Adaptive Learning on Time Series: Method and Financial Applications DRAFT

Parley Ruogu Yang

https://parleyyang.github.io/
Faculty of Mathematics, University of Cambridge
&
Department of Statistics, London School of Economics and Political Science

This version:	27th March 2022
Presentation time:	2pm, 28 March 2022
Presentation venue:	Department of Statistics, Oxford, UK

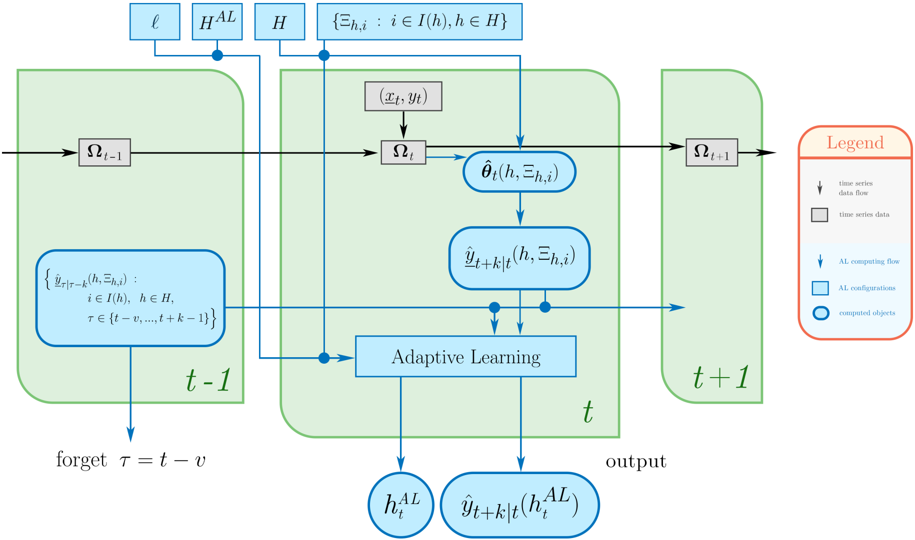
Initial Remarks

- Thanks for coming.
- Good to be back in person and in Oxford.
- Some maths will be noted on the whiteboard. Same information are available on the slides (Appendix) and the paper.
- This talk contains ongoing works. Currently revising the paper with focus on Empirical / Financial Time Series. (See the bibliography page for existing papers / pre-prints)
  - Many extensions are still Work-In-Progress. Do get in touch if you're interested to collaborate.
  - It will be great to have feedbacks during the talk and at the end!

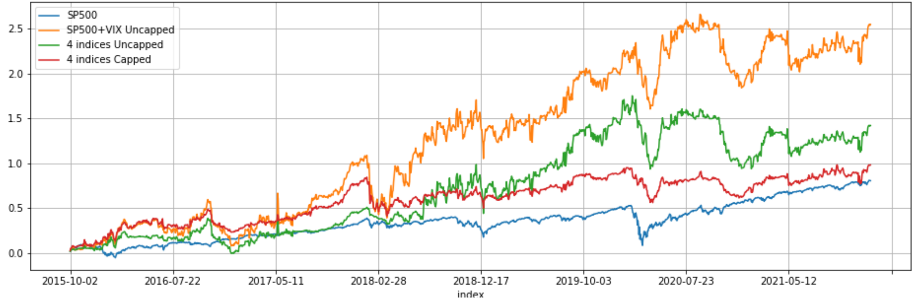
Plan of today's talk

1. Initial Empirical Investigation
   - Data: VIX, Yield, Asset Returns
   - Financial evaluations on standard models
   - This motivates the theoretical thoughts.
2. Motivation and Thoughts on Modern Time Series
   - Adaptive Learning: Theory and Algorithm
   - Loss functions specifications
3. Results
   - Overall performance
   - Dynamic Asset Allocation
4. Further Extensions

Highlights: Methodology



Highlights: Financial Applications



Contents

Return calculation

- Let p_t^A be the price of asset A at time t, define k-days ahead return as

$$r^A_{t:(t+k)} = \log(p^A_{t+k}) - \log(p^A_t) \tag{1}$$

- Remarks:
  - $r^A_{t:(t+k)}$ is only known after time $t+k$. When k and A are fixed, we write $y_{t+k} := r^A_{t:(t+k)}$
  - The choices of A are:
    - SP500 index
    - CBOE VIX index (note this is not usually tradable)
    - NASDAQ100 index
    - DJIA30 index

Illustrations: Baseline Cumulative Returns

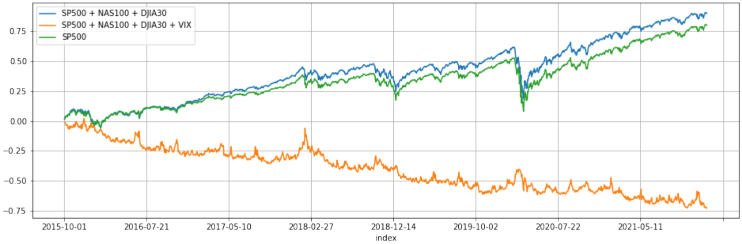


Figure: SP500 single-index long-only strategy and Equally-weighted strategies across 3 or 4 indices

Key problems: selection and combinations

- For a given (k, A), our target variable is

$$y_t = r^A_{(t-k):t} = \log(p^A_t) - \log(p^A_{t-k})$$

- Note that for each pair (k, A), y_t is defined differently.
- We want to forecast y_{t+k} at time t.
- Later on, we use Adaptive Learning (AL) to address the problem of dynamic model selection (DMS) and forecast combinations (Ensemble).

Key problems: evaluation and decision

- The goodness of forecast is assessed by
  - Statistical metrics: MSE and Percentage Correct (PC) of sign predictions
  - Financial statistics (Sharpe Ratio, Annualised Returns, and Max Drawdown) from the induced trading strategy: at time t, hold w_t of asset A where

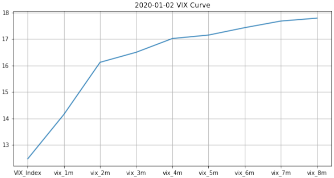
$$w_t^A := \frac{1}{k}\sum_{j=0}^{k-1}(\mathbb{1}[\hat{y}^A_{t+k-j|t-j} > 0] - \mathbb{1}[\hat{y}^A_{t+k-j|t-j} < 0]) \tag{2}$$

- Induced decisions on asset allocation
  - Which k to use / what portion to allocate?
  - What portion to allocate amongst A?

Data Information

- Daily frequency of asset prices / interest rates
- Data Range: Start of Year 2013 to End of Year 2021
- Evaluation Period: 1st August 2014 to End of Year 2021

Illustrations: Curves



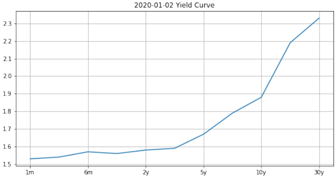
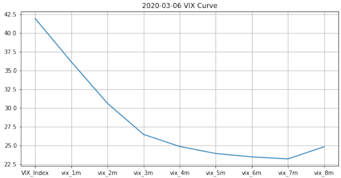


Figure: Pair of VIX and Yield Curves on a normal-day

Motivation towards curve slopes



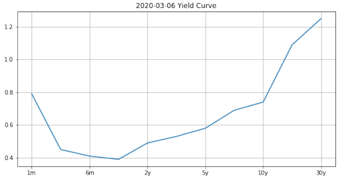


Figure: Pair of VIX and Yield Curves on the day before the first triggering of circuit breaker since 1997

- The trading day right before the 'Black Monday' in 2020, when a trading curb was triggered due to the global pandemic and its implications.
- Animated charts during those period are available at `https://optimalportfolio.github.io/subpages/Videos.html`

Term structures: intro

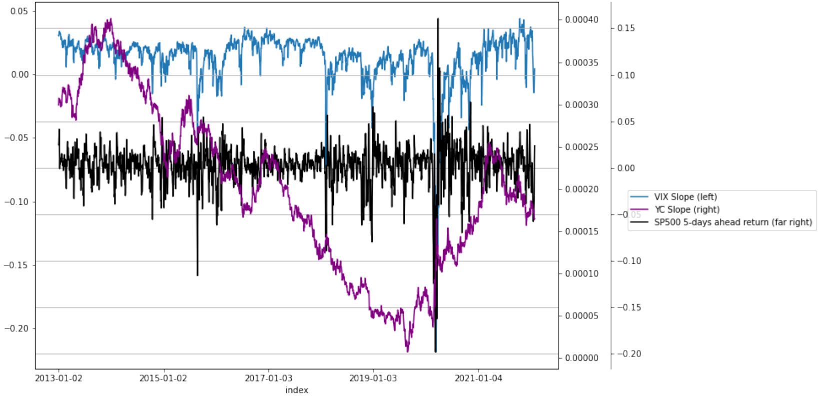
- At any time t, we have J futures: $\{(p_{t,j}, m_{t,j}) : j \in [J]\}$ where $p_{t,j}$ indicates the price of future with maturity $m_{t,j}$. Run regression

$$p_{t,j} = \alpha_t + \beta_t m_{t,j} + \varepsilon_{t,j} \quad \varepsilon_{t,j} \sim N(0, \sigma_t^2) \;\; \forall j \in [J] \tag{3}$$

- β_t is known as the slope of the curve.

Appendix: Data information

Illustrations: Full Sample Series



Fixed Model Class (MC)

We do rolling forecast on a w-windowed dataset.

- Class 1: AR(p) on returns

$$y_t = \alpha + \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t \tag{4}$$

  Note AR(0) means a constant model.

- Class 2: lagged linear regression with slope or spread (denoted s_t at time t)

$$y_t = \alpha + \beta s_{t-k} + \varepsilon_t \tag{5}$$

- Class 3: lagged linear regression with a pair of short-long rates (denoted $(short_t, long_t)$ at time t)

$$y_t = \alpha + \beta_1 short_{t-k} + \beta_2 long_{t-k} + \varepsilon_t \tag{6}$$

Fixed Model Results: Statistical vs Financial

k	SP500	VIX_Index	NAS100	DJIA30
1	MC1_(1, 252)	MC1_(0, 252)	MC1_(1, 252)	MC1_(1, 252)
2	MC1_(0, 252)	MC2_('VC_slope', 252)	MC1_(0, 252)	MC1_(0, 252)
3	MC1_(0, 252)	MC2_('V0-1', 252)	MC1_(0, 252)	MC1_(0, 252)
4	MC1_(0, 252)	MC2_('V0-1', 252)	MC1_(0, 252)	MC1_(0, 252)
5	MC1_(0, 252)	MC1_(0, 252)	MC1_(0, 252)	MC1_(0, 252)

k	SP500	VIX_Index	NAS100	DJIA30
1	MC3_126_('2y', '7y')	MC3_252_('VIX_Index', 'vix_6m')	MC1_(1, 252)	MC3_w63_('2y', '10y')
2	MC3_126_('2y', '7y')	MC3_126_('VIX_Index', 'vix_5m')	MC3_252_('2y', '30y')	MC3_w63_('2y', '3y')
3	MC3_252_('2y', '30y')	MC3_126_('VIX_Index', 'vix_6m')	MC3_252_('2y', '30y')	MC3_w63_('1m', '5y')
4	MC3_126_('2y', '30y')	MC3_252_('VIX_Index', 'vix_6m')	MC3_252_('3m', '5y')	MC3_252_('3m', '20y')
5	MC3_126_('2y', '7y')	MC3_252_('VIX_Index', 'vix_6m')	MC3_w63_('1m', '3y')	MC3_126_('2y', '5y')

Figure: Best Models by MSE (top) and SR (bottom)

$AR(0, 252)$ model: failed to perform in returns compared to various alternatives

Section conclusion

- A taste of the data / EDA
  - Divergence of results between different metrics
  - $AR(0)$ with window size 252 is commonly recognised as a 'good model' in terms of MSE, while others outperform in financial metrics
- 'A posteriori' type of evaluation method — not so valid in TS or Statistical Learning
  - Later on we consider dynamic evaluation, by engaging with asset allocation.
- A financial note: VIX index is not usually tradable. Practically one trades VIX futures. No trading frictions are considered.

Background: Statistical Learning

See whiteboard / appendix

Appendix: Background: Statistical Learning

Background: Time Series in a Statistical Learning context

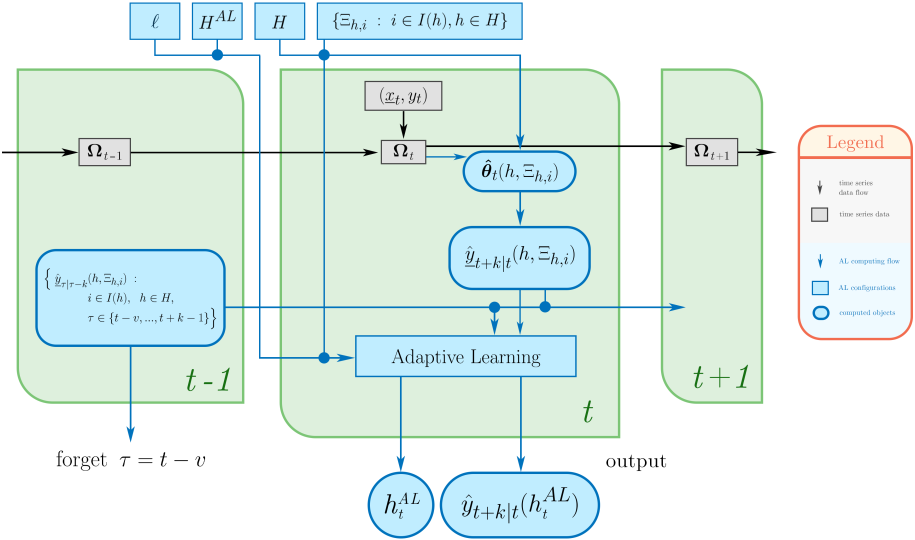
- Time-varying distributions: $P^t_{Y|X} \neq P^{t-1}_{Y|X}$
  - Motivates DMS to select model dynamically over time.
- Best model not computable, often due to information restriction: $g^{reg} \notin \mathcal{G}$
  - Motivates Ensemble of various 'top' models.
- Evaluation metrics of Financial Time Series
  - 'User-algorithm interaction' of loss formation and model choice. Further potential extensions to Automated Statistician / Automated TS Analyst.

Background: The value of forecasting

See whiteboard / appendix

Appendix: Background: The value of forecasting

Definition of AL via a computing graph



Algorithmic Example 1: Dynamic Model Selection ($H^{AL} = H$)

Mathematical definition & implementation:
see whiteboard / appendix

Appendix: Definition of AL Appendix: Implementation

Algorithm 1: AL implementation for DMS

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

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

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 $\hat{y}_{t+k|t}(h_t^{AL}) := \hat{y}_{t+k|t}(h^*, \Xi^*_{h,i})$ as the forecast

Figure: Dynamic Model Selection

Algorithmic Example 2: Ensemble ($H^{AL} = \Delta(N)$)

as desired. See algorithm 2 for a general algorithm.

Algorithm 2: AL implementation for Ensemble

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

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

1. Enumerate H to $[N]$.
2. 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 weights δ_t and align the forecast vector $\hat{y}^N_{t+k|t}$
   (c) Obtain and store $\hat{y}_{t+k|t}(h_t^{AL}) = \langle \delta_t, \hat{y}^N_{t+k|t} \rangle$ as the forecast

Figure: Ensemble

Summary of Results: MSE normalised to AR0

Appendix: MSE without normalisation

	SP500			VIX_Index			DJIA30			NAS100		
	Unrestricted Learning	Restricted Learning: AR	AR0	Unrestricted Learning	Restricted Learning: AR	AR0	Unrestricted Learning	Restricted Learning: AR	AR0	Unrestricted Learning	Restricted Learning: AR	AR0
1	1.096	1.072	1.000	1.047	1.013	1.000	1.080	1.068	1.000	1.070	1.052	1.000
2	1.072	1.048	1.000	1.032	1.024	1.000	1.080	1.051	1.000	1.060	1.030	1.000
3	1.142	1.049	1.000	1.061	1.026	1.000	1.089	1.057	1.000	1.117	1.040	1.000
4	1.140	1.042	1.000	1.102	1.023	1.000	1.154	1.056	1.000	1.113	1.032	1.000
5	1.151	1.045	1.000	1.138	1.032	1.000	1.158	1.052	1.000	1.132	1.033	1.000

Figure: Best Models by MSE

Summary of Results: Sharpe Ratio

	Unrestricted Learning				Restricted Learning to AR				AR0			
	SP500	VIX_Index	NAS100	DJIA30	SP500	VIX_Index	NAS100	DJIA30	SP500	VIX_Index	NAS100	DJIA30
1	0.759	0.926	0.545	0.693	0.423	0.131	0.444	0.335	0.014	0.157	0.199	0.032
2	0.712	1.299	0.513	0.482	0.021	0.677	0.646	-0.033	0.065	-1.039	0.054	0.094
3	0.690	1.176	0.740	0.603	0.454	0.602	0.878	0.186	-0.006	-0.970	0.049	-0.002
4	0.784	1.006	0.615	0.539	0.542	0.263	0.685	0.233	-0.012	-1.273	0.031	0.008
5	0.715	0.701	0.680	0.347	0.593	0.585	0.775	0.363	-0.035	-1.354	0.013	-0.020

Figure: Best Models by Sharpe Ratio

Summary of Results: Max Drawdown

	Unrestricted Learning				Restricted Learning				AR0			
	SP500	VIX_Index	NAS100	DJIA30	SP500	VIX_Index	NAS100	DJIA30	SP500	VIX_Index	NAS100	DJIA30
1	-0.135	-0.393	-0.200	-0.145	-0.292	-2.415	-0.300	-0.244	-0.614	-10.743	-0.400	-0.458
2	-0.134	-0.260	-0.168	-0.185	-0.345	-0.318	-0.137	-0.382	-0.386	-7.857	-0.276	-0.387
3	-0.147	-0.161	-0.132	-0.168	-0.175	-0.395	-0.139	-0.237	-0.388	-9.238	-0.330	-0.439
4	-0.118	-0.180	-0.175	-0.108	-0.133	-0.540	-0.139	-0.190	-0.392	-10.266	-0.347	-0.477
5	-0.109	-0.271	-0.156	-0.200	-0.163	-0.475	-0.157	-0.206	-0.416	-9.995	-0.326	-0.484

Figure: Best Models by MDD

Extension to asset allocation: Dynamic Asset Allocation

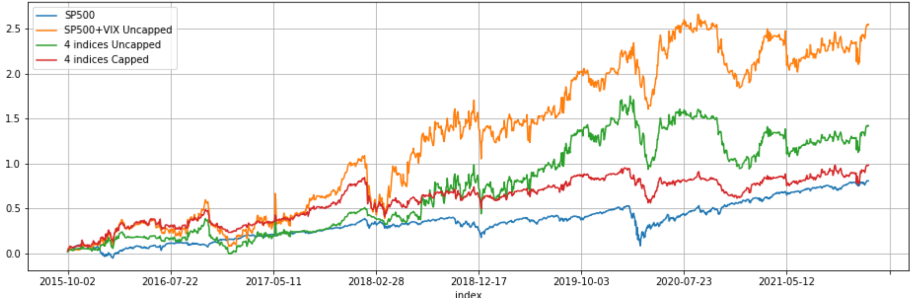
Algorithmic definition:
see whiteboard / appendix

Appendix: Algorithmic Definition of Dynamic Asset Allocation

	ANR	SR	MDD
3 indices Long-Only	0.145	**0.759**	-0.268
SP500 Long-Only	0.129	0.700	-0.293
VIX AL-DAA	**0.594**	0.611	-0.349
SP500+VIX AL-DAA	0.302	0.593	-0.292
SP500+VIX AL-DAA-Uncapped	0.412	0.578	-0.334
4 indices AL-DAA	0.158	0.551	**-0.235**
4 indices AL-DAA-Uncapped	0.229	0.444	-0.297

Table: Summary of Results

Illustrations: Cumulative Returns



Illustrations: Cumulative Returns compared to Restricted Learnings

Reminder: the unrestricted model space refers to the one that uses all classes of functions, whereas the restricted model space considers only autoregressive models.

Interpretability: Tracing back

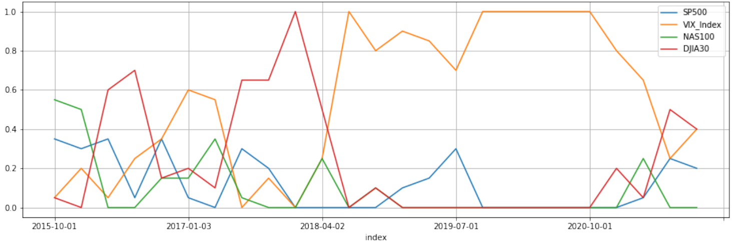


Figure: Quarterly-evaluated allocation decision on 4 indices (AL-DAA-Uncapped)

Key summaries

1. Empirical observations about financial time series:
   - Constant models can perform best by MSE, but these models perform badly in financial metrics.
   - The opposite is true of more complex specifications with meaning variables.
   - Yield Curve and VIX Curve has improved forecasting performance, especially in financial applications.
2. Reflections on time series methodologies and AL as one step to solve it:
   - AL addresses the problem of DMS and implementation of Ensemble in time series. Empirically it achieves good financial performance in a dynamic asset allocation framework.
   - Interpretable model selection and asset allocation frameworks.

Extension: Statistical and Financial

- Statistical:
  - Redesign loss function, e.g. $l = l^{PC}$.
  - Inference (e.g. distribution and testing) on selected functions and variables over time (see Yang (2021) for some possible ideas)
  - Using Neural Networks (Multi-layer Perceptron or Recurrent Layers) for non-linear learnings
- Financial:
  - Signal / cross-asset strategies
  - Interval estimation, use of smooth function instead of $sgn()$ functions for strategies

Bibliographical Remarks

References:

- 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
- 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
- Yang, Parley Ruogu and Ryan Lucas (2022). Adaptive Learning on Time Series: Method and Financial Applications https://arxiv.org/abs/2110.11156 Revision-In-Progress

Notations

- $[N] := \{1, 2, ..., N-1, N\}$
- y_{t+k} is the observed value of Y at time $t+k$
- $y_{t+k|t} = \mathbb{E}[Y_{t+k}|\Omega_t, \theta_t(h), h]$ is the forecast conditional on true parameter $\theta_t(h)$ with model h
- $\hat{y}_{t+k|t} = \mathbb{E}[Y_{t+k}|\Omega_t, \hat{\theta}_t(h, \Xi_{h,i}, \Omega_t), h]$ is the forecast made upon estimated parameter $\hat{\theta}_t(h, \Xi_{h,i}, \Omega_t)$

Term structures: remarks

- In the case of Yield Curve (YC), we have $J = 11$, whereas in the case of VIX Curve (VC), we have $J = 8$.
- The data we have are:
  - YC: one month, three months, six months, one year, two years, three years, five years, seven years, ten years, twenty years, and thirty years. These are obtained from Fed St Louis.
  - VC: Spot, one month, two months, ... , seven months. These are obtained from CBOE.

Background: Statistical Learning

- $y, x \sim P_{Y,X}$ with unknown distribution $P_{Y,X} = P_X \times P_{Y|X}$
- Learning Machine capable of implementing $\mathcal{G} \subset \{f : \mathbb{X} \to \mathbb{R}\}$
- $L(y, \hat{y})$ the loss, and risk

$$R(g) = \mathbb{E}_{X,Y}[L(Y, f(X))]$$

- Risk minimisation $g^* = \text{argmin}_{g \in \mathcal{G}} R(g)$
- Regression $g^{reg}(x) := \mathbb{E}[Y|X = x] \;\; \forall x \in \mathbb{X}$
- Well-known theorem: (e.g. Vapnik 1997)
  if $L = L^{MSE}$ and $g^{reg} \in \mathcal{G}$, then $g^{reg} = g^*$

Background: The value of forecasting

- Forecast error decomposition:

$$y_{t+k} - \hat{y}_{t+k|t} = (y_{t+k|t} - \hat{y}_{t+k|t}) + (y_{t+k} - y_{t+k|t})$$

- Example: AR(1)
  - $Y_{t+1} = c + \phi Y_t + \varepsilon_{t+1}$
  - $y_{t+1} - \hat{y}_{t+1|t} = (c - \hat{c}_t) + (\phi - \hat{\phi}_t)y_t + \varepsilon_{t+1}$

- Notice, at time t, you can draw $\hat{y}_{t|t-1}, \hat{y}_{t|t-2}, ..., \hat{y}_{t|t-k}$

Definition of AL (introductory)

- An estimation technique $\Xi_{h,i} : \Theta(h) \times \Omega_t \mapsto \hat{\theta}_t(h)$
  - E.g. Windowed OLS
- A learning function ℓ that takes past forecasts and other information and return a real number
  - E.g. ℓ^{MSE}
  - Further extensions: Financial metrics, e.g. induced Sharpe Ratio.
- A finite set of model specifications H.
- A set of AL models $H^{AL} \supseteq H$.

Definition of AL (main)

Fix t, k. An AL specification is a quadruple $(\ell, H^{AL}, H, \{\Xi_{h,i}\}_{i\in I(h), h\in H})$ such that, at time t:

1. The **set of functional specifications for AL**, denoted H^{AL}, should enable the AL forecasts to be adapted to the constituent models: $H^{AL} \supseteq H$
2. $\forall h \in H$, $I(h) \neq \emptyset$ and $\forall i \in I(h)$, $\Xi_{h,i} : \Theta(h) \times \Omega_t \mapsto \hat{\theta}_t(h)$ is well-defined.
3. $\forall h \in H$, $\forall i \in I(h)$, $\forall \tilde{k} \in [k]$, the forecast $\hat{y}_{t+\tilde{k}|t}(h, \Xi_{h,i})$ is well-defined.
4. Upon evaluating ℓ with the relevant information, $\hat{y}_{t+k|t}(h_t^{AL})$ and $h_t^{AL} \in H^{AL}$ are the outputs.

Loss functions

Consider a v-sized window up to time t, $\{y_\tau\}_{\tau=t-v+1}^{t}$, and a given pair $(h, \Xi_{h,i})$. Based on the third condition in the definition of AL, we have $\hat{y}_{\tau|\tau-\tilde{k}}$ for all $\tau \in \{t-v-k+1, ..., t-1, t\}$ and $\tilde{k} \in [k]$. Define $1_k := (1, 1, ..., 1, 1) \in \mathbb{R}^k$ and $\hat{\boldsymbol{y}}_{\tau+k|\tau} := (\hat{y}_{\tau+k|\tau+k-1}, \hat{y}_{\tau+k|\tau+k-2}, ..., \hat{y}_{\tau+k|\tau}) \in \mathbb{R}^k$. Now, define the formulations of ℓ as below:

$$\ell^{\text{Norm, }\texttt{single-valued}}(h, \Xi_{h,i}; \lambda, p) := \sum_{\tau=t-v+1}^{t} \lambda^{t-\tau} |\hat{y}_{\tau|\tau-k} - y_\tau|^p \tag{7}$$

$$\ell^{\text{Norm, }\texttt{multi-valued}}(h, \Xi_{h,i}; \lambda, p) := \sum_{\tau=t-v+1}^{t} \lambda^{t-\tau} ||\hat{\boldsymbol{y}}_{\tau|\tau-k} - y_\tau 1_k||_p^p \tag{8}$$

Implementation: Parameters

- Model space: two choices
  - H = {MC1} = $\{AR(w,p) : w \in \{22, 44, 63, 126, 252\}, p \in \{0, 1, ..., 5\}\}$
    We call this "Restricted Learning"
  - H = {MC1} ∪ {MC2} ∪ {MC3}
    We call this "Unrestricted Learning"
- Parameter space:
  - $\lambda \in \{0.8, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1\}$
  - $p \in \{1, 2\}$
  - $v = 100,\ v_0 = 50$
  - $K = 5$

Back to Results

Summary of Results: MSE

	SP500			VIX_Index			DJIA30			NAS100		
	Unrestricted Learning	Restricted Learning: AR	AR0	Unrestricted Learning	Restricted Learning: AR	AR0	Unrestricted Learning	Restricted Learning: AR	AR0	Unrestricted Learning	Restricted Learning: AR	AR0
1	2.859	2.798	2.609	20.475	19.804	19.548	2.915	2.883	2.698	3.277	3.220	3.061
2	3.562	3.481	3.321	27.598	27.361	26.730	3.735	3.634	3.458	4.167	4.048	3.929
3	4.711	4.326	4.124	33.933	32.794	31.970	4.712	4.572	4.327	5.351	4.984	4.791
4	5.467	4.997	4.795	40.355	37.471	36.629	5.821	5.325	5.044	6.147	5.700	5.524
5	6.077	5.517	5.279	45.659	41.399	40.129	6.448	5.860	5.568	6.922	6.318	6.114

Figure: Best Models by MSE, 10,000 times.

Back to Results

Algorithmic Definition of Dynamic Asset Allocation

- In plain language:
  - Period: 1 October 2015 to 31 December 2021
  - Review: we review the allocation at the start of the quarter
  - Output: An averaged holding of $N = K \times |\mathcal{A}|$ strategies.
  - Capped: each index is capped to an equal-weighted limit, e.g. the weights for any index in a 4-index basket would be 0.25.
  - Uncapped: no limit for any index in the basket.
- Algorithmically:
  - At all time t, we have $\{w_t^{k,A}(l) : k \in [K], A \in \mathcal{A}, l \in \mathcal{L}\}$
  - Consider quarter-evaluation dates $\{q_j\}_{j=1}^J$. At each $t = t_{q_j}$, we obtain the Sharpe Ratio of each strategy during the immediate one-year-behind window. Noted $\mathcal{S}_j := \{S_j^{k,A}(l) : k \in [K], A \in \mathcal{A}, l \in \mathcal{L}\}$
  - Uncapped: select the top N candidates in accordance with their $\mathcal{S}_j$.
  - Capped: for all $A \in \mathcal{A}$, consider $\mathcal{S}_j^A := \{S_j^{k,A}(l) : k \in [K], l \in \mathcal{L}\}$. Rank and select the top K candidates in accordance with their $\mathcal{S}_j^A$, repeat this for all $A \in \mathcal{A}$. Average the holdings across all A at the end.