International Conference on
"*High-Dimensional Econometric Modelling*"
3th – 4th December 2010

Venue: Cass Business School, 106 Bunhill Row, London, EC17 8TZ
Room 2002 (3 Dec) & Room 2003 (4 Dec)

Organiser: Giovanni Urga

Programme & Abstracts

Friday, 3th December 2010

08.30 - 09.15 Registration and Refreshments

09.15 Conference Opening

Session 1: Large Models
Chair: Giovanni Urga (Cass Business School, London, UK)

09:15-09:55
"Refitted Cross-Validation in Ultra High Dimensional Regression".
Jianqing Fan (Princeton University, USA), Shaojun Guo (Institute of Applied Mathematics, Academia Sinica, Taiwan), Ning Hao (Princeton University, USA).

Abstract: Variance estimation is a fundamental problem in statistical modeling. In ultrahigh dimensional linear regressions where the dimensionality is much larger than sample size, traditional variance estimation techniques are not applicable. Recent advances on variable selection in ultrahigh dimensional linear regressions make this problem more accessible. One of the major problems in ultrahigh dimensional regression is the high spurious correlation between the unobserved realized noise and some of the predictors. As a result, the realized noises are actually predicted when extra irrelevant variables are selected, leading to serious underestimate of the noise level. In this paper, we propose a two-stage refitted procedure via a data splitting technique, called refitted cross-validation (RCV), to attenuate the influence of irrelevant variables with high spurious correlations. Our asymptotic results show that the resulting procedure performs as well as the oracle estimator, which knows in advance the mean regression function. The simulation studies lend further support to our theoretical claims. The naive two-stage estimator which fits the selected variables in the first stage and the plug-in one stage estimators using LASSO and SCAD are also studied and compared. Their performances can be improved by the proposed RCV method. The methods are applied

to assess the forecasting errors of home price indices in the core based statistical areas in the US.

09.55-10.35
"Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain"
Alexandre Belloni, Daniel Chen, Victor Chernozhukov (MIT, USA), Christian Hansen

Abstract We develop results for the use of LASSO and Post-LASSO methods to form first-stage predictions and estimate optimal instruments in linear instrumental variables (IV) models with many instruments, p, that apply even when p is much larger than the sample size, n. We rigorously develop asymptotic distribution and inference theory for the resulting IV estimators and provide conditions under which these estimators are asymptotically oracle-efficient. In simulation experiments, the LASSO-based IV estimator with a data-driven penalty performs well compared to recently advocated many-instrument-robust procedures. In an empirical example dealing with the effect of judicial eminent domain decisions on economic outcomes, the LASSO-based IV estimator substantially reduces estimated standard errors allowing one to draw much more precise conclusions about the economic effects of these decisions. Optimal instruments are conditional expectations; and in developing the IV results, we also establish a series of new results for LASSO and Post-LASSO estimators of non-parametric conditional expectation functions which are of independent theoretical and practical interest. Specifically, we develop the asymptotic theory for these estimators that allows for non-Gaussian, heteroscedastic disturbances, which is important for econometric applications. By innovatively using moderate deviation theory for self-normalized sums, we provide convergence rates for these estimators that are as sharp as in the homoscedastic Gaussian case under the weak condition that $\log p = o(n^{1/3})$. Moreover, as a practical innovation, we provide a fully data-driven method for choosing the user-specified penalty that must be provided in obtaining LASSO and Post-LASSO estimates and establish its asymptotic validity under non-Gaussian, heteroscedastic disturbances.

10.35-11.00 Refreshments

Session 2: VAR Models

Chair: Richard Payne (Cass Business School, London, UK)

11.00-11.40.
"Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model"
Søren Johansen (University of Copenhagen and CREATES, Aarhus University, Denmark), Morten Ørregaard Nielsen (Queen's University, Canada, and CREATES, Aarhus University, Denmark)

Abstract: We consider model based inference in a fractionally cointegrated (or cofractional) vector autoregressive model based on the conditional Gaussian likelihood. The model allows the process X{t} to be fractional of order d and cofractional of order d-b; that is, there exist vectors β for which β X{t} is fractional of order d-b. The

parameters d and b satisfy either d≥b≥1/2, d=b≥1/2, or d=d0 ≥b≥1/2. Our main technical contribution is the proof of consistency of the maximum likelihood estimators on the set 1/2≤b≤d≤d1 fo r any d≥d0 . To this end, we consider the conditional likelihood as a stochastic process in the parameters, and prove that it converges in distribution when errors are i.i.d. with suitable moment conditions and initial values are bounded. We then prove that the estimator of β is asymptotically mixed Gaussian and estimators of the remaining parameters are asymptotically Gaussian. We also find the asymptotic distribution of the likelihood ratio test for cointegration rank, which is a functional of fractional Brownian motion of type II.

11.40-12.20

"Classical Time-Varying FAVAR Models - Estimation, Forecasting and Structural Analysis"

Sandra Eickmeier (European Central Bank and Deutsche Bundesbank, Germany), Wolfgang Lemke (Deutsche Bundesbank, Germany), and Massimiliano Marcellino (European University Institute, Italy and Bocconi University, Italy).

Abstract. We propose a classical, Kalman filter-based, approach to analyze factor-augmented vector autoregressive (FAVAR) models with time variation in the factor loadings, in the factor dynamics, and in their conditional variance-covariance matrix. When the time-varying FAVAR is estimated using a large dataset of US variables, the results indicate minor changes in the factor dynamics and more marked variation in their volatility and their impact on key macroeconomic variables. Forecasts from the time varying FAVAR are substantially more accurate than those from a constant parameter FAVAR, especially for several inflation, financial and monetary variables, also over the more recent period. Finally, when the time-varying FAVAR is used to assess how monetary transmission to the economy has changed, we find substantial time-variation in the volatility of monetary policy shocks and their transmission to real activity and inflation.

12.20-13.30 Lunch

Session 3: Factor Models and Panels

Chair: Lorenzo Trapani (Cass Business School, London, UK)

13.30-14.10

"Aggregation in Large Dynamic Panels"

Hashem M. Pesaran (Cambridge University and CIMF, UK, and USC, USA), Alexander Chudik (European Central Bank, Germany and CIMF, UK)

Abstract: This paper reviews Granger's contribution to the literature on aggregation and extends it by considering the case of large linear dynamic panels, where various interconnections between individual units are allowed. We notably permit distributed lags of all micro units to enter the individual micro relations, relax the independence assumption of the micro-distributed lag coefficients, and allow a general pattern of cross section dependence of micro innovations, which can be either strong or weak. Using the Pesaran's (2003) forecasting approach to derive the optimal aggregate model, our paper derives the optimal aggregate model of a factor-augmented VAR in N cross section units. The paper also discusses the aggregation error in this set-up, identify some of the distributional features of micro-parameters from aggregate relations, and proves Granger's (1980) conjecture about long memory properties of aggregate variables from a large scale dynamic econometric model. Monte Carlo experiments assess how these aggregate functions perform in small samples.

14.10-14.50

"Selecting the Correct Number of Factors in Approximate Factor Models: The Large Panel Case with Bridge Estimators"

Mehmet Caner (North Carolina State University, USA).

Abstract. This paper proposes Bridge estimators to correctly select the number of factors in an approximate factor model when there are latent factors. This contributes to both econometrics and statistics literature. Instead of the information based criterion penalty in econometrics, we propose a penalty based on factor loadings. So this extends the Bridge estimators in least squares context to factor models. This is a new approach in factor models, we show that oracle property of Bridge estimators is preserved, and hence we can select the correct number of factors through high dimension reduction. Simulations show that our technique can do a better job than information based criterion on both autocorrelated and cross section correlated data. A simple example about US macro factors in the last 25 years is supplied in the paper as well.

14.50-15.30

"Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data".

Marta Bańbura (European Central Bank, Germany) and Michele Modugno (European Central Bank, Germany)

Abstract. In this paper, we propose a methodology to estimate a dynamic factor model on data sets with an arbitrary pattern of missing data. We modify the Expectation Maximisation (EM) algorithm as proposed for a dynamic factor model by Watson and Engle (1983) to the case with general pattern of missing data. We also extend the model to the case with serially correlated idiosyncratic component. The framework allows to handle efficiently and in an automatic manner sets of indicators characterized by

different publication delays, frequencies and sample lengths. This can be relevant e.g. for young economies for which many indicators are compiled only since recently. We also show how to extract a model based news from a statistical data release within our framework and we derive the relationship between the news and the resulting forecast revision. This can be used for interpretation in e.g. nowcasting applications as it allows to determine the sign and size of a news as well as its contribution to the revision, in particular in case of simultaneous data releases. We evaluate the methodology in a Monte Carlo experiment and we apply it to nowcasting and backdating of euro area GDP.

15.30-16.10
"Factor Based Identification-Robust Inference in IV"
George Kapetanios (Queen Mary, University of London, UK), Lynda Khalaf (Carleton University, Canada), M. Marcellino (EUI and Bocconi University, Italy)

Abstract: Weak-instruments robust methods raise important size/power trade-offs resulting from omitted instruments biases. The popular Anderson-Rubin test has the right size when the underlying first-stage model (that is, the model linking the structural equations right-hand-side endogenous variables to available instruments) is closed or is incomplete. Alternative methods are available that may outperform this statistic assuming a closed first-stage specification (that is, assuming that all instruments are accounted for). We show that information-reduction methods provide a useful and practical solution to this problem. Formally, we propose factor-based modifications to three popular weak-instruments-robust statistics, and illustrate their validity asymptotically and in finite samples. Results are derived using asymptotic settings that are commonly used in both the factor-analysis and weak-instruments literatures. For the Anderson-Rubin statistic, we also provide analytical finite sample results under usual assumptions. An illustrative Monte Carlo study reveals the following. (1) Our factor-based corrections circumvent the size problems resulting from instrument omissions and improve the power of the Anderson-Rubin statistic. (2) Once corrected through factor reduction, all considered statistics perform equally well. Results suggest that factor-reduction holds promise as a unifying solution to the many instruments problem.

16.10-16.30 Refreshments

Session 4: Testing and Model Selection

Chair: Jennifer Castle (University of Oxford, UK)

16.30-17.10
"Testing Rates: Nonparametric Nonstationarity Tests"
Federico Bandi (The Johns Hopkins Carey Business School, Baltimore, USA) and Valentina Corradi (Warwick University, UK).

Abstract: We propose nonstationarity tests which exploit the different divergence rates of the occupation measures of a (possibly nonlinear) process under the null of nonstationarity (stationarity) versus the alternative of stationarity (nonstationarity). We consider both discrete- and continuous-time series. The discrete-time case covers Harris recurrent Markov chains (as in, e.g., Karlsen and Tjostheim, 2001, Guerre, 2004, and Schienle, 2008) and (near-)integrated processes (as in, e.g., Wang and Phillips, 2009a,

2009b). The continuous-time case focuses on diffusion processes. In addition to their simplicity and reliance on tabulated critical values, we show that the size and power properties of the proposed tests are satisfactory and generally superior to those of traditional approaches to nonstationarity testing.

17.10-17.50
"Model Selection in Equations with Many `Small' Effects"
Jennifer L. Castle (Oxford University, UK) and David F. Hendry (Oxford University, UK)

Abstract: Model selection from a general unrestricted model (GUM) can potentially confront many small relevant effects that may not be retained by selecting individual variables, but can be partially captured by latent factors, together with important individual determinants, while eliminating many irrelevant variables. We exploit the ability of automatic model selection procedures to handle perfect collinearity and more candidate variables, N, than observations, T to investigate including both the candidate regressors and all their principal components. This allows our approach to embed many salient explanatory variables, their lags, functional form transformations, and multiple location shifts together with `factor' structures, without being concerned that the resulting N>T. The factors might enter as `genuine' explanatory variables, or as approximations to the many small relevant effects that would not be retained individually. Potential applications to forecasting are noted.

19.00 Conference Dinner at "Cantina del Ponte" (http://www.cantina.co.uk/)

Saturday, 4th December 2010

08.30-09.00 Refreshments

Session 5: Large Cross-Sections

Chair: Lynda Khalaf (Carleton University, Canada)

09.00-09.40

"Summary Statistic for Cross-sectional Dependence in Large Datasets"
Natalia Bailey (Queen Mary University of London, UK) and George Kapetanios (Queen Mary University of London, UK)

Consistent estimation of the variance-covariance matrix of a large dimensional dataset is strongly affected, among other things, by the underlying structure of the data in question. In this paper we focus our attention on the identification of the strength of such a structure, when it exists. We claim that all information needed for this purpose is contained in the column-sum norm of the variance-covariance matrix. This approaches infinity at rate Na, $0 < a \leq 1$. The strength of structure can then be determined by the value of a; a low value ($a \rightarrow 0$) is indicative of weak interdependence between the elements of the dataset while a large value ($a \rightarrow 1$) corresponds to a strong inter-relationship. On this basis, we construct a summary statistic as a means of quantifying this parameter a. Our estimator is consistent for relatively small N=T ratios. For estimation purposes we favour non-linear optimisation. The accuracy of our statistic is checked by use of the non-parametric bootstrap approach. An application to S&P 500 stock market data, in the context of building our Sectoral Concentration Indicator (SCI), is also incorporated.

09.40-10.20
"Testing Linear Factor Pricing Models with Large Cross-Sections: a Distribution-Free Approach"
Sermin Gungor (Bank of Canada, Ottawa, Canada) and Richard Luger (Georgia State University, Atlanta, USA)

Abstract: We develop a finite-sample procedure to test the beta-pricing representation of linear factor pricing models that is applicable even if the number of test assets is greater than the length of the time series. Further, we make no parametric assumptions about the distribution of the disturbances in the factor model. Our framework leaves open the possibility of unknown forms of time-varying non-normalities, heteroskedasticity, and even outliers in the asset returns. The power of the proposed test procedure increases as the time series lengthens and/or the cross-section becomes larger. Finally, we illustrate the new procedure by testing the well-known Fama-French factor model over 5-year subsamples of monthly returns on 100 U.S. equity portfolios formed on size and book-to-market.

10.20-11.00
"Time-Varying Risk Premium in Large Cross-Sectional Equity Datasets"
Patrick Gagliardini (Swiss Finance Institute, University of Lugano, CH) Elisa Ossola (Swiss Finance Institute, University of Lugano, CH) and Olivier Scaillet (HEC Universite de Geneve and Swiss Finance Institute, CH)

Abstract. We develop an econometric methodology to infer the path of the risk premium from large unbalanced panel of individual stock returns. We estimate the time-varying risk premium implied by conditional linear asset pricing models, and show consistency and asymptotic normality under increasing cross-sectional and time series dimensions. We address consistent estimation of the asymptotic variance, and testing for the asset pricing restrictions. The empirical illustration on 15000 US stock returns from January 1960 to January 2010 shows that the risk premium is counter-cyclical. It is also larger and more volatile in recent years.

11.00-11.15 Refreshments

Session 6: Breaks/Changepoints

Chair: Elena Andreou (University of Cyprus, Nicosia, Cyprus)

11.15-11.55

"Detecting Structural Breaks in Large Dimensional Factor Models"

Liang Chen (Dept. of Economics, Universidad Carlos III de Madrid, Spain), Juan J. Dolado (Dept. of Economics, Universidad Carlos III de Madrid, Spain), and Jesús Gonzalo (Dept. of Economics, Universidad Carlos III de Madrid, Spain)

Abstract: Constant factor loadings is a standard assumption in the analysis of large dimensional factor models (LFM). Yet, this assumption may be restrictive when there are structural breaks in the economy, as e.g.,during the ongoing great recession. In this paper we propose a two-step test procedure to detect big structural breaks in factor loadings of LFM. Following Stock and Watson (2007), we define small and big breaks by specifying the statistical properties of the break sizes. Whereas under small breaks the number of factors can be consistently estimated by PC using the information criteria proposed by Bai and Ng (2002), we show that they will be over-estimated in the presence of big breaks. Our two-step procedure works as follows: (i) first, ˆr factors are estimated by PC and, (i) next, a sup type test is performed on the regression of one of the estimated factors on the remaining ˆr − 1 factors. If a break is detected in such a regression, we conclude that there exists a big break in the factor loadings at a common date for all variables ; otherwise we cannot reject small or no breaks at all and therefore the number of factors and the factor space can be consistently estimated. We also extend this approach to the case of unknown and/or variable-specific break dates, as well as discuss the empirically relevant case where the chosen number of factors differs in finite samples according to different criteria used. Finally, a large Monte Carlo study provides evidence that the proposed procedure behaves satisfactorily in terms of size and power while an empirical application illustrates that it is very easy to implement with standard econometric software.

11.55-12.35

"Testing for Instability in Covariance Structures"

Chihwa Kao (Syracuse University, USA), Lorenzo Trapani (Cass Business School, UK) and Giovanni Urga (Cass Business School, UK and Bergamo University, Italy)

Abstract: This paper proposes a test for the stability over time of the covariance matrix of multivariate time series of finite dimension. Our analysis is based on Principal Component Analysis (PCA), and thus it allows to ascertain whether changes in the covariance structure are due to instability in the eigenvalues, in the eigenvectors, or in the

principal components of the covariance matrix which are a combination of the two. The standard assumptions required to derive the asymptotics (time independence and normality) are relaxed, allowing for the data to be Near Epoch Dependent. Our test statistic is based on taking the supremum of a sequence of CUSUM-type statistics; we prove a strong invariance principle for the partial sample estimators of the covariance matrix, and thus we are able to calculate the supremum of the CUSUM-type statistics without having to trim at the beginning/end of the sample, enhancing the power of our test. The power properties of the test versus local alternatives and versus alternatives close to the beginning/end of sample are investigated theoretically and by simulation, and the testing procedure is validated through an application to the term structure of interest rates.

12.35-13.30 Lunch

Session 7: Dynamic Heteroskedasticity and Forecasting

Chair: Olivier Scaillet (HEC Universite de Geneve and Swiss Finance Institute, CH)

13.30-14.10

"Dynamic mean and volatility factors in large panels of high-frequency financial assets"

Elena Andreou (University of Cyprus, Nicosia, Cyprus), Eric Ghysels (University of North Carolina, Chapel Hill, USA)

Abstract. The objective of this paper is twofold. First we show how to extract high frequency financial factors that relate to the common component of both dynamic mean and volatility from a large panel of financial assets. We show how to extract these dynamic factors from a large panel of around one thousand daily financial assets over the last decade, spanning almost the entire cross-section of US assets from five major asset classes namely equities, commodities, corporate bonds, government securities and foreign exchange. In addition, we extract the corresponding mean and variance factors from these five homogeneous classes of assets such as, for instance, the equity mean returns and their volatility factors, corporate yields/spreads factors and corporate risk factors, among others. The existing literature neither addresses the estimation of dynamic factors from multiple high frequency financial asset classes nor the extraction of the common volatility factor. Instead it mainly deals with extracting factors from a smaller panel of financial assets at a lower (monthly) frequency (Ludvigson and Ng, 2007) and in particular the equity class (Da and Schaumburg, 2009), to address empirical asset pricing questions and forecasting. In addition the existing literature does not estimate the common volatility factor directly from a dynamic factor model but instead focuses on estimating the volatility of the common factor(s) as a second-step estimator that of alternative volatility estimators of the extracted common factor (Ludvigson and Ng, 2007, Anderson and Vahid, 2007). Notable exemptions to this are the dynamic factor multivariate stochastic volatility model in Han (2010) which is however applied to a smaller cross section of 36 stocks of of low frequency returns using Bayesian methods. Our approach addresses the following challenges: (i) We estimate a dynamic factor model for a large panel with time-varying loadings using the method of principal component and extract simultaneously the common dynamic mean and volatility factors extending the results in Stock and Watson (2002). (ii) We address the filtering problem that arises from constructing a panel of volatilities of individual assets in order to extract the relevant volatility factors. This relates to the optimality of the filtering and the asymptotic

efficiency of factors in the spirit of Sims (1974). (iii) Our common factors entail information about the entire cross-section or entire class and are at the same time less noisy estimators of volatility compared to the volatility of individual major financial assets and indices. Moreover, the homogeneity entailed by the factors extracted from each class allows us to identify the decomposition and driving forces of each daily factor. The second objective deals with the modeling aspects of the extracted daily financial factors. We show how these daily financial factors can be used in a parametric and semi-parametric model to study the dynamic relationship between daily mean factors and volatility factors. Using our factor approach we address two of the major challenges of multivariate dynamic volatility models, namely both the dependence and dimensionality reduction which thereby imply a reduction in the model's parameter space

14.10-14.50.

"Bayesian Adaptive quasi-Hamiltonian Monte Carlo with an Application to High-Dimensional BEKK GARCH Models"

Martin Burda (Department of Economics, University of Toronto, Canada) and John Maheu (Department of Economics, University of Toronto, Canada)

Abstract: Hamiltonian Monte Carlo (HMC) is a recent statistical sampling procedure that builds distant proposal draws in a sequence of steps following the Hamiltonian dynamics of the underlying parameter space. It is particularly suitable for drawing from high-dimensional joint posterior distributions where conditional sampling is intractable or results in strongly autocorrelated mixing. In this paper we propose a local adaptation of HMC whereby the proposal sequence is tailored to the local evolution of the parameter space. Such adaptability leads to higher acceptance rates and improved mixing of the resulting Markov chain compared to HMC. The benefits of adaptability become more pronounced with the degree of irregularity of the underlying model likelihood. In contrast to previous efforts in local adaptability of HMC, our procedure is volume-preserving and reversible, satisfying detailed balance needed for an invariant posterior distribution. We apply the procedure to a high-dimensional BEKK GARCH model and show that the model can be sampled well in higher dimensions than previously considered practical. We implemented the BEKK GARCH for 7 variables corresponding to 126 parameter dimensions and show that the resulting Markov chains of draws feature favourable properties.

14.50-15.30

"Macroeconomic Forecasting Using Kernel Ridge Regression"

Peter Exterkate (Econometric Institute, Erasmus University Rotterdam, Netherlands), Patrick J.F. Groenen, (Econometric Institute, Erasmus University Rotterdam, Netherlands) and Christiaan Heij (Econometric Institute, Erasmus University Rotterdam, The Netherlands).

Abstract. Nonlinear and high-dimensional relations can be estimated using kernels, which form a popular and well-established tool in the machine learning community. They allow modelling and estimating nonlinear relations by mapping the observed predictor variables nonlinearly into a high-dimensional space, where prediction takes place. To avoid overfitting a penalty term is added. This form of kernel ridge regression has been found to perform very well in many different contexts. Despite these attractive features, this method is not commonly used in time-series applications. In this paper, we study the forecasting performance of kernel ridge regression. We present Monte Carlo evidence

for the adequacy of this method for prediction purposes. Empirical results confirm that kernel regression produces more accurate forecasts for two key U.S. macroeconomic variables than traditional methods do, especially when forecasting at longer horizons.

15.30-16.10

"Conditional and Unconditional Forecasting with Large Vector Autoregressive Models"

Marta Banbura (European Central Bank, germany), Domenico Giannone (Université libre de Bruxelles, ECARES, Belgium, and CEPR), and Michele Lenza (European Central Bank, Germany)

Abstract: We construct a large Vector Autoregressive model for the Euro Area. The model includes the key macroeconomic , financial, credit and monetary variables and nests the specifications that are used by the main microfounded models adopted for the European Central Bank and Eurosystem macroeconomic projections, including the New Area Wide Model of Christoffel, Coenen, and Warne (2008) and the model by Christiano, Motto, and Rostagno (2010). In order to overcome the course of dimensionality, the model is estimated using Bayesian shrinkage as suggested by Banbura, Giannone and Reichlin (2010). The degree of shrinkage treated as an unknown parameter following the approach recently proposed by Giannone, Lenza and Primiceri (2010). We develop an efficient algorithm, based on Kalman filtering, to compute conditional density forecasts for large dynamic systems. We use the model to assess whether macroeconomic fundamentals can explain the dynamics of financial variables including stock prices and exchange rates.

16.10 End of Conference

Giovanni Urga, 29 November 2010.