

# Statistical Arbitrage in the U.S. Equities Market

Tinghao LI (tl1698)

New York University

May 2019

# Agenda

- 1 What is Statistical Arbitrage?
- 2 Mean Reversion Trading Models
- 3 Backtesting
- 4 Conclusion

# Introduction

- Statistical Arbs are systematic and rule-based
  - Pair-Trading. If stocks A and B have similar characteristics or cointegrated statistically, we could model the two time series via,

$$\ln(P_t^A/P_{t_0}^A) = \alpha(t - t_0) + \beta \ln(P_t^B/P_{t_0}^B) + X_t$$

,where  $X_t$  is a mean reverting/stationary process of interest.

- Generalized Pair-Trading. Trading stock A against a groups of other stocks or factors. Mathematically speaking,

$$\ln(P_t^A/P_{t_0}^A) = \alpha(t - t_0) + \sum_{i=1,2,\dots,N} \beta_i \ln(F_t^i/F_{t_0}^i) + X_t$$

- Where is the PnL coming from in essence?
  - **Market over-reaction & Noise trades**

## Relative-value model for equity pricing

- We model the residual with Ornstein–Uhlenbeck process

$$dX_i(t) = \kappa_i(m_i - X_i(t))dt + \sigma_i dW_i(t)$$

, with the assumption that in each estimation periods the  $\kappa_i$ ,  $m_i$  and  $\sigma_i$  are constant.

- **Trading signal** from a pure mean-reversion process

$$s_i = \frac{X_i(t) - m_i}{\sigma_{eq,i}}$$

, where  $\sigma_{eq,i} = \frac{\sigma_i}{2\kappa_i}$  computed from the OU process.

# Factor generation

- PCA Risk Factors
  - Computed the empirical correlation with the normalized stock return data
  - Estimated the eigen-values and eigen-vectors to form the **"eigenportfolio"**
  - One advantage regarding the formed portfolio is the independence of factors
  - But how many factors do we choose ? It depends.
- ETF Factors
  - Per stock, using one ETF return time series as only factor
  - Factor is intuitively appealing Vs PCA but biased against small capitalization companies.
  - Got good results but with most improving potential

## Trading Setting

- Estimation of parameters (betas, residuals), signal evaluations are using a 60-day trailing window for all stocks daily and portfolio re-balancing are performed daily
- Trade if the following signals being triggered (all-or-nothing strategy)

- Buy or short to open when :

$$s_i < -\hat{s}_{bo}$$

$$s_i > +\hat{s}_{so}$$

- And close position when :

$$s_i < +\hat{s}_{bc}$$

$$s_i > -\hat{s}_{sc}$$

, where  $\hat{s}_{bo} = \hat{s}_{so} = 1.25$  and  $\hat{s}_{bc} = 0.75$  and  $\hat{s}_{sc} = 0.5$  are estimated empirically and never changed for different stocks and times to avoid data mining

# Trading Signals from Calendar price

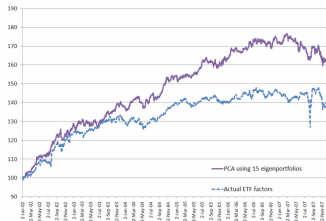


Figure 18: Comparison of strategies with 15 PCA factors and the using actual ETFs in the period 2002-2007. 15-PCA outperforms significantly the ETF strategy.

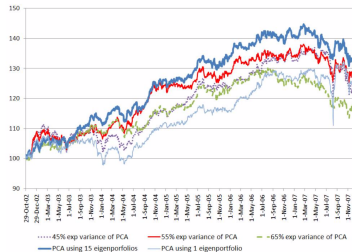


Figure 25: PNL for different variance truncation level: 2002-2007

(a) ETF Vs PCA factors

(b) With dif number of factors

- 15-PCA outperforms ETF factors by a large margin.
  - 1-PCA showed identical performances as ETF factors.
- why?

## Trading Signals from Volume-adjusted price

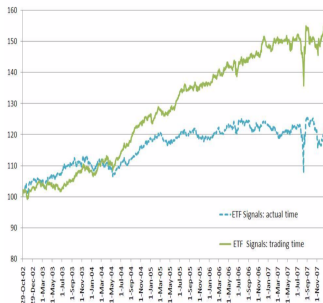


Figure 28: Comparison of signals in trading time vs. actual time using actual ETFs as factors : 2002-2007

Volume-adjusted price returns is defined :

$$\hat{R}_t = R_t \left( \frac{\bar{V}}{V(t + \Delta t) - V(t)} \right)$$

, where  $\bar{V}$  is the typical daily trading volume computed from a 60-day trailing window.

- It avoids trading against the short-term dominate patterns and only in if the potential returns are high



- It is a robust and not so data mining study for finding the pair trading strategies in US.
- PCA and ETF factors are essentially the two sides of coins.
- Empirically, they found the mean-reversion Statistics Arbs works best when 50% of variance can be explained via a **small** number of factors ( $\approx 10 - 20$ )
  - Too much variance and not enough factors, the residuals are not stationary and can have trends.
  - Too many factors, residuals fluctuates too little, less trading opportunities and overfitting further reduce the performance

**Further research** : Why a big drop for ETF-based factors in 2007 but not for optimal PCA factors ?

## References

Avellaneda, Marco, and Jeong-Hyun Lee. "Statistical Arbitrage in the U.S. Equities Market." SSRN Electronic Journal (2008). doi :10.2139/ssrn.1153505.