

Pairs Trading Based on Co-integration

NY1: Hang Ren, Zihan Zhou,
Lei Zhao, Yangfan Mu and
Sixuan Ren

Agenda

- Motivation
- Pair Selection Methods
- Trading Rules
- Data Description
- Empirical Results
- Statistical Arbitrage Test
- Conclusions and Future Research

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Motivation

- Pairs trading is one of the most popular statistical arbitrage strategies which is worth studying and researching
- Co-integration focuses on long-term relationships, meaning that deviations from such relationships can open up profitable trading opportunities



Pair Selection Methods

- Pre-selection
 - Correlation
 - Clustering
 - Graphical Lasso
- Co-integration
 - Engle-Granger Co-integration
 - Johansen Test
- In-sample Sharpe Ratio

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Pre-Selection Filtering

- Correlation
 - 0.97 correlation threshold to pre-select pairs
- Clustering
 - 11 default clustering groups, matching the number of S&P 500 sectors
- Graphical Lasso



Co-integration

- Engle-Granger Co-integration
- Johansen Test

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

In-sample Sharpe Ratio

- Significant co-integration \Rightarrow profitability
- Rank the stocks with in-sample simulated trading Sharpe Ratio
- Choose top 20 pairs or baskets

Correlation Matrix		Clustering		Graphical Lasso	
Co-integration test (Engle-Granger Test or Johansen Test)					
With SR Test	W/o SR Test	With SR Test	W/oSR Test	With SR Test	W/oSR Test



Trading Rules

- Basic Trading Idea
- Potential Risks
- Transaction Costs
- Trading Rules Formulation



Basic Trading Idea

- A stationary process formed by a linear combination of two or more series has a desirable property of mean-reversion for pairs trading strategy
- Trace the dynamics of the linear combination of logarithm price processes that forms a stationary process (spread)
- When spread diverges far enough from its mean we open trades and when it converges back or near to its mean we close the trades and realize profits



Potential Risks

- Spread goes too far away from its mean level but never comes back
- Set some bail-out levels to reduce such risks



Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Transaction Costs

- Bid-ask spread
 - use bid/ask prices when we short/long stocks
- Commission fees
 - 2* 0.01% brokerage fees and 0.02% rental costs
- Market impact costs
 - 2* 0.005% slippage costs

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Trading Rules Formulation

- Spread definition:

- $$LR = \log(ask_1) - \beta \log(bid_2) \rightarrow sLR = \frac{LR - Lm}{Ls}$$
- $$SR = \log(bid_1) - \beta \log(ask_2) \rightarrow sSR_s = \frac{SR - Sm}{Ss}$$

Moving Look-back Window	Formation Period
mv	rolling

- Trading signals:

- No open position + $sLR_t < Lo$ & $sLR_t > Lb \rightarrow$ long spread;
- Hold a long spread position + $sSR_t \geq Lc$ | $sSR_t \leq Lb \rightarrow$ close;
- No open position + $sSR_t > So$ & $sSR_t < Sb \rightarrow$ short spread;
- Hold a short spread position + $sLR_t \leq Sc$ | $sLR_t \geq Sb \rightarrow$ close.



Trading Rules Formulation

- Trading shares:
 - $1/P_1(t_0)$ and $\beta/P_2(t_0)$
- Close Types:

Position	Symbol	Meaning
Long pairs	du	Down and up, return to mean.
	dd	Down and down, bail out.
	dexc	Down and exceed the maximum holding length.
	dna	Down and encounter unavailable data.
Short pairs	ud	Up and down, return to mean.
	uu	Up and up, bail out.
	uexc	Up and exceed the maximum holding length.
	una	Up and encounter unavailable data.

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Data Description

- Quarterly updated NYSE listed stocks from 2001/01/02 to 2016/08/31
- Price, Ask or High, Bid or Low, Closing Bid, Closing Ask and Cumulative Factor; price/Cumulative Factor as the adjusted closing price
- Take absolute value of negative closing prices
- Unavailable data
 - formation period
 - trading period

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Empirical Results

- Trading and Testing Periods
- Parameter Tuning
- Stock Selection & Trading Strategies
- Performance Measurement
- In-sample Results
- Trading Period Results
- Visualization

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Trading and Testing Periods

- The training period is from the beginning of Jan 2001 to the end of Dec 2010
- The testing period is from the beginning of Jan 2011 to the end of June 2016
- Data is further cut into formation period and trading period for both periods
- Trading periods are not overlapped while formation periods are likely to overlap with each other



Parameter Tuning

- Number of days of formation periods J and that of trading periods K
 - J and K can take value of 60, 90 and 120
- Six trading thresholds
 - short bail-out level, short enter level, short exist level, long exist level, long enter level and long bail-out level

	Lb	Lo	Lc	Sc	So	Sb
1	-6	-2	0	0	2	6
2	-10	-2	0	0	2	10
3	-3.5	-2	0	0	2	3.5
4	-10	-2	-0.5	0.5	2	10
5	-6	-1	0	0	1	6

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Stock Selection & Trading Strategies

- Four stock selection methods
 - Correlation w/ & w/o in-sample Sharpe Ratio
 - Graphical Lasso w/ & w/o in-sample Sharpe
- Two trading rules
 - Spread standardized using moving windows
 - Spread standardized using formation statistics



In-sample Results

Pairs Selection Methods		Trading Rule	J	K	Sb	So	Sc	Lc	Lo	Lb	Sharpe Ratio	Average PnL	PnL Volatility	Cumulative PnL	Sharpe Ratio With TC	Average PnL With TC	PnL Volatility With TC	Cumulative PnL With TC
Correlation+Co-integration Test	with SR test	mv	120	120	10	2	0	0	-2	-10	0.09046	0.08	0.87	9.79	0.0903	0.078	0.866	9.770
		rolling	120	90	6	2	0	0	-2	-6	0.0521	0.03	0.52	3.44	0.0520	0.0274	0.527	3.428
	without SR test	mv	120	60	10	2	0	0	-2	-10	-0.0637	-0.81	12.84	-102.31	-0.064	-0.821	12.842	-102.675
		rolling	90	90	6	2	0	0	-2	-6	-0.1017	-12.67	124.55	-1583.12	-0.1017	-12.671	124.612	-1583.92
Graphical Lasso+Co-integration Test	with SR test	mv	120	120	6	1	0	0	-1	-6	-0.0054	-0.08	14.77	-10.00	-0.0056	-0.082	14.76	-10.21
		rolling	90	60	6	2	0	0	-2	-6	0.2535	1.23	4.84	153.61	0.2535	1.228	4.845	153.53
	without SR test	mv	120	120	6	1	0	0	-1	-6	-0.006	-0.09	14.77	-11.12	-0.006	-0.091	14.77	-11.33
		rolling	90	120	6	2	0	0	-2	-6	0.1844	2.29	12.40	286.09	0.1844	2.286	12.39	285.79

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

In-sample Results

Pairs Selection Methods		Trading Rule	J	K	Sb	So	Sc	Lc	Lo	Lb	Winning Percentage	Maximum Drawdown	Total Long Number	Total Short Number	Average Long Days	Average Short Days	Average PairsNum
Correlation +Co-integration Test	with SR test	mv	120	120	10	2	0	0	-2	-10	0.65	7.67	576	643	23.08	29.00	16.7
		rolling	120	90	6	2	0	0	-2	-6	0.30	3.22	36	56	28.03	48.63	16.44
	without SR test	mv	120	60	10	2	0	0	-2	-10	0.63	183.63	22976	22392	27.28	27.90	548.775
		rolling	90	90	6	2	0	0	-2	-6	0.26	1584.27	791	1301	25.68	31.07	659.184
Graphical Lasso+Co-integration Test	with SR test	mv	120	120	6	1	0	0	-1	-6	0.51	183.71	881	927	16.65	16.99	14.25
		rolling	90	60	6	2	0	0	-2	-6	0.41	6.25	164	147	9.30	8.32	13.39
	without SR test	mv	120	120	6	1	0	0	-1	-6	0.51	185.13	923	973	16.39	17.00	14.9
		rolling	90	120	6	2	0	0	-2	-6	0.39	45.76	670	773	11.65	8.98	14.09

Agenda

Motivation

Selection Methods

Trading Rules

Data Description

Empirical Results

Stat. Arb. Test

Conclusions

In-sample and Out-of-sample Comparison

Pairs Selection Methods		Correlation+Co-integration Test				Graphical Lasso+Co-integration Test			
		with SR test		without SR test		with SR test		without SR test	
Trading Rule		mv	rolling	mv	rolling	mv	rolling	mv	rolling
Sharpe Ratio	In-sample	0.09	0.05	-0.06	-0.10	-0.01	0.25	-0.01	0.18
With TC	Out-of-sample	0.12	0.12	0.25	-0.12	-0.03	-0.08	-0.03	-0.21
AveragePnL	In-sample	0.08	0.03	-0.82	-12.67	-0.08	1.23	-0.09	2.29
With TC	Out-of-sample	0.20	0.26	1.22	-0.09	-0.32	-0.31	-0.32	-2.25
		0.87	0.53	12.84	124.61	14.76	4.85	14.77	12.39
PnLVolatility	In-sample	1.72	2.07	4.80	0.81	9.24	3.91	9.24	10.84
With TC	Out-of-sample	9.77	3.43	-102.68	-1583.92	-10.21	153.53	-11.33	285.79
CumulativePnL	In-sample	14.10	18.27	86.42	-6.61	-22.79	-22.18	-22.79	-159.72
With TC	Out-of-sample	0.09	0.05	-0.06	-0.10	-0.01	0.25	-0.01	0.18

This chart shows the in-sample and out-of-sample performance comparison of the best in-sample set of parameters for each trading strategy

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Out-of-sample Results

Pairs Selection Methods		Trading Rule	J	K	Sb	So	Sc	Lc	Lo	Lb	Sharpe Ratio	Average PnL	PnL Volatility	Cumulative PnL	Sharpe Ratio With TC	Average PnL With TC	PnL Volatility With TC	Cumulative PnL With TC
Correlation+Co-integration Test	with SR test	mv	120	120	10	2	0	0	-2	-10	0.12	0.2	1.73	14.11	0.12	0.2	1.72	14.1
		rolling	120	90	6	2	0	0	-2	-6	0.12	0.26	2.07	18.29	0.12	0.26	2.07	18.27
	without SR test	mv	90	120	10	2	0	0	-2	-10	0.25	0.46	1.79	32.41	0.25	0.46	1.79	32.38
		rolling	90	90	6	2	0	0	-2	-6	-0.12	-0.09	0.81	-6.6	-0.12	-0.09	0.81	-6.61
Graphical Lasso+Co-integration Test	with SR test	mv	90	120	10	2	0	0	-2	-10	0.1	0.21	2.15	14.59	0.1	0.2	2.15	14.55
		rolling	90	60	6	2	0	0	-2	-6	-0.08	-0.31	3.91	-22.14	-0.08	-0.31	3.91	-22.18
	without SR test	mv	90	120	10	2	0	0	-2	-10	0.1	0.21	2.15	14.59	0.1	0.2	2.15	14.55
		rolling	90	120	6	1	0.5	-0.5	-1	-6	-0.21	-2.25	10.83	-159.56	-0.21	-2.25	10.84	-159.72

Agenda

Motivation

Selection Methods

Trading Rules

Data Description

Empirical Results

Stat. Arb. Test

Conclusions

Out-of-sample Results

Pairs Selection Methods		Trading Rule	J	K	Sb	So	Sc	Lc	Lo	Lb	Winning Percentage	Maximum Drawdown	Total Long Number	Total Short Number	Average Long Days	Average Short Days	Average Pairs Num
Correlation+Co-integration Test	with SR test	mv	120	120	10	2	0	0	-2	-10	0.78	4.01	213	221	28.35	32.4	11.91
		rolling	120	90	6	2	0	0	-2	-6	0.47	7.87	64	53	35.44	38.21	11.87
	without SR test	mv	90	120	10	2	0	0	-2	-10	0.81	5.53	1498	1341	24.28	23.12	62
		rolling	90	90	6	2	0	0	-2	-6	0.31	12.67	56	90	30.23	31.44	40.2
Graphical Lasso+Co-integration Test	with SR test	mv	90	120	10	2	0	0	-2	-10	0.71	9.22	177	181	28.57	31.61	8.5
		rolling	90	60	6	2	0	0	-2	-6	0.43	40.49	61	38	11.05	11.16	7.7
	without SR test	mv	90	120	10	2	0	0	-2	-10	0.71	9.22	177	181	28.57	31.61	8.58
		rolling	90	120	6	1	0.5	-0.5	-1	-6	0.37	178.45	274	348	9.95	10.57	8.58

Agenda

Motivation

Selection
Methods

Trading
Rules

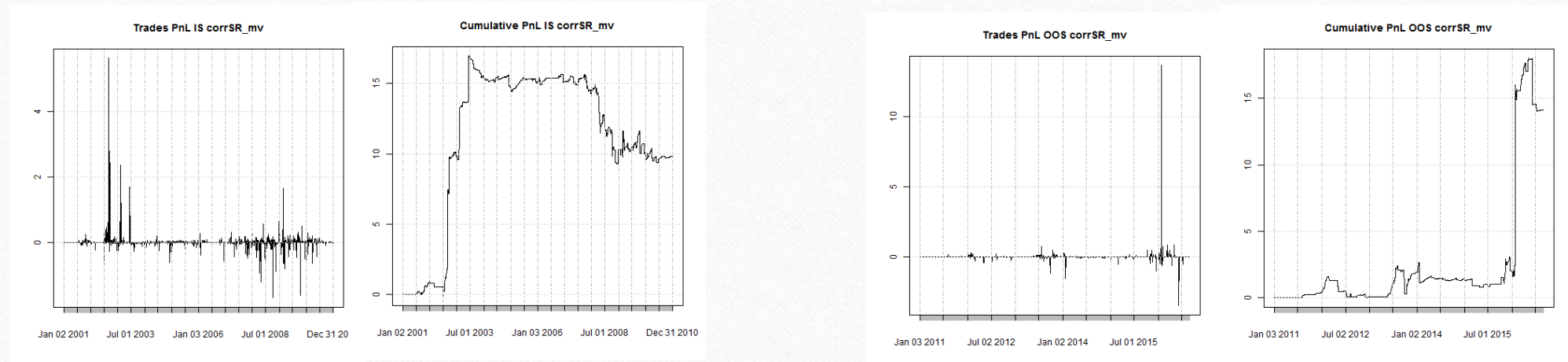
Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Performance Visualization



Best in-sample cumulative return and its out-of-sample performance

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

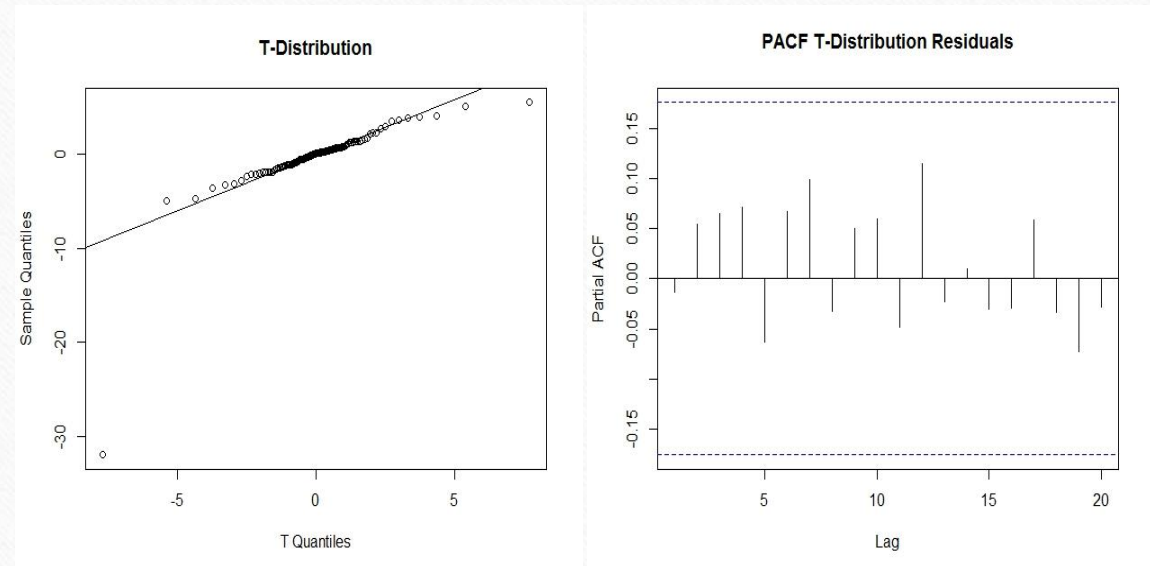
Stat. Arb.
Test

Conclusions

Statistical Arbitrage Test

The in-Sample with T-Statistical Residuals test gave the above p-value

	mu	lambda	df
graphSR_rolling	0.9746943	0.8619055	0.4813791



Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

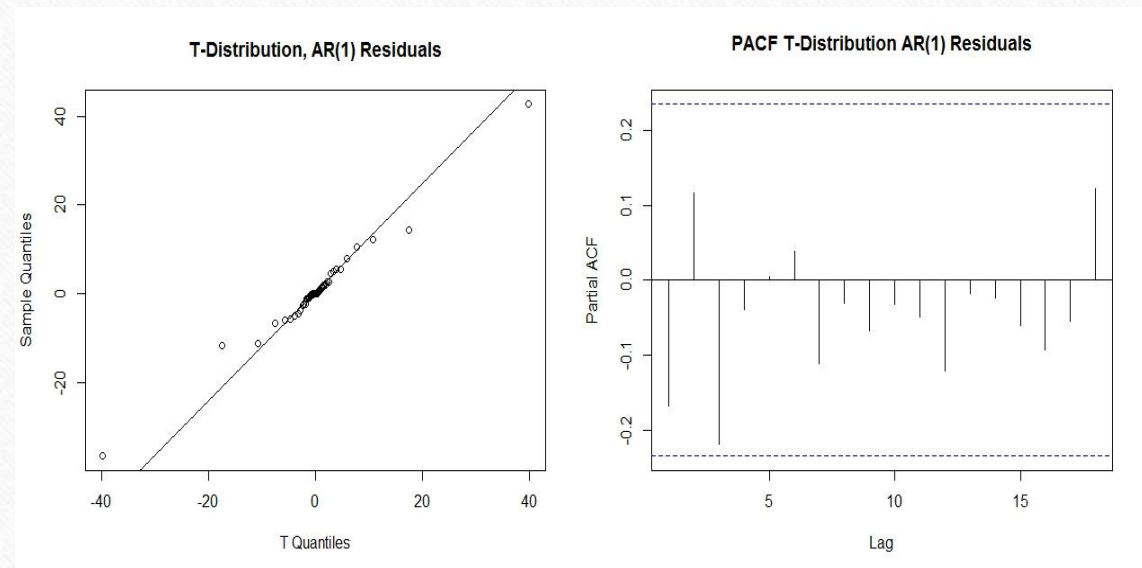
Stat. Arb.
Test

Conclusions

Statistical Arbitrage Test

The out-of-Sample with T-Statistical & AR(1) Residuals test gave the above p-value

	mu	lambda	df
Corr_MV	0.5653275	0.9616601	0.999771



Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Conclusions Future Research

- Sharpe Ratio test can help to significantly improve the Sharpe Ratio in the trading period under all cases.
- Although Graphical Lasso can pre-select stocks with higher profit potential, the performance is less stable and the drawdown is larger.
- Comparatively, performances of correlation matrix are more stable and more consistent in sample and out of sample.
- Using moving average windows to standardize spreads can realize a higher winning percentages than using formation periods.
- The higher cumulative profits by using formation periods are largely due to a few extremely large daily profits.
- Therefore, we prefer the strategy with correlation matrix as pre-selection measure and moving look-back windows as standardizing methods.

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Future Research

- 1. We should try different data frequencies to find the optimal choice for co-integration pairs-trading strategy to improve the Sharpe Ratio;
- 2. We need to find the reasons of extreme values in the profit series to come up with possible solutions to reduce the volatility;
- 3. We should further consider trading strategies with hedging options to determine whether hedging would improve the strategy performance;
- 4. We can use other information to determine the trade size instead of the price series, like volume data and minimum profit level we would like to realize.

Agenda

Motivation

Selection
Methods

Trading
Rules

Data
Description

Empirical
Results

Stat. Arb.
Test

Conclusions

Thank you

Thank you!

3:10 am. We finished the final project in this program. We appreciate the one and half year we being together and the memories we hold. All the best to everything!

