Time Series Adaptation and High Dimensionality

Parley Ruogu Yang
Faculty of Mathematics, University of Cambridge
https://parleyyang.github.io/

UNIVERSITY OF CAMBRIDGE
Faculty of Mathematics

Goldman Sachs

Data Science Seminar Series

Goldman Sachs (London), 20 Sep 2022

About myself

https://parleyyang.github.io

parleyyang.github.io

Parley Ruogu Yang

I am currently a *PhD student at the Cantab Capital Institute for the Mathematics of Information (CCIMI)*, **Faculty of Mathematics, University of Cambridge**. I am a founding member of the *Optimal Portfolio Research Group*, and within the faculty I volunteer at the *Stats Clinic.* I taught *ST456 Deep Learning* at the **London School of Economics** in 2022.

Recent Publications / Pre-prints

- *DMS, AE, DAA: methods and applications of adaptive time series model selection, ensemble, and financial evaluation*, 2022
- *Forecasting high-frequency financial time series: an adaptive learning approach with the order book data*, 2021
- *Using the yield curve to forecast economic growth*, *Journal of Forecasting*, 2020

Recent Presentations

- 2022 Sep: Data Science Seminar Series hosted by **Goldman Sachs (London)**

  My Slides

- 2022 Aug: Unconference Series hosted by **Oxford Machine Learning Summer School**

Parley Ruogu Yang

About
:: *LinkedIn*
:: *Google Scholar*

Teaching
:: *(LSE Postgrad Lent Term 2022) ST456 Deep Learning*
:: *(Cambridge Undergrad Summer Term 2022) Mathematical Foundation of Statistical Machine Learning*

Previous courseworks
:: *Cambridge first-year PhD courseworks*

https://parleyyang.github.io/ST

parleyyang.github.io/ST456/index.html

Deep Learning

Happy Deep Learning

ST456 @ LSE

LT 2022

Parley Ruogu Yang
Home

About
:: *LinkedIn*
:: *Google Scholar*

NOTE: This course has finished.

- **Weekly Blog Seris**
  Deep Learning with a tale of two cities

Blog I / IX: blogs from a Cambridge Mathematician
on teaching a Master course on Statistical Machine Learning at LSE

Blog II / IX: principle of ML and option pricing

Blog III / IX: down the hill of gradients

Blog IV / IX: convolve the pictures

Blog V / IX: wrapping up the Convolutions and reflecting upon Innovations

Blog VI / IX: the time

Blog VII / IX: time, coder, and generation

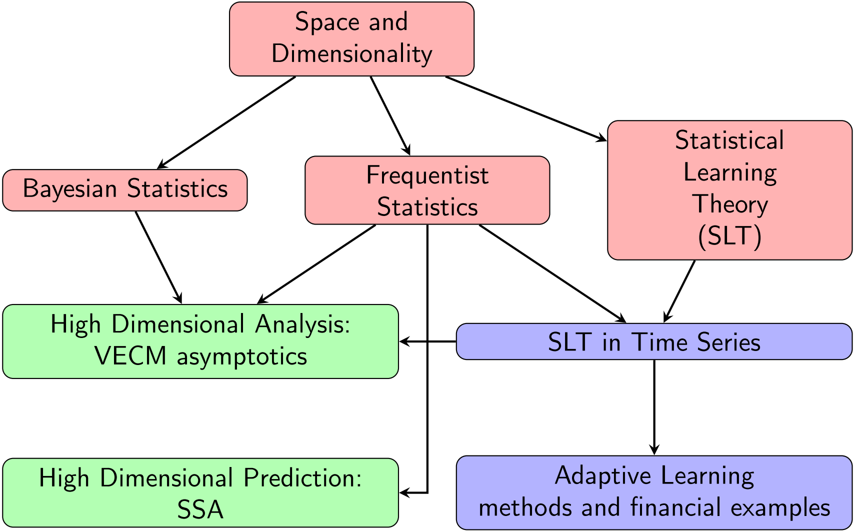
Interests

- Statistics
  - Time Series Forecasting
  - Interpretable Statistical Machine Learning
  - Application to Portfolio Management
- Mathematics
  - Functional Analysis
  - Mathematics of Machine Learning
  - History of Mathematics
- Economics
  - Modern Economic and Political History
  - Monetary Economics

What this talk is about

- A special appreciation on the mathematical development of space and dimensionality
- A number of research topics or literature I have engaged with.
- Some new methods and / or inference on time series:
    - Adaptive time series statistical methods
    - Big-data algorithms for time series analysis
- Some interesting applications:
    - Climate Statistics
    - Social Statistics
    - Macroeconomic Analytics
    - Financial Time Series

Plan


Space and Dimensionality
Bayesian Statistics
Frequentist Statistics
Statistical Learning Theory (SLT)
High Dimensional Analysis: VECM asymptotics
SLT in Time Series
High Dimensional Prediction: SSA
Adaptive Learning methods and financial examples

Be minded

- I will be using whiteboard (on the iPad)
- Questions and discussions are encouraged during the talk, try not to leave it at the end.

Background

- Space and Dimensionality: a historical perspective
- The Practice of Statistician: Bayesian vs Frequentist
- Statistical Learning Theory: an introduction and in a time series setting

Space and Dimensionality: a historical perspective

From Euclid to Hilbert: from $\mathbb{N}$ to $\mathbb{R}^{\mathbb{N}}$

Figure: Euclid, Descrates, Riesz, and Hilbert

Frequentist vs Bayesian

- Ronald Fisher (1925): sufficiency, efficiency, Fisher information, maximum likelihood theory

$$\hat{\theta}(D) = \arg\max_{\theta\in\Theta} l(D;\theta), \quad \hat{\theta}(D) \sim ?$$

- Thomas Bayes (1760s) and followers:

$$\pi_{post}(\theta|D) \propto l(D|\theta)\pi_{pri}(\theta)$$

  Applications include BIC (Schwarz 1978) and various Bayesian Machine Learning topics

- Reference: Efron and Hastie (2016, Ch.13 & Epilogue)

Frequentist vs Bayesian

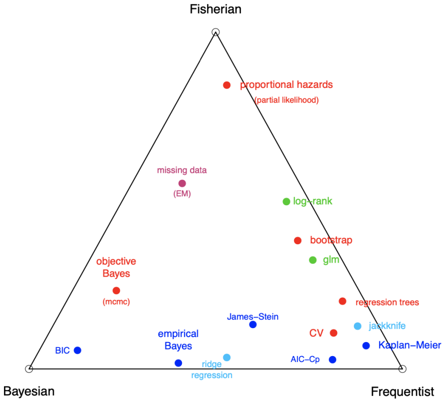


Figure 14.1 Bayesian, frequentist, and Fisherian influences, as described in the text, on 15 major topics, 1950s through 1990s. Colors indicate the importance of electronic computation in their development: red, crucial; violet, very important; green, important; light blue, less important; blue, negligible.

Statistical Learning Theory in a snapshot

- Input space $\mathbb{X}$ and output space $\mathbb{Y}$, often $\mathbb{X} = \mathbb{R}^p$ and $\mathbb{Y} = \mathbb{R}$
  Random Variables $X \in \mathbb{X}$, $Y \in \mathbb{Y}$
- Decision function $h : \mathbb{X} \to \mathbb{Y}$
- Loss function $l : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}$
- Risks $R(h) = \mathbb{E}[l(h(X), Y)]$
- Risk-minimisation: given a (large) set H, find $\arg\min_{h \in H} R(h)$
- Observation: squared loss $\implies$ regression
- Remark 1: sometimes $\mathbb{X} \subsetneq \mathbb{R}^p$ and more recently $\dim(\mathbb{X}) = \infty$ may also be analysed
- Remark 2: empirical risks $\hat{R}(h) = N^{-1} \sum_{i \in [N]} l(h(X_i), Y_i)$

Statistical Learning Theory in time series

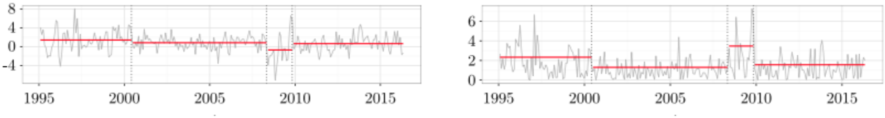
- For non-time-series, $(X, Y) \sim P_{X,Y} = P_{Y|X} \times P_X$
- In time series, we have $(X_t, Y_t) \sim P^t_{X,Y} = P^t_{Y|X} \times P^t_X$
- Forecasting: we try to learn $Y_{t+k}|(X_t, Y_t), (X_{t-1}, Y_t), \ldots$
- Remark 3: stationarity — a frequentist's concern
- Remark 4: more on k and the case of financial applications

Some relevant literature: change-point studies

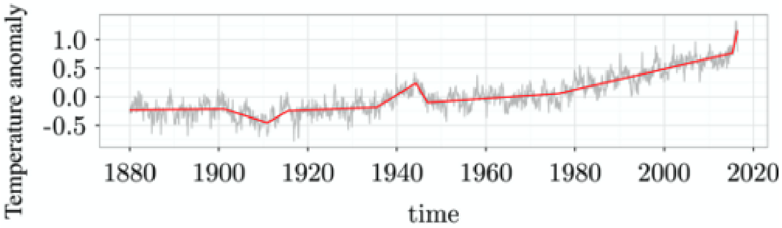
See whiteboard for maths demonstration.
Reference: Baranowski, Chen, and Fryzlewicz (2019)

Examples (Baranowski, Chen, & Fryzlewicz, 2019)

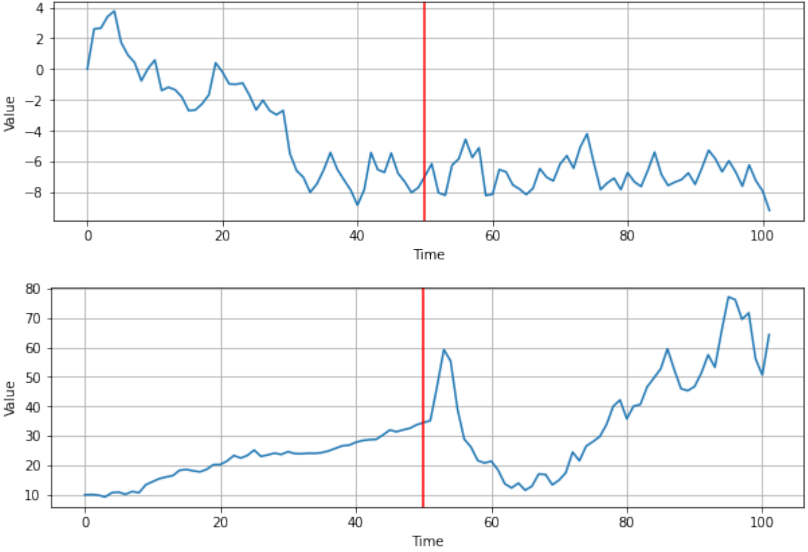
Example 1: London borough-level house price changes



Example 2: Global surface temperature anomalies



Stationarity and feature transformation problems

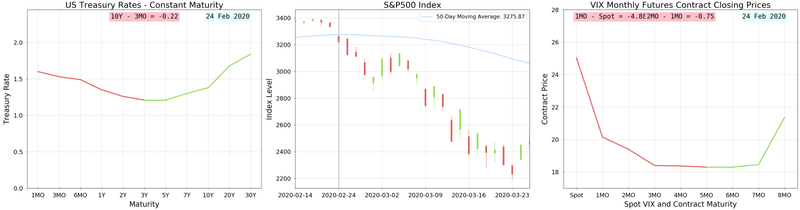


Pausing for any questions / discussions

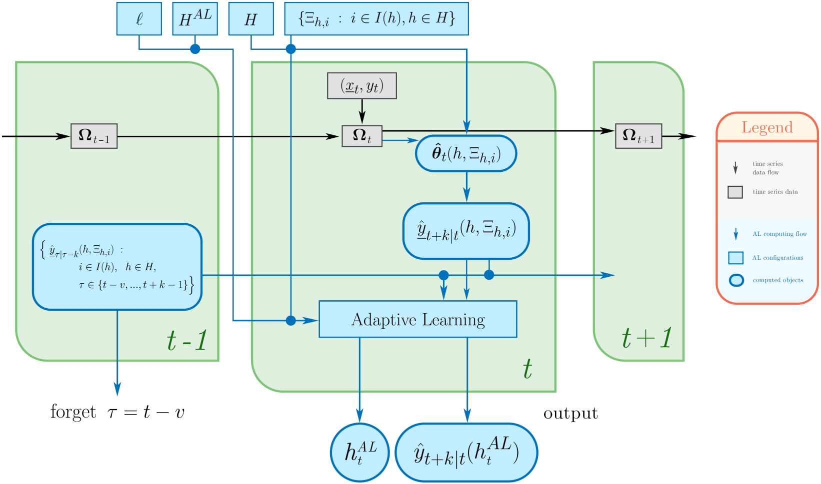
Motivation for Adaptive Learning: selection over time

- A financial example: https://optimalportfolio.github.io/subpages/Videos.html
- Problems at time t:
  - Temporary selection of variables and models
  - Temporary selection of estimation method
  - Temporary selection of forecasting method (out-of-sample)

Illustration of data utilised in Yang and Lucas (2022)



Method by computing graph



Algorithms as per Yang and Lucas (2022)

Algorithm 1: DMS

Input: Data, desired forecasting index set T, and hyperparameters $(\ell, H, \{\Xi_{h,i}\}_{i\in I(h),h\in H}, v)$

Output: Forecasts $\{\hat{y}_{t+k|t}(h_t^{DMS})\}_{t\in T}$ with the associated models $\{h_t^{DMS}\}_{t\in T}$

(1) For $t \in T$, repeat:
  (a) Evaluate ℓ given the information required. Then find $h^* \in H$ and $\Xi^*_{h,i}$ which minimises the loss.
  (b) Obtain and store $\hat{y}_{t+k|t}(h_t^{DMS}) := \hat{y}_{t+k|t}(h^*, \Xi^*_{h,i})$ as the forecast

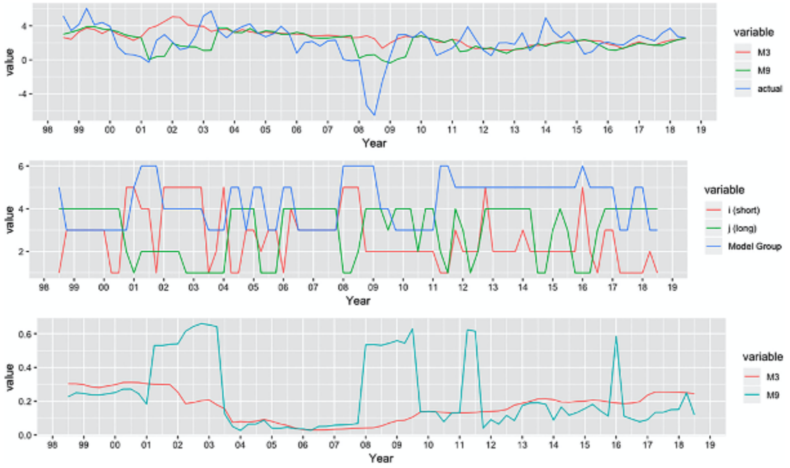
Algorithm 2: AE

Input: Data, desired forecasting index set T, and hyperparameters $(\ell, H, \{\Xi_{h,i}\}_{i\in I(h),h\in H}, v_0, v_1)$

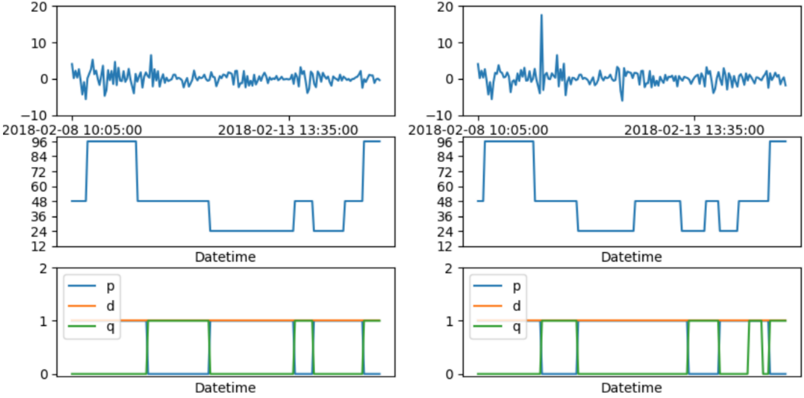
Output: Forecasts $\{\hat{y}_{t+k|t}(h_t^{AE})\}_{t\in T}$ with the associated models $\{h_t^{AE}\}_{t\in T}$

(1) Enumerate $\cup\{(h, \Xi_{h,i}) : i \in I(h), h \in H\}$ to $[M]$. For $t \in T$, repeat:
  (a) For $\tau \in \{t - v_0 + 1, ..., t\}$, repeat:
    (i) Evaluate ℓ given the information required. Then find $h^* \in H$ and $\Xi^*_{h,i}$ which minimises the loss.
    (ii) Allocate a weight of v_0^{-1} to the minimiser.
  (b) Collect the weight δ_t and align the forecast vector $\hat{y}^M_{t+k|t}$
  (c) Obtain and store $\hat{y}_{t+k|t}(h_t^{AE}) = \langle \delta_t, \hat{y}^M_{t+k|t}\rangle$ as the forecast

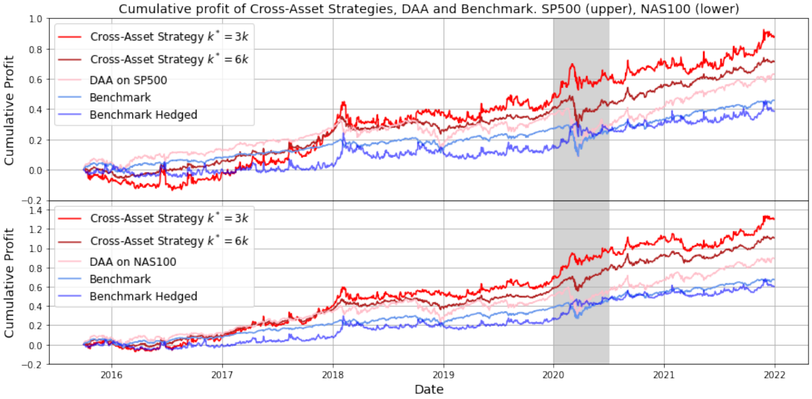
Result: Macroeconomic Analytics (Yang, 2020)



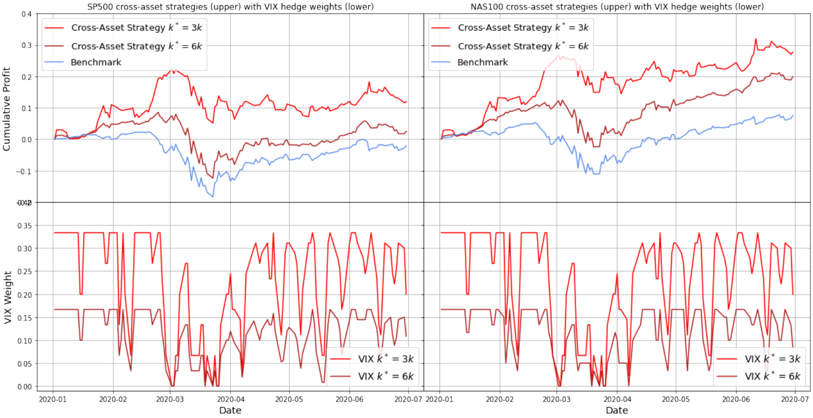
Result: Learning with regularisation over time (Yang, 2021)



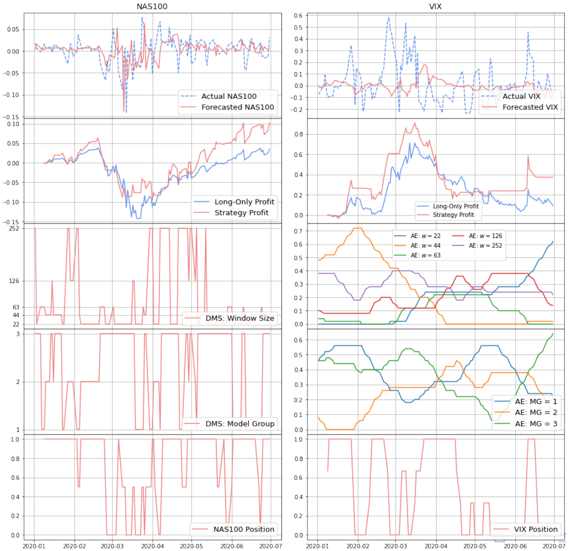
Result: Financial Time Series — Portfolio Management (Yang & Lucas, 2022)



Portfolio Management zoomed into 2020 Q1 and Q2



Result: Financial Time Series — Model Analytics (Yang & Lucas, 2022)



More on loss functions

- Generalised notion of aggregated loss over time (Yang & Lucas, 2022)

$$\ell(h, \Xi_{h,i}; \lambda, p) := \sum_{\tau=t-v+1}^{t} \lambda^{t-\tau} ||\hat{\boldsymbol{y}}_{\tau|\tau-k} - y_\tau 1_k||_p^p \qquad (1)$$

- Functional awards and penalties (Yang, 2021)

$$\ell^{\mathtt{total}}(h, \ldots, H \setminus \{h\}) = \hat{R}(h, \ldots) + D(h, h^*_{t-1}) \qquad (2)$$

- Call for further analysis (asymtptoics, inference, etc) on these

Pausing for any questions / discussions

Motivation in a VECM flavour

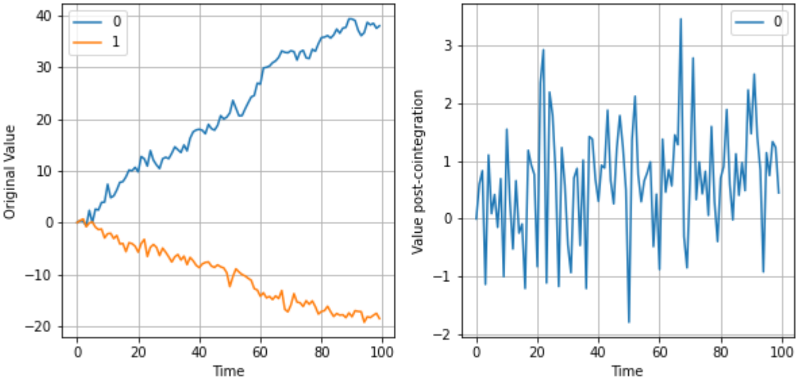


Figure: 2-dimensional cointegration

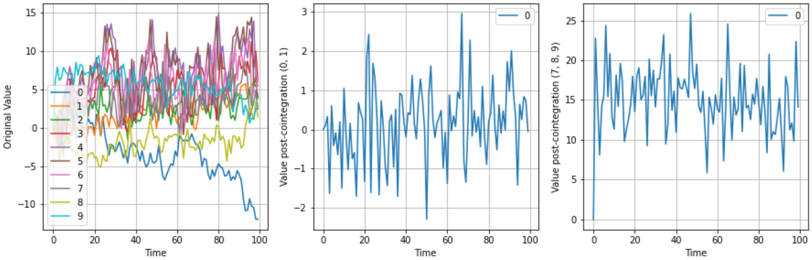


Figure: high-dimensional cointegration

Problem 'statement': let $Y_t \in \mathbb{R}^m$. Given

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i \in [p]} B_i \Delta Y_{t-i} + \varepsilon_t \tag{3}$$

'Understand' $r := rank(\Pi)$

Interactive question

Why 'Understand' but not 'Estimate'?

Bayesian and Frequentist Approaches

- Bayesian approach

$$\pi_{post}(\theta_-, r|D) \propto l(D|\theta_-, r)\pi_{pri,1}(\theta_-|r)\pi_{pri,2}(r) \quad (4)$$

$$\pi_*(r|D) \propto l_*(D|r)\pi_{pri,2}(r) \quad (5)$$

$$l_*(D|r) = \int l(D|\theta_-, r)\pi_{pri,1}(\theta_-|r)d\theta_- \quad (6)$$

References: Villani (2005) and Koop, Leon-Gonzalez, and Strachan (2011) for extension on dynamic VECM.

- Frequentist approach

$$\hat{R} = \arg\min_{R\in UT(m)} \sum_{t\in[T]} ||e_t||_2^2 + \sum_{k\in[m]} \frac{\lambda}{\tilde{\mu}_k^\gamma}||R(k,\cdot)||_2 \quad (7)$$

$$\hat{r} = rank(\hat{R}) \quad (8)$$

References: Liang and Schienle (2019) and very recently Chen and Schienle (2022).

Modern asymptotics on VECM

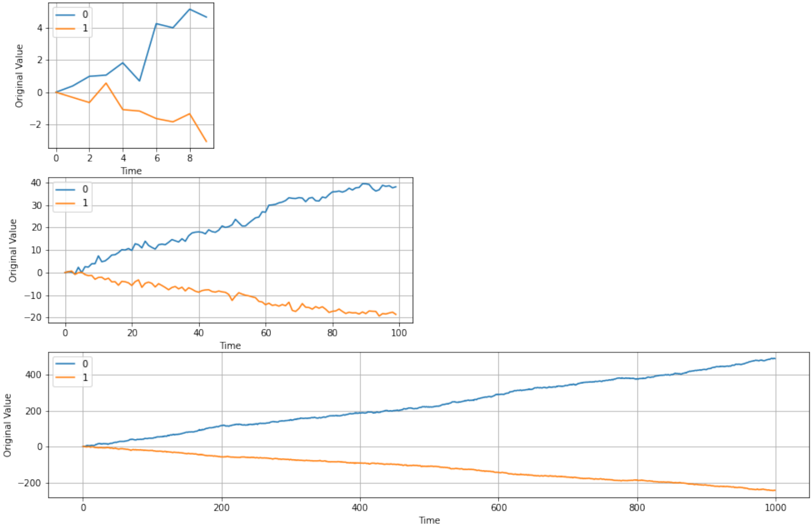


Figure: Classical asymptotics: $T \to \infty$

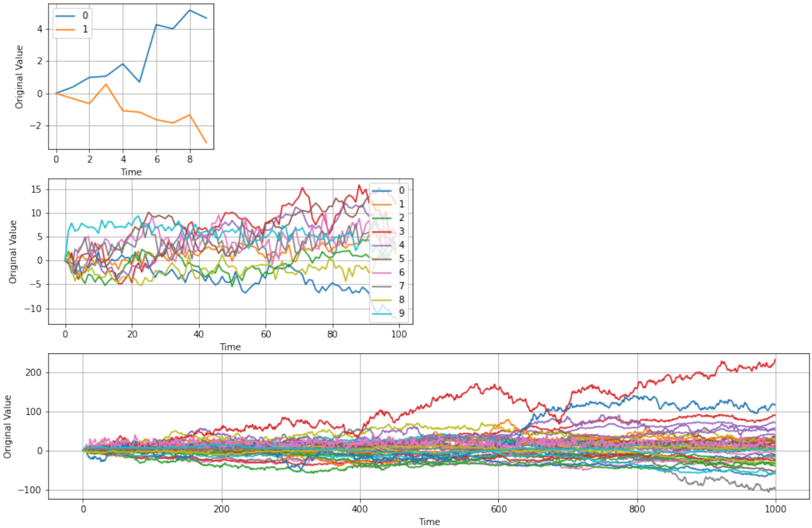
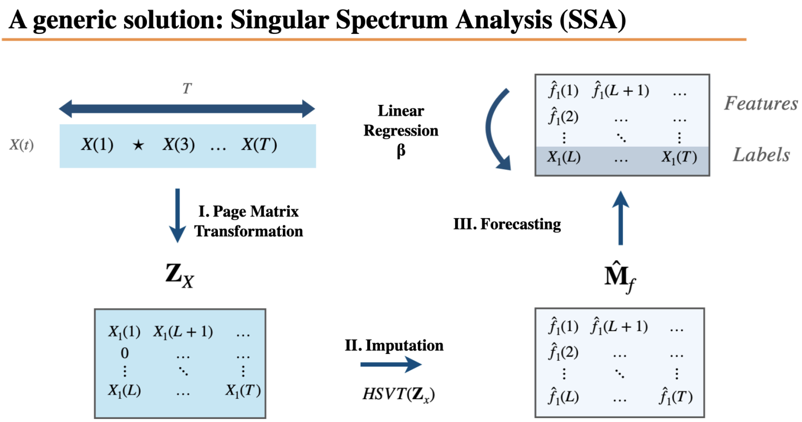


Figure: Big Data asymptotics: $T \to \infty$ while $m = \mathcal{O}(T)$

Another aspect of big data: Singular Spectrum Analysis (Agarwal, Alomar, & Shah, 2022)



More on the future

- Hilbert space machine learning
- Idea (could be sketchy):
  from

$$\arg\min_{\beta\in\mathbb{R}^m} \hat{R}(\beta) + \lambda||\beta||$$

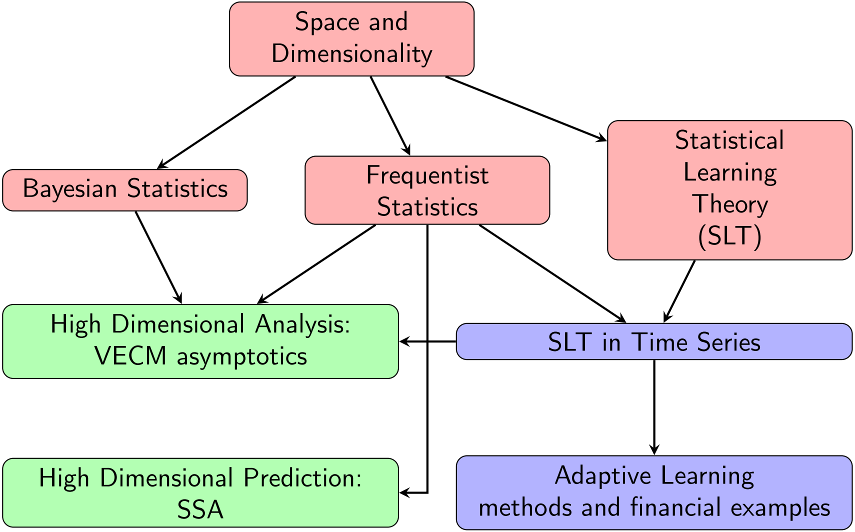
  to Reproducible Kernel Hilbert Space optimisation

$$h^* = \arg\min_{h\in\mathcal{H}} \hat{R}(h) + \Omega(h)$$

  and generalised notions of loss design

$$\arg\min_{\hat{R}\in\mathcal{R}} R(h^*(\hat{R}))$$

The End


Space and Dimensionality
Bayesian Statistics
Frequentist Statistics
Statistical Learning Theory (SLT)
High Dimensional Analysis: VECM asymptotics
SLT in Time Series
High Dimensional Prediction: SSA
Adaptive Learning methods and financial examples

Thank You

References I

Agarwal, A., Alomar, A., & Shah, D. (2022). On multivariate singular spectrum analysis and its variants. https://doi.org/10.48550/ARXIV.2006.13448

Baranowski, R., Chen, Y., & Fryzlewicz, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *81*(3), 649–672.

Chen, S., & Schienle, M. (2022). Large spillover networks of nonstationary systems. *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2022.2099870

Efron, B., & Hastie, T. (2016). *Computer age statistical inference : Algorithms, evidence, and data science*. CUP.

Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2011). Bayesian inference in a time varying cointegration model. *Journal of Econometrics*, *165*(2), 210–220.

References II

Liang, C., & Schienle, M. (2019). Determination of vector error correction models in high dimensions. *Journal of Econometrics, 208*(2), 418–441.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory*, *21*(2), 326–357. https://doi.org/10.1017/S026646660505019X

Yang, P. R. (2020). Using the yield curve to forecast economic growth. *Journal of Forecasting*, *39*(7), 1057–1080. https://doi.org/10.1002/for.2676

Yang, P. R. (2021). Forecasting high-frequency financial time series: An adaptive learning approach with the order book data. *arXiv:2103.00264*.

Yang, P. R., & Lucas, R. (2022). DMS, AE, DAA: Methods and applications of adaptive time series model selection, ensemble, and financial evaluation. *arXiv:2110.11156v3*.