

Book of Abstracts

Mathematics, Statistics and Operational Research Project B

This collection of abstracts represents student work over the course of two academic years. It highlights the wide variety of project topics that are run under the banner of MSOR (b) projects.

The motivation behind the research topics chosen by students doing this project module is varied. It can arise from personal interests, from placement experience, or research topics can be driven by collaborations with partners both inside and outside of UWE.

Some projects presented here using real data collaborating with partners have associated reporting restrictions. The others, either based on using simulations or publically available data, or based on theoretical work, have no such restrictions.

Each year has a separate table of contents. The advisors names have been included beside each project title to help inform potential students about the type of topics that staff members have previously advised upon.

We want to congratulate those graduates whose research is included in this collection and thank them for the time and effort they have given in undertaking their projects. I am honoured to share in our graduates' success and delighted to have this opportunity to open up their work to a wider audience. We wish all our graduates every success in their future.

Deirdre Toher

MSOR (b) Project Module Leader

2015-2016

2015 – 2016 Mathematics, Statistics and Operational Research Project B

Using Mathematical Programming for a Shift Allocation Problem in the Hospitality Industry
by Tom Arnall: advisor Alistair Clark

Use of Mathematical Programming to Model Traffic Flows on a Roundabout by Grace
Astbury: advisor Alistair Clark

Projectile Motion of a Sphere with Applications by Craig Davison: advisor Karen Henderson

Multivariate Analysis of Risk Factors for Graft Survival in UK Kidney Transplants with a
Focus on the Effect of Cold Ischaemia Time by Bethany Fisher: advisor Paul White

Can the NFL Combine Predict a Successful First Round Draft Pick in the National Football
League? By Benjamin Monk: advisor Deirdre Toher

Can the Success of Aviva Premiership Rugby Union Teams be Determined by Match Related
Statistics? By George Pawlett: advisor Robert Kelland

The Rate of Change of New HIV Infections in Africa by Susanna Skinner: advisor Deirdre
Toher

Identification and Exploitation of Inefficiencies in the Fixed Odds Betting Market by Sam
Spencer: advisor Paul White

The Effects of Socio-Economic Factors on Weight of Population on a Global Scale 2010 by
Maksim Zakarija: advisor Ana Sendova-Franks

Using Mathematical Programming for a Shift Allocation Problem in the Hospitality Industry

Tom Arnall advised by Alistair Clark

This ultimate aim of this project was to create a product for the hospitality industry that will optimize the allocation of staff to shifts. Having worked in the industry for over 3 years, one of the obvious problems was the old fashioned ways that staff rotas are written. Initially I interviewed an industry expert, to find out more about the current process and to find the requirements of a product of this nature.

An AMPL model was then developed to produce an optimal rota for a dummy restaurant, taking into account the various objectives, constraints and conditions identified by the expert. Once this development process was completed, we went back to the industry expert to test the product.

From this we evaluated the product, noted any particular strengths and limitations, and estimated a time saving that the product could achieve in the workplace. At the end of process, the product produced has the potential to be a commercially viable shift allocation tool for use in the industry, and is also capable of being adapted for other industries.

Use of Mathematical Programming to Model Traffic Flows on a Roundabout

Grace Astbury advised by Alistair Clark

The primary purpose of this project is to use A Mathematical Programme Language, AMPL, to model traffic flow on the Filton Abbey Wood roundabout, in Bristol. The objective function of the report is to minimise the amount of cars queuing at the entries to the roundabout and other key locations, which are represented by nodes in the model.

This has been achieved by firstly starting with a basic model (Model 1) which has then been improved upon to correctly represent the roundabout under investigation. This was followed by a rigorous evaluation of the limitations and the assumptions by using different modelling techniques, in order to develop better models. The main limitations are the mass amount of computer memory needed to solve Model 3 and having no means of simulating the randomisation of car allocation. The project concludes with an evaluation that Model 4 is the best representation to model traffic around this roundabout, and the software AMPL has a good ability to solve problems for such traffic modelling tasks.

Projectile Motion of a Sphere with Applications

Craig Davison advised by Karen Henderson

The primary aim of this report is to introduce projectile motion, which is an area of Applied Mathematics. This will be done by firstly explaining the basic equation and model. After which expanding to form a complex three-dimensional model which represents the trajectory of a projectile with external forces acting upon it such as drag, lift and spin. The solutions will be a mixture of analytical and numerical approximations as the more detailed models do not have an analytical solution. Maple will be used to approximate the solutions to the ordinary differential equations (ODE's) by using the numerical method of Runge-Kutta.

An application will then be demonstrated to show how projectile motion is used in real world scenarios. The application which will be shown in this report will be a corner kick of a football game. This uses the external forces of drag, lift and spin and it will allow for more complex computations in Maple, including adding a stopping criteria for when the ball reaches an optimal position for goal scoring. In order to achieve this optimal position, spin and lift coefficients must be in the range of 0.24-0.27 with a drag coefficient in the range of 0.25-0.28. However, these results vary depending on other conditions such as the launch angle and initial velocity, thus there are many ways in which a successful corner kick may be taken.

Multivariate Analysis of Risk Factors for Graft Survival in UK Kidney Transplants with a Focus on the Effect of Cold Ischaemia Time

Bethany Fisher advised by Paul White

Between 2006 and 2010 over 4,000 kidney transplants took place in the UK under the NHS. This project uses data collected by NHS Blood and Transplant in order to determine which factors regarding the transplant, donor or recipient, have a significant impact on the survival time of the kidney after implantation. A kidney has “survived” if it is still active at the end of the study.

Kaplan Meier was used as a form of exploratory data analysis to determine which categorical variables produce significantly different survival times. After which a Cox Proportional Hazards model was determined using backward elimination, and then tested to ensure it satisfied the necessary assumptions. The variables which proved to be significant were the donor age, recipient age, donor human leukocyte antigen (HLA) group, donor ethnicity and the cold ischaemia time of the graft. Both age variables were entered as continuous variables, and HLA group, ethnicity and cold ischaemia time were categorical.

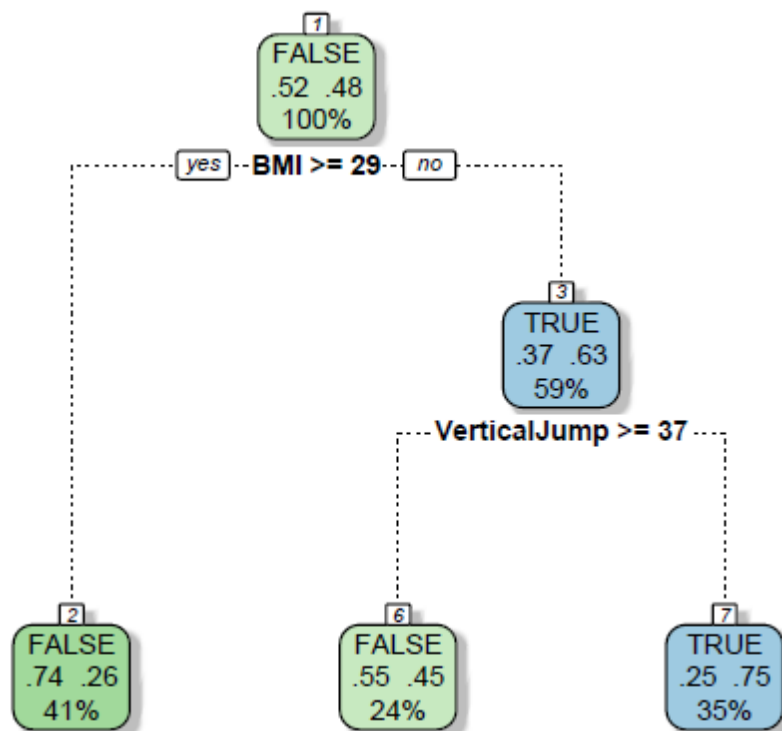
Cold ischaemia time was entered initially into the model as a continuous variable, however it did not produce significant results, so it was broken into categories using the “optimal” cut-points approach. This grouped the variable into four categories; ≤ 360 , 361-705, 706-1170 and ≥ 1171 minutes. The results of the Cox model showed that the survival times differed most significantly between the first two groups. The estimated survival time of a graft after a cold ischaemia time of between 361-705 minutes was approximately 360 days higher than those with a time of under 360 minutes. The relationship between this variable and survival time was found to be non-linear, however modelling using Regression Splines and Fractional Polynomials did not produce significant results in this case. Therefore cold ischaemia time was not as significant as expected with regards to this dataset, but still had an influence on graft survival time if grouped in this manner.

Can the NFL Combine Predict a Successful First Round Draft Pick in the National Football League?

Benjamin Monk advised by Deirdre Toher

The reverse order National Football League (NFL) draft enables the worst teams in the NFL, from the previous season, the chance to have a first pick of the best players coming out of College.

Each team will have intricate drafting methods that are used to predict the success of players in the NFL. These methods are based on College and Combine performances. This study focuses on the National Football League Combine and the first round selections of the NFL Draft between 2006 and 2015. This have been carried out by using Classification and Regression Trees Analysis and Binary Logistic Regression. An example of one such CART model is illustrated below. These results showed that for each of the four position groups, it was not possible to use the NFL Combine alone to predict a successful first round draft pick in the National Football League.



Classification Tree for Catchers and Runners

Can the Success of Aviva Premiership Rugby Union Teams be Determined by Match Related Statistics?

George Pawlett advised by Robert Kelland

The primary aim of this study is to identify if the success of Aviva Premiership rugby teams can be determined using match related statistics. The data for the study focuses on 3 teams of different standard within the Aviva premiership over the most recent four seasons (2011-2015) providing an overall sample size of 264 games. The dependant variable throughout the investigation is whether the game is won/drawn or lost. The list of match related statistics was originally formulated with the data obtained via the official Aviva Premiership website.

Initially univariate analysis in the form of the Mann-Whitney U test was performed. This provided key information regarding which variables are significantly different in matches that are won and matches that are lost. The focal point of the univariate analysis was to act as a variable reduction method, such that only those variables that were identified as statistically significant can be utilised in the multivariate analysis.

There was also concern the variables classified as point scoring variables may potentially overpower any multivariate analysis and mask the effects of other more interesting variables. The variables that were included in the multivariate analysis are; Meters made, Clean Breaks, Turnovers conceded, Penalties conceded, Scrums won and Kicks-in-play. Multivariate analysis was performed using binary logistic regression. Two separate models were produced. The first model consisted of the significant variables as previously mentioned. The second logistic regression model consisted of the component scores of each match as calculated using principal component analysis (PCA). The PCA was performed on all variables, excluding the point scoring variables. It was established that good maintenance of possession followed by a clever kicking game and a strong scrummage are essential to increase the odds of winning a match. It was determined that a clinical attacking game backed up by good discipline and not too much ball in hand is vital to increase the odds of winning a match.

The findings of each model complemented each other in many ways and are both very worthy models considering how dynamic the sport is. Ultimately these findings can provide direction to elite rugby staff as to what elements are characteristic of a winning performance. How valid these findings are to other teams in the Aviva Premiership is questionable as is the length of time these findings are valid for.

The Rate of Change of New HIV Infections in Africa

Susanna Skinner advised by Deirdre Toher

After the Millennium Summit in 2000 a series of eight goals were developed to improve the quality of life for all. One of these goals was focused solely on combating HIV/AIDS and other diseases, thus highlighting just how big a problem HIV is globally. Globally there were approximately 36.9 million people living with HIV/AIDS at the end of 2014 with over 70% of these people living in sub-Saharan Africa. An estimated 2 million people were newly diagnosed with HIV in 2014 contributing to the total of approximately 78 million people who have been infected with the virus since it was first discovered. Since the beginning of the pandemic, approximately 39 million people have died from AIDS related illnesses with approximately 1.2 million of these occurring in 2014, with 66% of the 2014 deaths being accounted for by adults and children living in sub-Saharan Africa. (amFAR, 2015)

This report will discuss the need for a reduction in the number of new HIV infections and propose models which can be used to explain the number of new infections in subsequent years by using data previously collected. It will show how a variety of variables influence the number of new infections: from the number of people who die from AIDS related illnesses to the gender parity index in primary education; and from HIV prevalence among young men to the proportion of the population using improved drinking water sources in rural areas. The report will also show how whilst there has been a decrease in the number of new infections since the introduction of the Millennium Development Goals, there is still a long way to go to putting a halt to HIV/AIDS.

References:

amFAR (2015) *The Foundation for AIDS Research: Statistics Worldwide*. Available from: <http://www.amfar.org/worldwide-aids-stats/> [Accessed 07/04/216].

Identification and Exploitation of Inefficiencies in the Fixed Odds Betting Market

Sam Spencer advised by Paul White

In this report we use a number of statistical techniques to inform whether any inefficiency in the fixed odds betting market can be found. Furthermore we use these techniques in an attempt to find a profitable betting rule can be found. The focus of this report is on the English Premier League and the English Championship with data collected from the 2014/15 and the 2015/16 season.

Variables focussing on recent form, seasonal points per game and measures of performance such as possession, shots on target and shots conceded were collected to aid in this. A number of regression techniques were trialled i.e. binary logistic, nominal logistic and ordinal logistic. These regression techniques were used with the 2014/15 data to find models used to predict the outcome of the football matches and to see if a profit was obtained.

Overall it was found that the variables which appeared most in the models were the bookmakers betting odds, and an interaction of the shots on target by the home side and shots conceded by the away side. The probabilities generated from the models were compared to the probabilities found by the bookmakers published odds to obtain a ratio. This ratio was used to see if a simple betting rule could be utilised to maximise the profits found by using the model.

These betting strategies were then tested on the data collected from the 2015/16 season to see whether they worked on a new data set. The findings were that the betting strategies constructed obtained a small profit when betting on the home win.

The Effects of Socio-Economic Factors on Weight of Population on a Global Scale 2010

Maksim Zakarija advised by Ana Sendova-Franks

The purpose of this project is to investigate the effects socio-economic factors have on body mass index (BMI) and presence of Obesity in average residents from countries around the world. 155 Countries are being investigated, with the information from 2010.

Exploratory data analysis shows that the continent with largest mean BMI is Oceania, which coincidentally also has the highest levels of sugar consumption per capita. Two types of statistical models have been constructed and investigated, the Multiple Linear Regression and Binary Logistic Regression models. The former was investigating effects if previously chosen factors on mean BMI, while the latter was investigating the effects those factors had on presence of obesity.

The multiple linear regression model has proven to be a viable model with sugar consumption, urban population prevalence, unemployment prevalence, continent, and GDP being statistically significant factors in predicting the BMI.

The Binary Logistic Regression model was of poor fit with none of its predictors being statistically significant.

2014-2015

2014 - 2015 Mathematics, Statistics and Operational Research Project B

Numerical Solution of a Stochastic Differential Equation by Alexandra Wood: advisor Kevin Golden

Spatial Interactions of House Prices in Contiguous USA Metropolitan Statistical Areas by James Dawes: advisor Deirdre Toher

An Investigation into the Social and Economic factors affecting Smoking Prevalence in England by Samuel Riches: advisor Iain Weir

Modelling Household Expenditure in the UK by Kieran Rudge: advisor Deirdre Toher

Relationship between Police Identity and Organisational Citizenship Behaviour by Hannah Scott: advisor Paul White

Modelling the Flight Path of a Ball with an Investigation into the Magnus Effect by Harry Reynolds: advisor Karen Henderson

An Investigation of the Distributional Assumptions Made on Asset Prices by the Black-Scholes Model by Edward Cross: advisor Deirdre Toher

Analysis and Modelling of the Population Genetics of X-Linked Genes through the use of Discrete Dynamical Systems Theory by Jo-Anne Williams: advisor Robert Laister

The Discrete Fourier Transform with Accompanying Methods: An Investigation into Particular Applications by Sam Gregory: advisor Jan Van lent

A Study of Numerical Solutions of Fredholm Equations by Chelsi Bainbridge: advisor Karen Henderson

A New Approach for Analysing Proportions in Partially Overlapping Samples by Anselma Dobson-McKittrick: advisor Paul White

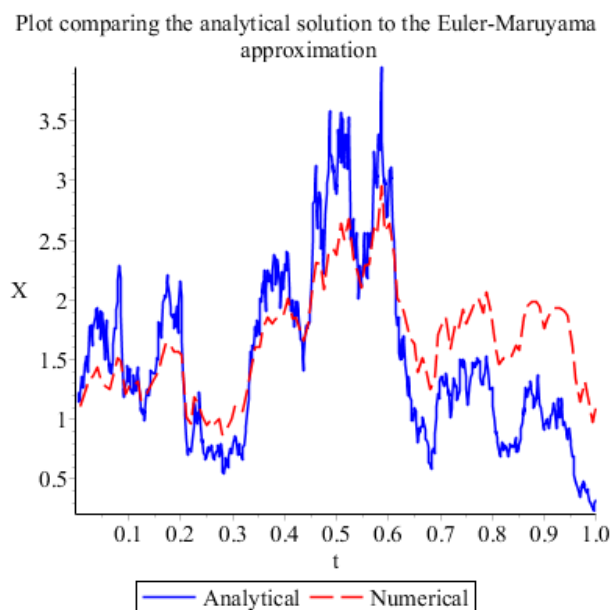
What is Data Compression, What are the Techniques Used and How Are These Applied in Practice? By Charles Day: advisor Catherine Hobbs

Numerical Solution of a Stochastic Differential Equation

Alexandra Wood advised by Kevin Golden

A stochastic differential can be solved analytically but in practice this is rarely possible and therefore an approximate solution can be found using a numerical scheme. To understand the accuracy of a numerical scheme the order of convergence can be calculated. Since the solution of a stochastic differential equation is a random process there are roughly two types of convergence: strong and weak.

This project is largely based on the work done by Sauer (2013) and Higham (2001). A stochastic differential equation will be solved analytically and its solution compared to the solutions obtained from the numerical schemes to find the order of convergence. It is found that the numerical schemes investigated have order of convergence that agree with those found in the previous studies done by Sauer (2013) and Higham (2001).

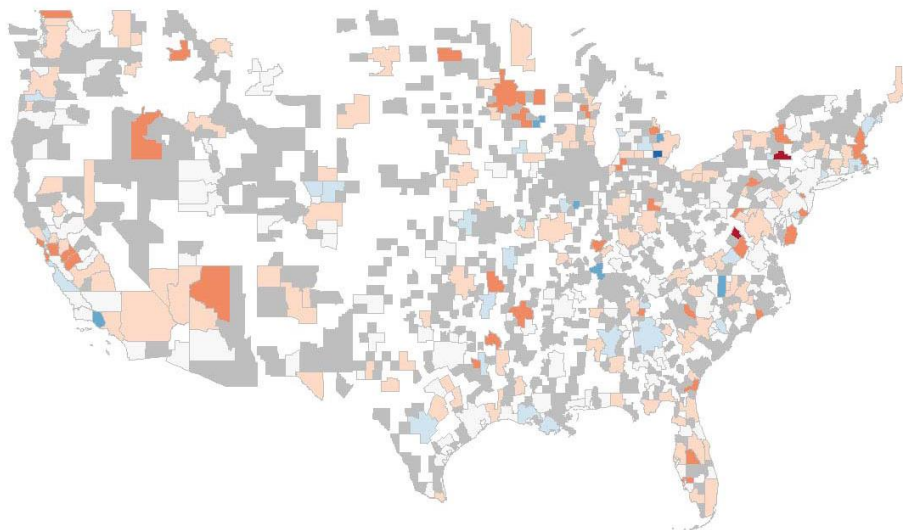


Spatial Interactions of House Prices in Contiguous USA Metropolitan Statistical Areas

James Dawes advised by Deirdre Toher

This project investigates whether there are any spatial effects of house prices in contiguous U.S.A. metropolitan statistical areas. The project uses house price index data from Freddie Mac for years 2009 to 2012. The research identifies that the crime data collected for analysis can be grouped into a single variable while retaining all the information using principal components analysis. The project also identifies macroeconomic, demographic and social characteristics which are significantly associated with house price indexes using regression analysis.

Finally a generalised additive model (GAM) and Moran's I were used to test and account for spatial auto-correlation. The formal test suggested that there was no significant global spatial relationship across the dependent variable or the final model's residuals. A GAM model proved to be unhelpful when modelling house price indexes as the smoothing term was non-significant. This indicates that, once crime, macroeconomic, demographic and social characteristics are taken into account, the remaining spatial relationship is negligible (as indicated by the lack of dark areas in the figure below).



An Investigation into the Social and Economic factors affecting Smoking Prevalence in England

Samuel Riches advised by Iain Weir

This investigation builds two models using regression techniques that will be able to be used to predict someone's likelihood of smoking and how much they smoke. The study should allow certain demographics to be targeted in order to reduce the prevalence of smoking in England. The motivation behind this project came when undertaking a work placement at the Office for National Statistics where a report was released that explained the link between different socio-economic groups and the amount of alcohol drunk and cigarettes smoked. Logistic regression and Generalised Linear Models were used to model the prevalence of smoking (in England). Throughout the literature review, results concluded that many things had a causal effect on someone's smoking habit: such as someone's socio-economic status, whether they had a relative who smokes or whether they drink more alcoholic units. The findings of the models produced supported this prior knowledge.

The overall results from logistic regression showed that as the age that someone left school increases, the odds of smoking consequently decrease. Also, for every £10,000 increase in income, the odds of smoking decrease by 6.6%. It was also shown that those whose mother smoked was 1.3 times more likely to smoke than if their mother did not smoke. Also if their father smoked they were 1.68 times more likely to smoke than if their father did not smoke.

The generalised linear model concluded that women smoke 17% less cigarettes than men. Being from a lower socio-economic background increases the number of cigarettes smoked. Someone that left school at 16 will smoke 51% more cigarettes than someone still in full-time education.

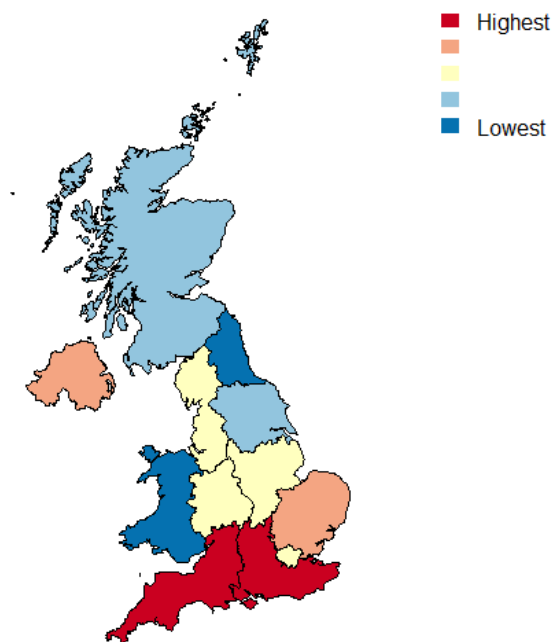
Modelling Household Expenditure in the UK

Kieran Rudge advised by Deirdre Toher

This dissertation project reviews the weekly expenditure of different households within the UK, using the Office for National Statistics 'Living Costs and Food Surveys' 2012 annual dataset.

This project describes how various attributes and characteristics of a particular household can be grouped together through the use of principal components analysis using parallel analysis to select and justify the number of components that should be retained. Furthermore, the resulting components are combined with other household variables and entered into various regression models, culminating in a prediction of average weekly expenditure on a household level. Transformations of the dependent variable were used to counteract the presence of skewness as is common with this type of data.

Both expenditure itself, and the residuals from the regression models are tested for spatial autocorrelation. From these, household expenditure is shown to vary depending on the regional location in the UK, as illustrated by the figure below (median expenditure).



Relationship between Police Identity and Organisational Citizenship Behaviour

Hannah Scott advised by Paul White

The Winsor review and government cuts to policing in England and Wales have led to lower morale within the police force. A dataset of approximately 13,500 police officers across 43 police forces in England and Wales was used to look at the relationship between Organisational Citizenship Behaviour (OCB) and Police Identity (PI). Factor analysis was conducted to measure the primary interests (PI and OCB) and a three-factor solution was found: Police Identity, Organisational Citizenship Behaviour towards Individuals (OCBI) and Organisational Citizenship Behaviour towards the Organisation (OCBO). Multiple regression analysis was then used to attempt to predict OCB within police officers, where PI is used as one of the predictor variables. Once a model was finalised, the same predictor variables were used to see whether they significantly predict OCBI and OCBO.

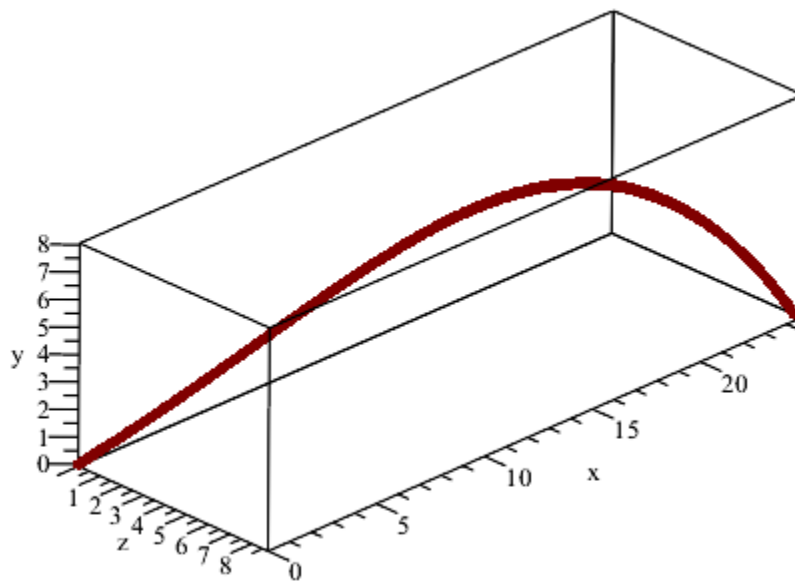
Findings show that Police Identity was the main predictor of OCB, OCBI and OCBO; police officers that show high levels of PI, tend to have higher levels of Organisational Citizenship behaviours. The nine predictors used for each model were: Policing is a vocation rather than a job, I am effectively on duty 24/7, I will not put myself at risk in case I get injured, Accountability is vital for the job of a police officer, The support of the public has a positive impact on my job, I am satisfied with my job, The goodwill of police officers is essential to the success of the police, I make sacrifices to be a police officer, and Police Identity.

Modelling the Flight Path of a Ball with an Investigation into the Magnus Effect

Harry Reynolds advised by Karen Henderson

This project aims to give an understanding of how mathematics can be used to model real world phenomenon, specifically the flight path of a spinning projectile. It first sets out an overall two dimensional model for predicting the flight path, and then gives a step by step construction of how this model is derived.

Where possible, analytical solutions have been found, otherwise the fourth order Runge-Kutta method has been used to numerically approximate solutions. Figures are produce throughout to visually display the results, an example of which is shown below. The model is then explored in three dimensions where an extra force is included.

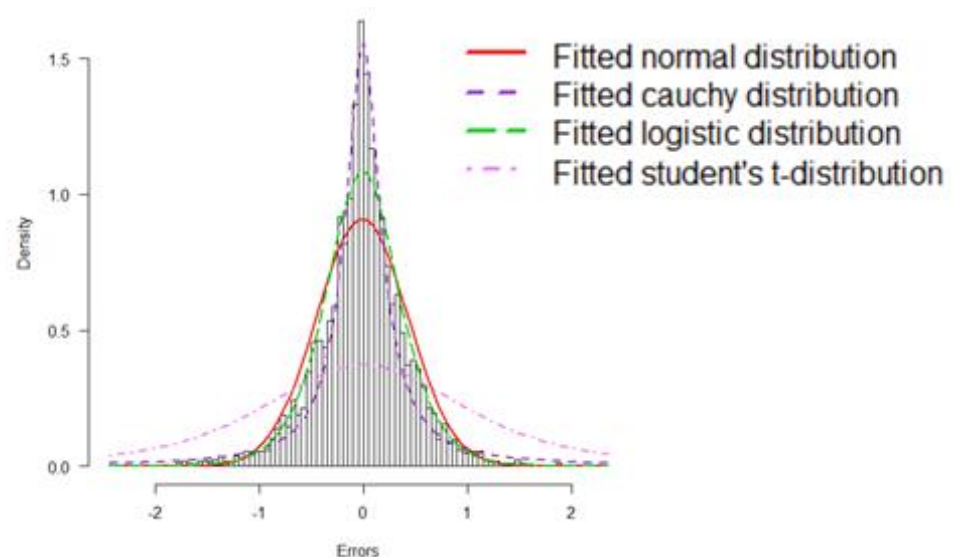


An Investigation of the Distributional Assumptions Made on Asset Prices by the Black-Scholes Model

Edward Cross advised by Deirdre Toher

In financial modelling, the Black-Scholes model introduced in 1973 is widely used today to help financial market participants' trade. The Black-Scholes model makes assumptions about market behaviour which do not accurately reflect true behaviour. The largest assumption is that the asset follows a geometric Brownian motion, implying that asset price returns follow a normal distribution.

This assumption, which forms the basis of this investigation, is largely criticised because large changes in the asset prices are not predicted by the model. An extremely prominent example is the market crash of September 2008 which, according to the model, was at odds of one in a billion. The normality assumption is challenged by Benoit B. Mandelbrot who proposes that the market prices follow a Cauchy distribution instead. In this study an alternative statistical approach is performed to investigate if there is a more appropriate distributional assumption for the price movements of assets. The results from this investigation generally support Mandelbrot's views.



Analysis and Modelling of the Population Genetics of X-Linked Genes through the use of Discrete Dynamical Systems Theory.

Jo-Anne Williams advised by Robert Laister

Population genetics involves the frequencies and distributions of the various possible forms of genes, known as alleles, within a genetic population. Dominance, mutation, migration and selection pressures, along with other variables, can affect the spread of alleles through a population.

This project analyses the behaviour of allele frequencies and distributions, of X-linked genes, which are only present on the X chromosome, and hence differ between males and females.

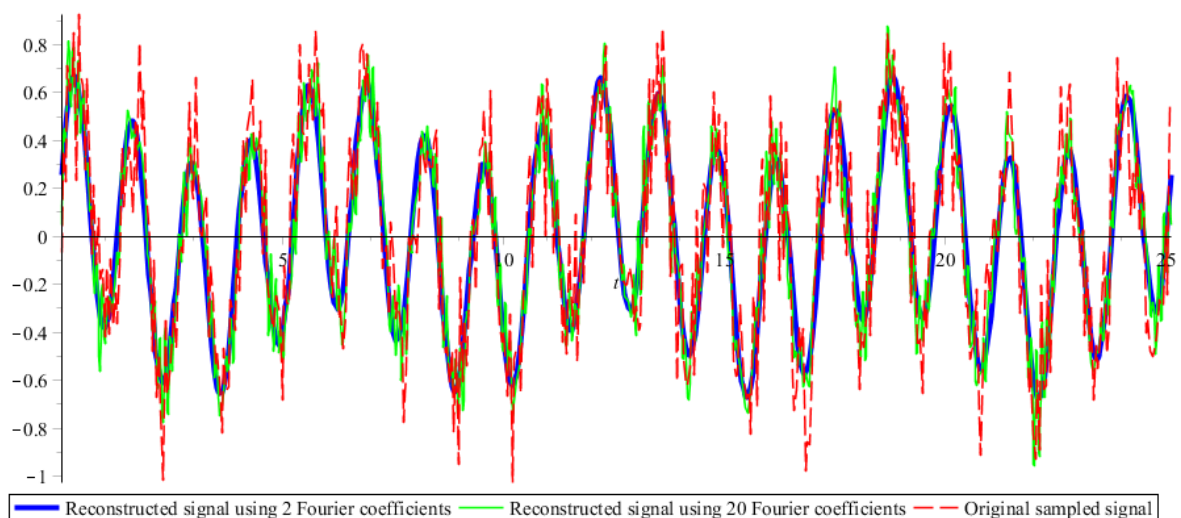
Basic theories of population genetics, such as Mendelian genetics, the Hardy-Weinberg law, Punnet Square and the Fisher-Haldane-Wright equation, will be used along with mathematical techniques within discrete dynamical systems - an area of mathematics that uses difference equations to model the behaviour of a system in discrete time steps.

The Discrete Fourier Transform with Accompanying Methods: An Investigation into Particular Applications

Sam Gregory advised by Jan Van lent

Discrete Fourier analysis is a broad branch of mathematics that likewise, has broad practical application. The essence of Fourier analysis is primarily concerned with ‘transformations’ that naturally, are represented by various transforms. This report will introduce and go some way to replicate these transforms in various applications. A basic example of signal reconstruction is shown below.

Methods under consideration include; the discrete Fourier transform in digital image processing, the discrete cosine transform in .JPEG compression and the discrete sine transform in solving difference equations.



A Study of Numerical Solutions of Fredholm Equations

Chelsi Bainbridge advised by Karen Henderson

In this report we will investigate different methods of solving integral equations, specifically Fredholm's equation of the second kind, beginning with simpler methods such as the Trapezoidal Rule and the Simpson's Rule, and then looking in to more complex enhanced methods such as Gaussian Quadrature.

We will begin by solving examples by hand and then go on to code the problems in Maple. We will look at the difference between the methods and investigate their accuracy.

A New Approach for Analysing Proportions in Partially Overlapping Samples

Anselma Dobson-McKittrick advised by Paul White

Within empirical research, comparing two groups on a binary variable which have partial overlapping samples is a common problem. The two most common situations are comparing two independent samples or two dependent samples. However there is no one recommended approach or method for comparing proportions between partially overlapping samples.

Consequently this project will propose four new test statistics for this situation, along with a review of previous methods and approaches. Within these new methods, the phi and the tetrachoric correlation coefficient will be used. Simulation techniques will be used to compare and contrast the behaviour of the four statistics. The simulation will be a designed factorial experiment with varying parameters of, proportions, correlation and sample sizes.

The main research questions of this project are: Do the four statistics work? Do the varying parameters affect the statistics? Is there a statistics that works the best? This analysis will all be carried out under the null hypothesis. The results of this analysis found that the four overlapping statistics work under the null hypothesis and there is very little effect on the performance of these statistics when there are small sample sizes, different amounts of overlap and levels of coefficients.

Furthermore, from the analysis, that both $Z1$ and $Z3$ work best out of the four test statistics with $Z3$ being superior for normality and robustness. Hence this indicates that when using \hat{p} in this case, it does not give the best results compared to when \hat{p}_1 and \hat{p}_2 are being used. Also, $Z1$ and $Z3$ have different types of correlation coefficients, phi and tetrachoric, yet both perform well. Therefore this backs up the agreement that whichever correlation coefficient is best to use is down to preference.

What is Data Compression, What are the Techniques Used and How Are These Applied in Practice?

Charles Day advised by Catherine Hobbs

The idea of this dissertation is to introduce the topic of data compression and explain it in simple terms to mathematicians who may not have come across the topic before. There are some compression techniques which have been analysed and compared, along with the basics of coding theory.

The final idea explored within this dissertation is to look at some of the different applications of compression techniques that are used in everyday life when it is not always obvious what techniques are being used.