# Uncertainty-Aware Lookahead Factor Models for Improved Quantitative Investing

Lakshay Chauhan, John Alberg, Zachary Lipton

# Overview

Improve quantitative investing by forecasting fundamentals and measuring uncertainty

## Quantitative Investing

- Portfolios are constructed by ranking stocks using a **factor**
- **factors** based on fundamentals such as Revenue, Income, Debt
- Standard quantitative investing uses current fundamentals
- Investment success $\longrightarrow$ what a company does in the future

Can we use forecast future fundamentals then?

# Overview

## Improve quantitative investing by forecasting fundamentals and measuring uncertainty

### Our Contribution

- Show value of forecasting fundamentals
- Forecast future fundamentals using neural networks and measure uncertainty
- Use uncertainty estimate to reduce risk as measured by Sharpe Ratio
- Portfolio return and risk are significantly improved

# Motivation

## Quantitative Investing

**Ranking Factor - Dividend Yield**

Southern
Hewlett Packard
Pfizer
Verizon
US Bancorp
Duke Energy
Leggett & Platt
Omnicom Group
⋮

Portfolio
Pick top N ranked by a **factor**

**Factors**

Dividend Yield
Earnings Yield
Book-to-Market
Momentum

**Value Factors**

$$\frac{Fundamental\ Item\ (net\ income, EBIT)}{Stock\ Price}$$

Value factors outperform market averages (SP500)

## Limitation

Factor models rely on current period fundamentals, but returns are driven by future fundamentals

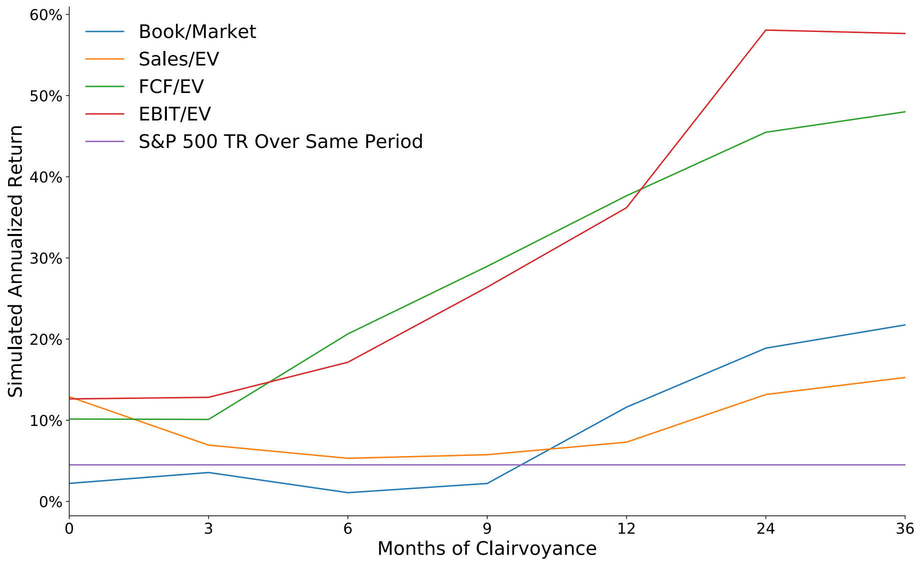
## Solution

Build factor models using forecast future fundamentals

# Motivation

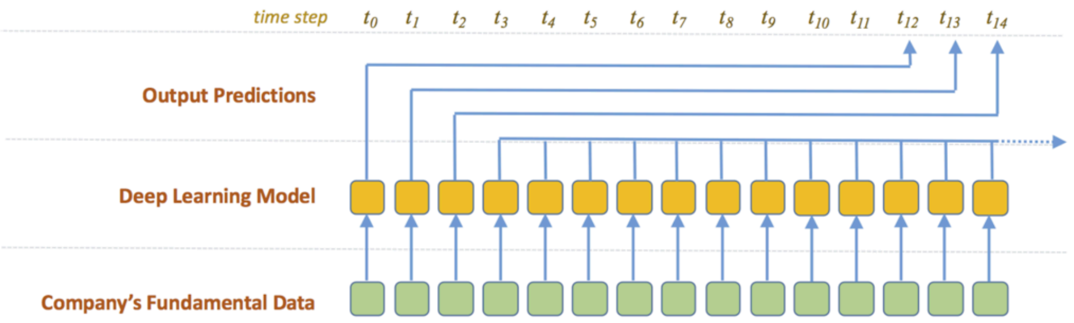
## Clairvoyant Factor Model

- Imagine we had access to **future** fundamentals
- Simulate performance with **future** fundamentals (2000-2019)
- Clairvoyant fundamentals offer substantial advantage
- This motivates us to forecast future fundamentals




## Problem Set up

- Use EBIT as the fundamental to create value-factor
- Forecast EBIT 12 months into the future



# Data Background

- US stocks from 1970-2019 traded on NYSE, NASDAQ and AMEX (~12,000), Market Cap > $100M
- Time series of 5 years with step size of 12 months

| | Input Series | | | | | Target |
|---|---|---|---|---|---|---|
| IBM | **Jan, 2000** | **Jan, 2001** | **Jan, 2002** | **Jan, 2003** | **Jan, 2004** | **Jan, 2005** |
| IBM | **Feb, 2000** | **Feb, 2001** | **Feb, 2002** | **Feb, 2003** | **Feb, 2004** | **Feb, 2005** |
| IBM | **Mar, 2000** | **Mar, 2001** | **Mar, 2002** | **Mar, 2003** | **Mar, 2004** | **Mar, 2005** |

- Feature Examples

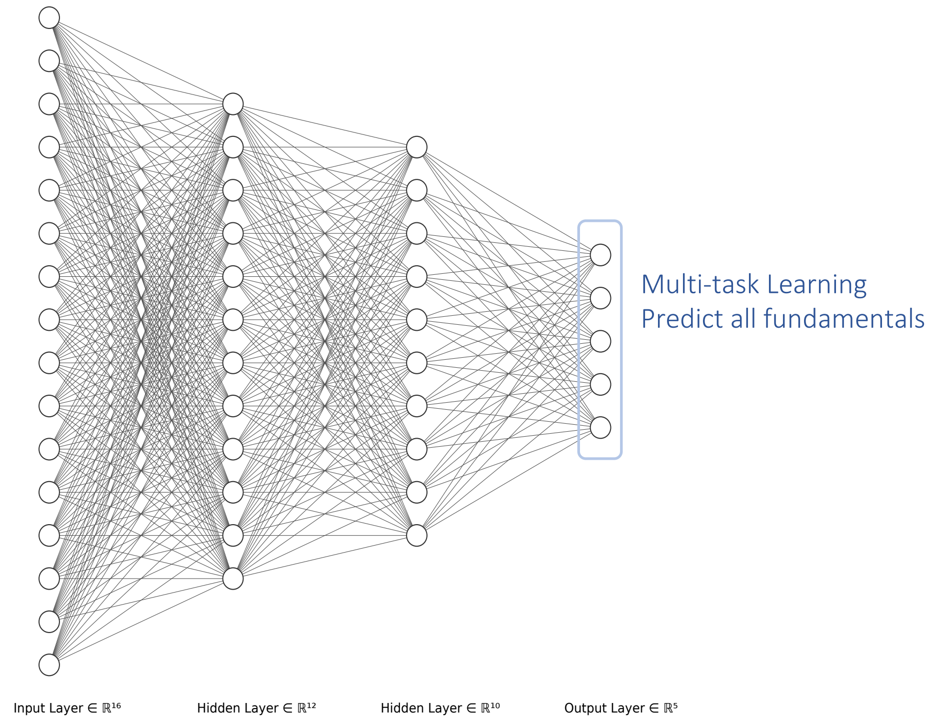
| **Fundamental** | **Momentum** | **Auxiliary** |
|---|---|---|
| Revenue | 1-month relative momentum | Short Interest |
| Cost of Goods Sold | 3-month relative momentum | Industry Group |
| Earnings Before Interest and Taxes (EBIT) | 6-month relative momentum | Company size category |
| Current Debt | 9-month relative momentum | |
| Long Term Debt | | |

# Forecasting Model

| 1970 | 2000 | 2019 |
|---|---|---|
| **Random 70-30 train-validation split** | **Out of Sample Test Set** | |

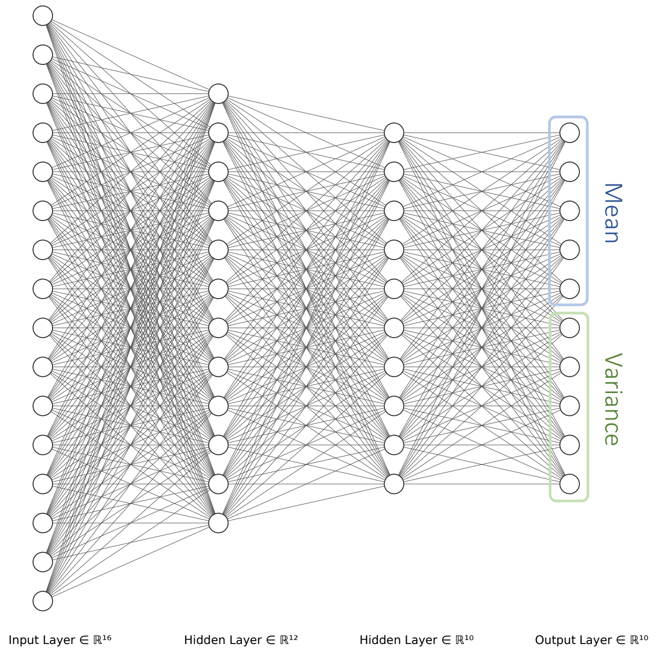
Training Set

- In-sample validation set is used for genetic algorithm based hyper-parameter tuning
- Multi-task learning to predict all fundamental features instead of just EBIT
  - Increases training signal
  - Improves generalization
- Use Max Norm and Dropout for regularization



# Uncertainty Quantification

- Financial data is heteroskedastic i.e. noise is data dependent
- Some companies will have more uncertainty in their earnings than others due to size, industry, etc.
- Jointly model mean and variance by splitting final layer
  - First half predicts means of targets $(f_\theta(\mathbf{x}))$
  - Second half predicts variance of the output values or aleatoric uncertainty $(g_\theta(\mathbf{x}))$




$$\boldsymbol{\theta}^{\mathrm{MLE}} = \max_\theta \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi g_\theta(\mathbf{x}_i)^2}} \exp\left(\frac{-(y_i - f_\theta(\mathbf{x_i}))^2}{2g_\theta(\mathbf{x}_i)^2}\right)$$

$$= \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(\underbrace{\log(g_\theta(\mathbf{x}_i))}_{\text{minimize uncertainty (narrow bounds)}} + \underbrace{\frac{\overbrace{(y_i - f_\theta(\mathbf{x_i}))^2}^{\text{prediction accuracy}}}{2g_\theta(\mathbf{x}_i)^2}}_{\text{penalize over-confident model}}\right)$$

Epistemic Uncertainty = Variance in outputs across Monte Carlo draws of dropout mask

Total Uncertainty = Aleatoric Uncertainty + Epistemic Uncertainty

# Constructing Factor Models

## Definitions

EV - Enterprise Value

$$market\ cap + net\ debt$$

QFM - Quantitative Factor Model

$$\frac{EBIT_{current}}{EV}$$

LFM - Lookahead Factor Model

$$\frac{EBIT_{forecast}}{EV}$$

LFM UQ – Uncertainty Quantified Model

$$\frac{EBIT_{forecast}}{\sigma^2 EV}$$

Companies with higher variance are riskier

- Higher variance = less certain about forecasts
- Therefore, scale factor in inverse proportion to variance

## Factor Models

QFM

LFM Auto Reg

LFM Linear

LFM MLP

LFM LSTM

LFM UQ-MLP

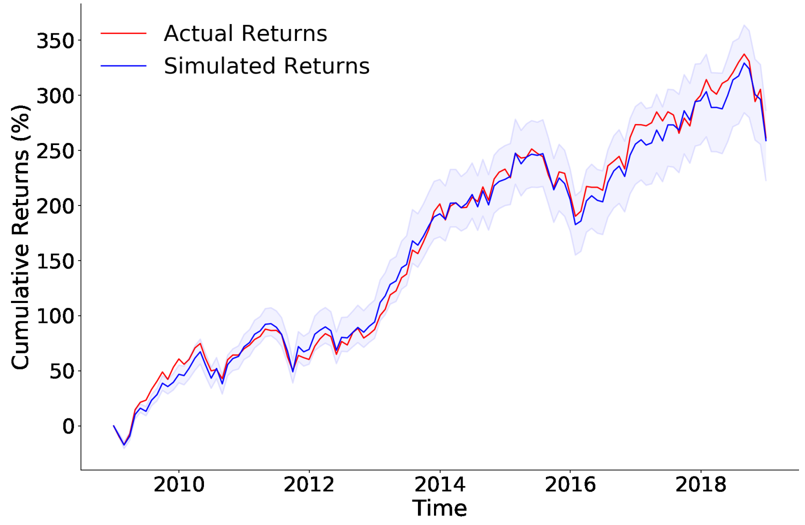
LFM UQ-LSTM

# Portfolio Simulation

- Industry grade, high fidelity investment portfolio simulator
- Portfolios formed of top 50 stocks ranked by factor
- Rebalance portfolio monthly
- Simulate 50 years of performance, many economic cycles
- Point-in-time data, no survivorship or look-ahead bias
- Include transactions cost, price slippage to reflect realistic trading
- Measure performance by Compound Annualized Return (CAR) and Sharpe Ratio

$$CAR = \left[\prod_{t}^{T}(r_t + 1)\right]^{12/T} - 1$$

$$\text{Sharpe Ratio} = \frac{CAR - R^f}{\sqrt{12}\sigma}$$




Simulated returns of a quantitative strategy vs. the real returns generated from live trading of the same strategy

# Results

## Out-of-Sample Performance 2000-2019

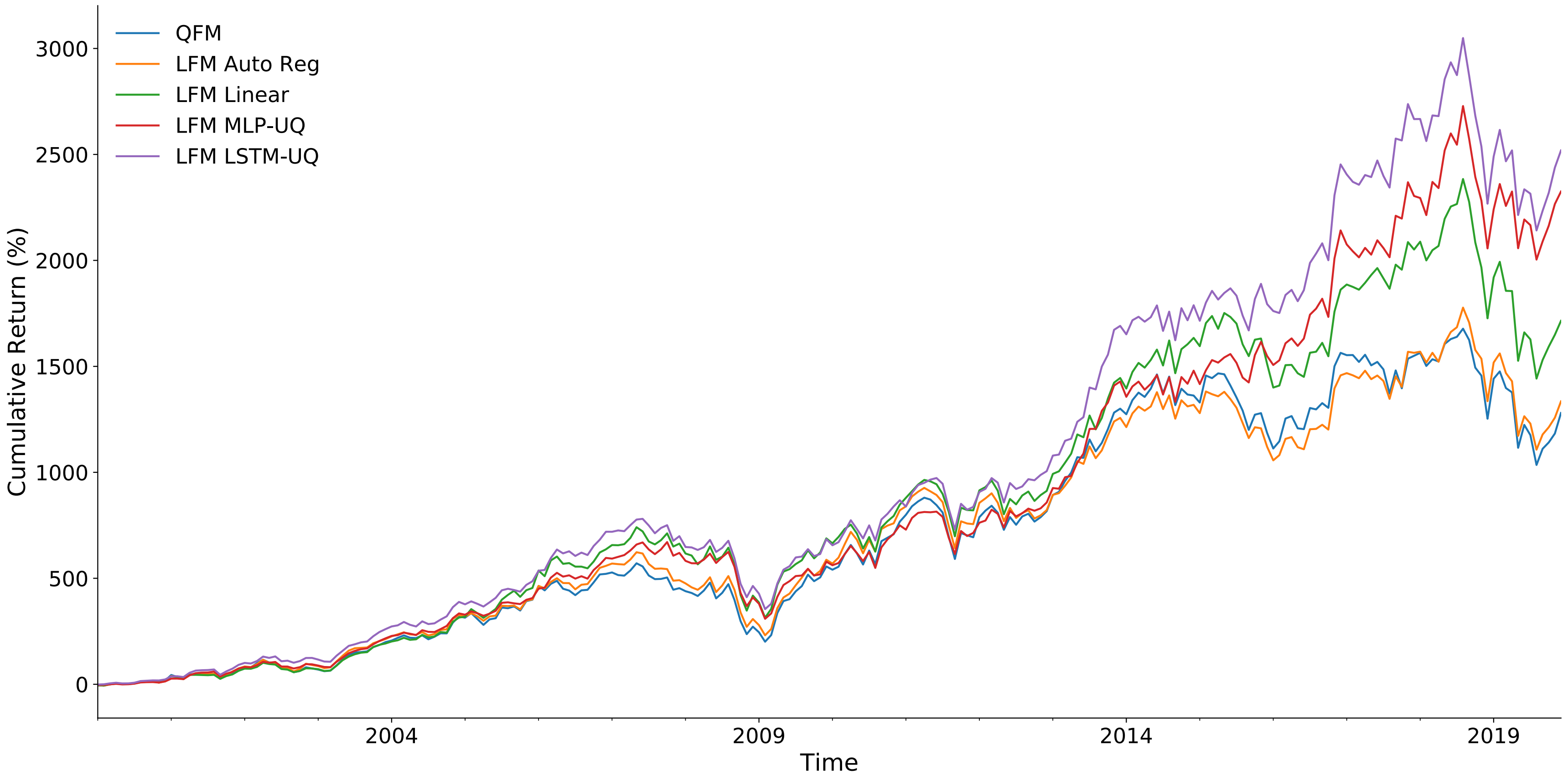
| Strategy | MSE | CAR | Sharpe Ratio |
|---|---|---|---|
| S&P 500 | n/a | 6.05% | 0.32 |
| QFM | 0.65 | 14.0% | 0.52 |
| LFM Auto Reg | 0.58 | 14.2% | 0.56 |
| LFM Linear | 0.52 | 15.5% | 0.64 |
| LFM MLP | 0.48 | 16.1% | 0.68 |
| LFM LSTM | 0.48 | 16.2% | 0.68 |
| LFM UQ-LSTM | 0.48 | **17.7%** | **0.84** |
| LFM UQ-MLP | 0.47 | **17.3%** | **0.83** |

## Pairwise t-statistic for Sharpe ratio with α=0.05

| | Auto-Reg | Linear | MLP | LSTM | UQ-LSTM | UQ-MLP |
|---|---|---|---|---|---|---|
| QFM | 0.76 | 2.52 | 2.93 | 2.96 | **5.57** | **6.01** |
| Auto Reg | | 1.89 | 2.31 | 2.36 | **5.10** | **5.57** |
| Linear | | | 0.36 | 0.46 | **3.12** | **3.66** |
| MLP | | | | 0.10 | **2.82** | **3.39** |
| LSTM | | | | | **2.66** | **3.22** |

# Results

## Cumulative return of different strategies from 2000 to 2019



# Conclusion

- Forecasting fundamentals is valuable in quantitative investing
- Use DNN to forecast future fundamentals and estimate uncertainty
- Improve return and Sharpe ratio

# Thank You

lakshay.chauhan@euclidean.com

john.alberg@euclidean.com

zlipton@cmu.edu