CCIMI Initial Research Presentation

Instabilities in statistical modelling and a proposal of the adaptive learning method in time series

Presentor: Parley Ruogu Yang

Supervisor: Dr Anders C Hansen

CCIMI Presentation Date: 7 June 2021

This version: 5 June 2021

The latest version can be found on my website.

Statistical contemporaries: what is regression? (Vapnik 1999)

Let y be the true value of response to a given input x, and $f(x,\alpha)$ be the response provided by the learning machine, with $\alpha \in A$, where A denotes the set of parameters. Then, given some specification of the loss function $L(\cdot,\cdot)$, we aim to minimise the risk functional

$$R(\alpha) = \mathbb{E}_{x,y}[L(y, f(x,\alpha))]$$

by finding over the class of functions $\alpha \in A$. Write $\alpha^* := \arg\min_{\alpha \in A} R(\alpha)$

Theorem (Regression estimation (Vapnik 1999))

Let the loss to take the form $L(a,b) = (a-b)^2$. Define the regression function as

$$f(x,\alpha_0) := \mathbb{E}[y|x]$$

Then

$$\alpha^* = \arg\max_{\alpha \in A} \mathbb{E}_x[(f(x,\alpha) - f(x,\alpha_0))^2]$$

In particular, if $\alpha_0 \in A$, then $\alpha^* = \alpha_0$

Overview

- Deterministic models:
  - Stability:

$$\forall \varepsilon > 0, \forall v \in B(\mathcal{D}_x, \varepsilon), \quad |g(v; \hat{\theta}) - f(v)| \leq \texttt{UpperBound}(f, \hat{\theta}, \varepsilon)$$

  - Instability:

$$\forall \varepsilon > 0, \exists v \in B(\mathcal{D}_x, \varepsilon) \texttt{ such that } |g(v; \hat{\theta}) - f(v)| \geq \texttt{LowerBound}(f, \hat{\theta}, \varepsilon)$$

- Probabilistic models: depend on distributions and the underlying Data Generating Process (DGP).
- Here we present two of the four analyses:
  - Deterministic instability in Single Layer Neural Network (SLNN)
  - Probabilistic stability in linear model with correct specification

Deterministic instability in SLNN [Whiteboard]

$$g(\cdot;\theta) \in \mathcal{G} \qquad \Longleftrightarrow \qquad g(x;\theta) = W_2\left(\rho\left(W_1\left(x\right)\right)\right)$$

Here the parameter $\theta = (W_1, W_2)$, which is two affine transformations. Write $\rho(x) = (\tilde{\rho}(x_1), ..., \tilde{\rho}(x_{N_1}))$ and in this example, we focus on twice continuously differentiable activation function with bounded derivatives.

Corollary (Instability by perturbation)

Let $\varepsilon > 0$. For all $M > 0$, there exists $\theta \in \Theta$ such that $\exists v \in B(\mathcal{D}_x; \varepsilon)$ satisfying

$$|g(v;\theta) - f(v)| \geq M\varepsilon$$

Probabilistic stability: linear model with correct specification [Whiteboard]

- DGP $y_i | x_i \sim N(\langle x_i, \beta \rangle, \sigma^2)$ independently $\forall i \in [n]$
- Functional class

$$g(\cdot; \beta) \in \mathcal{G} \qquad \Longleftrightarrow \qquad g(x; \beta) = \langle x, \beta \rangle$$

- Estimation: $\hat{\beta}$ by MLE

Lemma (Stability of linear model under correct specification)

Pick any positive real constant c. $\forall v \in \mathbb{R}^N$,

$$\mathbb{P}\left[|g(v; \hat{\beta}) - f(v)| < \sqrt{N} c ||v|| \middle| \boldsymbol{X} \right] \geq 1 - \sqrt{\frac{2}{\pi}} c^{-1} \sum_{i \in [N]} \exp\left(-\frac{c^2}{2 s_i(\boldsymbol{X}, \sigma^2)} \right)$$

End of Part I

```
if Presentation.time.length() <=10:
  Presentation.Q_&_A.open()
else:
  Presentation.Continue()
```

Time series introduction: by plots

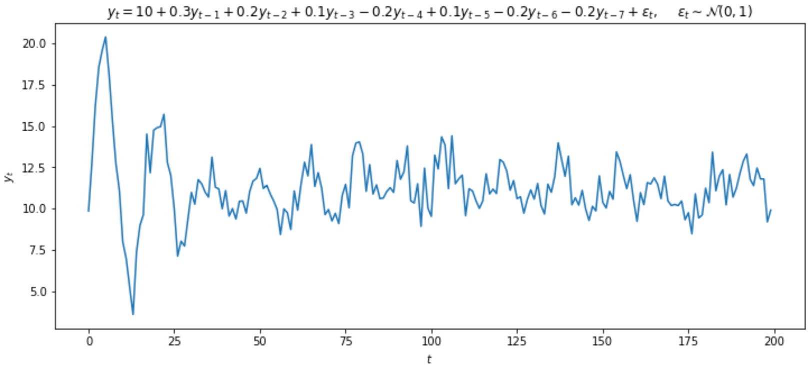

Figure: An example of simulated time series

Stationarity Check

In [10]:
```
np.random.seed(0)
random.seed(0)
all_ar_coef_7 = [0.3, 0.2, 0.1, -0.2, 0.1, -0.2, -0.2]
all_ar_coef_10 = [0.7, 0.1, -0.4, 0.3, 0.1, 0.05, -0.2, 0.1, 0.8, -0.7]

change_list_0 = [7,  0, all_ar_coef_7,  [], 150, 10, 1]
change_list_1 = [10, 0, all_ar_coef_10, [], 150, 6, 1]

plt.figure(figsize = (17,6))
plt.plot(gen_arma_change([change_list_0, change_list_1], 10))
```

Out[10]:
```
[<matplotlib.lines.Line2D at 0x23af983bf88>]
```

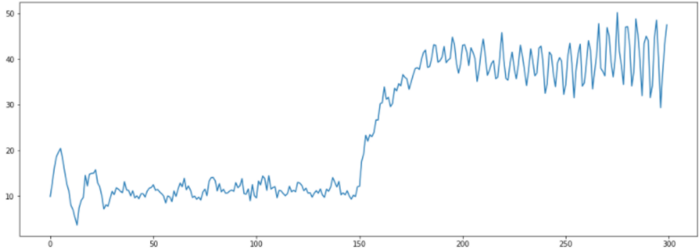

Figure: Another example of simulated time series

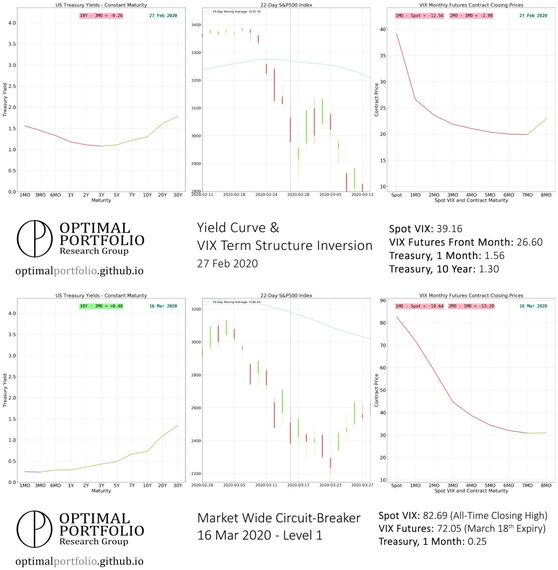

Figure: An example of empirical time series

Time series introduction: by definitions

- Hamilton (1994): Dynamic consequences of events over time.
- Prado and West (2010): A set of observations collected sequentially in time.
- Tsay (2010): Financial time series analysis is concerned with the theory and practice of asset valuation over time.
- Tsay and Chen (2018): Time series analysis is concerned with understanding the dynamic dependence of real-world phenomena.

Time series introduction: mathematical frameworks [Whiteboard]

- $y_t \in \mathbb{Y}$
  - Usually $\mathbb{Y} \subset \mathbb{R}^d$ for some $d \geq 1$
- ARMA modelling: [Assume $\mathbb{Y} = \mathbb{R}$]
  - An autoregressive moving average (ARMA) model of order p, q can be written as

$$\Phi_p(L)(y_t - \mu) = \Psi_q(L)\varepsilon_t \quad \text{where} \quad \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2) \quad \forall t$$

  - We denote an ARMA(p,0) model as AR(p), and an ARMA(0,q) model as MA(q).
- Forecasting:
  - $\Phi_t := \{y_\tau : \tau \in [t]\}$
  - let $k \geq 1$, denote the conditional forecast as

$$y_{t+k|t} = \mathbb{E}[y_{t+k}|\Phi_t]$$

- Evaluation:
  - AIC
  - Forecasting evaluation: MAE and RMSE

Time series introduction: forecasting example

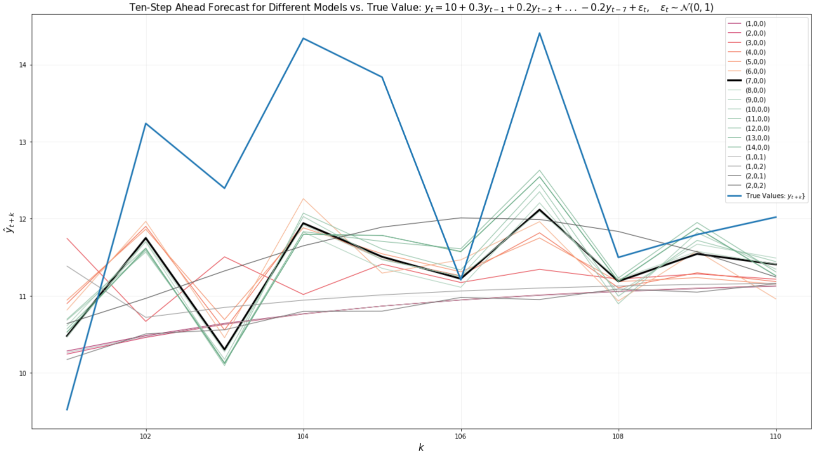

Figure: An example of forecasts by different models

Time series introduction: evaluation example

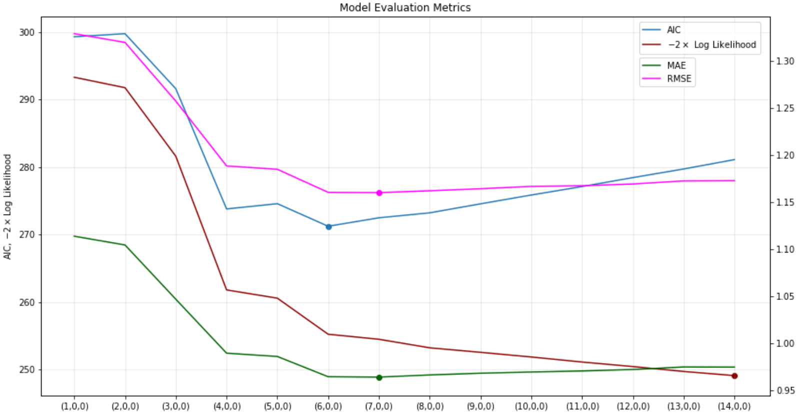

Figure: 1000 repetitions of an AR(7) model: AIC under-fits and MAE and RMSE evaluates appropriately

Time series introduction: evaluation example

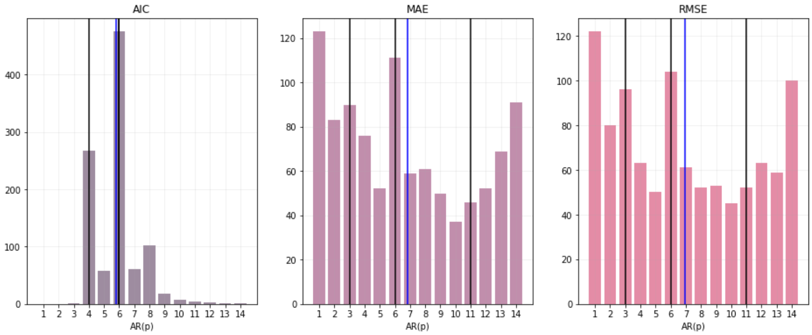

Figure: Count of the best models out of the 1000 repetitions induced by the in-sample AIC and out-of-sample MAE and RMSE. The three black lines indicates 25%, 50%, and 75% quantiles, and the blue line indicates the mean.

Time series introduction: stationarity [Whiteboard]

Definition ((Weak) stationarity)

Let $(x_t)_{t\in\mathbb{N}}$ be a real-valued time series. It is weakly stationary and denoted $x_t \sim I(0)$ if the first two moments of $(x_t)_{t\in\mathbb{N}}$ exists and time invariant. That is, for all $t, l \in \mathbb{N}$,

$$\mu := \mathbb{E}[x_t] < \infty$$
$$\gamma_l := Cov(x_t, x_{t+l}) < \infty$$

And note that both μ, γ_l are independent of t.
For a positive integer d, we denote $x_t \sim I(d)$ if $\Delta^d x_t := (1-L)^d x_t \sim I(0)$.

Example (Stationarity in ARMA(p,q))

Denote $\Phi_p^{-1}(L)\Psi_p(L) =: \Theta(L) := \sum_{i\in\mathbb{N}} \theta_i L^i$ Then:

$$\text{ARMA(p,q) stationary} \quad \Leftrightarrow \quad (\theta_i)_{i\in\mathbb{N}} \in \ell^2(\mathbb{N}) \tag{1}$$

Where stationarity tests fail

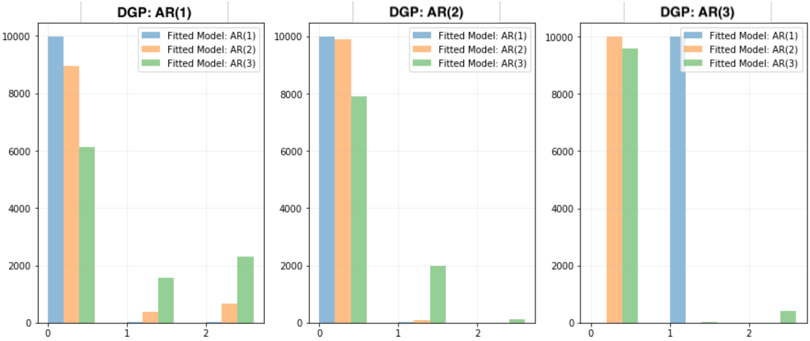

Figure: Model misspecification leads to false recommendation for non-stationarity

Inference on AR models: stability [Whiteboard]

Lemma (The consequence of correct specification)

Suppose y_t to have a true stationary AR(1) process

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \tag{2}$$

And that we use an AR(1) model to fit the data $\{y_0, ..., y_T\}$, noted

$$y_t = \rho y_{t-1} + e_t \quad , \quad e_t \sim N(0, \sigma^2) \quad \forall t \in \{0, 1, ..., T\} \tag{3}$$

Then, for the OLS estimate $\hat{\rho}$,

$$\hat{\rho} \xrightarrow[T\to\infty]{P} \phi_1 \tag{4}$$

$$T^{\frac{1}{2}}(\hat{\rho} - \phi_1) \xrightarrow[T\to\infty]{D} N(0, 1 - \phi_1^2) \tag{5}$$

Hence

$$\hat{y}_{T+1|T} - y_{T+1|T} \xrightarrow[T\to\infty]{P} 0$$

Inference on AR models: instability [Whiteboard]

Lemma (The consequence of misfit)

Suppose y_t to have a true stationary AR(2) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{6}$$

Suppose further that $v_1 := Var(y_t y_{t-1})$ is finite and time-independent. Denote $v_2 := Var(y_t) = (1-\phi_2)\sigma^2((1+\phi_2)((1-\phi_2)^2 - \phi_1^2))^{-1}$. Suppose that we use an AR(1) model to fit the data $\{y_0, ..., y_T\}$ as noted in Equation 3, then

$$\hat{\rho} - \phi_1 \xrightarrow[T\to\infty]{P} \frac{\phi_1\phi_2}{1-\phi_2} \tag{7}$$

$$T^{\frac{1}{2}}(\hat{\rho} - \phi_1 - \frac{\phi_1\phi_2}{1-\phi_2}) \xrightarrow[T\to\infty]{D} N(0, \frac{v_1\phi_2^2 + v_2\sigma^2}{v_2^2}) \tag{8}$$

Hence

$$\hat{y}_{T+1|T} - y_{T+1|T} - \left(\phi_2(\phi_1 y_T(1-\phi_2)^{-1} - y_{T-1})\right) \xrightarrow[T\to\infty]{P} 0$$

Adaptive learning: introduction

- Exploit the ordered nature of the time series, that because it is ordered, its pasts can verify forecasts that were produced based on previous observations, hence dynamically offer information which can be helpful for model determination.
- This can be similar to some basic statistical learning concepts including cross validation and can overlap with methods including model ensemble.
- First proposed in Yang (2019).

Adaptive learning: a sample algorithm

B.2 A sample algorithm for implementing adaptive learning

Input:

- Data: $\{y_t : t \in T\}$.
- Model specifications: H.
- Forecasting horizon k.
- Window size for the model estimations w.
- Specification of loss function: $l(\cdot)$. Default choice can be ARMSE($\tilde{T}$).

$$l(\{e_\tau\}_{\tau\in\tilde{T}}) = \sum_{\tau\in\tilde{T}} ||e_\tau||_2 \tag{21}$$

At time t, $\tilde{T} = \{t - v - k + 1, ..., t - k - 1, t - k\}$

Output:

- T^{test} the index of testing forecasts.
- Induced-optimal functional forms $(h_t^*)_{t\in T^{test}}$
- Induced-optimal forecasts for testing purposes: $\{\hat{y}_{t+k|t}(h_t^*)\}_{t\in T^{test}}$
- General evaluation of MAE and MSE at time $t \in T^{test}$ and the overall evaluations.

Figure: Input and output

Computing steps:

1. Call $T^{max} := \max(T)$. Define $T^{test} = \{w + v + k, ..., T^{max} - k - 1, T^{max} - k\}$
2. For all $t \in \{w + 1, ..., T^{max} - k\}$:
   (a) For all $h \in H$:
      i. Train h over the data $\{y_\tau : \tau \in [t] \setminus [t - w - 1]\}$
      ii. Obtain vector $\hat{y}_{t+k|t}(h)$
      iii. Obtain vector $e_t(h)$
3. For all $t \in T^{test}$:
   (a) Declare $\tilde{T}$.
   (b) For all $h \in H$:
      i. Collect the set$\{e_\tau(h)\}_{\tau \in \tilde{T}}$
      ii. Evaluate $\ell(h) = l(\{e_\tau(h)\}_{\tau \in \tilde{T}})$
   (c) Find $h^* := \arg\min\{\ell(h) : h \in H\}$.
   (d) Hence save this h_t^* and make and save the associated forecast $\hat{y}_{t+k|t}(h_t^*)$
4. Produce general evaluation for $\{\hat{y}_{t+k|t}(h_t^*)\}_{t \in T^{test}}$. Return the output.

Figure: Computational steps

Adaptive learning: formal definition

Definition (Adaptive learning)

An adaptive learning on a time series dataset $\mathcal{D}$ and forecasting horizon k, is a learning method in which there is a learning function l, a set of models H, and estimation techniques

$$\Lambda_h : \Theta(h) \mapsto \theta(h) \quad \forall h \in H$$

where $\Theta(h)$ is the domain of the parameter and $\theta(h)$ is the estimated parameter, such that for all time $t \in T^{\texttt{validation}} \cup T^{\texttt{test}}$, we are able to obtain the followings:

1. $\forall h \in H$, an estimated statistic $\theta_t(h)$
2. $\forall h \in H$, an estimated forecast $y_{t+k|t}(h)$
3. an optimal model $h_t^* \in H$ selected by l, with an induced optimal forecast $y_{t+k|t}(h_t^*)$

The design of $(l, H, \{\Lambda_h\}_{h \in H})$ should be made on the basis of theoretical validity and tuned in the validation data. Their ability to improve forecasts can then be demonstrated in the testing data.

Dynamic model selection (Yang 2020; Yang 2021)

Let H be finite. Each $h \in H$ gives $f_k(\cdot;\cdot;h)$ a functional specification. For instance, it could be $f_k(\Phi_t;\theta_t;h) = \theta_{t,1} + \theta_{t,2}x_t$. $\Theta(h) = \mathbb{R}^2$ in this case, and the training of $\theta(h)$ can be done by MLE over a w windowed training set at time t.

The above specifies H and $\{\Lambda_h\}_{h\in H}$. Now, we introduce an MSE-styled loss function:

$$l_t(h) = \sum_{\tau\in\tilde{T}} RMSE_\tau(h)$$

At time t, $\tilde{T} = \{t-v-k+1, ..., t-k-1, t-k\}$ indicating that we equally evaluate the loss over the v windowed history up to time $t-k$. Then, we can obtain an optimal model $h_t^* \in H$ by executing a finite minimisation operation:

$$h_t^* = \arg\min_{h\in H} l_t(h)$$

Contribution

- Section 1:
    - Provided a framework for analysing stability and instability of statistical modelling.
    - Engaged with the analysis of NN and adversarial attacks from a different dimension.
- Section 2:
    - Extended the framework to time series analysis.
    - Demonstrated the distributional results of autoregressive misfit, and proposed a new learning regime with potential extensions.

Further work

- Section 1:
  - Further diggings in to more distributions, analysis of NN, and adversarial attacks.
  - Relevant work-in-progress in AFHA: Bastounis, Hansen, and Vlacic (2021).
- Section 2:
  - More inference and computational experiments. Then upload to Arxiv.
  - Combine empirical results (collaboratively with OPRG) to submit to the BoE annual conference and RSSC.

Nobel Laureates

Robert F. Engle III
Facts

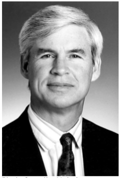

Photo from the Nobel Foundation archive.

Robert F. Engle III
The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003

Born: 10 November 1942, Syracuse, NY, USA

Affiliation at the time of the award: New York University, New York, NY, USA

Prize motivation: "for methods of analyzing economic time series with time-varying volatility (ARCH)."

Contribution: Developed methods to study the volatility properties of time series in economics, particular in financial markets. His method (ARCH) could, in particular, clarify market developments where turbulent periods, with large fluctuations, are followed by calmer periods, with modest fluctuations.

Prize share: 1/2

Clive W.J. Granger
Facts

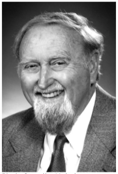

Photo from the Nobel Foundation archive.

Clive W.J. Granger
The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003

Born: 4 September 1934, Swansea, United Kingdom

Died: 27 May 2009, San Diego, CA, USA

Affiliation at the time of the award: University of California, San Diego, CA, USA

Prize motivation: "for methods of analyzing economic time series with common trends (cointegration)."

Contribution: Developed and applied new statistical methods, based on so-called "cointegration", to differentiate between, and combine the analysis of, short-term fluctuations and long-term trends.

Prize share: 1/2

Figure: The last time the Noble Prize in Economic Sciences was awarded to people who made landmark contribution to Time Series

References

Bastounis, A., A. C. Hansen, and V. Vlacic (2021). "The mathematics of adversarial attacks in AI." In: Work in progress.

Hamilton, James D (1994). *Time Series Analysis*. Princeton University Press.

Prado, Raquel and Mike West (2010). *Time Series: Modeling, Computation, and Inference*. 1st ed. Taylor and Francis Group.

Tsay, R.S. and R. Chen (2018). *Nonlinear Time Series Analysis*. Wiley.

Tsay, Ruey S. (2010). *Analysis of Financial Time Series*. Wiley.

Vapnik, Vladimir (1999). *The Nature of Statistical Learning Theory*. Springer.

Yang, Parley Ruogu (2019). *Bank of England: Modelling with Big Data and Machine Learning Conference*. URL: https://www.bankofengland.co.uk/events/2019/november/modelling-with-big-data-and-machine-learning.

– (2020). "Using the yield curve to forecast economic growth." In: *Journal of Forecasting* 39.7, pp. 1057–1080. DOI: 10.1002/for.2676.

– (2021). "Forecasting high-frequency financial time series: an adaptive learning approach with the order book data." In: eprint: arXiv:2103.00264.