Forecasting the Leading Indicator of a Recession: The 10-Year minus 3-Month Treasury Yield Spread by Sudiksha Joshi

Anirudha Akela

Indian Institute of Technology, Delhi

akelaanirudha@gmail.com

December 30, 2020

Overview

Abstract

In this paper, the author has applied various econometric time series and machine learning models to forecast the daily data on the yield spread.

- We do principal component decomposition and simulate the yield spread using the Vasicek model.
- We build and compare the following forecasting models: univariate ARIMA, multivariate ARIMAX, VAR, MLPs and LSTMs.
- We also study the impact of shocks through IRF and FEVD.

The results indicate that the parsimonious univariate ARIMA model outperforms the richly parameterized VAR method, and the complex LSTM with multivariate data performs equally well as the simple ARIMA model.

What is Yield Spread and why does it matter?

- The yield curve depicts the interest rates of treasury securities of various maturities that have equal credit quality and same risk characteristics.
- As a leading indicator of predicting a recession, economists typically incorporate the yield spread in probit models to forecast the probability of a recession one year from now.
- While the literature on yield curve modeling is prolific, not many economists have forecasted yield spread alone.

Normal and Inverted Yield Curves

- A normal yield curve slopes upward, reflecting the fact that short-term interest rates are usually lower than long-term rates. That is a result of increased risk and liquidity premiums for long-term investments.
- An inverted yield curve slopes downwards and represents a situation in which long-term debt instruments have lower yields than short-term debt instruments of the same credit quality.

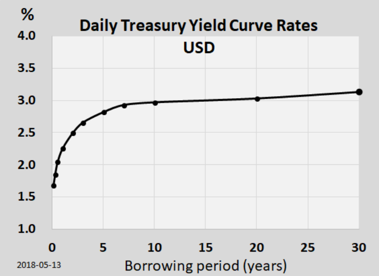


Figure: Normal Yield Curve

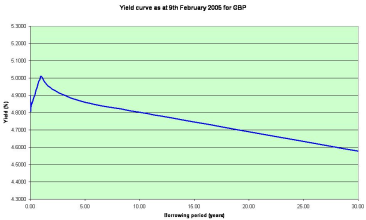


Figure: Inverted Yield Curve

Inverted Yield Curves precede Recessions

Historically, inverted yield curves have preceded every US recession.

An inverted yield curve means that most investors believe that short-term interest rates are going to fall sharply at some point in the future. As a practical matter, recessions usually cause interest rates to fall. Inverted yield curves are almost always followed by recessions.

The 3D yield curve

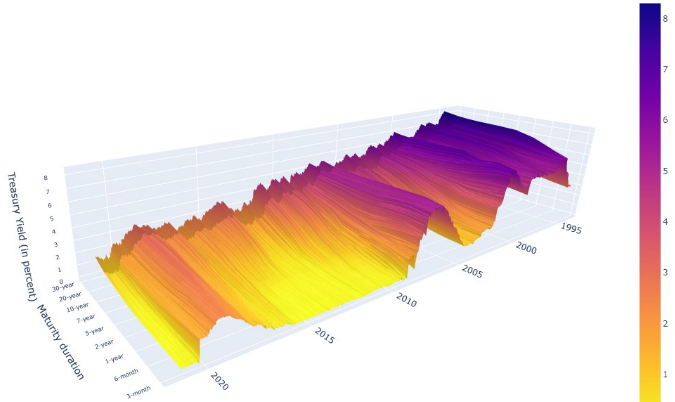


Figure: The 3D yield curve

Trubin and Estrella (2006)

- They emphasized that the yield curve is a more forward-looking leading indicator as the recession signals that it produces are more advanced than those produced by other variables.
- They claim that using treasury yields whose maturities are far apart generate accurate results in forecasting the real activity.
- The evidence connotes that more pronounced inversions are correlated with deeper downturns.

Time series models in literature

- Merterns and Bauer (2018) constructed probit models to quantify the probability of recession at t + 12 months based on the term at time t, contingent upon the term spread being above or below a particular threshold.
- Das and Sambasivan (2017) found that multivariate time series method performed well in forecasting the yields at the short-term region of the curve, whereas Gaussian Processes produced more accurate results in the long and medium term regions of the yield curve.

ML models in literature

- Bianchi et. al. (2018) examined the out-of-sample performance after modeling a diverse array of machine learning methods to forecast Treasury bonds across different maturities.
- Overall, their results point the neural network's success to its property of capturing convoluted non-linearities embedded in the data and non-linearities within categories such as labor market, output, etc.
- Lu et. al (2019) developed a Long Short Term Memory (LSTM) model with an out-of-sample accuracy of 88.9 percent in forecasting the next day change in the credit spread.
- The advantage of LSTMs are that they are able to gradually evolve and forget older data, while simultaneously learning patterns from recent history to extrapolate regime changes.

Principal Component Analysis (PCA)

- PCA extracts the signal and diminishes the dataset's dimensionality. This is because it finds the fewest number of variables that explain the largest proportion of the data.
- The author has decomposed the data into 5 principal components and has then calculated the proportion of total variance explained by each of the components.

Plotting the principal components

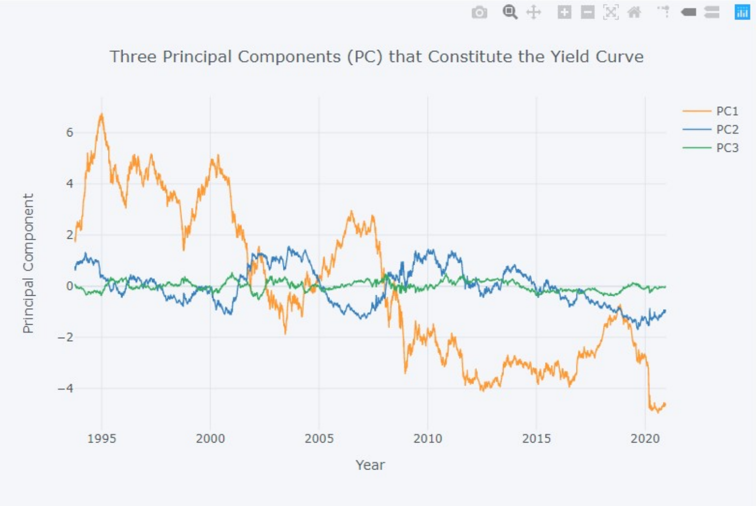


Figure: The 3 significant principal components

Scree Plot

Variance explained by principal component

We calculate the proportion of the variance explained by each PC_i using:

$$V(PC_i) = \frac{\lambda_i}{\lambda_1 + \lambda_2 + ... + \lambda_5}$$

Figure: Plotting the proportion of variance explained by each component

Vasicek Interest Rate Model

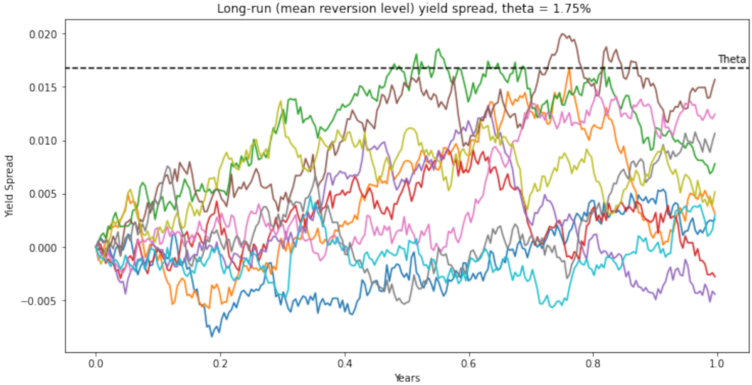
- A stochastic technique of modeling the instantaneous movements in the term structure of forward interest rates is the Vasicek interest rate model.
- The factors that describe the trajectory of interest rates are time, market risk, and equilibrium value, wherein the model assumes that the rates revert to the mean of those factors as time passes.

Closed form solution of Vasicek model

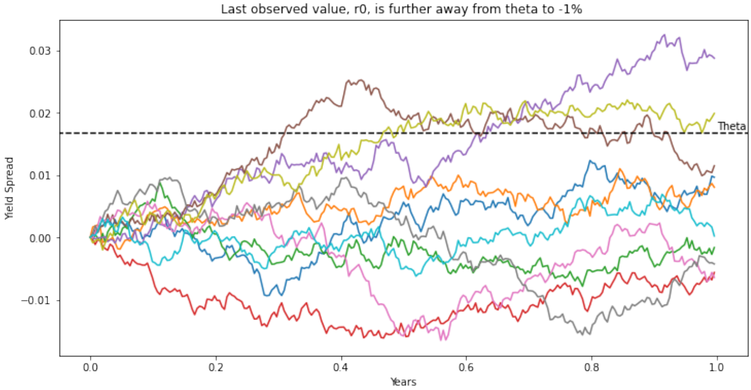
$$r_{t_i} = r_{t_{i-1}} e^{-k(t_i - t_{i-1})} + \theta(1 - e^{-k(t_i - t_{i-1})}) + Z\sqrt{\frac{\sigma^2(1 - e^{-2k(t_i - t_{i-1})})}{2k}}$$

Vasicek Interest Rate Model

Simulating from $r_0 =$ last observed value and assuming $\theta = 0.0175$, we have generated a sequence of 10 ex-ante trajectories of the yield spread.

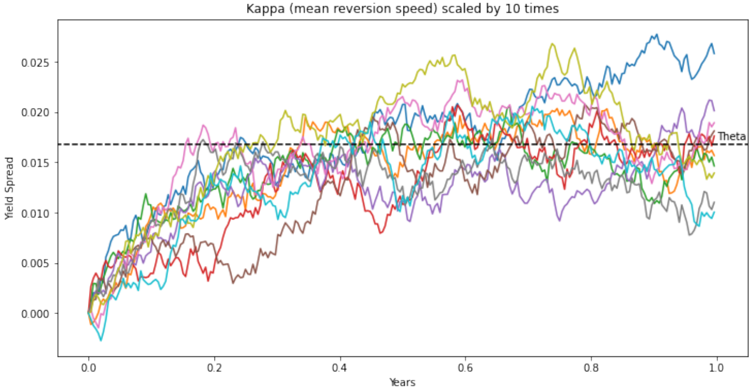


Effect of r_0



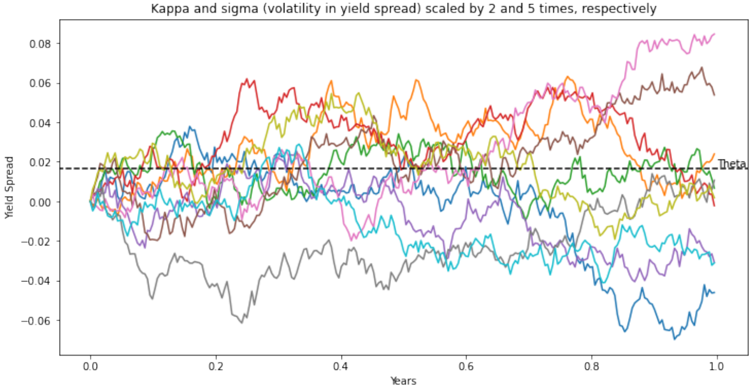
We observe the model's mean reverting nature by specifying r_0 further away from θ. Over time this pulls the yield spread towards .

Effect of κ



The magnitude of κ controls the speed of the reversion. As κ grows, mean reversion quickens.

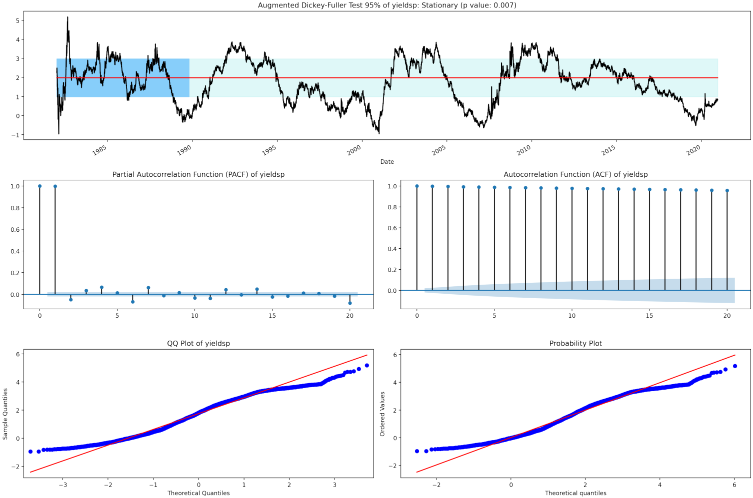
Effect of σ



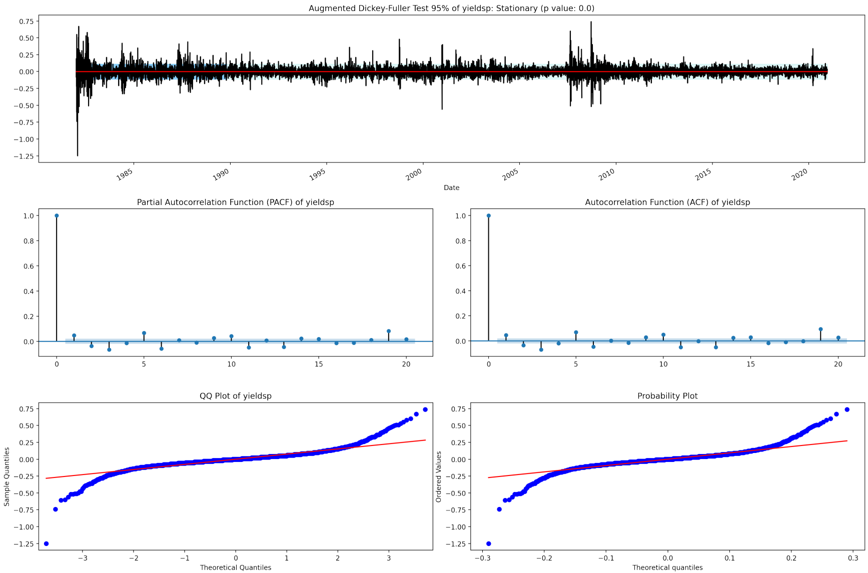
Likewise, larger σ widens the volatility and the potential distribution rate.

Stationarity of the Yield Spread

We first check for stationarity of the yield spread using the Augmented Dickey Fuller Test.



Stationarity of the differenced Yield Spread



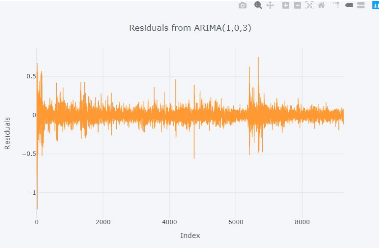
ARIMA(1,0,3) on yield spread

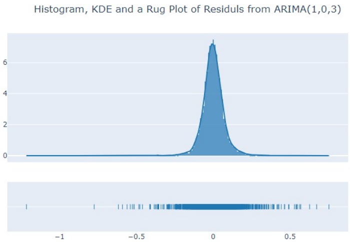
```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 9253
Model:                     ARMA(1, 3)   Log Likelihood               10592.283
Method:                       css-mle   S.D. of innovations              0.077
Date:                Tue, 29 Dec 2020   AIC                         -21172.567
Time:                        14:05:58   BIC                         -21129.771
Sample:                             0   HQIC                        -21158.025

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.7712      0.328      5.401      0.000       1.128       2.414
ar.L1.y        0.9978      0.001   1454.523      0.000       0.996       0.999
ma.L1.y        0.0438      0.010      4.231      0.000       0.024       0.064
ma.L2.y       -0.0211      0.011     -1.956      0.051      -0.042    4.58e-05
ma.L3.y       -0.0696      0.011     -6.319      0.000      -0.091      -0.048
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0022           +0.0000j            1.0022            0.0000
MA.1            2.4175           -0.0000j            2.4175           -0.0000
MA.2           -1.3604           -2.0240j            2.4386           -0.3442
MA.3           -1.3604           +2.0240j            2.4386            0.3442
-----------------------------------------------------------------------------
```

Fitting the ARIMA(1, 0, 3) model yields the following equation with an AIC of 20, 952.034 :

$$yieldsp_t = 1.7712 + 0.9978 yieldsp_{t-1} + 0.0438\epsilon_{t-1} - 0.0211\epsilon_{t-2} - 0.0696\epsilon_{t-3}$$

Residuals from ARMA(1,0,3)





The plots suggest that the errors are Gaussian with a mean centered at approximately zero (-0.000066), indicating no bias in the predictions. The line plot of residuals suggest that the model captures of the trend information.

Evaluation using walk forward validation

- We have fit ARIMA(1, 0, 3) model on the train set and predicted for each observation in the test set using a walk-forward validation scheme.
- The ARIMA(1,0,3) model achieves a RMSE of 0.05110.

ARIMA(3,1,3) on the differenced yield spread

ARIMA Model Results

Dep. Variable:	D.y	No. Observations:	9252
Model:	ARIMA(3, 1, 3)	Log Likelihood	10578.756
Method:	css-mle	S.D. of innovations	0.077
Date:	Tue, 29 Dec 2020	AIC	-21141.512
Time:	14:27:50	BIC	-21084.451
Sample:	1	HQIC	-21122.122

	coef	std err	z	P>|z|	[0.025	0.975]
const	-7.392e-08	2.8e-07	-0.264	0.791	-6.22e-07	4.74e-07
ar.L1.D.y	0.1647	0.009	18.811	0.000	0.148	0.182
ar.L2.D.y	0.1576	0.039	4.045	0.000	0.081	0.234
ar.L3.D.y	-0.0594	0.010	-5.657	0.000	-0.080	-0.039
ma.L1.D.y	-1.1186	nan	nan	nan	nan	nan
ma.L2.D.y	-0.0759	0.038	-2.014	0.044	-0.150	-0.002
ma.L3.D.y	0.1945	0.038	5.103	0.000	0.120	0.269

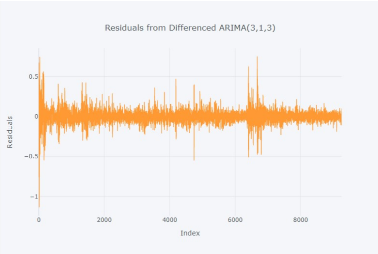
Roots

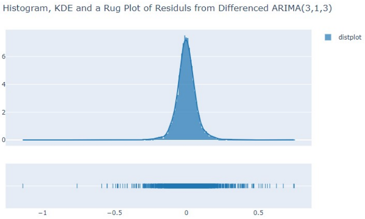
	Real	Imaginary	Modulus	Frequency
AR.1	-2.1763	-0.0000j	2.1763	-0.5000
AR.2	2.4153	-1.3803j	2.7819	-0.0826
AR.3	2.4153	+1.3803j	2.7819	0.0826
MA.1	1.0000	+0.0000j	1.0000	0.0000
MA.2	-2.5927	+0.0000j	2.5927	0.5000
MA.3	1.9827	+0.0000j	1.9827	0.0000

Fitting the ARIMA(3,1, 3) model on the differenced yield spread yields the following equation with an AIC of 21141.512 :

$$\Delta yieldsp_t = -7.392 * 10^{-8} + 0.1647\Delta yieldsp_{t-1} + 0.1576\Delta yieldsp_{t-2} - 0.0594\Delta yieldsp_{t-3} - 1.1186\epsilon_{t-1} - 0.0759\epsilon_{t-2} + 0.1945\epsilon_{t-3}$$

Residuals from ARMA(3,1,3)





Thus, the residuals from both the levelled and difference ARIMA models are stationary.

Evaluation using walk forward validation

- The ARIMA(3,1,3) model achieves a RMSE of 0.05147.
- So the ARIMA(3,1,3) model on the diffrenced yield spread does slighty better than the ARIMA(1,0 ,3) model on the yield spread.

ARIMA(1,0,3) vs ARIMA(3,1,3)

Figure: Actual vs. predicted yield spread in the test set from both levelled and difference ARIMA models

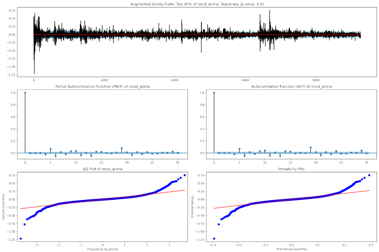
ARCH and GARCH Models

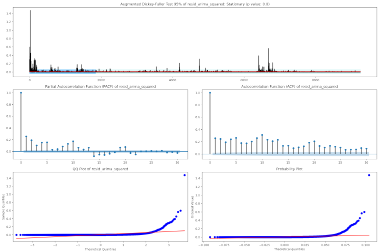
Mainly, the deficiencies of ARIMA models are ;

- They are not conditionally heteroskedastic i.e. they don't account for volatility clustering.
- The forecast width of ARIMA models is fixed as it linearly models the data.

In financial economics, a change in variance in one time may change the variance in the same direction in the subsequent time, as observed in the yield spread data.

Stationarity of `resid-arima` and `resid-arima-squared`





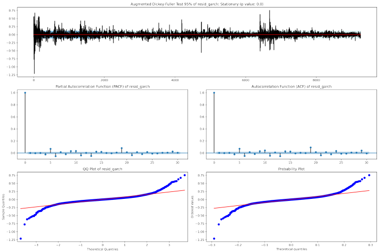
Whilst the residuals from ARIMA(1,0,3) appear to be white noise process, the squared residuals are highly auto-correlated.
The slow decay of successive lags indicate conditional hetero-skedasticity.

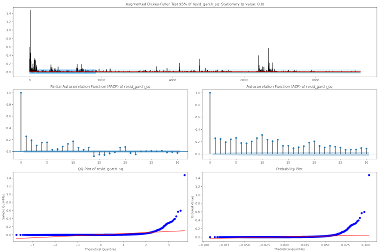
Fitting an ARCH(1,0,3) model

- Given that the lags in both the PACF and ACF plots are significant, the model would be better fit with both AR and MA parameters.
- So, we have fit a GARCH(1,3) model and checked how the residuals from GARCH(1,3) behave.

$$\sigma_t^2 = 1.7556 * 10^{-03} + 0.1142\sigma_{t-1}^2 + 0.8\epsilon_{t-1}^2 + 7.4291 * 10^{-12}\epsilon_{t-2}^2 + 0.0859\epsilon_{t-3}^2$$

Stationarity of `resid-garch` and `resid-garch-squared`





GARCH(1, 3) has not "explained" the serial correlation present in the squared residuals, inhibiting us from predicting in the test set.

Multivariate forecasting

The author uses the following exogenous variables to forecast the yield spread:

1. sahm : Real-time Sahm Rule recession indicator
2. forward1yr : Fitted instantaneous forward rate 1-year rate hence
3. recind : NBER based U.S. recession indicator
4. termpr : Term premium on a 10-year zero coupon bond
5. infexp : University of Michigan Inflation Expectations
6. vix : CBOE Volatility Index
7. ted : TED spread
8. 1yrff : 1-year Treasury constant maturity minus the federal funds rate

D'Agostino's K^2 Test

We use goodness-of-fit measure to check if the sample data on the yield spread arises from a Gaussian distributed population. Transforming the sample skewness and kurtosis derives the test statistic.

D'Agostino's K^2 Test Statistic

Ho: data is normally distributed
Ha: data is kurtic and/or skewed (not normally distributed)

The yield spread data is moderately skewed as its skewness is 0.0788 , and has heavy tails (very kurtotic) because the kurtosis is 1.0258.

Stationarity of the exogenous variables

Variable	p-Value	Stationary?
yieldsp	0.1268	No
ted	0.000	Yes
forward1yr	0.5191	No
recind	0.0035	Yes
1yr-ffr	0.000	Yes
termpr	0.595	No
sahm	0.000	Yes
infexp	0.000	Yes
vix	0.000	Yes

Table: p Values of the exogenous variables from Aug. DF test

So, we take the first difference for yieldsp, termpr and forward1yr. After taking the first difference, these variables become stationary,

ARIMA with Exogenous variables : SARIMAX(2,1,2)

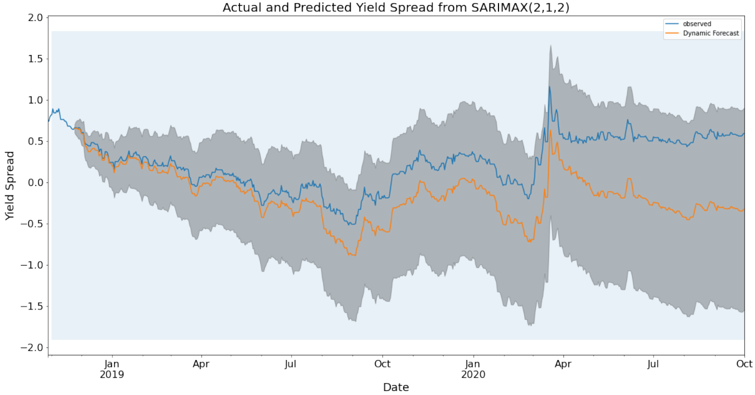
The author has modeled SARIMAX(2,1,2) without any seasonal component and incorporated the stationary exogenous variables.

The SARIMAX(2,1,2) model is:

$$\begin{aligned} yieldsp_t = {} & 0.4031 ted_t - 0.0072 \mathrm{rec\_ind}_t + 0.00651 \mathrm{yrffr}_t - 0.0002 vix_t \\ & + 2.4013 \Delta termpr_t - 0.4190 \Delta forward1yr_t - 0.0168 \mathrm{infexp}_t + 0.0259 sahm_t \\ & + 0.00651 \mathrm{yrffr}_t + 1.0625 yieldsp_{t-1} - 0.6652 yieldsp_{t-2} - 0.9914 \epsilon_{t-1} + 0.5403 \epsilon_{t-2} \end{aligned}$$

with an AIC of -49768.613

Forecasting using SARIMAX(2,1,2)



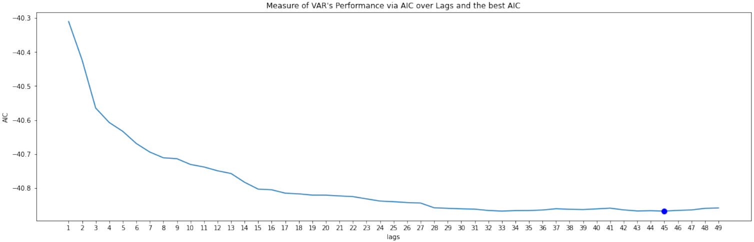
RMSE

The Root Mean Squared Error of the SARIMAX(2,1,2) model is 0.4744

Vector Autoregression (VAR)

- VAR(l) forecasts yieldsp using the l lagged values of itself and the stationary variables.
- To approximate the process well in the absence of MA terms, we may require a larger AR order.
- Thus, we have first determined the optimal VAR order from the lag order criteria AIC.

Determining the optimal lag order



Building VAR(45) model

VAR(45) is a seemingly unrelated regression (SUR) model with k = 7 deterministic regressors and l = 45 lagged variables of the form:

$$\Delta yieldsp_t = \alpha_1 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{\Delta yieldsp} \Delta yieldsp_{t-1} + \epsilon_{\Delta yieldsp_t}$$

$$\Delta termpr_t = \alpha_2 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{\Delta termpr} \Delta yieldsp_{t-1} + \epsilon_{\Delta termpr_t}$$

$$\Delta forward1yr_t = \alpha_3 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{\Delta forward1yr} \Delta yieldsp_{t-1} + \epsilon_{\Delta forward1yr_t}$$

$$\text{1yr-frr}_t = \alpha_4 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{\text{1yr-frr}} \Delta yieldsp_{t-1} + \epsilon_{\text{1yr-frr}_t}$$

$$\text{rec\_ind}_t = \alpha_5 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{\text{rec\_ind}} \Delta yieldsp_{t-1} + \epsilon_{\text{rec\_ind}_t}$$

$$ted_t = \alpha_6 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{ted} \Delta yieldsp_{t-1} + \epsilon_{ted_t}$$

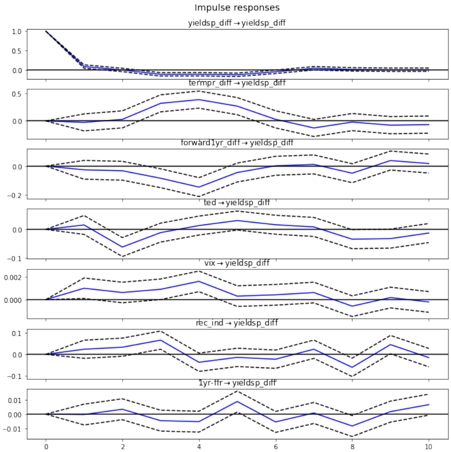
$$vix_t = \alpha_7 + \sum_{i=1}^{7}\sum_{l=1}^{45} \beta_{i,l}^{vix} \Delta yieldsp_{t-1} + \epsilon_{vix_t}$$

where $\epsilon_t = [\epsilon_{\Delta yieldsp_t} \;\; \epsilon_{\Delta termpr_t} \;\; \epsilon_{\Delta forward1yr_t} \;\; \epsilon_{1yrffr_t} \;\; \epsilon_{\text{rec\_ind}_t} \;\; \epsilon_{ted_t} \;\; \epsilon_{vix_t}]'$ is a white noise process of "innovations".

RMSE

The Root Mean Squared Error of the VAR(45) model is 0.50267

Impulse Response Function (IRF)



The impulse response of shocks on yieldsp decays to 0 by the tenth period, signifying that one-time innovations don't have long-term ramifications on the paths of yieldsp .

Multilayer Perceptrons

- A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN)
- MLPs use simple linear relations between consecutive layers to effectively model even the most complex functions.

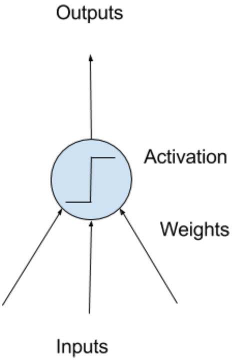


Figure: A model of a single neuron

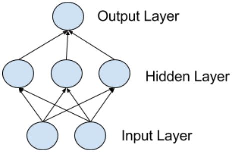


Figure: Model of a Simple Network

Building, training and evaluating the various MLP network

The author has made various MLP network by varying the number of neurons (1,3 and 5) with the lag period = 3. The author has also run every experiment 8 times.

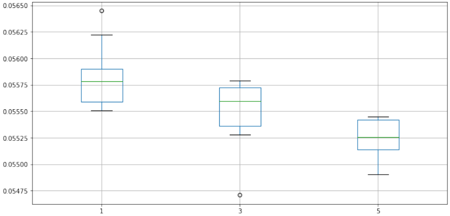


Figure: Box plot that shows the lowest RMSE occurs with 5 neurons in the hidden layer in MLP.

Overfitting?

We have plot the test and train RMSE for each of the 10 runs.

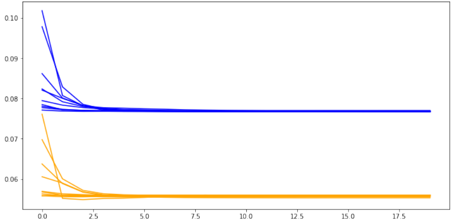


Figure: Box plot that shows the lowest RMSE occurs with 5 neurons in the hidden layer in MLP.

The results indicate that the model runs well without over-fitting as the line plot of the test RMSE (yellow color line) is lower than the train RMSE (blue color line).

Persistence issue in MLPs and RNNs

- Traditional neural networks do not have persistence, and it is a major shortcoming in time series applications.
- Recurrent neural networks address this issue. They are networks with loops in them, allowing information to persist.

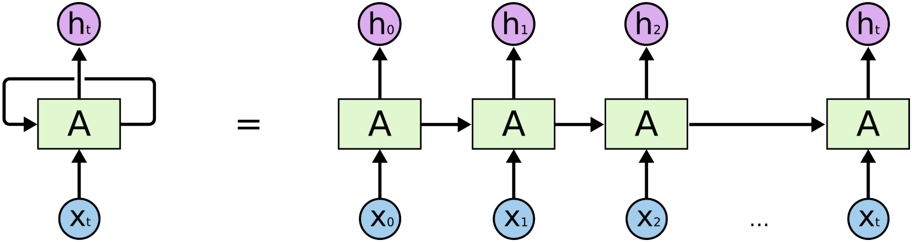


Figure: An unrolled recurrent neural network.

Long Short Term Memory Models

LSTMs are explicitly designed to avoid the long-term dependency problem.

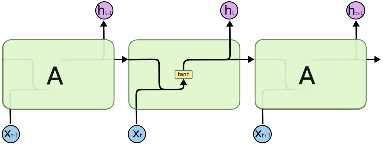


Figure: The repeating module in a standard RNN contains a single layer.

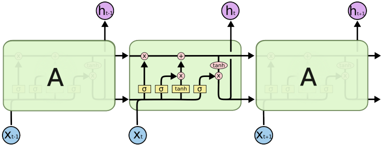
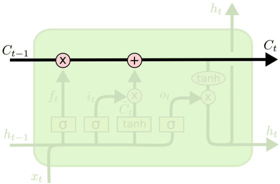


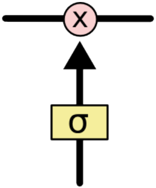
Figure: The repeating module in an LSTM contains four interacting layers.

Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

Core Idea behind LSTMs

The key to LSTMs is the cell state, the horizontal line running through the top of the diagram.

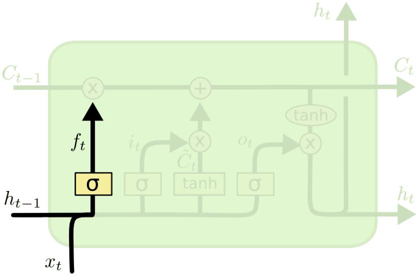




The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates.

Forget Gate

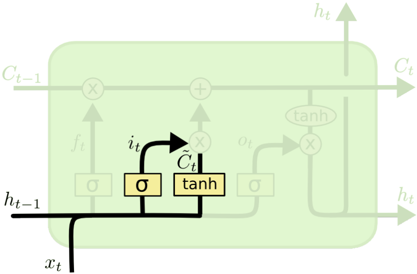
The first step in our LSTM is to decide what information we're going to throw away from the cell state. This decision is made by a sigmoid layer called the "forget gate layer."



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

Store state

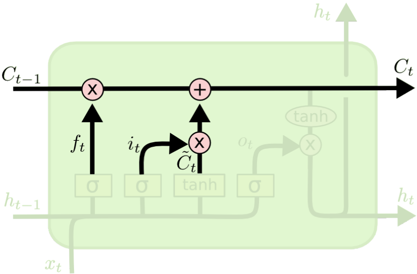
The next step is to decide what new information we're going to store in the cell state.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

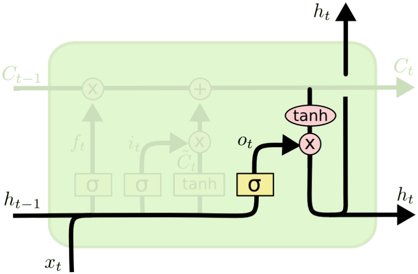
Updating the cell state



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Output Gate

Finally, we need to decide what we're going to output. This output will be based on our cell state, but will be a filtered version.



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

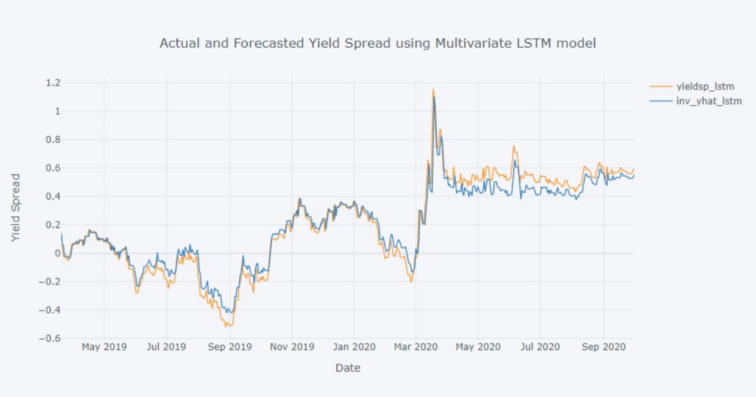
$$h_t = o_t * \tanh\left(C_t\right)$$

Modelling using stateful LSTMs

RMSE

The Root Mean Squared Error of the stateful LSTM model is 0.06081

LSTM using multivariate data



RMSE

The Root Mean Squared Error of the stateful LSTM model is 0.07118

Conclusion

Model	RMSE
ARIMA(1,0,3)	0.05110
Differenced ARIMA(3,1,3)	0.05147
SARIMAX(2,1,2)	0.4744
VAR(45)	0.50267
MLP	0.055239
Stateful stacked LSTM	0.06081
Multivariate LSTM	0.07118

Thank You!