MONASH University

MONASH BUSINESS SCHOOL

Monday 26 and Tuesday 27 November 2018

Room H1.16, Building H

Caulfield Campus

Monash University

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION AMBA ACCREDITED

International Workshop on High-Dimensional Time Series and Panel Data Program

Day One – Monday 26 November

Time	Program
9:00 – 9:15	Registration
9:15 – 9:30	Opening Session Chair: Jiti Gao, Monash Business School Speaker: Robert Brooks, Interim Dean, Monash Business School
9:30 – 10:30	**Keynote Session 1** Chair: Farshid Vahid, Monash Business School Speaker: Serena Ng, Columbia University Title: An Econometric View of Algorithmic Sampling
10:30 - 11:00	Morning tea break
11:00 - 12:45	**Invited Session 1** Chair: George Athanasopoulos, Monash Business School 11.00 – 11.35 Speaker: Rob Hyndman, Monash Business School Title: Feature–Based Time Series Analysis 11:35 – 12:10 Speaker: Didier Nibbering, Monash Business School Title: Forecasting Using Random Subspace Methods 12:10 - 12:45 Speaker: Anastasios Panagiotelis, Monash Business School Title: Probabilistic Forecasts in Hierarchical Time Series
12:45 - 14:15	Lunch break
14:15 - 16:00	**Invited Session 2** Chair: Vasilis Sarafidis, Monash Business School 14:15 - 14:50 Speaker: Catherine Forbes, Monash Business School Title: A Robust Bayesian Exponentially Tilted Empirical Likelihood Method 14:50 - 15:25 Speaker: Robert Kohn, The University of New South Wales Title: Efficiently Combining Pseudo Marginal and Particle Gibbs Sampling 15:25 - 16:00 Speaker: Gael Martin, Monash Business School Title: Loss-based Bayesian Prediction

16:00 - 16:30	Afternoon tea break
16:30 - 17:40	**Invited Session 3** Chair: Xueyan Zhao, Monash Business School 16:30 – 17:05 Speaker: David Harris, The University of Melbourne Title: INAR Models with Semi-parametrically Time-Varying Arrivals Distribution 17:05 - 17:40 Speaker: Bonsoo Koo, Monash Business School Title: Regularized Empirical Risk Minimization for Forecast Reconciliation of Hierarchical and Grouped Time Series
19:00 - 22:00	Dinner Session (By invitation only)

Day Two – Tuesday 27 November

Time	Program
9:30 - 10:30	**Keynote Session 2** Chair: Stan Hurn, Queensland University of Technology Speaker: Oliver Linton, University of Cambridge Title: Estimation of a Parsimonious Multiplicative Covariance Structure
10:30 - 11:00	Morning tea break
11:00 - 12:45	**Invited Session 4** Chair: Don Poskitt, Monash Business School 11:00 - 11:35 Speaker: Gianni Amisano, Federal Reserve Board, Washington DC Title: TBA 11:35 - 12:10 Speaker: Heather Anderson, Monash Business School Title: The Effects of Trade Size and Market Depth on Immediate Price Impact 12:10 - 12:45 Speaker: Adrian Pagan, The University of Sydney Title: Varieties of Cycles and Their Detection
12:45 - 14:15	Lunch break

14:15 - 16:00	**Invited Session 5** Chair: Xibin Zhang, Monash Business School 14:15 - 14:50 Speaker: Yundong Tu, Peking University Title: Sieve Estimation and Variable Selection in High–Dimensional Single Index Models 14:50 - 15:25 Speaker: Ying Wang, The University of Auckland Title: Spurious Functional-Coefficient Regression Models and a Robust Solution 15:25 - 16:00 Speaker: Weilun Zhou, Monash Business School Title: Semiparametric Single–Index Predictive Regression
16:00 - 16:30	Afternoon tea break
16:30 - 17:40	**Invited Session 6** Chair: Mervyn Silvapulle, Monash Business School 16:30 - 17:05 Speaker: Natalia Bailey, Monash Business School Title: Exponent of Cross-sectional Dependence for Residuals 17:05 - 17:40 Speaker: Yi He, Monash Business School Title: Locally Optimal Test in High-dimensional Linear Regression
17:40 - 17:50	**Concluding Session** Chair: Jiti Gao, Monash Business School Speaker: Heather Anderson, Monash Business School

Abstracts

Monday 26 November

9:30-10:30

Speaker: Serena Ng, Columbia University

Title: An Econometric View of Algorithmic Sampling

Abstract: Improved technology has dramatically reduced the cost of collecting data, and datasets that are terabytes in size are not uncommon. Not only can computation be slow with data of this size, memory and storage constraints may render analysis infeasible. This has motivated computer scientists to devise subspace reduction schemes while preserving the structure of the original data. We first review the foundation of these methods and highlight results mostly derived from an algorithmic perspective, void of any probabilistic structure. To gain insights from an econometric point of view, we study the implications of subspace reduction from the viewpoint of inference using the linear regression model as our framework. We also propose divide and conquer methods that can make efficient use of the data while minimizing the computational bottlenecks. These methods can be useful in estimation of simple models with lots of data, or in estimation of complex models when working with good sketches of the data can significantly reduce debugging time.

11:00-11:35

Speaker: Rob Hyndman, Monash Business School

Title: Feature-based Time Series Analysis

Abstract: It is becoming increasingly common for organizations to collect very large amounts of data over time. Data visualization is essential for exploring and understanding structures and patterns, and to identify unusual observations. However, the sheer quantity of data available means that new time series visualisation methods are needed. I will demonstrate an approach to this problem using a vector of features on each time series, measuring characteristics of the series. These feature vectors can then be mapped to a 2-dimensional space for visualization. The feature-based approach to time series can also be used to identify the best forecasting model using a pre-trained classifier, and to identify anomalous time series within a collection of time series.

11:35-12:10

Speaker: Didier Nibbering, Monash Business School

Title: Forecasting Using Random Subspace Methods

Abstract: Random subspace methods are a new approach to obtain accurate forecasts in high-dimensional regression settings. Forecasts are constructed by averaging over forecasts from many sub-models generated by random selection or random weighting of predictors. We provide a theoretical justification for these strategies by deriving upper bounds on the asymptotic mean squared forecast error, which show that the methods are particularly suitable for macroeconomic forecasting. An empirical application to the FRED-MD data confirms the theoretical findings, with random subspace methods outperforming all competing methods for at least 66% of the series.

12:10-12:45

Speaker: Anastasios Panagiotelis, Monash Business School

Title: Probabilistic Forecasts in Hierarchical Time Series

Abstract: Forecast reconciliation involves adjusting forecasts to ensure coherence with aggregation constraints. We extend this concept from point forecasts to probabilistic forecasts by redefining forecast reconciliation in terms of linear functions in general, and projections more specifically. New theorems establish that the true predictive distribution can be recovered in the elliptical case by linear reconciliation, and general conditions are derived for when this is a projection. A geometric interpretation is also used to prove two new theoretical results for point forecasting; that reconciliation via projection both preserves unbiasedness and dominates unreconciled forecasts in a mean squared error sense. Strategies for forecast evaluation based on scoring rules are discussed, and it is shown that the popular log score is an improper scoring rule with respect to the class of unreconciled forecasts when the true predictive distribution coheres with aggregation constraints. Finally, evidence from a simulation study shows that reconciliation based on an oblique projection, derived from the MinT method of Wickramasuriya et al (2017) for point forecasting, outperforms both reconciled and unreconciled alternatives.

14:15-14:50

Speaker: Catherine Forbes, Monash Business School

Title: A Robust Bayesian Exponentially Tilted Empirical Likelihood Method

Abstract: This paper proposes a new Bayesian approach for analysing moment condition models using data that may be contaminated by 'outliers'. Building on the Bayesian exponentially tilted empirical likelihood (BETEL) approach of Schennach (2005), auxiliary variables are used in conjunction with relevant moment conditions to stochastically trim potential outliers from the desired posterior distribution. We also demonstrate that both the BETEL and the new robust BETEL (RBETEL) posteriors may be linked to the recent work of Bissiri, Holmes and Walker (2016) who propose a general framework for updating prior belief via a specified loss function. In addition to an empirical illustration, the results of simulation experiments will be reviewed.

14:50-15:25

Speaker: Robert Kohn, University of New South Wales

Title: Efficiently Combining Pseudo Marginal and Particle Gibbs Sampling

Abstract: Particle Markov Chain Monte Carlo methods are used to carry out inference in non-linear and non-Gaussian state space models, where the posterior density of the states is approximated using particles. Deligiannidis 2017 introduce the correlated pseudo marginal sampler and show that it can be much more efficient than the standard pseudo marginal approach. Mendes 2018 propose a particle MCMC sampler that generates parameters that are highly correlated with the states using a pseudo marginal method that integrates out the states, while all other parameters are generated using particle Gibbs. Our article shows how to combine these two approaches to particle MCMC to obtain a flexible sampler with a superior performance to each of these two approaches. We illustrate the new sampler using a multivariate factor stochastic volatility model with leverage.

15:25-16:00

Speaker: Gael Martin, Monash Business School

Title: Loss-based Bayesian Prediction

Abstract: Bayesian predictive distributions quantify uncertainty about unknown (or out-of-sample) values of a random process conditional only on observed data; uncertainty regarding model-specific parameters being integrated out via the usual probability calculus. Uncertainty about the assumed model itself can, in turn, be accommodated via model-averaging, with the implicit assumption being that the true data generating process (DGP) is contained in the set over which the averaging occurs. We move away from this so-called M-closed world, in which the true DGP is assumed to be either known with certainty, or known to lie in a finite set of models. Operating with an M-open view of the world, we instead construct Bayesian predictives, not

based upon a given model, or set of models, but by adopting a user-supplied concept of predictive performance loss. To develop such machinery in the Bayesian paradigm, we rely on the principles underlying approximate Bayesian computation. Specifically, construction of prediction distributions is carried out using simulation, summary statistics that minimize predictive loss over a pre- specified training period, and a tolerance level that captures our risk aversion to predictive loss. Different methods of defining predictive loss are explored, including one in which performance is benchmarked against a parsimonious 'auxiliary predictive'. Comparison with the 'exact' Bayesian predictive based on a misspecified DGP is undertaken, in simulation scenarios. Using this new approach, we also propose a robust diagnostic procedure for detecting model misspecification. Empirical illustrations using both time series and cross-sectional data are provided.

16:30-17:05

Speaker: David Harris, The University of Melbourne

Title: INAR models with semi-parametrically time-varying arrivals distribution

Abstract: We consider extensions of the semi-parametric INAR(1) model of Drost, van den Akker and Werker (2009, JRSSB) for count data (e.g. of queue lengths) to permit the arrivals probabilities to vary over time as semi- or non-parametric functions of one or more covariates. Asymptotic and finite sample properties are considered for the NPMLEs. Hypothesis tests for time-variation in the arrivals probabilities are also constructed and analysed.

17:05-17:40

Speaker: Bonsoo Koo, Monash Business School

Title: Regularized Empirical Risk Minimization for Forecast Reconciliation of Hierarchical and Grouped Time Series

Abstract: Forecasting hierarchical time series requires good forecast accuracy at each level of the hierarchy, but it is also essential to satisfy the aggregation constraints. Given some base forecasts, the state-of-the-art reconciliation method adjusts these forecasts to ensure that the aggregation constraints are satisfied. However, the accuracy of this method relies on two unbiasedness assumptions which are hard to justify in practice. We propose a new forecast reconciliation algorithm based on empirical risk minimization which does not impose unbiasedness assumptions. To allow forecast reconciliation in high-dimensional settings, we also propose a regularization procedure and provide its theoretical justification. Furthermore, we identify similarities and differences between our approach and the state-of-the-art reconciliation method. Finally, we compare different reconciliation methods on both simulated data sets and an electricity smart meter data set. The results indicate that our approach works well with both simulated and real data.

Tuesday 27 November

9.30-10:30

Speaker: Oliver Linton, University of Cambridge

Title: Estimation of a Parsimonious Multiplicative Covariance Structure

Abstract: We propose a Kronecker product model for correlation or covariance matrices in the large dimension case. The number of parameters of the model increases logarithmically with the dimension of the matrix. We propose a minimum distance (MD) estimator based on a log-linear property of the model, as well as a one-step estimator, which is a one-step approximation to the quasi-maximum likelihood estimator (QMLE). We establish the rate of convergence and a central limit theorem (CLT) for our estimators in the large dimensional case. A specification test and tools for Kronecker product model selection and inference are provided. In an Monte Carlo study where a Kronecker product model is correctly specified, our estimators exhibit superior performance. In an empirical application to portfolio choice for S&P500 daily returns, we demonstrate that certain Kronecker product models are good approximations to the general covariance matrix.

11:00-11:35

Speaker: Gianni Amisano, Federal Reserve Board, Washington DC

Title: TBA

11:35-12:10

Speaker: Heather Anderson, Monash Business School

Title: The effects of trade size and market depth on immediate price impact

Abstract: We compare trade size to the prevailing market depth in the limit order book to detect and account for zero impact trades in an immediate price impact model. Our model also incorporates standard trade attributes (trade size, market capitalization and volatility) in a dynamic, nonparametric setting. The incorporation of market depth information reduces the mean absolute (squared) forecast error of an immediate price impact prediction by over 40% (20%). After controlling for trade attributes, market depth, price impact dynamics, nonparametric modelling, and intra-and inter-day periodicities (in order of relative importance) all improve the prediction of a trade's price impact.

12:10-12:45

Speaker: Adrian Pagan, The University of Sydney

Title: Varieties of Cycles and Their Detection

Abstract: From the history above we see that cycles involve recurrent events and the phases describing the recurrent events were separated by turning points. It is also clear that there was a great deal of emphasis upon the duration of the cycle and identifying them with the names of individuals, rather than upon what the cycles were measuring. Increasingly, however, cycles have become attached to the names of the series being investigated. Thus we now have financial cycles, leverage cycles, building cycles, commodity cycles, real estate cycles, inflation cycles, interest rate cycles. This has led to a great variety of cycles being described and classified in the literature.

A first question that needs to be addressed then is how one detects such cycles. This paper starts to look at a cycle in a single series. Here there are three types of methods. Although all involve locating turning points they differ in being are applied to the series after some transformation of it. With two of the common transformations one might look at what type of oscillation is in the transformed series, and this has been the approach captured by spectral methods. Because oscillations have turning points there is a connection between all of these different approaches, which we will discuss. The paper then moves on to how one detects oscillations and turning points when more than one series is involved. In that case we need to come up with a single set of turning points or a single oscillation. Finally, the paper discusses two applications - what has been termed the medium term cycle and the relationship between financial and real cycles.

14:15-14:50

Speaker: Yundong Tu, Peking University

Title: Sieve estimation and variable selection in high-dimensional single index models

Abstract: This paper considers sieve estimation in high-dimensional single index models. The use of Hermite polynomial in approximating the unknown link function provides a convenient framework to conduct both estimation and variable selection. The estimation of the index parameter is formulated as solutions obtained from the routine penalized linear regression procedure but with weight, in order to tackle unbounded support of the regressor. The resulting index parameter estimator is shown to be consistent and sparse. The asymptotic normality of the estimator of the index and that of the link function obtained by post-OLS are established. Numerical results show that both the variable selection procedure and the associated estimators perform well in finite samples.

14:50-15:25

Speaker: Ying Wang, The University of Auckland

Title: Spurious Functional-Coefficient Regression Models and A Robust Solution

Abstract: Functional-coefficient cointegrating models have become popular to model non-linear nonstationarity in econometrics (Cai et al., 2009; Xiao, 2009). However, there is rare study on testing the existence of functional-coefficient cointegration. Consequently, functional-coefficient regressions involving nonstationary regressors may be spurious. This paper investigates the effect that spurious functional-coefficient regression has on the usual diagnostics. We find that common characteristics of spurious regression are manifest, including divergent local significance tests, random local goodness-of-fit, and local Durbin-Watson ratio converging to zero, complementing those discovered in spurious linear and nonparametric regressions (Phillips, 1986, 2009). In addition, spuriousness causes the divergence of the global significance tests proposed by Xiao (2009) and Sun et al. (2016), which is likely to produce misleading conclusions for practitioners. To resolve the problems, we propose a simple-to-implement inference procedure based on a semiparametric balanced regression, by augmenting regressors of the original spurious regression with lagged dependent variable and independent variables. This procedure achieves spurious regression detection via standard inferential asymptotics, and is found robust to the true relationship between integrated processes. Monte Carlo simulations show that the balanced regression based tests have very good finite sample performance.

15:25-16:00

Speaker: Weilun Zhou, Monash Business School

Title: Semiparametric Single-Index Predictive Regression

Abstract: This paper studies a semi-parametric single-index predictive regression model with multiple nonstationary predictors that exhibit co-movement behaviour. Orthogonal series expansion is employed to approximate the unknown link function in the model and the estimator is derived from an optimization under the constraint of identification condition for index parameter. The main finding includes two types of super-consistency rates of the estimators of the index parameter along two orthogonal directions in a new coordinate system. The central limit theorem is established for a plug-in estimator of the unknown link function. In the empirical studies, we provide evidence in favour of nonlinear predictability of the stock return using long term yield and treasury bill rate.

16:30-17:05

Speaker: Natalia Bailey, Monash Business School

Title: Exponent of Cross-sectional Dependence for Residuals

Abstract: In this paper we focus on estimating the degree of cross-sectional dependence in the error terms of a classical panel data regression model. For this purpose we propose an estimator of the exponent of cross-sectional dependence denoted by α, which is based on the number of non-zero pair-wise cross correlations of these errors. We prove that our estimator, $\tilde{\alpha}$, is consistent and derive the rate at which $\tilde{\alpha}$ approaches its true value. We evaluate the finite sample properties of the proposed estimator by use of a Monte Carlo simulation study. The numerical results are encouraging and supportive of the theoretical findings. Finally, we undertake an empirical investigation of α for the errors of the CAPM model and its Fama-French extensions using 10-year rolling samples from S&P 500 securities over the period Sept 1989 - May 2018.

17:05-17:40

Speaker: Yi He, Monash Business School

Title: Locally Optimal Test in High-dimensional Linear Regression

Abstract: The number of variables in state-of-the-art economic datasets can grow as fast as the sample size. Applying classical econometric methods to big data may only improve in-sample fitness but, paradoxically, worsen the learning of the population data generating process. In particular, the curse of dimensionality can reduce the power of Wald-type tests down to a trivial level.

Using random matrix theory, we obtain the local asymptotic power of an universe of weighted quadratic tests in high-dimensional linear regression models when the number of variables is comparable to, even large than, the sample size. We identify an uniformly most powerful test for these local alternatives with robust power envelop against the dimensionality. Our approach requires only explicit weighting matrices in the quadratic test statistic, and allows the true model to deviate from the null hypothesis in dense, sparse, and often completely arbitrary direction. Simulation results agree with our asymptotic results.