HKUST IAS-SBM Joint Workshop on Financial Econometrics in the Big Data Era

May 21-22, 2024

Co-Sponsors:

HKUST JOCKEY CLUB INSTITUTE FOR ADVANCED STUDY

HKUST BUSINESS SCHOOL 香港科大商學院 WORLD CLASS IN ASIA

Invited Speakers:

- **Torben G. ANDERSEN (Northwestern University)**
- **Jinyuan CHANG (Southwestern University of Finance and Economics)**
- **Zhanhui CHEN (HKUST)**
- **Carsten H. CHONG (HKUST)**
- **Yi DING (University of Macau)**
- **Jean JACOD (Sorbonne University)**
- **Jia LI (Singapore Management University)**
- **Yingying LI (HKUST)**
- **Nour MEDDAHI (Toulouse School of Economics)**
- **Mathieu ROSENBAUM (École Polytechnique)**
- **Olivier SCAILLET (University of Geneva and Swiss Finance Institute)**
- **Liangjun SU (Tsinghua University)**
- **Viktor TODOROV (Northwestern University)**
- **Xiaoliang WANG (HKUST)**
- **Dacheng XIU (University of Chicago)**
- **Jun YU (University of Macau)**

May 21, 2024 (Tuesday)

Time	Event
09:00 – 09:20	Registration
09:20 – 09:30	Opening Remarks
Session 1	Chair: Xinghua ZHENG (HKUST)
09:30 – 10:00	**The Granular Origins of Tail Risk** Torben G. ANDERSEN, Yi DING, Viktor TODOROV
10:00 – 10:30	**Co-jump Networks, Mixed Membership and Beyond** Yi DING, Yingying LI, Guoli LIU, Changlei LYU, Xinghua ZHENG
10:30 – 11:00	Coffee Break
Session 2	Chair: Gavin FENG (CityU)
11:00 – 11:30	**Can Machines Learn Weak Signals?** Zhouyu SHEN, Dacheng XIU
11:30 – 12:00	**Learning the Stochastic Discount Factor** Zhanhui CHEN, Yi DING, Yingying LI, Xinghua ZHENG
12:00 – 13:30	Lunch
Session 3	Chair: Carsten CHONG (HKUST)
13:30 – 14:00	**High Frequency Returns Sign-Based Robust Inference** Jean JACOD, Nour MEDDAHI
14:00 – 14:30	**A General Test for Functional Inequalities** Jia LI, Zhipeng LIAO, Wenyu ZHOU
14:30 – 15:00	Coffee Break
Session 4	Chair: Merrick LI (CUHK)
15:00 – 15:30	**Testing for an Explosive Bubble Using High-frequency Volatility** H. Peter BOSWIJK, Jun YU, Yang ZU
15:30 – 16:00	**Non-linear Time Series Models and Machine Learning** Christian GOURIÉROUX, Nour MEDDAHI, Serge NYAWA
16:15	Excursion & Dinner

May 22, 2024 (Wednesday)

Time	Event
Session 5	Chair: Weichen WANG (HKU)
09:00 – 09:30	**Latent Factor Analysis in Short Panels** Alan-Philippe FORTIN, Patrick GAGLIARDINI, Olivier SCAILLET
09:30 – 10:00	**High-dimensional Covariance Matrix Estimation under Elliptical Factor Model with 2 + εth Moment** Yi DING, Xinghua ZHENG
10:00 – 10:30	**Semiparametric Conditional Factor Models: Estimation and Inference** Qihui CHEN, Nikolai ROUSSANOV, Xiaoliang WANG
10:30 – 11:00	Coffee Break
Session 6	Chair: Zhentao SHI (CUHK)
11:00 – 11:30	**High Dimensional Conditional Factor Model** Shang GAO, Zhonghao FU, Liangjun SU, Xia WANG
11:30 – 12:00	**Optimal Covariance Matrix Estimation for High-dimensional Noise in High-frequency Data** Jinyuan CHANG, Qiao HU, Cheng LIU, Cheng Yong TANG
12:00 – 13:30	Lunch
Session 7	Chair: Dachuan CHEN (Nankai University)
13:30 – 14:00	**Asymptotic Expansions for High-frequency Option Data** Carsten H. CHONG, Viktor TODOROV
14:00 – 14:30	**A Characterisation of Cross-impact Kernels** Mathieu ROSENBAUM, Mehdi TOMAS
14:30 – 15:00	**The Fine Structure of Volatility Dynamics** Carsten H. CHONG, Viktor TODOROV
15:00 – 16:00	Adjourn

The Granular Origins of Tail Risk

Torben Andersen
Kellogg School of Management, Northwestern University
(with Yi Ding and Viktor Todorov)

We study the tail risk in the cross-section of asset prices at high frequencies. We show that the cross-sectional tail behavior of asset returns depends on whether the price increment contains a systematic jump event or not. In case of systematic jumps, the cross-sectional asset return tail behavior is determined by the assets' exposures to the systematic event, while if the interval contains no systematic jump, it is determined by the tails of the idiosyncratic jumps. We develop an estimator for the tail shape of the cross-sectional asset return distribution with distinct asymptotic properties, depending on whether the interval contains a systematic jump or not. We show empirically that shocks to the tail shape parameters of the cross-sectional asset return distribution are source of priced risk. The price of this tail risk depends on its source: fat idiosyncratic tails are liked by investors while fat tails in assets' exposures to systematic jumps are disliked by them.

Co-jump Networks, Mixed Membership And Beyond

Yingying Li
School of Business and Management, HKUST
(with Yi Ding, Guoli Liu, Changlei Lyu and Xinghua Zheng)

This talk will be based on two recent works about stock co-jump networks. We propose a Degree-Corrected Block Model with Dependent Multivariate Poisson edges (DCBM-DMP) to study stock co-jump dependency. Both pure memberships and mixed memberships are studied. We provide algorithms and statistical properties to support the estimation of the community structure. Such communities exhibit different features than GICS. We further demonstrate the economic significance of these networks.

Can Machines Learn Weak Signals?

Dacheng Xiu
Booth School of Business, University of Chicago
(with Zhouyu Shen)

In high-dimensional regression scenarios with low signal-to-noise ratios, we assess the predictive performance of several machine learning algorithms. Theoretical insights show Ridge regression's superiority in exploiting weak signals, surpassing a zero benchmark. In contrast, Lasso fails to exceed this baseline, indicating its learning limitations. Simulations reveal that Random Forest generally outperforms Gradient Boosted Regression Trees when signals are weak. Moreover, Neural Networks with ℓ2-regularization excel in capturing nonlinear functions of weak signals. Our empirical analysis across six economic datasets suggests that the weakness of signals, not neces-sarily the absence of sparsity, may be Lasso's major limitation in economic predictions.

Learning the Stochastic Discount Factor

Zhanhui Chen
School of Business and Management, HKUST
(with Yi Ding, Yingying Li and Xinghua Zheng)

We develop a statistical framework to learn the high-dimensional stochastic discount factor (SDF) from a large set of characteristic-based portfolios. Specifically, we build on the maximum-Sharpe ratio estimated and sparse regression method proposed in Ao, Li and Zheng (RFS, 2019) to construct the SDF portfolio, and develop a statistical inference theory to test the SDF loadings. Applying our approach to 194 characteristic-based portfolios, we find that the SDF constructed by about 20 of them performs well in achieving a high Sharpe ratio and explaining the cross-section of expected returns of various portfolios.

High Frequency Returns Sign-Based Robust Inference

Jean Jacod
Department of Mathematics, Sorbonne University
(with Nour Meddahi)

The paper derives the limit and the infill asymptotic distribution of the sum of positive returns in a given period of time. The framework is multivariate and quite general; it allows for the presence of leverage effects and jumps. In a second step, the results are used to estimate the drifts of the continuous part of the processes.

A General Test for Functional Inequalities

Jia Li
School of Economics, Singapore Management University
(with Zhipeng Liao and Wenyu Zhou)

This paper develops a nonparametric test for general functional inequalities that include conditional moment inequalities as a special case. It is shown that the test controls size uniformly over a large class of distributions for observed data, importantly allowing for general forms of time series dependence. New results on uniform growing dimensional Gaussian coupling for general mixingale processes are developed for this purpose, which readily accommodate most applications in economics and finance. The proposed method is applied in a portfolio evaluation context to test for “all-weather” portfolios with uniformly superior conditional Sharpe ratio functions.

Testing for An Explosive Bubble Using High-Frequency Volatility

Jun Yu
Faculty of Business Administration, University of Macau
(with H. Peter Boswijk and Yang Zu)

Based on a continuous-time stochastic volatility model with a linear drift, we develop a test for explosive behavior in financial asset prices at a low frequency when prices are sampled at a higher frequency. The test exploits the volatility information in the high-frequency data. The method consists of devolatizing log-asset price increments with realized volatility measures and performing a supremum-type recursive Dickey-Fuller test on the devolatized sample. The proposed test has a nuisance-parameter-free asymptotic distribution and is easy to implement. We study the size and power properties of the test in Monte Carlo simulations. A real-time date-stamping strategy based on the devolatized sample is proposed for the origination and conclusion dates of the explosive regime. Conditions under which the real-time date-stamping strategy is consistent are established. The test and the date-stamping strategy are applied to study explosive behavior in cryptocurrency and stock markets.

Non-Linear Time Series Models and Machine Learning

Nour Meddahi
Toulouse School of Economics
(with Christian Gouriéroux and Serge Nyawa)

We recently observed the irruption and rapid development of machine learning (ML) methods in econometrics and statistics, especially for forecasting purposes. For instance, ML methods have been recently used in several studies for forecasting economic and financial variables like assets returns (Gu, Kelly, and Xiu, 2020), stock and bond returns (Bianchi, Buchner, and Tamoni, 2021), volatility (Patton and Simsek, 2023), inflation (Medeiros, Vasconcelos, Veiga, and Zilberman, 2021), and macroeconomic variables (Goulet Coulombe, 2021; Goulet Coulombe, Leroux, Stevanovic, and Surprenant, 2022). An important common conclusion of these studies is that ML methods are successful in forecasting because they account for non-linearities that popular time series models do not. The first goal of the paper is to highlight the non-linearities that ML methods capture and connect them with traditional non-linear time series modeling. The second goal of the paper is to modify some traditional non-linear time series model by including insights from the ML literature. Applications to the Euro-US dollar exchange rate and the SPYDER index are provided.

Latent Factor Analysis in Short Panels

Olivier Scaillet
University of Geneva and Swiss Finance Institute
(with Alain-Philippe Fortin and Patrick Gagliardini)

We develop inferential tools for latent factor analysis in short panels. The pseudo maximum likelihood setting under a large cross-sectional dimension *n* and a fixed time series dimension *T* relies on a diagonal $T \times T$ covariance matrix of the errors without imposing sphericity nor Gaussianity. We outline the asymptotic distributions of the latent factor and error covariance estimates as well as of an asymptotically uniformly most powerful invariant (AUMPI) test for the number of factors based on the likelihood ratio statistic. We derive the AUMPI characterization from inequalities ensuring the monotone likelihood ratio property for positive definite quadratic forms in normal variables. An empirical application to a large panel of monthly U.S. stock returns separates month after month systematic and idiosyncratic risks in short subperiods of bear vs. bull market based on the selected number of factors. We observe an uptrend in the paths of total and idiosyncratic volatilities while the systematic risk explains a large part of the cross-sectional total variance in bear markets but is not driven by a single factor. Rank tests show that observed factors struggle spanning latent factors with a discrepancy between the dimensions of the two-factor spaces decreasing over time.

High-dimensional Covariance Matrix Estimation under Elliptical Factor Model with 2 + εth Moment

Yi Ding
Faculty of Business Administration, University of Macau
(with Xinghua Zheng)

We study the estimation of high-dimensional covariance matrices under elliptical factor models with 2 + εth moment. For such heavy-tailed data, robust estimators like the Huber-type estimator in Fan et al. (2018) can not achieve sub-Gaussian optimal convergence rates. We develop an idiosyncratic-projected self-normalization (IPSN) method to remove the effect of heavy-tailed scale parameter, and propose a robust pilot estimator for the scatter matrix and show that the estimator enjoys the optimal sub-Gaussian rate. We further develop a consistent generic POET estimator of the covariance matrix and show that it achieves a faster convergence rate than the generic POET estimator in Fan et al. (2018).

Semiparametric Conditional Factor Models: Estimation and Inference

Xiaoliang Wang
School of Business and Management, HKUST
(with Qihui Chen and Nikolai Roussanov)

We develop a simple and tractable sieve estimation of semiparametric conditional factor models with latent factors. We establish large-N asymptotic properties of the estimators without requiring large T. We also develop a simple bootstrap procedure for conducting inference about the conditional pricing errors as well as the shapes of the factor loading functions. We employ this method on the large cross-sectional individual stocks, corporate bonds and options. A common risk factor structure prominently emerges across asset classes. Several common factors explain a substantial amount of time-series variation of individual asset returns across all three asset classes, and have sizable Sharpe ratios. Several of our factors are highly correlated with some of asset-class-specific factors as well as macroeconomic and financial variables. However, a small set of common factors does not fully capture the cross-section of average returns. A mean-variance efficient portfolio that utilizes asset characteristics achieves a high Sharpe ratio as different asset classes hedge each other's exposures to the common factors.

High Dimensional Conditional Factor Model

Liangjun Su
School of Economics and Management, Tsinghua University
(with Shang Gao, Zhonghao Fu, and Xia Wang)

This paper studies the estimation and variable selection for conditional factor models with high dimensional instruments. We employ a nuclear-norm regularized regression and adaptive group LASSO regression to consistently estimate the row-sparse coefficients and latent factors, and to select the relevant instrument characteristics. We establish the estimation consistency of factor estimators, selection consistency, and oracle property of the row-sparse coefficient estimators. Additionally, a singular value thresholding procedure is applied to determine the correct number of factors with probability approaching one. Simulations demonstrate that our estimators perform well in finite samples, both in terms of estimation accuracy and selection consistency. In an empirical application, we employ our estimation framework to predict asset returns, highlighting its effectiveness in financial applications.

Optimal Covariance Matrix Estimation for High-dimensional Noise in High-frequency Data

Jinyuan Chang
Southwestern University of Finance and Economics
(with Qiao Hu, Cheng Liu and Cheng Yong Tang)

We consider high-dimensional measurement errors with high-frequency data. Our objective is on recovering the high-dimensional cross-sectional covariance matrix of the random errors with optimality. In this problem, not all components of the random vector are observed at the same time and the measurement errors are latent variables, leading to major challenges besides high data dimensionality. We propose a new covariance matrix estimator in this context with appropriate localization and thresholding, and then conduct a series of comprehensive theoretical investigations of the proposed estimator. By developing a new technical device integrating the high-frequency data feature with the conventional notion of alpha-mixing, our analysis successfully accommodates the challenging serial dependence in the measurement errors. Our theoretical analysis establishes the minimax optimal convergence rates associated with two commonly used loss functions; and we demonstrate with concrete cases when the proposed localized estimator with thresholding achieves the minimax optimal convergence rates. Considering that the variances and covariances can be small in reality, we conduct a second-order theoretical analysis that further disentangles the dominating bias in the estimator. A bias-corrected estimator is then proposed to ensure its practical finite sample performance. We also extensively analyze our estimator in the setting with jumps, and show that its performance is reasonably robust. We illustrate the promising empirical performance of the proposed estimator with extensive simulation studies and a real data analysis.

Asymptotic Expansions for High-frequency Option Data

Carsten Chong
School of Business and Management, HKUST
(with Viktor Todorov)

We derive a nonparametric higher-order asymptotic expansion for small- time changes of conditional characteristic functions of Itô semimartingale increments. The asymptotics setup is of joint type: both the length of the time interval of the increment of the underlying process and the time gap between evaluating the conditional characteristic function are shrinking. The spot semimartingale characteristics of the underlying process as well as their spot semimartingale characteristics appear as leading terms in the derived asymptotic expansions. The analysis applies to a general class of Itô semi- martingales that includes in particular Lévy-driven SDEs and time-changed Lévy processes. The asymptotic expansion results are subsequently used to construct a test for whether volatility jumps are of finite or infinite variation. In an application to high-frequency data of options written on the S&P 500 index, we find evidence for infinite variation volatility jumps.

A Characterisation of Cross-impact Kernels

Mathieu Rosenbaum
Centre de Mathématiques Appliquées, Ecole Polytechnique
(with Mehdi Tomas)

Trading a financial asset pushes its price as well as the prices of other assets, a phenomenon known as cross-impact. We consider a general class of kernel-based cross-impact models and investigate suitable parametrisations for trading purposes. We focus on kernels that guarantee that prices are martingales and anticipate future order flow (martingale admissible kernels) and those that ensure there is no possible price manipulation (no-statistical-arbitrage-admissible kernels). We determine the overlap between these two classes and provide formulas for calibration of cross-impact kernels on data. We illustrate our results using SP500 futures data.

The Fine Structure of Volatility Dynamics

Viktor Todorov
Kellogg School of Management, Northwestern University
(with Carsten H. Chong)

We develop a nonparametric test for deciding whether volatility of an asset follows a standard semimartingale process, with paths of finite quadratic variation, or a rough process with paths of infinite quadratic variation. The test utilizes the fact that volatility is rough if and only if volatility increments are negatively autocorrelated at high frequencies. It is based on the sample autocovariance of increments of spot volatility estimates computed from high-frequency asset return data. By showing a feasible CLT for this statistic under the null hypothesis of semimartingale volatility paths, we construct a test with fixed asymptotic size and an asymptotic power equal to one. The test is derived under very general conditions for the data-generating process. In particular, it is robust to jumps with arbitrary activity and to the presence of market microstructure noise. In an application of the test to SPY high-frequency data, we find evidence for rough volatility.