Forecasting Yield Curves with Neural Nets - Improvement on DNS

Dr. Matthew Lightwood & Anna Knezevic

Conning GmbH *M&A Solutions LTD*

About the speaker

- **Matthew Lightwood**
- Director Quantitative Finance
- Expert in quantitative and economic modelling
- GEMS™ Economic Scenario Generators

CONNING®

- **Conning**
- GEMS™ Economic Scenario Generator (ESG) provider
- Asset allocation and fund management specialist
- Risk manager

About the speaker

- **Anna Knezevic**
- Managing Director
- Works in quantitative aspects of profit/capital optimization based on incomplete data; utilizing data analytics to drive business insights.

- **M&A Solutions Ltd**
- Provider of customized quantitative solutions for:
  - Capital optimization,
  - Risk management, and
  - Strategy

Forecasting Yield Curves Motivation

- Many places where a central assumption of the future value of interest rates is required
  - Investment decision making
  - Regulatory internal models (e.g. Solvency II)
  - Policy decision making (e.g. Central Banks)
  - Assesment of risk in pensions funds (e.g. With profits funds)
- A robust, automatable, repeatable, explainable, justifiable approach is required

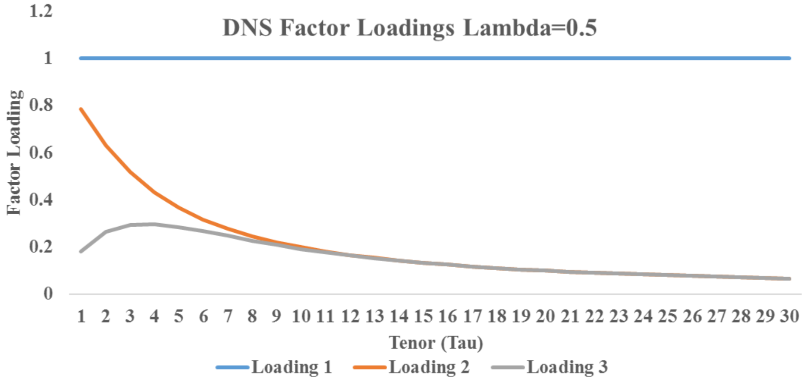
Dynamic Nelson Siegel Model

- Dynamic Nelson Siegel Model (DNS) is a popular framework for analysing and forecasting interest rates
  - Backed by a large body of research (e.g. Diebold and Li 2005/2006)
  - Outperforms other methods on data from multiple economies
  - Parsimonious, intuitive, relatively simple to estimate

Dynamic Nelson Siegel (Basic Idea)

- Three factor model
- Fix λ and fit β's to historical yield curves (OLS)
- For example with German yields.....

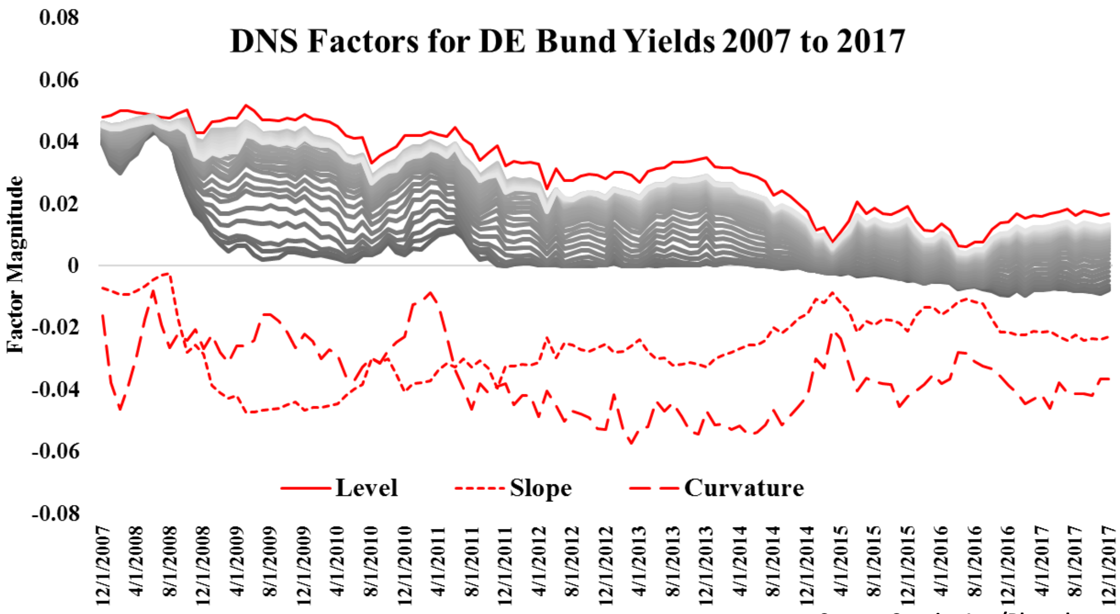
$$y_t(\tau) = \underbrace{\beta_{1,t}}_{\text{Level}} + \beta_{2,t}\left[\underbrace{\frac{1-\exp(-\lambda_t \tau)}{\lambda_t \tau}}_{\text{Slope}}\right] + \beta_{3,t}\left[\underbrace{\frac{1-\exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau)}_{\text{Curvature}}\right] + \varepsilon_t.$$



Source: Conning Inc.

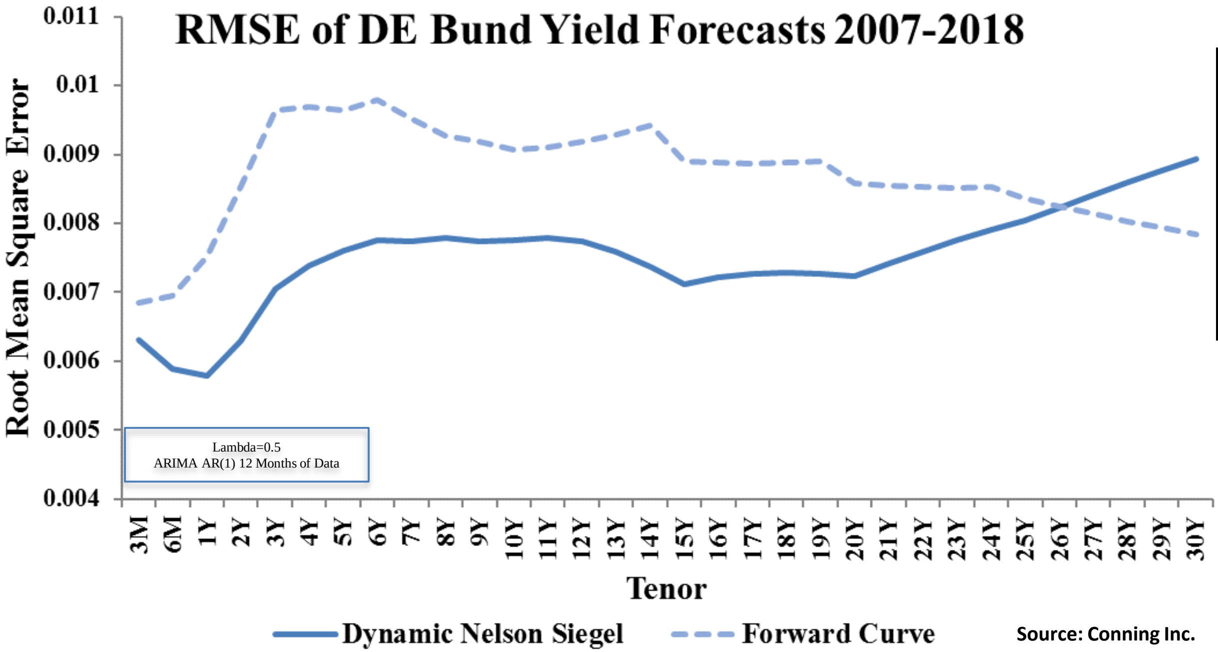
Dynamic Nelson Siegel (Basic Idea)

- Factors β are dynamic
- B1,t closely follows the yield levels as expected
- "Shape" factor movements track term structure movements
- Build ARIMA model to forecast future yields curves



Source: Conning Inc./Bloomberg

Dynamic Nelson Siegel - Performance



Source: Conning Inc.

	DNS		*Forward Curve*	
	RMSE	STDER	RMSE	STDER
Tenor				
3 Month	*63.03*	*3.77*	*68.53*	*4.36*
1 Year	*57.91*	*3.58*	*75.24*	*4.91*
5 Year	*75.98*	*4.18*	*96.48*	*5.54*
10 Year	*77.52*	*4.05*	*90.65*	*5.14*
30 Year	*89.31*	*4.53*	*78.46*	*4.36*

Thought process for Neural Net

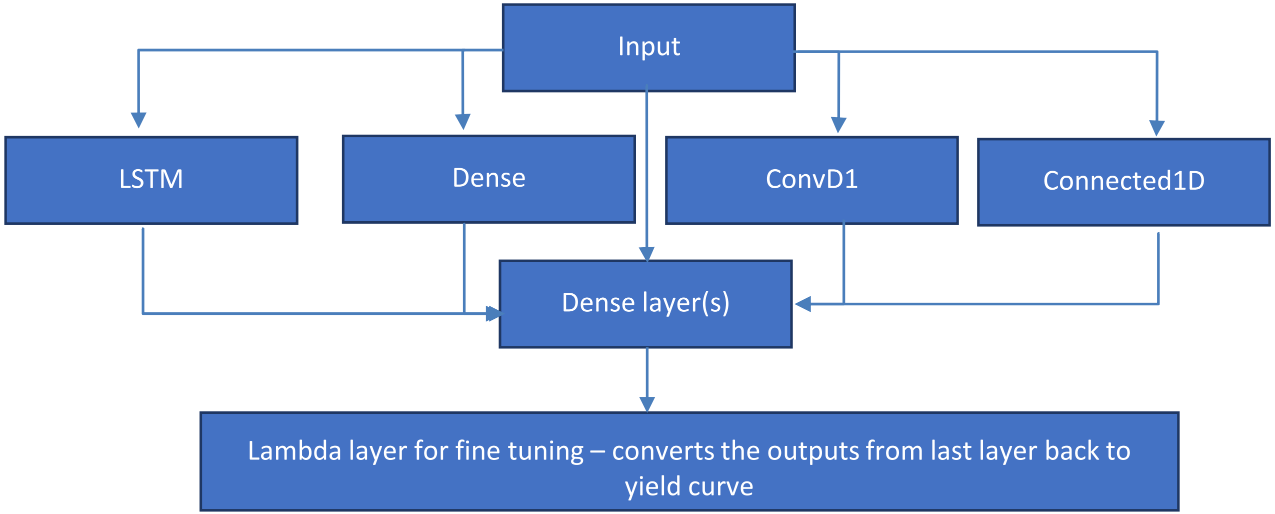
- We considered that DNS doesn‘t take into consideration
  - long term shifts in the regimes;
  - Interaction between three components;
  - Actual underlying yield curves.

- Stipulation: Neural Net that DOES do all of the above should improve on DNS performance, ceterus paribas.

How and why does the wide model work?

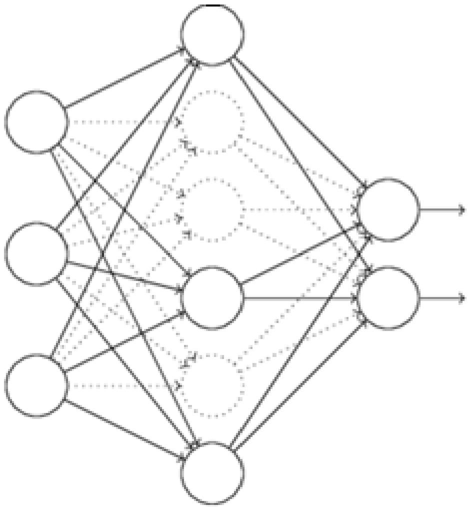
- Previous attempts to predict yield curves have noted that deep neural nets struggle to forecast data
- Data contains multiple features that the wide model is able to detect because of multiplicity of different „approaches“
- Gating unit (i.e. Initial values) has been added to enable network to identify under which scenarios some experts may perform better

Implementation



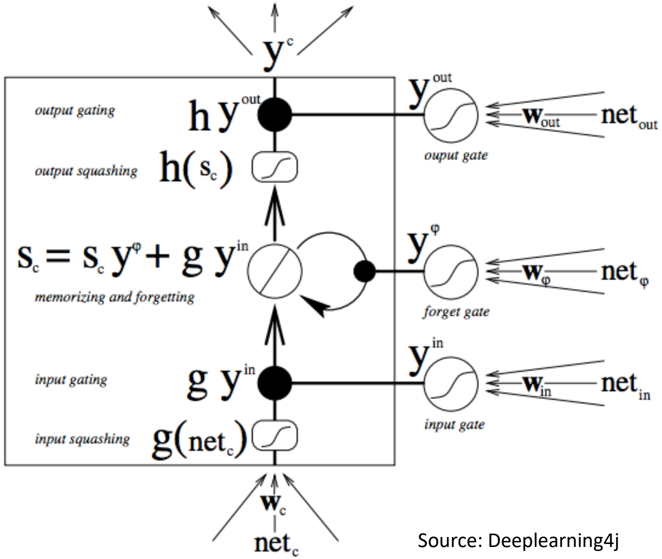
What do the layers look like?

- Dense Layer



Source: Quora- Mike West

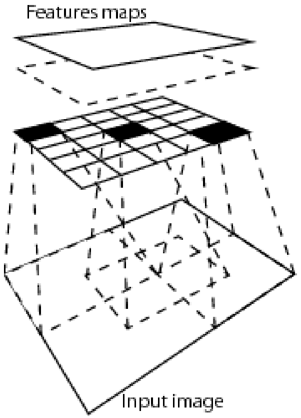
- LSTM Layer



Source: Deeplearning4j

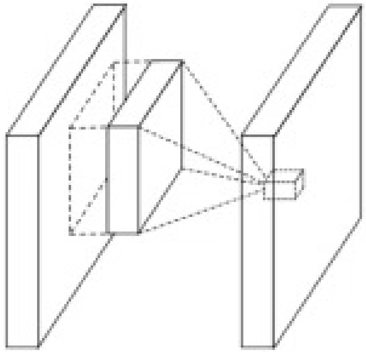
What do the layers look like?

- Conv1D Layer



Source: Zhu, W.W. et al. 2014 HEP

- LocallyConnected1D Layer



Source: Blog by Adrian Colyer, 2017

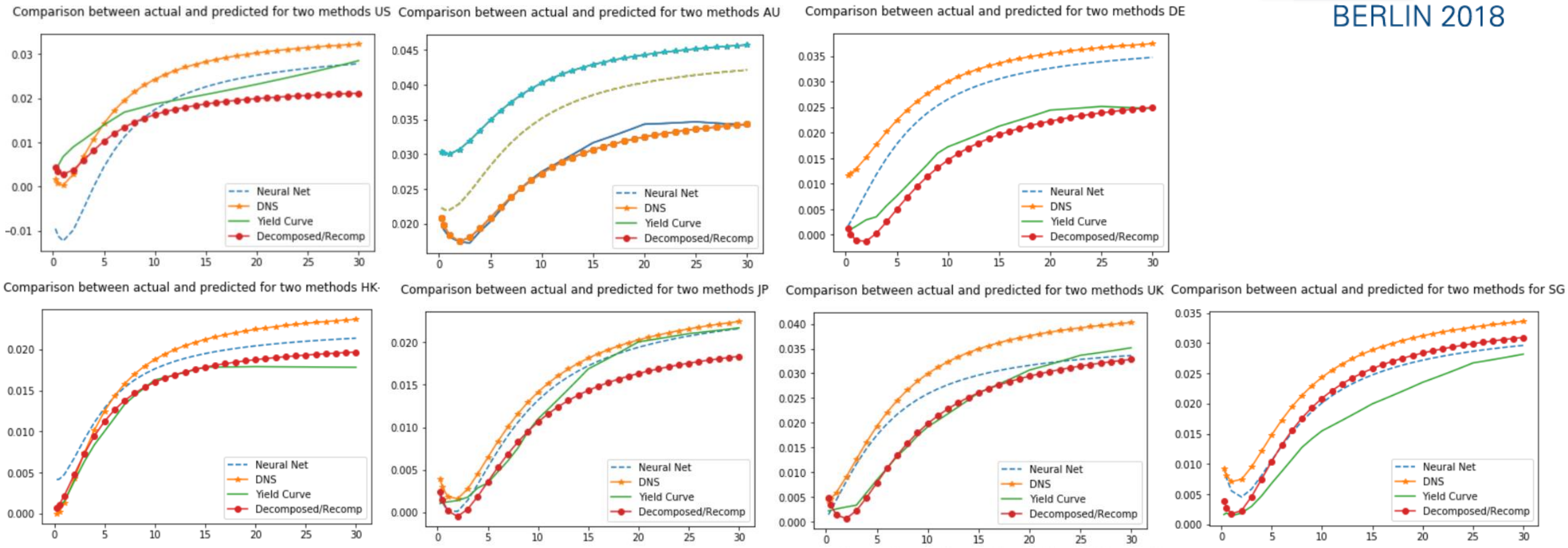
What does the math look like?

- For $\beta_{1,2,3} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ $where\ z = \{z_1,\ z_2,\ z_3\}$
- For $z = w_1 \sum_{60}^{0} gating\ +\ w_2 \sum_{120}^{0} LSTM\ +\ w_3 \sum_{9}^{1} Conv1D\ + w_4 \sum_{9}^{1} LocallyConnected$ $+\ w_5 \sum_{6}^{0} Dense$ *« (β as above)*;
- Where $w_1 \ldots\ w_5$ are matricies of weights,
  - gating is the original input;
- $LSTM = h(s_c\, y^{\varphi} + g\, y^{in})$
- $Conv1D = \ f(b_i^{(\ell)}\ + \sum_{t'=1}^{d} \left\langle W_{i,t',\cdot}^{(\ell)}, E_{\cdot,t+d+t'}^{(\ell-1)} \right\rangle)$
- $LocallyConnected = \ f(b_i^{(\ell,i)}\ + \sum_{t'=1}^{d} \left\langle W_{i,t',\cdot}^{(\ell,i)}, E_{\cdot,t+d+t'}^{(\ell-1)} \right\rangle)$

Results

Country	% improvement	Absolute error DNS	Absolute Error Neural Net
United States	**-4.2%**	4.5834	4.3923
United Kingdom	**-9.2%**	4.4786	4.0666
Germany	**-21.4%**	7.1098	5.5848
Australia	**-27.0%**	6.2658	4.5742
Japan	**-34.8%**	1.4191	0.9247
Singapore	**-20.1%**	3.4951	2.7930
Hong Kong	**-2.7%**	2.0795	2.0242

Sample yield curves by country



Thank you very much for your attention!

Contact details:

Dr. Matthew Lightwood

address: Augustinerstr. 10,
50667 Cologne,
Germany

phone: +49 (0) 22178800405

mail: matthew.lightwood@conning.com

web: www.conning.com

Contact details:

Anna Knezevic

address: 31 Robinson Rd
Hong Kong, Hong Kong
China

phone: +(852) 6461 8307

mail: anna.k@m8asolutions.com

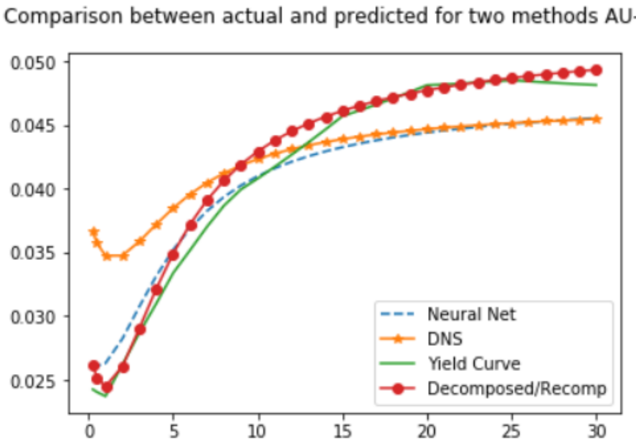
web: www.m8asolutions.com

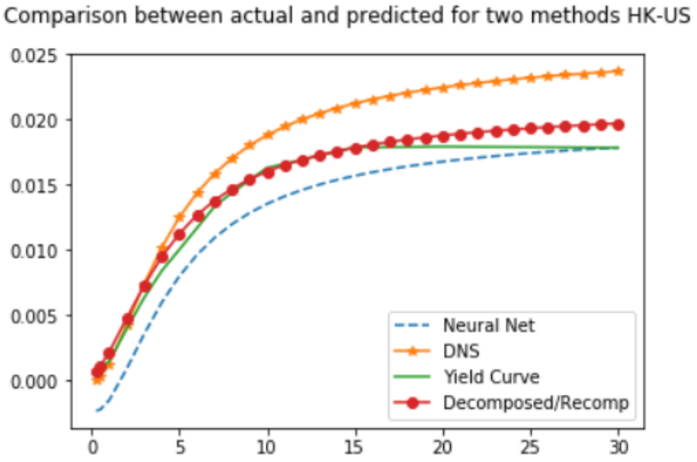
Q&A: Overfitting, scarce data and others

- Forcing model to generalise
- Early stopping times
- Rolling rather than non-overlapping
- Dropout
- Normalisation
- Data compression

Q&A: Cross economy prediction

Country	DNS	Usual NN	US as input to NN
Australia	6.2658	4.5742	4.6814
Hong Kong	2.0795	2.0242	2.0242





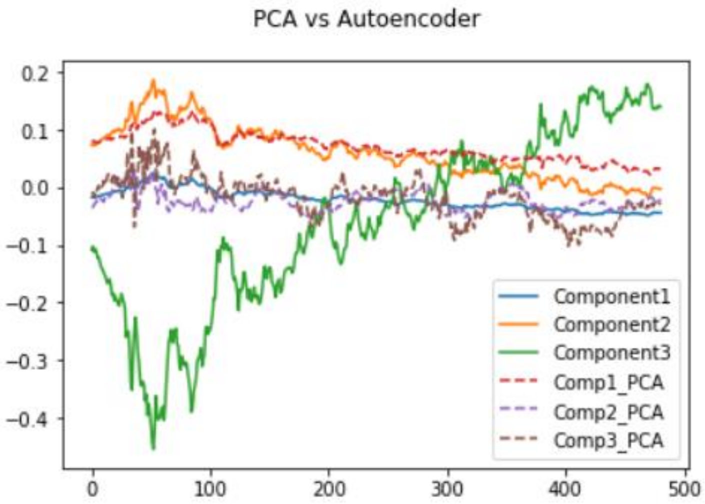
Q&A: Technology

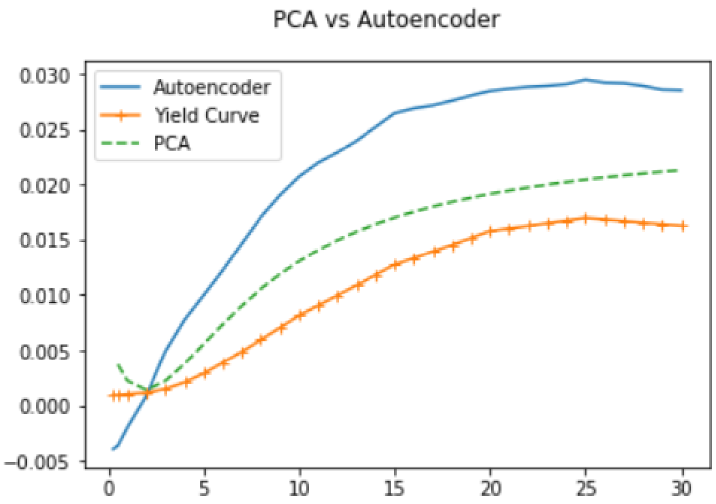
- Matlab (Octave)
- R-Studio
- Python (Jupyter)

- Cloud (GCP)

Q&A: Autoencoders and PCA

- PCA works better! (Although we haven’t tried causal encoders)





Q&A: Further improveNets

- Additional data sources
  - Fed statements, markets, GDP…
- Residual nets (some of the errors will have correlation with time)