Adaptive Learning on Time Series: Method and Financial Applications

Parley Ruogu Yang[1], Ryan Lucas[2] and Camilla Schelpe[3]

24 November 2021

Isaac Newton Institute, Cambridge

[1] CCIMI, University of Cambridge and OPRG. Acknowledgement to Raena McElwee (UCD and OPRG) for visualisations and videos.

[2] University College Dublin and OPRG

[3] OPRG

References:

[1] Yang, Parley Ruogu (2020). Using The Yield Curve To Forecast Economic Growth. *Journal of Forecasting*. 2020; 39: 1057– 1080. https://doi.org/10.1002/for.2676

[2] Yang, Parley Ruogu (2021). Forecasting High-Frequency Financial Time Series: An Adaptive Learning Approach With the Order Book Data. https://arxiv.org/abs/2103.00264

[3] Yang, Parley Ruogu, Ryan Lucas, and Camilla Schelpe (2021). Adaptive Learning on Time Series: Method and Financial Applications https://arxiv.org/abs/2110.11156

Time-varying distribution

Background

- Data $\{(x_t, y_t) : t \in [T]\}$
- Data Generating Process $y_t \sim P^t_{Y|X}(\cdot|x_t;\theta)\ \forall t \in [T]$
- Ordinary statistics: $P^1_{Y|X} = P^2_{Y|X} = \ldots = P^t_{Y|X} = P^{t+1}_{Y|X} = \ldots$
- Time-varying distribution: $P^t_{Y|X} \neq P^{t+1}_{Y|X}$ for many $t \in [T]$

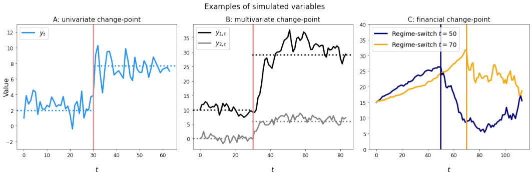
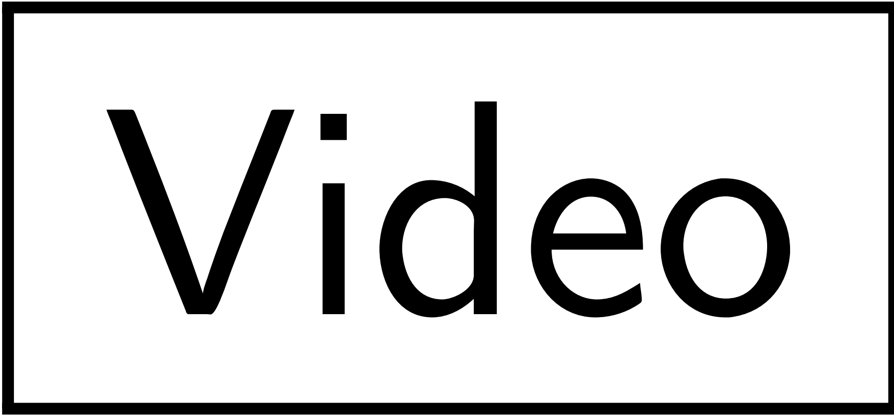


Figure: Simulated change point systems studied in [3]

Problems

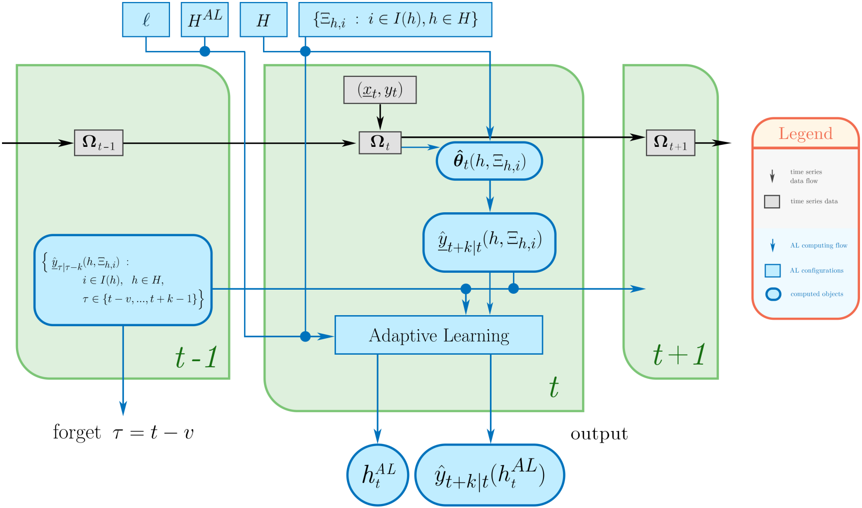
- Modelling & Forecast: which model and estimation shall be used?
- Evaluation: how to evaluate the performance of models under time-varying distribution?

Example: VIX, Yield Curve and S&P 500 during 2020



For full video, visit
https://optimalportfolio.github.io/subpages/Videos.html

Algorithmic introduction



Loss functions

- Basic forms used in [1, 2, 3]:

$$\ell^{\texttt{Norm, single-valued}}(h, \Xi_{h,i}; \lambda, p) := \sum_{\tau=t-v+1}^{t} \lambda^{t-\tau} |\hat{y}_{\tau|\tau-k} - y_\tau|^p$$

  where v is a hyperparameter for window size
- Basic forms with tricks used in [3]:

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

  where $\hat{\boldsymbol{y}}_{\tau|\tau-k} = (\hat{y}_{\tau|\tau-1}, \hat{y}_{\tau|\tau-2}, \ldots, \hat{y}_{\tau|\tau-k})$
- Interesting Penalisation used in [2]:

$$\ell^{\texttt{Penalised}}(h, \Xi_{h,i}; \lambda, p) := \ell^{\texttt{Global}}(h, \Xi_{h,i}; \lambda, p) + D(h, h^*_{t-1})$$

  where D measures distance, e.g. in terms of complexity and parametric dimensions
  Remark: looks like fused / adaptive Lasso.

Improved forecasts through ensemble

k	Type	Configuration	$1000 \times$ MSE	CS	SR	$100 \times$ ANR	MDD
3	AL	Ensemble-MC	**2.984**	**0.500**	**-0.958**	**-51.591**	**-0.220**
	Fixed	MG3T, AR(4), $w = 252$, VIX 3-8m	4.102	0.409	-2.081	-81.887	-0.224
5	AL	Ensemble-MC	**7.067**	0.515	**-1.415**	**-70.035**	**-0.214**
	Fixed	MG3N, $w = 252$, VIX 3-6m	11.526	**0.529**	-2.264	-100.219	-0.288

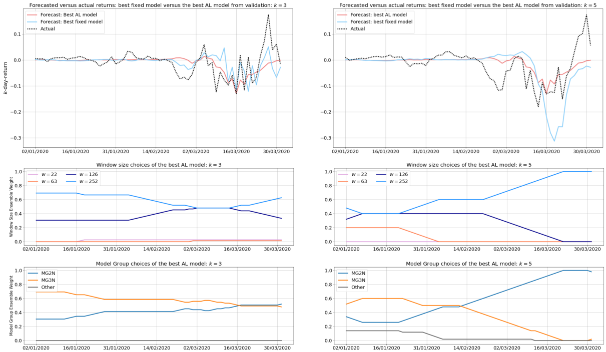


Figure: 2020 Q1 testing performance, models selected based on data in 2019 [3]

Statistical testing

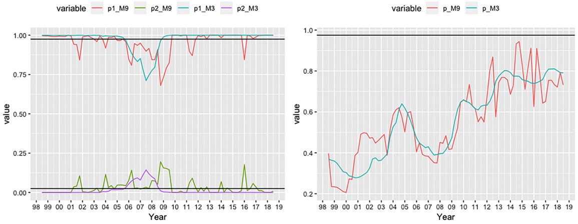


Figure: Standard parametric testing [1]

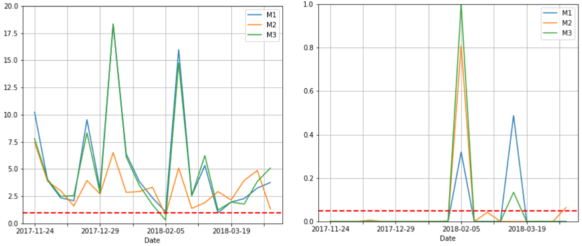


Figure: Model testing: Bayesian Factor (left) and Frequentist (right) [2]

Evaluation in Financial Time Series: the failure of MSE

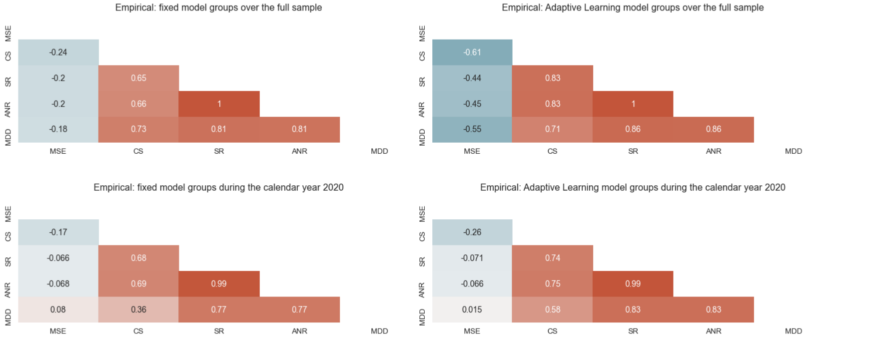


Figure: Correlation amongst statistical (MSE, CS) and financial (SR, ANR, and MDD) metrics [3]

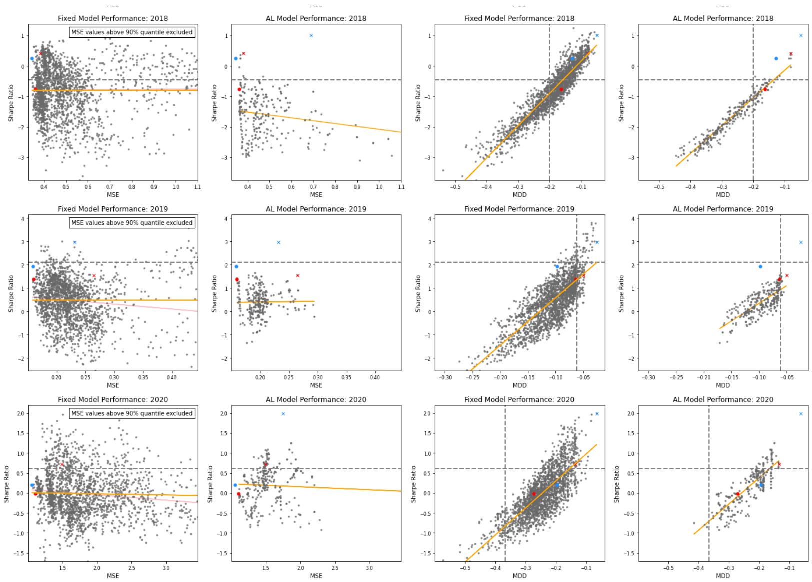


Figure: Time-varying relationships amongst MSE, SR, and MDD [3]

Thank you for listening!
See you at the Poster Session!
For further information: https://parleyyang.github.io