CONFERENCE AGENDA - AT A GLANCE

Conference Venue: Federation Hall, University of Waterloo,
200 University Avenue West, Waterloo,
Ontario, N2L 3G1, Canada

Friday December 1, 2017

Time	Room	Event
07:45-17:00	Lobby	Registration
07:45-08:15	Lobby	Refreshments
08:15-08:30	Main Hall	Opening Session
08:30-09:10	Main Hall	Invited Session 1
09:10-10:25	Main Hall	Panel Discussion in Life Insurance
10:25-10:50	Lobby	Refreshment Break
10:50-12:05	Main Hall	Parallel Session 1A
	Westmount Boardroom	Parallel Session 1B
12:05-13:35	Main Hall	Lunch and President's Speech
13:35-14:55	Main Hall	Invited Session 2
14:55-15:20	Lobby	Refreshment Break
15:20-16:10	Main Hall	Parallel Session 2A
	Westmount Boardroom	Parallel Session 2B
16:10-17:25	Main Hall	Panel Discussion in P&C Insurance
17:30-18:30	Columbia A&B	Pre-banquet Reception
18:30-21:30	Main Hall	Conference Banquet & President's Speech

Saturday December 2, 2017

Time	Room	Event
07:50-08:20	Lobby	Registration and Refreshment
08:20-09:00	Main Hall	Invited Session 3
09:00-09:50	Main Hall	Parallel Session 3A
	Westmount Boardroom	Parallel Session 3B
09:50-10:15	Lobby	Refreshment Break
10:15-11:30	Main Hall	Parallel Session 4A
	Westmount Boardroom	Parallel Session 4B
11:30-13:00	Main Hall	Lunch
13:00-14:15	Main Hall	Case Study Session
14:15-14:25	Main Hall	Closing Remarks

CONFERENCE AGENDA—DETAILED

Friday December 1, 2017

Time	Room	Event
07:45-17:00	Lobby	Registration
07:45-08:15	Lobby	Refreshments
08:15-08:30	Main Hall	**Opening Session** Moderator: Chengguo Weng Speakers: • Stephen M. Watt, Dean of the Faculty of Mathematics • Stefan Steiner, Chair of the Department of Statistics and Actuarial Science
08:30-09:10	Main Hall	**Invited Session 1** Moderator: Chengguo Weng Speaker: Patrick L. Brockett, "*Rapid Estimation of Disaster Relief Fund Distribution: Iterative Learning with Diverse Geospatial Data Inputs*", University of Texas at Austin, USA
09:10-10:25	Main Hall	**Panel Discussion in Life Insurance** Moderator: Kevin Pledge, Claim Analytics Panelists: • Ian Bancroft, Sun Life Financial • Kevin Pledge, Claim Analytics • Jean-Yves Rioux, Deloitte Canada • Eugene Wen, Manulife Financial
10:25-10:50	Lobby	Refreshment Break
10:50-12:05	Main Hall	**Parallel Session 1A:** Modern Analytics for Insurance Premium Rating Moderator: Jose Garrido Speakers: • Mathieu Boudreault, "*Pricing River Flood Risk Using a Hierarchical Model Based Upon Physics, Statistics and Actuarial Science*", The Université du Québec à Montréal, Canada • Liqun Diao, "*Regression Tree Credibility Model*", University of Waterloo, Canada • Jose Garrido, "*Bayesian Credibility for GLMs*", Concordia University, Canada
10:50-12:05	Westmount Boardroom	**Parallel Session 1B:** Predictive Analytics for Variable Annuities Moderator: Guojun Gan Speakers: • Ou Dang, "*Efficient Nested Simulation for Conditional Tail Expectation of Variable Annuities*", University of Waterloo, Canada • Mingbin Feng, "*Efficient Valuation of Large Portfolio of Variable Annuities via Simulation*", University of Waterloo, Canada • Guojun Gan, "*An Introduction to Data Clustering with Actuarial Applications*", University of Connecticut, USA

Time	Room	Event
12:05-13:35	Main Hall	**Lunch and President's Speech** Moderator: Mary Hardy Speaker: Jeremy J. Brown, Immediate Past-President of the Society of Actuaries
13:35-14:55	Main Hall	**Invited Session 2** Moderator: Tony Wirjanto Speakers: • Greg Taylor, "*The Long Road to Enlightenment: Loss Reserving Models from the Past, with Some Speculation on the Future*", University of New South Wales, Australia • Edward W. (Jed) Frees, "*Predictive Analytics and Medical Errors*", University of Wisconsin-Madison, USA
14:55-15:20	Lobby	Refreshment Break
15:20-16:10	Main Hall	**Parallel Session 2A:** Neural Network for Insurance Applications Moderator: Hongjun Ha Speakers: • Betül Zehra Karagül, "*Motor Own Damage Insurance Pricing with Neural Networks*", Hacettepe University, Turkey • Hongjun Ha, "*A Neural Network Monte Carlo Evaluation of Withdrawal Benefits in Variable Annuities*", Saint Joseph's University, USA
15:20-16:10	Westmount Boardroom	**Parallel Session 2B:** Predictive Modelling for Insurance Applications Moderator: Brian Hartman Speakers: • Lina Xu, "*Approaches to Validating Methodologies and Models with Insurance Applications*", Columbia University, USA • Brian Hartman, "*Estimating Loss Reserves Using Hierarchical Bayesian Gaussian Process Regression with Input Warping*", Brigham Young University, USA
16:10-17:25	Main Hall	**Panel Discussion in P&C Insurance** Moderator: Chris Fievoli, Canadian Institute of Actuaries Panelists: • Jeffrey Baer, Economical Insurance • Denise Cheung, Aviva Canada • Sebastien Bernard, Intact Insurance • Frédérick Guillot, The Co-operators Insurance and Financial Services
17:30-18:30	Columbia A&B	Pre-banquet Reception
18:30-21:30	Main Hall	**Conference Banquet and President's Speech** Moderator: Chris Fievoli Speaker: Sharon Giffen, President of the Canadian Institute of Actuaries

Saturday December 2, 2017

Time	Room	Event
07:50-08:20	Lobby	Registration and Refreshment
08:20-09:00	Main Hall	**Invited Session 3** Moderator: Ken Seng Tan Speaker: Ian Duncan, "*Healthcare: How Can We Harness Predictive Analytics for Patients, Providers and Payers?*", University of California Santa Barbara, USA
09:00-09:50	Main Hall	**Parallel Session 3A:** Data Mining in Insurance and Finance Moderator: Emiliano A. Valdez Speakers: • Thomas W. Sproul, "*Simulated Posterior Predictive Checks for Mixture Models*", University of Rhode Island, USA • Emiliano A. Valdez, "*Data Mining Techniques for Actuaries: An Overview*", University of Connecticut, USA
09:00-09:50	Westmount Boardroom	**Parallel Session 3B:** Predictive Modelling in Insurance and Finance Moderator: Mostafa Pouralizadeh Jobejarkoli Speakers: • Himchan Jeong, "*A Predictive Random Effects Model of Dependent Claims Frequency and Severity*", University of Connecticut, USA • Mostafa Pouralizadeh Jobejarkoli, "*Forecasting Implied Volatility in a Risk Management Problem*", Allameh Tabataba'i University, Iran
09:50-10:15		Refreshment Break
10:15-11:30	Main Hall	**Parallel Session 4A:** Predictive Analytics for Longevity Risk Management Moderator: Min Ji Speakers: • Adelaide Di Wu, "*Non-parametric Bühlmann Credibility-based Approaches to Modeling Multi-Population Mortality Rates*", Simon Fraser University, Canada • Rui Zhou, "*Drivers of Mortality Dynamics: Identifying Age/Period/Cohort Components of Historical US Mortality Improvements*", University of Manitoba, Canada • Min Ji, "*Predictive Analytics for Modeling Threshold Life Tables*", Towson University, USA
10:15-11:30	Westmount Boardroom	**Parallel Session 4B:** Predictive Analytics for AgriInsurance and AgriBusiness Moderator: Hibod Assa Speakers: • Kai Liu, "*Predicting Forage Crop Loss in the Presence of Systemic Weather Risk Using Remote Sensing Data: An Insurance Application*", University of Prince Edward Island, Canada • Wenjun Zhu, "*Factor Forecasting for Agricultural Production Process*", Nanyang Technological University, Singapore

Time	Room	Event
		• Hirbod Assa, "*How to Design an Optimal Insurance on Price Fluctuation: Example from Agricultural Market*", University of Liverpool, UK
11:30-13:00	Main Hall	Lunch
13:00-14:15	Main Hall	**Case Study Session** Moderator: Chengguo Weng Speakers: • Jeffrey Baer, Economical Insurance • Dragos Capan, Manulife Financial • Denise Cheung, Aviva Canada
14:15-14:25	Main Hall	Closing Remarks by Ken Seng Tan

Invited Sessions

Invited Session 1:

Title: Rapid Estimation of Disaster Relief Fund Distribution: Iterative Learning with Diverse Geospatial Data Inputs
Authors: Patrick Brockett*, Rajiv Garg, Linda Golden, and Yuxin Zhang
University of Texas at Austin
Abstract: Recovery efforts after a natural disaster require rapid deployment of both volunteer support and financial resources. While financial resources for impacted regions may be available from federal and state governments, charitable organizations, private entities, and global citizens, the allocation and distribution of these funds is challenging. This challenge is further aggravated when the magnitude of the natural disaster is more severe and spans multiple regions where financial support is not sufficient for complete recovery of all local regions. Thus, in this study we provide a framework for rapid geospatial estimation of the losses in all local regions with a learning process to improve accuracy iteratively as more disaster data becomes available. This framework is valuable not only to predict the extent of losses (even before actual data becomes available) but also to provide guidance for how to rationally distribute relief funds to local governments for rapid recovery. This predictive/probabilistic modeling approach utilizes all three characteristics of big data (volume, variety, and velocity) to improve the accuracy in damage prediction. While this study focuses on individual and home assistance, we also provide discussion on extending this work to business and infrastructure losses.

Invited Session 2:

Title: The Long Road to Enlightenment: Loss Reserving Models from the Past, with Some Speculation on the Future
Authors: Greg Taylor*, University of New South Wales
Abstract: The presentation will survey the evolution of models of claim data arrays, with particular reference to loss reserving. Generally, the focus will not be on the technicalities and fine detail of these models, but rather on their broad categories, how one category evolves from another, and the extent to which they affect actuarial practice. The pace of evolution has increased in recent decades, and increased computing power has enabled the development of more comprehensive and detailed models. These include models that fall into the “Predictive Analytics” and “Machine Learning” categories. There are a number of contenders for the crown among future categories of model, and the desirability of various possible future directions will be discussed.

Title: Predictive Analytics and Medical Errors
Authors: Edward W. (Jed) Frees*, University of Wisconsin-Madison
Abstract: Insurance as a discipline has long embraced analytics and market trends signal an even stronger relationship going forward. This presentation describes contributions of analytics and statistical methods that further our understanding of insurance operations and markets. I will provide an introduction and a summary of selected resources for learning more about insurance analytics. To be specific, this talk illustrates the use of predictive analytics in the context of medical errors. Analyzing medical errors helps improve healthcare systems and, through a type of insurance known as "medical professional liability" or "medical malpractice," we have the ability to analyze medical errors using data from outside the healthcare system. In the spirit of modern analytics, this talk describes the application of data from several different sources. These different

sources give different insights into a specific problem facing the medical malpractice community familiar to actuaries, the relative importance of upper limits (or caps) on insurance payouts for non-economic damages (e.g., pain and suffering). This topic is important to the industry in that many courts are considering the legality of such limitations. All stakeholders, including patients, physicians, hospitals, lawyers, and the general public, are interested in the implications of removing limitations on caps. This talk demonstrates how we can use data and analytics to inform the many different stakeholders on this issue.

Invited Session 3:

Title: Healthcare: How Can We Harness Predictive Analytics for Patients, Providers and Payers?
Authors: Ian Duncan*, University of California Santa Barbara
Abstract: In the United States healthcare is approaching 20% of GDP; other countries are not far behind. Yet budgets are under considerable pressure and demand for healthcare services outstrips supply and our ability to pay. Predictive Analytics offers potential to help identify patients before they become high-cost consumers, yet the promise of big data and predictive analytics somehow seems to be always in the future. Why is this? This talk will discuss predictive analytics and the use of big data in healthcare and the problems and promise of their implementation.

Panel Discussion and Case Study Sessions

Panel Discussion in Life Insurance

Session description: This panel offers practitioner perspectives on applications of predictive analytics in the life insurance industry. Speakers will address such questions as: What are the key success factors for a life insurance predictive analytics team, and a data scientist? Is an actuarial data scientist different than other data scientists? What are the typical hurdles in setting up an insurance analytics function? How do you maintain the balance between organizational agility and compliance? What types of techniques are used for solving life insurance problems with analytics? How much effort is typically spent on preparing/cleaning the data? What are the key metrics used to assess the predictive power of a model?

Panelists:

- Ian Bancroft, Sun Life Financial
- Kevin Pledge, Claim Analytics
- Jean-Yves Rioux, Deloitte Canada
- Eugene Wen, Manulife Financial

Panel Discussion in P&C Insurance

Session description: Predictive analytics is gaining increased importance in the property and casualty insurance industry, as P&C companies are looking for ways better integrate diverse sources of data into their operations. Our panelists, each practitioner in this area, will present their insights on how predictive analytics are being used in their own companies, and how future research initiatives could potentially help.

Panelists:

- Jeffrey Baer, Economical Insurance
- Denise Cheung, Aviva Canada
- Sebastien Bernard, Intact Insurance
- Frédérick Guillot, The Co-operators Insurance and Financial Services

Case Study Session

Panelists:

- Jeffrey Baer, Economical Insurance
- Denise Cheung, Aviva Canada
- Dragos Capan, Manulife Financial

Parallel Sessions

Parallel Session 1A: Modern Analytics for Insurance Premium Rating

Title: Pricing River Flood Risk Using a Hierarchical Model Based upon Physics, Statistics and Actuarial Science
Authors: Mathieu Boudreault*, Mathieu Pigeon, Jean-Mathieu Potvin,
Abstracts: Following from the Québec (2011) and Alberta (2013) floods, private insurance companies have been invited to play an important role in covering flood risk in Canada, as it is already the case in the other G8 countries. In this context, we propose a hierarchical approach to generate reliable simulations of flood losses from the Chaudière River (Québec). Scenarios from a chain of physically-based models representing the climate, hydrology and hydraulicity as well as simulations from an econometric model are both used to represent water discharge at a specific location on the river. Then, using the distribution of water discharge, a hydrologic/hydraulic model (HEC-RAS) calculates water levels (stage) in a 3D map along the river. Moreover, we employ civil engineering data to map the level of damage as a function of water level. Finally, we determine the distribution of the total loss for a portfolio of insureds as well as the distribution of the loss for each individual policy using Monte Carlo simulations. This project is part of a much larger one by which we seek to assess flood risk in the entire province to further analyze the feasibility of various risk sharing plans between individuals, governments (federal, provincial and municipal) and the private sector. This is joint work with Ouranos, a consortium on regional climatology and adaptation to climate change, and with the Quebec Water Expertise Center from the Government of Quebec (Direction de l'expertise hydrique du Québec du Ministère du développement durable, de l'environnement et de la lutte contre les changements climatiques, MDDELCC).

Title: Regression Tree Credibility Model
Authors: Liqun Diao*, Chengguo Weng
Abstracts: This paper applies learning techniques to credibility theory and proposes a regression trees based algorithm to integrate covariate information for credibility premium prediction. The algorithm recursively binary partitions a collective of individual risks into mutually exclusive sub-collectives, and consequently applies the classical Bühlmann-Straub credibility formula for the prediction of individual net premium within each sub-collective. It provides a flexible way to integrate covariate information into individual net premium prediction. It is appealing for capturing non-linear and/or interaction covariate effects. It automatically selects influential covariate variables for premium prediction with no additional ex ante variable selection procedure required. The superiority in prediction accuracy of the proposed model is demonstrated by extensive simulation studies. An application to the U.S. Medicare data is presented.

Title: Bayesian Credibility for GLMs
Authors: Oscar Alberto Quijano Xacur, Jose Garrido*
Abstracts: We revisit the classical credibility results of Jewell (1974) and Bühlmann (1969) to obtain credibility premiums for a GLM severity model using a modern Bayesian, computational approach. Here prior distributions are chosen from out-of-sample information, without restrictions to be conjugate to the severity distribution. Then we use as a loss function the relative entropy between the "true"' and the estimated models, without restricting credibility premiums to be linear. A numerical illustration on real data shows the feasibility of the approach, now that computing power is cheap, and simulations software readily available.

Parallel Session 1B: Predictive Analytics for Variable Annuities

Title: Efficient Nested Simulation for Conditional Tail Expectation of Variable Annuities
Authors: Ou Dang*, Mingbin Feng, Mary Hardy
Abstracts: For valuation of Variable Annuity contracts with a dynamic hedging program using Monte Carlo methods, nested simulation is often required. The process is computationally challenging, sometimes prohibitively so, in many practical applications.
We propose a simulation procedure for estimating the Conditional Tail Expectation (CTE) of liabilities of a Variable Annuity dynamic hedging strategy. In a CTE calculation, tail scenarios, i.e., the scenarios that result in extreme losses, are most relevant. Thus, correctly identifying those scenarios would greatly improve the efficiency in a nested simulation. The proposed procedure takes advantage of the special structure of the CTE by first identifying a small set of potential tail scenarios from the first tier of simulation. We then focus the simulation budget on only those scenarios. We conduct extensive numerical experiments on different guarantee types and different stochastic stock return dynamics. The numerical results show that, when given a fixed simulation budget, the proposed procedure can improve the accuracy of CTE estimation by an order of magnitude compared to a standard nested simulation.

Title: Efficient Valuation of Large Portfolio of Variable Annuities via Simulation
Authors: Mingbin Feng*, Zhenni Tan, Jiayi Zheng
Abstracts: Variable Annuities (VAs) have been popular insurance products in practice and have attracted significant research attention in the last few decades. From the insurer's perspective, it is essential to valuate a large portfolio of VAs for different strategic goals such as those in enterprise risk management. For accurate valuation of individual VA contracts, Monte Carlo simulation is usually required due to the contracts complexities. However, the computational resources required for valuing all contracts in a large VA portfolio using standard Monte Carlo could be prohibitively expensive. In the last few years, there have been numerous research efforts which apply machine learning methods to the valuation of large portfolios of VAs. All of the proposed methods show superior computational efficiencies compared to the standard Monte Carlo experiment. However, it is unclear whether these proposals have leveraged the full power of the employed machine learning methods. The current research aims to provide a comprehensive comparison among some of recently proposed machine learning methods for large VA portfolio valuation. In particular, we identify pitfalls in some of the methods and propose corresponding improvements. Moreover, we propose, analyze, and test a new valuation method based on our suggested improvements. We show that resulting procedure has both higher accuracy and lower computational requirement than the previously proposed methods.

Title: An Introduction to Data Clustering with Actuarial Applications
Authors: Guojun Gan*, Emiliano Valdez
Abstracts: Data clustering refers to the process of dividing a set of objects into homogeneous groups or clusters such that the objects in each cluster are more similar to each other than to those of other clusters. As one of the most popular tools for exploratory data analysis, data clustering has been applied into many scientific areas. In this presentation, we give a review of the basics of data clustering, such as distance measures and cluster validity, and different types of clustering algorithms. We also demonstrated the applications of data clustering in insurance by using two scalable clustering algorithms, the TFCM algorithm and the hierarchical k-means algorithm, to select representative variable annuity contracts, which are used to build predictive models. We found that the hierarchical k-means algorithm is efficient and produces high quality representative variable annuity contracts.

Parallel Session 2A: Neural Network for Insurance Applications

Title: Motor Own Damage Insurance Pricing with Neural Networks
Authors: Betül Zehra Karagül*, Manuel Morales
Abstract: Pricing is one of the most important actors which impacting insurer's ability to compete and profit realizations in the insurance sector and many factors must be taken into account in this process whether it is about the insured itself or not. Generalized Linear Models (GLM) are popular statistical models that are often used for non-life insurance pricing. Today, the use of the internet and fast access to the database have made the world smaller and data bigger. These changes have made collating and understanding data, predictive modelling and transforming huge repositories of data into actionable forms more difficult for actuaries. Neural Networks (NN), one of the deep learning methods, have become very popular in complex system modeling. They operate the systems in a similar fashion to the human brain and can be used in any context that you want to represent a set of data by a model to make some prediction. In this study our aim is to use NN for motor own damage (MOD) insurance pricing and compare the results with GLM.

Title: A Neural Network Monte Carlo Evaluation of Withdrawal Benefits in Variable Annuities
Authors: Hongjun Ha*, Daniel Bauer
Abstracts: Advanced life insurance products with exercise-dependent financial guarantees present challenging problems in view of pricing and risk management. In particular, due to the complexity of the guarantees and since practical valuation frameworks include a variety of stochastic risk factors, conventional methods that are based on the discretization of the underlying (Markov) state space may not be feasible. As a practical alternative, this paper explores the applicability of the Neural Network Monte Carlo (NNM) method familiar from machine learning community for predictions in this context. Unlike previous literature of American option pricing, we consider optionality beyond surrendering the contract, where we focus on popular withdrawal benefits --- so-called GMWBs --- within Variable Annuities. We introduce an approximation method for a value function via the neural network with a single layer and focus on how to train coefficients of a neural network based on sample paths generated from Monte Carol simulation. We commence our numerical analysis in a basic Black-Scholes framework, where we compare the NNM results to those from a discretization approach. We then extend the model to include various relevant risk factors and compare the results to those from the basic framework.

Parallel Session 2B: Predictive Modelling for Insurance Applications

Title: Approaches to Validating Methodologies and Models with Insurance Applications
Authors: Lina Xu*, Victor de la Pana, Shaun Wang
Abstracts: The paper introduces a new approach/method to validate and certify analytical models that used in the banking and insurance applications. A statistical approach, described in de la Pena et al (Journal of Risk 2006), of assessing the quality of risk measures: Quality Control of Risk Measures (QCRM), is applied to the over-dispersed Poisson (ODP) bootstrap model for unpaid loss reserve for a property-casualty insurance companies. Empirical tests have been conducted to assessing if the reserves generated from the ODP bootstrap of the paid chain-ladder method would be acceptable. In this paper we perform back-testing using actual historical reserve data of property-casualty insurance companies, as used in Leong/Wang/Chen (Variance, 2015). Our empirical test using QCRM measure rejects the ODP bootstrap model. However, the ODP bootstrap model with Wang's transform adjustment for systemic risks is accepted using the QCRM measure.

Title: Estimating Loss Reserves Using Hierarchical Bayesian Gaussian Process Regression with Input Warping
Authors: Nathan Lally, Brian Hartman*
Abstracts: We visualize the loss reserve runoff triangle as a spatially-organized data set. We apply Gaussian Process (GP) regression with input warping and several covariance functions to estimate ultimate claims. We then compare our results over a range of product lines, including workers' comp, medical malpractice, and personal auto. Even though the claims development of the lines are very different, the GP method is very flexible and can be applied without much customization. We find that our model generally outperforms the classical chain ladder model as well as the recently proposed hierarchical growth curve model of Guszcza (2008) in terms of point-wise predictive accuracy and produces dramatically better estimates of ultimate reserves.

Parallel Session 3A: Data Mining in Insurance and Finance

Title: Simulated Posterior Predictive Checks for Mixture Models
Authors: Thomas W. Sproul*, Joshua D. Woodard
Abstracts: Recent agricultural economics research has used mixture models in domains including crop yield modeling, reinsurance ratemaking, and even farmer behavior, because they can fit multi-modal heterogeneity in data. The number of mixture components is a critical aspect of model selection in insurance related applications, because missing or additional peaks in the estimated density lead to mispricing. Model selection is commonly achieved via cross validation (CV), via information criteria such as AIC and BIC, or via entropy based criteria such as the Integrated Completed Likelihood Criterion (designed for robustness to violating modeling assumptions). Unfortunately, the best approach may be data- and/or model-dependent, and no approach is known to be uniformly best for identifying the number of mixture components. Recent (somewhat ad hoc) approaches include establishing consensus across different criteria or reconciling a lack thereof using posterior class membership probabilities from Expectation-Maximization. We propose a new Bayesian method for model selection inspired by posterior predictive checks, where the checks are simulated to avoid double use of data. Given a set of candidates (already fitted) models, we simulate from each and estimate the distribution of models "chosen" by each selection procedure (CV, AIC, BIC, ICL), conditional on one model being true. We then estimate posterior probabilities for each model being true, conditional on the choice of procedure. Our methodology handles both model selection and finding the most informative selection procedure, given data. It also identifies cases where the selection procedures are unable to differentiate effectively between models, a sign that the models considered may be uninformative.

Title: Data Mining Techniques for Actuaries: An Overview
Authors: Banghee So, Guojun Gan, Emiliano A. Valdez*
Abstracts: Data mining involves the computational process of exploring and analyzing large amounts of data to uncover hidden and useful information. Such information is useful to process and efficiently reduce data into a more summarized, analytical representation. The ultimate goal of data mining is to be able to deliver predictive models applicable to new data. Predictive modeling is increasingly becoming an important function of an actuary in all areas of insurance: life, health, pensions, property and casualty. In this survey article, we explore and describe the data mining tasks associated with supervised and unsupervised learning. There are generally four primary data mining tasks: association rule learning, clustering, classification, and regression. With each data mining task, we illustrate, using real data whenever available, its potential applications in actuarial science and in different areas of insurance. We further demonstrate the

usefulness of these data mining techniques for actuaries to perform predictive analytics. Additionally, we briefly describe the emerging development of a new class of machine learning algorithms called deep structured learning. This is joint work with Banghee So and Guojun Gan, both from the University of Connecticut.

Parallel Session 3B: Predictive Modelling in Insurance and Finance

Title: A Predictive Random Effects Model of Dependent Claims Frequency and Severity
Authors: Himchan Jeong*, Emiliano A. Valdez
Abstracts: In the two-part regression model, it has conventionally assumed independence between claims frequency and severity, but there is an increase in interest in developing models to capture the possible dependence. In this paper, we explore the benefits of using random effects for predicting insurance claims observed longitudinally, or over a period of time, within a two-part framework relaxing the assumption of independence between claims frequency and severity. Especially, the (average) severity component can be modeled using any positive and continuous distributions but with covariates added to capture heterogeneity and dependence with frequency, and simultaneously introducing random effects to generate the unobserved structure of association when claims are observed over time. In constructing our predictive random effects model with dependence, we focus on families of distributions with conjugate random effects in order to be able to derive explicit moment formulas for the aggregate claims. In our search for such models for claims severity, we discovered that the family of GB2 distributions could be derived with an underlying distribution based on generalized gamma (G-Gamma) and a random effect based on generalized inverse gamma (GI-Gamma). This resulting family provides a flexible parametric distribution with four parameters describing scale and various shapes.

Title: Forecasting Implied Volatility in a Risk Management Problem
Authors: Hirbod Assa, Mostafa Pouralizadeh Jobejarkoli*, Abdolrahim Badamchizadeh
Abstracts: Implied volatility modeling helps to anticipate the future price fluctuation and has a crucial role in derivative pricing. However, further to derivative pricing forecasting implied volatility can have a great impact on asset risk management. In this paper, by setting a simple framework for asset risk management we will observe the impact of implied volatility modeling on optimal contracting. We consider a risk management problem for hedging volatile stock prices in a static framework. First, we show that to efficiently manage the risk, we need to take care of static arbitrages. That leads us to the second part of the paper where by ruling out the static arbitrages we propose a machine learning polynomial approach for parametrizing the implied volatility and we also control high bias and high variance during the fitting procedure. Using a regularization term, we observe that this term has a crucial role in removing butterfly arbitrage. The numerical results illustrate that the proposed algorithm most of the time yields an implied volatility smile which is also free from static arbitrage for out-the-money European call options. Finally, we compare the optimal design of contracts on a S&P 500 investment and see how prices evolve by considering different models. Also, SVI is a precious model of parameterization of total implied variance with nice properties, our proposed polynomial approach provides a cheap and simple model with high accuracy (low cross-validation error) mostly resulting in the absence of static arbitrage. The results show that using the proposed machine learning approach to parameterize total implied variance results in remarkable gain over a widely-used Stochastic Volatility Inspired (SVI) parameterization, since most of the time it makes a volatility surface free from static arbitrage with no need of repicking the parameters.

Parallel Session 4A: Predictive Analytics for Longevity Risk Management

Title: Non-parametric Bühlmann Credibility-based Approaches to Modeling Multi-population Mortality Rates
Authors: Cary Chi-Liang Tsai and Adelaide Di Wu*
Abstracts: Credibility theory is a set of quantitative tools which allow an insurance company to perform prospective experience rating on a risk or a group of risks. Inspired by the ideas of the joint-k, the co-integrated and the augmented common factor Lee-Carter modes (three extensions of the Lee-Carter model for multiple populations), in this paper we propose three corresponding non-parametric Bühlmann credibility based approaches to modeling multi-population mortality rates, the joint-k Bühlmann credibility model, the co-integrated Bühlmann credibility model, and the augmented common factor Bühlmann credibility model. We use the mortality data of both genders of the U.K., the U.S.A., and Japan from the Human Mortality Database. Mortality data for a selected age span and each of a series of year spans for two populations (both genders of a country) and six populations (both genders of all of the three countries), respectively, are first fitted to each of the six models (three Lee-Carter-based and three Bühlmann credibility-based). We then compare the forecasting performances of the six underlying mortality models with the measure of MAPE (mean absolute percentage error) between the predicted and observed mortality rates for three forecasting year spans (10, 20 and 30 years), respectively.

Title: Drivers of Mortality Dynamics: Identifying Age/period/cohort Components of Historical US Mortality Improvements
Authors: Johnny Siu-Hang Li, Rui Zhou*, Yanxin Liu, George Graziani, Jennifer Haid, Dale Hall, Andrew Peterson, Larry Pinzur
Abstracts: The goal of this paper is to obtain an Age/Period/Cohort (A/P/C) decomposition of historical U.S. mortality improvement. Two different routes to achieving this goal have been considered. In the first route, the desired components are obtained by fitting an A/P/C model directly to historical mortality improvement rates. In the second route, an A/P/C model is estimated to historical crude death rates and the desired components are then obtained by differencing the estimated model parameters. For each route, various possible A/P/C model structures are experimented, and are evaluated on the basis of their robustness to several factors (e.g., changes in the calibration window) and their ability to explain historical changes in mortality improvement. Based on the evaluation results, an A/P/C decomposition for each gender is recommended. The decomposition will be examined in a follow-up project, in which the linkages between the A/P/C components and certain intrinsic factors will be identified.

Title: Predictive Analytics for Modeling Threshold Life Tables
Authors: Min Ji*, Mostafa Aminzadeh, Min Deng
Abstracts: Advances in health care and improved living standards have led to growing number of people beyond age 90. Accurate modeling of mortality at advanced ages is crucial to the valuation of pension plans and life annuities, especially longevity annuities where annuity payments start very late in life, say 85. Parametric regression models extrapolate lifetime distribution to the advanced ages but do not guarantee goodness-of-fit. Threshold life tables offer a promising solution to this problem using piecewise distribution via the peaks-over-threshold (POT) approach in the extreme value family. The POT approach describes the behavior of large observations which exceed same high threshold, and the generalized Pareto distribution as the fore of the POT provides a unifying approach that can be used without making any assumption about the underlying distribution. However, parameter estimation for a threshold life table is challenging and it does not guarantee a smooth life table. In this research, we impose parameter constraints to achieve a smooth threshold life table and propose a Bayesian approach to derive predictive density of lifetime distribution. We consider several parametric prior distributions as well as Jeffreys priors to derive an appropriate predictive density of lifetime random variable. The predictive density is used to compute life expectancy and other measures of interest in Bayesian

framework. Simulations are also conducted to assess the accuracy of the estimates of interest based on the predictive density.

Parallel Session 4B: Predictive Analytics for AgriInsurance and AgriBusiness

Title: Predicting Forage Crop Loss in the Presence of Systemic Weather Risk Using Remote Sensing Data: An Insurance Application
Authors: Kai Liu*, Wenjun Zhu, Ken Seng Tan, Brock Porth, Milton Boyd, Lysa Porth
Abstracts: Agriculture lays the foundation for feeding the world population and serving the global economy. Agriculture industry is exposed to pronounced systemic weather risk (Okhrin, et al, 2013). It is estimated that adverse weather may be responsible for at least 70% of agricultural loss, including crop and livestock production (USDA, 2014). In addition, the possible threat of climate change is increasing frequency and severity of extreme weather events, and affecting more people (Hellmuth et al., 2009; IPCC, 2007). It is widely known that the traditional crop insurance design is subject to various challenges, including adverse selection and moral hazard (Chambers, 1989; Nelson and Loehman, 1987; Quiggen et al., 1994). To alleviate these issues, recent research has focused on index-based insurance (IBI). One critical issue is the basis risk which refers to the mismatch between the actual loss on the farm and the payment to the farmer based on the index. The objective of this research is to design a remote sensing index insurance contract which is able to improve agricultural losses forecasting, and hence reduce basis risk as much as possible. The index we are interested in is called Normalized Difference Vegetation Index (NDVI). NDVI is one of the most widely used vegetation indices to detect vegetation present and crop production with remote sensing data (Macdonald and Hall, 1980; Sellers, 1985; Tucker et al, 1983). By incorporating reporting system and remote sensing index, the optimal insurance design, which has the best case of zero basis risk, is suggested.

Title: Factor Forecasting for Agricultural Production Process
Authors: Hong Li, Wenjun Zhu*, Ken Seng Tan, Lysa Porth
Abstracts: Central to risk management at the farmer, insurer, reinsurer and government level is a robust and accurate yield forecasting model. In this paper, we propose a dynamic factor approach to develop a new approach for predicting crop yields. The proposed dynamic factor approach is able to summarize a large set of regressors by a much smaller set of latent factors, and hence achieve efficient dimension reduction. In addition, this approach determines the "key" factors to be included in the model through a data-driven method to attain effective variable selection. Using models in Lobell and Burke (2010) as a benchmark, we consider both the time-series model specification as well as the cross-section model specification. It is shown that the proposed dynamic factor model approach is able to improve the county-level time-series model in-sample forecasting ability by 65.31%, the state-level time-series model in-sample forecasting by 22.73%, and the cross-section model in-sample forecasting performance by 48.28%. In terms of the out-of-sample forecasting, we apply the leave-one-out cross-validation method to compare the relative root mean square errors (RMSE). The results indicate that compared to the benchmark models, our proposed dynamic factor approach is able to improve the out-of-sample forecasting accuracy (i.e., reduce relative RMSE) by 50%, 48.15%, and 43.75% for the county-level time-series model, state-level time-series model, and the cross-section model, respectively.

Title: How to Design an Optimal Insurance on Price Fluctuation: Example from Agricultural Market
Authors: Hirbod Assa*, Tianyuan (Nina) Ni, Meng (Simon) Wang

Abstracts: We introduce a framework to find the optimal insurance contracts to hedge against agricultural good volatile prices. While most of the papers in the literature find optimal contracts through optimizing an objective (e.g., minimizing risk, maximizing dividend or profit), we adopt a different approach by forecasting prices and filtering different sets of contracts based on a few Key Performance Indicators (or KPIs). Given most of the results in the literature indicate a two-layer policy (also called two-triggered or call-put spread) is optimal, we consider two-layer policies and introduce an approach for finding the best choices for the layers in a contract to keep the performance consistent. We develop a cost-efficient framework to find the best parametric or nonparametric model in predicting pricing data that are related to agriculture products. The framework considers linear models, seasonal adjustment models, time series models, long memory models, stochastic models and recurrent neural network models. Different agriculture data sets are used to test the performance. Our results can be used as a benchmark for common forecasting models in actuarial science, with agriculture price data to find the optimal agriculture insurance contracts.