SOUTH AFRICAN ACTUARIAL JOURNAL
SAAJ **20** (2020) 211–17

Abstracts of recent postgraduate theses and dissertations at South African universities

The profile and cost of end-of-life care in South Africa—the medical scheme's experience

By P Botha for MCom (Actuarial Science) at the University of Cape Town, 2020

South African medical schemes spend billions of Rands each year on the medical care costs of beneficiaries near their end of life. Hospi-centric benefit design, fee-for-service reimbursement arrangements and fragmented, silo-based delivery of care result in high, often unnecessary spending on beneficiaries near their end of life. Factors including an ageing population, increasing incidence rates of cancer and other non-communicable diseases, and high levels of multi-morbidity among beneficiaries near their end of life further drive up end-of-life care costs. Low levels of hospice and palliative care utilisation, a high proportion of deaths in-hospital and significant chemotherapy use in the last weeks of life point to potentially poor quality care near the end of life. The usual care pathway for serious illness near the end of life is care in private hospitals. This often entails resource-intensive care that includes aggressive care interventions right up until death. The result is potentially sub-optimal care and poor healthcare outcomes for many scheme beneficiaries and their surviving relatives. Understanding the complex nature of the end of life, the different care pathways, the available insurance benefits, the interactions between key stakeholders and the multitude of factors that are associated with end-of-life care costs are vital for end-of-life care reform. In order to increase value at the end of life, i.e. to increase quality and/or to reduce costs, benefit design reform, alternative reimbursement strategies, effective communication and multi-stakeholder buy-in is key.

Avoiding data-mining bias when testing technical analysis strategies—a methodological study

By R Douglas for MCom (Actuarial Science) at the University of Cape Town, 2020

When seeking to identify a profitable technical analysis (TA) strategy, a naive investigation will compare a large number of possible strategies using the same set of historical market data. This process can give rise to a significant data-mining bias, which can cause spurious results. There are various methods which account for this bias, with each one providing a different set of advantages and disadvantages. This dissertation compares three of these methods, the

step-wise Superior Predictive Ability (step-SPA) method of P-H Hsu, Y-C Hsu and Kuan (2010), the False Discovery Rate (FDR) method of Benjamini and Hochberg (1995) and the Monte Carlo Permutations (MCP) method of Masters (2006). The MCP method is also extended, using a step-wise algorithm, to allow it to identify multiple profitable strategies. The results of the comparison show that while both the FDR and extended MCP methods can be useful under certain circumstances, the step-SPA method is ultimately the most robust, making it the best choice in spite of its significant computational requirements and stricter set of assumptions.

Impact of claim intervention on employability of large SA insurer's group disability income claimants

By L du Toit for MCom (Actuarial Science) at the University of Cape Town, 2020

Long-term disability income (DI) claims account for a significant portion of life insurance companies' risk product policy liabilities. The annuity type benefits offered by these products, typically paid until a policyholder reaches retirement age, dies or returns to work, results in liabilities of a long-term nature often with experience other than originally assumed. It is thus crucial for insurance companies to manage the incidence as well as terminations of their long-term DI claims. From the literature reviewed, data analysed and latest industry DI claims experience, through formal and informal channels, this research comes at a crucial time, with DI claims experience, though improved over the last couple of decades, showing deterioration. Speculation is that this is due to various reasons, including a struggling economy, as well as recent DI tax changes resulting in previously take-home benefits becoming tax-free. Nevertheless, it is an optimal time to reconsider disability income claims and potential improvement in ongoing claims management. Group DI claims, incurred over 20 years, by one of the largest insurers in SA, were analysed, considering the case management information available to identify claimant intervention. Results found claim intervention to be significant in explaining increased probability of RTW. Six claim factors, in addition to intervention, were identified as significantly explanatory of RTW. Claim intervention also appeared to make a significant positive impact in specific pockets of incurred claims. This included increasingly improved RTW relative to no intervention, as claimants' age at disability increase; for claimants with higher replacement ratios; as well as for cancer and respiratory claims. In the absence of a controlled trial with randomly allocated intervention, improved data collection, in particular clinical information of intervention type and severity, will go a long way in answering how to intervene to improve RTW. Additional detail regarding claims, in particular disability severity, and disability definitions, will improve interpretation of RTW drivers. Ultimately, intervention experience over longer durations, will improve quantifying ultimate intervention impact on RTW. To reduce the impact of right-censoring on intervened claims, it is recommended that, in addition to increased data collection, a similar study be conducted once more run-off data is available for intervened claims.

The Application of Machine Learning to Predict the Behaviour of New Generation Risk Policyholders for Setting a Valuation Basis

By L Botha for Masters in Information Technology at the University of Pretoria, 2020

Life insurance policyholders of new generation risk policies display behaviours of lapsing and altering their contracts. Predicting this policyholder behaviour is not only a requirement when reserving for this business, but affects profitability, pricing, liquidity, solvency, risk management and capital. Lapse risk is considered the most significant underwriting risk to a life insurer. There is thus an incentive to improve on the predictions of these behaviours. Machine learning methods offer automated feature importance ranking to understand the top drivers of policyholder behaviour. Machine learning also shows improved predictive capability over traditional aggregate methods used to set policyholder behaviour assumptions and thus a valuation basis. Machine learning methods are benchmarked against the traditional approach, and include logistic regression, decision trees, random forests, neural networks and clustering. The 'black-box' algorithms are seen to outperform the simpler and more interpretable models, such as logistic regression and decision trees. Lapses are best modelled using a neural network, and negative alterations using a random forest algorithm. The features most important in predicting each behaviour include whether a client defaulted on a previous premium, and whether the client has annual cover growth on their policy, for lapses and alterations respectively.

Use of machine learning techniques to better utilise multiple underwriting factors in mortality pricing and reserving of life insurance risk products

By C van Zyl for Masters in Information Technology at the University of Pretoria, 2020

Actuaries need to graduate or adjust standard tables for the expected experience of the insured population and then make further adjustments for the various selected underwriting factors when pricing and calculating liabilities (reserving) for life insurance benefits. This is a time-consuming process that requires actuarial judgement at various steps. Automating portions of the process and introducing statistical tests to assess fit would allow more accurate pricing, inclusion of more rating factors and more regular review of pricing bases. This research explores the use of statistical and machine learning approaches to pricing the mortality cost of life insurance benefits (for use in pricing and reserving). In particular, the research explores how generalised linear models and tree-based machine learning models can be used to develop entity-specific mortality tables that account for a large number of underwriting factors. To allow for regular review and the application of actuarial judgement at various stages, the methodology allows for the following discrete steps: Derivation of a population-specific standard table by adjusting an industry standard table using approaches described by Gschlössl et al. (2011) and Tomas & Planchet (2014); Adjustment to the standard table for common underwriting factors. For this step both generalised linear models as well as tree-based models (Model-Based Recursive Partitioning and Gradient Boosting Machines) are used; Adjustment to the standard table using additional underwriting factors only present in a smaller subset of the data using similar models. The research also aims to assess whether tree-based models can improve on simpler generalised linear models. The

results of the research illustrate the relative ease and effectiveness of the approach to derive a standard table and/or test the current tables used for an insured population for their fit, which is useful when assessing the need to make assumption or basis changes as part of the actuarial control cycle. The results further demonstrate the effective allowance for more traditional underwriting factors such as income, smoking status and education levels for all models tested. Results using more underwriting factors were less intuitive or suitable for use in practice despite apparent better performance of the models using criteria such as mean square error. This highlights the need for actuarial judgement and explains the lower use of such rating factors in the industry other than for initial underwriting. Options for including such factors in pricing models is discussed. In all cases, the tree-based machine learning models (despite extensive parameter tuning) did not meaningfully outperform generalised linear models suggesting that generalised linear models are better suited given their easier application and interpretation.

Pricing and hedging variance swaps using stochastic volatility models
By K Bopoto for MSc Financial Mathematics at the University of Pretoria, 2020

In this dissertation, the price of variance swaps under stochastic volatility models based on the work done by Barndorff–Nielsen and Shepard (2001) and Heston (1993) is discussed. The choice of these models is as a result of properties they possess which position them as an improvement to the traditional Black–Scholes (1973) model. Furthermore, the popularity of these models in literature makes them particularly attractive. A lot of work has been done in the area of pricing variance swaps since their inception in the late 1990s. The growth in the number of variance contracts written came as a result of investors' increasing need to be hedged against exposure to future variance fluctuations. The task at the core of this dissertation is to derive closed or semi-closed form expressions of the fair price of variance swaps under the two stochastic models. Although various researchers have shown that stochastic models produce close to market results, it is more desirable to obtain the fair price of variance derivatives using models under which no assumptions about the dynamics of the underlying asset are made. This is the work of a useful analytical formula derived by Demeterfi, Derman, Kamal and Zou (1999) in which the price of variance swaps is hedged through a finite portfolio of European call and put options of different strike prices. This scheme is practically explored in an example. Lastly, conclusions on pricing using each of the methodologies are given.

A study of the relationship between economic and technical aspects of Bitcoin
By J Kirsten for MSc (Financial Engineering) at the University of Pretoria, 2020

This study investigates the cryptocurrency called Bitcoin. A cryptocurrency is a type of currency that depends on cryptography to issue new units instead of depending on government decree like fiat currencies. The study will first explain some of the technical details that make bitcoin work. This is necessary to lay groundwork to get to the actual aim of the study, namely investigating the economic aspects of bitcoin. The study will evaluate Bitcoin, and other cryptocurrencies, along with fiat currencies against certain definitions. In the process

it will introduce a new subclass of cryptocurrency—the sovereign cryptocurrency. Bitcoin's implied monetary policy will also be discussed, as well as the problems it creates for central banks. A hypothesis on the behaviour of the bitcoin price will be explained and research will be provided to support the acceptance of the hypothesis. Using this hypothesis, a stochastic pricing model for bitcoin will be derived. Arbitrage trading strategies will also be provided that explain certain price constraints that operate in the bitcoin market. The dissertation will also introduce a means to improve the anonymity of a user of bitcoin and will reason that improvements such as these and others will increase the use of bitcoin. Therefore, improvements to anonymity will increase the economic relevance of bitcoin and increase its competitive edge over the traditional banking system.

Stochastic modelling of temperature derivatives in South Africa: A pricing framework
By T Mhazo for MSc (Financial Mathematics) at the University of Pretoria, 2020

The continuous-time auto-regressive $CAR(p)$) model was fitted to daily average temperature (DAT) data. A mean-reversion Lévy temperature dynamics model that captures normal and extreme variations in temperature was developed to represent the stochastic dynamics of temperature. The property of mean-reversion was included in the model because daily temperatures will never deviate from the mean temperature for a long period. Our DAT data had extreme points which are caused by abrupt changes in temperature, thus the stochastic Lévy process was chosen to fit these anomalies.

A study of 40 years' historical DAT data for Bloemfontein, Cape Town, East London, Kimberley, Mafikeng and Polokwane showed that the deseasonalised DAT data possesses the mean-reversion property. Implementing the Shapiro–Wilk test and histograms showed that the standardised residuals of the deseasonalised DAT data was not normally distributed. The histograms for the standardised residuals showed that data was skewed and had semi-heavy tails. To capture skewness and semi-heavy tails in the residuals the generalised hyperbolic distribution was employed.

The parameters of the $CAR(p)$ model were identified from the estimated parameters of the $AR(p)$ model. The $CAR(p)$ model driven by Lévy process was implemented for the derivation of temperature futures prices. The theory of arbitrage-free dynamics of temperature futures prices was adopted through the application of the Esscher transform. The Esscher transform introduced a flexible class of risk-neutral measures that are suitable for pricing temperature futures under the theory of arbitrage. Explicit pricing dynamics for CDD futures under the risk-neutral probability, were derived.

We also introduced temperature derivatives to the agricultural sector in South Africa. Since maize is the biggest grain production, an analysis on the relationship between temperature and maize yield was done. The results from this analysis showed that maize yield is negatively correlated to temperature, meaning that an increase in temperature results in a decrease in yield. It is therefore viable to introduce temperature derivatives to the agricultural sector in South Africa to help farmers hedge against yield risk. However, temperature derivatives will not only benefit farmers, they could also be sold through banks, micro-financing institutions, and/or cooperatives. Banks and financial markets can protect their portfolios against default

risk caused by event risk by purchasing temperature derivatives. Protecting borrowers and lenders against adverse weather events could be instrumental in improving credit markets in South Africa and other developing countries. The implementation of temperature derivatives in the country is not only limited to the agricultural industry, temperature derivatives can also be expanded to energy industries, theme parks and/or event organisers.

In conclusion, we found that the $CAR(p)$ model is sufficient to model temperature evolution and it is acquiescent for pricing futures under the theory of arbitrage. To develop temperature derivatives historical records must be adequate and available.

Time series clustering of financial returns

By CL Bezuidenhoudt for MSc (Statistics) at the University of Pretoria, 2020

The interconnectedness of the world's economies and stock markets has resulted in spillover effects that are far reaching. Determining financial returns with similar volatility structures provides an avenue to inform on possible spillover effects among markets. However, the vast amount of data available often makes exploring the different relationships more challenging. Time series clustering is a solution to this problem by providing homogenous groups that have maximum similarity. This mini-dissertation explores time series clustering of financial returns, with specific focus on utilising the generalised autoregressive conditional heteroscedasticity (GARCH) model and its various extensions to produce optimal clusters. The usefulness of financial time series clustering is explored using data of the Euro exchange rate of developed countries. Additionally, the usefulness of time series clustering to determine different risk profiles of stocks on the Johannesburg Stock Exchange is included.

Shrinkage estimation in ARMA-GARCH regression models with an application in Bitcoin returns

By Z Sibanda for MSc (Statistics) at the University of Pretoria, 2020

In this document, we focus on the extensions of autoregressive conditional heteroscedastic (ARCH) models and the generalised autoregressive conditional heteroscedastic (GARCH) models applied to financial data. Volatility is observed in financial time series as a response to information or news, which in most cases is unknown beforehand. Although, in certain situations, the timing of information provided may not be a surprise (e.g. announcements of mergers or initial public offerings (IPOs), etc.), giving rise to some aspects of volatility being predictable. Even though volatility is a latent measure in that it is not directly observable but given ample information, it can be estimated. With the uncertainty of risk on financial assets, it would be an inadequate assumption that a constant variance exists over a given time period which is assumed when using ordinary least squares estimation. In the past, linear regression models were used to predict relationships between macro-economic variables but when heteroscedasticity is present, one might still obtain unbiased regression parameter estimates with too low standard errors, which will influence the true sense of precision. The ARMA-GARCH regression model is one of many extensions of the GARCH process with respect to the conditional mean. This dynamic model allows for both the conditional mean and conditional variance to be modelled by the ARMA process and the GARCH process

respectively. More specifically, in this mini-dissertation, we develop shrinkage estimation techniques for the parameter vector of the linear regression model with ARMA-GARCH errors. For the purpose of shrinkage estimation, we will be assuming that some linear restrictions hold on the regression parameter space. From a practical point of view, specifying a set of logical restrictions plays an important role in economic and financial modelling. We conducted an extensive Monte Carlo simulation study to assess the relative performance of the proposed estimation techniques compared to the existing likelihood-based estimators. The application of our research is considered in the estimation and modelling of Bitcoin returns and testing the significance of the interest in the topic of cryptocurrencies as well as the impact of which traditional financial markets may have on Bitcoin and the cryptocurrency market.