Invited Sessions V - 12 December 2024, 14:30 – 16:00

Session 11	**Recent Advances in Time Series Analysis** *Organizer/Chair: Zhou Zhou, University of Toronto*	E22-2002	12 Dec, 14:30 - 16:00

On threshold nonlinear time series with a buffer zone

Wai Keung Li, The Education University of Hong Kong

A time series with a buffer (hysteretic) zone was first introduced in the autoregressive form in Li et al. (2015) and Zhu et al. (2014). The time series is an extension of the classical nonlinear threshold time series by allowing the current time series structure to be unchanged if its immediate past lies within a buffered region. Since its introduction it has found applications in many fields including exchange rates and environmental modelling. It has also been extended in various directions covering, for example, smooth transition and non-stationarity structures. This talk will give an updated review on the developments of the buffered time series models including the most recent moving average version. This will be illustrated by applications to real data.

Sparse Matrix Estimation Based on Greedy Algorithms and Information Criteria

Ching-Kang Ing, Tsing Hua University

We address the challenge of estimating the covariance matrix of serially correlated vectors, wherein the dimension of the vector can be much larger than the sample size. We approach this estimation problem by reframing it as a variable selection challenge. This novel perspective can leverage the advantages of the orthogonal greedy algorithm (OGA) and the high-dimensional Akaike's information criterion (HDAIC) in matrix estimation. By demonstrating that our proposed estimate attains rate optimality under a sparsity condition more flexible than those in the existing literature, we substantially improve the convergence rate for thresholding-type estimators over a class of sparse matrices characterized by polynomial decay. If the covariance matrix exhibits a certain bandable structure, we introduce a banding/tapering estimate whose parameters are chosen using a novel two-index information criterion. We establish the rate optimality of this estimate under the operator norm, which, to the best of our knowledge, is the first optimal adaptive estimation result for banding-type estimators under the operator norm.

Statistical inference for weakly stationary time series

Yunyi Zhang, The Chinese University of Hong Kong, Shenzhen

The literature often adopts two types of stationarity assumptions in the analysis of time series, i.e., the weak stationarity, suggesting that the mean and the autocovariance function of a time series are time invariant; and strict stationarity, indicating that the marginal distributions of the time series are time invariant. While the strict stationarity assumption is vital from theoretical aspect, it is hard to verify in practice. On the other hand, the weak stationarity is relatively feasible to ensure and verify, as it only relies on the second--order structures of the time series. Concerning this, while sorts of weak stationarity assumptions are typically adopted in time series modeling, statisticians may want to avoid relying on strict stationarity assumptions during statistical inference.

This presentation focuses on the analysis of quadratic forms within a weakly, but not necessarily strictly stationary (vector) time series. In the context of scalar time series, it establishes the Gaussian approximation for quadratic forms of a short--range dependent weakly stationary scalar time series. Building upon this result, it derives the asymptotic distributions of the sample autocovariances, the sample autocorrelations, and the sample autoregressive coefficients. Transitioning to vector time series, this presentation tackles statistical inference within high--dimensional vector autoregressive models

with white noise innovations. Given the complicated covariance structures inherent in non-stationary time series, this presentation adopts the dependent wild bootstrap method to facilitate statistical inference. Numerical results verifies the consistency of the proposed theories and methods.
Strict stationarity is hard to ensure and verify for a real--life dataset. Therefore, our work should be able to assist statisticians in capturing the inherent non--stationarity of real--life time series.

Convex and Strong Gaussian Approximations for Non-Stationary Time Series of Diverging Dimensionality

Zhou Zhou, University of Toronto

The central limit theorem (CLT) and its extensions, such as the Berry-Esseen theorem, are among the most useful results in classic statistics. In high dimensions, the analogous results to the CLT are Gaussian approximation (GA) schemes on various collections of subsets of the multi-dimensional Euclidean space. In this talk, by considering two important collections of subsets: the convex sets and the Borel sets, we establish general GA theory for a very general class of high-dimensional non-stationary (HDNS) time series. Our approximation rates are nearly optimal with respect to both dimension and time series length. A block multiplier bootstrap procedure is theoretically verified for the implementation of our GA theory. We demonstrate by applications the use of the GA and bootstrap theory as a unified tool for a wide range of statistical inference problems of HDNS time series

Session 21	**Advances in High-Dimensional Modeling and Optimization** *Organizer/Chair: Zhaoxing Gao, Zhejiang University*	E22-2007	12 Dec, 14:30 - 16:00

Tests for Alpha in Linear Factor Models via Maximum Random Sparse Portfolios

Guanhao (Gavin) Feng, City University of Hong Kong

Given the enlarging factor zoo, a fundamental question is to evaluate whether a specific linear factor model statistically explains cross-sectional expected asset returns and compare which model is the best among a set of candidate pools. The pioneering GRS test requires the construction of a portfolio that maximizes the Sharpe ratio to test and compare factor models. However, when there are a large number of assets, the sample covariance matrix estimator becomes inconsistent and nontrivial biased, hindering the calculation of the maximum Sharpe ratio portfolio and making the GRS test ineffective. The Maximum Random Sparse Portfolios (MRSP) test aims to construct multiple random sparse portfolios, allowing the selection of the most effective one for testing. It will be shown rigorously that the proposed MRSP test converges to a type-I extreme value distribution as min (T, N) approaches infinity, under mild conditions.

The performance in finite samples is examined through extensive Monte Carlo simulations. Our proposed MRSP demonstrates a reasonable empirical size. Regarding empirical power, the MRSP test preserves the power of the naive portfolio test against sparse alternatives and improves the power of the maximum-type test against dense alternatives. Empirically, the MRSP test is applied to evaluate the U.S. stock market's efficiency and compare the seven workhorse factor models. Overall, the results demonstrate that the proposed MRSP test outperforms existing tests in terms of both statistical power and economic implications.

Coefficient Shape Alignment in Multiple Functional Linear Regression

Shuhao Jiao, City University of Hong Kong

In multivariate functional data analysis, different functional covariates often exhibit homogeneity. The covariates with pronounced homogeneity can be analyzed jointly within the same group, offering a parsimonious approach to modeling multivariate functional data. In this paper, a novel grouped multiple functional regression model with a new regularization approach termed "coefficient shape alignment" is developed to tackle functional covariates homogeneity. The modeling procedure includes two steps: first aggregate covariates into disjoint groups using the new regularization approach; then the grouped multiple functional regression model is established based on the detected grouping structure. In this grouped model, the coefficient functions of covariates in the same group share the same shape, invariant to scaling, and we develop a novel method to capture shape discrepancy, leading to robust grouping structure detection. We establish conditions under which the true grouping structure can be accurately identified and derive the asymptotic properties of the model estimates. This work offers a novel framework for analyzing the homogeneity of functional covariates and constructing parsimonious models for multivariate functional data.

Panel Quantile GARCH Models under Homogeneity

Qianqian Zhu, Shanghai University of Finance and Economics

Empirical evidence indicates that the estimates of GARCH parameters cluster in a panel of financial assets, potentially due to assets with similar exposure to common market risks. To capture the subgroup effect on conditional quantiles of financial asset returns and improve estimation efficiency by pooling information across individuals within the same group, this paper introduces the panel quantile GARCH model with homogeneous structures in the coefficient functions. We propose a three-stage estimation procedure to detect the grouping structures using a binary segmentation algorithm and estimate the

coefficient functions under detected homogeneity by quantile regression. Asymptotic properties are established for both group detection and the coefficient estimators. The proposed model and estimation procedure are further extended to allow for factor structures in the conditional quantiles. Simulation experiments are conducted to evaluate the finite-sample performance of the proposed estimation procedure. The results indicate that the final estimator, which utilizes group panel information, is more efficient than the initial estimator that relies on individual information alone, particularly when a subgroup effect exists. An empirical example on a panel of U.S. financial returns is presented to illustrate the usefulness of the proposed methodology in pursuing homogeneity, as well as its superior performance in forecasting value-at-risks at tail quantiles compared to quantile GARCH models that do not utilize any homogeneous information in the panel.

Adjusted Expected Improvement for Cumulative Regret Minimization in Noisy Bayesian Optimization

Shouri Hu, University of Electronic Science and Technology of China

The expected improvement (EI) is one of the most popular acquisition functions for Bayesian optimization (BO) and has demonstrated good empirical performances in many applications for the minimization of simple regret. However, under the evaluation metric of cumulative regret, the performance of EI may not be competitive, and its existing theoretical regret upper bound still has room for improvement. To adapt the EI for better performance under cumulative regret, we introduce a novel quantity called the evaluation cost which is compared against the acquisition function, and with this, develop the expected improvement-cost (EIC) algorithm. In each iteration of EIC, a new point with the largest acquisition function value is sampled, only if that value exceeds its evaluation cost. If none meets this criteria, the current best point is resampled. This evaluation cost quantifies the potential downside of sampling a point, which is important under the cumulative regret metric as the objective function value in every iteration affects the performance measure. We establish in theory a high-probability regret upper bound of EIC based on the maximum information gain, which is tighter than the bound of existing EI-based algorithms. It is also comparable to the regret bound of other popular BO algorithms such as Thompson sampling (GP-TS) and upper confidence bound (GP-UCB). We further perform experiments to illustrate the improvement of EIC over several popular BO algorithms.

Session 22	**Inference for High-dimensional Time Series and Survival Data** *Organizer/Chair: Rongmao Zhang, Zhejiang University*	E22-2009	12 Dec, 14:30 - 16:00

Fiducial inference in survival analysis

Yifan Cui, Zhejiang University

In this talk, we introduce novel nonparametric and semiparametric fiducial approaches to censored survival data. We propose Gibbs samplers and establish Bernstein-von Mises theorems. We also demonstrate our estimators by extensive simulations and real data applications.

High-dimensional Vector Autoregressions: A Ridge Dynamic Mode Decomposition Approach

Liyuan Cui, City University of Hong Kong

Dynamic mode decomposition (DMD), a commonly employed technique for dimension reduction and factorization in complex dynamical systems within fluid mechanics, is still relatively underdeveloped in its application and theoretical properties in statistical analysis. Given the intrinsic relationship between dynamical systems and autoregressive models, this paper delves into the applications and benefits of DMD in high-dimensional vector autoregressive (VAR) models. Existing high-dimensional VAR models often face deteriorated performance due to challenges in rank determinations and curse of dimensionality. We propose a truncated-ridge regularized DMD method with a proven upper error bound for the transition matrix, whose convergence rate can be further enhanced with our newly proposed rank-determination method.

To address the influence of noise, we introduce a bagging truncated ridge DMD method based on block bootstrap, significantly enhancing the stability of ridge DMD. Through an extensive set of simulation studies, we confirm that our DMD-based VAR estimation methods excel in providing more accurate estimates of the transition matrix and enhanced prediction performance compared to existing approaches. As an empirical application, we utilize our DMD-based VAR methods to predict the inflation rates of 25 countries from 1995 to 2023. Our new methods demonstrate improved predictability and desirable interpretability, presenting an alternative approach to examining changes in the transmission mechanism of international business cycle fluctuations.

Sparse Asymptotic PCA: Identifying Sparse Latent Factors Across Time Horizon

Zhaoxing Gao, University of Electronic Science and Technology of China

This paper proposes a novel method for sparse latent factor modeling using a new sparse asymptotic Principal Component Analysis (APCA). This approach analyzes the co-movements of large-dimensional panel data systems over time horizons within a general approximate factor model framework. Unlike existing sparse factor modeling approaches based on sparse PCA, which assume sparse loading matrices, our sparse APCA assumes that factor processes are sparse over the time horizon, while the corresponding loading matrices are not necessarily sparse. This development is motivated by the observation that the assumption of sparse loadings may not be appropriate for financial returns, where exposure to market factors is generally universal and non-sparse. We propose a truncated power method to estimate the first sparse factor process and a sequential deflation method for multi-factor cases. Additionally, we develop a data-driven approach to identify the sparsity of risk factors over the time horizon using a novel cross-sectional cross-validation method. Theoretically, we establish that our estimators are consistent under mild conditions. Monte Carlo simulations demonstrate that the proposed method performs well in finite samples. Empirically, we analyze daily stock returns for a balanced panel of S\&P 500 stocks from January 2004 to December 2016. Through textual analysis, we examine specific events associated with the identified sparse factors that systematically

influence the stock market. Our approach offers a new pathway for economists to study and understand the systematic risks of economic and financial systems over time.

Factor-based Estimation for High Dimensional Reduced-rank Time Series

Rongmao Zhang, Zhejiang University

High-dimensional time series are very common in reality. Analyzing each series separately may not be a good strategy, as it may miss some important information and result in a less optimal outcome. Even worse, in some cases, it may not even provide an answer to the question of interest. Multi-response time series emerges as an important tool for joint analysis. While the methodology and theory of classic vector time series models are well established, they may not apply to high-dimensional cases. In this paper, we develop a new and powerful method for estimating the coefficient matrix of high-dimensional reduced-rank time series, based on factor-augmented models. With the help of the estimated factors, we also propose two statistics to test for the dependence of high-dimensional time series. Asymptotic results for the proposed estimators and tests are established. Intensive simulation studies show that the proposed procedure is more powerful than its alternatives. We also apply the proposed method to a real dataset to illustrate its usefulness in solving real-life problems.

Session 25	**New Frontiers in Business Intelligence: Innovative Insights and Cutting-Edge Methods** *Organizer/Chair:* *Lingzhou Xue, The Pennsylvania State University;* *Weichen Wang, The University of Hong Kong*	E22-2010	12 Dec, 14:30 - 16:00

Technology's March, Inclusiveness in the Larch: Unveiling GAI's Aspiring Minority Arch for Online Labor Markets

Yanzhen Chen, The Hong Kong University of Science and Technology

Recent progress in Generative Artificial Intelligence (GAI) have introduced disruptive changes to the nature of work. By improving language and reducing overall costs of efforts, GAI has reshaped the dynamics of the online labor market. While prior research predominantly focused on understanding how GAI empowers people in general, there exists a discernible void in discussions concerning its inclusivity, particularly for minority groups—an area that our study endeavors to address. Using data from a large online labor market, we find that aspiring minority professionals on the platform have witnessed a notable decline in their job market prospects and performance subsequent to the introduction of GAI. This inadvertent and adverse consequence becomes more conspicuous when interacting with employers who possess limited attention when sifting through candidates and have a restricted understanding of their qualifications. Our findings emphasize the importance of both embracing innovation and maintaining a vigilant awareness of inclusiveness concerns, particularly for employers and the online labor market. Additionally, our research sheds light on how individuals from minority backgrounds and are building reputations, can effectively navigate and cope with the new challenges introduced by GAI.

Knowledge Transfer Federated Learning in the Era of LLMs

Yang Liu, Tsinghua University

In an era where Artificial Intelligence (AI) is increasingly integrated into our daily lives, bridging the power of large language models (LLMs) to private domain remains a crucial challenge. Federated learning (FL) emerges as a promising paradigm for developing private intelligence, enabling AI models to be trained collaboratively on decentralized devices without exposing private data. This talk will delve into recent advances in federated learning, with a focus on key techniques and applications to foster collaboration between LLMs and small domain models.

Going Beyond Black Box Models by Leveraging Behavioral Insights: an Intent-Based Recommendation Framework Authors

Yuyan Wang, Stanford University

Most platforms today rely on large-scale machine learning (ML) models to predict consumer choices. However, these systems are black-box in nature, lacking generalizability and understanding of consumer behaviors. In this work, we demonstrate how behavioral insights can be leveraged to improve these black-box ML systems. In the context of recommender systems, we introduce an intent-based recommendation framework that incorporates consumer's intent-driven decision making process into the system's design. Unlike traditional recommender systems, which rely on a single-step, black-box prediction, our framework employs a two-phase approach: an intent prediction phase that dynamically predicts consumer intents, followed by an intent-based planning phase that integrates these real-time predictions across various stages of the recommender system. By moving beyond traditional item-level predictions to a \emph{higher-order} modeling of intent-driven behaviors, our approach aligns the ML system more closely with the underlying data generation process, thereby improving its performance \emph{without} the need for extra data. We theoretically prove the optimality of our proposed framework

by considering consumers' scrolling behaviors, establishing a solid mathematical foundation for large-scale industrial applications.
Our framework was validated through extensive A/B testing on YouTube, the world's largest video recommendation platform. By incorporating consumer intents related to novelty and familiarity, we achieved a 0.05\% increase in daily active users (DAU), one of the most significant business metric improvements observed in recent YouTube experiments. Our work provides empirical evidence in support of how behavioral insights can be utilized to improve ML systems. Contrary to the common belief that introducing structure to ML systems reduces their flexibility, our findings show that imposing a structure which aligns with the underlying data generation process can, in fact, improve the performance of these systems.

Preferential Latent Space Models for Networks with Textual Edges

Emma Jingfei Zhang, Emory University

Many real-world networks contain rich textual information in the edges, such as email networks where an edge between two nodes is an email exchange. The useful textual information carried in the edges is often discarded in most network analyses, resulting in an incomplete view of the relationships between nodes. In this work, we represent each text document as a generalized multi-layer network, and introduce a new and flexible preferential latent space network model that can capture how node-layer preferences directly modulate edge probabilities. We establish identifiability conditions for the proposed model and tackle model estimation with a computationally efficient projected gradient descent algorithm. We further derive the non-asymptotic error bound of the estimator from each step of the algorithm. The efficacy of our proposed method is demonstrated through simulations and an analysis of the Enron email network.

Session 26	**Innovative Statistical Approaches for Complex Data** *Organizer/Chair:* *Weixin Yao, University of California, Riverside;* *Chuoxin Ma, Beijing Normal University-Hong Kong Baptist University United International College (UIC)*	E22-2011	12 Dec, 14:30 - 16:00

Robust model averaging prediction of longitudinal response with ultrahigh-dimensional covariates

Jialiang Li, National University of Singapore

Model averaging is an attractive ensemble technique to construct fast and accurate prediction. Despite of having been widely practiced in cross-sectional data analysis, its application to longitudinal data is rather limited so far. We consider model averaging for longitudinal response when the number of covariates is ultrahigh. To this end, we propose a novel two-stage procedure in which variable screening is first conducted and then followed by model averaging. In both stages, a robust rank-based estimation function is introduced to cope with potential outliers and heavy-tailed error distributions, while the longitudinal correlation is modelled by a modified Cholesky decomposition method and properly incorporated to achieve efficiency. Asymptotic properties of our proposed methods are rigorously established, including screening consistency and convergence of the model averaging predictor, with uncertainties in the screening step and selected model set both taken into account. Extensive simulation studies demonstrate that our method outperforms existing competitors, resulting in significant improvements in screening and prediction performance. Finally, we apply our proposed framework to analyse a human microbiome dataset, showing the capability of our procedure in resolving robust prediction using massive metabolites.

A single index varying coefficient panel data model with interactive fixed effects

Rui Li, Shanghai University of International Business and Economics

In this paper, we propose a single index varying coefficient panel data model with the unobservable multiple interactive fixed effects that are allowed to be correlated with covariates. Integrating B-spline approximation with the least squares approach, we get estimates of link function and coefficient functions through robust iterative algorithm. Under some regular conditions, we show that the estimators are consistent and asymptotically normal. Further, we provide two generalized likelihood ratio statistics from residual-based block bootstrap procedure to identify that whether the link function is linear and whether the coefficient function is some constant respectively. To circumvent the impact of block length on hypothesis testing, a Warp-speed approach is introduced for selecting block lengths. Monte Carlo simulations are conducted to illustrate the good performance of our proposed methods. Finally, two real data applications are implemented using our proposed method, in which the impact of cash flow volatility and earnings management on enterprise is explored from a CSMAR dataset, while the influence factors on temperature are also analyzed from a Met Office climate dataset.

Nonparametric Variable Selection Using Deep Differentiable Neural Networks

Xingqiu Zhao, The Hong Kong Polytechnic University

To effectively identify significant predictors nonparametrically, we propose a unified framework for variable selection. This framework employs a general loss function relating to an outcome and a completely unspecified regression function. Our methodology involves approximating the unknown function using deep differentiable neural networks and formulating a penalized loss function by imposing penalties on the first partial derivatives of the unknown function with respect to predictors. We establish the nonasymptotic error bound for the resulting penalized estimator based on this loss function. Furthermore, we derive nonasymptotic error bounds for the estimated derivatives and ensure variable

selection consistency. Through extensive simulation studies and real-world applications, we demonstrate the effectiveness of our proposed approach.

Identifying important factors for depressive symptom dynamics in Chinese middle-aged and older adults using a multi-state transition model with feature selection

Chuoxin Ma, Beijing Normal University-Hong Kong Baptist University United International College (UIC)

Depressive symptoms are increasingly common in middle-aged and older adults and have become a major public health problem. People may experience transitions across different underlying states due to symptom variability over a course of many years. And risk factors may have different impact on different symptom states. However, existing research rarely consider the identification of important factors related to symptom conversion. The purpose of this study was to examine the risk associated with transitioning between various stages of depressive symptoms and their influencing factors, utilizing a multi-state model with simultaneous feature selection method. Multi-state models were used to evaluate the risks of state transitions. The maximum L1-regularized (adaptive Lasso) partial likelihood approach was employed to simultaneously identify the important risk factors, estimate their impact on the state transition rates and determine their statistical significance. We identified several important risk factors including sex, education level, income, number of chronic conditions, body pain, difficulty with basic daily activities, and marital status.

Session 29	**Research and Applications of Microeconometric Models** *Organizer/Chair: Yan Sun, Shanghai University of Finance and Economics*	E22-2013	12 Dec, 14:30 - 16:00

Identification and Estimation of Average Causal Response Function in a high-dimensional Sample Selection Model

Yahong Zhou, Shanghai University of Finance and Economics

Average causal response function (ACRF) is a useful tool to assess treatment effect with dose functions, especially when the treatment is endogenous. This paper presents the identification and estimation of an ACRF with sample selection and a high dimensional controls. We derive the Ney-man orthogonal moments with multiple nuisance parameters and utilize double machine learning method and typical nonparametric techniques to estimate the proposed estimators. Asymptotics for proposed estimators are derived and Monte Carlo simulations demonstrate their good finite sample properties. Our identification and estimation results could be readily extended to the case with more complex sample selection mechanisms. We apply the proposed method to US Job Corps data to evaluate the heterogeneous effect of residential components, which yields new insights for policy makers.

The Flexibility and Labor Supply of the Gig Economy : Based on Micro Big Data of the Ride-Hailing Market

Pingfang Zhu, Shanghai Academy of Social Science

The gig economy, characterized by its flexibility, has altered traditional labor market dynamics, including employment patterns, work styles, and job structures. This study represents worker welfare as the labor supply surplus, defined as the difference between wages and reservation wages. The ability of workers to adjust their labor supply in response to fluctuations in expected wages and reservation wages is crucial for the enhancement of welfare through flexibility. Utilizing micro-level big data from the Shanghai ride-hailing market, this paper classifies different types of ride-hailing drivers and establishes a high-dimensional labor supply model. By employing high-dimensional MCMC (Markov Chain Monte Carlo) methods to estimate reservation wages, the impact of flexibility on worker welfare is quantitatively evaluated. The findings indicate that flexibility significantly enhances labor supply and worker welfare, nearly doubling the welfare compared to fixed working hours. This study represents a cutting-edge practice in labor supply research using big data, providing an empirical framework applicable to the study of other new employment forms in the gig economy.

Efficient and Sequential Estimation of High-order Spatial Dynamic Panels with Time-varying Dominant Units

Xiaoyi Han, Xiamen University

This paper studies a high-order spatial dynamic panel data model (SDPD) that incorporates time-varying dominant units and unknown heteroskedasticity, where the column sums of the spatial weights matrices may not be uniformly bounded due to the presence of dominant units. The number of dominant units can be finite or infinite, and may change over time and across different spatial weights matrices. We develop new central limit theorems (CLTs) for linear-quadratic forms that allow for strongly dominant units, where the column sum magnitude of the matrix in the linear-quadratic form can diverge to infinity in the same speed as n, and also accounts for unknown heteroskedasticity. We propose two moment-based estimation methods: the generalized method of moments estimator (GMME) and the root estimator (RE). The RE has a closed-form solution that does not involve any optimization algorithm, and can achieve the same asymptotic efficiency as the best GMME. We establish the consistency and asymptotic normality of these estimators when both n and T are large, and show that the convergence rate of GMME may depend on the column sum magnitudes of the weights matrices

appearing in the moment conditions. We further extend the model to accommodate endogenous spatial weights and introduce a two-step RE. Monte Carlo simulations demonstrate that our proposed estimators have satisfactory finite sample performance. An empirical application to the peer effects of financial decisions among Chinese listed firms is provided to illustrate the merit of our proposed models and estimation methods.

The impact of robots on unemployment duration: Evidence from the Chinese General Social Survey

Linhui Wang, Jilin University

The issue of the impact of robot applications on unemployment duration remains relatively unexplored. Using individual-level data from the Chinese General Social Survey (CGSS), we shed new light on the effects of robots on unemployment likelihood and duration. Major findings include that: (1) robots prolong the duration of unemployment, particularly for workers in routine occupations compared to those in non-routine occupations. This phenomenon can be understood as a form of structural unemployment; (2) workers with low skills, low economic and occupational status, and parenting responsibilities suffer more from the robot shock. In particular, robots adversely affect women's careers, and traditional perceptions amplify this effect to dampen their willingness and efforts to find job searching, resulting in longer periods of unemployment. These heterogeneous impacts of robots on unemployment duration highlight the need for greater support and policy interventions for vulnerable groups; (3) access to information from digital channels plays a pivotal role in alleviating the adverse impacts of robots on vulnerable groups. Reducing information asymmetry and nurturing job search capabilities help the vulnerable better adapt to evolving labor market conditions, thereby mitigating the negative consequences of robots.

Session 30	**Predictive Regression, Rough Volatility and Tail Risk** *Organizer/Chair: Yang Zu, University of Macau*	E22-2014	12 Dec, 14:30 - 16:00

Consistent Inference for Long-Horizon Predictive Regressions

Yubo Tao, University of Macau

This paper considers the estimation and inference problem of long-horizon predictive regression models where the state variables are fractionally integrated. We develop a consistent local spectrum (LCM) estimation procedure that delivers asymptotic Gaussian inference. Monte Carlo analysis suggests that the proposed test offers improved finite sample properties compared to the leading tests in the literature. As an empirical application, we reexamine the predictive ability of various popular predictors for aggregate equity premium.

Exploiting the errors and jumps: A time-varying rough volatility model for improved forecasts

Xiaohu Wang, Fudan University

This paper introduces a time-varying rough volatility model to mitigate the adverse impacts of measurement errors and jumps in forecasting. The specification is a discretized fractional Ornstein-Uhlenbeck (fOU) process with a time-varying persistency depending on realized quadraticity. Under the in-fill asymptotic scheme, the process has a stochastic-unit-root specification, with the error term being a fractional Gaussian noise. So that fractional integration is possible. Methods are proposed to estimate parameters in the model. The asymptotic theory is developed for the estimators. Empirical estimates from RV of 110 cryptocurrencies suggest strong evidence of time-varying persistency and roughness. When using the proposed model to forecast the RV of 110 cryptocurrencies, we find evidence of superior forecasting performance of the proposed model relative to other popular models in the literature.

Evaluating parameter change in a heteroskedastic predictive regression model

Yang Zu, University of Macau

A nuisance parameter-free evaluation test in a heteroskedastic predictive regression model is difficult to construct. We develop a class of such tests based on the recursive residuals computed from a feasible weighted least squares regression. We show that the resulting CUSUM process of the recursive residuals admits a standard Brownian motion limit. Our class of tests is constructed based on this weak convergence result, and we show in Monte Carlo simulation that our tests enjoy good size and power properties.

The Granular Origins of Tail Dispersion Risk

Yi Ding, University of Macau

We study tail risk in the cross-section of asset prices at high frequencies. The tail behavior of the cross-section depends on whether a systematic jump event occurred. If so, the cross-sectional return tail is governed by assets' exposures to the systematic event while, otherwise, it is determined by idiosyncratic jumps. An estimator for the tail shape of the cross-sectional distribution displays distinct properties with and without systematic jumps. We show empirically that shocks to the cross-sectional tail shape are a source of priced risk: fat idiosyncratic tails are favored by investors, while fat-tailed exposures to systematic jumps are disliked.

Session 46	**Supervised and Semi-supervised Learning for Estimation and Decision-Making** *Organizer/Chair:* *Xinyuan Song, The Chinese University of Hong Kong;* *Kai Kang, Sun Yat-sen University*	E22-2015	12 Dec, 14:30 - 16:00

No-lose Converging Kernel Estimation of Long-run Variance

Kin Wai Chan, The Chinese University of Hong Kong

Kernel estimators have been popular for decades in long-run variance estimation. To minimize the loss of efficiency measured by the mean-squared error in important aspects of kernel estimation, we propose a novel class of converging kernel estima- tors with three major properties: (1) the optimal bandwidth choice is model-free; (2) positive-definiteness is ensured through a principle-driven aggregation technique with no loss of theoretical efficiency; and (3) potentially misspecified prewhitening models and transformations of the time series do not harm the asymptotic efficiency. A shrinkage prewhitening transformation is proposed for more robust finite-sample performance. The estimator has a positive bias that diminishes with the sample size so that it is more conservative compared with the typically negatively biased classical estimators. The proposal improves upon standard kernel functions and can be well generalized to the multivariate case. We discuss its performance through simulation results and a real-data application in the forecast breakdown test.

Inference on Potentially Identified Subgroups in Clinical Trials

Xinzhou Guo, The Hong Kong University of Science and Technology

When subgroup analyses are conducted in clinical trials with moderate or high dimensional covariates, we often need to identify candidate subgroups from the data and evaluate the potentially identified subgroups in a replicable way. The classical statistical inference applied to the potentially identified subgroups, assuming the subgroups are the same as what we observe from the data, might suffer from bias issue when the regularity assumption that the boundaries of the subgroups are negligible is violated. In this talk, we will introduce a shift-based method to address nonregularity bias issue and combining it with cross-fitting and subsampling, develop a de-biased inference procedure for potentially identified subgroups. The proposed method is model-free and asymptotically efficient whenever it is possible, and can be viewed as an asymmetric smoothing approach. The merits of the proposed method are demonstrated by re-analyzing the ACTG 175 trial. This talk is based on joint work with Shuoxun Xu.

Semisupervised Score Based Matching in Public Health Intervention Evaluation and Decision Making

Jiasheng Shi, The Chinese University of Hong Kong, Shenzhen

Multivariate matching algorithms "pair" similar study units in an observational study to remove potential bias and confounding effects caused by the absence of randomizations. In one-to-one multivariate matching algorithms, a large number of "pairs" to be matched could mean both the information from a large sample and a large number of tasks, and therefore, to best match the pairs, such a matching algorithm with efficiency and comparatively limited auxiliary matching knowledge provided through a "training" set of paired units by domain experts, is practically intriguing.

We proposed a novel one-to-one matching algorithm based on a quadratic score function $S_{\beta}(x_i,x_j)= \beta^T (x_i-x_j)(x_i-x_j)^T \beta$. The weights β, which can be interpreted as a variable importance measure, are designed to minimize the score difference between paired training units while maximizing the score difference between unpaired training units. Further, in the typical but intricate case where the training set is much smaller than the unpaired set, we propose a \underline{s}emisupervised \underline{c}ompanion \underline{o}ne-\underline{t}o-\underline{o}ne \underline{m}atching \underline{a}lgorithm (SCOTOMA) that makes the best use of the unpaired units.

The proposed weight estimator is proved to be consistent when the truth matching criterion is indeed the quadratic score function. When the model assumptions are violated, we demonstrate that the proposed algorithm still outperforms some popular competing matching algorithms through a series of simulations. We applied the proposed algorithm to a real-world study to investigate the effect of in-person schooling on community Covid-19 transmission rate for policy making purpose.

Efficient Estimation for Functional Accelerated Failure Time Model

Kin Yat Liu, The Chinese University of Hong Kong

We introduce a functional accelerated failure time model designed to analyze the impact of both functional and scalar covariates on the time until a specific event occurs. We establish regularity conditions to ensure the model's identifiability. To efficiently estimate the model parameters, we propose a sieve maximum likelihood method that integrates parametric and nonparametric coefficients along with an unknown baseline hazard function into the likelihood framework. This integration presents significant numerical challenges and complicates theoretical development. By creating a comprehensive theoretical framework, we address these issues and determine the convergence rate of our estimator. We also demonstrate that the finite-dimensional estimator is root-n consistent, asymptotically normal, and meets the semiparametric information bound. Moreover, we establish the nonparametric optimality of the functional estimator and formulate an asymptotic simultaneous confidence band. Our inference methods are validated through extensive simulations and applied to data from the National Health and Nutrition Examination Survey.

Session 48	**Exploring the Intersection of AI: Implications of Artificial Intelligence in Academical, Financial and Digital Industries** *Organizer/Chair: Yingpeng Zhu, University of Macau*	E22-2017	12 Dec, 14:30 - 16:00

Use of Large Language Models in Survey Development

Ping Fan Ke, Singapore Management University

This study investigates the use of Large Language Models (LLMs) in survey development for business and research through the creation and evaluation of the Behavioral Research ASSistant (BRASS) Bot. We began by examining the traditional scale development process to pinpoint tasks that could benefit from LLM integration and created corresponding LLM prompts for the BRASS Bot. We then tested its deployment and performance across various LLMs. Additionally, we showcased its practical value with a predictive validity simulation. Our research offers both theoretical and practical insights by uncovering previously missed patterns, saving significant time and resources, and enhancing scale validity. This work guide for future research on the broader application of LLMs as assistants and collaborators in survey analysis and behavioral research, underscoring their transformative potential for the field.

Responsive Yet Non-human: The Dual Influence of Financial Chatbots on Empathic Engagement During Stock Price Plunge

Lingfei Deng, Sun Yat-sen University

Empowered by technical innovation, AI-based chatbots possess enhanced interactive capabilities to facilitate investor services in the finance field. An online experimental survey (n=395) was conducted, focusing on a stock price plunge scenario, to investigate the influence of financial chatbots on investor emotion and behavior. Our findings suggest that perceived empathy is a key factor in investor interactions, which further affects the mitigation of negative emotions and changes in investment intentions. Then, we uncover a dual pathway through which financial chatbots, compared to senior executives, influence investors' perceived empathy. Specifically, financial chatbots can enhance empathy through the mediation of excellent responsiveness, while its direct impact on empathy is negative, possibly stemming from investors' subjective biases against non-human entities. Mediation analysis highlights the significance of several theoretical pathways. We also find partial heterogeneity effects of interaction orientation (information/emotion). These findings offer theoretical contributions to relevant literature and valuable guidance for listed companies.

Effects of Autonomous Algorithmic Moderation on User Content Contribution: A Natural Experiment in an Online Knowledge-Sharing Community

Yingpeng Zhu, University of Macau

Online digital platforms are increasingly relying on artificial intelligence (AI) technologies to perform platform governance. We study a growing application of AI in platform governance: automated content moderation. Platforms thriving on user-generated content (UGC) commonly deploy sophisticated AI algorithms to screen content uploaded by users and impose punishment (e.g., folding) on low-quality content. We leverage the sudden implementation of an automated answer-folding AI system in a large online knowledge-sharing community to investigate how automated AI content moderation affects users' content-generating behaviors in terms of both content quantity and quality (as measured by the number of voteups received and usage of cognitive language). Drawing on deterrence theory and related vicarious punishment theory, we hypothesize and empirically examine how punished users (i.e., users who received at least one folding punishment given by the algorithm) and observing users (i.e., users who did not receive a folding punishment given by the algorithm) might differentially respond to the introduction of the AI moderation algorithm. We find novel deterrence effects of AI punishment contingent on user type: Punished users refrained from generating new answers (i.e., an inhibition effect)

and their content quality did not improve, while observing users also showed an inhibition effect but they were motivated to reallocate more efforts on producing higher-quality content (i.e., a displacement effect). An additional online experiment is conducted to corroborate the empirical findings and explore why the deterrence effects of AI punishment are different from those of human punishment. Our research makes important theoretical contributions by extending the deterrence theory into the AI-punishment context and also provides important practical implications for online platforms that use, or consider using, AI algorithms to perform platform governance.

Session 51	**Machine Learning in Economics and Finance** *Organizer/Chair: Qingliang Fan, The Chinese University of Hong Kong*	E22-2018	12 Dec, 14:30 - 16:00

On the CCE Estimation of Three-Dimensional Panels with Multi-level Factors

Xun Lu, The Chinese University of Hong Kong

We consider the common correlated effects (CCE) estimation of a three-dimensional (3D) panel data model with heterogeneous slope coefficients and multi-level interactive fixed effects. We study the asymptotic properties of the heterogeneous slope coefficient estimators and the mean group (MG) estimators. We find that bias-correction is indispensable for the MG estimator and show that the MG estimators exhibit different convergence rates depending on the degree of heterogeneity in the slope coefficients. We propose a bias correction procedure and inference method that is uniformly valid irrespective of the degree of heterogeneity. Monte Carlo simulations demonstrate excellent performance of our estimators and inference methods in finite samples. We apply our new methods to international trade data and find significant heterogeneity in the gravity model even after we control for the multi-level factors.

Low-Frequency Risk Factors and Their Fundamental Drivers

Sicong Li, The Chinese University of Hong Kong

There is a "zoo" of factors that capture systematic risk premia and a large number of economic variables that explain their time variation, which poses a doubly high-dimensional challenge to understanding how economic fundamentals relate to the time-varying dynamics of risk premia. I propose a method to regularize this problem by identifying low-frequency risk factors, whose risk premia are driven by latent low-frequency state variables. Empirically, one below-business-cycle-frequency factor and one business-cycle-frequency factor, whose variation concentrates on cycles longer than eight years and between 1.5 and eight years, explain the expected returns of individual stocks and characteristic-managed portfolios. The below-business-cycle-frequency factor has a high Sharpe ratio, and stocks whose current size is small compared to their long-term average load on it. Moreover, selected macroeconomic and financial variables have statistically and economically significant out-of-sample predictive power for the returns of the two low-frequency factors.

A Unified Framework for Estimation of High-dimensional Conditional Factor Models

Qihui Chen, The Chinese University of Hong Kong, Shenzhen

This paper presents a general framework for estimating high-dimensional conditional latent factor models via constrained nuclear norm regularization. We establish large sample properties of the estimators and provide an efficient algorithm for their computation. Additionally, we propose a cross-validation procedure for selecting the regularization parameter, thereby enhancing the practical applicability of the method. Our framework enables the estimation of various conditional factor models in a unified manner and facilitates the derivation of novel asymptotic results. Empirical analyses of the cross section of individual US stock returns suggest that imposing homogeneity may improve the model's out-of-sample predictability.

Estimating the Efficient Frontier

Leheng Chen, The Hong Kong University of Science and Technology

This paper introduces an estimator named CORE (COnstrained sparse Regression for Efficient portfolio), designed to estimate large mean-variance efficient portfolios consisting of only risky assets. These portfolios collectively form an estimated efficient frontier. Our method relies on a novel linear constrained regression representation of the optimization problem. Theoretically, we establish the

CORE portfolio's asymptotic mean-variance efficiency under a high dimensional setting where both the number of assets and the sample size can go to infinity. We also consider a scenario when factors involve long-short portfolios and develop CORE-LS (CORE for Long-Short factors) to estimate the efficient portfolio in this scenario. Extensive simulation and empirical studies demonstrate the favorable performance of our proposed method.

Session 54	**Return Predictability and the Entrepreneurship Studies** *Organizer/Chair:* *Wenjin KANG, University of Macau;* *Rachel Ma University of Macau*	E22-G008	12 Dec, 14:30 - 16:00

Factor Zoo: Omitted Variable Bias Revisited

Wai Kong Cheung, City University of Macau

Factor zoo is large as many factors have been proposed to explain stock returns in the literature. These factors are either too many (high-dimensional) for classical statistical approaches to be applicable or their effect on stock returns may not be satisfactorily modeled by parametric (linear) functions. In addition, the impact of these factors may be heterogeneous. We examine the performance of a particular factor in the presence of other factors using double machine learning method. Among 94 factors, our preliminary result shows that only three factors can explain stock returns and they are beta, book to market ratio, and size, respectively. This suggests that only the three factor model can survive the test of time.

Intra-Family Relationships and Female Entrepreneurship

Mingzhu Tai, The University of Hong Kong

We study how intra-family relationships affect female entrepreneurship. By utilizing a major change in marriage law in the U.S., we find that female entrepreneurship increases significantly when the divorce regime changes from mutual consent to unilateral, especially when property rights are more favorable to women. This effect is mainly from an increase in the survival of existing female entrepreneurs rather than the entry of previous housewives. We consider a variety of potential mechanisms and conclude that the increase in female entrepreneurship after the law change is most likely driven by an increase in women's intra-family bargaining power against spouses.

Wall Street to Hong Kong: The Value of Dual Listing for China Concept Stocks

Zhuo Chen, Tsinghua University

The U.S. stock market has long been the most popular venue for both foreign companies and global investors. The recent cross-border regulation tensions between the U.S. and China, however, have exposed many U.S.-listed China Concepts Stocks (CCS) to substantial de-listing risks, forcing them to pursue dual listings on the Hong Kong Stock Exchange (HKEX). In this paper, we quantify the economic value of dual-listing, using the SEC's adoption of the final amendments implementing mandates of the Holding Foreign Companies Accountable Act (HFCAA) on December 2, 2021 as a natural experiment. We estimate that CCS with pre-shock dual-listing status on average have 14.88% higher returns, or USD 8 billion in market capitalization, than their peers listed only on the U.S. exchanges during a three-month period after the shock. Our findings survive a set of robustness checks, including parallel trends test, alternative treatment and control groups based on the qualified but not yet dual-listed CCS, and various sub-sample and placebo analyses. In addition to stock returns, dual-listed CCS are also less adversely affected in trading volume, volatility, and liquidity. Our findings highlight the large economic impact of the escalating political U.S.-China tensions on the global financial markets.

The Predictive Power of Capital Flows in Commodity Futures and the Impact of the New Securities Law Regulatory Reforms: Evidence from Futures Companies on Longhu List

Biao Guo, Renmin University of China

This paper aims to investigate the predictive ability of capital flows from futures companies on the Longhu List (Longhu List Members, LLM)on the returns of the commodity futures and to explore their performance under different market environments. We construct a TopOI factor by using futures position

data of LLM. Our analysis reveals that the capital flows of LLM can significantly positively predict future returns of commodity futures. A Long-Short portfolio based on the TopOI factor achieves an annualized excess return of 9.776%, with particularly notable performance before the release of important information and among state-owned LLM. Further analysis indicates that state-owned LLM hold net short positions most of the time under various macroeconomic conditions. However, after the implementation of the new Securities Law in 2020, their short-selling activities significantly reduced, and the predictive power of the TopOI factor markedly declined, especially for commodities with high private information. This research not only provides a new "position trend" perspective for understanding the relationship between capital flows and futures returns but also offers empirical evidence for the formulation of relevant regulatory policies.