Topics in Modern Time Series: High Dimensionality, Modelling, and Applications

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

Oxford Machine Learning Summer School Unconference Series

4 Aug 2022

What this talk is not about

- I am not going to discuss one particular paper in depth.
- I am not going to discuss one sole empirical example in depth.
- I am not going to introduce one particular methodology in depth.

What this talk is about

- A number of research topics I have been working on.
- A collection of literature that I engage in.
- Some new methods and / or inference on time series:
  - Adaptive time series methods
  - Change-points
  - Big-data algorithms: PCR, SSA, and more
- Some interesting applications:
  - Climate Statistics
  - Macroeconomic Analytics
  - Financial Time Series

Remarks

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

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: empirical risks $\hat{R}(h) = N^{-1} \sum_{i \in [N]} l(h(X_i), Y_i)$

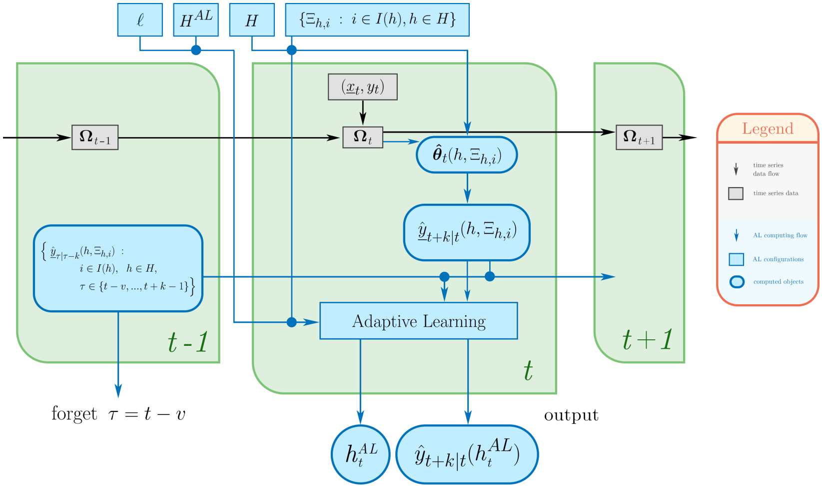
Statistical Learning Theory in time series

- 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: Bayesian? Maybe, the slides here are mostly frequentists' , but suggestions / discussions are welcome.

Motivation: variable 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)

Method by computing graph



Method by 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:

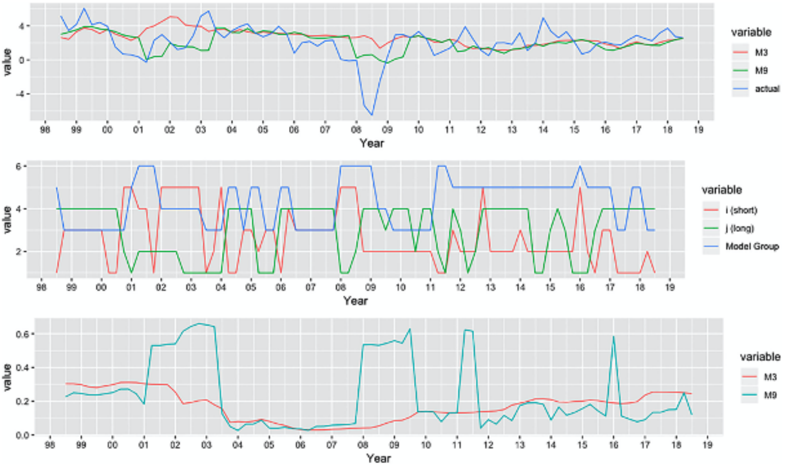
(i) Evaluate ℓ given the information required. Then find $h^* \in H$ and $\Xi^*_{h,i}$ which minimises the loss.

(ii) Allocate a weight of v_0^{-1} to the minimiser.

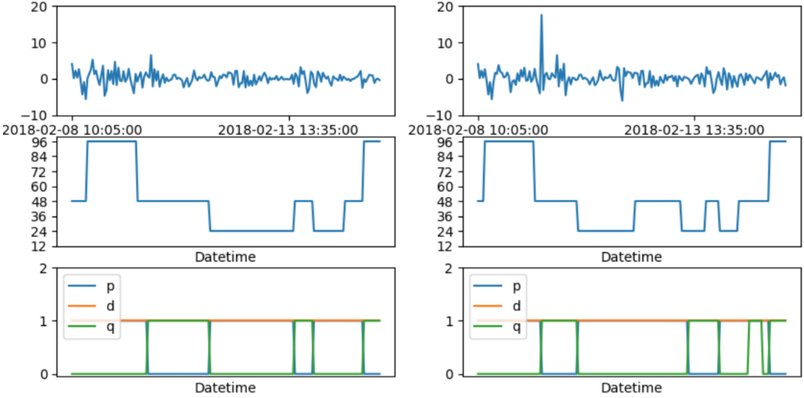
(b) Collect the weight δ_t and align the forecast vector $\hat{y}^M_{t+k|t}$

(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

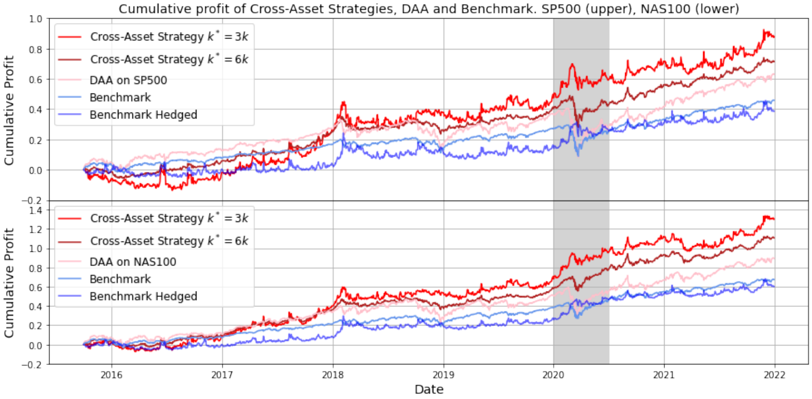
Result: Macroeconomic Analytics (Yang, 2020)



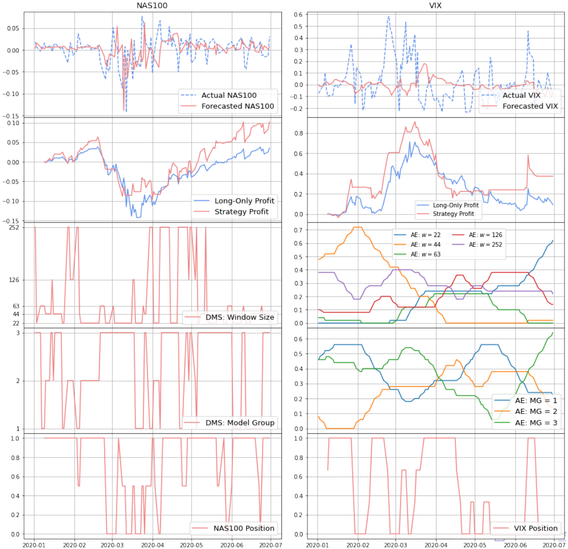
Result: Learning with regularisation over time (Yang, 2021)



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



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)

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

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

An example (Baranowski, Chen, & Fryzlewicz, 2019)

Specifications: for $t \in \{\tau_{j-1}+1, ..., \tau_j\}$ and $j \in [q+1]$

- Piecewise constant variance, piecewise constant mean:

$$X_t = \theta_{j,1} + \theta_{j,2}\varepsilon_t \tag{3}$$

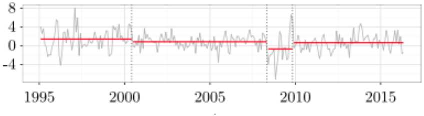
- Constant variance σ_0 fixed with continuous and piecewise linear mean:

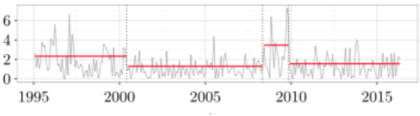
$$X_t = \theta_{j,1} + \theta_{j,2}t + \sigma_0\varepsilon_t \tag{4}$$

where $\varepsilon_t \overset{iid}{\sim} N(0,1) \forall t$, $\theta_j := (\theta_{j,1}, \theta_{j,2})$ satisfies $\theta_j \neq \theta_{j-1}$ for all j, and that we aim to search for an unknown amount (q) of change-points noted $\tau_1, ..., \tau_q$ with $\tau_0 = 0$, $\tau_{q+1} = T$.

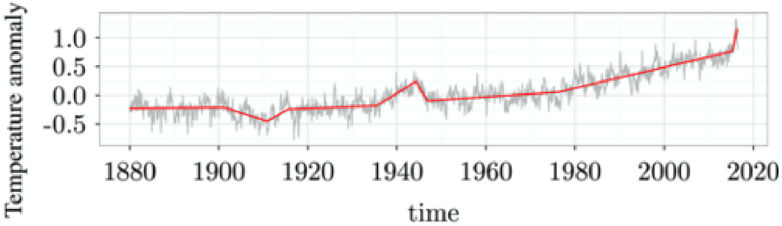
Example continued (Baranowski, Chen, & Fryzlewicz, 2019)

Example 1: London borough-level house price data (equation 3)

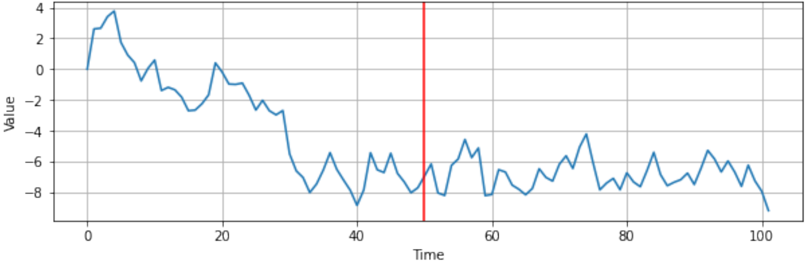


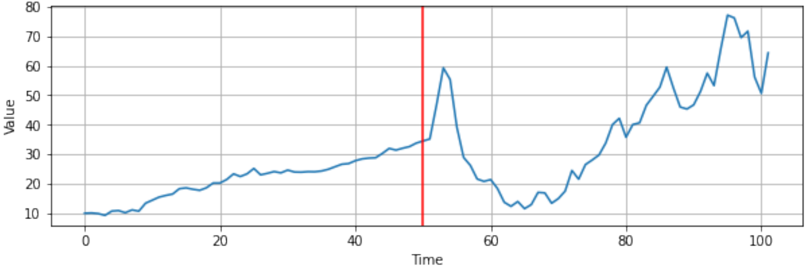


Example 2: Global surface temperature anomalies data (equation 4)



Stationarity and feature transformation problems





Principle Components Regression (PCR)

- Let Z be a $N \times p$ matrix, then SVD transforms $Z = \sum_{i=1}^{N} s_i u_i v_i^T = USV^T$ where $s_1 \geq s_2 \geq \ldots \geq s_N$
- Fix $k \in [N]$, then annotate $V_k := [v_1, \ldots, v_k]$,

$$Z^{PCR,k} := ZV_k \tag{5}$$

$$\hat{\beta}^{PCR,k} := \arg\min_{\beta \in \mathbb{R}^k} \sum_{i \in I} (Y_i - Z_i^{PCR,k} \beta)^2 \tag{6}$$

$$= \left(\sum_{i=1}^{k} s_i^{-1} v_i u_i^T\right) Y \tag{7}$$

Properties of PCR

- Equivalence with Hard Singular Value Thresholding (HSVT)

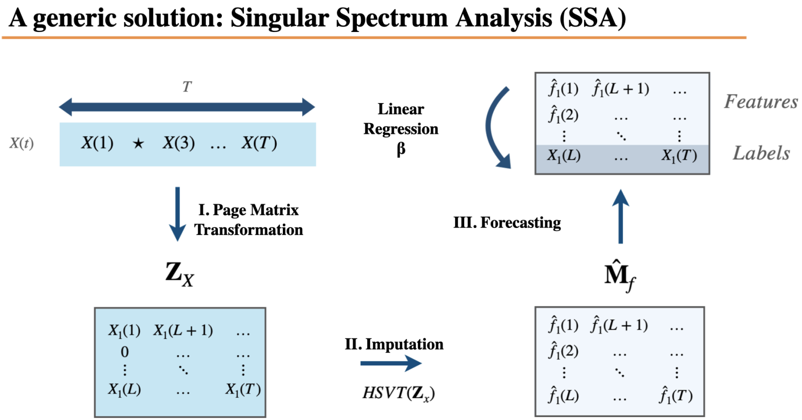
$$Z^{HSVT,k} := \sum_{i=1}^{N} \mathbb{1}[s_i \geq s_k] s_i u_i v_i^T$$

$$\hat{\beta}^{HSVT,k} := \arg\min_{\beta \in \mathbb{R}^k} \sum_{i \in I} (Y_i - Z_i^{HSVT,k} \beta)^2$$

$$Z^{HSVT,k} \hat{\beta}^{HSVT,k} = Z^{PCR,k} \hat{\beta}^{PCR,k}$$

- Key works: Agarwal et al. (2019) and Agarwal, Shah, and Shen (2020) show bounds on $||\hat{\beta}^{PCR,k} - \beta||$ and prediction and forecasting errors.

Singular Spectrum Analysis (Agarwal, Alomar, & Shah, 2022)



Special acknowledgement to Anish Agarwal for sharing this slide with me

More on the future

- Tree-based algorithms with robust statistics (if time allows)
- Hilbert space machine learning
- Idea (could be sketchy):
  from

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

  to Reproducible Kernel Hilbert Space optimisation

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

  and generalised notions of loss design

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

References

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

Agarwal, A., Shah, D., & Shen, D. (2020). On principal component regression in a high-dimensional error-in-variables setting. *arXiv:2010.14449.*

Agarwal, A., Shah, D., Shen, D., & Song, D. (2019). On robustness of principal component regression. *Advances in Neural Information Processing Systems, 32.*

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.

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.*