

Abstracts of recent postgraduate theses and dissertations at South African universities

Using DEA to profile in-hospital surgeon services: a South African funder perspective

By Matan Abraham for MPhil (Actuarial Science) at the University of Cape Town, 2014

The comparative assessment of physician performance, also known as ‘profiling’ is frequently used by healthcare funders. It aims to identify and improve the resource efficiency and quality of physician care. South African private healthcare funders use a wide range of profiling techniques; however, currently the use of frontier analysis is absent. This study explores the use of the nonparametric frontier analysis technique called Data Envelopment Analysis (DEA) for the profiling of physicians in South Africa. This is investigated by following a DEA profiling approach to evaluate the performance of 403 general/paediatric surgeons in providing in-hospital services in 2012. A 7-input 1-output VRS DEA model is used to determine the efficiency of the surgeons. The profiling results are then analysed to determine their usefulness. Its results reveal that 58 surgeons are efficient, representing only 14.4% of surgeons profiled. Therefore, the DEA approach reveals a large potential for efficiency improvements. The average efficiency score of inefficient surgeons is found to be 0.68. This means that, on average, inefficient surgeons have to decrease resource utilisation by 32% to achieve efficiency. The DEA approach is also found to be proficient at identifying the physicians presenting the most severe levels of inefficiency. 37 surgeons are found to be significantly inefficient. The approach also allows for the identification of peers against which inefficient surgeons are able to directly compare their practices. These results are determined to be of significant potential use to South African private healthcare funders. It is, however, noted that the analysis and results obtained was solely of a statistical nature. Closer consideration of the clinical appropriateness of the results is essential. In any case, this study concludes that a DEA profiling approach can be considered a useful technique in the comparison of physician performance in South Africa.

The effect of security return dispersion on performance measurement in a South African context

by Bryce Gething for MPhil (Mathematical Finance) at the University of Cape Town, 2014

This work replicates a similar study performed by de Silva et al. (2001). Our study was performed on the South African market. De Silva et al. (2001) studied the effect of cross-sectional volatility (CSV) on fund managerial skill measurement. This led to the conjecture that increased fund performance dispersion was primarily due to higher CSV, and not changes in informational efficiency or ranges in managerial talent. In this dissertation we firstly critique the CSV-adjusted alpha as a measure of fund performance and show that it can only be used as a means of normalising fund performance, yet reveals very little with regard to managerial talent. Since fund performance is intrinsically linked to CSV, we find it difficult to disentangle the effects of CSV and managerial talent dispersion. Adjusting for CSV therefore also implies adjustment for managerial talent, and we conclude with ideas for how a CSV-adjusted alpha may be used to assess manager talent.

A comparative analysis of non-linear techniques in South African stock-selection

by Nikhil Hutharam for MPhil (Mathematical Finance) at the University of Cape Town, 2015

Forecasting stock performance has long been one of the primary objectives of financial practitioners. Literature has shown that the classical linear approach to modelling the interactions among company-specific factors and its stock market returns in time have become less suited for capturing the movements of the stock market. Hence, attempts to predict the performance of a stock have become associated with additional layers of complexity. This has led to the adoption of non-linear approaches to forecast stock performance. This dissertation explores the performance of some non-linear models in the South African market. These were classification and regression trees (CART), logistic regression and a random forest approach compared against a linear regression model. Moreover, a hybrid model between CART and logistic regression was considered. The models fell into two categories (i.e., static and dynamic models). Using a set of classification and portfolio performance metrics it was found that a dynamic modelling approach outperformed a static approach. Overall, the logistic and linear regression models dominated in terms of performance against the tree-based models and hybrid approaches. The results also demonstrated that a hybrid approach offered an improvement over a stand-alone CART.

Hedging techniques for guaranteed annuity options

by Kathryn Janssen for MPhil (Mathematical Finance) at the University of Cape Town, 2015

This dissertation investigates the hedging of a guaranteed annuity option (GAO) using three different hedging methods: PV01 hedging, forward rate hedging and hedging using principal component analysis (PCA). The GAO is priced using martingale pricing theory and by implementing the Two Additive-Factor Gaussian (G2++) Model interest-rate model, under the assumption that survival probabilities evolve deterministically over time. Monte

Carlo simulation is then used to price the GAO. To compare the performances of the hedge portfolios, static one-day and one-week profit-and-loss distributions are generated using market data and model simulated data. The mean, standard deviation, 1% and 5% Value at Risk (VaR) and Expected Tail Loss (ETL) are used as risk metrics to compare the hedge performances. In all of the cases, the PCA hedge portfolio outperforms the PV01 and forward rate hedge portfolios. Stress tests are also conducted to test the robustness of the portfolios; the PCA portfolio is determined to be the most robust of the three portfolios. Lastly, a dynamic hedge is simulated using market data, assuming weekly re-balancing over a one-year period. The PCA portfolio proves to be the optimal portfolio for hedging a guaranteed annuity option in comparison to PV01 and forward rate hedging portfolios.

Infant and under-five mortality in South Africa: perspectives from the 2011 census and the 2012 HSRC survey

By Boboh Kamangira for MPhil (Demography) at the University of Cape Town, 2014

This research focuses on estimating infant and under-five mortality in South Africa for the period 1998–2012, both to update previous estimates taking into account new data and to assess the reasonableness of all estimates. Data from the 2011 Census and the 2012 HSRC survey were used for this purpose. The 2011 Census provided data from deaths reported by households as well as the survival of the most recent births. The 2012 HSRC survey provided full birth history data for women aged 15–49 which were used for direct estimation of childhood mortality. Deaths reported by households together with census estimates of the number of children under the age of five are used to produce estimates of infant and under-five mortality using the synthetic cohort life table approach. Blacker and Brass's previous birth technique is used to provide an estimate of infant mortality based on the survival of the most recent birth in the 24 months preceding the census after correcting for the bias in the proportion dead among most recent births relative to the proportion dead among all births. The under-five mortality rate corresponding to this infant mortality rate is estimated using the ratio of under-five mortality to infant mortality as observed in the Princeton West level 19 model life tables. The direct method for estimating childhood mortality is applied to the 2012 HSRC full birth history data. After imputing the exact dates of birth and death, locating deaths in time and calculating the exposure to risk of dying, the method of deriving period life tables is then followed to estimate $q(1)$ and $q(5)$. Results show that the correction for the bias in the proportion dead among the most recent births relative to the proportion dead among all births in the Previous Birth Technique estimates did not work as estimates were still too low relative to those produced by other researchers. Also, the direct estimates from the 2012 HSRC survey were lower than those produced by other researchers owing to too few deaths recorded in this survey and a large proportion of these missing ages at death. Only the estimates from deaths reported by households were found to be useful. Thus, it was concluded that the data from the 2012 HSRC survey and the survival of most recent births from the 2011 Census do not produce reliable estimates of childhood mortality. It is recommended that a DHS-type survey should be conducted with one of its purposes being

to investigate the issues around childhood mortality estimation in South Africa, particularly to improve on the quality of data available for estimation of childhood mortality and to investigate further why the current methods are failing to produce reasonable estimates of childhood mortality. Further research to investigate the extent of the bias in the proportion of children dead among most recent births relative to the proportion dead among all births is also needed.

Analysis of fertility estimates in Zimbabwe: a comparison of the census and DHS data

By Zvikomborero TR Madari for MPhil (Demography) at the University of Cape Town, 2014

Analysis of census data is important to uncover new insights as well as highlight where improvements in future data collection are required. The study provides an assessment of the fertility estimates derived from census data in comparison to those derived from the Zimbabwe Demographic and Health Surveys. Robust methods are used to estimate fertility levels and to identify the trends in fertility in Zimbabwe. Fertility decline in Zimbabwe is observed to have started in the early 1980s. The greatest level of decline occurred between the 1980s and the mid-1990s. In more recent years fertility in Zimbabwe has stalled at roughly four children per woman. Using projected parity progression ratios fertility decline has been observed to be in part a result of parity limitation, as fewer women progress to higher parities. A comparison of the census and Zimbabwe Demographic and Health Survey fertility measures show that, for the same cohort of women, the measures of fertility are strongly congruent. While there are problems with census data, it has been shown that using robust estimation the census fertility estimates are comparable to those from the Demographic and Health Surveys.

Reinsurance and dividend management

By Humphrey Marufu for MPhil (Mathematical Finance) at the University of Cape Town, 2014

In this dissertation we set out to find the dual optimal policy of a dividend payout scheme for shareholders with a risk-averse utility function and the retention level of received premiums for an insurance company with the option of reinsurance. We set the problem as a stochastic control problem. We then solve the resulting second-order partial differential equation known as the Hamilton–Jacobi–Bellman equation. We find out that the optimal retention level is linear with the current reserve up to a point whereupon it is optimal for the insurance company to retain all business. As for the optimal dividend payout scheme, we find out that it is optimal for the company not to declare dividends and we make further explorations of this result.

Statistical arbitrage in South African equity markets

by Kuthadzo Masindi for MPhil (Mathematical Finance) at the University of Cape Town, 2014

The dissertation implements a model-driven statistical arbitrage strategy that uses the principal components from Principal Component Analysis as factors in a multi-factor stock model, to isolate the idiosyncratic component of returns, which is then modelled as an

Ornstein-Uhlenbeck process. The idiosyncratic process (referred to as the residual process) is estimated in discrete-time by an auto-regressive process with one lag (or AR(1) process). Trading signals are generated based on the level of the residual process. This strategy is then evaluated over historical data for the South African equity market from 2001 to 2013 through backtesting. In addition, the strategy is evaluated over data generated from Monte Carlo simulations as well as bootstrapped historical data. The results show that the strategy was able to significantly out-perform cash for most of the periods under consideration. The performance of the strategy over data that was generated from Monte Carlo simulations demonstrated that the strategy is not suitable for markets that are asymptotically efficient.

Mean-variance hedging in an illiquid market

By Melusi Mavuso for MPhil (Mathematical Finance) at the University of Cape Town, 2015

Consider a market consisting of two correlated assets: one liquidly traded asset and one illiquid asset that can only be traded at time 0. For a European derivative written on the illiquid asset, we find a hedging strategy consisting of a constant (time 0) holding in the illiquid asset and dynamic trading strategies in the liquid asset and a riskless bank account that minimises the expected square replication error at maturity. This mean-variance optimal strategy is first found when the liquidly traded asset is a local martingale under the real world probability measure through an application of the Kunita–Watanabe projection onto the space of attainable claims. The result is then extended to the case where the liquidly traded asset is a continuous square integrable semi-martingale, and we again use the Kunita–Watanabe decomposition, now under the variance optimal martingale measure, to find the mean-variance optimal strategy in feedback form. In an example, we consider the case where the two assets are driven by correlated Brownian motions and the derivative is a call option on the illiquid asset. We use this example to compare the terminal hedging profit and loss of the optimal strategy to a corresponding strategy that does not use the static hedge in the illiquid asset and conclude that the use of the static hedge reduces the expected square replication error significantly (by up to 90% in some cases). We also give closed form expressions for the expected square replication error in terms of integrals of well-known special functions.

Using vital registration data to track mortality in Zimbabwe's metropolitan populations: 2000–2012

By Elton Mukondo for MPhil (Demography) at the University of Cape Town, 2015

The vital registration system in Zimbabwe is incomplete and mortality estimates produced from these data might not give a true representation of mortality in the population. However, it may be assumed that vital registration data for urban areas is more complete than for the country as a whole. This research was, therefore, conducted in an attempt to answer the question of whether vital registration data can be used to track the mortality of Zimbabwe's metropolitan populations. To answer this question, direct and indirect estimates from census and Demographic and Health Survey (DHS) data were used to decide on the viability of using

these vital registration data to estimate mortality. Estimates of under-five mortality between 2001 and 2011 from vital registration data ranged from around 50 to 80 deaths per thousand for Harare while Bulawayo's estimates were generally between 55 and 105 deaths per thousand in the same period. Bulawayo's vital registration data appeared to produce reasonable estimates of under-five mortality, while Harare's vital registration data underestimated both infant and under-five mortality when compared to the other supporting estimates from the alternative data sources. For adult mortality, ${}_{45}q_{15}$, completeness of vital registration was estimated at 102 and 99 percent for Harare males and females respectively, and 95 percent for Bulawayo using the Death Distribution methods. The level of adult mortality from vital registration data in the middle of the period between the 2002 and 2012 census was 48.4 percent for Harare males, 46.2 percent for Harare females and 55.2 percent for Bulawayo (males and females combined). Comparison of the annualised rates and the supporting estimates revealed that Harare's vital registration data produced reasonable estimates of male adult mortality for the population, while the level of female adult mortality was slightly underestimated. For Bulawayo, adult (male and females combined) mortality estimates were found to slightly overestimate the mortality of the population when compared to the estimates from the other data sources. The trend suggested by vital registration estimates, however, was similar to that suggested by the other data sources and methods, implying that vital registration data can be used, at the very least, to follow the trend of infant, under-five and adult mortality in urban areas.

Assessment of the robustness of recent births in estimating infant mortality using multi-country Demographic Health Survey data

By Malvern Munjoma for MPhil (Demography) at the University of Cape Town, 2014

This dissertation investigates the robustness of recent births in estimating infant mortality rates from the proportion of deaths observed among births reported in a 24-month period. The Blacker–Brass technique is applied to all births reported in the 24-month period and to most recent births in the 24-month period. The study uses birth history data from 76 Demographic and Health Surveys conducted in 16 countries across the developing world between 1986 and 2011. All births (and the deaths of those births) occurring in five 2-year periods before each survey were extracted to obtain five estimates of infant mortality using the Blacker–Brass and direct estimation methods from each dataset. This allows trends in infant mortality for the 10-year period before the survey to be compared and relative errors to be calculated. The results showed a decline in infant mortality in most datasets and are consistent with the United Nations and the World Health Organization 2013 estimates. The relative errors did not indicate any systematic bias of the Blacker–Brass method applied to all births; however, further investigations showed that the method underestimated infant mortality in the period closest to the survey date in most datasets. Furthermore, the relative errors were positively correlated with the directly estimated level of infant mortality. There were, however, no significant differences in the relative errors across countries. Blacker–Brass estimates were also derived using last births reported in the 24-month period of each DHS dataset. Unlike the estimates which were derived from all births in the 24-month periods most of these

estimates were below the approximate 95 percent confidence intervals of the direct estimates. Therefore, the Blacker–Brass estimates, which used recent births only had more bias. DHS data on last births only, in each of the 24-month period, were used to calculate infant mortality using the Blacker–Brass method as in census datasets. This evidence suggests that applying the Blacker–Brass method to census data underestimates infant mortality. We suggest a modification of census questions on recent births to include all births in the past 24 months and vital status of these children on census date to get more robust infant mortality rates using census data. We further propose a simple improvement to the method using all births in the 24-month period that reduces the error by almost 35% relative to the direct method estimates when compared to the original Blacker–Brass method. The revised method is validated using DHS data from four countries that were not part of the initial data used.

Estimating dynamic affine term structure models

By Zachry Pitsillis for MPhil (Mathematical Finance) at the University of Cape Town, 2015

Duffee and Stanton (2012) demonstrated some pointed problems in estimating affine term structure models when the price of risk is dynamic, that is, risk factor dependent. The risk-neutral parameters are estimated with precision, while the price of risk parameters are not. For the Gaussian models they investigated, these problems are replicated and are shown to stem from a lack of curvature in the log-likelihood function. This geometric issue for identifying the maximum of an essentially horizontal log-likelihood has statistical meaning. The Fisher information for the price of risk parameters is multiple orders of magnitude smaller than that of the risk-neutral parameters. Prompted by the recent results of Christoffersen et al. (2014) a remedy to the lack of curvature is attempted. An unscented Kalman filter is used to estimate models where the observations are portfolios of FRAs, Swaps and Zero Coupon Bond Options. While the unscented Kalman filter performs admirably in identifying the unobserved risk-factor processes, there is little improvement in the Fisher information.

Stochastic time-changed Lévy processes with their implementation

by Odwa Sihlobo for MPhil (Mathematical Finance) at the University of Cape Town, 2014

We focus on the implementation details for Lévy processes and their extension to stochastic volatility models for pricing European vanilla options and exotic options. We calibrated five models to European options on the S&P500 and used the calibrated models to price a cliquet option using Monte Carlo simulation. We provide the algorithms required to value the options when using Lévy processes. We found that these models were able to closely reproduce the market option prices for many strikes and maturities. We also found that the models we studied produced different prices for the cliquet option even though all the models produced the same prices for vanilla options. This highlighted a feature of model uncertainty when valuing a cliquet option. Further research is required to develop tools to understand and manage this model uncertainty. We make a recommendation on how to proceed with this research by studying the cliquet option's sensitivity to the model parameters.

An empirical study between customer satisfaction, customer loyalty and profitability within the short-term insurance sector

by Lucy de Canha for MSc (Mathematical Statistics) at the University of Free State, 2014

Customer satisfaction has been a popular topic in many research studies and finding the relationship between customer satisfaction and profits or company revenue has been important as it helps stakeholders understand the importance of satisfying customers and how this reflects financially. This study looks at the relationship between customer satisfaction, loyalty and profitability within the short-term insurance industry in South Africa. A survey was sent out to all clients from five insurance companies in South Africa. This was sent via email once the insured had made an amendment to his/her policy. Using the satisfaction attributes within the survey a regression analysis was used and the coefficients were applied to the data to develop a one-dimensional customer satisfaction rating for each respondent. Two main hypotheses were looked at to prove the relationship between customer satisfaction and profitability. The first was a direct approach performing regression and correlation analysis between the overall satisfaction rating and the profitability variable which was developed by the insurance company's statistical team. The second hypothesis looked at the relationship between customer satisfaction and loyalty (tenure and number of products). Additional analysis was completed by looking at the first hypothesis across different demographics which included age, employment status of the insured, insured's gender, area of product purchased, length of the relationship between the insured and the company and whether the insured used a broker or direct services. The results were interesting and the main outcome was that there was no relationship between customer satisfaction and profitability or customer satisfaction and loyalty. The results may at first seem surprising; however, the short-term insurance industry is different to other industries when it comes to satisfying their customers. The design of short-term insurance products is such that it should not relate to satisfaction, but rather to retention. A customer in the short-term insurance industry is neither satisfied nor dissatisfied until an event that leads to a claim occurs. At this point a number of factors will lead to the client becoming satisfied or dissatisfied, with one factor being whether the claim is paid out or not. Should the insured be paid out the entire claim he/she will be more satisfied than if the claim was not paid out at all. Paying this claim will therefore make the client less profitable even though the insured may be more satisfied with the company. Further implications and recommendations are also discussed.

Investigating the threshold of event detection with application to operational risk theory

by CJ de Witt for MSc (Actuarial Science) at the University of Pretoria, 2014

This paper provides systematic analysis of points of structural change in probability distributions. In observed frequency data of earthquakes, such a threshold exists due to the non-detection of events below a certain magnitude. By examining the factors influencing the operational risk exposure of institutions, a similar threshold is hypothesised to exist in operational loss data. This threshold is termed the threshold of completeness, above which

100% of events are detected. External factors may cause this level of completeness to shift over time. The level of complete recording influences the volume of data that can be consistently incorporated in a study of operational risk losses. Such analysis can be used by financial institutions that need to model operational risk for regulating purposes, such as the Basel and Solvency regimes. Historically, a variety of methods have been proposed by authors in an attempt to gauge the location of the threshold of completeness in the case of earthquake data. This study aims to evaluate the efficacy of some of the most prominent methods under differing assumptions regarding the incomplete portion of operational risk data. Furthermore, a new threshold estimation scheme (MITC) is developed and tested against the prevailing methods.

The effect of observation errors on the assessment of insurance losses as a result of seismic activity

by S Pretorius for MSc (Actuarial Science) at the University of Pretoria, 2014

This article analyses the effect of errors in earthquake magnitude determination on the estimation of key seismic hazard parameters. In addition, the implications for the insurance industry of estimators that include observational error are discussed. The probabilistic seismic-risk assessment methodology and an application to the South African short-term property insurance industry underpin the above investigation. Probabilities of property losses attributable to seismic activity are estimated. The sensitivity of the estimated losses to changes in the underlying earthquake parameters is tested. The assumption that the estimated earthquake magnitude is a sum of two random variables, namely the actual (true) value of the earthquake magnitude and the observation error, is considered. Most often, the error is assumed to follow a normal distribution; however, when significant error outliers are present, such as in historic earthquake catalogues, the Laplace distribution may offer better approximations. The two error models of magnitude determination, i.e. the assumptions that errors follow normal and Laplace distributions, are applied to estimate the b-value of the frequency-magnitude Gutenberg–Richter relation that describes the distribution of different earthquake magnitudes, alongside methods of estimation that do not take errors into account.

Artificial neural networks and their application to modelling South African market returns

by Matthew Lee Smith for MSc (Actuarial Science) at the University of Pretoria, 2014

The modelling technique known as Artificial Neural Networks (ANNs) is investigated. ANNs have the ability to detect and project non-linear relationships between variables. Further, they can adapt in dynamically changing environments while providing accurate results. A method of constructing ANNs in order to form a forecasting system is presented here. In addition, detailed investigations into parameter estimation for ANNs are performed. ANNs and traditional models (ARIMA, seasonal smoothing, geometric Brownian motion, etc.) are constructed to forecast monthly inflation and the average monthly return on the

money, bond and equity markets in South Africa from 1975 to 2010. The ANNs constructed are done through an integrated and isolated approach. The performance of the traditional and ANN models are compared. No general conclusion, as to which model is superior for all the applications considered, can be made. This suggests that ANNs perform as well as traditional models when forecasting financial markets. Further, it is found that the money market and inflation are forecast efficiently through all the models over a single month. As the forecast period extends to three months the money market favours the traditional model. However, a forecast period of twelve months leads to the preference of ANNs in the case of the money market. Neither technique can forecast the equity or bond market accurately, as these require additional explanatory variables to those considered. As the forecast period increased, the forecast accuracy decreased for all the models. The integrated ANNs, which allow interaction between the markets, do not lead to improved forecasts which indicates that the relationships between the markets have a limited effect on the future values of the markets. Hybrid models are constructed, trained and tested for the money market and inflation. They are found to add value to traditional models when forecasting inflation, but not the money market. The sensitivity of the performance of ANNs and the traditional model to different subsets of the inflation data is tested. No statistical difference between the models is found. The implementation advantages of ANNs are also described.

Conditional large deviations for losses on portfolios with heterogeneity between constituent groups

by Nadine Mari Walters for MSc (Actuarial Science) at the University of Pretoria, 2015

The research provides an estimate of the distribution of losses between homogeneous classes in a large financial portfolio, given that a large total loss was experienced. For example, in the case of a corporate bond portfolio, individual constituents could be grouped into classes according to their credit rating. In order to calculate the amount of capital required for supporting risky portfolios, financial institutions often use scenario analysis, particularly stress testing, as part of the risk management procedure. When simulating stressed scenarios it is useful to have this additional information in order to allocate appropriate proportions of the total loss between the constituents of the portfolio. Institutions can therefore obtain additional information to aid in capital allocation and the setting of exposure limits to the different classes. The approach taken to derive the probabilities makes use of large deviations theory with techniques from measure- and probability theory. The assumption of independent and identically distributed random variables is relaxed in that constituents in different portfolio groups are not assumed to be identically distributed. Furthermore, the results can easily be applied to the case where constituents are dependent on some external factor, thereby relaxing the independence assumption as well. The asymptotic joint conditional distribution of the individual portfolio constituents given a specified large loss was found to take the form of a Gibbs measure. This approximation performed well even for small or almost completely heterogeneous portfolios. Furthermore, we provide an illustration of how to use these results in order to allocate large losses among portfolio groups.

Mortality models: comparison and application in old-age populations of selected countries

By Brian Jin-Wei Hu for MSc (Actuarial Science) at the University of the Witwatersrand, 2014

This research examined which of the five well-known chosen extrapolative mortality models best captured the trends in old-age population mortality for different age groupings in four different countries. Mortality rates from the Human Mortality Database for the United Kingdom, Poland, Japan and Taiwan were used, encompassing males and females in the 65–89 age group. This allowed assessments to be made across developed and emerging economies, and across Europe and Asia. Comparisons were made across models to understand why some work better for some age groupings in some countries. The research considered the goodness-of-fit of these well-known mortality models to historical population mortality rates, assessed the range of projected future mortality rates, and evaluated the financial impact of mortality uncertainty on annuity prices across the subject populations. Some of the findings which emerged were that the Booth–Maindonald–Smith model tended to work best for most of the selected populations, particularly for female or Asian populations. Perhaps surprisingly, retiring females in the emerging economies can be expected to possibly outlive males in the developed economies selected. In a low-yield environment, uncertainty around mortality has a noticeable impact on the range of pricing of annuities. The extent of mortality uncertainty is expected to be less for developed than in emerging economies, and less for females than males.

Modelling internal migration in South Africa

By Xolani Jozi for MSc by Coursework and Research Report at the University of the Witwatersrand, 2015

The aim of this study was to model internal migration in South Africa using the 2011 Census data. The net-internal migration was modelled in the district municipalities of South Africa using Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR). In this study, the following global and local modelling techniques were used: Gravity, Poisson, Negative Binomial (NB), Gamma, and GWR model (local model). Poisson and NB failed to fit the migration data, while the Gamma model managed to fit the data reasonably well. The GWR model performed better than OLS regression in modelling net-internal migration in district municipalities of South Africa. The results from these models revealed that there was a strong relationship between internal migration and economic variables, as well as living conditions and demographic variables. The Monte Carlo significance test results showed that the parameters of the white population vary significantly across space. The results of the study signal that the differences in social and economic disparities in the district municipalities of South Africa are the drivers of internal migration.

Single equation models for inflation forecasting in Rwanda

By Pascal Kayisire for MSc by Dissertation at the University of the Witwatersrand, 2014

This study evaluates Phillips Curve forecasts of inflation for Rwanda. The study relies on the use of various single equation prototype Phillips curve models, as described by Stock and Watson (2008). Pseudo out-of-sample comparison tests are used to evaluate the forecast performance of these Phillips curve forecasts relative to the AR (auto-regression) benchmark forecasts. In this regard, tests of equal forecast accuracy based on mean square forecast error and those based on forecast encompassing as used by several scholars (for example, Clark and McCracken (2001, 2005), Rapnach and Weber (2004)) are reported. Furthermore, the results from forecasts using inflation in levels and in differences as the dependent variable are reported, to check the sensitivity to this specification issue. The study finds that the Phillips curve and augmented Phillips curve benchmarks outperform the AR benchmark forecasts at one- and two-quarter horizons. The output gap, exchange rate and money supply (M3) are found to be good predictors of inflation in Rwanda in the generalised Phillips curve context. It is therefore strongly recommended that Rwandan economic policymakers take into consideration these variables when forecasting inflation.

Survival analysis of bank loans and credit risk prognosis

By Mercy Marimo for MSc by Dissertation at the University of the Witwatersrand, 2015

Standard survival analysis methods model lifetime data where cohorts are tracked from the point of origin, until the occurrence of an event. If more than one event occurs, a special model is chosen to handle competing risks. Moreover, if the events are defined such that most subjects are not susceptible to the event(s) of interest, standard survival methods may not be appropriate. This project is an application of survival analysis in a consumer credit context. The data used in this study were obtained from a major South African financial institution covering a five-year observation period from April 2009 to March 2014. The aim of the project was to follow up on cohorts from the point where vehicle finance loans originated to either default or early settlement events and compare survival and logistic modelling methodologies. As evidenced by the empirical Kaplan–Meier survival curve, the data typically had long-term survivors with heavy censoring as at March 2014. Cause-specific Cox regression models were fitted and an adjustment was made for each model, to accommodate a proportion p of long-term survivors. The corresponding Cumulative Incidence Curves were calculated per model, to determine probabilities at a fixed horizon of 48 months. Given the complexity of the consumer credit lifetime data at hand, we investigated how logistic regression methods would compare. Logistic regression models were fitted per event type. The models were assessed for goodness of fit. Their ability to differentiate risk was determined using the model Gini Statistics. Model assessment results were satisfactory. Methodologies were compared for each event type using Receiver Operating Characteristic curves and area under the curves. The results show that survival methods perform better than logistic regression methods when modelling lifetime data in the presence of competing risks and long-term survivors.

Modelling for the optimal product to offer a financial services customer

By John Shingirai Mukomberanwa for MSc by Coursework and Research Report at the University of the Witwatersrand, 2014

This study illustrates how various statistical classification models can be compared and utilised to resolve cross-selling problems encountered in a financial services environment. Various statistical classification algorithms were deployed to model for the appropriate product to sell to a financial services customer under a multi-classifier setting. Four models were used, namely: multinomial logistic regression, multinomial bagging with logistic regression, multinomial random forests with decision trees, and error correcting output coding. The models were compared in terms of predictive accuracy, generalisation, interpretability, ability to handle rare instances and ease of use. A weighted score for each model was obtained based on the evaluation criteria stated above and an overall model ranking thereof. In terms of the data, banked customers who only had a transactional account at the start of the observation period were used for the modelling process. Varying samples of the customers were obtained from different time points with the preceding six to twelve months information being used to derive the predictor variables and the following six months used to monitor product take-up. Error correcting output coding performed the best in terms of predictive accuracy but did not perform as well on other metrics. Overall, multinomial bagging with logistic regression proved to be the best model. All the models struggled with modelling for the rare classes. Weighted classification was deployed to improve the rare-class prediction accuracy. Classification accuracy showed significant limitation under the multi-classifier setting as it tended to be biased towards the majority class. The measure of area under the receiver operating characteristic curve (AUC) as proposed by Hand and Till (2001) proved to be a powerful metric for model evaluation.