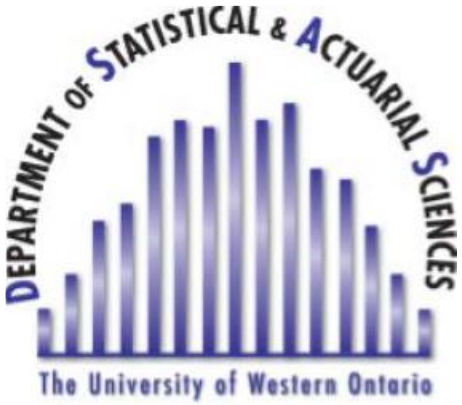
DEPARTMENT OF STATISTICAL & ACTUARIAL SCIENCES

Master's Day

July 28, 2022



18th Annual Master's Day

Contents

Schedule of Events

Location: Physics & Astronomy Building (PAB) Atrium

9:00 a.m. Student and Faculty coffee and mingle

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

9:15 a.m. **Session One**

Dr. Jiandong Ren
- Peixen Fang (MSc AS)

Dr. Marcos Escobar-Anel
- Samira Alipour (MSc FM)
- Sara Ansari (MSc FM)

Dr. Shu Li
- Anbang Liang (MSc FM)
- Yihui Lin (MSc FM)

Dr. Ricardas Zitikis
- Lingling Lin (MSc SS)
- Yanzhi Wang (MSc SS)

10:15 a.m. Break and set up for session two

10:30 a.m. **Session Two**

Dr. Jiandong Ren
- Pedro Assunção Rangel (MSc SS)
- Sanghyun Jung (MSc SS)

Dr. Marcos Escobar-Anel
- Aichen Liu (MSc FM)
- Yixuan Liu (MSc FM)

Dr. Ricardas Zitikis
- Yao Li (MSc SS)
- Braedan Walker (MSc SS)

11:30 a.m. Closing remarks

11:35 a.m. Light lunch and awards

Titles and Abstracts

Samira Alipour, Financial Modelling, supervised by Dr. Cristiàn Bravo Roman

Modelling Probability of Default Using Macro Variables

Financial institutions are required to estimate their future credit risk using the Probability of Default (PD) calculations. PDs have a significant systematic component which is affected by the variability of economic, political, and social conditions. Understanding the systematic component of default rates helps to better estimate the future credit risk which a financial institution could be exposed to.

In this study we used 93 macro-economic variables to find this empirical relationship. Out of 469 models considered, the best model was formed using three indices of Unemployment Rate, Median Weeks Unemployed and Gross Domestic Product.

We further improved our predictions using a SARIMAX model and stepwise variable selection. This approach resulted in the selection of time series model with parameters (0, 1, 1) x (1, 0, 1, 12)12 and exogenous variables of Sticky Price Consumer Price Index, Bank Prime Loan Rate and Total Borrowings from the Federal Reserve.

Sara Ansari, Financial Modelling, supervised by Dr. Ricardas Zitikis

Anomaly detection in financial markets

Anomaly detection is the process of finding observations which are different from the normal behaviour of data. Applications include detecting fraud, network intrusions, and other rare events which make data unusual. In Finance, market manipulation is one of the common fraudulent activities in the financial sector. Stock manipulation is the result of changing the behaviour of others to make money. My summer project 1) evaluates three machine learning algorithms to recognize anomalies on five large real-world, labelled datasets of anomalous stock market data where market manipulation has occurred and 2) compares the algorithms from an array of different perspectives. We also consider some popular technical indicators for anomaly detection and price forecasting. When comparing the results of different methods we have found that the LSTM method is more capable of anomaly detection but the difference in methods is not significant. Finally, using NN deep learning, we forecast stocks by considering and resolving anomalies.

Pedro Assunção Rangel, Statistics, supervised by Dr. Camila de Souza

Model-based Clustering for Zero-inflated Poisson Data via an EM Algorithm

The motivation for this project comes from the problem of clustering single-cell RNA sequencing (scRNA-seq) data. Clustering single-cell RNA sequencing data allows researchers to characterize the gene expression profiles of different cell subtypes. However, challenges arise due to the high number of zeros present in the dataset. These zeros can be true biological zeros (no expression) or caused by sequencing errors. Therefore, there is a need to develop clustering methods that can handle the excess of zeros (zero-inflation) in scRNA-seq data. In this project, we consider a model-based approach to cluster cells based on scRNA-seq raw counts (number of reads aligned to each gene) assuming zero-inflated Poisson (ZIP) distributions. We implement an expectation-maximization (EM) algorithm in R to infer cluster assignments and model parameters. A simulation study is conducted to assess the performance of the proposed EM algorithm across different scenarios.

Peixin Fang, Actuarial Science, supervised by Dr. Kristina Sendova

A study on how long one's saving lasts after retirement.

Saving enough to maintain the same standard of living after retirement has been an important topic for Canadians. Traditionally, a pre-fixed replacement ratio is used to determine the amount of wealth a person would need to maintain their pre-retirement lifestyle in retirement. In Canada, the ratio is cited as around 70 to 85 percent of the income before retirement. In this research, we first model the frequency and the severity (net amount) of the trades for each retiree by summarizing their portfolios. After estimating and testing the parameters of the frequency and severity models, an aggregate ruin model is obtained. In addition to the discrete frequency model, a ruin model with the given initial capital of each client is simulated under the assumption of a homogeneous Poisson process. The expected time of ruin is obtained through the simulation study. Subsequently, the age when the retirees deplete their accounts is calculated based on their age and their wealth. This research provides a perspective of how prepared the retirees in Canada are for their future lives.

Sanghyun Jung, Statistics, supervised by Dr. Cristiàn Bravo Roman

The Impact of Publicly-Funded Business Advisory Services on Growth and Financial Wellness of Small and Medium-Sized Enterprises.

In this work, we present a causal analysis for the impact of business advisory services against training services from government programs that support small businesses on new entrepreneurs in Chile. The Chilean institution SERCOTEC provides two programs, differing in reach and cost. The first is a program focused on training entrepreneurs in different business skills. In contrast, the second assigns a direct advisor to each entrepreneur and offers personalized support during a set period. Comparing with the course service, we measure the impact the personalized support had one year after the support took place in terms of new contracts, new businesses, an increase in sales, and financial inclusion effects (new current accounts, new debt, and positive and negative financial events). We found that there is a positive average treatment effect (ATE) of the advisory services on investment value, while advisory services do not affect contract value. Also, the advisory services have positive ATE on debt, which can be explained by the increased investment value. Lastly, the advisory services increase the probability to be in good financial standing in terms of debt if an SME was in a bad financial status when they received the treatment, while the services also increase the probability to worsen their status if an SME was initially in a good financial status. Overall, the advisory services provide dynamism in the Chilean economy at the expense of a small increase in financial strain.

Yao Li, Statistics, supervised by Dr. Katsu Goda

Tsunami Early Warning Algorithm Based on Ocean Bottom Sensor S-net system with Statistical Methods in Tohoku, Japan

A tsunami is one of the most serious natural disasters, posing a great threat to the safety of people's lives and properties in coastal communities. The purpose of this study is to develop an effective and robust algorithm for tsunami early warning using tsunami wave data from ocean bottom sensor (OBS) arrays with statistical methods. This study focuses on the Tohoku region of Japan, where a new OBS S-net system, consisting of 150 sensors, has been deployed. To calibrate the tsunami early warning system using realistic tsunami wave profiles at the S-net stations, 4000 stochastic tsunami simulations have been generated. In this study, Multiple Linear Regression (MLR) and Random Forest (RF) are used. MLR analysis is first used to build models, and then AIC forward selection and knee-point are applied to select sensors that improve the accuracy the most significantly. The model performance is compared against the base model, which only uses the earthquake magnitude and location. Then, the model performance with respect to different model complexity and methods are compared. The result indicates that estimating tsunami height and loss via S-net improves accuracy. An optimal waiting time is identified to be five minutes, which contains enough information to estimate the tsunami height and loss and leaves relatively plenty of time for people evacuating from the coast. Besides, adding five stations from the S-net improves the accuracy of the estimation most significantly and this number of stations can be more easily deployed globally from economical perspectives. Moreover, RF has better performance at forecasting the tsunami than MLR, especially when the number of input variables is relatively small.

Anbang Liang, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

A comparison of existing Affine GARCH models

Option pricing models continue to evolve due to the increased complexity of data. Researchers have to reduce the gap between model assumptions and reality in the search for better models. 20 years after the seminal work of Black-Scholes (1973) based on a Geometric Brownian Motion, Heston (1993) proposed a continuous-time stochastic volatility model closer to reality. The main drawback of this model is that it is difficult to estimate. To solve this problem, Heston and Nandi (2000) developed a closed-form option pricing model, known as the HN-GARCH model. Christoffersen and Jacobs (2006) extended the model to allow for non-Gaussian innovations, using the Inverse-Gaussian(IG)-GARCH model, which boasted conditional skewness, and leverage effect; more recently, Ornthanalai (2014) proposed a model with Lévy jump innovations, known as the Lévy GARCH. These models capture the typical fact of seasonal volatility, and they are more easily implemented than continuous-time models. More importantly, they all lead to closed-from moment generating functions, therefore permitting analytical pricing of financial derivatives.

Lingling Lin, Statistics, supervised by Dr. Jörn Diedrichsen

Taking into account the covariance of noise for inference on multivariate data: an application on functional magnetic resonance imaging (fMRI) data

Recent analysis methods aim to compare different statistical models of brain computation by investigating how brain activity patterns between those models relate to experimental conditions, with the number of features (voxels) often being larger than the number of observations. Most of these methods assume that the noise is independent across voxels. However, this is not the case: the noise covariance of neighboring brain locations (voxels) is strongly correlated, and this covariance structure is also stable across different experimental runs for the same subject. Ignoring the covariance of noise in multivariate analysis has been shown to lead to sub-optimal or, worse, erroneous inference. To obtain better inference, we explored ways to provide better estimates of the noise covariance structure by combining information from the residuals of the first level Generalized Linear Model, spatial location of the voxels, and anatomical structure of the brain. Next, we evaluated the impact of noise covariance estimates on statistical inference in Representational similarity analysis (RSA), a popular method for multivariate analysis of fMRI data. For our evaluation criterion, we chose the reliability of the estimated second moment of activity patterns obtained by "prewhitening" the data using the noise covariance estimates between voxels. Our results indicate that the proposed estimates improve the reliability and hence the inference (e.g., model selection accuracy) of RSA.

Yihui Lin, Financial Modelling, supervised by Dr. Marcos Escobar-Anel

Review and improvement of hedging approaches

Hedging is a way to reduce business risk while still making a profit on an investment. This project focuses on comparing the actual hedging effects of various delta hedging approaches. We use the a dataset of options on APPL (Apple Inc) to compare the hedging effectiveness of the various approaches. We consider a total of seven scenarios: the traditional Black Scholes approach, the minimum variance delta hedging approximation, the upper and lower bounds of the delta band of the Whalley-Wilmott approach, and the upper, median, and lower bounds of the delta band of the Zakamouline approach. In our sample dataset, the Whalley-Wilmott and Zakamouline approaches have better performance over other approaches.

Aichen Liu, Financial Modelling, supervised by Dr. Shu Li

The Impact of Investment Behaviors on Investors' Financial Wellness

In Canada, most portfolio management is done by financial advisors or dealers. Compared to institutional investors, retail investors face disadvantages such as weak purchasing power and inexperience, which increase their probability of loss. Hence, given limited resources, how should retail investors trade to achieve better portfolio performance? In this project, we investigate two main factors of portfolio trading: trading frequency and portfolio allocation strategy. In particular, we compare the effectiveness of three investment strategies (namely, growth strategy, value strategy and no-trading strategy) under three different trading frequency schemes: random trading, high-frequency systematic trading and low-frequency systematic trading via the simulation approach. A combined dataset of 14,767 available securities information is used to calibrate the parameters for our stochastic models. Our numerical study shows that the static no-trading strategy is the best among the three trading strategies in terms of returns. However, depending on the investor's trading strategy, the high/low-frequency trading may also achieve high returns. We also present sensitivity analysis for key parameter assumptions

Yixuan Liu, Financial Modelling, supervised by Dr. Reg Kulperger & Dr. Hao Yu

A Comparison Study on Traditional versus Robust Portfolio Optimization on Risky Assets.

Market volatility is inevitable. When making investments, investors either want a higher return or a lower risk. Under the traditional portfolio optimization theory, we assume the portfolio's volatility follows a multivariate normal distribution. This report introduces another model called Robust optimization, which considers the model-level uncertainty with semi-G-normal. The model displays the volatility as a varied number fluctuates within an interval. The Robust model generates a portfolio with the minimum variance center and fluctuating range. This report discusses two risky assets and uses two ten-year stock movements (2010 to 2020) as training set. The risk profile investigates that Robust optimization leads to a smaller variance uncertainty which illustrates that robust optimization performs better than traditional optimization. By observing two additional years' stock (2020 to 2022) as testing set and considering a half-year rebalancing, Robust optimization significantly overcomes the traditional method in terms of ROI, volatility center and volatility ambiguity.

Braedan Walker, Statistics, supervised by Dr. Simon Bonner

A Spike-and-Slab Poisson Regression Approach to Lipid Quantification

Due to their critical roles in cell membrane structure or as an energy source, lipids have been a point of interest for many applied chemists. Mass spectrometry (MS) is a chemical method that involves identifying and quantifying the different compounds present within a biological sample. This is a challenge because there are many lipids that have masses that are too close to be distinguished. One potential way to differentiate between them is through their isotope patterns. The goal of this project was to develop a model that improves the identification and quantification of lipids in a biological sample by incorporating the information in the isotope patterns. To do this, we developed a spike-and-slab Poisson GLM. The spike-and-slab model is a Bayesian regression approach that models the contribution of candidate lipids in a sample in two respects: the probability that a compound is present in the sample and its relative abundance given that it is present. Models were fit on real MS data via MCMC sampling implemented with JAGS. Beyond the probabilities and abundances given by the model parameters, the model also allows us to estimate the noise within an MS scan, performs variable selection, and allows us to observe omissions within the set of candidate lipids while achieving low Pearson residuals on areas of the spectrum without gaps.

Yanzhi Wang, Statistics, supervised by Dr. Wenqing He

The Vaccine Effectiveness Against SARS-CoV-2 Omicron Variant: A Meta-Analysis

With the Omicron SARS-CoV-2 variant rapidly spreading worldwide, it is of great interest to evaluate how effective the vaccine is in protecting COVID-19 patients from severe disease outcomes. In this project, we use a meta-analysis approach to explore the effectiveness of different vaccines and vaccination stages against different disease outcomes. The results show that one-dose booster increases the vaccine effectiveness for the same vaccine type of full vaccination regarding each disease outcome, the vaccine effectiveness of the vaccine against severe disease outcomes is higher than that against non-severe outcomes. Aside from the scenario using full vaccination to protect severe disease outcomes, the vaccine effectiveness of the mRNA vaccine for the other scenarios is higher than that of the non-mRNA vaccine.