HONOURS PROJECTS OFFERED IN 2020.

Topic: Sea Level Rise and Housing Prices
Supervisors: Dr Patricia Menendez (Clayton) and Dr Daniel Melser (Caulfield)
Email: patricia.menéndez@monash.edu and daniel.melser@monash.edu

The focus of the project is on identifying the extent to which prospective sea level rise (SLR), as a result of climate change is reflected in house prices. Various authors have examined the effect of specific climate events on house price sales such as Ortega Taşpinar (2018) that studied the impact of Hurricane Sandy on New York real estate prices. More recently, work by Bernstein et al (2019) looked at the effect of sea level rise in the US and found evidence of price reduction in properties that could potentially be affected by SLR. Even though part of the Australian coast line could possibly be affected by SLR, this type of study has not been done.

This research aims at exploring different methodological approaches to identifying the effects of sea level rise on house prices and the data include housing transaction prices for NSW since 2000 and records for SA. The data consist of not only dwelling prices but also house attributes. The goal here is to model house prices using those house characteristics together with some measure of exposure to SLR such as elevation and coast distance as explanatory variables. The main focus is to understand whether more exposure to SLR would have a detrimental impact on house prices.

The model implementation for this project will be carried out using the statistical software R and will involve managing spatial large data files (shape files and/or master files) as well as data collection and preparation. The elevation and distance to coastline data would need to be assembled and compiled. The project in general will require good econometrics skills and R advanced expertise to managing complex data wrangling, merging and advanced data manipulation. The project reporting will be set up using reproducible techniques including Rmarkdown files and github for version control.

This project would suit a student with strong modelling skills and advanced skills in R, Matlab and GIS.

References:

Ortega F. and S. Taşpinar. Rising sea levels and sinking property values: Hurricane Sandy and New York's housing market. Journal of Urban Economics, 106:81–100, 2018.

Bernstein, A., Gustafson, M. T. and R. Lewis. Disaster on the horizon: The price effect of sea level rise. Journal of Financial Economics, 134(2):253–272, 2019.

Topic: The impact of time effects modelling on multi-population mortality projection

Supervisor: Associate Professor Athanasios A. Pantelous
Email: Athanasios.Pantelous@monash.edu

Stochastic mortality projection plays an important role in quantifying longevity risk and its potential financial consequences for individuals, institutional investors (e.g., pension funds), social welfare institutions among others. Recently, a considerable amount of literature has been published around the theme of multi-population mortality forecasting. Further, the Augmented Common Factor (ACF) model proposed by Li and Lee (2005) has been increasingly attracting the attention of researchers in the multi-population mortality modelling community. However, much of the research up to now has almost ignored the dynamic specification on the hidden time effects (except for selecting the same time series models as Li and Lee did to inhibit divergence in mortality projection in the long run). Systematic testing of the parametric formulation on time effects is not available before performing Bayesian inference, which is a challenge for forecasters. Consequently, the main aim of this project is to investigate the impact of latent states modelling on multi-population mortality projection. Specifically, we will adopt the nonparametric Bayesian framework, and introduce a state-space model where the functional of the evolutional equation is assumed to be unknown. Therefore, this sort of nonparametric method will generalize the parametric state-space model, as well as release us from the restrictive assumption of the predetermined, but not optimal, functional form of the time effects. Bayesian estimation using Markov Chain Monte Carlo will be exploited to improve computing efficiency. The proposed model will be illustrated with real mortality data including countries possessing similar demographic characteristics, and the mortality experience of both genders in a country. Model comparison with the original ACF model as the benchmark will be conducted at the end.

References:

Lee, R.D. and Carter, L.W., 1992. Modelling and forecasting US mortality (with discussion). Journal of the American Statistical Association, 87(419), pp.659-675.

Li, N. and Lee, R., 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography, 42(3), pp. 575-594.

Fung, M.C., Peters, G.W. and Shevchenko, P.V., 2017. A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. Annals of Actuarial Science, 11(2), pp. 343-389.

Antonio, K., Bardoutsos, A. and Ouburg, W., 2015. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. European Actuarial Journal, 5(2), pp. 245-281.

Orbanz, P., 2014. Lecture Notes on Bayesian Nonparametrics.
http://www.gatsby.ucl.ac.uk/~porbanz/papers/porbanz_BNP_draft.pdf

Topic: Factor selection in asset pricing: an application to US stock markets
Supervisors: Dr Natalia Bailey (Clayton) and Dr David Frazier (Clayton)
Email: natalia.bailey@monash.edu and david.frazier@monash.edu

The capital asset pricing model (CAPM) and its multi-factor extension in the context of the Arbitrage Pricing Theory (APT) are the leading theoretical contributions implemented widely in modern empirical finance to analyse the cross-sectional differences in expected returns. Both approaches imply that expected returns are linear in asset betas with respect to fundamental economic aggregates, and the Fama-MacBeth two-pass procedure (Fama and MacBeth (1973)) is one of the most broadly used methodologies to assess these linear pricing relationships.

The first stage in the Fama-MacBeth procedure entails choosing the risk factors to be included in the asset pricing model. Given the upsurge in the number of factors deemed relevant to asset pricing in the past few years, a rapidly growing area of the finance literature has been concerned with evaluating the contribution of potential factors to these models. Predominantly, such methods consider the problem of factor selection using penalised regressions. In view of recent theoretical results, the importance of first evaluating the strength of these factors becomes evident, since factor selection is only meaningful for asset pricing if the set of factors under consideration are sufficiently strong.

This project focuses on determining the strength of prospective factors as a means of evaluating whether their risk can be priced correctly, and then proceeds to select the most appropriate out of these to be included in the asset pricing model. Such ``factor selection'' procedures can be implemented using existing machine learning methods such as elastic nets, random forests, or the Dantzig selector. Further, since the strength of these factors can change over time it is of interest to investigate the evolution of factor choice over the past 30 years.

Students with some programming experience in R are preferred. This topic would suit a student interested in (learning) advanced regression techniques and their applications to capital asset pricing models.

References:

Anatolyev, S. and A. Mikusheva (2019). Factor models with many assets: Strong factors, weak factors and the two-pass procedure. Available at arXiv: https://arxiv.org/abs/1807.04094

Bailey, N., G. Kapetanios, and M. H. Pesaran (2020). Measurement of factor strength: theory and practice. CESifo Working Paper.

Harvey, C.R. and Y. Liu (2019). A census of the factor zoo. Available at SSRN: https://ssrn.com/abstract=3341728 or http://dx.doi.org/10.2139/ssrn.3341728

Fama, E.F. and J. MacBeth (1973). Risk, returns and equilibrium: empirical tests. Journal of Political Economy 81(3), 607-636.

Feng, G., S. Giglio and D. Xiu (2020). Taming the factor zoo: a test of new factors. Journal of Finance, forthcoming.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1, no. 10. New York: Springer series in statistics, 2001.

Lettau, M. and M. Pelger (2018). Estimating latent asset pricing factors. NBER Working Paper No. 24618.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47 (1), 13-37.

Pesaran, M.H., and R. Smith (2019). The role of factor strength and pricing errors for estimation and inference in asset pricing models. CESifo Working Paper No. 7919.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341-360.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk.

Topic: To replicate or not to replicate? Recovery of latent variable models under different experimental schemes.

Supervisor: Dr Emi Tanaka (Clayton)
Email: emi.tanaka@monash.edu

Latent variable models (LVM), including the special case of factor analytic models when the responses are conditionally normally distributed, are gaining traction in various scientific fields from econometrics to agriculture. Model selection for LVMs has a striking, additional twist: as the name suggests, the latent variables in LVMs also have to be estimated from the data, as opposed to covariates in the standard regression settings which are directly observed. Model selection in LVMs thus amounts to order selection and variable estimation, where the former involves choosing the number of latent variables. Hui et al. (2018) presented one such method capable of performing both, using the ordered factor LASSO, however, little consideration was given for the structure of the data to predict the latent variables.

In this project, we will consider a multi-environmental field trials (MET) that plant breeders often use to select the optimal variety from predictions of a latent variable model. Particularly pertinent to the selection of optimal variety is the reliable predictions of latent variables. Plant breeders often wish to test more varieties in a MET at a cost to replication of varieties. This project will explore simulations on how different experimental schemes for MET performs for the recovery of latent variable models, and would suit a student with strong computational skills, particularly in R. Knowledge in EM algorithm is desirable but not necessary.

Reference: Hui, F.K.C., Tanaka, E. & Warton, D. (2018) Order selection and sparsity in latent variable models via the ordered factor LASSO. Biometrics. doi.org/10.1111/biom.12888

Topic: Optimal distributional forecasts of financial returns and volatility

Supervisors: Professor Gael Martin (Clayton) and Dr Ruben Loaiza Maya (Clayton)
Email: gael.martin@monash.edu and ruben.loaizamaya@monash.edu

Over the past decade, the use of scoring rules to measure the accuracy of distributional forecasts as become ubiquitous. In brief, a scoring rule rewards a probabilistic forecast for assigning a high density ordinate to the observed value, so-called `calibration', subject to some criterion of `sharpness', or some reward for accuracy in a part of the predictive support that is critical to the problem at hand. See Gneiting, Balabdaoui and Raftery (2007) and Gneiting and Raftery (2007) for early extensive reviews, and Diks, Panchenko and Van Dijk (2011) and Opschoor, van Dijk and van der Wel (2017) for examples of later developments. In the main, scoring rules have been used to compare the relative predictive accuracy of probabilistic forecasts produced by di_erent forecasting models and/or methods, with little or no attention given to the relationship between the manner in which the forecast is produced, and the way in which its accuracy is assessed. Exceptions to this include, Gneiting et al. (2005), Gneiting and Raftery (2007), Elliott and Timmermann (2008), Patton (2019) and Loaiza-Maya, Martin and Frazier (2020), and related work on the scoring of point forecasts in Gneiting (2011). In this work, focus is given to deriving forecasts that are, in some sense, optimal for the particular empirical problem and - as part of that - deliberately matched to (or made `consistent' with) the score used to evaluate out-of-sample performance; the idea here being that the forecast so chosen will, by construction, perform best out-of-sample according to the scoring rule that matters.

This project will pursue this idea in the context of producing accurate forecasts of extreme values of asset returns and return volatility. This goal will be achieved by producing predictive distributions for both random variables that are optimal according to a scoring rule that rewards accurate prediction of extreme values. The method will be applied to various financial series, including returns on the S&P500 stock index and the VIX index constructed from option prices. Success of the predictive approach will be gauged in terms of various important financial measures, as they relate to both the financial return itself and its volatility.

This project is based - in part - on ongoing research by the supervisors with David Frazier and other co-authors. It is designed for an Honours student who has completed ETC3460 and is enrolled in ETC4460 for Semester 2. Template Matlab programs will be provided; so the student needs to either be familiar with Matlab or willing to learn it. Of course, the programs can be translated into another language of the student's choice (e.g. R) if so desired.

References:

Diks, C., Panchenko, V., and Van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. Journal of Econometrics, 163(2):215-230.

Elliott, G. and Timmermann, A. (2008). Economic forecasting. Journal of Economic Literature, 46(1):3-56.

Kneiting, T. (2011). Making and evaluating point forecasts. Journal of the American Statistical Association, 106(494):746-762.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(2):243-268.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477):359-378.

Gneiting, T., Raftery, A. E.,Westveld III, A. H., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review, 133(5):1098- 1118.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile weighted scoring rules. Journal of Business & Economic Statistics, 29(3):411-422.

Loaiza Maya, R., Martin, G.M. and Frazier, D.F. (2020), Focused Bayesian Prediction, https://arxiv.org/abs/1912.12571.

Maneesoonthorn, W., Martin, G.M, Forbes, C.S. and Grose, S. (2012). Probabilistic Forecasts of Volatility and its Risk Premia. Journal of Econometrics, 171, 217-236.

Opschoor, A., Van Dijk, D., and van der Wel, M. (2017). Combining density forecasts using focused scoring rules. Journal of Applied Econometrics, 32(7):1298-1313.

Patton, A. J. (2019). Comparing possibly misspecifed forecasts. Journal of Business & Economic Statistics, (published on-line).

Topic: Detecting internet anomalies worldwide: A global anomaly detector

Supervisor: Dr Klaus Ackermann (Clayton)
Email: Klaus.ackermann@monash.edu

The Monash IP-Observatory (https://ip-observatory.org/) collects worldwide Internet data for social good. Uninterrupted access to the Internet is crucial for many aspects of the civil society or business operations. In times of crises, such as hurricanes, earthquakes or electricity outages, detecting which regions are affected and need urgent assistance, is an important piece of information. (e.g. latest earthquake in Puerto Rico) The goal of this project is to build a global anomaly detector with either a method from time-series analysis or machine learning. Interested students should have good data-wrangling and visualization skills.

References:

Talagala, P. D., Hyndman, R. J., Smith-Miles, K., Kandanaarachchi, S., & Muñoz, M. A. (2019). Anomaly detection in streaming nonstationary temporal data. Journal of Computational and Graphical Statistics, 1-21.

Wang, R., Nie, K., Wang, T., Yang, Y., & Long, B. (2020, January). Deep Learning for Anomaly Detection. In Proceedings of the 13th International Conference on Web Search and Data Mining (pp. 894-896)

Topic: Time2Vec embedding for time series
Supervisor: Dr Klaus Ackermann (Clayton)
Email: Klaus.ackermann@monash.edu

Embedding is a popular technique for neural networks for dimensionality reductions, such as text or images. The idea for this project is to create embedding for time-series and use this dimension reduction then for time series clustering, forecasting and transfer learning of time-series, with application to electricity data in Europe and Australia. Embedding's for time series have been explored by the literature, here the idea is to incorporate external predictor series, such as high frequency Internet data, at a different timing resolution and experiment if it improves the forecasting accuracy. A preferable student background would include exposure to machine learning, or completion of ETC3555.

References:

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

Yi, X., Zhang, J., Wang, Z., Li, T., & Zheng, Y. (2018, July). Deep distributed fusion network for air quality prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 965-973).

Topic: Forecasting with economic theory
Supervisor: Dr Didier Nibbering (Clayton)
Email: Didier.nibbering@monash.edu

Many macroeconomic forecasts do not satisfy conditions imposed by economic theory. For instance, inflation and output forecasts are rarely coherent with the Taylor rule, which establishes the relation between inflation, output and the interest rate set by the central bank. Robertson et al. (2005) show that reconciling inflation and output forecasts with the Taylor rule can improve forecast accuracy.

This project examines optimal reconciliation of inflation and output forecasts with the Taylor rule. The forecasts can be reconciled with the Taylor rule from the perspective of the central bank, in which the interest rate is known, or from the perspective of the forecasters, for which we have to set up a time series model to forecast the interest rate.

The project aims to extend the forecast reconciliation methods in Wickramasuriya et al. (2019) for hierarchical and grouped time series to the linear restriction imposed by the Taylor rule. The methods are then applied to forecasts of the Survey of Professional Forecasters. Possible extensions are to take the uncertainty into account in the; inflation and output forecasts, the coefficients in the Taylor rule, the reconciliation rule.

A good understanding of forecasting and coding is essential.

References:

https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/

Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. Journal of Money, Credit and Banking, 383-401.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association, 114(526), 804-819.

Topic: The impact of children on a household's housing expenditure
Supervisor: Dr Daniel Melser (Caulfield)
Email: Daniel.melser@monash.edu

The focus of this topic is on identifying the impact of having children on a household's housing expenditures. This is an interesting phenomenon to measure but is also useful in terms of better informing government about the appropriate level of the housing supplement for households that receive welfare. The basic approach would be to run a regression of the form:

$$Housing\ Expenditure = \boldsymbol{x}T\boldsymbol{\beta} + \alpha \cdot Number\ of\ Children + \epsilon$$

This would use household panel data such as that available in the Household, Income and Labour Dynamics in Australia (HILDA) dataset. There is also the possibility of using US data. The main estimation challenge is that the variable 'Number of Children' is likely to be endogenous. That is, there might be omitted factors which are related to both housing expenditure and the number of children that families have. This means that the results from OLS estimation will be biased.

A potential solution to this problem is to find an instrumental variable (IV) for number of children. One approach that could be explored is that of Angrist and Evans (1998). They argued that a good IV for whether a couple has a third child is whether the first two children are of the same sex. That is, families are more likely to have a third child if the first two were either both boys or both girls—i.e. families prefer a gender mix. This instrument could be used to derive a consistent estimate of the impact of children on housing expenditure.

The HILDA data is relatively easily available. A form needs to be completed, you wait a few days and you get the data. This data is quite large and complex, so it takes a while to get familiar with it, but it is well documented. The project involves using a software package such as Stata or R.

References

Joshua D. Angrist and William N. Evans (1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", The American Economic Review, Vol. 88, No. 3, pp. 450-477

Bradbury, B. (2008), "Time and the Cost of Children", Review of Income and Wealth 54: 305-323. https://doi.org/10.1111/j.1475-4991.2008.00277.

Topic: Modelling Housing Careers in Australia
Supervisor: Dr Daniel Melser (Caulfield)
Email: Daniel.melser@monash.edu

The focus of this project would be to better understand the way in which Australians transition across housing statuses during their lifetime and what the key factors are driving these changes. We can think of a person as having various dimensions to their housing status; whether they own or rent, whether they live in a house or apartment, whether they are living in the city or a rural/regional area and whether the home is large or small. The interaction of each of these possibilities can define a particular housing status. These statuses, and transitions between them, can be modelled and explored.

The best data set for this project is probably the Household, Income and Labour Dynamics in Australia (HILDA) Survey. This is an annual household panel data set on a sample of Australian households. Multinomial logit models (or equivalent) would be estimated on the transition probabilities between housing statuses. This would facilitate the construction of a micro-simulation model to look at the impact of the various factors. First, we could examine the extent to which aging drives housing transitions, such as elderly people downsizing or young couples with children moving to the suburbs. Second, we could explore the extent to which people from different ethnic/birth-country groups have different preferences for housing types and transitions. Third, it would enable us to look at the impact of affordability/price on housing transitions.

References

Kendig, H. L. (1984). "Housing Careers, Life Cycle and Residential Mobility: Implications for the Housing Market", Urban Studies 21(3), 271-283. https://doi.org/10.1080/00420988420080541

Clark, W. A. V., Deurloo, M. C., & Dieleman, F. M. (2003). "Housing Careers in the United States, 1968-93: Modelling the Sequencing of Housing States", Urban Studies 40(1), 143-160. https://doi.org/10.1080/00420980220080211

Supervisor: Dr Denni Tommasi (Caulfield)
Email: denni.tommasi@monash.edu
https://sites.google.com/site/dennitommasi/

Topic: The elephant in the room: Local average treatment effects with misclassification

In the estimation of causal effects, treatment is often endogenous and measured with error. The latter is also called misclassification error in case of discrete treatments. While the endogeneity problem is well understood, how to deal with misclassification in the recorded treatment status is

far less obvious and it is rarely discussed in empirical papers. This is a problem of primary importance because, even with infrequent arbitrary errors in the binary treatment indicator, empirical results can be severely biased (Millimet, 2011). Tommasi and Zhang (2020) develop a framework to recover the weighted average of local average treatment effects (LATE) of Imbens and Angrist (1994) in a context of endogenous and misclassified binary treatment. The main purpose of their work is to provide a simple tool that can be used by applied researchers in any setting where the endogenous binary treatment is not well measured and instrument(s) are available. They call the associated estimator P-LATE, for Partially Identified LATE. This is a very useful (yet, simple) identification result because there are many examples potential treatments of clear economic significance that are rarely analyzed causally because the treatments themselves are not observable.

In this project, you will apply the P-LATE framework to study an important topic in applied microeconomics: the effects of the Supplemental Nutrition Assistance Program (SNAP) on elderly health status and adult obesity rates. Specifically, you will have access to a (ready-to-analyse) dataset and codes to identify the effects of interest. The dataset contains information about SNAP participation, health status, and many other characteristics. You will also have access to other datasets to potentially analyse different applied questions of interest. In this project, you will learn how to identify and frame policy-relevant economic questions, formulate solid identification strategies, and analyse cross-sectional data in a partial identification setting. Please e-mail me if you want to discuss more about this.

The project involves theoretical skills and informatics/software skills. The ideal candidate should have a strong understanding of micro-econometric theory, particularly instrumental variables (IVs) and partial identification techniques. The candidate should be proficient with Matlab (and coding in general). The use/knowledge of a different software (e.g. STATA) is very much welcome

References

Imbens, G. W. and J. D. Angrist (1994): Identification and estimation of Local Average Treatment Effects," Econometrica, 62, 467–475.

Milliment, D. (2011): "The elephant in the corner: a cautionary tale about measurement error in treatment effects models," in Missing Data Methods: Cross-Sectional Methods and Applications. In: Advances in Econometrics, Emerald Group Publishing Limited, vol. 27, 1–39, 1 ed.

Tommasi, D. and L. Zhang (2020): "The elephant in the room": Local average treatment effects with misclassification". Working Paper.

Topic: A new recursive estimation method for a class of panel data models
Supervisor: Professor Jiti Gao (Caulfield)
Email: jiti.gao@monash.edu

This project proposes a new recursive estimation method for a class of panel data models. It is expected that the newly proposed method has advantages in both theory and practice. In theory, it is anticipated that the estimation method will outperform existing methods proposed in Bai (2009), and Pesaran (2006) as well as in Jiang, et al (2017). In practice, several empirical datasets will be used to demonstrate the practical relevance and superiority of the proposed method over existing methods.

References:

Bai, J. (2009), Panel data models with interactive fixed effects. Econometrica, 77(4): 1229–1279.

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2017), Recursive Estimation in Large Panel Data Models: Theory and Practice. Available at https://ideas.repec.org/p/msh/ebswps/2017-5.html

Pesaran, M. H. (2006), Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, 74(4):967–1012.

Topic: An integrated panel data modelling of an OECD health care expenditure data set
Supervisor: Professor Jiti Gao (Caulfield) & Dr Bin Peng
Email: jiti.gao@monash.edu

This project proposes an integrated panel data model to take into account possible cross-sectional dependence, heterogeneity and trending nonstationarity. Both estimation theory and computational algorithm will be developed for practical implementations. An empirical application of the proposed estimation method and the implementational procedure will be on modelling the relationship between health care expenditure per capita and a vector of explanatory variables, including GDP per capita, the population of over 65 and below 15 divided by the population between 15 to 65, physicians per 1000 persons and government financed ratio in health care expenditure. It is expected that the empirical results may lead to some policy implications.

References:

Baltagi, B. H. and Moscone, F., (2010), Health care expenditure and income in the OECD reconsidered: evidence from panel data. Economic Modelling 27, 804-811.

Chen, J., Gao, J. and Li, D. (2012), Semiparametric trending panel data models with cross-sectional dependence. Journal of Econometrics 171, 71-85.

Feng, G., Gao, J. and Peng, B. (2019), An Integrated Panel Data Approach to Modelling Economic Growth. Available at https://ideas.repec.org/p/arx/papers/1903.07948.html.

Topic: Econometric estimation of multivariate time series models with structural breaks
Supervisors: Professor Jiti Gao (Caulfield) and Dr Wei Wei (Caulfield)
Email: jiti.gao@monash.edu and weiwei2@monash.edu

This project considers a multivariate time series model with structural breaks in an additive form. There are several unknown quantities to be estimated, including the form of the conditional mean function, the locations of possible structural breaks and the levels of the breaks. We propose using a simultaneous estimation methods for all unknown quantities. The estimation method and its asymptotic theory will both be new. Applications of the proposed estimation method and computational procedure include in climatology, energy econometrics, finance and machine learning.

References:

Gong, X. and Gao, J. (2018), Nonparametric kernel estimation of the impact of tax policy on the demand for private health insurance in Australia. Australian & New Zealand Journal of Statistics 60, 374-393.

Gao, J., Gijbels, I. and Van Bellegem, S., (2008), Nonparametric simultaneous testing for structural breaks. Journal of Econometrics 143, 123–142.

Gao, J., Tong, H. and Wolff, R. (2002), Adaptive series estimation in additive stochastic regression models. Statistica Sinica 12, 409–428

Topic: Patterns in labour productivity in Timor-Leste
Supervisor: Prof Brett Inder (Clayton)
Email: brett.inder@monash.edu

It is widely recognised that improvements in labour productivity provide a valuable indicator of development of an economy. This project will work with a time series of 9 years of data from the Timor-Leste Business Activity Survey (2010-2018) to identify patterns and trends in Labour Productivity. Of particular interest is the variation in productivity across sectors, and productivity dispersion within industries, as well as identifying any trends showing improvements in productivity over time.

Reference:

Campbell, S. Nguyen T,. Sibelle, A., and F. Soriano: Measuring productivity dispersion in selected Australian industries. Treasury–ABS Working Paper, 2019-02

Topic: High Frequency trading profits around index rebalancing dates
Supervisor: Assoc. Professor Paul Lajbcygier (Clayton)
Email: paul.lajbcygier@monash.edu.

Passive investors rely on index investing. By investing in indexes, passive investors obtain market returns at low cost, which mean that in the long run they outperform most active managers. As a consequence, hundreds of billions of dollars is invested in indexing across the globe. The low cost of indexing is achieved not only because of the low management fees charged by indexers, but also because investing in indexes results in low portfolio. Although some index rebalancing is required in order to track the index this is small compared to the typical active investor whose turnover is 100% of their portfolio per annum, in contrast to the indexer which is 20%. Such low index turnover results in low trading and hence low trading costs. However, these costs can be reduced further: in order to reduce tracking error indexers must rebalance as the index composition changes, which for the ASX200 is once a quarter. Indexers effectively have no choice, they must rebalance at such times in order to eliminate (or at least reduce) index tracking error. This means that nimble predatory traders, such as high frequency traders, know which stock and on what days the index behemoths must trade and can exploit that information to 'front run' the indexers and make quick profits from the natural demand created in these stocks from indexers. Our aim is to study if HFTs trade and profit from indexers on index rebalancing dates and ascertain how much profits HFTs obtain at these times.

Topic: Sampling with alternative indexation
Supervisor: Assoc. Professor Paul Lajbcygier (Clayton)
Email: paul.lajbcygier@monash.edu.

An equal weight index outperforms a value weight index before costs, but not after. We introduce a simple, theoretically motivated, enhanced equal weighted index that outperforms a value weight index, even after transaction costs. It removes the smallest stocks in each industry from the index, thus removing much of the market impact costs, and as theory would suggest, provides a lower tracking error than simply removing the smallest stocks in the entire index. We test the approach by comparing equal and value weight indexes of various depths and show that such an enhanced equal index outperforms a value index. Thus, our contribution is to show that equal weight indexing is viable when stratified industry sampling is utilized.

Topic: Portfolio rebalancing and abnormal returns in the Russell 1000 and 2000 indices.
Supervisor: Assoc. Professor Paul Lajbcygier (Clayton)
Email: paul.lajbcygier@monash.edu.

There are numerous research studies that have been executed with regard to the stock price responses related to large capitalization stocks especially in the S&P 500 index. However, there are few studies that asses the stock returns of small capitalization indexes, like Russell 2000 index. The aim of this study is to: (1) examine the source of abnormal returns during index rebalancing activities; (2) assess the stock returns of the organizations that are added to or deleted from the Russell 1000 and Russell 2000 index; (3) examine the price pressure effect for both the Russell 1000 index and the Russell 2000 index; and, (4) investigate which factor is more significant in explaining the abnormal returns.

Topic: Computational and visual methods for exploring high dimensional functions

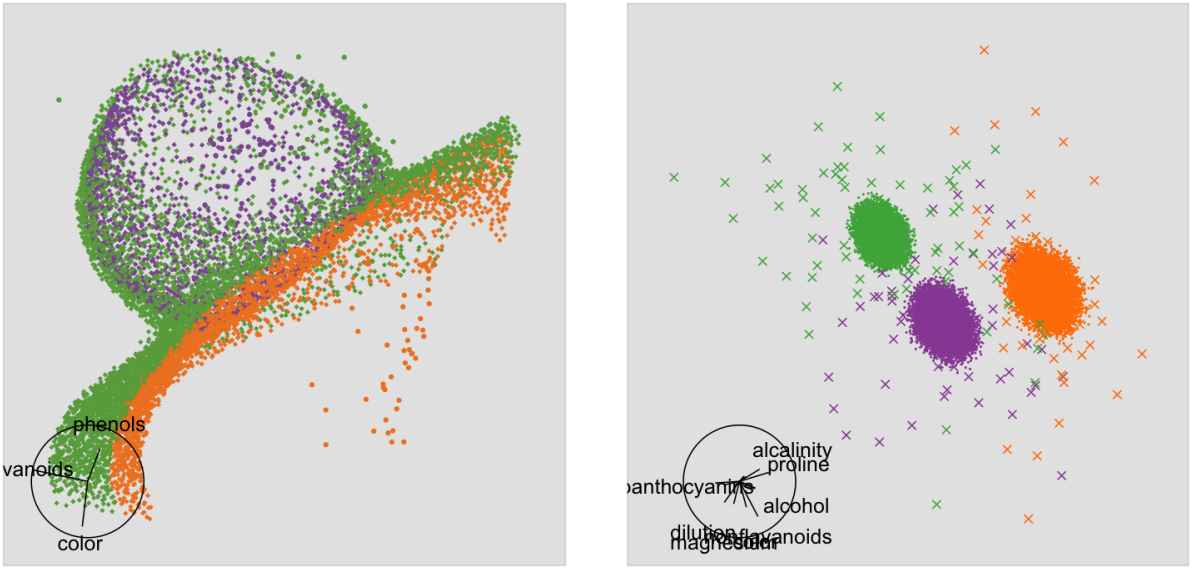
Supervisors: Prof Di Cook (Clayton) and Ursula Laa

email: dicook@monash.edu

Overview

In data analysis, models are commonly functions defined on high-dimensional spaces. A regression model is a hyper-plane, and if nonlinear terms are included it is a nonlinear surface in high-dimensions. Multivariate confidence regions are mostly geometric shapes in high-dimensions. Bayesian posterior distributions are geometric shapes of parameter spaces, and will have as many dimensions as parameters. Supervised classification induces a partitioning of the high-dimensional space that can be considered to be a function in high-dimensions.

This project will take new methods for slicing through high dimensions, and apply them to function visualisation. The actual nature of the problem will depend on the interests of the student. You could chose from any of the areas above to motivate and guide the work.



Skills

Strong working knowledge of R, math background including linear algebra, and multivariate analysis, some background in machine learning and data analysis desirable

Reading

Hadley Wickham, Dianne Cook, Heike Hofmann. Visualizing statistical models: Removing the blindfold. Statistical Analysis and Data Mining: The ASA Data Science Journal, vol. 8, no. 4, pp. 203–225, 2015.

Hadley Wickham, Dianne Cook, Heike Hofmann, Andreas Buja (2011). tourr: An R Package for Exploring Multivariate Data with Projections. Journal of Statistical Software, 40(2), 1-18. URL http://www.jstatsoft.org/v40/i02/.

Topic: Is it arson, carelessness or lightening? Understanding 2019-2020 Australian bushfires
Supervisors: Prof Di Cook (Clayton) and Emily Dodwell

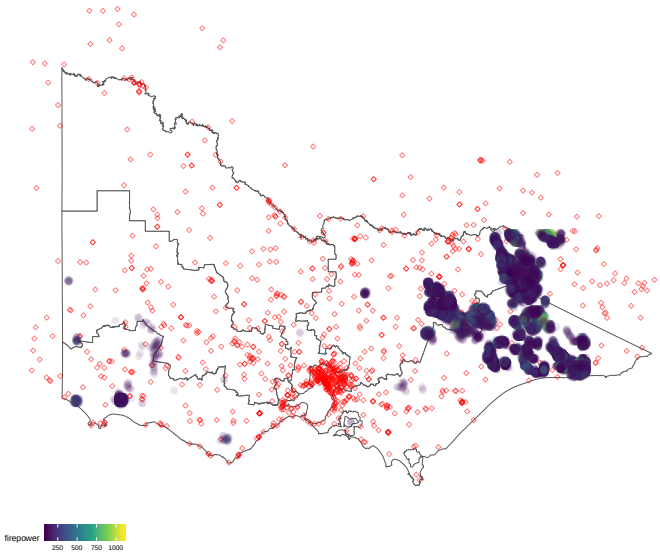
email: dicook@monash.edu

Overview

The summer of 2019-2020 Australia experienced probably the worst bushfires on record, that received attention across the globe. The devastation to property, forests and wildlife was immense. Major cities were choked in smoke. In the social media a cultural war was waged between climate change deniers and activists. Deniers strongly spread stories of arson events on a massive scale, that was then relaxed with calls for returning to preliminary burns to reduce dry vegetation. This was roundly countered with the arguments that vegetation this year was dry and crackling due to an extended drought, brought on by the warming of the Australian temperatures and decrease in precipitation as a result of climate change.

Satellite techology may allow investigating the location and potential sources of fire ignition. Hotspot data from the JAXA's Himawari-8 satellite is available for several years, and can be downloaded with R code similar to https://gist.github.com/ozjimbob. There is an R package, called "bomrang" which allows historical weather data to be downloaded from the Bureau of Meteorology. Fusing these two data sets, with other possible data such as campsite locations, distance to roads, will allow investigation of fire ignition.

This project will involve data construction, visualisation and modeling to tackle this problem. An ideal outcome will be a shiny app to allow others to explore bushfire location and intensity data. Developing models to predict fire risk of neighbourhoods and at different times would also be interesting.



Skills

Strong working knowledge of R, data analysis and statistical modeling are needed. Knowledge of the tidyverse and reproducible research practice is expected.

Reading

Wang, E., Cook, D. and Hyndman, R. J. (2019) A new tidy data structure to support exploration and modeling of temporal data. {*Journal of Computational and Graphical Statistics*} *Available at https://doi.org/10.1080/10618600.2019.1695624*.

Cheng, X., Cook, D. and Hofmann, H. (2016) Enabling Interactivity on Displays of Multivariate Time Series and Longitudinal Data, {Journal of Computational and Graphical Statistics}, {**25}(4):1057–1076.**