Research Showcase - Time Series and Statistical Learning

Dr Yining Chen

Department of Statistics
London School of Economics and Political Science

LSE Statistics PhD Open Day 2022

Outline

Outline

TSSL Members

- Mona Azadkia - Assistant Professor
- Yining Chen - Associate Professor
- Piotr Fryzlewicz - Professor
- Kostas Kalogeropoulos - Associate Professor
- Clifford Lam - Professor
- Xinghao Qiao - Associate Professor
- Chengchun Shi - Assistant Professor
- Tengyao Wang - Associate Professor
- Qiwei Yao - Professor

General Research Focus

- Theoretical developments and practical applications of various time series and statistical learning problems.
- Develops statistical methods in high-dimensional inference and dimension reduction for various kinds of data.
- E.g. panel and tensor time series data, inferential analysis on dynamic networks and spatio-temporal processes, functional data analysis, functional time series analysis, shape-constrained estimation, change-point detection and multiscale modelling and estimation for high dimensional non-stationary time series.

Outline

Research highlight - Dr Mona Azadkia

- Mona's research focuses on developing new methodologies to understand and quantify the dependency structure in data.
- Finding interpretable measures of the degree of dependence between the variables is a fundamental task in Statistics. Such measures are the core ingredient of many areas, such as variable selection, dimensionality reduction, sensitivity analysis, and causal inference.

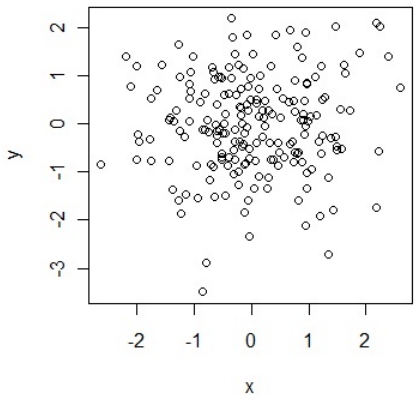
Research highlight - Dr Mona Azadkia

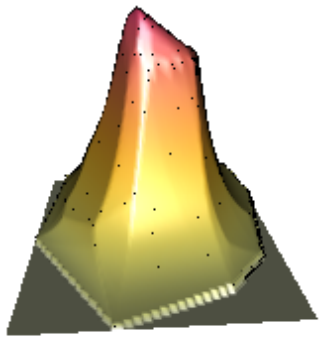
- Mona is also working on a generalization of this problem, measuring conditional dependence and the hypothesis testing problem of conditional independence with applications in causal inference and graphical models. She is also interested in non-parametric statistics and problems in high-dimension.
- Before joining LSE, Mona earned her Ph.D. in Statistics at Stanford University in 2020 and was an FDS Postdoc Fellow at Seminar for Statistics at ETH Zürich.

Outline

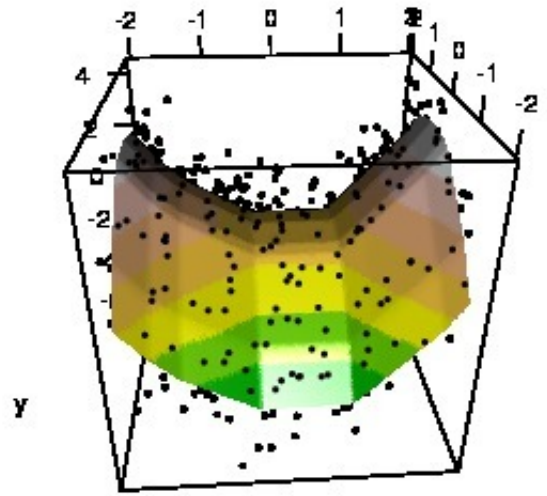
Research highlight - Dr Yining Chen

- Nonparametric statistics (especially shape-constrained methods)
- Change-point detection/estimation
- Computational statistics



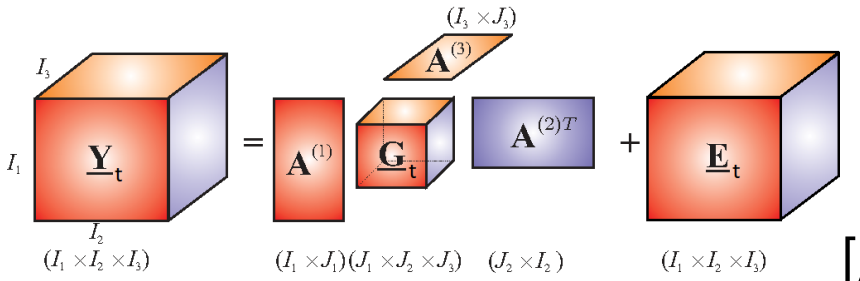


Research highlight - Dr Yining Chen



Outline

Research highlight - Prof Clifford Lam

- Tensor Time Series Analysis:
- E.g. (Economics) A study of fuel consumption: 3 different fuel types, 7 different types of cars in 12 cities $\Rightarrow$ At time t, $\mathcal{X}_t \in \mathbb{R}^{3\times 7\times 12}$ is an order 3 tensor.
- Tensor factor model:
- Order 3 tensor: 

[A. Phan and A. Cichocki (2011)]

Research highlight - Prof Clifford Lam

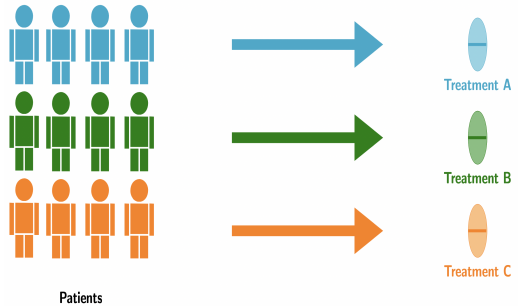
- High frequency data analysis: Market microstructure noise modelling for e.g. in portfolio allocation, integrated volatility + risk estimation
- Spatial Econometrics: Estimation of spillover effects from data - a type of network effects.
- Current projects include: Tensor factor model estimation including rank estimation, time-varying spatial weight matrix in spatial lag models as semiparametric time series factor model.

Outline

Research highlight - Dr Chengchun Shi

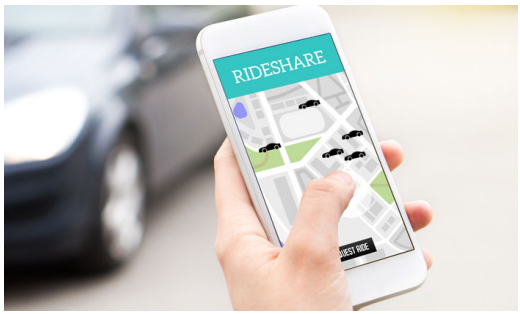
Overview

My research primarily focus on developing **statistical methods** (e.g., estimation & inference) in **reinforcement learning** (RL) with applications to



(a) Precision Medicine

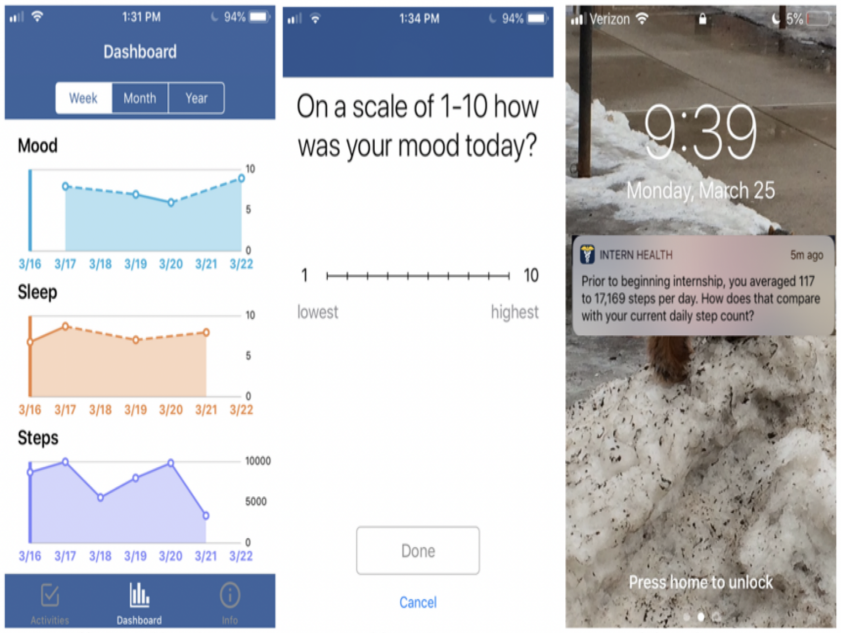
(b) Mobile Health (mHealth)



Research highlight - Dr Chengchun Shi

Applications in mHealth

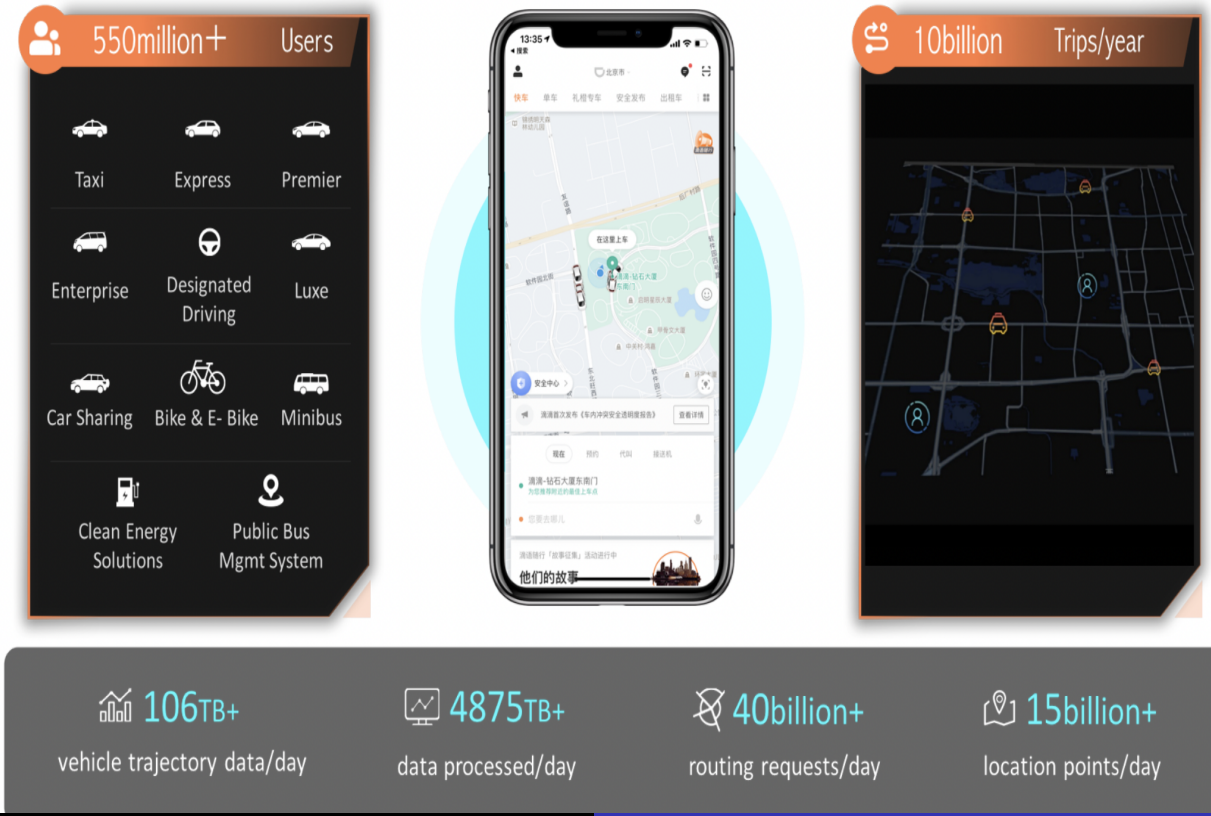
- An **emerging** area that uses smart phones, computer tablets or wearable devices in healthcare
- **Examples**: management of type-I diabetes (Shi et al., 2022); intern health study (NeCamp et al., 2020; see also the figure on the right)
- **Objective**: develop RL algorithms to learn data-adaptive treatment interventions to improve subjects' health conditions



(i) App Dashboard (ii) Mood EMA (iii) Notifications

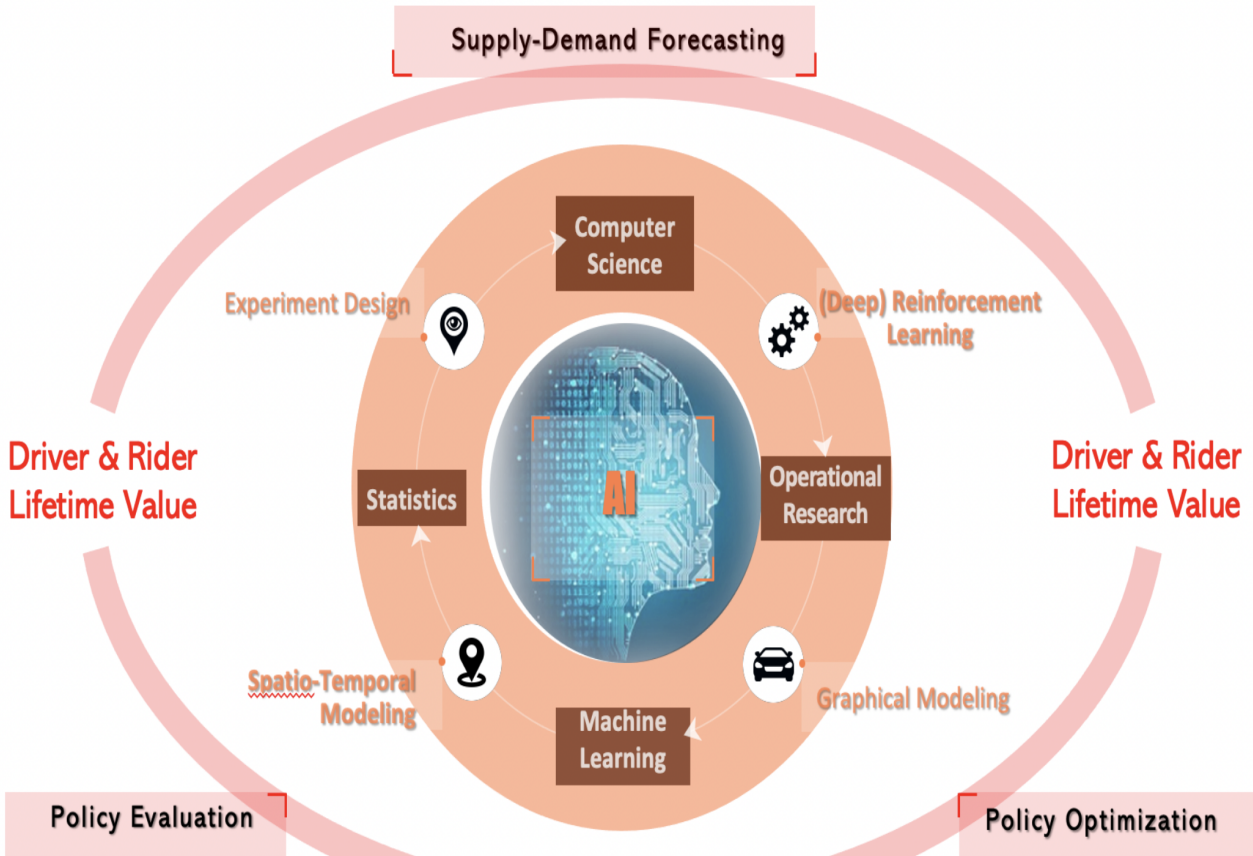
Research highlight - Dr Chengchun Shi

Applications in Ridesharing



Research highlight - Dr Chengchun Shi

Applications in Ridesharing (Cont'd)



Research highlight - Dr Chengchun Shi

Topics

- **Objectives**: policy learning, policy evaluation (causal inference, A/B testing), model validation/selection
- **Data**: high-dimensional, massive, non-markovanity, doubly inhomogeneous (e.g., nonstationary over time & heterogeneous over population), multi-relational, nonregular, unmeasured confounding

Outline

Research highlight - Dr Tengyao Wang

Sparse signal detection

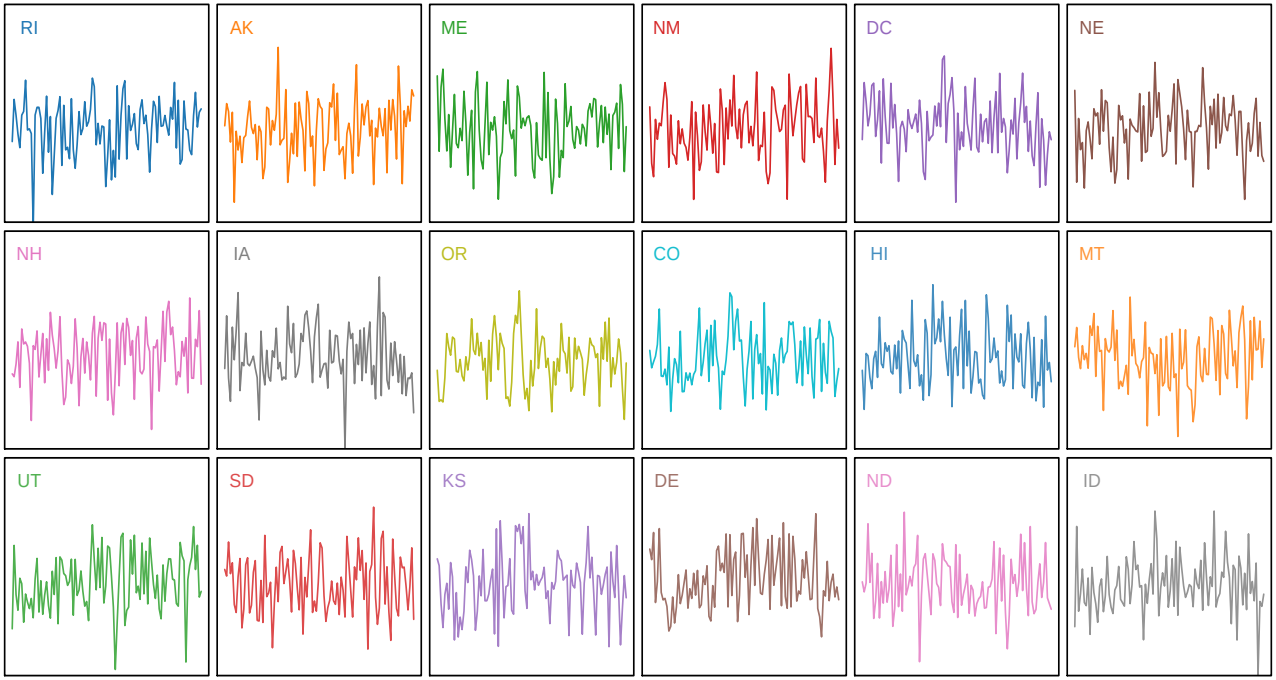
Research area: identifying sparse signal in high-dimensional data

- High-dimensional statistics study data whose dimension is of comparable order or even larger than the number of observations.
- High-dimensional data typically possess some low-dimensional structure
- The challenge is to develop a method that can identify and aggregate over such low-dimensional signal.

Research highlight - Dr Tengyao Wang

An example: sparse change detection

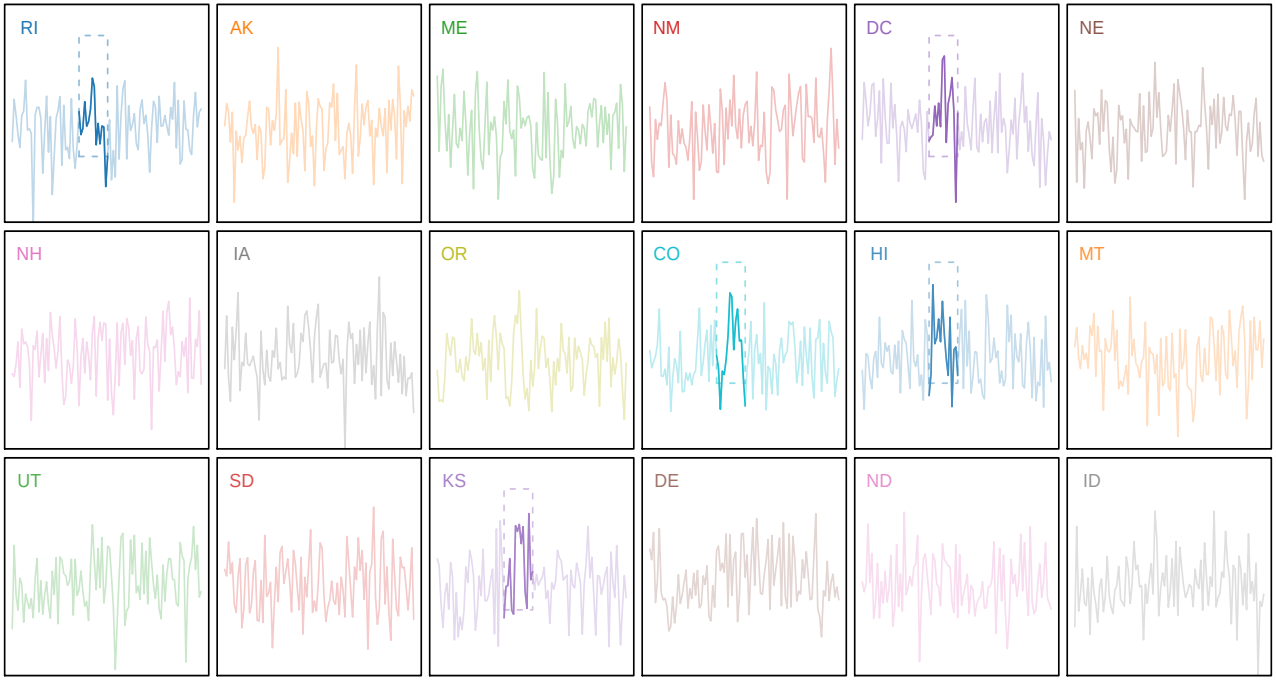
- Weekly standardised excess death numbers in US states between Jan 2017 and Dec 2018.
- Anything unusual?



Research highlight - Dr Tengyao Wang

An example: sparse change detection

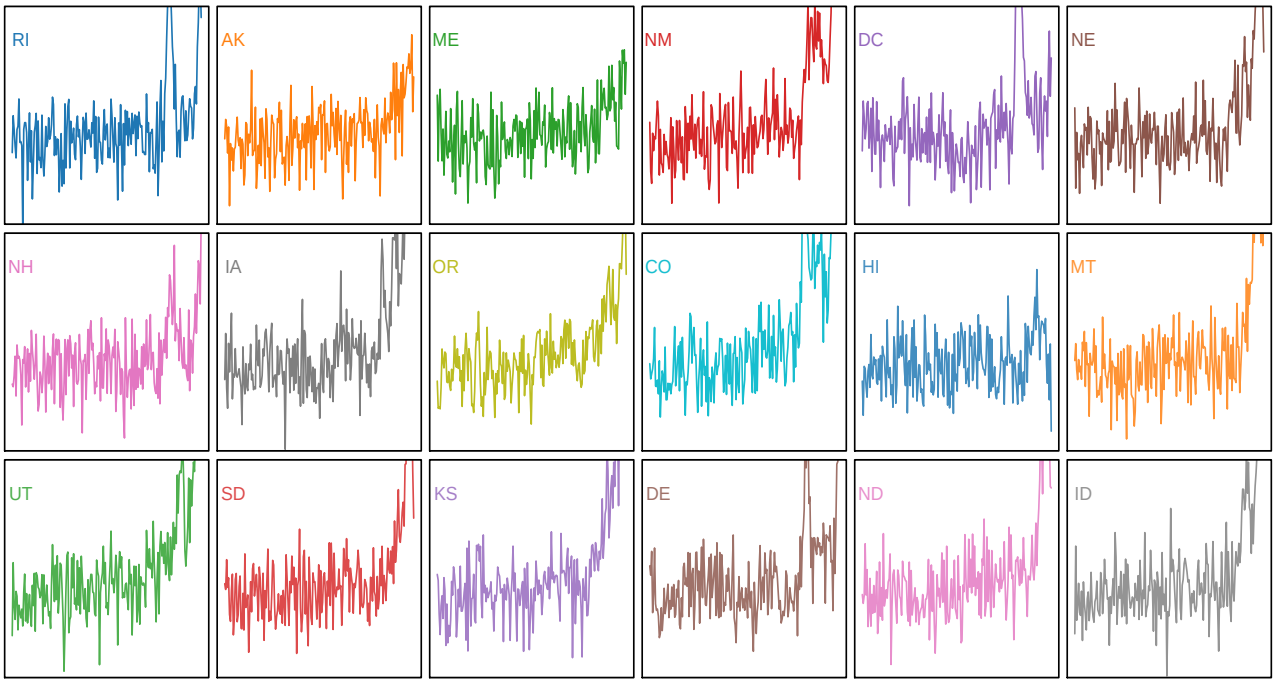
- Weekly standardised excess death numbers in US states between Jan 2017 and Dec 2018.
- Bad flu season in 2017 causing a spike in death numbers in a few states



Research highlight - Dr Tengyao Wang

An example: sparse online change detection

- Weekly standardised excess death numbers in US states between Jan 2017 and Dec 2020
- Declare change as soon as possible when there is a disease outbreak



Research highlight - Dr Tengyao Wang

Some research topics

Some research topics on sparse signal detection

- Sparse PCA: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N_p(0, \Sigma)$, where the leading eigenvector v of Σ is sparse. How do we estimate v?
- High-dimensional clustering: high-dimensional observations $X_1, \ldots, X_n$, when restricted to some (unknown) low-dimensional subspace, are distributed as mixture of Gaussians. How do we perform classification? Clustering? Semi-supervised learning?
- High-dimensional changepoints: $X_1, \ldots, X_z \overset{\text{iid}}{\sim} N_p(\mu_1, \Sigma)$ and $X_{z+1}, \ldots, X_n \overset{\text{iid}}{\sim} N_p(\mu_2, \Sigma)$ with $\theta = \mu_1 - \mu_2$ sparse. How do we test if $\theta = 0$? Estimate θ? Estimate z? Do it in an online algorithm?
- Two sample testing of regression coefficients: $Y_1 = X_1\beta_1 + \epsilon_1$, $Y_2 = X_2\beta_2 + \epsilon_2$. Difference between two regression coefficients $\theta = \beta_1 - \beta_2$ is sparse. How do we test $\theta = 0$? Estimate θ?

Outline

Research highlight - Prof Qiwei Yao

Complex Time Series Analysis

Qiwei Yao

Department of Statistics, London School of Economics
q.yao@lse.ac.uk

- High-dimensional time series
- Functional time series
- Spatio-temporal processes
- Dynamic networks

Research highlight - Prof Qiwei Yao

Distinctive feature of TS data: Dependence across times makes data analysis more exciting and challenging, and future forecasting possible.

Challenges:

1. Complex and high-dimensional data
2. Serial dependence in addition to cross-sectional dependence
3. Non-stationarity
4. Dependence beyond linear correlations

Prerequisite: Solid background in Mathematical Statistics/Mathematics

- Creativity in developing machine/statistical learning methods/algorithms
- Complex programming

Research highlight - Prof Qiwei Yao

Some past and on-going projects:

- Independent component analysis for high-D TS, matrix TS, or high-D random fields
- Dimension-reduction and factor modelling for high-D TS, or spatial-temporal processes
- Testing for high-D white noise
- Cointegration for high-D TS
- Intrinsic spatial processes and spatial cointegration
- High-D unit-root tests
- AR(1) dynamic network models
- Matrix TS modelling based on tensor CP-decomposition
- Clustering high-D TS based on cluster-specific factors