Introduction of Quantitative Investing with Machine Learning

Professor Haifeng YOU

2021.10.21

Plan

- Three pillars of quantitative investing
- Evaluation of alpha factors
- Framework for factor discovery
- Application of machine learning in factor discovery

An Example of Quantitative Equity Fund: GS US Equity Insights Fund

Goldman Sachs
U.S. Equity Insights Fund

A Strong Foundation

Seeks long-term growth of capital and dividend income

Fundamentally Based

We forecast expected returns on over 4,000 stocks within the U.S. on a daily basis. Stock return forecasts are based on six investment themes - Valuation, Momentum, Sentiment, Profitability, Quality and Management.

Quantitatively Constructed

Our proprietary risk model helps construct a well-diversified U.S. equity portfolio that seeks to fluctuate in price at the same rate as the market, has similar sector, style and capitalization characteristics to the S&P 500 Index, and maximizes return potential from our six fundamental themes.

Efficiently Traded

Transaction costs are considered at every step of the process, from the weighting of investment themes, to portfolio optimization, to trading. We seek to trade with maximum efficiency using integrated trading systems and sophisticated transaction cost-management techniques.

Key Element of Quantitative Investment

Quantitative Equity Investing

Alpha

Return Prediction

Alpha model (α)

Forecast and maximize future expected return

Most important element (our focus)

Risk

Risk model (σ)

Forecast and control for portfolio risk

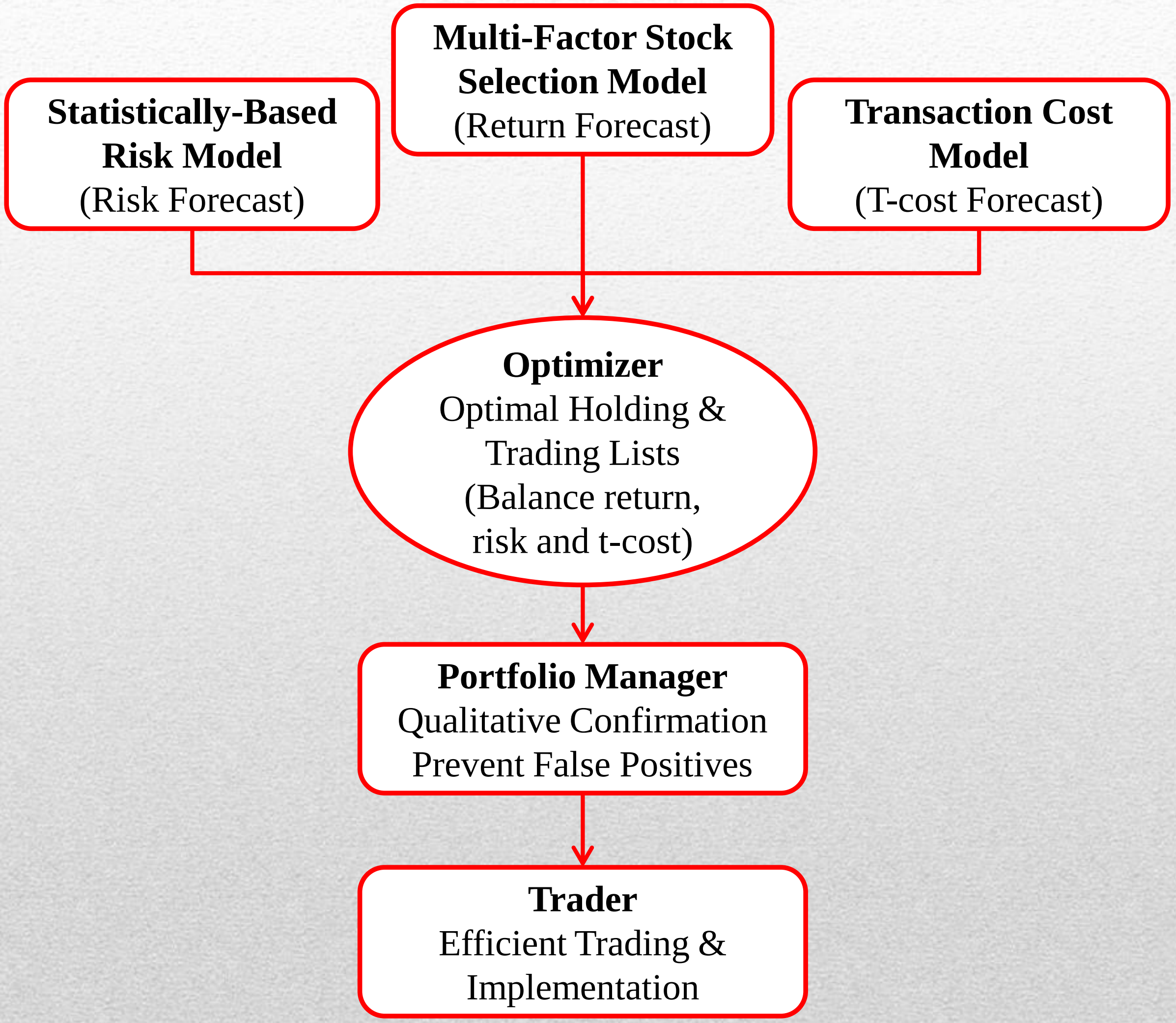
T-cost

Transaction cost

Transaction cost model (c)

Forecast and minimize transaction cost

Quantitative investment process



Putting Them Together to Build Portfolio

- In constructing optimal portfolio, portfolio managers maximize the following objective function:

$$U = \quad E(R) \quad - \; TC \quad -\lambda\sigma^2$$

Investor utility

$E(R)$ =

Expected active return of portfolio

(alpha)

Main Job

TC =

Transaction cost

σ^2 = Portfolio risk

λ = risk aversion parameter

(how much you hate volatility)

Commercially available from MSCI Barra, Axioma, Northfield, etc.

Alpha (α) Model

- Alpha is an aggregate of individual factors/signals that predict future returns (i.e. *ex ante* return forecasts)

- Factors/signals
  - Firm characteristics expected to be predictive of future returns
  - e.g. Book-to-market ratio

Factor-Based Strategies (Factor Investing)

- A *factor* is a variable or characteristic that drives, or correlates with asset returns.

- When such factors are identified, they can be used to rank stocks for investment with the aim of predicting future returns or risks.

- Typical examples are the size and value factors introduced by Fama and French (1993) in their multifactor model.

- They noticed that smaller companies tend to offer higher returns than larger companies (the *size* factor), and stocks with higher book values relative to market values tended also tended to outperform (the *value* factor).

Portfolio Approach for Factor Evaluation

1. Identify a quantifiable stock characteristics expected to predict future stock returns (e.g. B/M)

2. Sort all stocks in the investment universe by the characteristics into deciles (i.e. ten equally sized portfolios) each month/year/day

3. Calculate the time-series of portfolio returns for the 10 decile portfolios over the sample period

4. Compute the average return, volatility, turnover, and other characteristics for the 10 decile portfolios as well as the hedge portfolio that takes long (short) position in the top (bottom) decile

Portfolio Approach for Factor Evaluation

- Do the top deciles have better returns than the bottom deciles?

- Does the hedged portfolio, i.e. the top-minus-bottom portfolio generate consistent return performance over time?

- Is the relationship (between decile rank and average returns) monotonic?

Discover Return Predictive Factors: Guidance from Valuation Theory

- Value of a security should equal to the present value of future cash distributions:

$$V_t = \sum_{\tau=1}^{\infty} \frac{E_M\left[CF_{t+\tau}\right]}{(1+r_M)^{t+\tau}}$$

where

- V_t = Value of a stock at time t
- $E_M[CF_{t+\tau}]$ = Consensus market expectation at time t of cash distribution at time t+τ
- r_M = market's required rate of return at time t

What Drives Changes of Stock Prices?

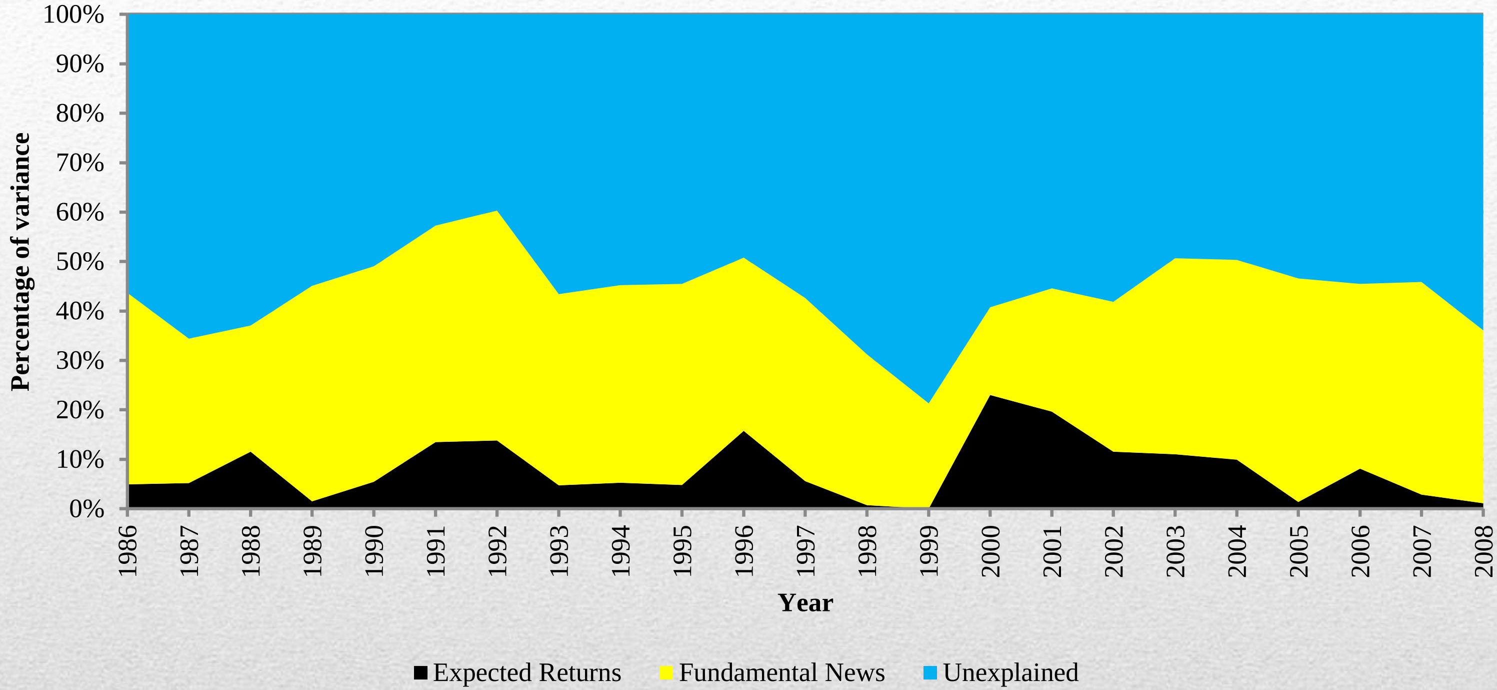
- Cum-dividend price change between period *t* and period *t+1* has 3 components:

  - r_M : The expected return that was priced into the stock at period t ("Expected Return")

  - $\Delta_{t,t+1}\, E_M[CF_{t+\tau}]$: News causing revisions to the market's cash flow expectations ("Fundamental News")

  - $\Delta_{t,t+1}\, [r_M]$: Changes in the market's required rate of return ("Expected Return News")

Relative Importance of Return Components



Source: Richardson, Sloan and You, 2012, What makes Stock Prices Move? Financial analyst Journal 68(2).

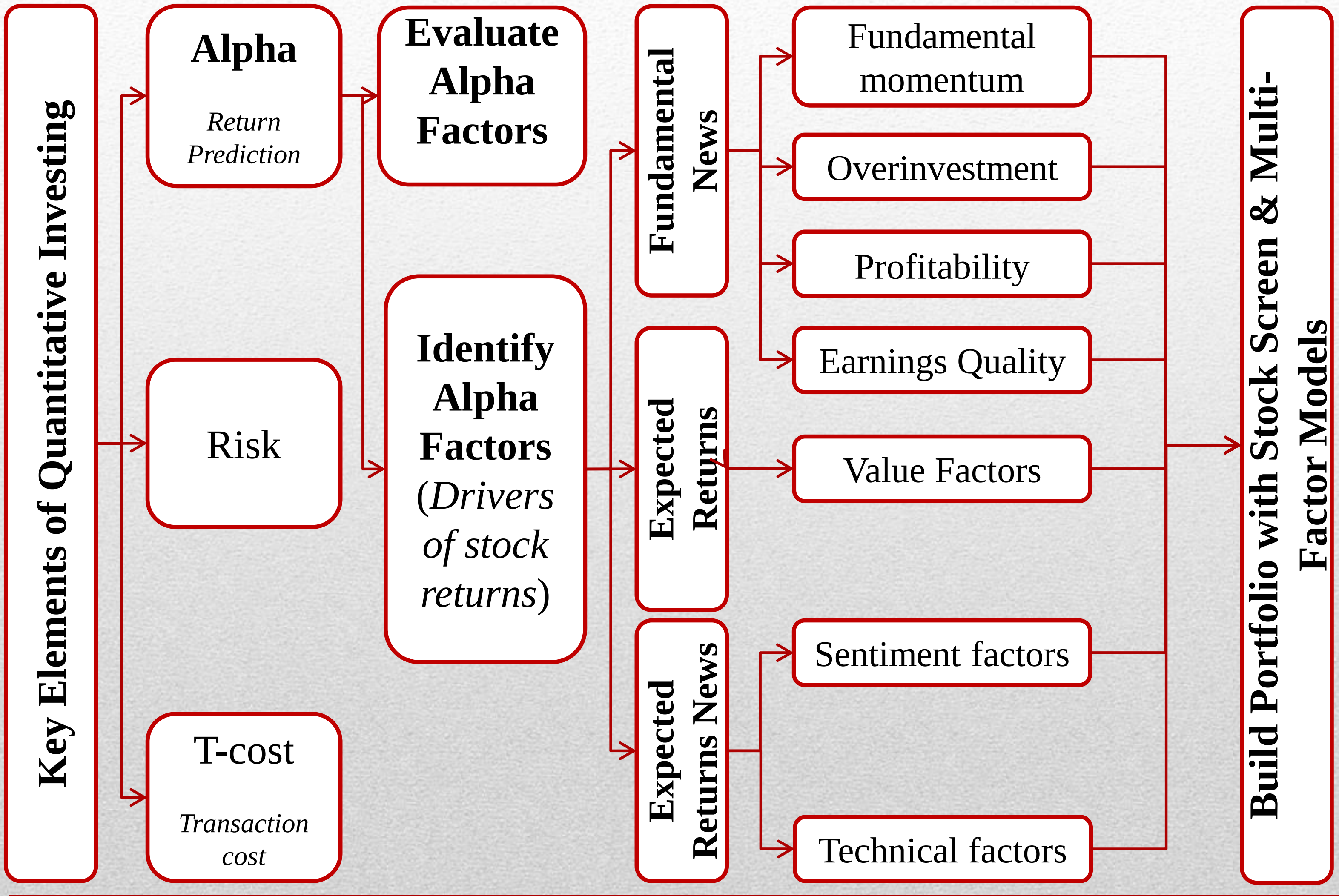
- Expected returns, together with fundamental news explains about 40% of the cross-sectional variation of annual stock returns

- Expected returns news is an important driver of the “unexplained area”

Discover Return Predictive Factors: Implication of the Valuation Theory

- Three types of potential factors:
  - Factors that capture the expected returns
  - Factors that help predict future fundamental news
  - Factors that help predict future expected return news

Multifactor Quantitative Investing



Example: JP Morgan Composite Factor

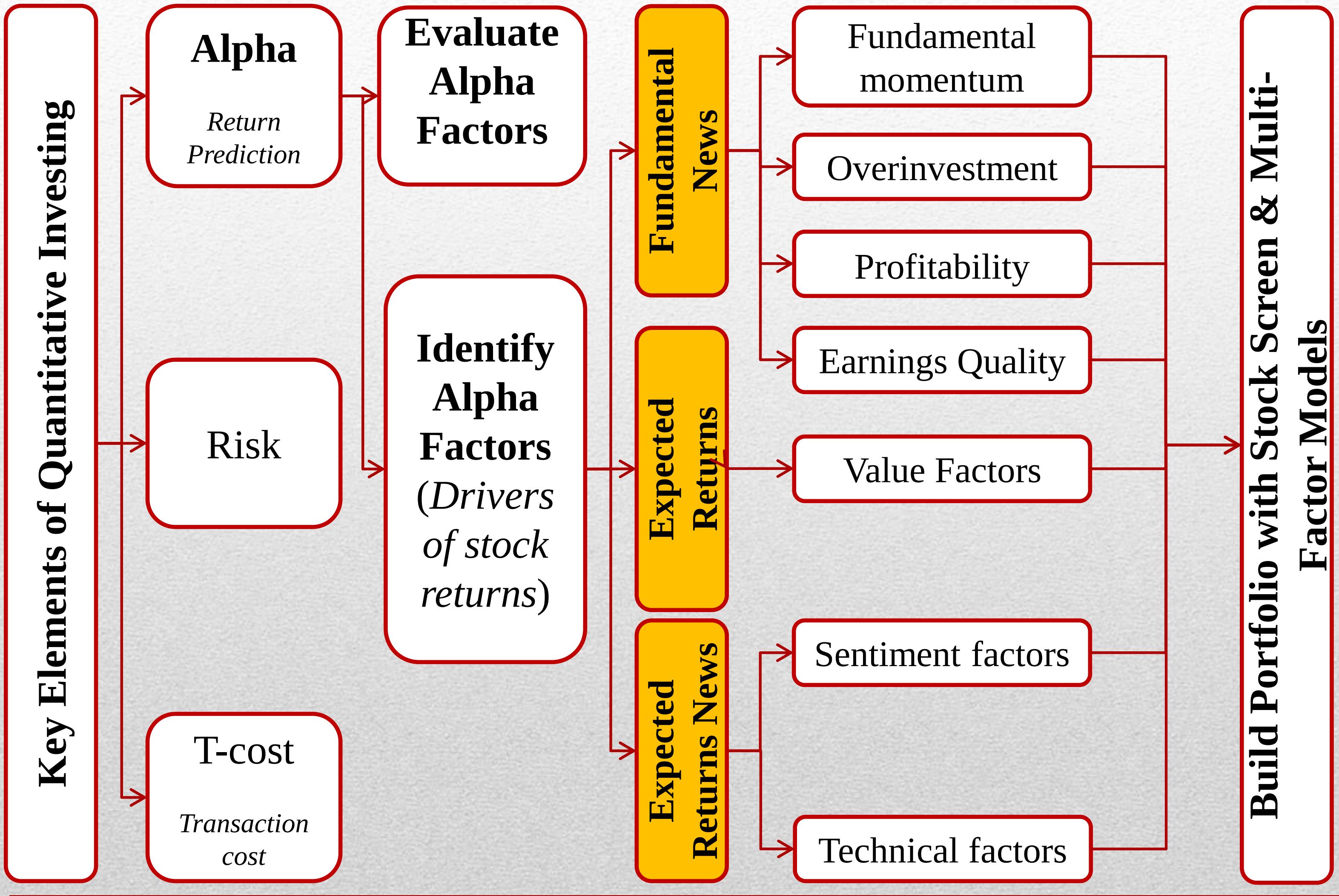
COMPOSITE FACTORS

Composite General (Growth, Value, Quality, Momentum)	Price (20%), Earnings (30%), Value (30%), and Quality (20%) Composites are combined to create a JMPQ Composite Model. The model score provides a useful quantitative benchmark signal and is used by JPMQ as a reference benchmark for factor analysis.
Composite General (Growth, Value, Quality, Momentum, Reversion)	Long-Term Price Momentum and Short-Term Price Reversal (20%), Earnings (30%), Value (30%), and Quality (20%) Composites are combined to create a JMPQ Composite Model. The model score provides a useful quantitative benchmark signal and is used by JPMQ as a reference benchmark for factor analysis.
Composite Value	Price to Earnings, Price to Sales, and Price to Cash Flow ratios are equal weighted and combined to create the JMPQ Composite Value. The model score provides a useful quantitative benchmark signal and is used by JPMQ as a reference benchmark for factor analysis.
Composite General Blend (Value, Momentum)	JPMQ Composite Momentum and JPM Composite Value are combined equally to create a Value Momentum Composite. The model score provides a useful/typical quantitative benchmark signal.
Composite General Blend (Value, Growth)	JPMQ Composite Momentum and JPMQ Composite Growth are combined equally to create a Value Growth Composite. The model score provides a useful/typical quantitative benchmark signal.
Composite Quality	JPMQ Composite Quality combines 2 flavors of Value measures. ROE and Earnings Risk are normalized and combined equally to form the Composite.
Composite Sentiment	JPMQ Composite Sentiment Change equally combines the Analyst Recommendation Level, 3-Month Change in Analyst Recommendation, and Change in 6-Month Target Price.
Composite Recommendation Change	JPMQ Composite Recommendation Change equally combines the 1 Month and 3 Month Change in Analyst Recommendations factors.
Composite Price Momentum with ST Reversal	JPMQ Composite Price combines a volatility normalized 12-Month Price Momentum factor (75%) with a 1-Month Price Reversion factor (i.e., negative of 1-Month Price Momentum factor) (25%).
Composite Price Momentum	JPMQ Composite Price equally combines a volatility normalized 12-Month Price Momentum factor with a 6-Month Price Acceleration factor to form the Composite.
Composite Earnings Momentum	JPMQ Composite Momentum combines three flavors of momentum measure. Risk-adjusted 3-Month EPS Momentum, FY2 Net Revisions, and 1-Month Change in Recommendation are all normalized and combined equally to form the Composite.

Application of ML in Quantitative Investing

- Alpha/factor discovery
- Alpha aggregation
- Portfolio optimization

Multifactor Quantitative Investing



Factor Discovery with Machine Learning

- Fundamental news
  - Cao and You (2021)
  - Binsbergen, Han, and Lopez-Lira (2021)

- Expected returns news
  - Technical analysis with ML (e.g. Murray et al. 2020; Jiang et al. 2020)
  - Sentiment analysis with NLP (e.g. Jegadeesh and Wu 2013; Huang et al. 2020; Ke et al. 2020)

- Expected returns
  - Bartram and Grinblatt (2018)
  - Geertsema and Lu (2020)

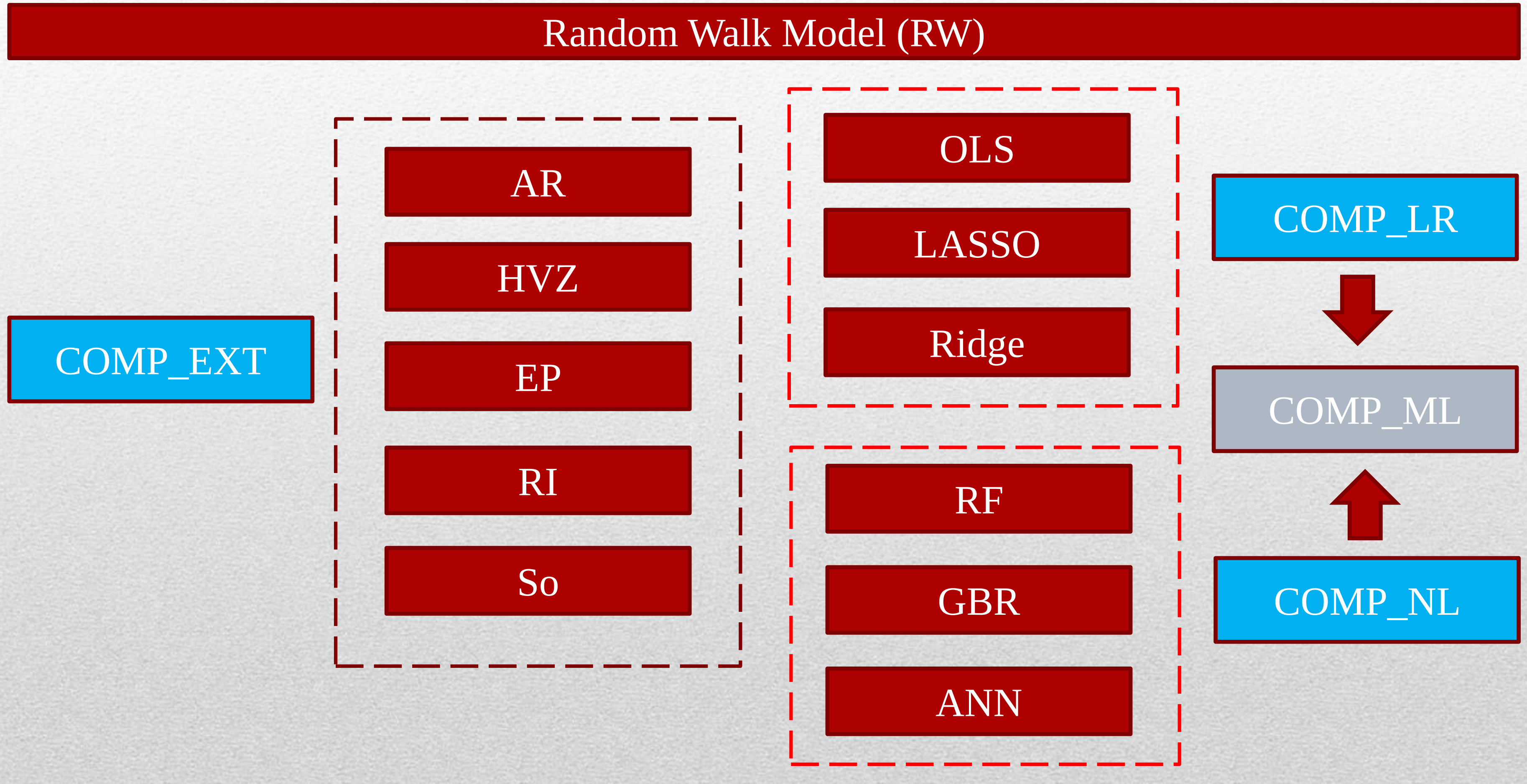
ML based Fundamental Analysis

Cao and You (2021)

- Examine whether machine learning extracts useful information from financial statements and generates better earnings forecasts
  - Accuracy, information content, proximity to the market expectation
  - Potential reasons for the difference in performance
  - Economic significance of the difference in performance

- Shed light on the usefulness of financial statement information and machine learning in fundamental analysis

Model Comparison Framework



Feature Selection

Income statement items (# = 12):	
$SALE_t$	Sales (sale)
$COGS_t$	Cost of goods sold (cogs)
$XSGA_t$	Selling, general, and administrative expenses (xsga)
XAD_t	Advertising expense (xad)
XRD_t	Research and development (R&D) expense (xrd)
DP_t	Depreciation and amortization (dp)
$XINT_t$	Interest and related expense (xint)
$NOPIO_t$	Non-operating income (expense) – other (nopio)
TXT_t	Income taxes (txt)
$XIDO_t$	Extraordinary items and discontinued operations (xido)
E_t	Earnings (ib - spi)
DVC_t	Common dividend (dvc)

Balance sheet items (# = 15):	
CHE_t	Cash and short-term investments (che)
$INVT_t$	Inventories (invt)
$RECT_t$	Receivables (rect)
ACT_t	Total current assets (act)
$PPENT_t$	Property, plant, and equipment – Net (ppent)
$IVAO_t$	Investments and advances – other (ivao)
$INTAN_t$	Intangible assets (intan)
AT_t	Total assets (at)
AP_t	Accounts payable (ap)
DLC_t	Debt in current liabilities (dlc)
TXP_t	Income taxes payable (txp)
LCT_t	Total current liabilities (lct)
$DLTT_t$	Long-term debt (dltt)
LT_t	Total liabilities (lt)
CEQ_t	Common/Ordinary equity (ceq)

Cash flow statement items (# = 1):	
CFO_t	Cash flow from operating activities (oancf - xidoc); if missing, it is computed using the balance sheet approach (ib - accruals)

First-order differences of the above 28 items (# = 28):	
$\Delta CHE_t \sim \Delta CFO_t$	Computed as the corresponding item in year t less the same item in year t - 1

Table 2: Comparison of forecast accuracy

	Mean absolute forecast errors				Median absolute forecast errors			
	Average	Comparison with RW			Average	Comparison with RW		
		DIFF	t-stat	%DIFF		DIFF	t-stat	%DIFF
Benchmark model								
RW	0.0764				0.0309			
Extant models								
AR	0.0755	-0.0009	-2.51	-1.15%	0.0308	-0.0001	-0.22	-0.24%
HVZ	0.0743	-0.0022	-3.63	-2.82%	0.0311	0.0002	0.64	0.76%
EP	0.0742	-0.0022	-2.79	-2.85%	0.0313	0.0004	1.02	1.42%
RI	0.0741	-0.0023	-3.15	-3.07%	0.0311	0.0002	0.66	0.74%
SO	0.0870	0.0105	5.19	13.78%	0.0347	0.0039	5.50	12.56%
Linear machine learning models								
OLS	0.0720	-0.0045	-5.04	-5.83%	0.0306	-0.0002	-0.60	-0.73%
LASSO	0.0716	-0.0048	-5.32	-6.31%	0.0304	-0.0004	-1.11	-1.43%
Ridge	0.0718	-0.0047	-5.19	-6.11%	0.0305	-0.0003	-0.87	-1.08%
Nonlinear machine learning models								
RF	0.0698	-0.0066	-6.44	-8.64%	0.0296	-0.0012	-3.10	-3.97%
GBR	0.0697	-0.0068	-6.08	-8.86%	0.0292	-0.0016	-4.23	-5.34%
ANN	0.0713	-0.0051	-5.38	-6.67%	0.0310	0.0001	0.24	0.38%
Composite models								
COMP_EXT	0.0737	-0.0027	-3.89	-3.58%	0.0311	0.0002	0.56	0.66%
COMP_LR	0.0717	-0.0047	-5.25	-6.16%	0.0305	-0.0004	-1.02	-1.33%
COMP_NL	0.0689	-0.0075	-6.99	-9.87%	0.0292	-0.0017	-3.92	-5.55%
COMP_ML	0.0693	-0.0071	-7.12	-9.35%	0.0294	-0.0015	-3.75	-4.81%

Information Content Analysis

The ability of forecasted earnings change (FECH) to predict actual earnings change (ECH):

- Pearson and Spearman correlations
- Univariate regressions of *ECH* on *FECH*.
- Multivariate regressions:

$$ECH = \beta_0 + \boldsymbol{\beta_1 FECH_{ML}} + \beta_2 FECH_{AR} + \beta_3 FECH_{HVZ} + \beta_4 FECH_{EP} + \beta_5 FECH_{RI} + \beta_6 FECH_{SO} + \varepsilon$$

Table 4, Panel B

Panel B: Incremental information content of the machine learning models

	Multivariate regression: $ECH = \beta_0 + \beta_1 FECH_{ML} + \beta_2 FECH_{AR} + \beta_3 FECH_{HVZ} + \beta_4 FECH_{EP} + \beta_5 FECH_{RI} + \beta_6 FECH_{SO} + \varepsilon$							
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	Avg. R^2 (%)
Linear machine learning models								
OLS	0.0016	0.0432	0.0107	-0.0058	0.0004	-0.0098	0.0251	18.99
	(0.57)	(11.90)	(1.56)	(-1.42)	(0.03)	(-0.82)	(8.82)	
LASSO	0.0016	0.0458	0.0085	-0.0072	0.0017	-0.0111	0.0251	19.09
	(0.57)	(15.45)	(1.28)	(-1.72)	(0.13)	(-0.87)	(8.72)	
Ridge	0.0016	0.0453	0.009	-0.0068	0.0019	-0.0113	0.0251	19.09
	(0.57)	(12.19)	(1.36)	(-1.66)	(0.14)	(-0.89)	(8.71)	
Nonlinear machine learning models								
RF	0.0016	0.049	0.0105	-0.0072	-0.0043	-0.0014	0.0146	19.53
	(0.57)	(16.83)	(1.60)	(-1.71)	(-0.30)	(-0.12)	(3.89)	
GBR	0.0016	0.0497	0.0086	-0.0079	-0.0005	-0.006	0.0183	19.63
	(0.57)	(16.40)	(1.42)	(-1.91)	(-0.03)	(-0.54)	(5.54)	
ANN	0.0016	0.0466	0.0078	-0.0047	0.0111	-0.0137	0.0176	20.20
	(0.57)	(16.24)	(1.29)	(-1.17)	(0.78)	(-1.17)	(5.15)	
Composite models								
COMP_LR	0.0016	0.045	0.0094	-0.0068	0.0016	-0.011	0.025	19.08
	(0.57)	(12.27)	(1.41)	(-1.64)	(0.12)	(-0.88)	(8.82)	
COMP_NL	0.0016	0.059	0.0075	-0.0087	0.0053	-0.0144	0.0132	20.84
	(0.57)	(17.91)	(1.30)	(-2.11)	(0.36)	(-1.25)	(3.92)	
COMP_ML	0.0016	0.0593	0.0071	-0.0104	0.0081	-0.0199	0.0175	20.80
	(0.57)	(16.22)	(1.20)	(-2.44)	(0.63)	(-1.71)	(6.24)	

Usefulness of Machine learning Forecasts for Return Prediction

- New information uncovered by the machine learning models
  - Residuals from the cross-sectional regression of the machine-learning-based forecasts against the forecasts generated using the RW model and the extant models.
- Fama-MacBeth regression analysis

$$EXRET12M_{i,t+1} = \beta_0 + \beta_1 ML\_RESD_{i,t} + \beta_2 SIZE_{i,t} + \beta_3 BM_{i,t} + \beta_4 MOM_{i,t} + \beta_5 ROE_{i,t} + \beta_6 INV_{i,t} + \beta_7 ACC_{i,t} + IndustryFE + \varepsilon_{i,t+1}$$

- Portfolio analysis
  - Equal-weighted portfolios and value-weighted portfolios

Table 8: Regression of analyst forecast errors on the new information uncovered using the machine learning models

Multivariate regression: $FERR_{i,t+1} = \beta_0 + \beta_1 ML\_RESD_{i,t} + \beta_2 SIZE_{i,t} + \beta_3 BM_{i,t} + \beta_4 MOM_{i,t} + \beta_5 ACC_{i,t} + \beta_6 LTG_{i,t} + \varepsilon_{i,t+1}$

	β_0	$\boldsymbol{\beta_1}$	β_2	β_3	β_4	β_5	β_6	Avg. R^2(%)
			Linear machine learning models					
OLS	-0.053	**0.242**	0.008	-0.028	0.030	0.014	-0.031	13.20
	(-6.03)	**(4.00)**	(5.68)	(-4.91)	(3.76)	(1.47)	(-2.98)	
LASSO	-0.053	**0.272**	0.008	-0.028	0.029	0.014	-0.032	13.25
	(-6.03)	**(3.84)**	(5.68)	(-4.89)	(3.77)	(1.46)	(-2.94)	
Ridge	-0.053	**0.258**	0.008	-0.028	0.030	0.014	-0.032	13.19
	(-6.04)	**(4.16)**	(5.68)	(-4.90)	(3.76)	(1.47)	(-2.95)	
			Nonlinear machine learning models					
RF	-0.053	**0.248**	0.008	-0.028	0.029	0.018	-0.030	12.93
	(-5.96)	**(3.34)**	(5.69)	(-4.86)	(3.84)	(1.79)	(-3.03)	
GBR	-0.053	**0.184**	0.008	-0.028	0.030	0.017	-0.030	12.91
	(-5.97)	**(3.11)**	(5.67)	(-4.83)	(3.86)	(1.72)	(-3.12)	
ANN	-0.053	**0.204**	0.008	-0.028	0.030	0.017	-0.029	13.12
	(-6.03)	**(3.44)**	(5.71)	(-4.84)	(3.83)	(1.76)	(-2.87)	
			Composite models					
COMP_LR	-0.053	**0.259**	0.008	-0.028	0.030	0.014	-0.031	13.21
	(-6.03)	**(4.01)**	(5.68)	(-4.90)	(3.77)	(1.47)	(-2.95)	
COMP_NL	-0.053	**0.251**	0.008	-0.028	0.029	0.018	-0.029	13.04
	(-5.99)	**(3.27)**	(5.68)	(-4.85)	(3.86)	(1.81)	(-2.97)	
COMP_ML	-0.053	**0.282**	0.008	-0.028	0.029	0.017	-0.030	13.15
	(-6.01)	**(3.55)**	(5.68)	(-4.88)	(3.82)	(1.71)	(-2.92)	

Table 7: Portfolio analysis of the new information uncovered using the machine learning models

Panel A: Equal-weighted portfolios

	OLS	LASSO	Ridge	RF	GBR	ANN	COMP_LR	COMP_NL	COMP_ML
Mean Return	0.6185	0.6262	0.6346	0.5962	0.6795	0.7185	0.6402	0.7203	0.7720
	(8.65)	(8.89)	(8.85)	(7.49)	(8.73)	(8.12)	(9.29)	(8.05)	(9.50)
CAPM Alpha	0.6817	0.6856	0.6989	0.6328	0.7110	0.7784	0.7022	0.7695	0.8372
	(9.96)	(10.46)	(10.48)	(7.82)	(9.07)	(8.89)	(10.87)	(8.78)	(10.73)
FF3 Alpha	0.6538	0.6597	0.6758	0.6062	0.6733	0.7247	0.6761	0.7279	0.8033
	(9.71)	(9.88)	(10.18)	(8.54)	(9.90)	(9.63)	(10.46)	(9.61)	(11.39)
Carhart4 Alpha	0.5938	0.5921	0.6178	0.5166	0.5934	0.6558	0.6137	0.6448	0.7134
	(9.08)	(9.03)	(9.49)	(7.29)	(8.57)	(8.50)	(9.66)	(8.35)	(10.23)
FF5 Alpha	0.5371	0.5488	0.5655	0.4312	0.4828	0.5286	0.5613	0.5143	0.6096
	(7.96)	(8.21)	(8.48)	(5.97)	(7.08)	(7.18)	(8.64)	(6.63)	(8.59)

Panel B: Value-weighted portfolios

	OLS	LASSO	Ridge	RF	GBR	ANN	COMP_LR	COMP_NL	COMP_ML
Mean Return	0.2239	0.2484	0.2674	0.3177	0.4163	0.4747	0.2677	0.4568	0.3831
	(1.99)	(2.19)	(2.27)	(2.74)	(3.50)	(4.08)	(2.29)	(3.74)	(3.60)
CAPM Alpha	0.3571	0.3778	0.3969	0.3775	0.4797	0.5914	0.3954	0.5490	0.4884
	(3.30)	(3.57)	(3.53)	(3.05)	(4.01)	(5.07)	(3.58)	(4.34)	(4.66)
FF3 Alpha	0.3237	0.3552	0.3667	0.4478	0.5505	0.6325	0.3663	0.6217	0.5289
	(3.34)	(3.53)	(3.54)	(3.75)	(4.60)	(5.52)	(3.65)	(5.19)	(5.15)
Carhart4 Alpha	0.2829	0.2999	0.3320	0.3081	0.4316	0.5605	0.3247	0.4768	0.4558
	(3.08)	(3.06)	(3.41)	(3.07)	(3.70)	(4.70)	(3.37)	(4.49)	(4.23)
FF5 Alpha	0.1222	0.1205	0.1634	0.2810	0.4142	0.4358	0.1575	0.4119	0.3715
	(1.42)	(1.40)	(1.90)	(2.57)	(3.80)	(4.40)	(1.85)	(3.54)	(3.89)

Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases

- By Jules H. van Binsbergen, Xiao Han, and Alejandro Lopez-Lira
  - Forecast earnings using information in financial statements, macroeconomic variables, and analysts' predictions
  - Machine learning model: Random forest regression
  - Assess bias in analyst forecasts by comparing analysts' forecasts to machine learning forecasts
  - Biases increase with forecast horizon
  - High bias predicts lower future stock returns

Variables Used for Earnings Forecasts-Firm Fundamentals:

1. Realized earnings from the last period. Earnings data are obtained from /I/B/E/S
2. Earnings growth, defined as the growth rate in earnings
3. Sales growth, defined as the growth rate in sales and obtained from COMPUSTAT
4. Asset growth, defined as the growth rate in total assets and obtained from COMPUSTAT
5. Investment growth, defined as the growth rate in capital expenditure and obtained from COMPUSTAT
6. Monthly stock prices and returns from CRSP
7. Sixty-seven financial ratios such as book-to-market ratio and dividend yields obtained from the Financial Ratios Suit by Wharton Research Data Services.

Variables Used for Earnings Forecasts- Macroeconomic Variables & Analyst Forecasts:

Macroeconomic variables from the Federal Reserve Bank of Philadelphia:

1. Consumption growth, defined as the log difference of consumption in goods and services
2. GDP growth, defined as the log difference of real GDP
3. Growth of industrial production, defined as the log difference of Industrial Production Index (IPT)
4. Unemployment rate

Analysts' one-year ahead EPS forecasts

Random Forest Regression

- Random forest regression model to forecast earnings from January 1987:

$$E_t[eps_{i,t+\tau}] = RF[Fundamentals_{i,t}, Macro_t, AF_{i,t}].$$

- Hyperparameters are determined using data of 1986

Number of Trees	2000
Maximum Depth	7
Sample Fraction	1%
Minimum Node Size	5

- Model retrained each month from January 1987
- Define bias as:

$$BiasedExpectation_{i,t}^{t+h} = \frac{Analysts'Forecasts_{i,t}^{t+h} - MLForecast_{i,t}^{t+h}}{Price_{i,t-1}}$$

Table 4: Fama-MacBeth Regression of Future Returns on BE and Control Variables

$$R_{i,t+1} = \alpha + \beta_1 BE_{i,t} + \gamma_i \sum_{i=1}^{8} Control_{i,t} + \epsilon_{i,t+1}$$

	Panel A: Average BE		Panel B: BE Score	
	(1)	(2)	(1)	(2)
BE	−0.0808	−0.0852	−0.0279	−0.0456
	(−4.61)	(−5.30)	(−6.57)	(−15.99)
Lnsize		−0.0009		−0.0029
		(−2.46)		(−8.37)
Lnbeme		0.0012		0.0019
		(2.00)		(3.16)
Ret12_7		0.0038		0.0011
		(2.44)		(0.73)
Ret1		−0.0284		−0.0313
		(−6.62)		(−7.29)
IA		−0.0007		−0.0007
		(−2.60)		(−2.73)
Ivol		−0.1941		−0.1743
		(−1.72)		(−1.53)
Retvol		0.1339		0.1982
		(1.13)		(1.67)
Turnover		−0.0006		−0.0005
		(−1.38)		(−1.18)
Intercept	0.0078	0.0213	0.0215	0.0675
	(2.74)	(3.98)	(8.70)	(13.69)
R^2	0.0105	0.0604	0.0156	0.0629

BE Score: the arithmetic average of the percentile rankings of the five conditional biases

Portfolio Analysis of BE Score

Table 6: Portfolios sorted on conditional bias

Notes: This table reports the time series average of excess returns (in percent) on value-weighted portfolios formed on the conditional bias in different forecast horizons. Panel A looks at "Average BE", defined as the average of conditional bias at different forecast horizons. Panel B presents the sorts based on "BE score", defined as the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. The sample period is 1987 to 2019.

Quintile	1	2	3	4	5	1-5
Panel A: Average BE						
Mean	1.07	0.70	0.46	-0.04	-0.88	1.95
t-stat	5.03	3.17	1.82	-0.12	-2.05	5.88
CAPM Beta	0.92	0.98	1.11	1.28	1.58	-0.66
Panel B: BE Score						
Mean	0.96	0.66	0.43	0.07	-0.57	1.53
t-stat	4.76	2.93	1.64	0.22	-1.38	4.90
CAPM Beta	0.89	1.01	1.14	1.28	1.53	-0.63

Time-Series Tests with Multi-Factor Models

Table 7: Time series tests with common asset-pricing Models

Notes: This table reports the regression of stock returns (in percent) on the long-short portfolio sorted with the conditional bias, on the CAPM, the Fama-French three-factor model (FF3), and the Fama-French five-factor model (FF5). Panel A looks at average conditional bias at different forecast horizons. Panel B presents the sorts based on "BE score", defined as the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. The sample period is 1987 to 2019. The t-statistics are adjusted by the White's heteroscedasticity robust standard errors.

$$LS\_Port_t = \alpha + \sum_{i=1}^{5} \beta_i F_{i,t} + \epsilon_t$$

	CAPM		FF3		FF5	
	Coeffi	t-stat	*Coeffi*	t-stat	*Coeffi*	t-stat
Panel A: Average BE						
Intercept	2.39	8.15	2.52	9.70	2.02	7.21
Mkt_RF	-0.66	-7.81	-0.61	-7.52	-0.42	-5.34
SMB			-0.86	-6.33	-0.62	-4.33
HML			-0.60	-4.10	-1.01	-6.10
RMW					0.84	4.07
CMA					0.53	1.79
Panel B: BE Score						
Intercept	1.94	7.02	2.03	8.01	1.53	5.73
Mkt_RF	-0.63	-7.50	-0.56	-6.58	-0.37	-4.62
SMB			-0.83	-6.89	-0.57	-4.39
HML			-0.44	-3.07	-0.83	-4.93
RMW					0.90	4.63
CMA					0.48	1.63

Other Approaches

- Using machine learning based earnings forecasts as valuation inputs (Binz, Schipper and Standridge 2021)
- Relative Valuation with Machine Learning (based on fundamentals (Geertsema and Lu 2020)
- Predicting accounting fraud/misstatement (earnings quality) using machine learning (Bao et al. 2019; Bertomeu et al. 2020)

ML based Technical Analysis

Research on Technical Analysis

- Stock Selection:
  - Short-term reversal (e.g. Fama 1965; Jegadeesh 1990)
  - Medium-term momentum (e.g. Jegadeesh and Titman 1993)
  - Long-term reversal (De Bondt and Thaler 1985)
  - Trend factor with moving averages (Han, Zhou, Zhou 2016)
- Market timing:
  - Moving average, momentum and volume based indicators (Neely, Rapach, Tu and Zhou 2014)
- …
- Machine learning based technical analysis
  - Charting By Machines, by Murray, Xiao and Xia (2020)
  - (Re-)Imag(in)ing Price Trends, by Jiang, Kelly and Xiu (2020)

Murray et al. (2020)

Forecast future stock returns from historical price plots using machine learning

- Features: Cumulative returns of individual stocks over the month t-12 through t-1:
  - CR1: t-12
  - CR2: t-12 to t-11
  - …
  - CR12: t-12 to t-1
- Machine Learning models:
  - Feed-forward neural network (FNN)
  - Convolutional neural network (CNN)
  - Long-short term memory (LSTM)
  - Convolutional neural network with long-short term memory (CNNLSTM)

Murray et al. (2020)
Research Design Issue

- What to forecast?
  - r: excess stock return
  - r_{Std}: standardized excess return (z-score transformation)
  - r_{Norm}: normalized excess return (change to normal dist.)
  - r_{Pctl}: percentiles of a stock's return in a month
- Loss function
  - MSE:$\mathcal{L} = \sum w_j\, \varepsilon_j^2$
  - MAE:$\mathcal{L} = \sum w_j\, |\varepsilon_j|$
- Loss function weighting (w_j) schedule
  - EW: equal weighting
  - EWPM: equal weighting per month
  - EWPMVW: equal weighting per month, but weight each stock in a month based on its market cap
- In total, 96 models: 4 (ML)*4(Target)*2(Loss)*3(Weights)

Murray et al. (2020) Research Design Issue

- Optimization period: 192701-196306
- Training Sample: Even months from even years & odd months from odd years
- Validation sample: Other months in the optimization period
- Testing period: 196307-201912
- Model evaluation/selection: time-series average of the monthly cross-sectional Spearman rank correlation (i.e. Spearman IC)

Table 1: ML Process Optimization

Dependent Variable	Weighting Methodology	FNN MAE	FNN MSE	CNN MAE	CNN MSE	LSTM MAE	LSTM MSE	CNNLSTM MAE	CNNLSTM MSE
r	EW	4.6	1.6	4.0	1.0	5.5	−1.8	5.3	0.8
	EWPM	4.6	1.4	3.8	1.2	6.7	0.8	5.0	2.6
	EWPMVW	0.9	−0.5	0.5	−2.3	2.5	0.9	1.0	2.0
r_{Std}	EW	3.4	2.3	9.7	7.3	9.8	8.5	10.4	9.8
	EWPM	5.0	1.7	8.9	7.4	9.8	7.6	10.5	9.9
	EWPMVW	4.7	2.0	4.7	4.6	4.7	5.1	6.0	4.7
r_{Norm}	EW	7.5	1.5	9.7	8.8	10.3	10.2	10.4	10.7
	EWPM	6.9	1.4	9.1	8.0	10.1	10.2	10.6	10.8
	EWPMVW	3.7	0.4	4.3	4.3	5.2	6.3	6.9	7.5
r_{Pctl}	EW	−2.3	−1.6	7.7	−2.7	9.8	9.4	10.2	10.1
	EWPM	0.7	−3.0	7.8	−3.5	9.6	9.3	10.2	10.2
	EWPMVW	−0.1	−3.1	1.2	−1.7	4.9	5.7	7.3	7.8

Generate machine learning based forecast, *MLER*, using (CNNLSTM, MSE, EWPM, and r_{Norm})

$$MLER_{i,t} = \begin{cases} MLER_{i,t}^{192701,196306}, & \text{if } 196307 \le t \le 197412; \\ MLER_{i,t}^{192701,197412}, & \text{if } 197501 \le t \le 198412; \\ MLER_{i,t}^{192701,198412}, & \text{if } 198501 \le t \le 199412; \\ MLER_{i,t}^{192701,199412}, & \text{if } 199501 \le t \le 200412; \\ MLER_{i,t}^{192701,200412}, & \text{if } 200501 \le t \le 201412; \\ MLER_{i,t}^{192701,201412}, & \text{if } 201501 \le t \le 201912. \end{cases}$$

Table 3: Portfolio Analysis

Value	$MLER\,1$	$MLER\,2$	$MLER\,3$	$MLER\,4$	$MLER\,5$	$MLER\,6$	$MLER\,7$	$MLER\,8$	$MLER\,9$	$MLER\,10$	$MLER\,10-1$
Excess Return	−0.13	0.31	0.40	0.51	0.51	0.65	0.68	0.66	0.75	0.93	1.06
	(−0.50)	(1.35)	(1.98)	(2.62)	(2.80)	(4.03)	(3.99)	(3.91)	(4.19)	(5.00)	(5.38)
α^{CAPM}	−0.80	−0.32	−0.17	−0.06	−0.02	0.13	0.17	0.16	0.25	0.38	1.18
	(−6.20)	(−3.48)	(−2.27)	(−0.97)	(−0.42)	(2.19)	(3.43)	(2.43)	(3.68)	(3.86)	(6.17)
α^{FF}	−0.86	−0.37	−0.21	−0.09	−0.05	0.10	0.15	0.12	0.25	0.36	1.22
	(−7.80)	(−4.19)	(−2.83)	(−1.63)	(−0.85)	(1.80)	(3.02)	(2.47)	(4.06)	(3.80)	(7.00)
α^{FFC}	−0.58	−0.20	−0.04	0.07	0.05	0.13	0.12	0.05	0.12	0.19	0.78
	(−6.21)	(−2.38)	(−0.44)	(1.29)	(0.84)	(2.09)	(2.26)	(0.98)	(1.89)	(2.11)	(5.03)
α^{FFCLIQ}	−0.58	−0.20	−0.05	0.08	0.04	0.14	0.11	0.06	0.12	0.21	0.78
	(−5.89)	(−2.28)	(−0.59)	(1.39)	(0.67)	(2.19)	(2.01)	(1.10)	(1.88)	(2.16)	(4.84)
α^{FF5}	−0.68	−0.25	−0.17	−0.06	−0.06	0.09	0.10	0.03	0.18	0.31	0.99
	(−6.27)	(−2.77)	(−2.07)	(−0.89)	(−0.94)	(1.67)	(2.01)	(0.63)	(2.65)	(3.56)	(6.27)
α^{Q}	−0.50	−0.15	−0.06	0.05	0.03	0.13	0.08	−0.04	0.12	0.21	0.71
	(−4.27)	(−1.70)	(−0.61)	(0.59)	(0.38)	(2.03)	(1.32)	(−0.70)	(1.57)	(1.89)	(3.66)
S.D.	6.31	5.52	5.04	4.93	4.53	4.45	4.36	4.28	4.38	4.98	
Skewness	−0.10	−0.23	−0.34	−0.21	−0.41	−0.49	−0.58	−0.49	−0.03	−0.07	
P_1	−17.95	−13.69	−13.04	−12.22	−10.87	−10.73	−12.08	−10.24	−10.63	−13.40	
P_5	−10.72	−8.89	−7.96	−7.65	−7.11	−6.56	−6.50	−6.37	−6.54	−6.96	
ES_1	−20.35	−18.22	−17.52	−15.86	−15.14	−15.40	−15.11	−14.77	−14.29	−17.13	
ES_5	−14.90	−12.45	−11.19	−10.88	−10.09	−10.02	−9.83	−9.78	−9.41	−11.21	

Table 9:
Non-Linearity of ML-based Forecasts

Panel A: FM Regresssions of $MLER$

CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}	Adj. R^2
0.030	0.023	0.006	0.009	0.004	0.002	0.011	0.033	−0.067	0.135	0.255	−0.283	37.46%
(11.61)	(12.90)	(2.74)	(6.14)	(2.35)	(1.11)	(3.06)	(9.07)	(−11.39)	(17.57)	(17.09)	(−15.68)	

Panel B: FM Regresssions of Future Excess Returns - Equal-Weighted

$MLER$	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}
6.80												
(10.05)												
6.07	0.01	0.00	−0.00	0.00	−0.00	−0.00	−0.00	−0.00	0.00	−0.00	0.02	−0.01
(9.91)	(1.98)	(1.32)	(−1.76)	(1.07)	(−1.12)	(−0.25)	(−0.48)	(−0.68)	(0.38)	(−0.70)	(4.37)	(−2.60)

Panel C: FM Regresssions of Future Excess Returns - Value-Weighted

$MLER$	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}
3.24												
(5.37)												
2.68	0.00	0.01	−0.01	0.01	0.00	−0.01	0.01	0.00	−0.00	0.00	0.02	−0.01
(5.15)	(0.40)	(1.52)	(−3.56)	(3.93)	(0.09)	(−3.00)	(2.37)	(0.54)	(−1.38)	(0.41)	(4.16)	(−2.70)

Table 13: Fama and MacBeth Regressions with Momentum and Reversal

	EW	EW	EW	EW	VW	VW	VW	VW
$MLER$	6.804	6.603	4.536	3.883	3.237	2.731	3.299	2.147
	(13.78)	(14.14)	(7.22)	(6.87)	(5.47)	(4.92)	(4.49)	(3.17)
Mom		0.002		0.004		0.005		0.006
		(0.98)		(2.42)		(2.68)		(3.37)
Rev			−0.028	−0.033			−0.001	−0.010
			(−5.23)	(−6.61)			(−0.11)	(−1.58)

From Features to Pictures

- Feature engineer (e.g. breaking down stock returns into predefined intervals)
  - Helps reduce the dimension of predictors and overfitting
  - It may also lead to loss of information
- Can machine learning extract more flexible/subtle patterns from historical price/volume that are useful for return prediction?
- (Re-)Imag(in)ing Price Trends, by Jiang, Kelly and Xiu (2020)

Jiang, Kelly and Xiu (2020)
Research Design: Inputs

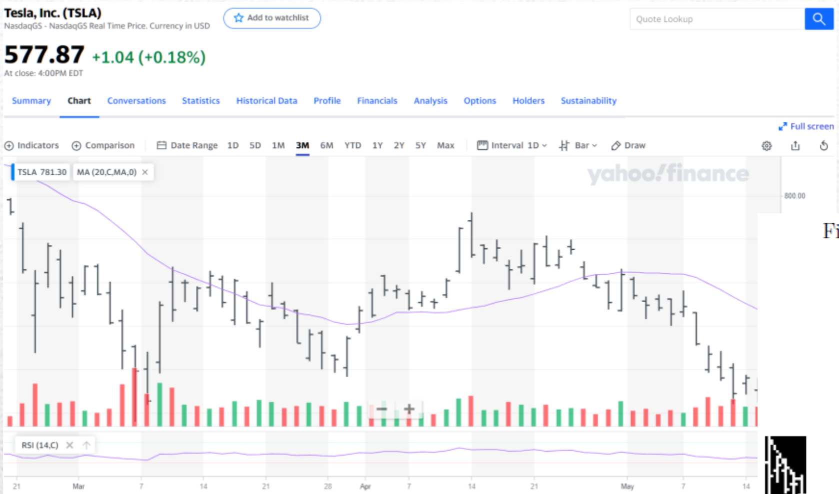
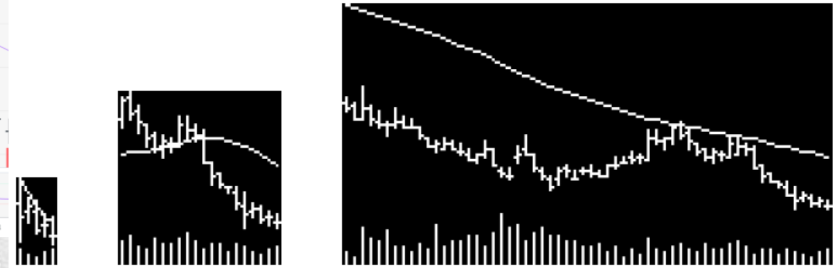


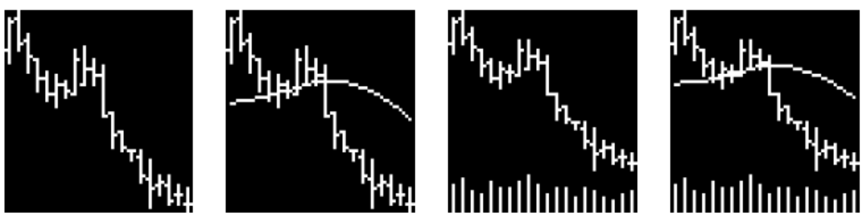
Figure 4: Generated OHLC Images with Volume Bar and Moving Average Line



(a) 5 (b) 20 (c) 60

Note: Market data images for 5, 20, and 60 days of data.

Figure 3: Examples of 20-day Image under Different Settings



(a) w/o VB, w/o MA (b) w/o VB, w/ MA (c) w/ VB, w/o MA (d) w/ VB, w/ MA

Note: From left to right are 20-day images (a) without volume bar and moving average line, (b) without volume bar but with moving average line, (c) with volume bar but without moving average line, and (d) with volume bar and moving average line.

Jiang et al. (2020)
Research Design

- Sample: NYSE, AMEX, and NASDAQ
- Sample period: 1993-2019
- Training & Validation:
  - 1993 to 1999
  - 70% training & 30% for validation (randomly)
- Test sample: 2000-2019
- Target variable: y=1 if subsequent return is positive and y=0 otherwise

Out-of-Sample Classification Accuracy

Table 2: Out-of-Sample Classification Accuracy

	Return horizon			
	20-day		60-day	
Image size	Acc.	Corr.	Acc.	Corr.
5-day	52.1%	3.2%	52.5%	2.0%
20-day	52.5%	3.2%	52.9%	2.6%
60-day	52.5%	3.1%	53.5%	3.1%
MOM	52.2%	1.9%	52.2%	1.7%
STR	50.4%	1.4%	49.7%	1.2%
WSTR	51.1%	2.8%	50.6%	2.6%

Note: The table reports out-of-sample forecast performance for image-based CNN models and benchmark signals. We calculate classification accuracy and correlation cross-sectionally each period then report time series averages over each period in the test sample.

Portfolio Analysis (Equal Weight Portfolios)

Table 3: Performance of Equal Weight Portfolios

	I5/R20		I20/R20		I60/R20		MOM/R20		STR/R20		WSTR/R20	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	-0.03	-0.21	-0.03	-0.16	-0.04	-0.19	0.07	0.18	0.04	0.18	-0.00	-0.02
2	0.02	0.13	0.03	0.14	0.03	0.16	0.05	0.20	0.06	0.34	0.04	0.21
3	0.05	0.29	0.05	0.29	0.06	0.31	0.06	0.31	0.08	0.48	0.07	0.41
4	0.05	0.30	0.07	0.38	0.07	0.37	0.06	0.33	0.08	0.52	0.07	0.46
5	0.08	0.41	0.08	0.45	0.09	0.49	0.05	0.35	0.08	0.53	0.08	0.54
6	0.09	0.50	0.09	0.50	0.10	0.56	0.08	0.56	0.07	0.47	0.09	0.56
7	0.09	0.47	0.10	0.58	0.11	0.63	0.10	0.69	0.09	0.58	0.09	0.55
8	0.12	0.62	0.12	0.65	0.11	0.68	0.12	0.81	0.08	0.45	0.09	0.52
9	0.14	0.72	0.13	0.77	0.12	0.76	0.12	0.79	0.08	0.35	0.11	0.54
High	0.20	1.03	0.16	0.95	0.14	0.96	0.14	0.70	0.15	0.48	0.18	0.63
H-L	0.23***	2.47	0.19***	2.18	0.19***	1.63	0.07	0.26	0.11**	0.56	0.19***	1.25
Turnover	181%		179%		160%		74%		174%		164%	

	I5/R60		I20/R60		I60/R60		MOM/R60		STR/R60		WSTR/R60	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	0.06	0.25	0.06	0.24	0.06	0.21	0.10	0.23	0.06	0.23	0.05	0.19
2	0.08	0.34	0.07	0.30	0.08	0.30	0.08	0.27	0.09	0.43	0.09	0.38
3	0.08	0.38	0.09	0.38	0.08	0.35	0.10	0.40	0.10	0.53	0.08	0.43
4	0.10	0.47	0.09	0.42	0.09	0.39	0.09	0.43	0.09	0.52	0.09	0.48
5	0.09	0.41	0.10	0.46	0.10	0.47	0.08	0.45	0.10	0.56	0.09	0.56
6	0.10	0.45	0.09	0.41	0.10	0.48	0.09	0.52	0.09	0.48	0.09	0.54
7	0.10	0.45	0.10	0.52	0.11	0.56	0.09	0.56	0.10	0.53	0.09	0.53
8	0.10	0.50	0.11	0.56	0.11	0.58	0.11	0.68	0.09	0.43	0.09	0.47
9	0.12	0.54	0.11	0.56	0.12	0.69	0.12	0.68	0.10	0.38	0.11	0.44
High	0.13	0.65	0.12	0.71	0.13	0.79	0.12	0.53	0.13	0.37	0.15	0.45
H-L	0.07***	1.16	0.06**	0.47	0.07*	0.45	0.02	0.06	0.07	0.34	0.10***	0.66
Turnover	63%		61%		54%		36%		57%		54%	

Note: Performance of equal-weighted decile portfolios sorted on out-of-sample predicted up probability. Each panel reports the average annualized holding period return and Sharpe ratio. Average returns accompanied by ***,**,* are significant at the 1%, 5% and 10% significance level, respectively. We also report monthly turnover of each strategy.

Portfolio Analysis (Valule Weight Portfolios)

Table 4: Performance of Value Weight Portfolios

	I5/R20		I20/R20		I60/R20		MOM/R20		STR/R20		WSTR/R20	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	0.03	0.20	0.02	0.09	0.02	0.11	-0.01	-0.01	0.03	0.14	0.00	0.00
2	0.02	0.14	0.03	0.15	0.05	0.27	-0.00	-0.00	0.04	0.26	0.03	0.16
3	0.03	0.18	0.03	0.19	0.04	0.23	0.02	0.07	0.03	0.22	0.02	0.16
4	0.04	0.26	0.03	0.19	0.04	0.21	0.04	0.22	0.06	0.42	0.04	0.28
5	0.05	0.34	0.06	0.33	0.04	0.27	0.04	0.27	0.06	0.45	0.07	0.46
6	0.05	0.30	0.07	0.42	0.05	0.31	0.06	0.39	0.06	0.44	0.06	0.43
7	0.06	0.42	0.05	0.36	0.06	0.41	0.06	0.47	0.09	0.54	0.09	0.63
8	0.06	0.39	0.08	0.52	0.06	0.42	0.08	0.58	0.08	0.46	0.08	0.51
9	0.09	0.55	0.07	0.46	0.05	0.37	0.08	0.57	0.06	0.27	0.08	0.41
High	0.08	0.50	0.08	0.51	0.08	0.55	0.13	0.63	0.04	0.13	0.06	0.21
H-L	0.05*	0.42	0.06**	0.56	0.06**	0.51	0.13*	0.38	0.01	0.03	0.06	0.30
Turnover	195%		180%		173%		100%		189%		189%	
	I5/R60		I20/R60		I60/R60		MOM/R60		STR/R60		WSTR/R60	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	0.05	0.23	0.03	0.15	0.05	0.26	0.07	0.16	0.05	0.20	0.03	0.10
2	0.07	0.36	0.06	0.31	0.08	0.38	0.01	0.05	0.07	0.37	0.05	0.26
3	0.06	0.34	0.07	0.38	0.04	0.23	0.07	0.28	0.06	0.37	0.06	0.35
4	0.07	0.41	0.05	0.28	0.07	0.36	0.07	0.32	0.07	0.46	0.07	0.48
5	0.08	0.44	0.08	0.40	0.08	0.48	0.06	0.37	0.08	0.56	0.09	0.57
6	0.07	0.41	0.07	0.39	0.05	0.33	0.08	0.52	0.07	0.47	0.07	0.45
7	0.07	0.46	0.07	0.43	0.06	0.38	0.07	0.54	0.09	0.54	0.09	0.54
8	0.06	0.39	0.08	0.50	0.07	0.43	0.07	0.47	0.08	0.42	0.08	0.44
9	0.08	0.51	0.08	0.53	0.09	0.60	0.08	0.51	0.07	0.32	0.07	0.36
High	0.07	0.43	0.09	0.63	0.08	0.55	0.11	0.52	0.07	0.22	0.07	0.23
H-L	0.02	0.20	0.06**	0.59	0.03	0.25	0.04	0.12	0.02	0.08	0.04	0.19
Turnover	67%		62%		58%		48%		61%		63%	

Note: Performance of value-weighted decile portfolios sorted on out-of-sample predicted up probability. Each panel reports the average holding period return and annualized Sharpe ratios. Average returns accompanied by ***,**,* are significant at the 1%, 5% and 10% significance level, respectively. We also report monthly turnover of each strategy.

Short-horizon Portfolio Analysis

Table 6: Short-horizon (One Week) Portfolio Performance

	Equal Weight											
	I5/R5		I20/R5		I60/R5		MOM/R5		STR/R5		WSTR/R5	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	-0.29	-2.04	-0.34	-2.14	-0.22	-1.21	0.14	0.41	-0.02	-0.10	-0.09	-0.41
2	-0.07	-0.44	-0.06	-0.36	-0.01	-0.07	0.08	0.36	0.04	0.22	0.02	0.12
3	0.00	0.00	0.01	0.06	0.06	0.30	0.08	0.37	0.06	0.41	0.05	0.32
4	0.04	0.22	0.07	0.37	0.08	0.40	0.07	0.41	0.08	0.49	0.06	0.41
5	0.08	0.43	0.10	0.51	0.11	0.60	0.07	0.44	0.08	0.50	0.07	0.42
6	0.10	0.51	0.13	0.66	0.13	0.72	0.09	0.57	0.09	0.53	0.08	0.49
7	0.15	0.73	0.16	0.84	0.15	0.81	0.10	0.66	0.09	0.50	0.11	0.62
8	0.20	0.96	0.20	1.01	0.18	0.97	0.12	0.77	0.10	0.51	0.12	0.62
9	0.28	1.38	0.26	1.31	0.21	1.15	0.14	0.82	0.14	0.62	0.17	0.74
High	0.53	2.79	0.50	2.67	0.32	1.78	0.16	0.74	0.38	1.16	0.45	1.53
H-L	0.82***	6.99	0.85***	6.89	0.55***	5.17	0.02	0.07	0.40***	1.78	0.54***	2.88
Turnover	847%		820%		764%		130%		358%		725%	

	Value Weight											
	I5/R5		I20/R5		I60/R5		MOM/R5		STR/R5		WSTR/R5	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	-0.06	-0.37	-0.06	-0.37	-0.05	-0.28	0.01	0.02	0.02	0.09	-0.04	-0.16
2	-0.01	-0.09	0.01	0.06	0.00	0.02	-0.01	-0.02	0.02	0.10	0.00	0.03
3	0.03	0.19	0.02	0.11	0.02	0.13	0.04	0.17	0.03	0.20	0.04	0.21
4	0.02	0.11	0.02	0.14	0.01	0.06	0.05	0.23	0.07	0.41	0.04	0.22
5	0.06	0.33	0.04	0.22	0.03	0.15	0.05	0.28	0.07	0.43	0.05	0.33
6	0.05	0.28	0.06	0.32	0.05	0.29	0.05	0.31	0.08	0.44	0.08	0.47
7	0.09	0.51	0.08	0.47	0.06	0.34	0.06	0.38	0.07	0.39	0.09	0.52
8	0.11	0.57	0.09	0.51	0.08	0.45	0.08	0.50	0.11	0.53	0.13	0.64
9	0.13	0.67	0.11	0.57	0.10	0.53	0.09	0.53	0.11	0.43	0.16	0.65
High	0.19	0.86	0.17	0.86	0.13	0.73	0.13	0.59	0.15	0.42	0.17	0.54
H-L	0.25***	1.63	0.24***	1.69	0.19***	1.57	0.13	0.36	0.13**	0.45	0.21***	0.78
Turnover	979%		869%		895%		121%		430%		840%	

Note: Performance of equal-weighted (top panel) and value-weighted (bottom panel) decile portfolios sorted on out-of-sample predicted up probability. Each panel reports the average holding period return and annualized Sharpe ratios. Average returns accompanied by ***,**,* are significant at the 1%, 5% and 10% significance level, respectively. We also report monthly turnover of each strategy.

Transfer Learning and International Market Performance

Table 11: International Transfer and H-L Decile Portfolio Sharpe Ratios (I5/R5)

	Stock Count	Equal Weight			Value Weight		
		Re-train	Direct Transfer	Transfer−Re-train	Re-train	Direct Transfer	Transfer−Re-train
Global	17206	0.18	5.20	5.03***	0.46	-3.05	-3.50
Japan	3056	3.56	5.68	2.12***	0.96	1.23	0.27
Canada	2924	9.01	12.12	3.11***	2.98	5.34	2.36***
India	1861	2.52	-1.46	-3.98	0.67	-1.08	-1.75
UnitedKingdom	1783	0.03	-0.23	-0.26	1.04	0.98	-0.06
France	955	2.47	4.09	1.63***	1.12	2.10	0.98***
SouthKorea	911	3.64	1.66	-1.97	1.74	2.39	0.65***
Australia	886	8.28	11.37	3.09***	2.78	3.48	0.70***
Germany	868	-0.29	2.43	2.72***	-0.01	2.93	2.94***
China	662	2.26	-2.19	-4.45	0.66	-0.95	-1.62
HongKong	543	1.97	5.35	3.37***	0.72	2.08	1.36***
Singapore	284	6.98	6.79	-0.20	2.48	3.94	1.46***
Sweden	260	5.43	6.99	1.56***	0.83	2.37	1.54***
Italy	241	2.14	3.55	1.40***	0.76	1.60	0.84***
Switzerland	240	0.48	0.67	0.19	1.30	2.62	1.33***
Denmark	223	1.94	3.56	1.62***	1.18	1.85	0.68***
Netherlands	212	-0.30	3.75	4.05***	0.11	1.67	1.56***
Greece	201	2.74	3.26	0.51**	0.98	1.88	0.90***
Belgium	171	0.73	4.34	3.60***	0.73	2.88	2.15***
Spain	170	1.62	0.28	-1.35	0.68	1.02	0.34*
Norway	169	0.79	3.38	2.59***	1.11	2.88	1.77***
Portugal	121	0.30	2.64	2.33***	0.93	1.40	0.47**
NewZealand	114	0.50	2.34	1.84***	0.65	1.19	0.54***
Finland	113	2.66	5.38	2.72***	0.95	2.55	1.60***
Austria	110	0.14	0.67	0.53**	0.66	1.05	0.39**
Ireland	75	0.47	1.80	1.34***	0.31	1.99	1.69***
Russia	53	-0.72	2.19	2.91***	-0.13	0.44	0.57***
Average	1274	2.21	3.54	1.34	0.99	1.73	0.75
Average (excluding Global)	661	2.28	3.48	1.19	1.01	1.92	0.91

Note: The table reports annualized out-of-sample Sharpe ratios for H-L decile spread portfolios within each country. We report the average monthly stock count by country, the image-based strategy from re-training the I5/R5 CNN using local data, and the image-based strategy directly transfers the I5/R5 model estimated in US data without re-training. Sharpe ratio gains (Transfer−Re-train) accompanied by ***,**,* are significant at the 1%, 5% and 10% significance level, respectively.

Sentiment Analysis with NLP

Textual Data and NLP

- According to IDC, the size of digital data will be 40 zettabytes by 2020, more than 5,200 gigabytes for every person in the world.
- Much of its is text from various sources such as web, social media, newswire, emails, regulatory documents…
- How do investors make sense of text data?
- Natural Language Processing (NLP) helps to convert texts (unstructured) into an easier to use format (structured).

NLP and Sentiment Analysis

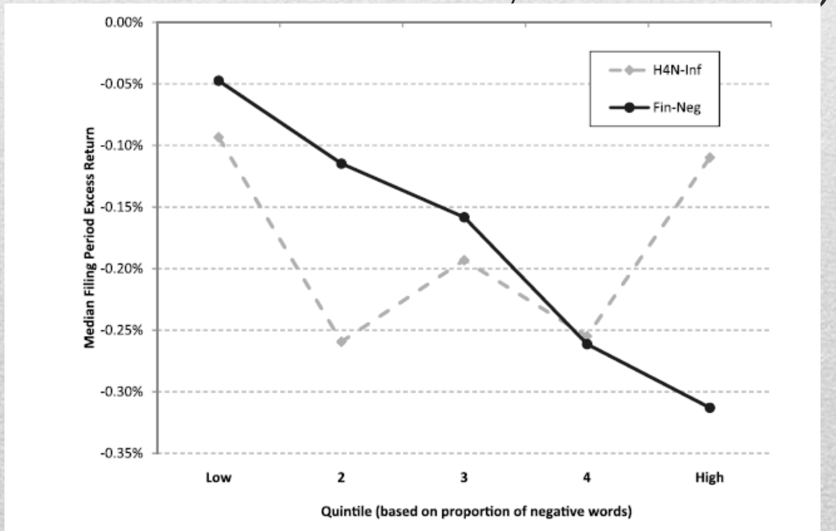
- Data Preprocessing
  - Tokenization: covert sentences to words
  - Remove stop words-frequent words such as “the”, “is”, etc.
  - Stemming and lemmatization: reduce words to its root (playing, plays, played=> play)
- Sentiment Analysis
  - Dictionary based approach: positive/negative words: https://sraf.nd.edu/textual-analysis/resources/
  - Machine learning approach:
    - Feature extraction: mapping text to real value vector (Bag of Words and Word2vec etc.)
    - Train a machine learning algorithm

Selected Research on Sentiment Analysis

- Sentiment in news predict short-term stock returns:
  - Tetlock (2007): WSJ's "Abreast of the Market" column
  - Tetlock, Sarr-Tsechansky, and Macskassy (2008): all WSJ and Dow Jones News Service news
- Tone in earnings press release (Henry 2008)
- Tone in 10-K/Q filing:
  - Management discussion and analysis (MD&A) section of : (Feldman, Govindaraj, Livnat, and Segal 2010; Li 2010)
  - 10-K Filings (Loughran and McDonald 2011)
- Conference calls (prepared remarks and Q&As) (Brockman, Li, and Price 2015)
- Social media sentiment:
  - Bollen, Mao and Zeng (2011): Twitter feeds
  - Chen, De, Hu and Hwang (2014): Seeking Alpha websites

Dictionary based measure of sentiment

- Harvard General Inquirer list: http://www.wjh.harvard.edu/~inquirer
- Loughran and McDonald (2011)
  - A word list developed for psychology and sociology may not translates well into business, for example, t*ax*, *cost*, *capital*, *board*, *liability*, *foreign*, and *vice* are negative on the Harvard list
  - Create a list of 2,354 words that typically have negative implications in a financial sense, and a list of 354 positive words (https://sraf.nd.edu/textual-analysis/resources/)



Word power: A New approach for content analysis, by Jegadeesh and Wu (2013)

- Not all positive (negative) words are equally good (bad), e.g. bad vs worst, thus, the weight on each word should be different too
- When assign a positive/negative sentiment score, it should satisfy the following properties:
  - The score should be positively related to the number of occurrence of each positive or negative word
  - The score should be positively related to the strength of the positive or negative words
  - The score should be inversely related to the total number of words in the documents.

We propose the following functional form for the score for document i that satisfies the above properties:

$$Score_i = \sum_{j=1}^{J} (w_j F_{i,j}) \frac{1}{a_i}, \tag{4}$$

where w_j is the weight for word j and $F_{i,j}$ is the number of occurrences of word j in document i. The term $1/a_i$ reflects the fact that the score is negatively related to the total number of words in the document. To the extent that the

Estimation of the weight/strength

- Determine the weight/strength of words based on market reaction to 10-K filings.
- Assumption: the market reaction would be more positive for filings with more positive overall sentiment.
- Using both the Loughran and McDonald (2011) wordlist and the global list, which combine i) LM list, ii) the Harvard IV-4 dictionaries, the top and bottom 200 words from the word list developed by Bradley and Lang (1999).

$$r_i = a + b\left(\sum_{j=1}^{J}(w_j F_{i,j})\frac{1}{a_i}\right) + \epsilon_i = a + \left(\sum_{j=1}^{J}(bw_j F_{i,j})\frac{1}{a_i}\right) + \epsilon_i, \quad (5)$$

where r_i is the abnormal return when the ith document is released.

While we can directly compute $F_{i,j}$ and a_i, we have to estimate the weights associated with each word. To do so, we fit the regression

$$r_i = a + \left(\sum_{j=1}^{J}(B_j F_{i,j})\frac{1}{a_i}\right) + \epsilon_i. \quad (6)$$

In this regression, we treat B_j's as regression coefficients and the estimated values of these coefficients provide unbiased estimates of bw_j. We cannot separately estimate b and w_j at this stage because the weights measure the relative strength of each word in the lexicon and the weights can be scaled arbitrarily. We standardize the estimates of B_j's to obtain an estimate of the weight for each word. Specifically,

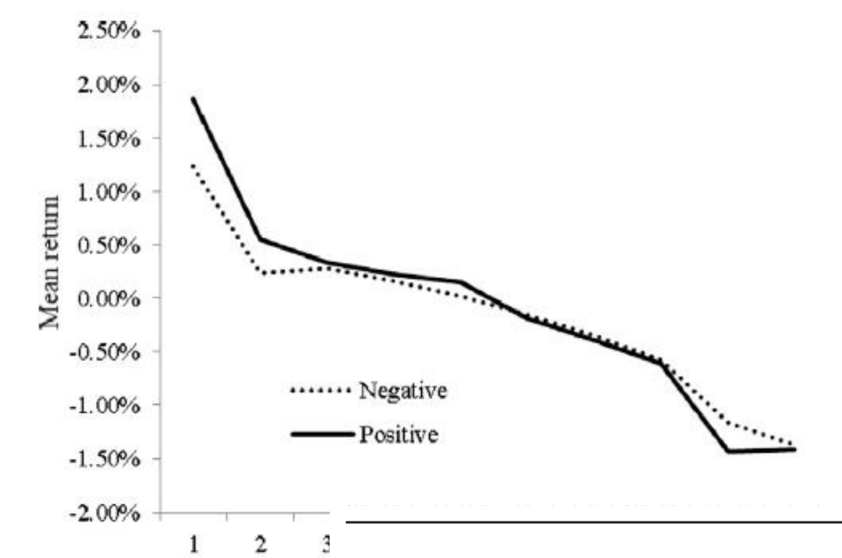
$$\hat{w}_j = \frac{\hat{B}_j - \overline{B}}{\text{Standard Deviation}(\hat{B}_j)}, \quad (7)$$

where $\hat{w}_j$ is our estimate of w_j, $\hat{B}_j$ is the slope coefficient estimate in from Eq. (6), and $\overline{B}$ is the mean of $\hat{B}_j$ across all words.

To examine whether our estimate of score is related to returns, we fit the regression

$$r_i = a + b\left(\sum_{j=1}^{J}(\hat{w}_j F_{i,j})\frac{1}{a_i}\right) + \epsilon_i. \quad (8)$$

Sentiment Score/Word Power and 10-K Filing Period Abnormal Returns



Panel C: Both Positive and Negative Scores

(Rank correlation of positive

	Models	
	(7)	(8)
Term weighting scheme		
WP (positive)	0.300	0.191
	(2.45)	(2.74)
WP (negative)	0.219	0.132
	(2.64)	(3.84)
Control variables		
Size		−0.018
		(−0.21)
BM		2.330
		(1.37)
Volatility		−0.238
		(−1.68)
Turnover		−0.109
		(−1.48)
EADRet		0.572
		(5.75)
		−0.225
		(−1.53)

	Models			
	Combined LM lexicon		Global lexicon	
	(1)	(2)	(3)	(4)
Term weighting scheme				
WP	0.343	0.192	0.294	0.190
	(2.67)	(3.81)	(2.44)	(3.58)
Control variables				
Size		−0.018		−0.019
		(−0.21)		(−0.14)
BM		2.631		2.461
		(1.45)		(1.00)
Volatility		−0.312		−0.334
		(−1.75)		(−1.84)
Turnover		−0.117		−0.123
		(−1.56)		(−1.62)
EADRet		0.575		0.546
		(5.80)		(6.38)
Accruals		−0.312		−0.312
		(−1.75)		(−1.62)

Sentiment Score and Future Stock Returns

Table 9
Document tone and future returns.

This table reports the slope coefficient of the regression of future stock returns against document score. Market-adjusted returns is stock return minus contemporaneous CRSP value-weighted index return, and size-adjusted return is stock return minus the contemporaneous return on matched size decile portfolio (available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The dependent variable is the abnormal returns computed within the event windows specified at the top of the respective columns. The independent variables in all regressions are the word power (WP) score calculated using lists of positive and negative words. We compute the word power weights for each year using Eqs. (6) and (7) over the sample period prior to the filing of 10-Ks, and we compute positive and negative WP scores for each 10-K using Eq. (4). The estimates use a sample of 45,860 10-Ks over 1995–2010. The independent variables are standardized to a mean of 0 and standard deviation of 1. The table reports the coefficients and *t*-statistics computed using the Fama-MacBeth approach with annual regressions.

	Event windows		
Dependent variable	+5 to +9	+5 to +14	+5 to +26
Panel A: Positive words			
Market-adjusted returns	0.132	0.200	0.228
	(2.06)	(1.81)	(0.07)
Size-adjusted returns	0.093	0.123	0.130
	(1.98)	(1.80)	(0.25)
Panel B: Negative words			
Market-adjusted returns	0.101	0.132	0.191
	(1.93)	(1.51)	(0.83)
Size-adjusted returns	0.111	0.127	0.144
	(1.90)	(1.44)	(0.45)

FinBert by Huang, Wang and Yang (2020)

- BERT (Bidirectional Encoder Representations from Transformers), Google's state-of-the-art language model for NLP, which learn the language model by:
  - Masked Language Modeling (LM): randomly mask 15% of the words with a [MASK] token, and then attempts to predict the original value of the masked words, based on the context provided by the other, non-masked words in the sequence
  - Next Sentence Prediction (NSP): the model receives pairs of sentences as input and learns to predict whether the 2nd sentence in the pair is the subsequent sentence in the original document.

BERT Fine-Tuning for Specific Tasks

- Google pre-trained two BERT models using general text copus from Wikipedia and BooksCorpus with a total of 3.3 billion word tokens:

$BERT_{BASE}$	$BERT_{LARGE}$
Layers = 12	Layers = 24
Hidden size = 768	Hidden size = 1024
self-Attention heads = 12	self-Attention heads = 16
Total parameters = 110M	Total parameters = 340M

- Using transfer learning, users can fine-tune the pre-trained model for specific tasks such as sentiment analysis, question-answering tasks, and named entity recognition etc.
- **Sentiment analysis:** adding a classification layer on top of the transformer output to predict sentiment labels (by human), just like the Next Sentence classification
- Huang et al. (2020)
  - Pre-train the FinBERT based on the pretrained BERT by google using financial text in 10-K, 10-Q, Earnings conference call and Analyst Report
  - Fine-tune the FinBERT model for sentiment classification using a sample of 10,000 pre-labeled sentences from financial text

Performance of Sentiment Score of FinBERT

- Sentiment classification accuracy
  - FinBERT: 88.4%, Loughran and McDonald: 61.7%, BERT: 85.5%, Naïve Bayes: 82.7%, Word2Vec 50.9%
- FinBERT based sentiment score has higher association with market reaction to conference calls and abnormal trading volume

Panel A: Regression of cumulative abnormal return on textual sentiments

	(1)	(2)	(3)	(4)	(5)
Dependent Variable			*CAR*		
$Tone_{FinBERT}$	0.734***				
	(15.01)				
$Tone_{BERT}$		0.709***			
		(15.08)			
$Tone_{LM}$			0.464***		
			(9.77)		
$Tone_{NB}$				0.369***	
				(8.10)	
$Tone_{W2V}$					0.175***
					(3.86)

- FinBERT based sentiment score also predict future earnings better than the sentiment score based on the LM dictionary

Other Approaches

- *SESTM* by Ke, Kelly, and Xiu (2020)
  - Identify the sentiment-charged dictionary S using frequency and return thresholds.
  - Estimate the vectors of positive and negative sentiment topics by regressing word frequencies on sentiment ranks.
  - Predict sentiment score of a new article using Maximum Likelihood Estimation (MLE) with a penalty term.
- *FarmPredict* by Fan, Xue and Zhou (2021)
  - Extract hidden topics (factors) from all words (PCA)
  - Screen the idiosyncratic variables by their correlation with stock returns.
  - Apply simple LASSO to predict asset price using hidden factors and screened idiosyncratic components.

Thank You!