# Use of Hypothetical Data in Machine Learning Trading Strategies


Ankit Awasthi

Quantitative Portfolio Manager

qplum

# What will we cover today

Evolution in trading and machine learning

Data Augmentation in Deep Learning

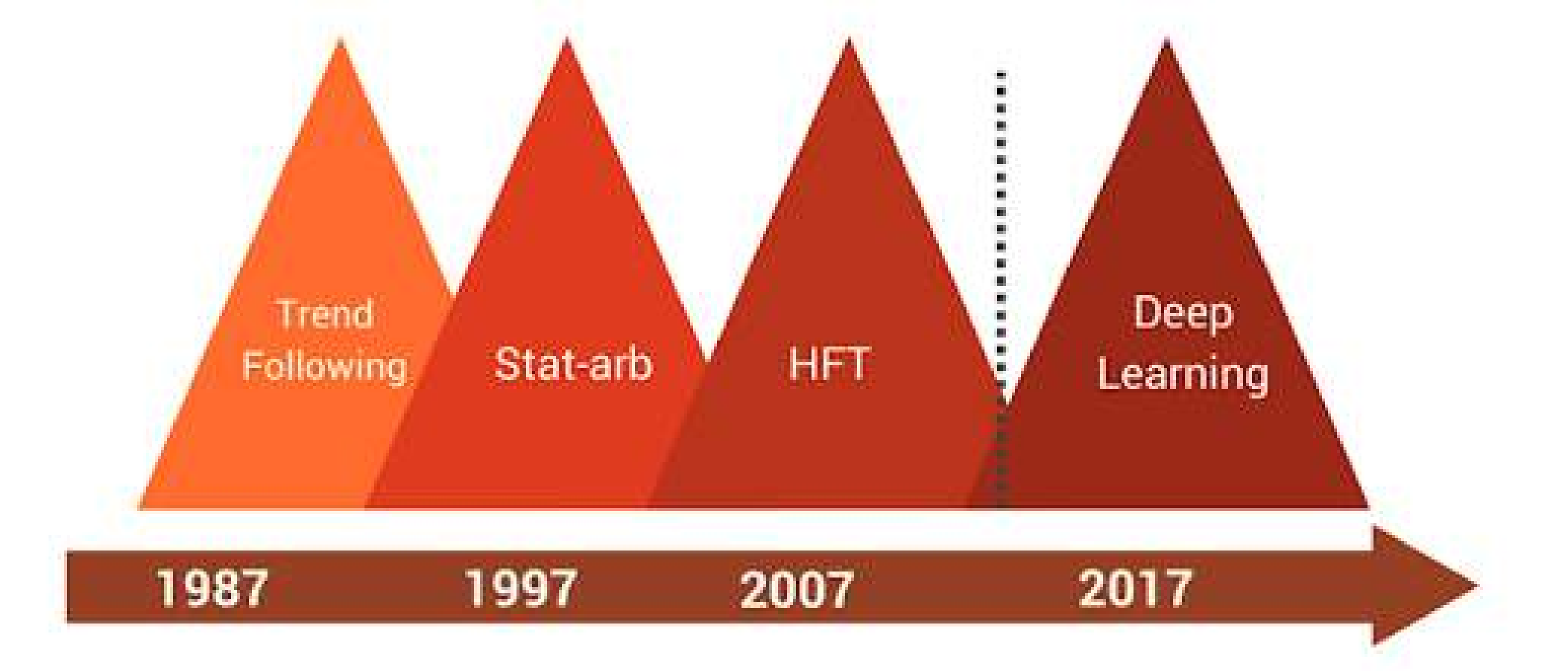
How to generate hypothetical data for trading strategies

Stylized facts of time series of asset returns

Improving deep learning models for Finance

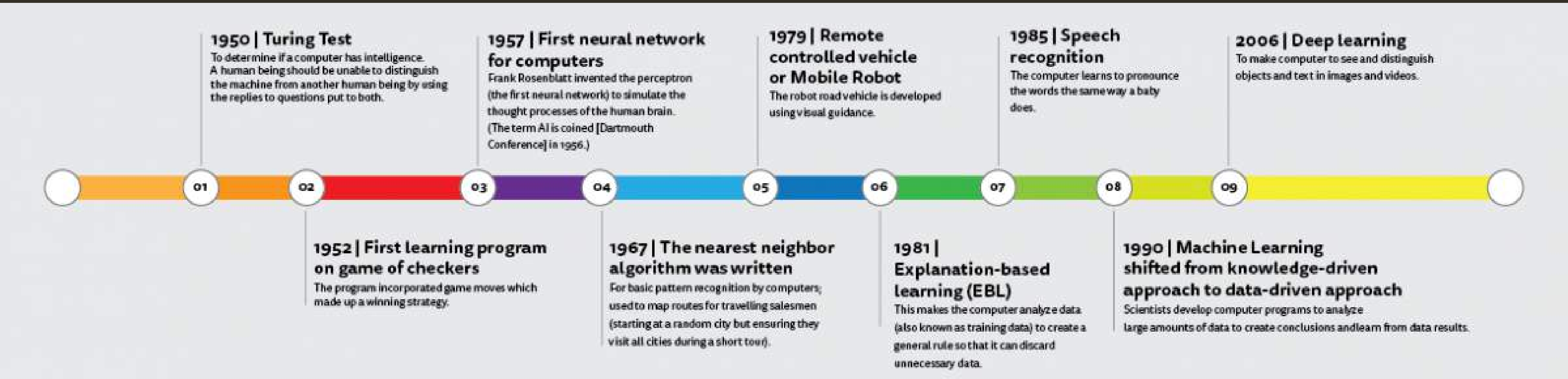
Future Testing Strategies

# Ten Year Cycles of Innovations in Trading




Source : May 2017 Research, qplum

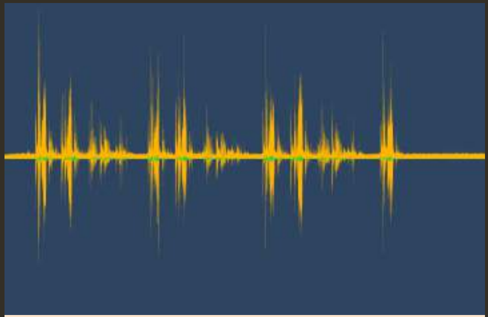
# Evolution of Machine Learning




Source : Evolution of Machine Learning, Synectics

# Tremendous Success of Deep Learning

Image

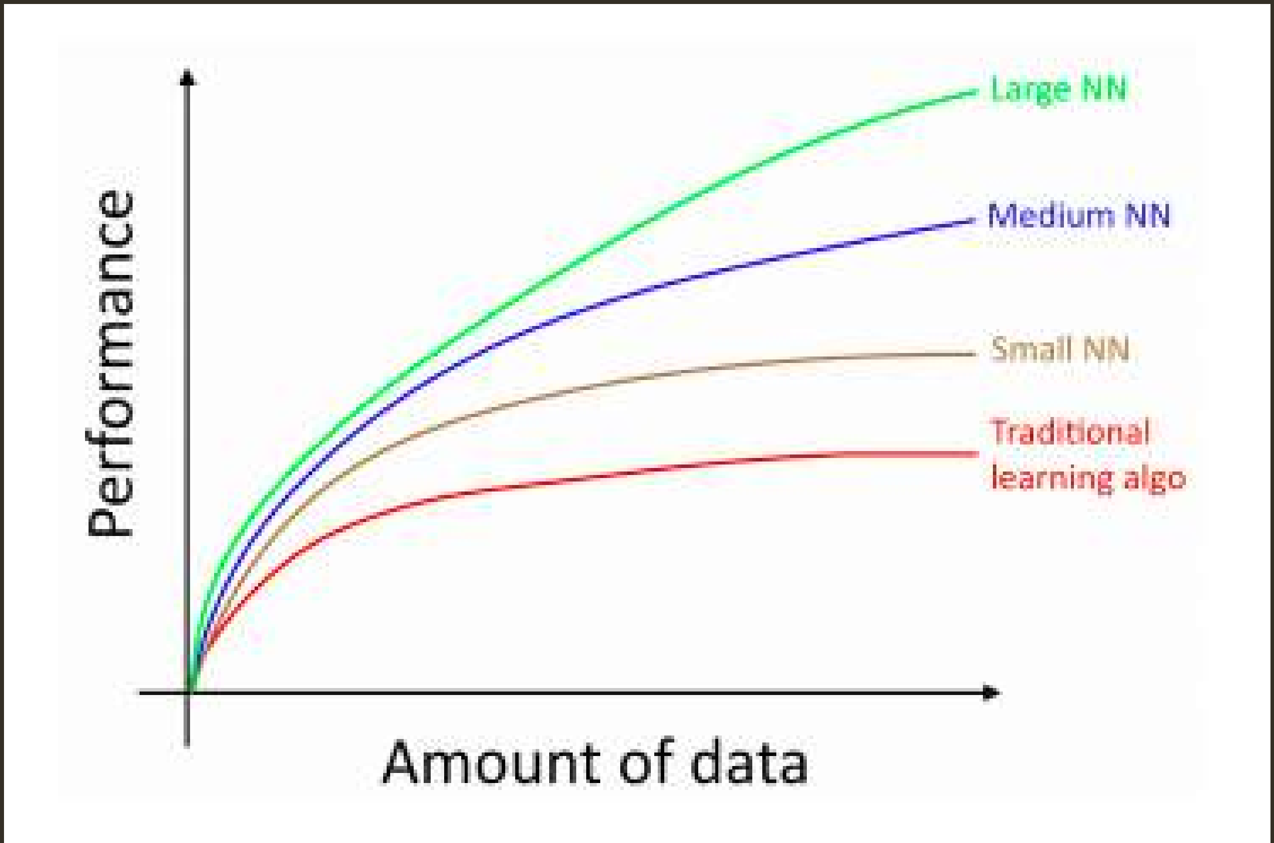


Speech


Text

## Abundance of Data!

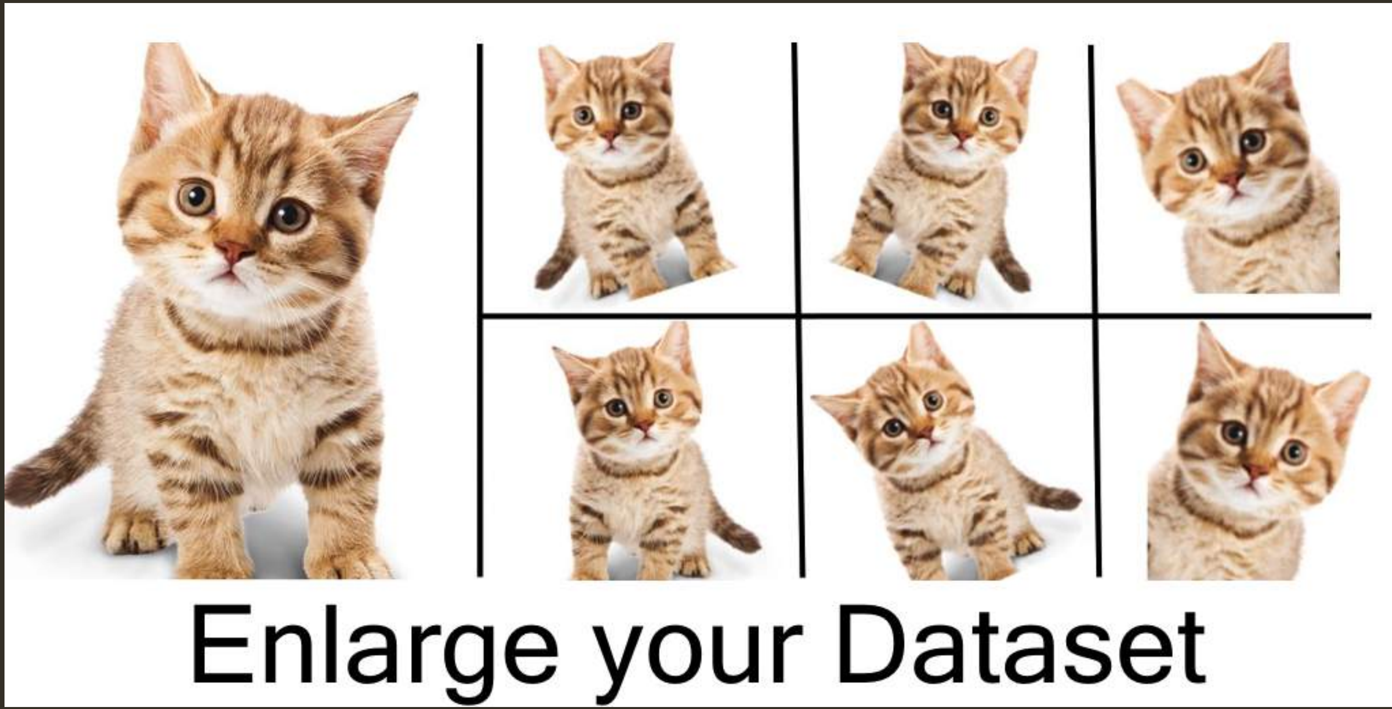
# Deep learning scales with data




Source : Andrew Ng, Coursera Deep learning specialization

# Data Augmentation for Deep Learning

Even more data!




Much harder in Finance!

Source : Bharat Raj, Data Augmentation I How to use Deep Learning when you have Limited Data

# Generating Hypothetical Data for Trading

- Adding noise or transformations
  - Data generated is similar to original so doesn't lead to great generalization
  - Higher magnitude noise can distort properties of financial time series
- Theoretical models
  - Hard to estimate and sample from
  - Multi-variate distributions are problematic
  - Hard to account for all empirical properties of financial time series
- Appropriating High Frequency Data
  - Shows remarkable similarity to Daily Data
  - Data acquisition is hard and costly

# Using High Frequency Data as Daily Data

Step 1 : Generate 15-minute contiguous,non-overlapping samples from high frequency tick-data

Step 2 : Compute Mean and Variance for daily returns as Daily_Mean and Daily_Variance

Step 3 : Transform High Frequency Data to have the mean as Daily_Mean and variance as Daily_Variance

# Stylized Effects of Financial Time Series

Absence of Autocorrelations

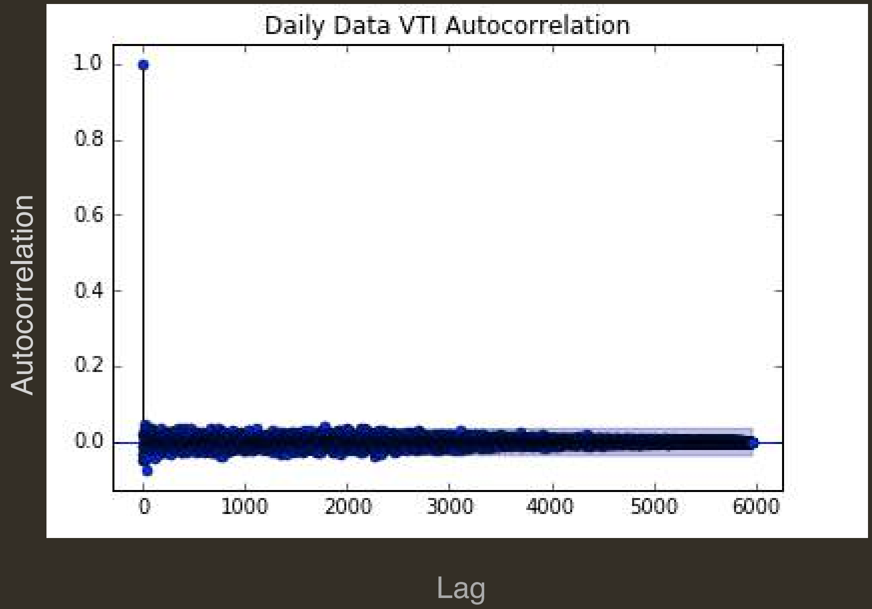
Heavy Tails

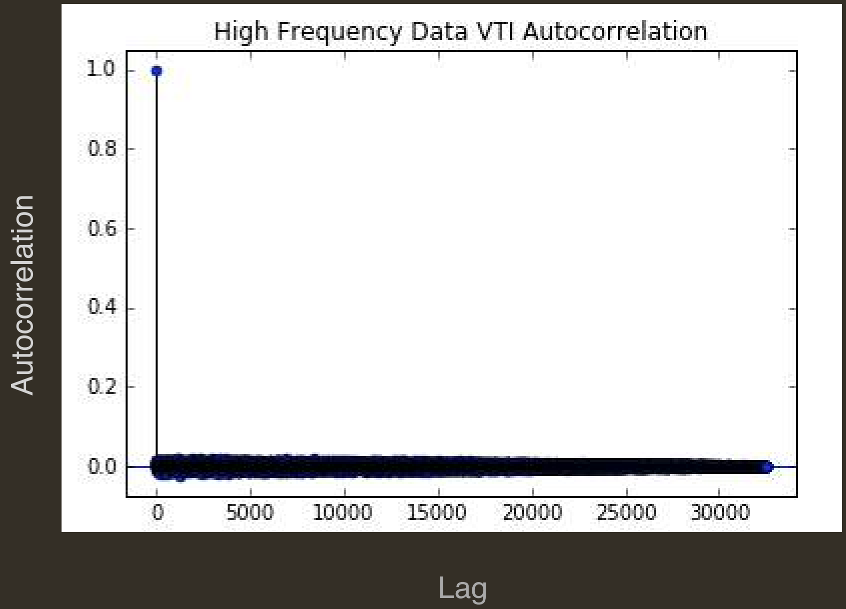
Volatility Clustering

Leverage Effect

Cross Correlation vs Volatility

# Absence of Autocorrelations

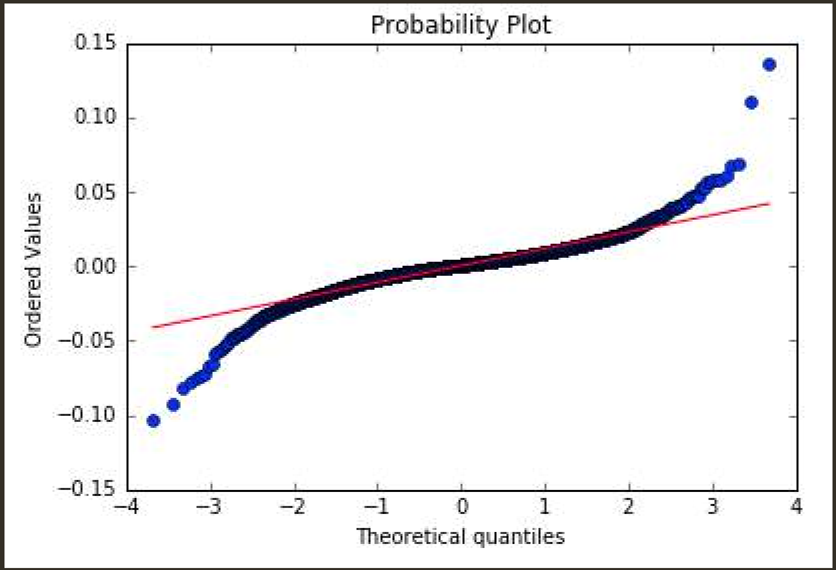


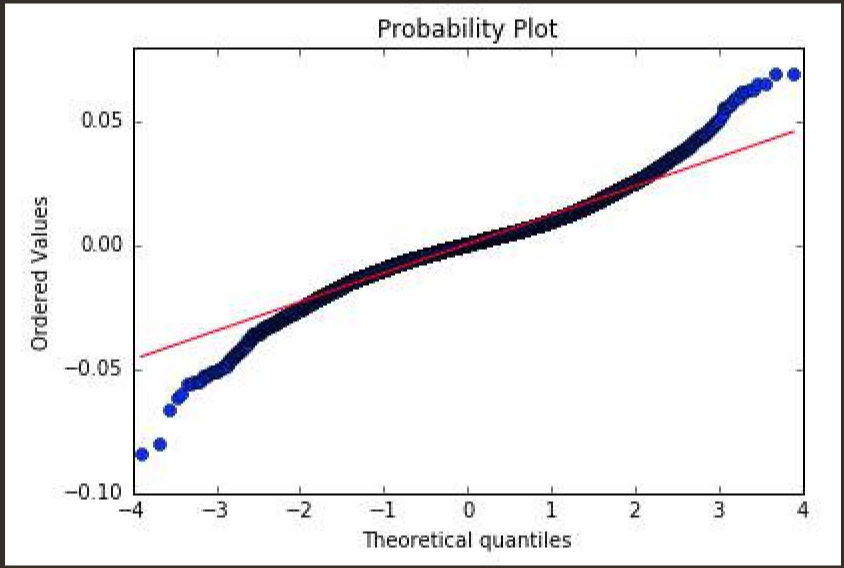




Autocorrelation of returns is low both in daily and high frequency data

Source : qplum Research

# Heavy Tails



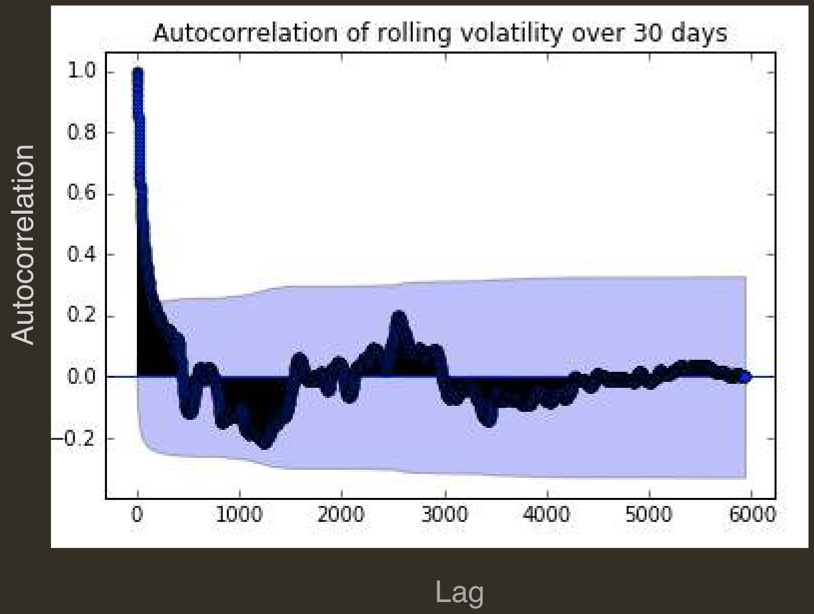



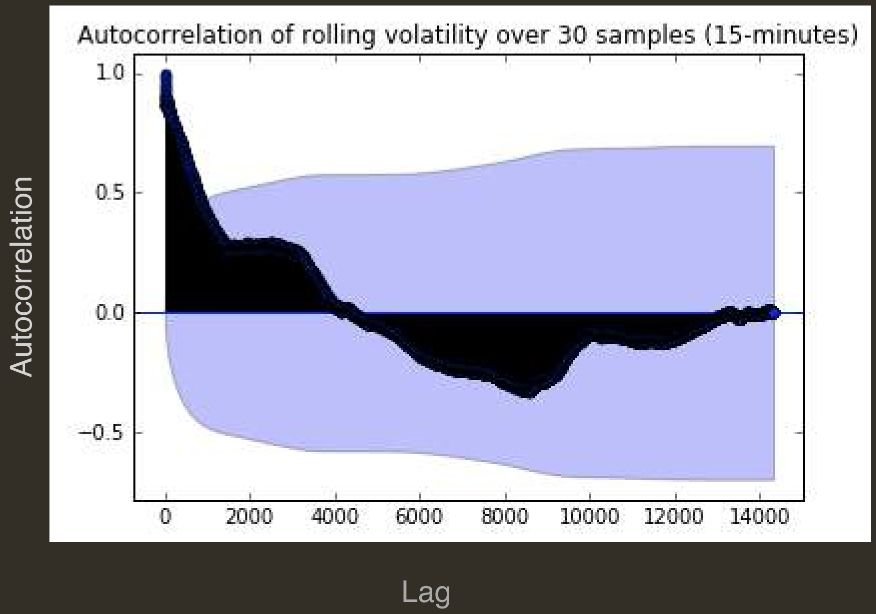

High frequency data exhibits divergence from normal similar to daily returns

The magnitude of the divergence, however, is much more pronounced for daily returns

Source : qplum Research

# Volatility Clustering



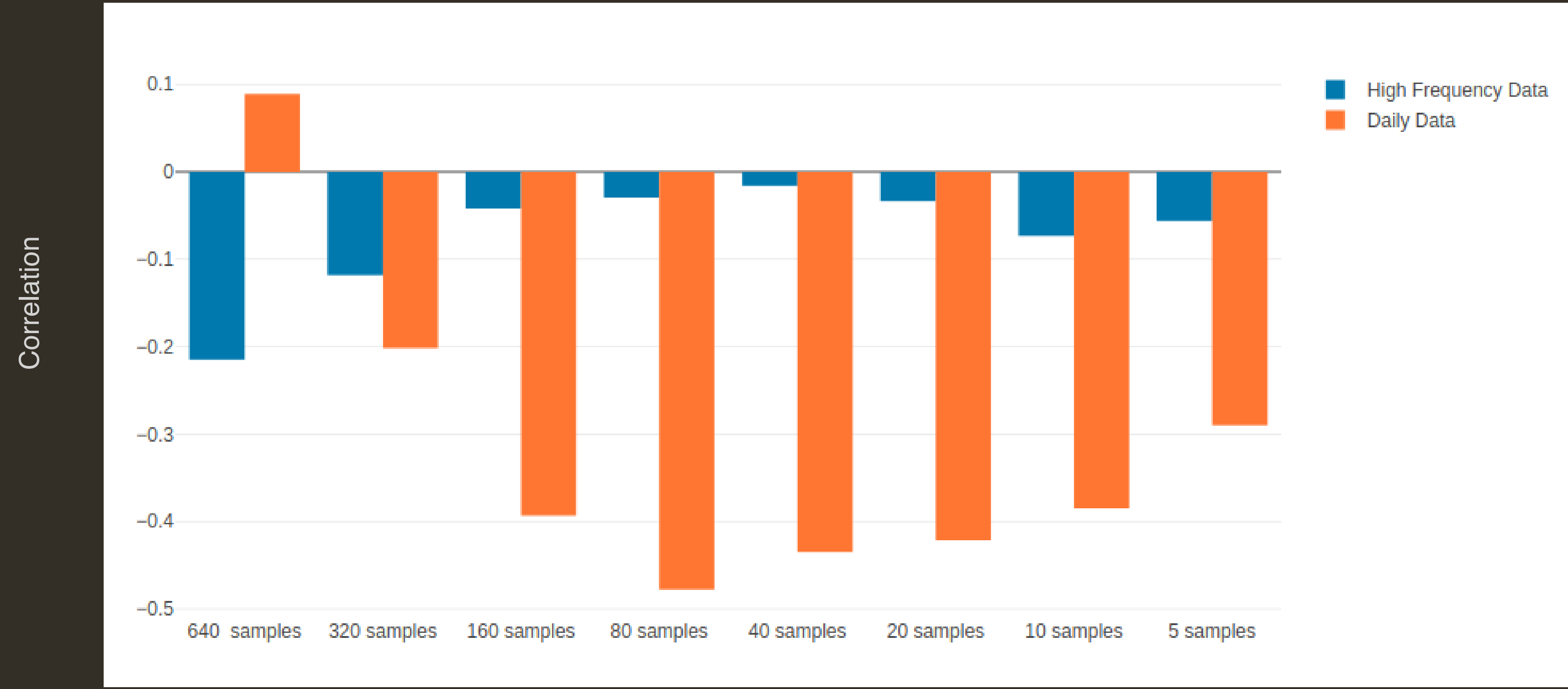




As opposed to returns, volatility of daily returns shows significant autocorrelation, which is retained in high frequency data as well

Source : qplum Research

# Leverage Effect

...refers to the observed tendency of an asset's volatility to be negatively correlated with the asset's returns

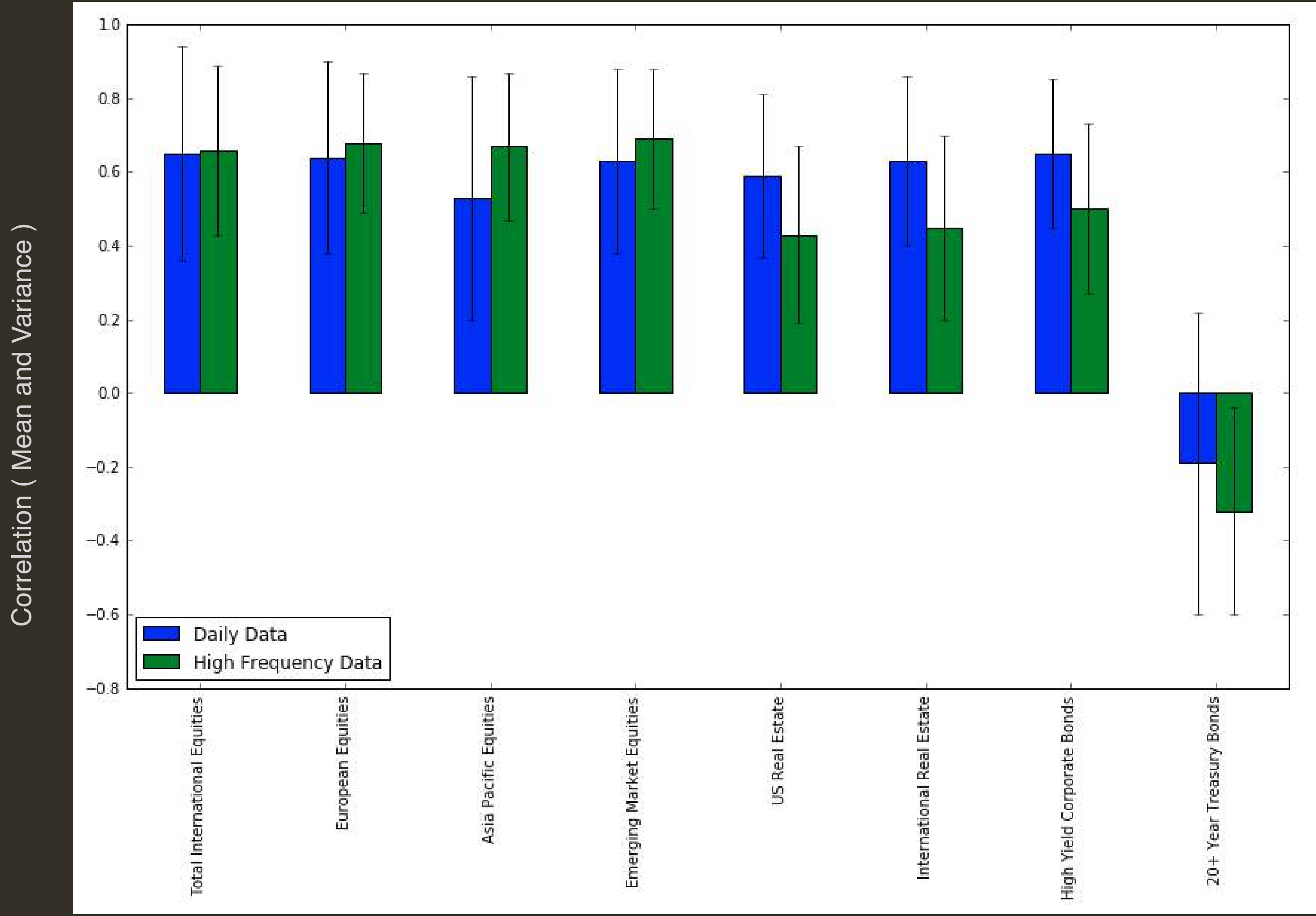



Correlation of volatility with returns for US Total Market Index (VTI)

Leverage effect is also exhibited by high frequency data but to a smaller extent

Source : qplum Research

# Cross-Correlation Vs Asset Volatility

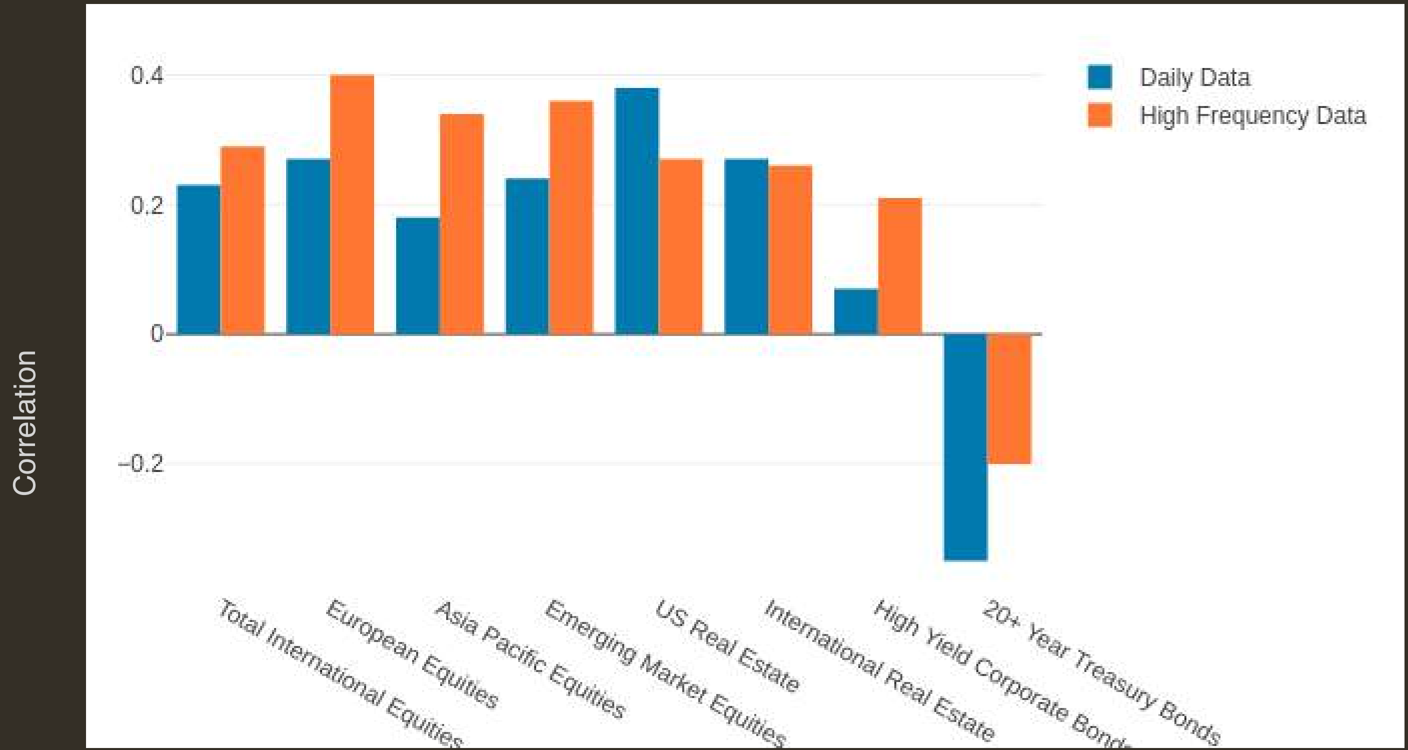



Mean of 30-day rolling correlation of different assets with US Total Stock Market (VTI)

Standard Deviation is shown as error bars.

Source : qplum Research

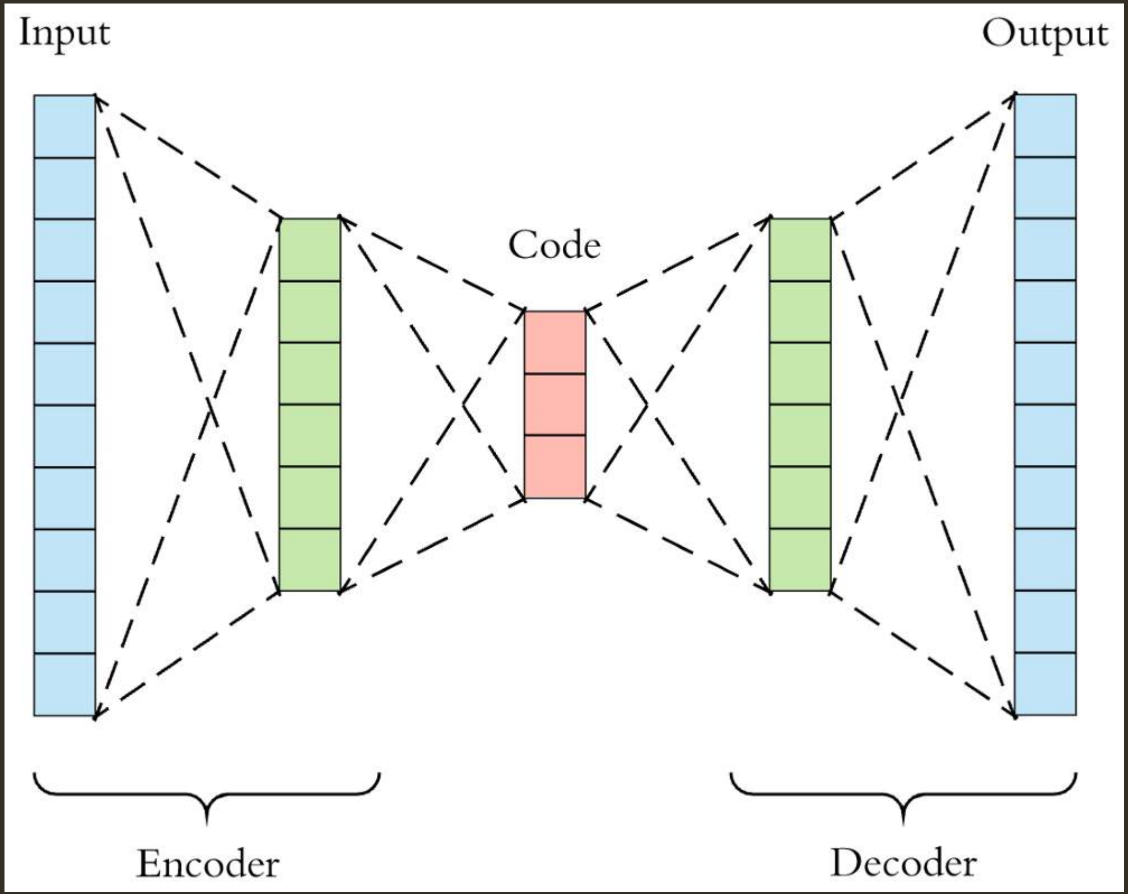
# Correlation increases with volatility




Correlation of 30-day rolling correlation of log returns of different assets vs US Total Stock Market against the volatility of US Total Stock Market.

Source : qplum Research

# A simple example - Autoencoders



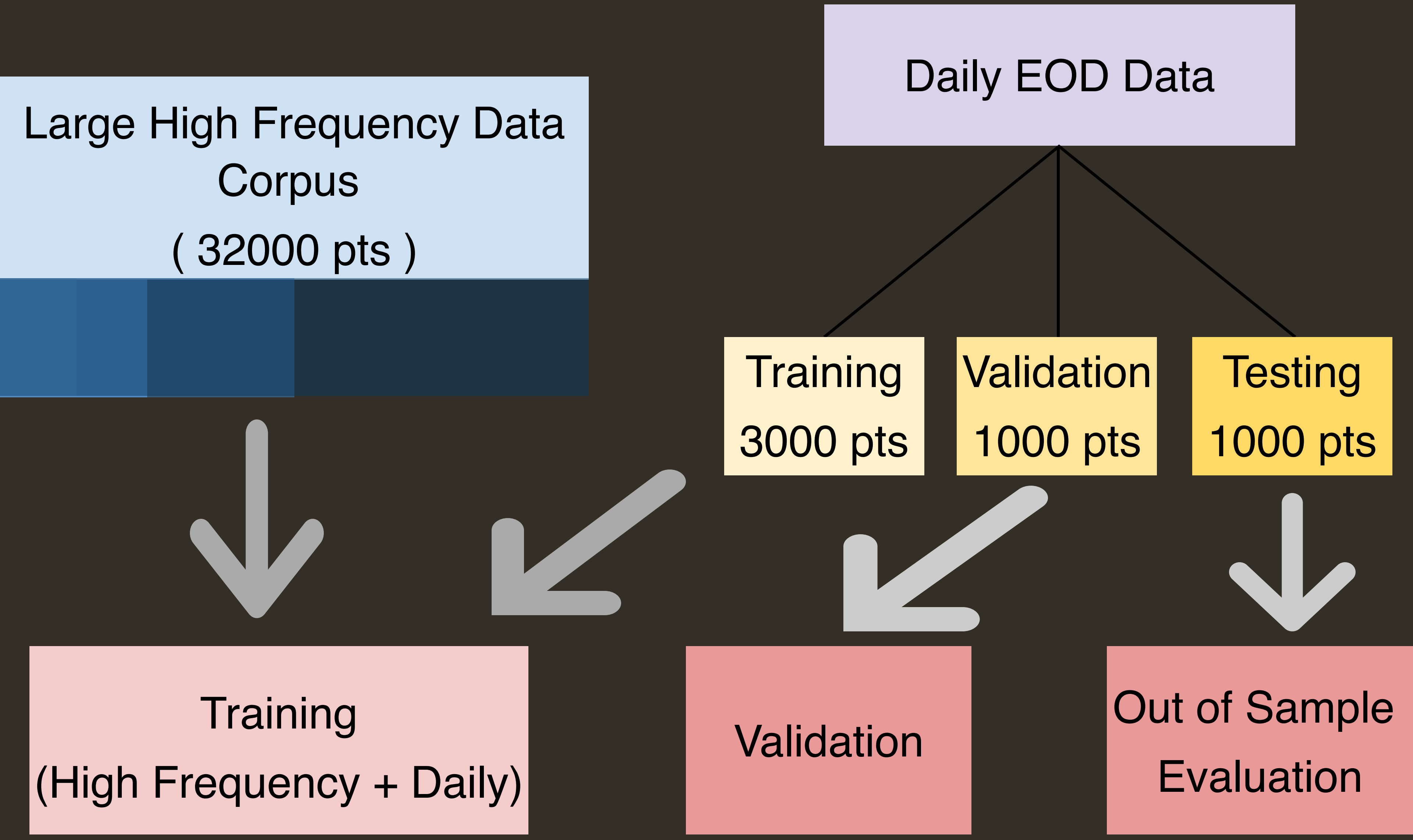

Source : Arden Dertat, Applied Deep Learning

# Feature Learning

- Factorization of price-volume features across multiple securities
- Unsupervised - It is not used to predict returns, or maximize performance metric
- Information bottleneck forces the network to learn the most important factors
- Captures non-linearities across time and cross section of securities

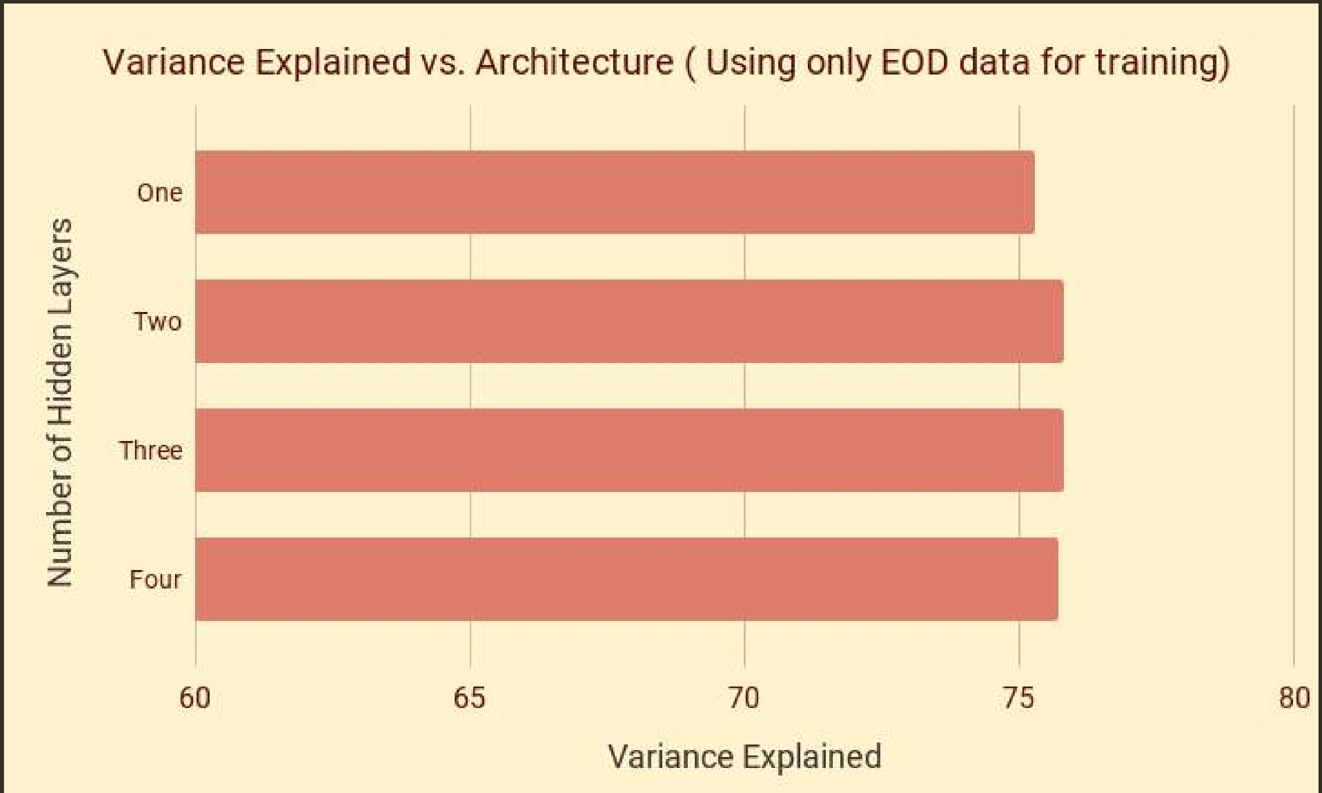
Source : Arden Dertat, Applied Deep Learning

# Experiment




Source : qplum Research

# Model complexity alone doesn't help much!




Source : qplum Research

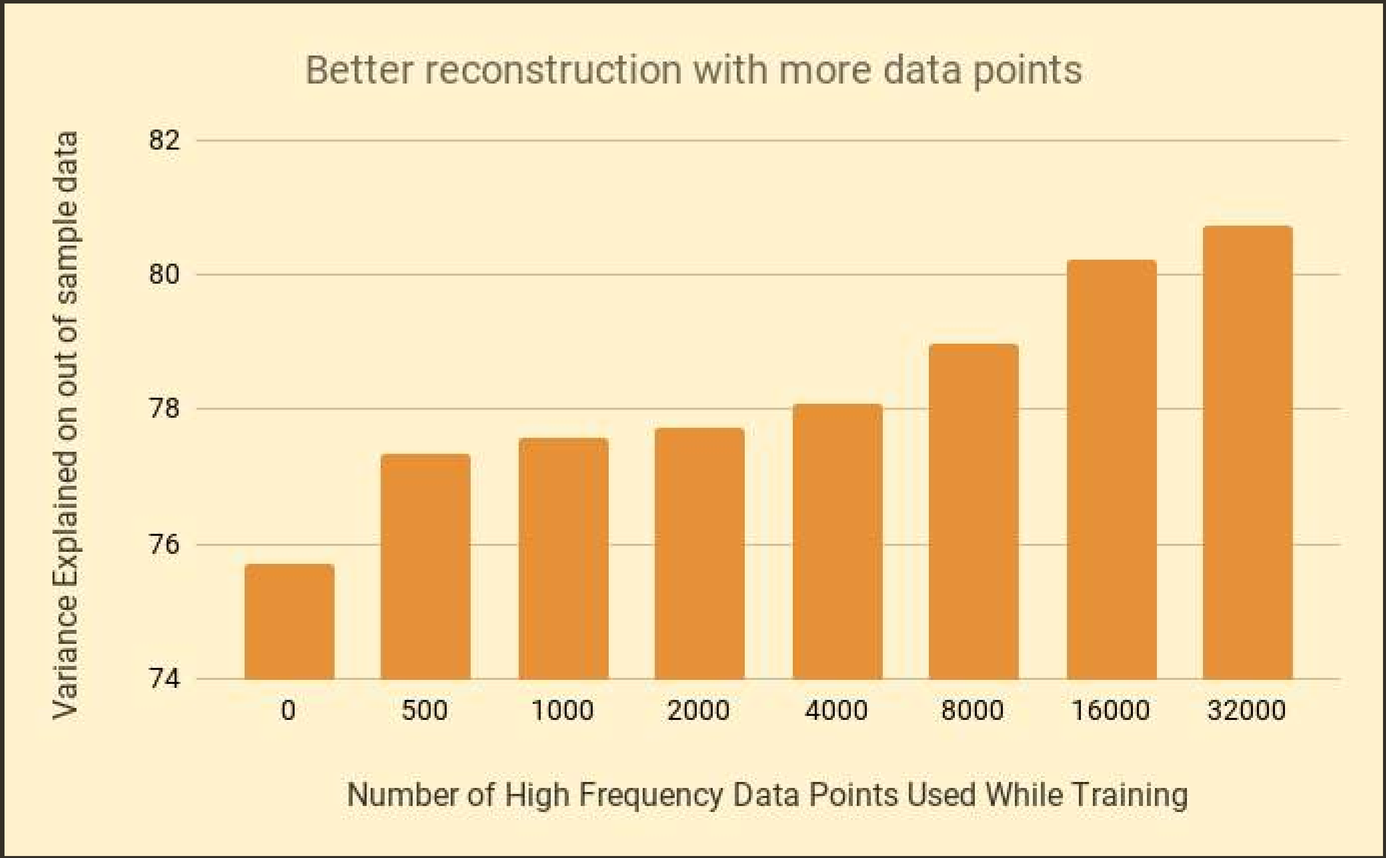
One Hidden Layer : 62 -> 3 -> 62

Two Hidden Layers : 62 -> 10 -> 3 -> 10 -> 62

Three Hidden Layers : 62 -> 20 -> 10 -> 3 -> 10 -> 20 -> 62

Four Hidden Layers : 62 -> 100 -> 20 -> 10 -> 3 -> 10 -> 20 -> 100 -> 62

# However....more data does!




Architecture Used : 62 -> 100 -> 20 -> 10 -> 3 -> 10 -> 20 -> 100 -> 62

Source : qplum Research

# Scenario Analysis

Helps to be data-driven with more data!

- Multiple instances of *'market crashes'*
- Easy to find custom scenarios e.g., fixed income positively correlated to equities
- Periods of high or low divergence across different sectors in equity markets
- Multiple inflationary/deflationary periods as measured through commodity prices

# Cross-Validating Hyper-parameters

## Helps to be data-driven with more data!

- Risk model parameters
  - Sensitivity to lookback duration
  - Combinations of different metrics like expected shortfall, current drawdown
- Regularization constants
  - Soft constraints in multi-objective portfolio optimization - constraining active risk and risk of the portfolio
- Tax Optimization parameters
  - Thresholds for deferring gains, booking losses as a function of volatility

# Future Testing Strategies

*" It is far better to forsee even without certainty than not to forsee at all " - Henri Poincare*

- Collect different capital market assumptions
  - Qplum's Internal Expected Returns
  - JP Morgan Capital Market Assumptions
  - Research Affiliates Capital Market Assumptions
  - ....
- Transform high frequency data
  - Set mean of returns such that annualized returns match the respective capital market assumptions
  - Set variance of returns such that annualized volatility matches the respective capital market assumptions
- Evaluate performance of strategies under different capital market assumptions

# Key Takeaways

Deep learning works best when you have lots of data

Adapting high frequency returns is one of the easiest ways of augmenting daily returns data

Hypothetical data can be used to stress test strategy performance

Hypothetical data can be used to develop better and more robust strategies

# Questions?

contact@qplum.co

Disclosures: All investments carry risk. Information presented is for educational purposes only and does not intend to make an offer or solicitation for the sale or purchase of any specific securities, investments, or investment strategies. Investments involve risk and are never guaranteed. Be sure to first consult with a qualified financial adviser and/or tax professional before implementing any strategy discussed herein. Past performance is not indicative of future performance. Important information relating to qplum and its registration with the Securities and Exchange Commission (SEC), and the National Futures Association (NFA) is available here and here.