



# 15<sup>th</sup> OxMetrics User Conference

## Abstracts

### Session 1: Modelling

#### *Milton Friedman as a Statistician and Econometrician*

**Neil R. Ericsson** (Board of Governors of the Federal Reserve System), David F. Hendry, and Stedman B. Hood

**Abstract.** Milton Friedman began his career as a statistician, making original contributions to several areas of statistics, before turning to economics as his primary focus. His first few research experiences led him to accord importance to errors of measurement, both with respect to conceptual variables, as well as inaccurate recordings, and that problem remained a life-long interest. We focus on Friedman's empirical studies of money, and address a number of issues about his general approach to modeling and to data adjustment, for example by using phase averages rather than the original annual series. We contrast that with using model augmentation, which Friedman undertook additional to data adjustments, rather than just the former alone.

#### *Using Forecasting to Detect Corruption in International Football*

**James J. Reade** (University of Reading)

**Abstract.** Corruption is hidden action aimed at influencing the outcome of an event away from its competitive outcome. It is likely common in all walks of life yet its hidden nature makes it difficult to detect, while its distortionary influence on resource allocation ensures the importance of trying to detect it both practically and economically. This paper further develops methods to detect corrupt activity contained in Olmo et al. (2011) and Reade (2013) that make use of different forecasting methods and their information sets to detect corruption. We collect data from 63 bookmakers covering over 9,000 international football matches since 2004 and assess a claim made in early 2013 by Europol that the outcomes of almost 300 international matches since 2009 were fixed. Our collected data consists of match outcomes and prematch bookmaker odds, which we use to explore the divergence between two kinds of forecasts of match outcomes: those by bookmakers, and those constructed by econometric models. We argue that in the absence of corrupt activity to fix outcomes these two forecasts should be indistinguishable as they are based on the same information sets, and hence any divergence between the two may be indicative of corrupt activity to fix matches. Such an assertion is conditional on the quality of the econometric model and in this paper we discuss the peculiarities of modelling international football match outcomes. In the absence of corroborating evidence we cannot declare any evidence procured in our manner as conclusive regarding the existence or otherwise of corruption, but nonetheless we argue that it is indicative. We conclude that there is mild evidence regarding potentially corrupt outcomes, and we also point towards yet more advanced strategies for its detection.

## Session 2: Volatility

### *On the Formulation of ARCH in Mean Models*

**Andrew Harvey** (Cambridge University) and Rutger-Jan Lange.

**Abstract.** Volatility of a stock may incur a risk premium, leading to a positive correlation between volatility and returns. On the other hand the leverage effect, whereby negative returns increase volatility, acts in the opposite direction. We propose a two component ARCH in Mean model to separate the two effects; such a model also picks up the long memory features of the data. An exponential formulation, with the dynamics driven by the score of the conditional distribution, is shown to be theoretically tractable as well as practically useful. In particular it enables us to write down the asymptotic distribution of the maximum likelihood estimator, something that has not proved possible for standard formulations of ARCH in Mean. A model in which the returns have a skewed generalized-t distribution is shown to give a good fit to S&P500 excess returns over a 60-year period (4 Jan 1954- 30 Dec 2013).

### *Modeling Dynamic Diurnal Patterns in High Frequency Financial Data*

**Ryoko Ito** (University of Cambridge)

**Abstract.** We develop the spline-DCS model to forecast the conditional distribution of high-frequency financial data with periodic behavior. The dynamic cubic spline of Harvey and Koopman (1993) is applied to allow diurnal patterns to evolve stochastically over time. The model is robust to outliers as the dynamics of scale is driven by the score. An empirical application illustrates the practicality and impressive predictive performance of the model. It also illustrates that allowing dynamic diurnal patterns can lead to an improvement in the quality of the fit of the model to the empirical distribution of data, especially around the upper extreme quantiles.

## Session 3: (SPEED)

### *Capital gain: The Firm Value of Operating Next to the Government*

**Jörg Stahl** (Ph.D. student, University Pompeu Fabra)

**Abstract.** The literature identifies various city characteristics that attract firm head-quarters. However, the extraordinary high degree of concentration of headquarters in capital cities has not received much attention. This suggests that the literature may be missing an important determinant of headquarters agglomeration - the presence of the government. Geographic proximity to a country's key politicians may be of advantage for firms' decision-makers in their attempt to influence the policy-making process. In this paper, I examine a unique event - the relocation of the German Federal Government from Bonn to Berlin - to determine the firm value effects of locating headquarters close to the government. I apply a Fama-French Multi-Factor framework, and find that firms with operational headquarters in Berlin experienced mean cumulative abnormal returns of about 3% within the two days following the relocation decision. These returns were even higher two weeks later, do not seem to be driven by industry composition, and are robust to different model specifications.

### *Unemployment Hysteresis and Cycle Asymmetry - A Case Study*

**António Neto** (Ph.D. student, Faculdade de Economia da Universidade do Porto)

**Abstract.** This paper aims to capture the dynamics of the Portuguese unemployment rate. We first assess if the series follows a unit root process as to confirm the hysteresis hypothesis. Then, we develop a nonlinear model to test for the asymmetric behavior of unemployment across cycle phases. Our results lend support for hysteresis and show that the Portuguese unemployment dynamics is better described by a nonlinear rather than by a linear model. Thus strong enough short-run increases in unemployment, as those observed during the recent fiscal consolidation effort, have non-negligible impacts on raising the Portuguese natural rate of unemployment.

### *The Pairwise Approach to Model a Large Set of Disaggregates with Common Trends*

**Guillermo Carlomagno** (Ph.D. student, Universidad Carlos III de Madrid)

**Abstract.** The objective of this paper is to model and forecast all the components of a macro or business variable. Our contribution concerns cases with a large number (hundreds) of components where multivariate approaches are not feasible. We extend in several directions the pairwise approach originally proposed by Espasa and Mayo-Burgos (2013) and study its statistical properties. The pairwise approach consists on performing common features tests between the  $N(N-1)/2$  pairs of series that exist in a group of  $N$  of them. Once this is done, groups of series that share common features can be formed. Next, all the components are forecast using single equation models that include the restrictions derived by the common features. In this paper we focus on discovering groups of components that share single common trends. The asymptotic properties of the procedure are studied analytically. Monte Carlo evidence on the small samples performance is provided and a small samples correction procedure designed. A comparison with a DFM alternative is also carried out, and results indicate that the pairwise approach dominates in many empirically relevant situations. A relevant advantage of the pairwise approach is that it does not need common features to be pervasive. A strategy for dealing with outliers and breaks in the context of the pairwise procedure is designed and its properties studied by Monte Carlo. Results indicate that the treatment of these observations may considerably improve the procedure's performance when series are 'contaminated'.

### *Joint modelling UHF Financial Data under Autoregressive Conditional Duration model specification*

**Handy Tanuwijaya** (Ph.D. student, University of Essex)

**Abstract.** The aim of this paper is to model the relationships of irregularly spaced prices and irregularly spaced volume trades of ultra high frequency data on the forecasting ability of assets prices. Microstructure literatures have shown that, individually, both prices and volumes exhibit a Martingale process and contain information about future prices, specifically the unobservable information flow that can be used to forecast prices. The next step is to test intertemporal prices and volumes jointly using a Autoregressive Conditional Duration (ACD) model (ACD henceforth) to forecast stochastic returns. Using Dow Futures tick-by-tick data, provided by Tick Data Corp. ([www.tickdata.com](http://www.tickdata.com)), the goal is to decipher whether the Joint ACD model contains more information on price prediction than the individual duration models would have predicted.

## Session 4: Ana Timberlake Memorial Lecture

*The Theory of Statistics: Some History and the Current Position*

Sir David Cox (University of Oxford)

## Session 5: Estimation and testing

*Endogeneity in Parametric Duration Models with Applications to Clinical Risk Indices*

Anand Acharya, **Lynda Khalaf** (Carleton University), Marcel Voia, David Wensley

**Abstract.** We consider the problem of endogenous predictors and unmeasured confounding in an accelerated life regression model. An observational Canadian clinical patient data set provides a unique experimental-like setting, in which we propose the trauma status of a critical care patient as an instrument to correct for the endogenous risk of mortality index in a commonly specified length of stay regression. Specifically, we provide an exact, identification-robust method to correct for endogeneity and unmeasured confounding in the location-scale family of parametric duration models using instrumental variables and a generalized Anderson Rubin statistic. Our analysis suggests that increased risk of mortality results in an increased length of stay in intensive care. In particular, the methods developed here correct for the bias of conventional estimates. Using the instrument and methods suggested in this paper, routinely collected patient risk indices are meaningful for informing policy in the health care setting.

*The Trajectory of Psycho-Social Depression in Ukraine following the Chernobyl Nuclear Accident*

**Bob Yaffee** (New York University) T. B. Borak, R.M. Perez Foster, R. Frazier, M. Burdina, V. Chtenguelov, and G. Prib

**Abstract.** Our objectives were to examine predictive parameters of psychological impacts, resulting from the Chernobyl accident, on residents living in the oblasts of Kiev and Zhitomyr. We tested drivers for psycho-social depression based on estimates radiological dose received from radioactivity release during the accident and the perception of increased health effects associated with this radiation. To obtain a representative sample of individuals, we attached computer generated random numbers to area codes provided by the telephone company. In January 2009, Russia created an intervening crisis by interrupting supplies of natural gas to the Ukraine. We employed modified scenario forecasting to circumvent crisis effects that could otherwise undermine the internal validity of our study. State space methods were used to model and graph trajectories of psycho-social depression reported by male and female respondents. Results of the dose reconstruction process revealed that the dose received by this population was too low to identify pathological disease or injury. From our empirical analysis, we found that the psychological impacts of the nuclear incident stemmed from perceived risks, rather than actual exposure to radiation directly associated with the Chernobyl nuclear accident. Work funded by NSF HSD Grant 082 6983.

### *Time-Varying Temporal Dependence in Autoregressive Models: An Observation Driven Approach*

F. Blasques, **Siem Jan Koopman** (VU University Amsterdam, Netherlands, & CREATES, Aarhus University, & Tinbergen Institute) and A. Lucas

**Abstract.** The temporal dependence in an autoregressive process is determined by the autoregressive coefficients which are typically estimated by the method of maximum likelihood. We develop a novel and flexible method for the estimation of time-varying temporal dependence by adopting a recursive updating procedure based on the score of the predictive likelihood function at each time point. The resulting autoregressive model can be expressed in reduced form as a nonlinear dynamic model that relates to the class of threshold and smooth-transition autoregressive models. Furthermore we establish the information theoretic optimality of the score driven update of the autoregressive coefficient. The behavior of the new model is studied in a Monte Carlo exercise. An empirical application to macroeconomic data shows that the model outperform its most direct competitors, in-sample and out-of-sample.

### *Maximum Non-extensive Entropy Bootstrap*

Michele Bergamelli, **Jan Novotný** (Cass Business School, City University London, UK and CERGE-EI, CZ), and Giovanni Urga

**Abstract.** In this paper, we employ a generalized concept of non-extensive entropy to extend a maximum entropy (ME) bootstrapping method. Such a procedure, called the maximum non-extensive entropy (MNE) bootstrap, preserves all the main features of the ME bootstrap available in the literature, while it provides more flexibility in modelling the fat tails. The evaluation of the overall performance of the proposed procedure is undertaken by an extensive simulation exercise, where we compare behavior of both the MNE vs ME, and the behavior of the MNE vs existing methods (Continuous Path Block Bootstrap and Residuals Bootstrap) by considering a unit root test p-values of I(1) series.

## **Session 7: Model selection**

### *Model Selection with 'Big Data'*

**Sir David Hendry** (University of Oxford)

**Abstract.** Big Data come in many shapes and sizes, offering many potential benefits but confronting a range of problems, including an excess of false positives, mistaking correlations for causes, ignoring sampling biases, and analyzing by inappropriate methods. We consider the many important requirements of any procedure searching for a data-based relationship using 'big data', and the possible role of Autometrics in that context. Paramount considerations include embedding relationships in general initial models, using good quality data on all variables, analyzed with tight significance levels by a powerful selection procedure, restricting the number of variables to be selected over using non-statistical criteria, testing for relationships being invariant to shifts in explanatory variables, and checking for hidden dependence and endogeneity.

### *Robust Estimation of Real Exchange Rate Process Half-life*

**Michele Bergamelli** (Cass Business School)

**Abstract.** In this paper, we show that the data generating process of the real exchange rate is likely to include outliers that, if not accounted for, produce unreliable half-lives estimates. In order to obtain robust estimates of the half-life, we propose to detect outlying observations by means of an extension of the Dummy Saturation approach (Hendry et al., 2008; Johansen and Nielsen, 2009) to ARMA processes, considering additional and innovative outliers as well as level shifts in the real exchange rate process. An empirical application involving US dollar real exchange rates allows us to conclude that the estimated half-lives are consistently shorter after outlying observations are correctly modelled, thus shedding some light on the PPP puzzle.

### *Outlier Detection Algorithms for Least Squares Time Series Regression*

Søren Johansen and **Bent Nielsen** (Nuffield College & Department of Economics, University of Oxford & Programme on Economic Modelling, INET, Oxford)

**Abstract.** We review recent asymptotic results on some robust methods for multiple regression. The regressors include stationary and non-stationary time series as well as polynomial terms. The methods include the Huber-skip M-estimator, 1-step Huber-skip M-estimators, in particular the Impulse Indicator Saturation, iterated 1-step Huber-skip M-estimators and the Forward Search. These methods classify observations as outliers or not. From the asymptotic results we establish an asymptotic theory for the gauge of these methods, which is the expected frequency of falsely detected outliers. The asymptotic theory involves normal distribution results and Poisson distribution results. The theory is applied to two time series data sets.

### *Estimating Regression Models with More Variables than Observations.*

**Jurgen Doornik** (Oxford University, UK)

**Abstract.** Autometrics incorporates an algorithm that allows the estimation of regression models that have more variables than observations. This has turned out to be very useful: it allows, e.g., estimating models for developing economies where the sample size is small. It has also acted as a spur for research into saturation estimators such as impulse indicator saturation (IIS), step indicator saturation (SIS), differenced IIS and others. We review the algorithm, as well as the simulation results obtained so far. Potential modifications are considered.

## Session 8: High frequency econometrics

### *Persistence Through Correlation*

**Guillaume Chevillon** (ESSEC Business School & CREST) Alain Hecq (Maastricht University) and Sébastien Laurent (Aix-Marseille University & GREQAM)

**Abstract.** This paper analyzes a novel source of long memory though multicollinearity. We consider a vector autoregression of order one, a VAR(1) of large dimension. We use a final equation representation to show that as the VAR dimension tends to infinity while the proportion of stochastic trends remains constant, individual variables may exhibit strong persistence akin to fractional integration whose degree corresponds to the fraction of unit roots in the system. We consider the implications of our findings for the volatility of asset returns where the so-called golden-rule of realized volatility states that they always exhibit fractional integration of degree close to 0.4. Hence, this empirical feature can be related to the correlation of the many financial assets.

### *System-wide Tail Comovements: a Bootstrap Test for Cojump Identification on the S&P 500, US bonds and Exchange Rates*

Jean-Yves Gnaboy, Lyudmyla Hvozdyk, and **Jerome Lahaye** (Fordham University)

**Abstract.** This paper studies bivariate tail comovements on financial markets that are of crucial importance for the world economy: The S&P 500, US bonds, and currencies. We propose to study that form of dependence under the lens of cojump identification in a bivariate Brownian semimartingale with idiosyncratic jumps, as well as cojumps. Whereas univariate jump identification has been widely studied in the high-frequency data literature, the multivariate literature on cojump identification is more recent and scarcer. Cojump identification is of interest, as it may identify comovements which are not trivially visible in a univariate setting. That is, price changes can be small relative to local variation, but still abnormal relative to local covariation. This paper investigates how simple parametric bootstrapping of the product of assets' intraday returns can help detect cojumps in a multivariate Brownian semi-martingale with both idiosyncratic jumps and cojumps. In particular, we investigate how to disentangle idiosyncratic jumps from common jumps at an intraday level for pairs of assets. The approach is exible, trivial to implement, and yields good power properties. It allows to shed new light on extreme dependence at the world economy level. We detect cojumps of heterogeneous size which are partly undetected with a univariate approach. We find an increased cojump intensity after the crisis on the S&P 500-US bonds pair before a return to normal.

### *Roughing it Up Some More: Jumps and Co-Jumps in Vast-Dimensional Price Processes*

Kris Boudt, **Sébastien Laurent** (Aix-Marseille School of Economics, CNRS & EHES, France) and Rogier Quaadvlieg

**Abstract.** An estimator of the ex-post covariation of log-prices under asynchronicity, microstructure noise and finite activity jumps is proposed. The estimator is based on the CholCov introduced in Boudt et al. (2014), and therefore ensures a positive semidefinite estimate. Monte Carlo simulations confirm good finite sample properties. In the application we demonstrate the benefits of decomposing the quadratic covariation into a continuous and a jump component. Based on adaptations of a range of popular multivariate volatility models, we show the decomposition leads to improved fit and forecasting abilities.

### *Comparing Alternative Integrated Covariance Estimators*

Simona Boffelli (Bergamo University, Italy) and **Giovanni Urga** (Cass Business School, UK and Bergamo University, Italy)

**Abstract.** The presence of non-synchronous tradings and market microstructure noise in tick-by-tick data greatly affect the performance of covariance estimators. In this paper, the performance of several integrated covariance estimators is evaluated via a comprehensive Monte Carlo analysis by considering alternative synchronization methods and different levels of microstructure noise and liquidity of transactions. The best performance, in terms of RMSE, is showed by the estimators proposed by Aït-Sahalia, Fan and Xiu (2010) and by Shephard and Xiu (2012). The optimal performance of the Aït-Sahalia, Fan and Xiu (2010) estimator is achieved in combination with the refresh time synchronization procedure, while the Shephard and Xiu (2012), relying on Kalman filter and disturbance smoother, is directly applied to non-synchronized data. Finally, we report a backtesting risk management exercise based on a portfolio of European government bonds.