MS&E 448
Statistical Arbitrage

—

Group 5 2017:
Carolyn Soo, Zhengyi Lian, Jiayu Lou, Hang Yang

The Problem

- Stock returns: systematic and idiosyncratic components

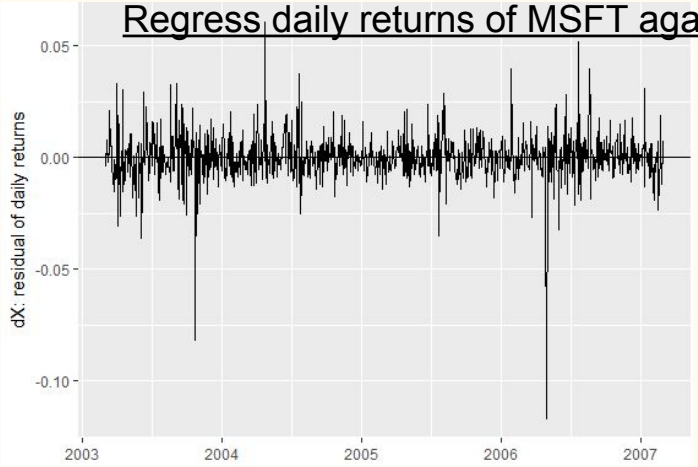
$$\frac{dP_t}{P_t} = \alpha dt + \sum_{j=1}^{n} \beta_j F_t^{(j)} + dX_t$$

- Idiosyncratic residual (dX) and spread ($X = X_0 + \int dX$)

How to systematically find stationary spreads X(t) and profit from their mean reversion?

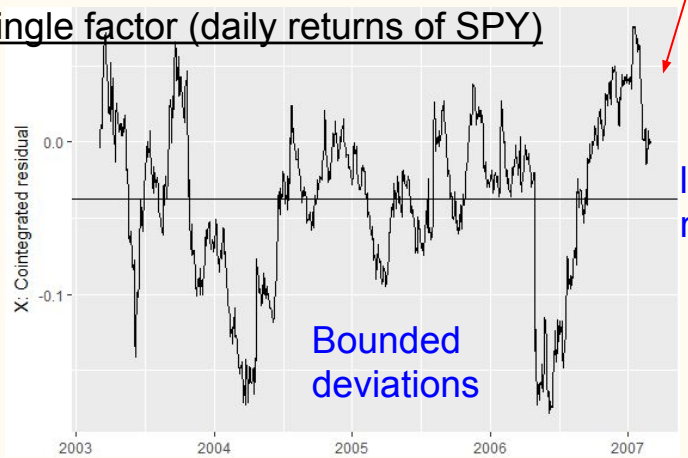
Regress daily returns of MSFT against a single factor (daily returns of SPY)

dX



time

X



time

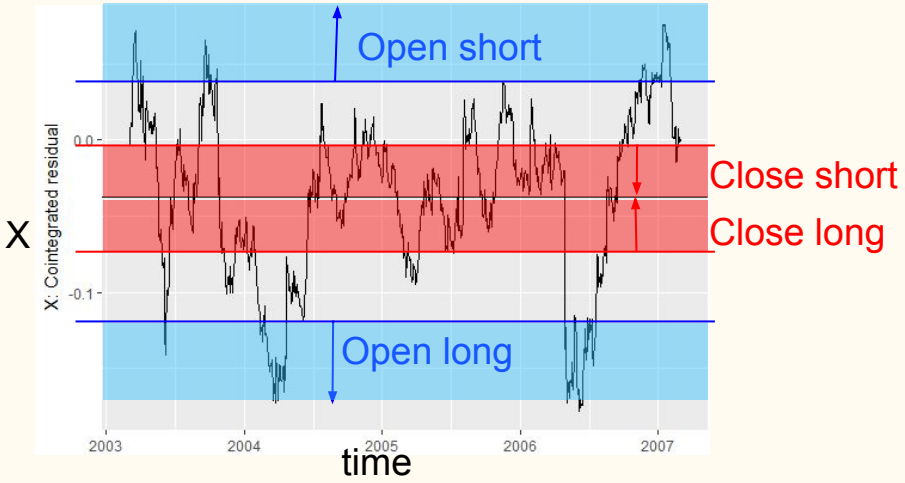
Foundational Assumptions and Notes

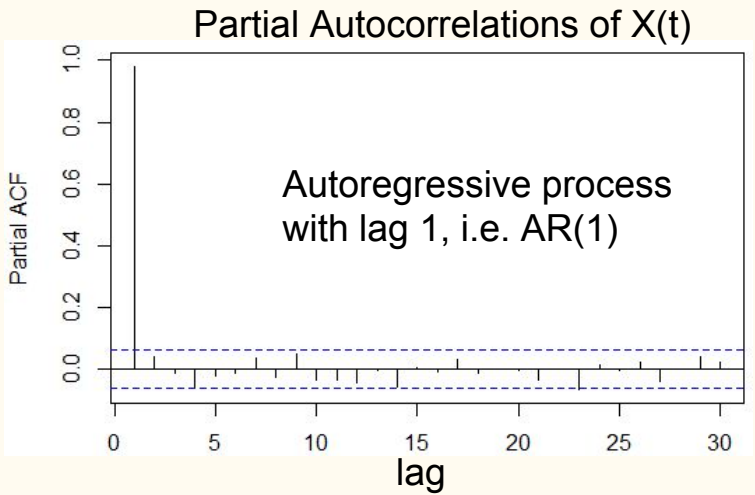
The spread between what the stock price is - and what it should be - is driven by its idiosyncratic returns. So if we can study the behavior of these idiosyncratic returns, we can also **identify when and how the spread reverts** to its equilibrium... and **profit when the gap closes!**

- Compute before market open, using trailing window of data
- Execute trades daily, at market open, at market price, iff mean reversion “rate” is “fast enough” *and* signal cutoffs met
- Trading US stocks and ETFs only

The Intuition - Statistical Arbitrage

- Profit from mean reversion of many independent spread processes X(t)
- while remaining market neutral:
  - Open long: for every $1 long in stock, short $β of factor (e.g. market ETF)
  - Open short: for every $1 short in stock, long $β of factor (e.g. market ETF)





Model - Concept

- **X(t)** for each stock modelled as an **Ornstein-Uhlembeck (OU) process**

$$dX(t) = \kappa\Big(m - X(t)\Big)dt + \sigma dW(t)$$

W_t : Wiener process

- Assuming κ, m, σ for each stock stays constant over a 60-day trailing window,

$$X(t+\Delta t) = a + bX(t) + \zeta$$

- where

$$\left.\begin{aligned} a &= m\left(1 - e^{-\kappa\,\Delta t}\right) \\ b &= e^{-\kappa\,\Delta t} \\ \zeta &\sim N\left(0, \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta t}\right)\right) \end{aligned}\right\} \quad \left.\begin{aligned} \kappa &= -ln(b)/\Delta t \\ m &= \frac{a}{1-b} \\ \sigma &= \sqrt{\frac{\text{Variance}(\zeta)\cdot 2\kappa}{1-b^2}} \end{aligned}\right.$$

In equilibrium
*(*i.e. as $\Delta t \to \infty$*),*

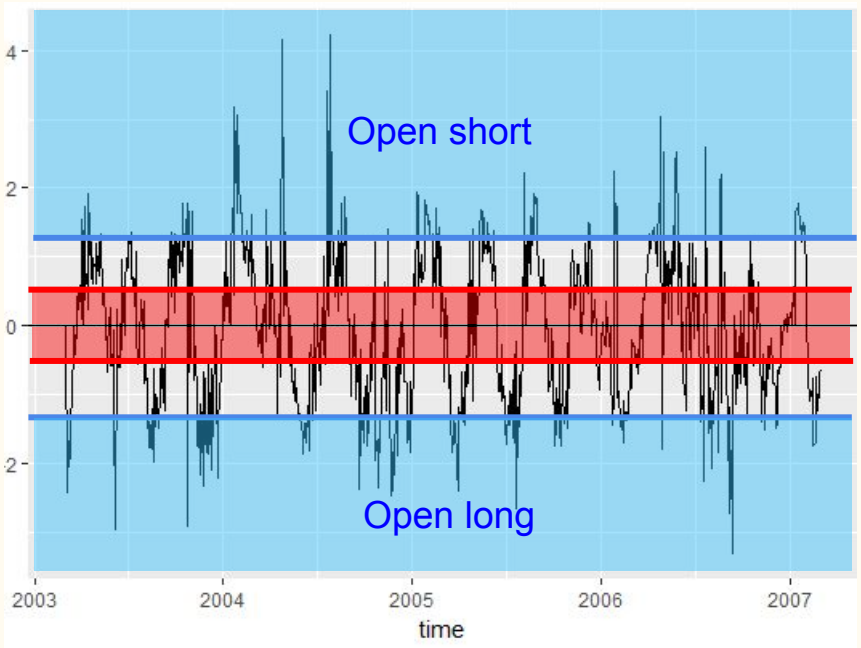
$$\mathbb{E}(X) = m$$

$$\mathbb{V}(X) = \sigma_{eq} = \frac{\sigma^2}{2\kappa}$$

Reference: Avellaneda & Lee, 2008

Model - Concept

- Signal: $$S = \frac{X - \mathbb{E}(X)}{\mathbb{V}(X)} = \frac{X - m}{\sigma_{eq}} = \frac{X - m}{\sigma/\sqrt{2\kappa}}$$



Model Adjustments

60 day trailing window to estimate *X* and parameters (*κ, m, σ*)

- reject if mean reversion time constant = 1/*κ* > 30 days
- adjust for bias in *m*

$$bias(\hat{m}) \approx \frac{1}{N_s}\sum_{s=1}^{N_s}\hat{m}_s$$

$$\hat{m}_{adj} = \hat{m} - bias(\hat{m})$$

- volume adjusted returns to reduce signal on high trading volumes.

Reference: Avellaneda & Lee, 2008

Weights

- **Main Idea: ETFs will eventually cancel out**
  - Treat as a stock portfolio; only consider the weights of stocks
- **Limit the maximum number of long stocks and short stocks to 20 and 20**
  - Invest cash in long stock
  - Set aside the money obtained from short-sell for future buying back to avoid excessive leverage
  - Choice of maximum limit: too large (inefficient use of capital); too small(lack of diversification)
- **Invest equally across the stocks**
  - Each long/short stock accounts for around $\pm 5\%$ of the total portfolio value
  - When executing buying/short-selling decisions, dynamically scale up the value of individual stock to match the updated portfolio value

Trading Steps

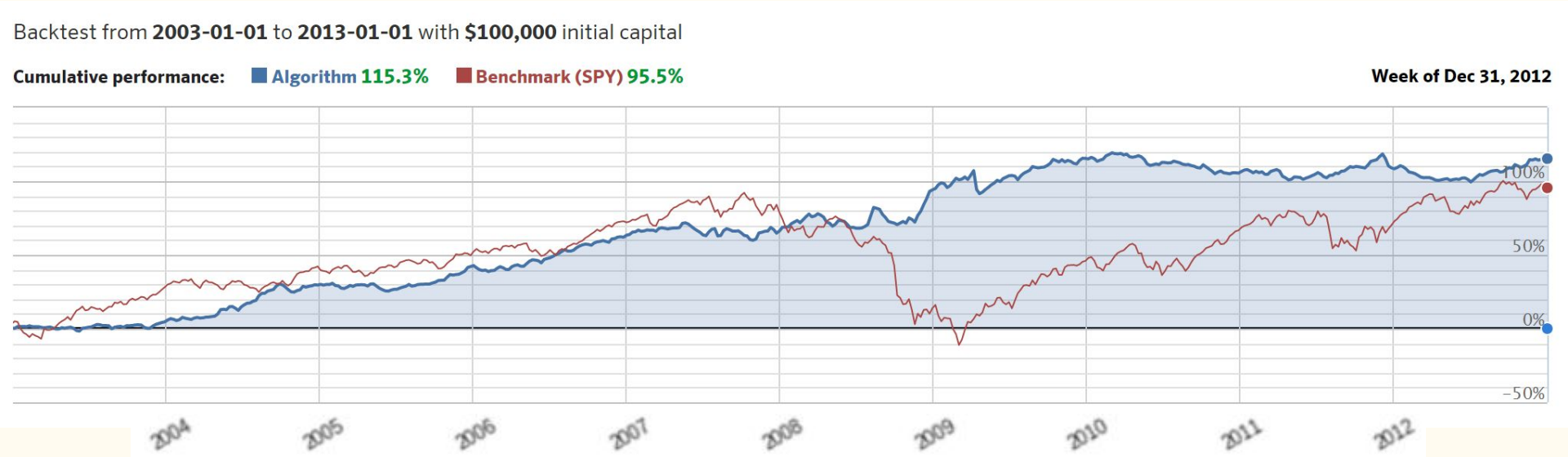
- **Before trading start: parameter estimation**
- $$R_n^S = \beta_0 + \beta R_n^I + \epsilon_n, \quad n = 1, 2, ..., 60.$$
- **Compute signals based on last market close prices**
- **First examine all the close signals and close the long/short positions**
  - Make room for potential open opportunities
- **Compute open positions**
  - Store the relationship of {stock: [list of (ETF, ETF share)]} in a map
  - In the sequential order until the maximum limit is reached
  - Might miss some good opportunities because of lack of slots
- **After the market opens: execute the orders with market price**

Preliminary Results

- Properties:
  - Training Period: 2003-2012
  - Holding Out Period: 2013 and forward
  - Universe: top 100 stocks in market cap at the beginning of 2003 and 1 ETF (SPY)
- Results for Training Period:
  - Net Gain: 115% ($100,000)
  - Annual Return: 8%
  - Annual Volatility: 0.08
  - Shape Ratio: 0.99

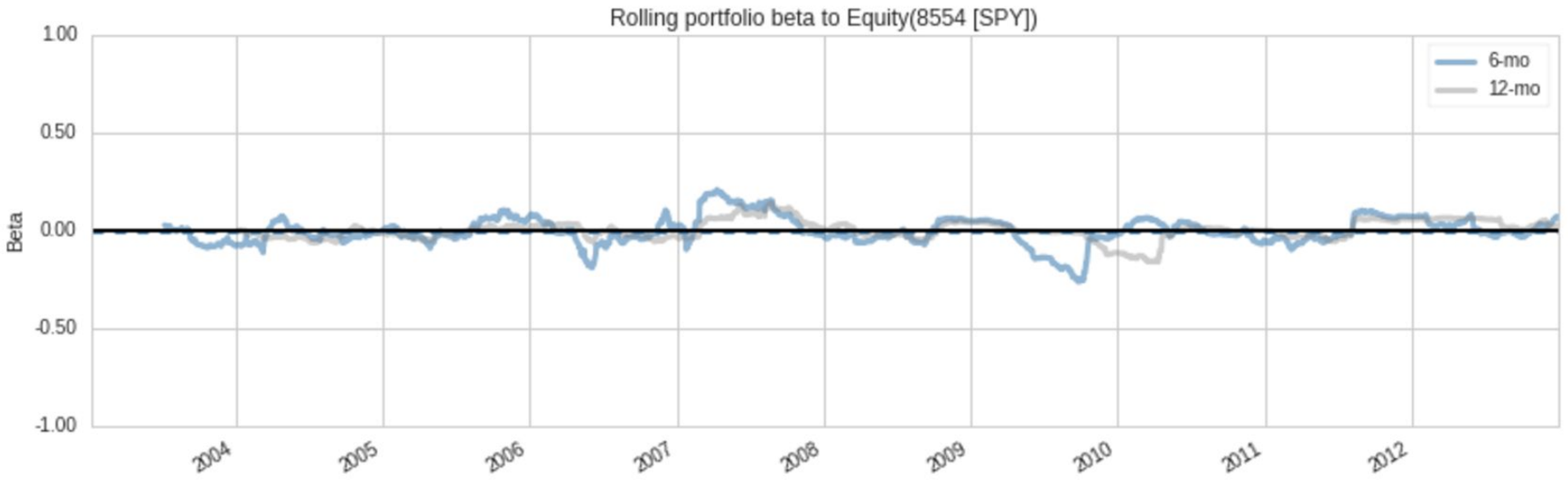
Preliminary Results

Graph - Portfolio Net Gain from 2003 to 2012



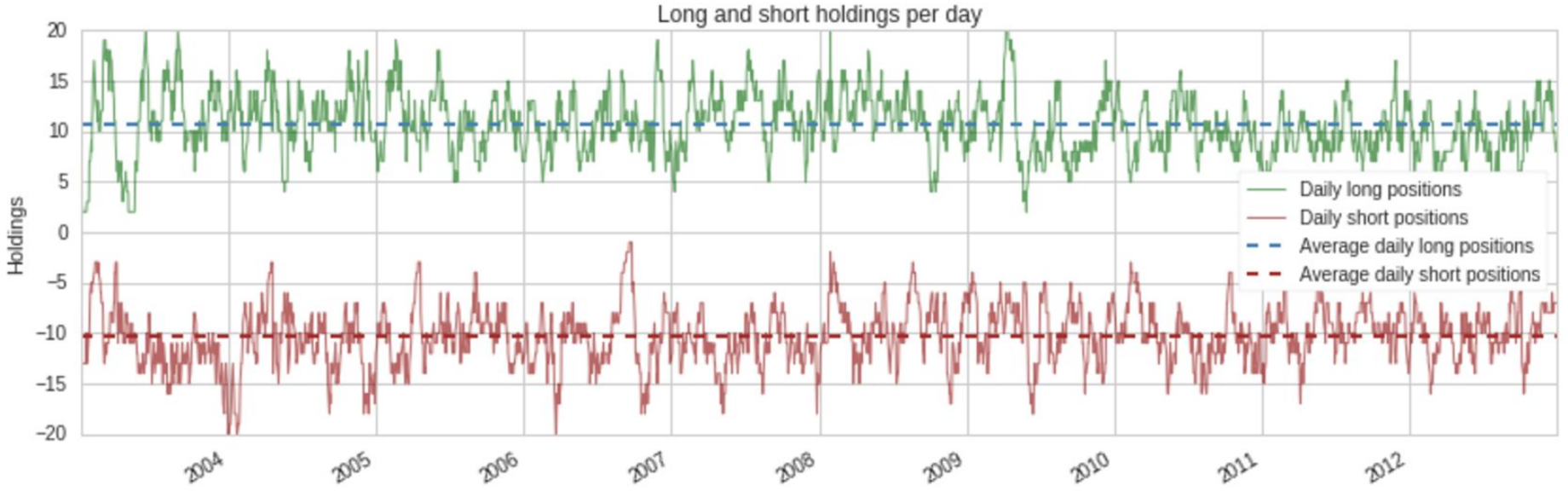
Preliminary Results

Graph - Rolling Portfolio Beta from 2003 to 2012



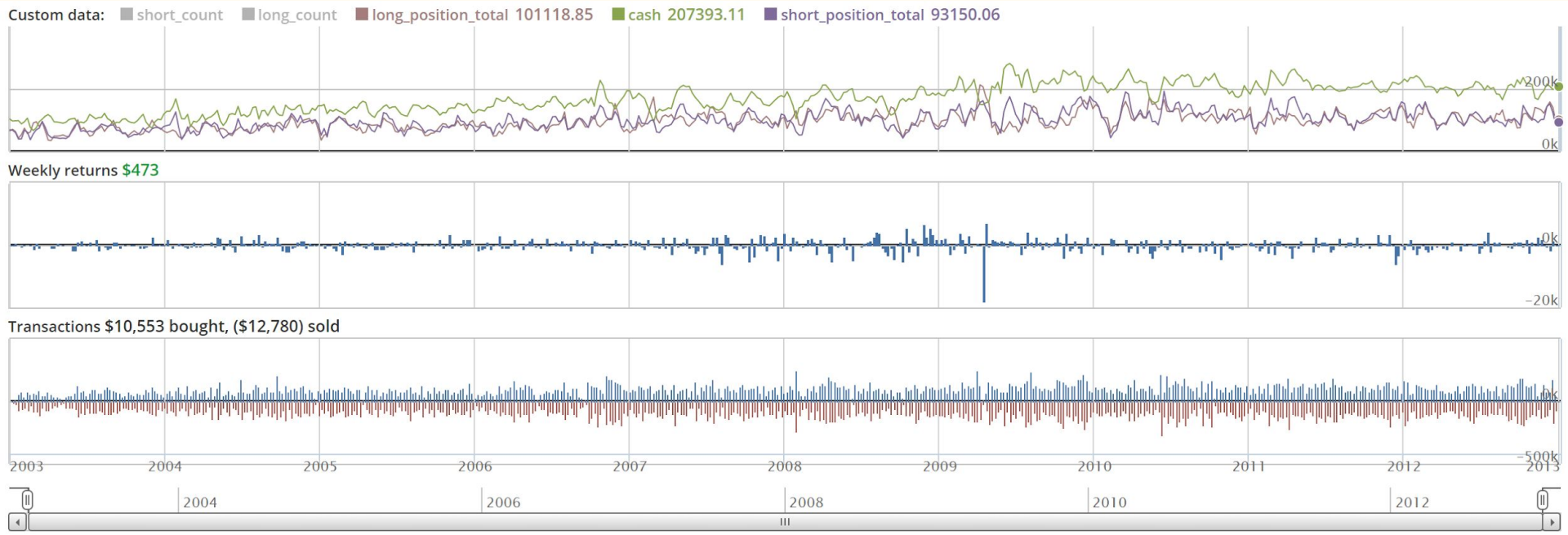
Preliminary Results

Graph - Long and Short Holdings from 2003 to 2012



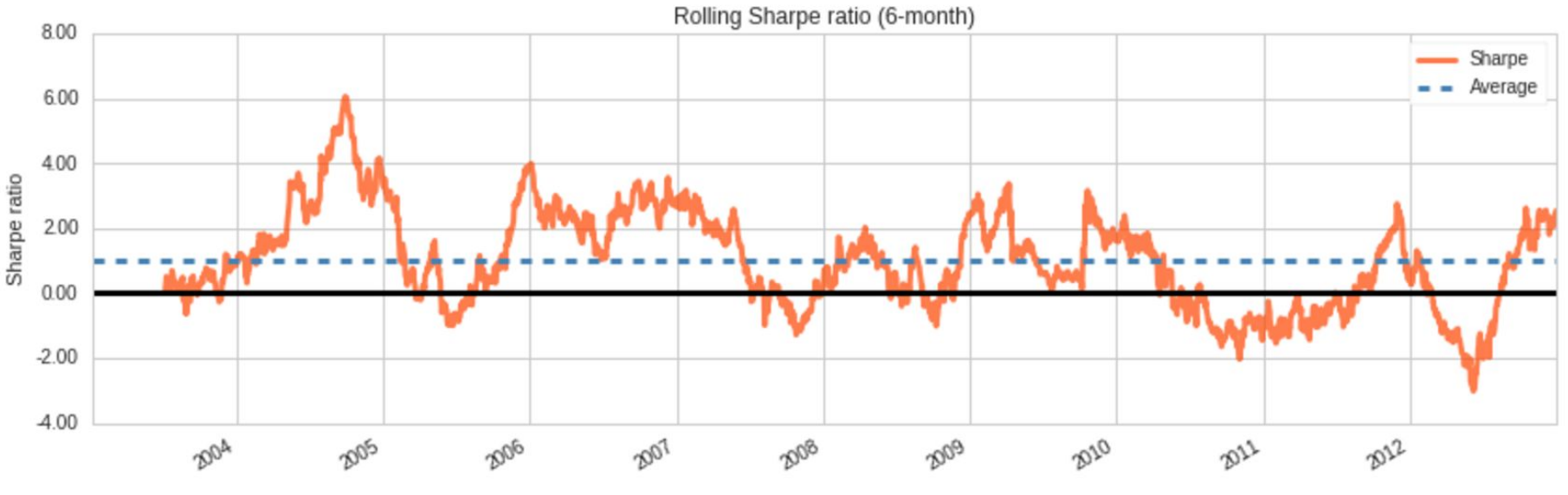
Preliminary Results

Graph - Positions, Weekly Returns, and Transactions from 2003 to 2012



Preliminary Results

Graph - Rolling Sharpe Ratio from 2003 to 2012



Complications

- Not truly market neutral: Orders are sometimes not completely filled
  - Not enough volume available on the market
  - Messes up the goal of beta neutrality
  - Happens during trading time, so not fixable in the calculation step
  - Still working on better fixes, reduced starting cash as a temporary solution

- When too many opportunities are present, we don't always choose the best ones
  - Arbitrary limit in total number of long and short stocks
  - Sequential selection of stocks according to signals received, instead of ranking by signal strengths

Caveats

- We used default signal cutoffs
  - Open Long: $<$-1.25, Open Short: $>$1.25, Close Long: $>$-0.5, Close Short: $<$0.5
  - Original paper only verifies these for pre-2008
- We haven't optimized our weights
  - Currently using about equal weights

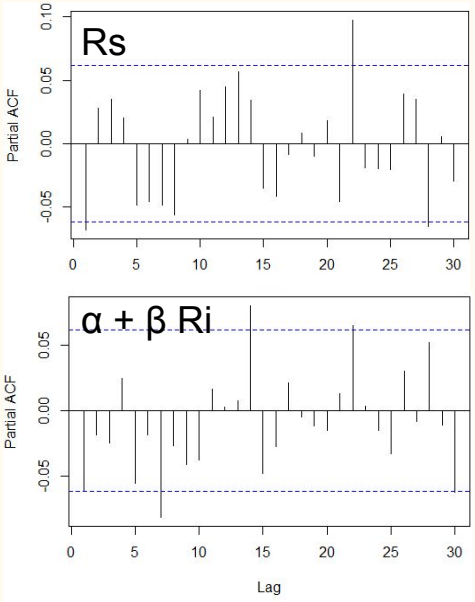
Next Steps

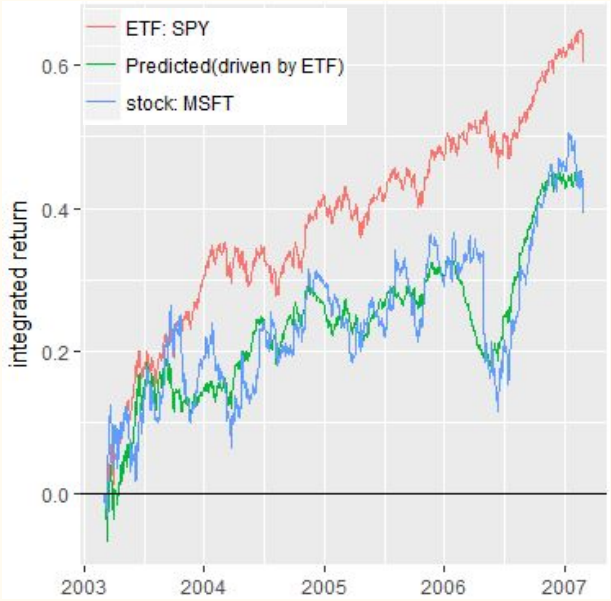
- Weight optimization
- Mapping stocks to ETFs to decide what regressors are appropriate
- Lack of opportunities (when market stagnates, not enough fluctuations)
  - ★ Dynamic signal cutoffs?
  - ★ Expand universe? (right now is 2003 SPY's 100 tickers)
- Robustness and sensitivity tests? E.g. Sharpe vs signal cutoff
- Walk-forward validation?
- How to manage sudden changes in trading environment (e.g. beta)?

Q + A

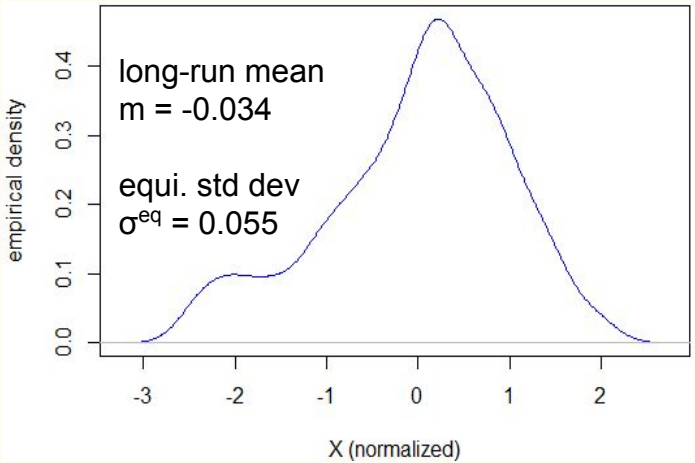
Extra Stuff

Link to Paper (View Only): https://www.overleaf.com/read/zbmtcjwmbvpp

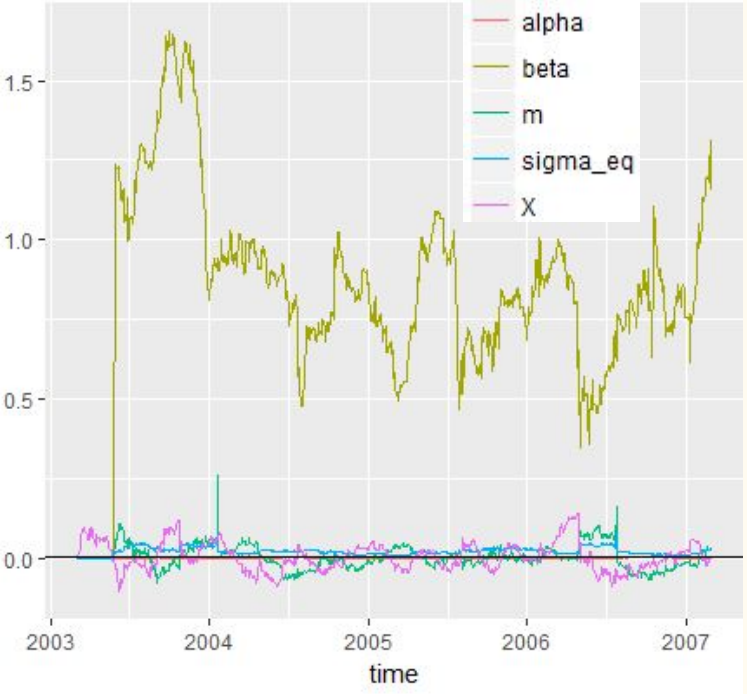




Distribution of Normalized X = (X - m)/σ^{eq}



Verifying the relative insignificance of the drift term



Model - Concept

Entry/Exit rules: Use signal cutoffs to determine when to either enter / exit a long position ("buy to open", denoted “bo” / "sell to close" “sc”resp.) or enter / exit a short position (“so” / “bc” resp.):

$$\overline{s_{bo}} = \overline{s_{so}} = 1.25$$
$$\overline{s_{bc}} = \overline{s_{sc}} = 0.5$$

Parameter Estimation

- Regress daily returns of 100 stocks against daily returns of selected ETFs

$$R_n^S = \beta_0 + \beta R_n^I + \epsilon_n, \quad n = 1, 2, ..., 60.$$

- Compute cumulative residuals

$$X_k = \sum_{j=1}^{k} \epsilon_j \quad k = 1, 2, ..., 60.$$

- Solve the order(1) autoregressive model

$$X_{n+1} = a + bX_n + \zeta_{n+1}, \quad n = 1, ..., 59.$$

- Calculate parameters k, m, σ