IEOR E4650 Business Analytics

Session 5: Financial Analytics

Spring 2018

Prof. Adam Elmachtoub

Quantitative Investment Strategies: Objective

How can analytics capture value in the investment process?

Goal: make money! (without too much risk)

1. Use data to predict future prices
2. Make trading decisions based on predictions

Data

What data can be used to predict a stock price in the future?

- Own price history
- Cross-sectional price history (e.g., AAPL vs. GOOG)
- Fundamental data (sales, earnings, supply chain indicators, etc.)
- News
- Analyst ratings, sentiment, earnings revisions

Measuring Financial Performance

Performance measures

- Average return
- Average return relative to risk
- Return relative to a benchmark
  Compare return to a passive strategy using the same or similar assets

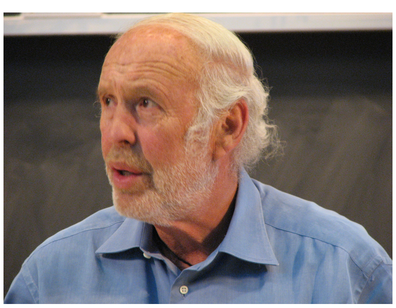
Can we predict the stock market at all?

What do you think, Jon Stewart?

`http://www.cc.com/video-clips/3h5dk0/the-daily-show-with-jon-stewart-stock-market`

Quantitative Investment Strategies: Theory versus Practice

Theory: Markets are efficient ⇒ no arbitrage opportunities



Practice:
"Patterns of price movements are not random. However, they're close enough to random so that getting some excess, some edge out of it, is not easy and not so obvious, thank God."

Jim Simons

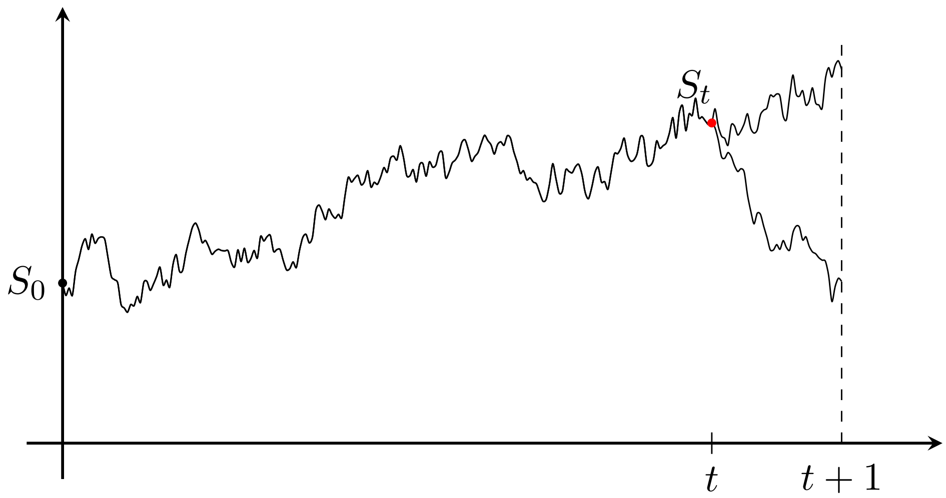
Jim Simons:

- former mathematician, founder of Renaissance Technologies
- net worth: \$12.5 billion

Medallion [Renaissance Fund] has scored average annual returns of about 35%, after fees, since its inception in 1988, with only one money-losing quarter since 1995, a slight 0.5% drop in the first quarter of 1999, the firm has told investors.

`http://on.wsj.com/1hBVRbE`

Patterns: Momentum and Reversals



Is the stock price more likely to increase or decrease after a recent run-up?

Momentum and Reversals

Momentum: price trends in the same direction

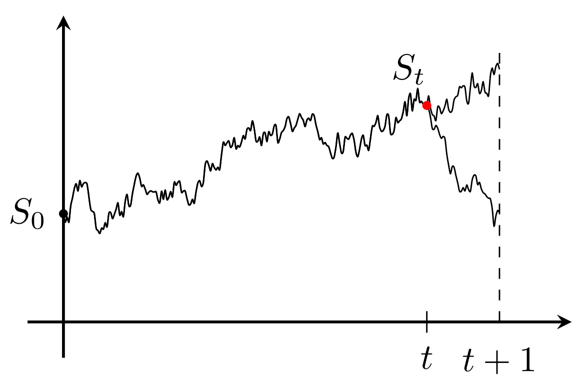
- Information
- Behavioral

Reversal: price trend changes direction

- Over-reaction
- Temporary supply-demand imbalance

Evidence of Momentum or Reversals in the Data?

Predictive Variable: One-Day Lagged Return



Return: $= \dfrac{S_{t+1} - S_t}{S_t}$

Prediction equation: $\underbrace{\text{Return}}_{Y} = \beta_0 + \beta_1 \times \underbrace{(1\text{D})}_{X} + \text{error}$

Returns are predicted using the previous day's return: (1D)

IBM Returns: Regression Model Fit to Training Data

Prediction equation: $\underbrace{\text{Return}}_{Y} = \underbrace{\beta_0 + \beta_1 \times (\text{1D})}_{\text{predicted return}} + \text{error}$

	A	B	C
6			
7	Date	Return	1D
8	7/2/2012	0.12%	2.19%
9	7/3/2012	0.05%	0.12%
10	7/5/2012	-0.33%	0.05%
11	7/6/2012	-1.99%	-0.33%
12	7/9/2012	-0.91%	-1.99%
13	7/10/2012	-1.80%	-0.91%
14	7/11/2012	-0.54%	-1.80%
15	7/12/2012	-1.17%	-0.54%
16	7/13/2012	1.60%	-1.17%
17	7/16/2012	-0.66%	1.60%
18	7/17/2012	-0.62%	-0.66%

What are β_0, β_1 and R^2?

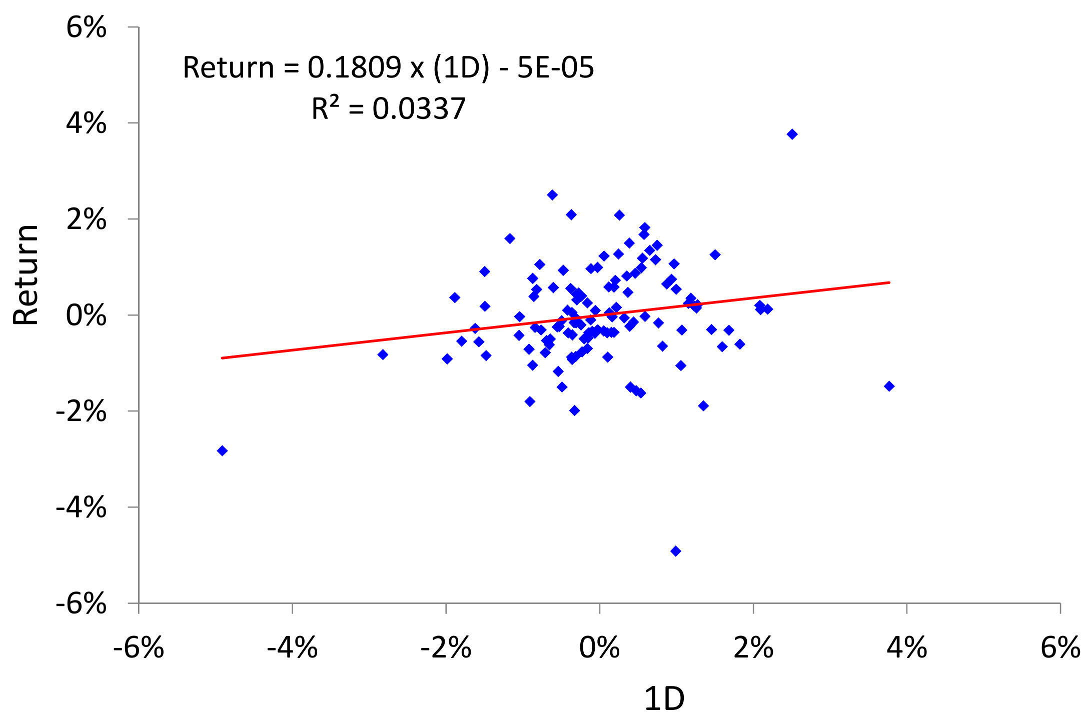
$$\beta_0 = -0.000053$$
$$\beta_1 = 0.180859$$
$$R^2 = 3.4\%$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.1834
R Square	0.0337
Adj R Square	0.0258
Standard Error	0.0104
Observations	125

	Coefficients	*Std Error*	*t Stat*	*P-value*
Intercept	-0.00005	0.00	-0.06	0.95
1D	0.18086	0.09	2.07	0.04

Prediction (Regression) Equation Plot of Training Data



From Predictions to Trading Decisions

1-Variable Regression to Predict Returns and to Trade

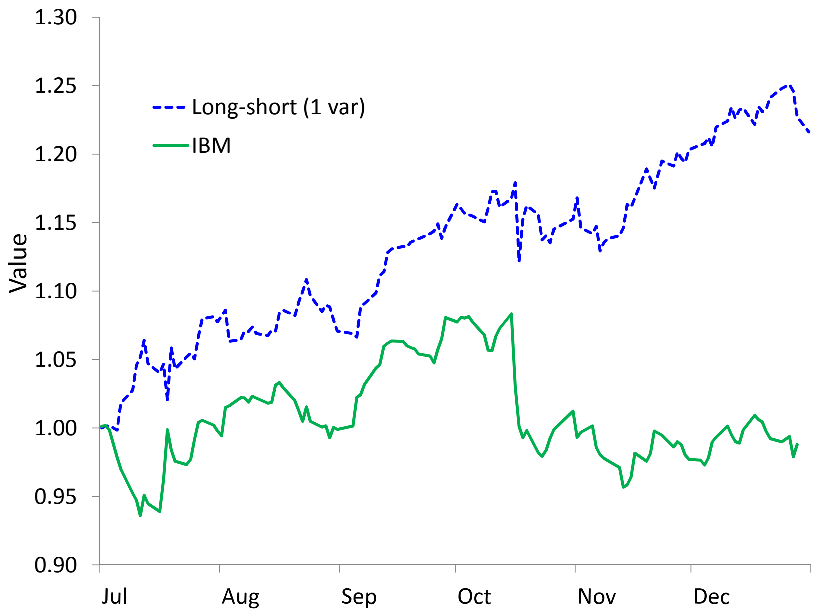
Prediction equation: $\underbrace{\text{Return}}_{Y} = \underbrace{\beta_0 + \beta_1 \times (\text{1D})}_{\text{predicted return}} + \text{error}$

Date	1D	Predicted Return	Trade	Return	Long-short Return	Long-short value
7/2/2012	2.19%	0.39%	Long	0.12%	0.12%	1.0012
7/3/2012	0.12%	0.02%	Long	0.05%	0.05%	1.0018
7/5/2012	0.05%	0.00%	Long	−0.33%	−0.33%	0.9985
7/6/2012	−0.33%	−0.06%	Short	−1.99%	1.99%	1.0183
7/9/2012	−1.99%	−0.36%	Short	−0.91%	0.91%	1.0276
7/10/2012	−0.91%	−0.17%	Short	−1.80%	1.80%	1.0460

What is the long-short fund value at the close on 12/31/2012? 1.216

Does this represent impressive performance? IBM value: 0.988

Training Data: Value versus Time

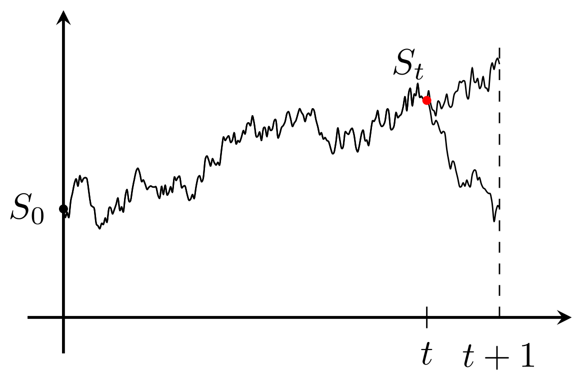


Why does this strategy do so well when the R^2 is so low?

Can We Do Better?

Multiple Regression: Adding Variables

Prediction equation: $Y = \beta_0 + \beta_1(\text{1D}) + \beta_2(\text{3D}) + \beta_3(\text{1W}) + \cdots + \beta_{14}(\text{1Y}) + \text{error}$



- Momentum and reversion might operate on different time scales
- Which time scale predicts best isn't known in advance: let the data tell us
- Try fourteen predictor variables: 1D, 3D, 1W, 2W, 3W, 1M, 6W, 2M, 3M, 4M, 5M, 6M, 9M and 1Y

Estimate β_0, β_1, $\ldots$, β_{14} using 6 months of daily data for IBM

14-Variable Regression Equation

Prediction equation: $Y = \beta_0 + \beta_1(\text{1D}) + \beta_2(\text{3D}) + \beta_3(\text{1W}) + \cdots + \beta_{14}(\text{1Y}) + \text{error}$

SUMMARY OUTPUT

Regression Statistics	
R Square	18.1%
Observations	125

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0058	0.0046	1.2699	0.2068
1D	0.2495	0.1072	2.3265	0.0218
3D	-0.1946	0.2459	-0.7914	0.4304
1W	0.0892	0.3312	0.2694	0.7881
2W	-0.3846	0.4685	-0.8209	0.4135
3W	0.7076	0.7067	1.0014	0.3188
1M	1.2222	0.9033	1.3530	0.1788
6W	-1.0421	1.3014	-0.8007	0.4250
2M	2.4480	1.8490	1.3239	0.1883
3M	0.5060	2.3677	0.2137	0.8312
4M	-5.6408	3.0995	-1.8199	0.0715
5M	-4.0379	4.7791	-0.8449	0.4000
6M	-14.9719	6.5775	-2.2762	0.0248
9M	2.5847	5.2152	0.4956	0.6212
1Y	-9.5631	8.6525	-1.1052	0.2715

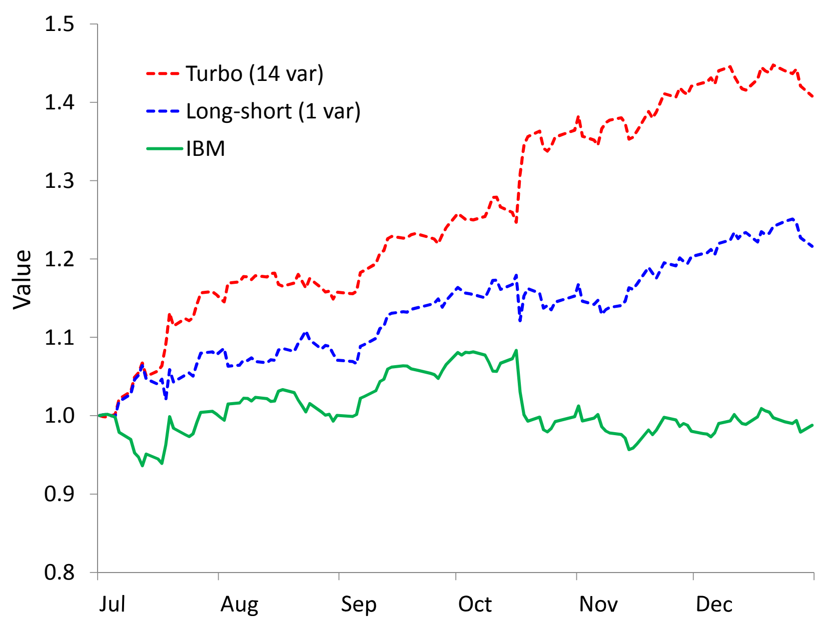
14-Variable Regression to Predict Returns and to Trade

Prediction equation: $Y = \beta_0 + \beta_1(\text{1D}) + \beta_2(\text{3D}) + \beta_3(\text{1W}) + \cdots + \beta_{14}(\text{1Y}) + \text{error}$

Date	Predicted Return	Trade	Return	Turbo Return	Turbo value
7/2/2012	−0.37%	Short	0.12%	−0.12%	0.9988
7/3/2012	−0.47%	Short	0.05%	−0.05%	0.9982
7/5/2012	−0.58%	Short	−0.33%	0.33%	1.0015
7/6/2012	−0.64%	Short	−1.99%	1.99%	1.0214
7/9/2012	−0.92%	Short	−0.91%	0.91%	1.0307
7/10/2012	−0.67%	Short	−1.80%	1.80%	1.0492

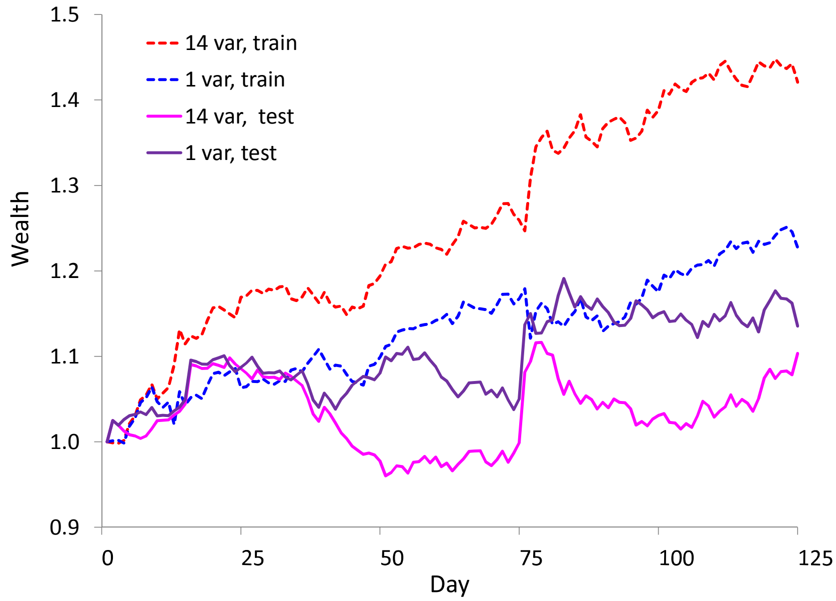
What is the Turbo Long-Short fund value at the close on 12/31/2012? 1.408

Value versus Time: 1 and 14 Predictive Variables



Would you invest in the Turbo Long-Short fund?

Value versus Time: Next Six Months

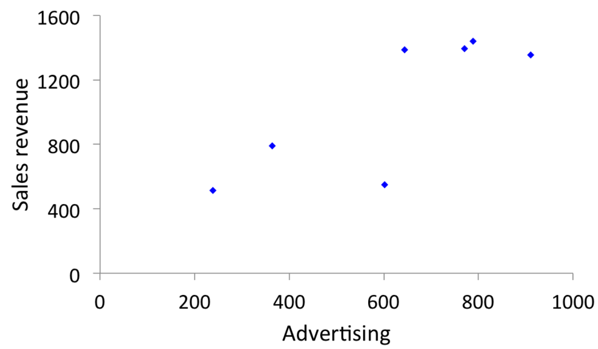


What went wrong?

Performance Evaluation: Train versus Test Sets

Example

Data:



Two possible models:

$y = 1.4234x + 182.78$

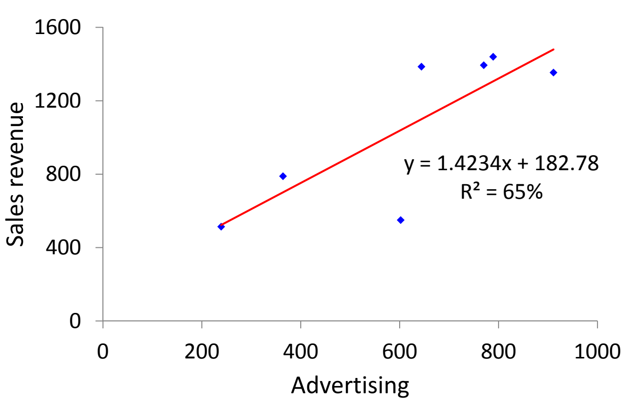
$R^2 = 65\%$

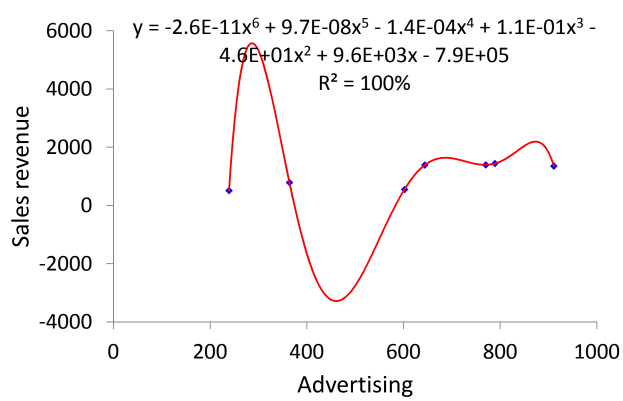
$y = -2.6\text{E-}11\, x^6 + 9.7\text{E-}08\, x^5 - 1.4\text{E-}04\, x^4 + 1.1\text{E-}01\, x^3 - 4.6\text{E+}01\, x^2 + 9.6\text{E+}03\, x - 7.9\text{E+}05$

$R^2 = 100\%$

Which model do you select?

Overfitting: Too Many Predictive Variables





Adding predictive variables gives a better and better fit to the data.

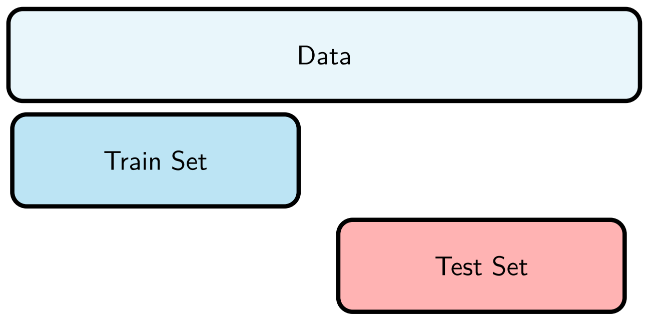
So are more variables better?

More variables may or may not do a better job of predicting out of sample

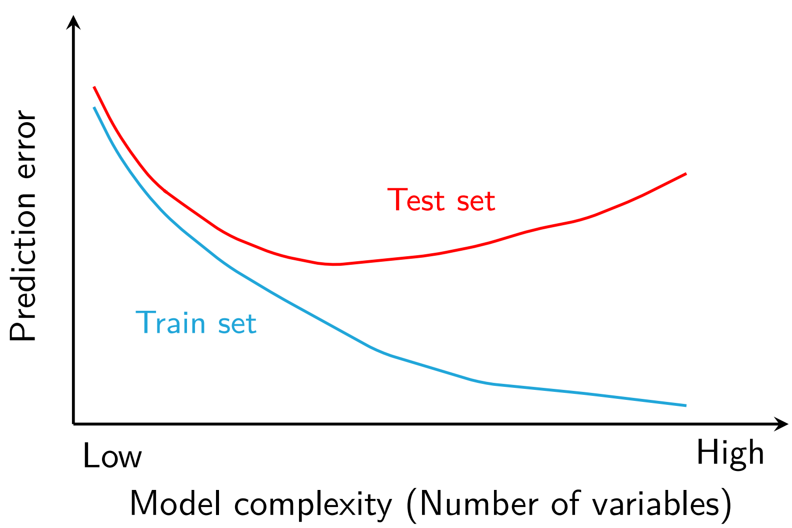
Key Idea: Need to Test Performance "Out of Sample"

Divide data into two sets:

- Train set: use this data to fit the model (I suggest 75%)
- Test set: use this data to assess the quality of the model's predictions (I suggest 25%)



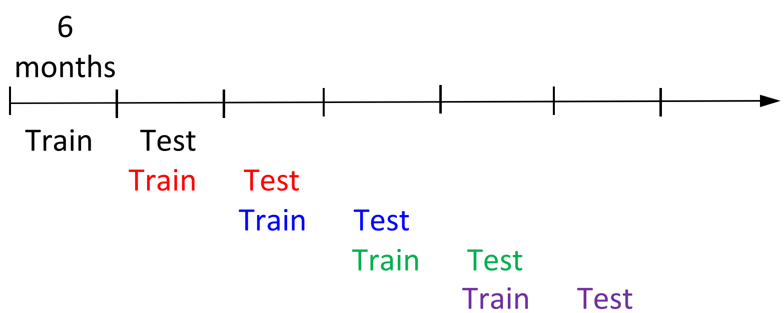
Key idea: Prediction Error versus Model Complexity



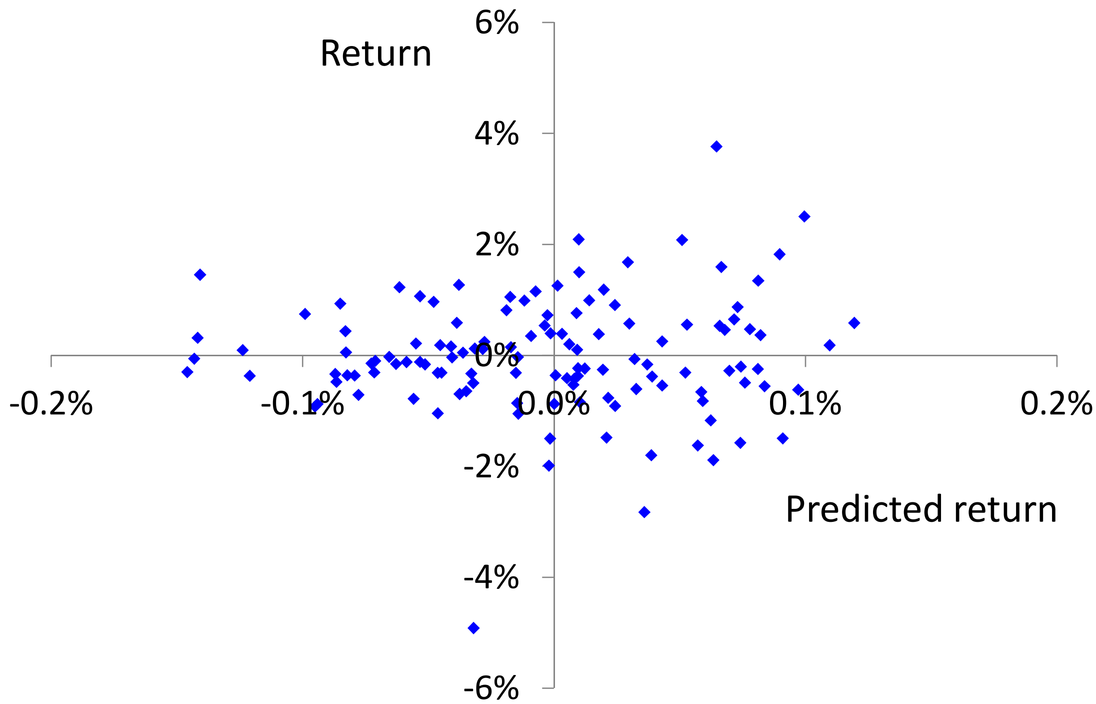
Using prediction error in training data to evaluate models will lead to overly optimistic performance assessments and wrong model choices

Testing the Robustness of Predictions

- To prevent overfitting: use fewer variables
  We'll try: 1D, 1W and 2W (just one example)
- To increase power: use more stocks
  We'll try (approximately) 50 stocks, separate regressions for each stock
- Test out of sample: longer time horizon
  - Training set: Use 6 months of data to fit regression equation
  - Test set: Use previous model for next 6 months of trades
  - Update the regression equations every 6 months and repeat

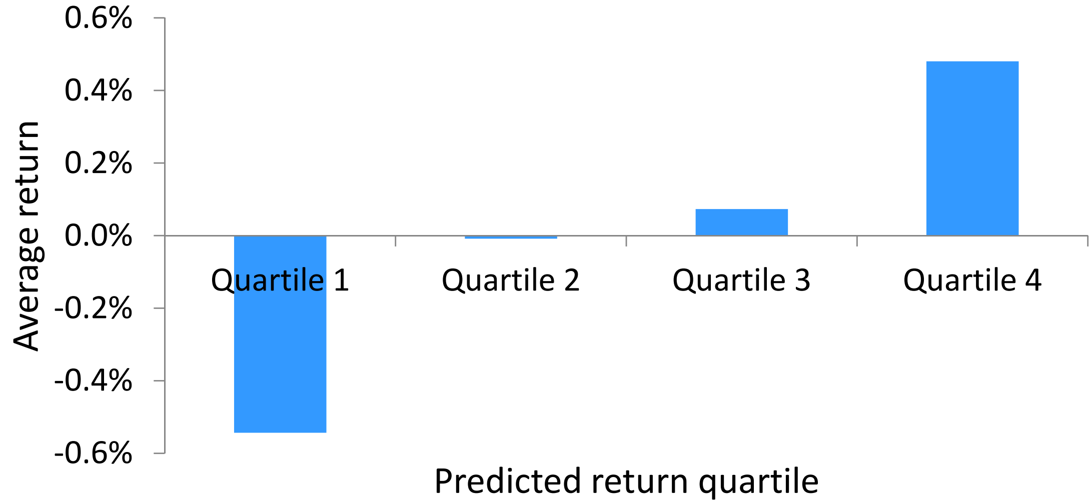


Training Data: Return versus Predicted Return



Does a positive predicted return lead to a positive actual return?

Actual Return by Predicted Return Quartile



Grouping predicted returns shows the potential of the prediction equation

Portfolio Construction Example

Ticker	Predicted return	Ticker	Predicted return
AAPL	-0.94%	JPM	-0.25%
ABT	-0.57%	KO	0.41%
AMGN	-1.25%	MCD	0.01%
AXP	-0.50%	MDT	0.11%
BA	-0.11%	MMM	-0.58%
BAC	-0.59%	MO	0.15%
BMY	-0.06%	MSFT	-1.16%
C	-0.88%	OXY	-0.62%
CL	-0.40%	PEP	0.09%
COP	0.61%	PFE	-0.91%
CVS	-0.18%	PG	-0.52%
CVX	-0.86%	SLB	0.56%
DIS	-0.23%	T	0.23%
GE	-0.14%	UTX	0.59%
HD	-0.43%	VZ	0.59%
HPQ	-0.83%	WFC	0.16%
IBM	-0.42%	WMT	-0.36%
JNJ	-0.25%	XOM	0.44%

What trades to make based on these predicted returns on one representative day?

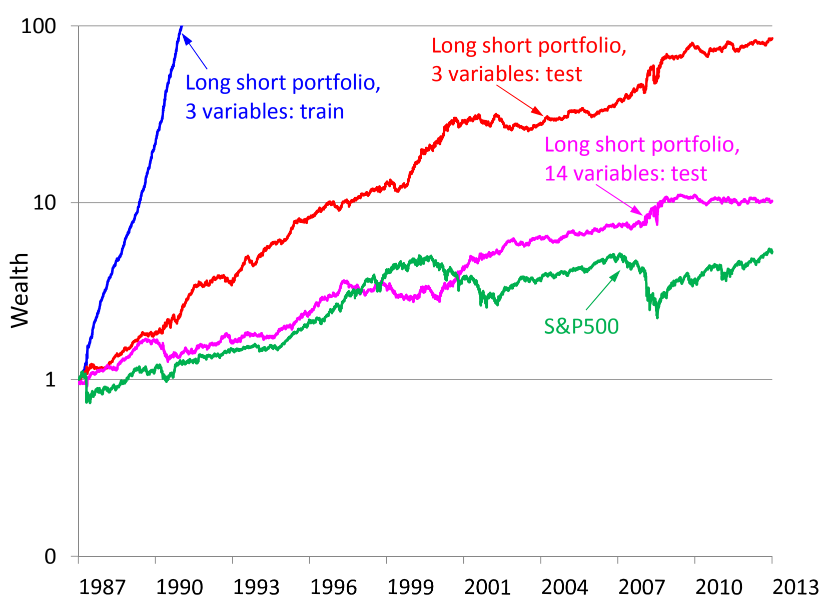
Portfolio Construction

Ticker	Predicted return	Ticker	Predicted return
COP	0.61%	JNJ	-0.25%
VZ	0.59%	WMT	-0.36%
UTX	0.59%	CL	-0.40%
SLB	0.56%	IBM	-0.42%
XOM	0.44%	HD	-0.43%
KO	0.41%	AXP	-0.50%
T	0.23%	PG	-0.52%
WFC	0.16%	ABT	-0.57%
MO	0.15%	MMM	-0.58%
MDT	0.11%	BAC	-0.59%
PEP	0.09%	OXY	-0.62%
MCD	0.01%	HPQ	-0.83%
BMY	-0.06%	CVX	-0.86%
BA	-0.11%	C	-0.88%
GE	-0.14%	PFE	-0.91%
CVS	-0.18%	AAPL	-0.94%
DIS	-0.23%	MSFT	-1.16%
JPM	-0.25%	AMGN	-1.25%

One approach to constructing a portfolio

- Sort the stocks in order of decreasing predicted returns
- Neutral portfolio: Buy (go long) the top 5 stocks, sell (short) the bottom 5 stocks
- What is the portfolio return the next day?
- Portfolio return: average of the top 5 actual returns minus the average of the bottom 5 actual returns

Training vs. Test: Stock Portfolio Results



Results from 1987-2013

- Long short portfolio, 3 variables, test: Avg return: 18%, Std dev: 13%
- Long short portfolio, 14 variables, test: Avg return: 10%, Std dev: 13%
- S&P500: Avg return: 9%, Std dev: 19%

Quantitative Strategies Today

> *One of the datasets we purchased is cloud cover data. It turns out that when it's cloudy in Paris, the French market is less likely to go up than when it's sunny in Paris. That's true in Milan, it's true in Tokyo ... It's just true.*
>
> *Now, you can't make a lot of money from that data because it's only slightly more likely to go up. But it is statistically significant ...*
>
> *The idea is that any one these [weak signals] or any handful of these wouldn't be enough to overcome transaction costs, but if you combined lots and lots and lots of them together, then you can overcome transaction costs and you have something. So it's a lot of weak signals like this cloud cover data, but that's the only one I'm going to disclose no matter how hard you push me.*
>
> Peter Brown, Renaissance Technologies `http://cs.jhu.edu/~post/bitext/`

New frontiers of quantitative strategies

- High frequency trading (HFT)
- Natural language processing (NLP) to process news

Wrap-Up

Analytics is central to modern finance

- Analyzed stock predictions and quantitative strategies
  - Evidence that there were market inefficiencies
  - Challenges in practice: transactions costs and price impact
- Will return to portfolio optimization later in the course

Performance evaluation of predictions

- Estimated performance can be overly optimistic!
- Need to separate training and testing
  - training set for fitting and hold-out sample for testing
- Even if one separates the two, there is still a risk of customizing the predictions too much for the test (hold-out) set
- Next time we discuss model selection and assessment formally