap sampling to compute standard errors and confidence intervals. Simulated datasets are used to illustrate the application of this developed package.

---

## Zihan Zhang, Statistics, supervised by Dr. Grace Yi

### *Enhancing Understanding of Emergency Care Management Using Machine Learning Methods*

Predicting the length of stay (LOS) in the emergency department is critical for healthcare professionals to effectively assess patients' conditions. In this study, utilizing the MIMIC-IV-ED demo data dataset, we evaluate the performance of logistic regression, random forest, and XGBoost models in predicting both the need for hospitalization and the LOS for patients. Among these, the random forest model demonstrates superior performance, achieving an AUC value of 0.85, together with accuracy, specificity, and sensitivity values of 0.86, 0.66, and 0.92, respectively; the random forest model can reasonably well predict patients' LOS based on their early physiological parameters. Our findings suggest the usefulness of the random forest approach in enhancing the allocation of medical resources. However, comprehensive validation of these results requires further data collection from additional hospitals.

---

## Kasiwatte Kankanamge Dinithi De Silva, Statistics, supervised by Dr. Douglas Woolford

### *Modeling Microplastic and Macroplastic Abundance in Urban Stormwater Ponds*

Plastic pollution is a major environmental concern on a global scale. Stormwater runoff plays a crucial role in transporting plastic waste from various sources into aquatic environments, especially in urban environments. This study focuses on modelling the abundances of microplastics (MPs; 53  $\mu\text{m}$  to 5 mm) and macroplastics (MaPs; >5 mm) in urban stormwater ponds, aiming to identify and assess the factors influencing their accumulation and distribution. Bottom sediment was sampled for MPs from 28 stormwater ponds located in London, Ontario, Canada. Macroplastic debris items were recorded from 10m<sup>2</sup> quadrats along the pond banks. The number of sampling sites for each pond depended on factors, such as the number of inlets and outlets,

pond size, and water depth. There is interest in modelling not only the abundance of plastic pollution, but what leads to certain locations containing certain types of microplastic pollution. Consequently, we plan to employ a hurdle modelling framework. For a given MaP type of interest, this two-stage approach first models the probability it is present and then models its abundance. This presentation provides an overview of the preliminary analysis of the MaPs data, highlighting current research directions including exploratory data analysis, preliminary modelling, and exploring for correlation between counts of different types of MaP, which may be important for the joint modelling of abundances.

---

## Luc John-Paul Benoit, Statistics - Hazards, Risks and Resilience, supervised by Dr. Douglas Woolford

### *Exploring Dispatch System Success in Wildland Fire Management*

Although wildland fires are an important ecological process with benefits to ecosystems, as the wildland-urban, infrastructure and industry interfaces continue to expand there are an increasing number of values at risk. Effective fire management seeks to mitigate losses caused by wildland fires. As a system, it comprises linked subsystems of detection, dispatch, initial attack and sustained action. An important key performance measure of fire management is initial attack (IA) success, which is typically defined as keeping a fire below a certain size or achieving a stage of control of “Being Held” before peak burning time the day following its report. The success of this subsystem has been explored previously with various modelling methodologies. In addition, the effectiveness of the IA subsystem may depend on the performance of the subsystems preceding it. Our research currently focuses on the dispatch subsystem, which represents the fire management agencies' deployment of resources in response to detected fires. Using historical data provided by the Ontario Ministry of Natural Resources and Forestry, we evaluate the effectiveness of the dispatch system and the influence of covariates such as seasonality, weather conditions, fire load and location. Through generalized additive modelling, we establish baseline results that summarize the impact of these covariates. These baseline results will guide future modelling efforts in quantifying dispatch system success. Performance measures of the dispatch system will provide fire management officials with additional information to support decision-making and assess the effectiveness of operational response to wildland fire.

---

## Jessica Dobbin, Statistics - Hazards, Risks and Resilience, supervised by Dr. Douglas Woolford

### *Investigating Demanding Wildland Fires in Ontario*

Every year, Canadian wildland fire management agencies spend hundreds of millions of dollars responding to fires on the landscape. Fire managers must anticipate fire activity and allocate limited resources to strategically protect values while limiting the risk to fire management personnel. Over the last twenty years, Canada has experienced fire disasters such as the Fort McMurray wildfire in 2016, which required integrated and collaborative efforts from municipal, provincial, as well as federal organizations and direct costs of over \$5 billion. Ontario has not

recently had a fire with an impact of a similar magnitude, though had conditions varied slightly, this may not be the case. With recent research on climate change projecting an increase in frequency, duration and severity wildfires, the need to quantify the risk to communities in the Wildland Urban Interface and other important values experiencing a fire that requires resources with limited availability has become more apparent. In this presentation, the background and initial efforts into exploring this problem are discussed, with some discussion on future work.