Text as Data and Sparse Portfolios

Week 7

Daniele Bianchi[1]

whitesphd.com

[1] School of Economics and Finance
Queen Mary, University of London

Summary

This week we expand the use of machine learning methods for the predictability of aggregate stock. In particular, we are going to cover three main papers in the academic literature which opens up the possibility of using both linear and non-linear methods to improve returns forecasting.

Contents

Text as Data

Text Analytics in Finance

- Content Analysis: Mine for significant correlations (Who is interested in what?)
- Topic detection and classification (What are these texts about?)
- Sentiment analysis: Attitudes (positive or negative), Emotion classification (Fear, Joy, Surprise, Risk, Intent, and so on).
- Deception detection: What is the speaker/writer trying to hide?

Who use text analytics?

- Retail industry: Brand management, trend tracking.
- Securities markets: Pricing news, opinion, emotions.
- Financial regulators: Fraud detection.
- Public health: Search patterns for medical suggestions.
- What kind of texts?
  1. New articles.
  2. Blogs, social media (e.g. Twitter).
  3. Company filings.

Sentiment Analysis: Tetlock et. al. (2008)

- Quantify language to predict earnings and stock returns.
- The tone and sentiment of financial news stories may say something about firms.
- Example: "Consumer Groups Say Microsoft Has Overcharged for Software".
- Sentences in the article: "The alleged 'pricing abuse will only get worse if Microsoft is not disciplined sternly by the antitrust court,'..."

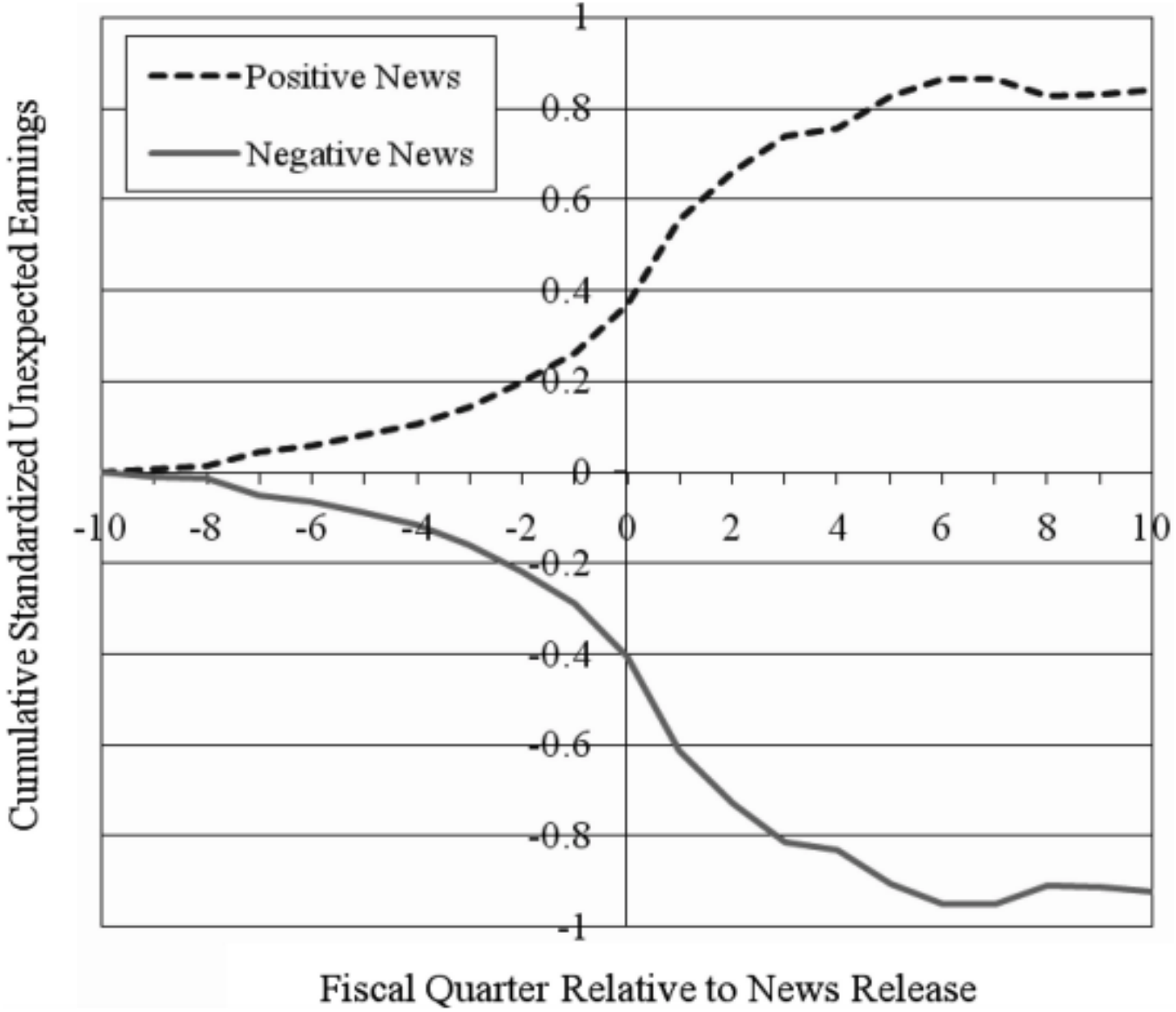
Sentiment Analysis: Tetlock et. al. (2008)

- Use classification dictionary available to train computers (Example: Harvard-IV-4 classification dictionary).
- Based on the classification, this sentence's fraction of negative words rank very high.
- Microsoft's cumulative abnormal stock returns were $-42, -141$ and -194 for the three trading days around this news event!
- Key question: Can we systematically use such data to understand cash flows, and therefore stock returns?
- Caveat: The dictionary used for these purposes matter! Loughran and McDonald (2011) show that 75% of the words in the Harvard classification need not necessarily be negative in finance context!

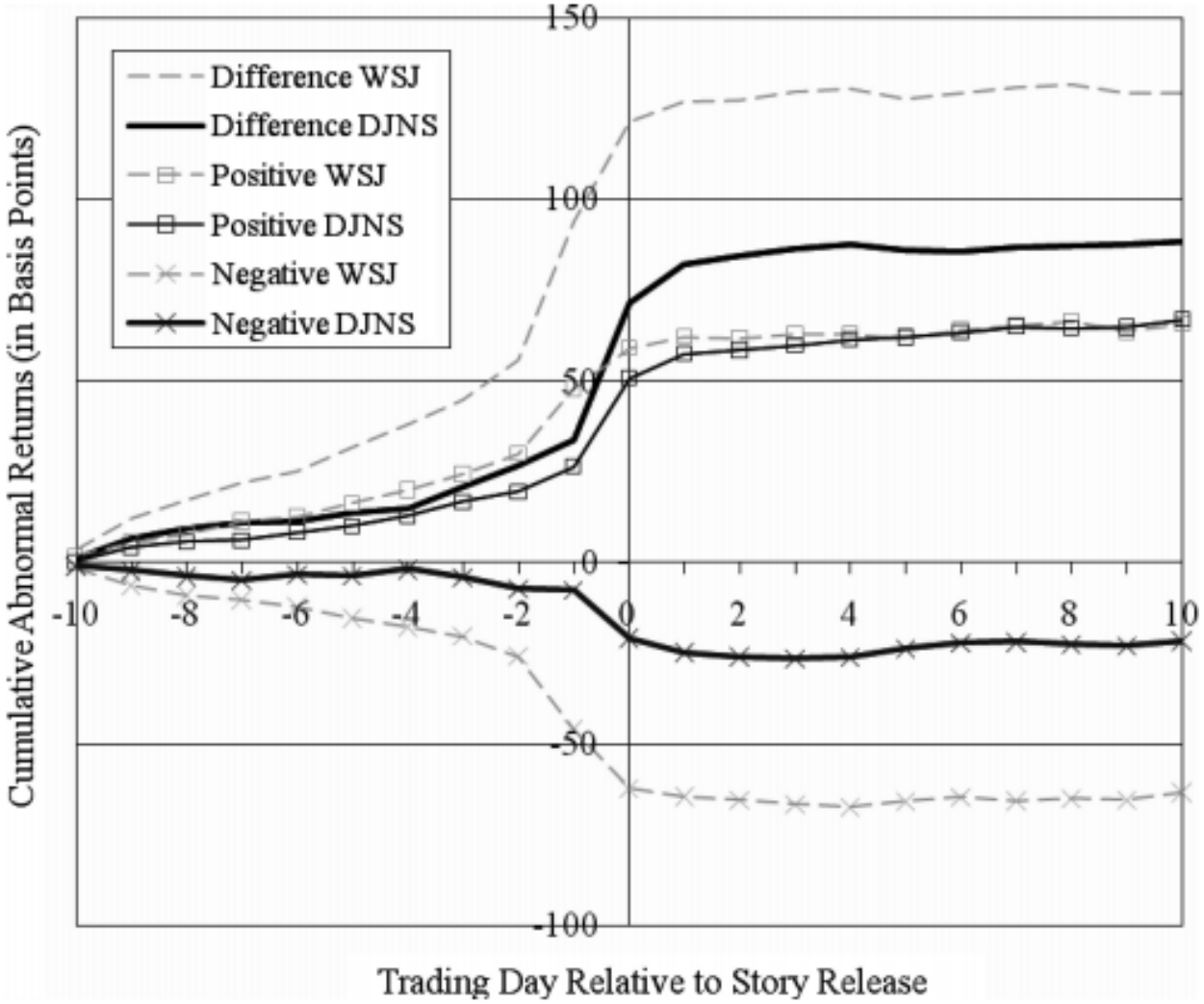
Predicting Earnings Using Negative Words

	SUE				*SAFE*	
Stories Included	*DJNS*	*WSJ*	Before Forecasts	All Stories	Before Forecasts	All Stories
$neg_{-30,-3}$	−0.0584	−0.1083	−0.0640	−0.0637	−0.0192	−0.0197
	(−4.42)	(−5.28)	(−3.95)	(−4.69)	(−3.79)	(−4.44)
Lag(Dependent Var)	0.2089	0.2082	0.2042	0.2101	0.2399	0.2523
	(11.82)	(11.83)	(11.90)	(11.98)	(7.82)	(8.74)
Forecast Dispersion	−0.9567	−1.0299	−0.9634	−0.9373	−0.2984	−0.3076
	(−9.84)	(−9.59)	(−9.21)	(−10.20)	(−5.34)	(−6.34)
Forecast Revisions	20.2385	18.0394	20.4855	19.5198	0.5111	0.7580
	(8.89)	(7.91)	(8.51)	(8.94)	(0.68)	(1.19)
Log(Market Equity)	−0.0071	0.0003	−0.0043	−0.0037	0.0258	0.0289
	(−0.40)	(0.01)	(−0.24)	(−0.21)	(4.79)	(5.32)
Log(Book / Market)	0.0173	0.0182	0.0221	0.0204	−0.0162	−0.0110
	(0.62)	(0.56)	(0.77)	(0.75)	(−1.97)	(−1.41)
Log(Share Turnover)	−0.1241	−0.1348	−0.1095	−0.1261	0.0274	0.0254
	(−3.09)	(−2.90)	(−2.75)	(−3.20)	(2.69)	(2.61)
$FFAlpha_{-252,-31}$	1.9784	1.9711	1.9770	2.0015	0.2199	0.2382
	(9.14)	(9.90)	(10.01)	(9.50)	(4.17)	(4.36)
$FFCAR_{-30,-3}$	0.0119	0.0129	0.0117	0.0116	0.0062	0.0071
	(6.76)	(6.33)	(6.28)	(6.64)	(10.17)	(11.38)
$FFCAR_{-2,-2}$	0.0104	0.0103	0.0177	0.0118	0.0053	0.0037
	(1.65)	(1.37)	(2.40)	(1.91)	(2.30)	(1.89)
Observations	16755	11192	13722	17769	12907	16658
Clusters	80	79	78	80	78	79
Adjusted R^2	0.1177	0.1204	0.1158	0.1187	0.1163	0.1244

Predicting Earnings



Predicting Returns

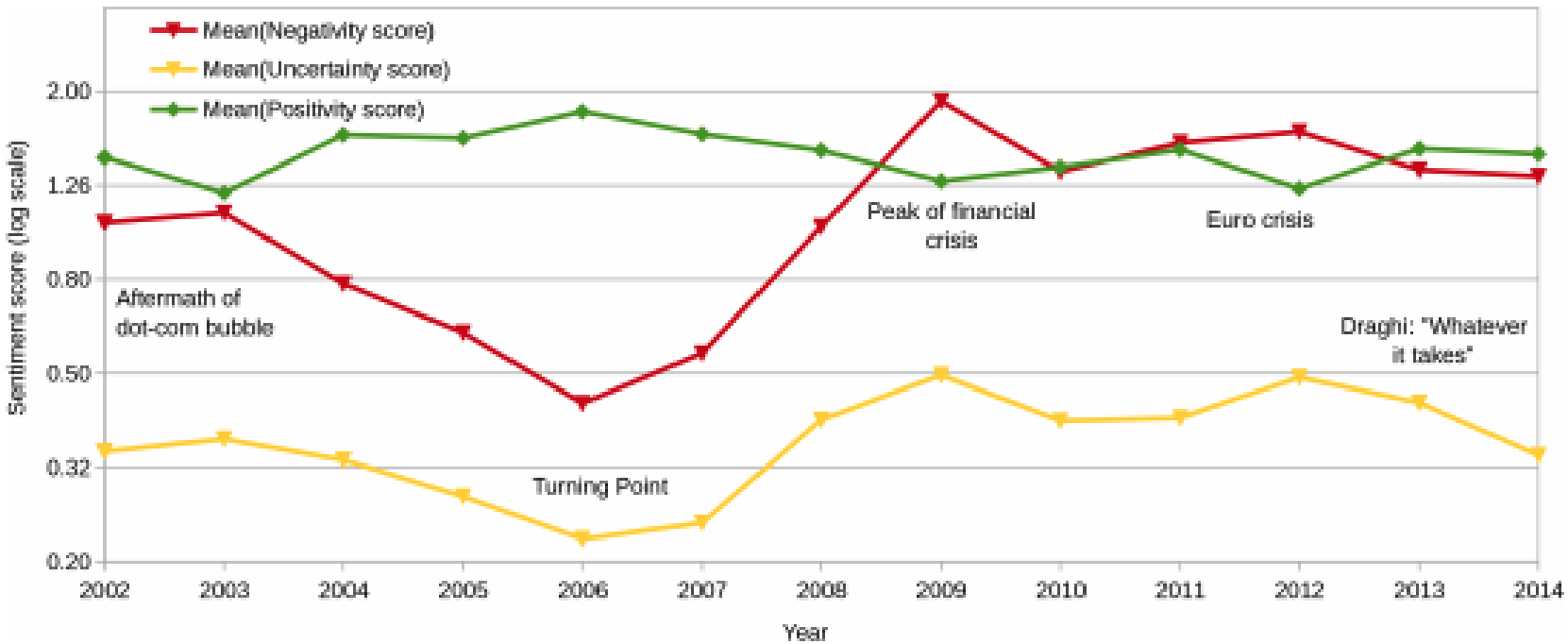


Deception Detection

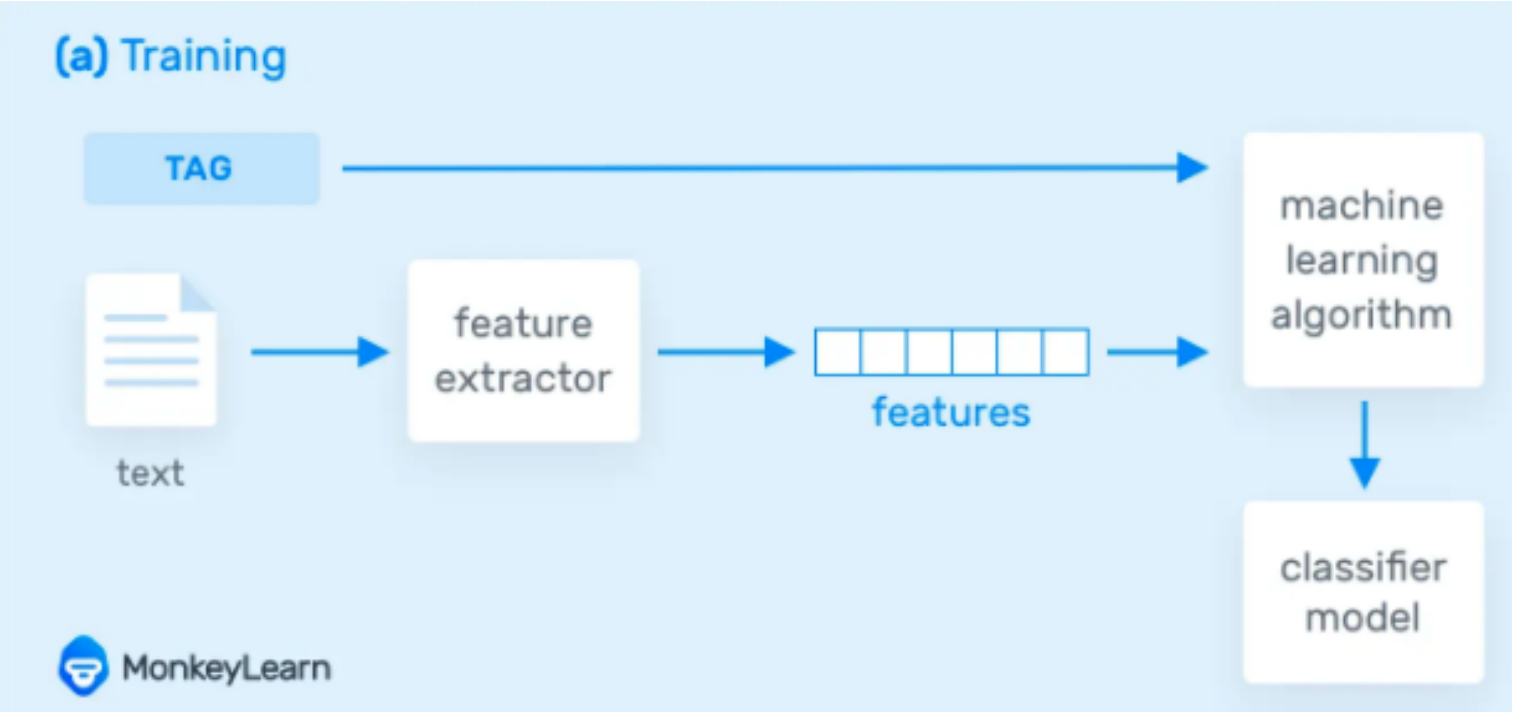
- Text based models can help classify information as "truthful" or "deceptive".
- Larcker and Zakolyukina (2012) use CEO and CFO statements during quarterly earnings calls narratives.
- For example: Label a statement as deceptive if there is disclosure of material weakness, auditor change, late filing etc.
- Deceptive executives exhibit more general knowledge, few nonextreme positive emotions, and fewer reference to shareholder value.
- Deceptive CEOs use fewer anxiety words, use extreme positive emotion etc.
- Portfolio formed from firms with highest deception scores produce negative alpha (-4% to -11%!)

Detecting Risks in the Banking System

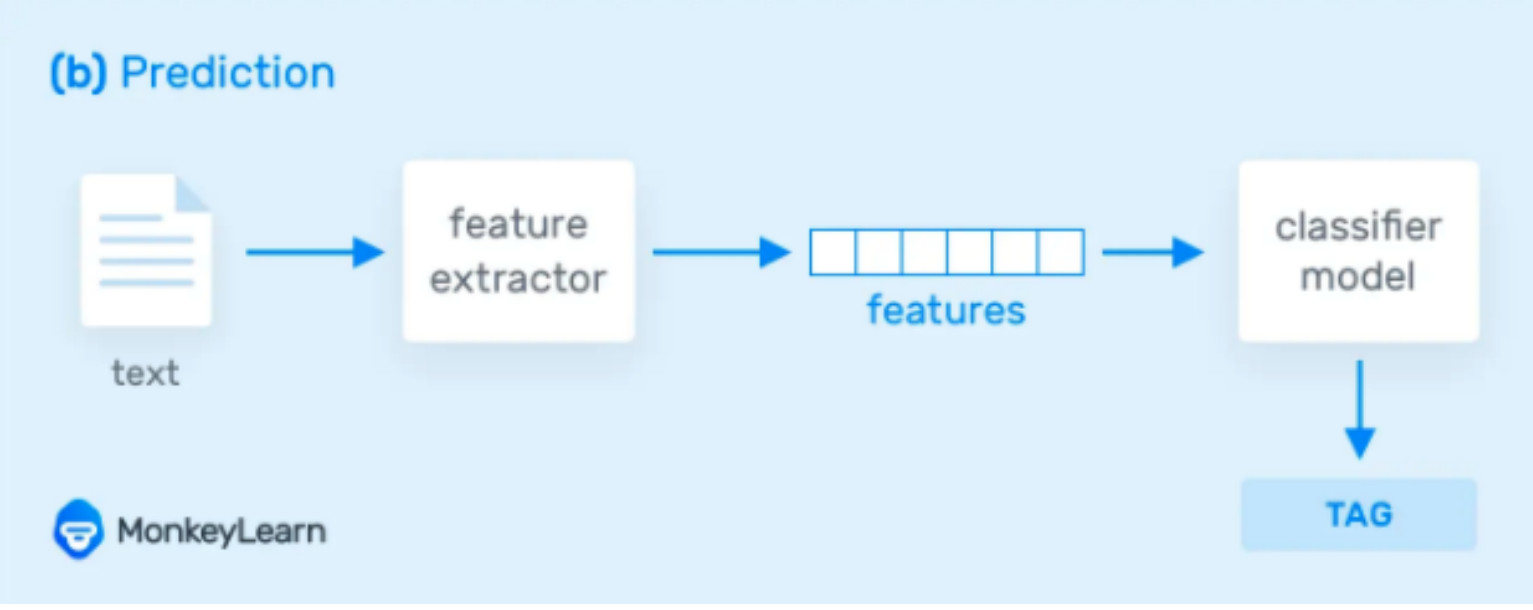
Nopp and Hanbury (2015): CEO Letters



Using Big Data Techniques with Text Data



Using Big Data Techniques with Text Data



Portfolio Management and Big Data Tools

Portfolio Construction

- The Markowitz Framework.
- Purely pursuing high returns constitutes a poor strategy.
- Rational investors: Risk-return tradeoff. Balance expectation of high returns and low risk, measured by variability of returns.
- Markowitz in the modern world: Any of the portfolio selection algorithm available do not consistently outperform a simple naive approach of equally weighting all stocks.
- Why?
- The standard optimization at the core of Markowitz scheme is empirically unstable:
- Small changes in asset returns, volatilities, or correlations can have large effects on optimal portfolio!

Enter Brodie et. al. (2008)

- Use "Penalized Regression".
- Their proposal is very similar to what we have already studied early on this term with big data techniques: add a penalty term proportional to the sum of the absolute values of the portfolio weights.
- This is the "$\ell 1$" norm!
- The rest of today's lecture will set this up and show you how the penalized regression approach can improve your good old portfolio optimization strategy!

Portfolio Construction

- N securities, with returns at time t is an $N \times 1$ vector $\boldsymbol{r}_t = (r_{1,t}, ..., r_{N,t})^\top$. $r_{1t} = r'_{1,t} - r_f$.
- $E[\boldsymbol{r}_t] = \boldsymbol{\mu}$. In words, the expected return vector is a vector of returns for each stock $i \in N$.
- Variance Covariance Matrix of returns:

$$E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_t - \boldsymbol{\mu})^\top] = \Omega \tag{1}$$

$$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \cdots & \sigma_1\sigma_N \\ \sigma_2\sigma_1 & \sigma_2^2 & \cdots & \sigma_2\sigma_N \\ \vdots & & \ddots & \\ \sigma_N\sigma_1 & \sigma_N\sigma_2 & \cdots & \sigma_N^2 \end{pmatrix} \tag{2}$$

A portfolio

A portfolio is a list of weights ω_i for each asset $i \in N$. To simplify the setup, let us assume we have one pound to be fully invested, so $\sum_{i=1}^{N} \omega_i = 1$ The vector

$$\boldsymbol{w} = (\omega_1, ..., \omega_N)^\top \tag{3}$$

In matrix terminology, the normalization constraint can be rewritten as

$$\boldsymbol{w}^\top 1_N = 1 \tag{4}$$

where 1_N is an $N \times 1$ vector in which every entry is 1. Portfolio Expected return:

$$\boldsymbol{w}^\top \boldsymbol{\mu} \tag{5}$$

Portfolio Variance:

$$\boldsymbol{w}^\top \Omega \boldsymbol{w} \tag{6}$$

Traditional Markowitz Optimization

- Find a portfolio that has the lowest variance for a given level of expected returns.

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}}(\boldsymbol{w}^\top \Omega \boldsymbol{w}) \tag{7}$$

- Subject to the following constraints:

$$\boldsymbol{w}^\top 1_N = 1 \tag{8}$$

$$\boldsymbol{w}^\top \boldsymbol{\mu} = \Delta \tag{9}$$

Empirical Implementation

- Replace expectations by sample averages: $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t$ Then, $\boldsymbol{R}$ is a $T \times N$ matrix of returns.
- Replace variance covariance matrix by sample variance covariance matrix!
- This problem is analytically super simple. However, very challenging in practice.

Key Challenge

- **Condition Number:** Measure how much the output value can change for a small change in the input argument.
- A problem with low condition number is well conditioned and stable. A problem with high condition number is unstable.
- Condition number = Ratio of largest to smallest singular values of a matrix.
- Asset returns tend to be highly correlated, and the smallest singular values tend to be quite small.
- What is the solution: A regularization procedure.

Rewrite Optimization Procedure

Replacing with empirical estimates (sample averages, covariance matrix), and rewriting the problem as follows:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \frac{1}{T} \mid\mid \Delta \mathbf{1}_T - \boldsymbol{R}\boldsymbol{w} \mid\mid_2^2 \tag{10}$$

subject to the following constraints:

- $\boldsymbol{w}^\top \mu = \Delta$.
- $\boldsymbol{w}^\top 1_N = 1$

Note here that $\mid\mid \boldsymbol{x} \mid\mid_2^2$ is the sum $\sum_{t=1}^{T} \boldsymbol{x}_t^2$. In order to improve the solution, we can augment this optimization problem to penalize using an ℓ_1 penalty.

Regularization

- Regularization is a way to avoid overfitting by penalizing high-valued coefficients.
- ℓ_1 regularization prevents the coefficients to fit so perfectly to the data. It is the sum of weights.
- So, with ℓ_1 regularization, the new optimization problem is:

$$\hat{\boldsymbol{w}}^{[\tau]} = \arg\min_{\boldsymbol{w}} \frac{1}{T} \mid\mid \Delta \mathbf{1}_T - \boldsymbol{R}\boldsymbol{w} \mid\mid_2^2 + \tau \mid\mid \boldsymbol{w} \mid\mid_1 \tag{11}$$

subject to the following constraints: $\boldsymbol{w}^\top \mu = \Delta$ and $\boldsymbol{w}^\top 1_N = 1$. Note that the ℓ_1 norm of a vector $\boldsymbol{w}$ is defined by $\mid\mid \boldsymbol{w} \mid\mid := \sum_{i=1}^{N} \mid w_i \mid$. Finally, τ is the parameter that allows you to adjust the relative importance of the ℓ_1 penalty.

Our old friend: LASSO

- The entire procedure is the same as our old friend LASSO.
- The problem of minimizing an objective function with the ℓ_1 penalty helps "reduce" the number of weights that are non-zero, and therefore create "sparse portfolios".
- Investors typically want to limit the number of positions held, monitored and liquidated.
- Since we do not constrain the weights to be non-negative, in this setup, you can short stocks! With the ℓ_1 penalty, you limit how much shorting you do in the market.
- It stabilizes the problem and therefore you will find a solution to the Markowitz portfolio!

Out of Sample Performance

Fama-French 48 Industry Portfolios:

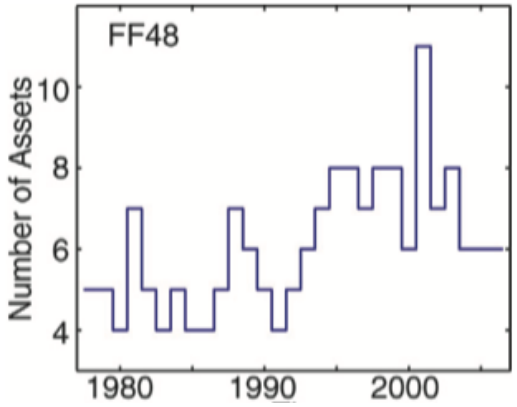
Evaluation	Eq. weight			$w_i \geqslant 0$			bin=8–16			bin=17–24			bin=25–32			bin=33–40			bin=41–48		
period	m		S	m		S	m		S	m		S	m		S	m		S	m		S
07/76–06/06	17	61	27	17	41	41	16	40	40	14	40	34	12	43	28	12	47	26	12	54	22
07/76–06/81	29	66	44	23	48	49	20	41	50	18	39	46	19	40	49	22	43	50	23	50	46
07/81–06/86	18	58	31	23	41	57	25	42	58	23	44	52	24	46	52	23	50	46	22	56	39
07/86–06/91	5	72	7	9	45	20	8	43	18	7	44	15	4	47	9	4	51	7	5	57	8
07/91–06/96	18	41	44	16	26	62	15	26	57	13	27	47	12	33	36	12	41	30	11	52	21
07/96–06/01	11	67	17	16	40	40	16	41	38	9	42	22	3	50	6	2	54	4	0	61	0
07/01–06/06	18	60	30	13	43	30	13	42	30	12	41	29	10	38	28	10	37	27	12	43	27

Fama-French 100 Industry Portfolios:

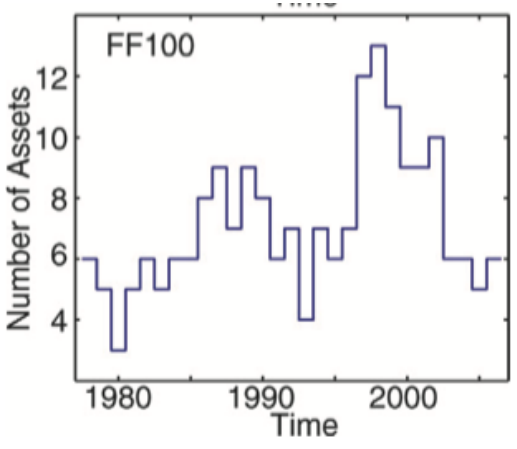
Evaluation	Eq. weight			$w_i \geqslant 0$			bin=11–20			bin=21–30			bin=31–40			bin=41–50			bin=51–60		
period	m		S	m		S	m		S	m		S	m		S	m		S	m		S
06/76–06/06	17	59	28	16	53	30	16	50	33	19	48	39	19	49	40	20	52	39	21	60	34
07/76–06/81	23	61	38	12	59	21	11	52	22	12	51	24	12	55	22	11	56	20	7	66	10
07/81–06/86	20	53	38	24	49	49	26	41	64	31	38	81	31	40	77	31	43	72	33	49	67
07/86–06/91	9	71	13	10	65	15	9	63	14	9	61	16	11	62	18	12	64	19	12	71	17
07/91–06/96	18	34	53	19	31	61	20	29	70	22	25	86	20	28	73	22	31	70	25	36	67
07/96–06/01	16	63	26	18	52	35	18	53	35	23	52	44	29	47	61	31	50	62	34	63	54
07/01–06/06	12	64	19	11	55	21	11	53	22	15	51	29	13	51	26	13	56	23	14	64	22

Number of Assets in Portfolio

Fama-French 48 Industry Portfolios



Fama-French 100 Industry Portfolios



Wrap-up

- Big-data tools can benefit greatly even in "small data" problems!
- Text as data is perhaps the single largest transformation of the finance industry in the recent past.
- A great deal of work can be done using big data techniques in every type of role in the finance industry.
- This is an introductory course on Big Data Techniques and therefore we have only merely scratched the surface.
- Next week: Wrap up lecture that tie together all of the sessions so far, and details on final examination.

References

1. Brodie, J., Daubechies, I., De Mol, C., Giannone, D., Loris, I. (2009). Sparse and stable Markowitz portfolios. Proceedings of the National Academy of Sciences, 106(30), 12267-12272.

2. Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. "More than words: Quantifying language to measure firms' fundamentals." The Journal of Finance 63, no. 3 (2008): 1437-1467.

3. Larcker, David F., and Anastasia A. Zakolyukina. "Detecting deceptive discussions in conference calls." Journal of Accounting Research 50, no. 2 (2012): 495-540.