Applying Machine Learning Techniques to Everyday Strategies

Ernie Chan, Ph.D.
QTS Capital Management, LLC.

About Me

- Previously, researcher at IBM T. J. Watson Lab in machine learning, researcher/trader for Morgan Stanley, Credit Suisse, and various hedge funds.
- Managing Member of QTS Capital Management, LLC., a commodity pool operator and trading advisor.
- Author:
  - *Machine Trading: Deploying Computer Algorithms to Conquer the Markets* (Wiley 2017).
  - *Algorithmic Trading: Winning Strategies and Their Rationale* (Wiley 2013).
  - *Quantitative Trading: How to Build Your Own Algorithmic Trading Business* (Wiley 2009).
- Blogger: epchan.blogspot.com

Limitations of Financial Data

- Machine learning depends on training data with stable statistics
  - E.g. Facial recognition, speech recognition, playing GO.
- Financial data is anything but stable.
  - Regime changes regularly.
  - Anomalies disappearing due to increasing arbitrage activities.
  - In contrast, cats don’t change their faces because computers start to recognize them on YouTube.

Limitations of Financial Data

- Regime changes render older data unsuitable for testing new strategies.
  - 2001 Decimalization of stock prices.
  - 2008 Financial crisis and start of quantitative easing.
- Only about 2,000 rows of daily data since 2008.
  - Google (Le *et al*, 2012) used 10 million YouTube videos to train neural network to recognize cats' faces.

Limitations of Financial Data

- Seasonality further limits size of data sets.
  - Options expirations strategies rely on weekly or monthly data.
  - Earnings strategies rely on quarterly data.
  - Seasonal physical commodity futures strategies rely on annual cycles.

Overcoming Data Scarcity

- Use **high frequency** data
  - Arbitrage opportunities depend on time scale.
  - Inapplicable to strategies for longer time scales.
  - There is seasonality even intraday.
- **Aggregate** data from multiple instruments
  - E.g. run same trading model on all Russell 3,000 stocks.
  - 2,000 rows x 3,000 columns = 6 million data points.

Reducing Overfitting

- **Bagging**
  - Over-sampling existing data to create more rows.
- **Random Subspace**
  - Under-sampling existing predictors to limit overfitting.
- **Stepwise Regression**
  - Sequentially adding predictors, then sequentially removing predictors.
- **Random Forest**
  - Combining bagging and random subspace.

Example: Factor Model on SPX Stocks

- Predict quarterly returns on stocks using simple linear model:

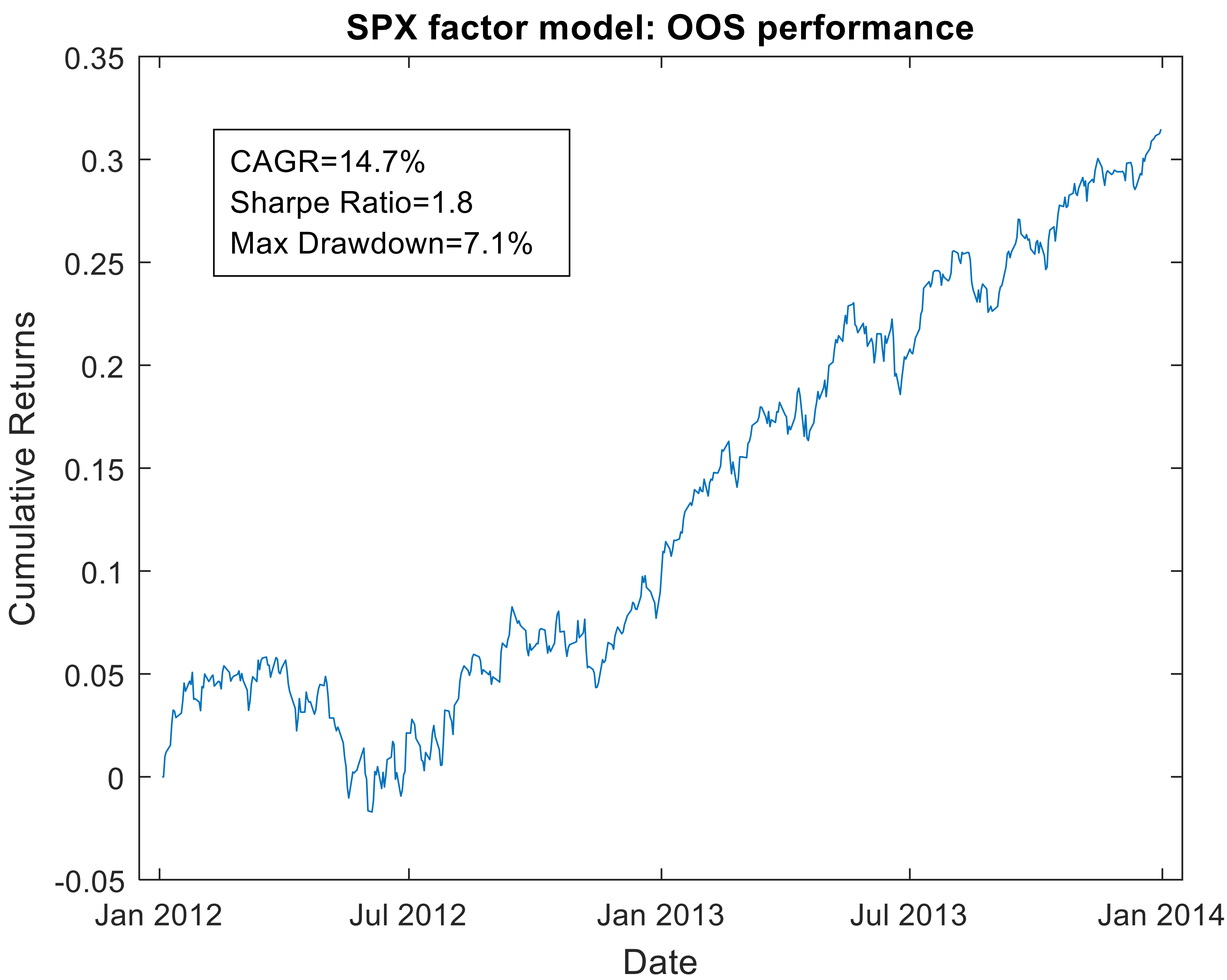
$$Return(t+1,s) = \alpha + \beta_1(t,s) * Factor_1 + \beta_2(t,s) * Factor_2 + \cdots + \varepsilon(\mathrm{t},\mathrm{s})$$

- Use fundamental factor loadings $\beta_{\mathrm{i}}(t,s)$ extracted from quarterly financial statements as predictors.
- Restrict ourselves to factor loadings that do *not* scale with a firm’s size.
  - There are about 27 such factor loadings.
  - Source: Sharadar’s Core US Fundamentals database via Quandl.com.

Variable name	**Explanation**	**Period**
CURRENTRATIO		Quarterly
DE	Debt to Equity Ratio	Quarterly
DILUTIONRATIO	Share Dilution Ratio	Quarterly
PB	Price to Book Value	Quarterly
TBVPS	Tangible Asset Book Value per Share	Quarterly
ASSETTURNOVER		Trailing 1 Year
EBITDAMARGIN		Trailing 1 Year
EPSGROWTH1YR		Trailing 1 Year
EQUITYAVG	Average Equity	Trailing 1 Year
EVEBIT	Enterprise Value over EBIT	Trailing 1 Year
EVEBITDA	Enterprise Value over EBITDA	Trailing 1 Year
GROSSMARGIN		Trailing 1 Year
INTERESTBURDEN	Financial Leverage	Trailing 1 Year
LEVERAGERATIO		Trailing 1 Year
NCFOGROWTH1YR		Trailing 1 Year
NETINCGROWTH1YR	Net Income Growth	Trailing 1 Year
NETMARGIN	Profit Margin	Trailing 1 Year
PAYOUTRATIO		Trailing 1 Year
PE	Price Earnings Damodaran Method	Trailing 1 Year
PE1		Trailing 1 Year
PS		Trailing 1 Year
PS1	Price Sales Damodaran Method	Trailing 1 Year
REVENUEGROWTH1YR		Trailing 1 Year
ROA		Trailing 1 Year
ROE		Trailing 1 Year
ROS		Trailing 1 Year
TAXEFFICIENCY		Trailing 1 Year

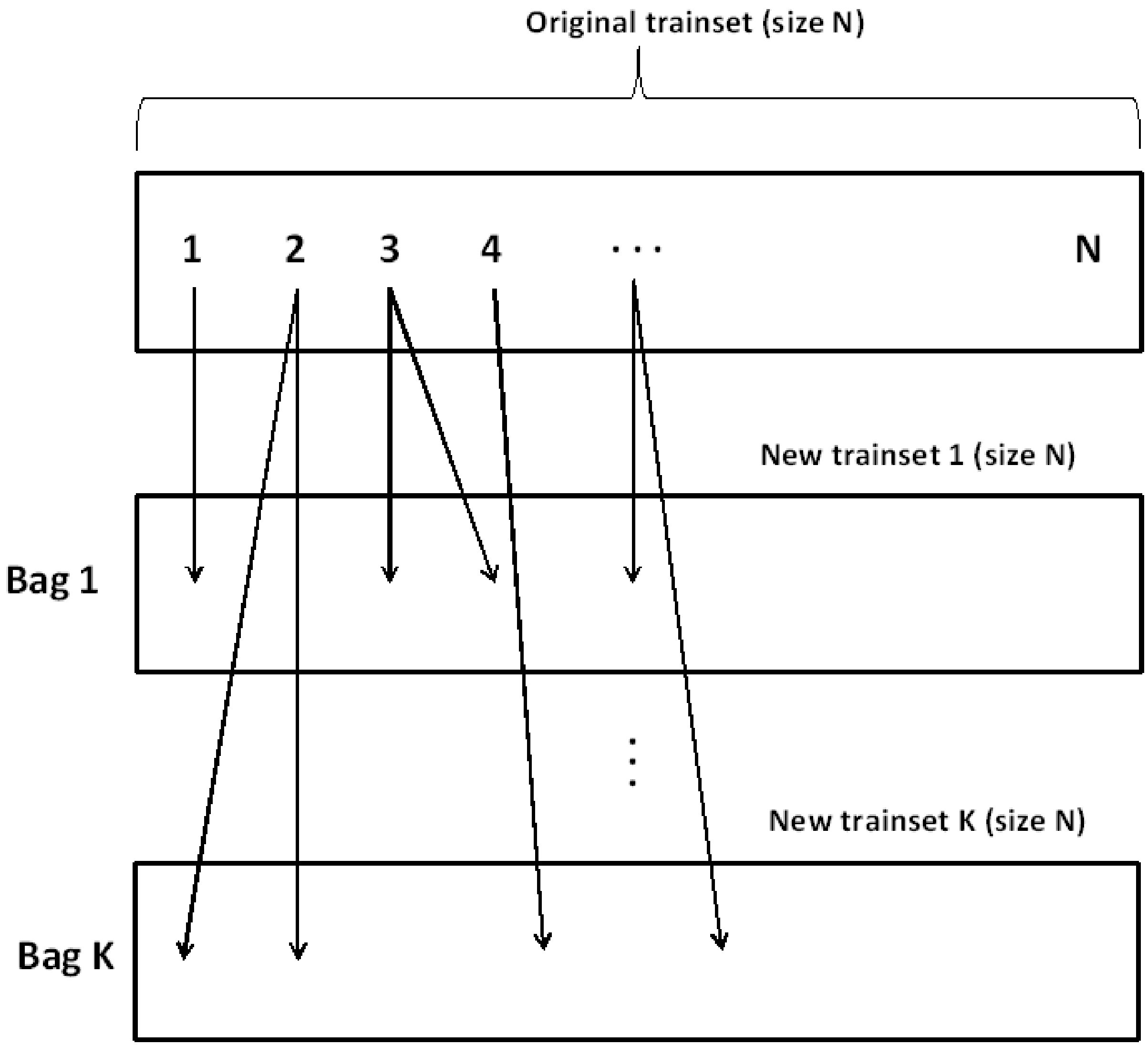
Factor Model on SPX Stocks

- $Factor_{\mathrm{i}}$ are the regression coefficients: assumed fixed across all *stocks* and all *time*.
  - **Aggregation** in action!
  - Training data: 200701-201112.
  - 1,260x 500 data points (instead of just 1,260).
- Trading strategy: At end of each day
  - Buy if predicted return > 0, *vice versa* for short.
  - Hold for a quarter.


SPX factor model: OOS performance
CAGR=14.7%
Sharpe Ratio=1.8
Max Drawdown=7.1%
Cumulative Returns
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
Jan 2012
Jul 2012
Jan 2013
Jul 2013
Jan 2014
Date

Bagging

- Increasing the training set (size *N*) by over-sampling data.
  - i.e. Resampling with replacement.
- Re-sample *N* data points to become *K* bags of *N* data points.
  - Total: *K* x *N* data points.
- Train separate model for each bag.
- Take average predicted returns of *K* models.


Original trainset (size N)
1
2
3
4
. . .
N
New trainset 1 (size N)
Bag 1
New trainset K (size N)
Bag K

Random Subspace

- Randomly select subset of predictors to train *K* models.
- Similar to bagging, take average predicted returns of *K* models.

Random Forest

- Combine both bagging and random subspace:
  – Over-sample data
  – Under-sample* (in our case) predictors

*In other applications such as classification and regression trees, we can over-sample and re-use predictors too (*i.e.* sampling with replacement)

Out-of-Sample Results

	CAGR	Sharpe Ratio	Calmar Ratio
Base model (aggregated)	14.7%	1.8	2.1
Bagging (K=100)	15.1%	1.8	2.1
Random Forest (K=100, 14 predictors)	16.7%	1.7	2.1

Interpretation

- Glass half empty:
  - Random forest does not improve performance significantly.
- Glass half full:
  - Random forest shows that original performance is robust with respect to re-sampling.
  - I.e. original results are **statistically significant**!

Stepwise Regression

- Random subspace/forest randomly picking predictors.
- Stepwise regression picked them step-by-step based on BIC: essentially maximizing log-likelihood while penalizing number of variables.
  - BIC is proportional to negative log likelihood.
- Stop adding variables when BIC is minimized, then start deleting them until BIC increases.

Out-of-Sample Results

	CAGR	Sharpe Ratio	Calmar Ratio
Base model (aggregated)	14.7%	1.8	2.1
Bagging (K=100)	15.1%	1.8	2.1
Random Forest (K=100, 14 predictors)	16.7%	1.7	2.1
Stepwise Regression	19.1%	1.8	2.2

Stepwise Regression

- Just 2 variables generate all the predictive power of the factor model.
  - Gross margin (trailing 1 year)
  - Price-to-Sales (trailing 1 year)

- Similar result may be generated by “L1 regularization” (LASSO regression)?

Market Neutral Version

- Is good return due to a net long exposure during the bull market?
- Modified trading strategy:
  - Buy 50 stocks with the top predicted returns
  - Short 50 stocks with the bottom predicted returns
  - Hold for 1 quarter.
- Out-of-sample: CAGR=5.54%, Sharpe=1.4, Calmar=1.4.
- Model has real alpha!

Conclusion

- Aggregating data across time or instruments usually a good idea.
- Selectively reduction of variables produce slightly better results than oversampling training data.
- Reduction of variables produces a more parsimonious model with more intuitive meaning.

Thank you for your time!

www.epchan.com

Twitter: @chanep

Blog: epchan.blogspot.com