Department of Statistics: PhD Presentation Event

Thursday 10th and Friday 11th May 2012
Leverhulme Library (COL 6.15), Columbia House

Abstracts

YEHUDA DAYAN	TBC

BAOJUN DOU	Sparse factor modelling for high dimensional time series

Analyzing multiple time series via factor model is one of the frequently used methods to achieve dimension reduction. Modern time series analysis concentrates on the situation when the number of time series *p* is as large as or even larger than the sample size *n*. Under the stationary settings, Lam and Yao (2011) estimate the factor loading matrix and factor process itself by carrying out an eigenanalysis on a *p* by *p* non-negative definite matrix. Based on this method, we develop an estimation approach for sparse factor model. The motivation for such a sparse model is that for some sort of data set, like environmental time series, each common factor drives the dynamics of only part of the original time series. In light of the success of the lasso (elastic net) in regression, we transform the eigenanalysis problem to a regression-type optimization problem. Then sparse solutions are obtained by adding the lasso (elastic net) penalty. Future work will focus on investigating convergence rate of the estimated factor loadings (for both large *p* and *n*) and model selection consistency (or strongly, sign consistency) of our method.

JOSEPH DUREAU	A Bayesian approach to estimate time trends in condom use following a targeted HIV prevention programme

While significant resources are being invested to implement interventions that "should" work, an increasing attention is given to intervention monitoring. As condom use remains the principal mean to mitigate HIV transmission, public deciders are interested in the impact of past intervention on this quantity and in its present level, in order to design future policies. However, as the measure of condom use is impossible or subject to important declarative bias, dynamic transmission models are used to explore different scenarios of HIV propagation in the population, to perform Bayesian inference from HIV prevalence observations. This approach reconstructs the trajectory of condom use over the last 25 years from a limited number of observations, hence estimates could strongly depend on how this trajectory is modelled: different alternatives are discussed and their properties are explored in details in a the perspective of informed and conservative decision making.

MAI HAFEZ	A latent variable model for multivariate longitudinal data subject to dropout

Longitudinal data are collected for studying changes across time. It is very common that the phenomenon of interest is a latent "unobserved" attitude or construct (e.g. democracy, happiness…etc.). When this is the case, a number of observed variables (e.g. items from a questionnaire) are measured at each time point and the latent variable of interest is assumed to be underlying those observed variables. Dropout is a very common problem in longitudinal studies where subjects exit the study prematurely. Ignoring the dropout mechanism can lead to biased estimates, especially when the dropout is "non-ignorable". Our proposed approach uses latent variable models to capture the evolution of the latent phenomenon over time while accounting for dropout (possibly non-random). A latent variable, which accounts for dependencies among items, is used to summarise an attitude or a construct within each time point. These attitudinal latent variables are linked to account for dependencies among items between different time points. Random effects are also included to account for the repetition of items over time. Another latent variable is introduced to summarise dropout. Different relationships among these latent variables are studied via three different model specifications.

SARAH HIGGINS

How skilful are seasonal probability forecasts constructed from multiple models?

Ensemble forecasting on a lead time of seconds over several years generates a large forecast-outcome archive, which can be used to evaluate and weight "models." In weather forecasting one typically has only thousands of forecasts however those launched 6 hours apart are not independent of each other, nor is it justified to mix seasons with different dynamics. Seasonal forecasts from ENSEMBLES typically have less than 64 unique launch dates; decadal forecasts less than eight, and long range forecasts arguably none.

The impact of using different skill scores is examined with a large forecast-outcome archive. The robustness of the results in large archives is demonstrated using imperfect models of perfectly known non -linear (chaotic) dynamical systems. The implications these results hold for distinguishing the skill of a forecast from its value to a user of the forecast are discussed. It is shown that blending with climatology (J Brocker and L A Smith, Tellus A, 60(4), 663-678, (2008)) tends to increase the robustness of the results.

NA HUANG

Precision matrix estimation via pairwise tilting

We propose a *tilting*-based method to estimate the precision matrix of a *p*-dimensional random variable, **X**, when p is possibly much larger than the sample size n. Each 2 x 2 block indexed by (*I*, *j*) of the precision matrix can be estimated by the inversion of the pairwise sample conditional covariance matrix of X_i and X_j controlling for all the other variables. However, in the high dimensional setting, including too many or irrelevant controlling variables may distort the results. To determine the controlling subsets in high dimensional scenarios, the proposed method applies the *tilting* technique to measure the contribution of each remaining variable to the variance-covariance matrix of X_i and X_j, and only puts the (hopefully) highly relevant remaining variables into the controlling subsets. The simulation results will be presented under different scenarios for the underlying precision matrix. Comparison with other competing methods will also be given.

References:

[1] H. Cho and P. Fryzlewicz. High-dimensional variable selection via tilting. To appear in the *Journal of the Royal Statistical Society Series B*. 2012.
[2] J. Peng, P. Wang, N. Zhou and J. Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*. 104: 735-745, 2009.
[3] S. L. Lauritzen. *Graphical models.* Oxford University Press, Oxford, 1996.

ALEX JARMAN

Misleading estimates of forecast quality: quantifying skill with sequential forecasts

Quantifying the skill in probability forecasts is complicated by the fact that the behaviour of many physical systems changes slowly over long timescales. This can lead to unreliable evaluation of the skill and the value of a forecast system. Demonstrating robust out-of-sample skill requires a sufficient sample of probability forecasts and corresponding verifications. Larger samples of evaluations result in more precise confidence intervals but what often appears to be overlooked (D. Wilks. QJRMS, 136(653), pp2109-2118, 2010) is the effect of serial correlation in a forecast/outcome archive on skill score statistics. Linear autocorrelation in a forecast time series can contribute to variance inflation in the sampling distribution, and thus result in falsely narrowed confidence intervals. The investigation is applied to the ignorance score in a broader range of systems where sample variance inflation effects have been observed. Sampling variances can sometimes be deflated in a linearly auto-correlated time series. In fact, it is not merely linear correlation but the lack of independence of consecutive forecasts which can lead to a misleading estimate of skill. This is demonstrated using a chaotic time series which lacks independence yet has no linear autocorrelation in the forecast/observation time series. These results support the need for sample size corrections to avoid overconfidence in forecast skill but also indicate that a forecast user should be aware of the implications of any serial correlation for statistical inference with skill scores.

KAROLOS KORKAS

Adaptive estimation for locally stationary autoregressions

We present an adaptive nonparametric method that estimates time-varying AR(p) coefficients. For this purpose we adapt the Fussed Lasso method of Tibshirani et al (2005), an approach in penalised least squares problem which uses simultaneously the total variation and the L1-norm penalties. The additional penalty aims to penalise differences between neighbouring coefficients when there is some natural ordering of those coefficients and hence favour solutions that are both sparse and blocky. We present the properties of the estimator and two algorithms which solve the above problem.

JIA WEI LIM	Parisian option pricing: A recursive solution for the density of the Parisian stopping time

In this paper, we obtain the density function of the single barrier one-sided Parisian stopping time. The problem reduces to that of solving a Volterra integral equation of the first kind, where a recursive solution is consequently obtained. The advantage of this new method as compared to that in previous literature is that the recursions are easy to program as the resulting formula only involves a finite sum and does not require a numerical inversion of the Laplace transform. For long window period, an explicit formula for the density of the stopping time can be obtained. For shorter window length, we derive a recursive equation from which numerical results are computed. From these results, we compute the prices of one-sided Parisian options.

YANG YAN	Efficient estimation of conditional risk measures in a semiparametric GARCH model

This paper proposes efficient estimators of risk measures in a semiparametric GARCH model defined through moment constraints. Moment constraints are often used to identify and estimate the mean and variance parameters and are however discarded when estimating error quantiles. In order to prevent this efficiency loss in quantile estimation, we propose a quantile estimator based on inverting an empirical likelihood weighted distribution estimator. It is found that the new quantile estimator is uniformly more efficient than the simple empirical quantile and a quantile estimator based on normalized residuals. At the same time, the efficiency gain in error quantile estimation hinges on the efficiency of estimators of the variance parameters. We show that the same conclusion applies to the estimation of conditional Expected Shortfall. Our comparison also leads to interesting implications of residual bootstrap for dynamic models. We find that these proposed estimators for conditional Value-at-Risk and expected shortfall are asymptotically mixed normal. This asymptotic theory can be used to construct confidence bands for these estimators by taking account of parameter uncertainty. Simulation evidence as well as empirical results are provided.
(With Dajing Shang and Oliver Linton)