

# 통계적 차익거래 기법 보고서

조규영(20213858)

## 변경부분

데이터 기간 변경(코로나 기간으로 인해 노이즈가 많다고 판단.)  
PCA를 통해 팩터를 만들 때 사용한 Eigen vector 수에 대한 분석추가  
팀 프로젝트에서 반영했던 Transaction cost, Return 적용 방식 변경  
Signal 변경  
Pair volume에 대한 추가적인 고찰  
cointegration p-value변경(좀 더 많은 페어를 뽑기 위함)

## ■ Abstract

페어 트레이딩은 통계적 차익거래기법으로, 서로 연관된 자산의 가격 차이에서 거래 기회를 포착하는 기법이다. 이러한 자산 간의 가격 차이가 장기적으로 일정하게 유지된다는 가정으로 시작한다. 즉, 두 종목으로 구성된 스프레드가 일시적으로 원래의 움직임에서 벗어날 때 스프레드의 저평가 및 고평가 유무에 따라 해당 자산을 매입하고 매도하여 수익을 창출한다. 매수와 매도를 동시에 하기 때문에 포트폴리오의 베타를 중립화 할 수 있어서, 시장의 위험과 관련 없는 수익이 발생할 것이라 생각하기 때문에 대표적인 시장 중립 전략 중 하나이다.

## ■ Principal Component Analysis(주성분 분석)

페어를 구성하기 위해 주성분 분석을 통해 주식 시장의 개별 주식 수익률에 내재되어 있는 체계적 위험 요인을 추출한다. Eigen vector를 통해 팩터를 만들고 이를 통해 각 개별 주식이 가지고 있는 시장 위험을 추정. 그 후 유의한 베타에 대한 주식들끼리 1차 군집화를 진행.

### ✓ Factor Mimicking Portfolio

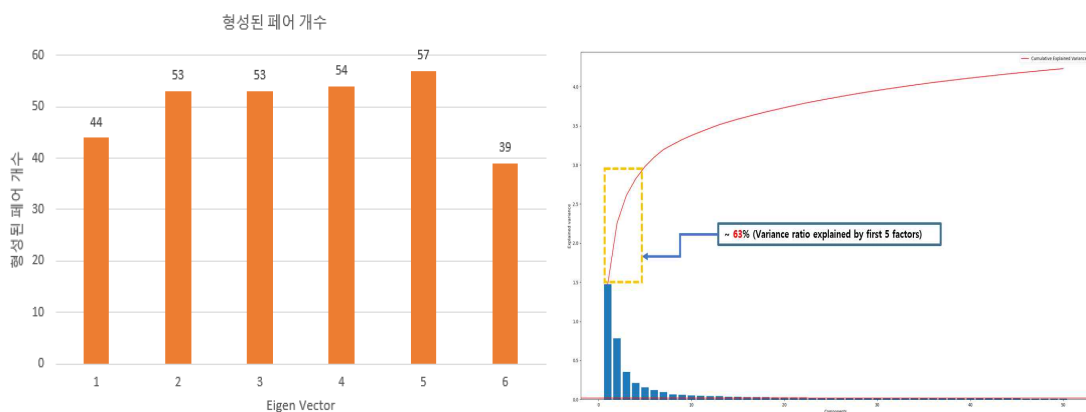
$$F_t^k = \sum_{i=1}^N \frac{v_i^k}{\bar{\sigma}_i} R_{i,t} = \sum_{i=1}^N Q_i^k R_{i,t}$$

- $v_i^k$  = Eigen Vector for  $F_t^k$  (고유 포트폴리오)
- $\bar{\sigma}_i$  = Standard deviation of individual stocks

### ✓ PCA Factor Model

$$R_{i,t} = \alpha_i + \sum_{k=1}^K \beta_{ik} F_t^k + \epsilon_{i,t}$$

- $R_{i,t}$ : the i-th stock return at time t
- $F_t^k$ : time t eigenportfolios with weight  $Q_i^k$
- $\beta_{ik}$ : the i-th asset's factor loading to the k-th factor
- $\epsilon_{i,t}$ : residual



훈련기간을 통해 형성된 페어 개수를 보면 고유 벡터를 5개를 선택했을 때 가장 많은 페어가 성되어 상위 5개의 벡터를 사용했다. 상위 5개의 고유 벡터를 선택했을 때, **전체 변동성의 약 63%를 설명함**을 확인할 수 있다. 이는 MKF500이 상위 시가총액 기준 상위 500개 기업으로 이루어진 인덱스이므로 적은 고유벡터로도 전체 변동성의 대부분을 설명할 수 있기에 나타난 결과이다.

## ■OPTICS

밀도 기반의 클러스터를 탐색하기 위한 알고리즘으로써 데이터들을 정렬하여 가까운 point들이 이웃이 될 수 있도록 하며, 각 point가 더 밀집된 cluster에 포함될 수 있도록 한다. 밀도뿐만 아닌 계층까지 유지하기 때문에 더 효과적인 군집을 할 수 있다.

## ■공적분(Cointegration)

$$\ln\left(\frac{P_t}{P_{t_0}}\right) - \beta \ln\left(\frac{Q_t}{Q_{t_0}}\right) = \alpha(t - t_0) + X_t$$

여기서 P와 Q는 각각 자나 P와 Q의 t시점에서의 가격이다. X는 수익률의 잔차로 해석할 수 있으며 평균 회귀 경향을 갖는 정상 확률 과정이다. 두 자산의 계수를 조정한 스프레드가 통계적 평형 근처에서 진동한다는 것이다. 이러한 관계를 갖는 자산을 찾아 페어를 구성하고 그 스프레드가 확대되었을 때 고평가된 자산을 매도, 저평가된 자산을 매수한다면 평균회귀 경향을 갖는 X항으로 인해 수익을 얻을 수 있다.

## ■Signals

거래 신호로는 고정 임계 값 기반의 방법을 사용했다. Z score를 사용했고 **페어 개방 구간은 0.8~2.5**, **페어 정리를 위한 임계점 0.5**, **페어 손절매를 위한 임계점을 3**으로 설정했다. 페어 개방 구간을 설정한 이유는 **스프레드가 급격하고 과도하게 벌어졌을 시 공적분 관계가 깨질 수 있는 위험을 줄이고 페어 개방 구간으로 들어오게 된다면 어느 정도 노이즈가 사라지게 된다고 판단했다**. 즉, 안정되고 스프레드가 평균으로 회귀할 수 있는 구간이라 생각했다.

## ■수익률 측정방법(수정주가 & 거래비용 10bp 적용)

전에 발표했던 군집화 기반 앙상블 페어 선택 알고리즘을 사용한 페어 트레이딩 전략의 성과 분석(박찬주)에서 사용한 개별 수익률 대신 **페어가 개방되는 시점과 닫히는 시점에서 베타를 조정하지 않고 거래비용을 고려**. 또한 한 **페어가 개방되면 베타의 계수의 크기와 상관없이 leverage를 고려해 1만원의 비용으로 고정한다는 가정** 후 진행했습니다. 이는 좀 더 보수적으로 개별 수익률을 측정하기 위함이다. 또한 투하자본 수익률로 포지션을 개방하지 않은 페어의 기회비용을 고려한 보수적인 페어 포트폴리오 수익률 성과 측정방법을 사용했다.

## ■Time period 및 Selected pair

	Formation period	Trading period
Senario1	2018/01/01~2018/12/31(12개월)	2019/01/01~2019/07/01(6개월)
Senario2	2019/07/01~2019/07/01(12개월)	2019/07/01~2019/12/31(6개월)

	Selected Pair
Senario1(8 pairs) (# of total pair = 38)	['한국금융지주', '키움증권'], ['동진씨미캠', '테스'], ['KB금융', '한화생명'], ['만도', '오리온홀딩스'], ['유한양행', '대웅제약'], ['디앤씨미디어', '일진다이아'], ['삼성엔지니어링', '씨에스윈드'], ['솔브레인홀딩스', '에프에스티']
Senario2(9 pairs) (# of total pair = 19)	['넷마블', '제주항공'], ['유진테크', '원익홀딩스'], ['한국가스공사', '현대일렉트릭'], ['LG전자', '금호석유'], ['S-Oil', '교보증권'], ['한국조선해양', '대우조선해양'], ['미코', '에이프로젠 MED'], ['RFHIC', '네패스'], ['씨에스윈드', '미원에스씨']

$2^5 - 1 = 16$ 개에 대한 각각의 군집 중에서 가장 p-value가 낮은 1개의 페어만 선택 후 시뮬레이션 진행.(분산효과고려)  
16개가 아닌 것은 군집에서 페어가 유의하지 않다고 결과로 인한 현상.

Pair Selection 예시)

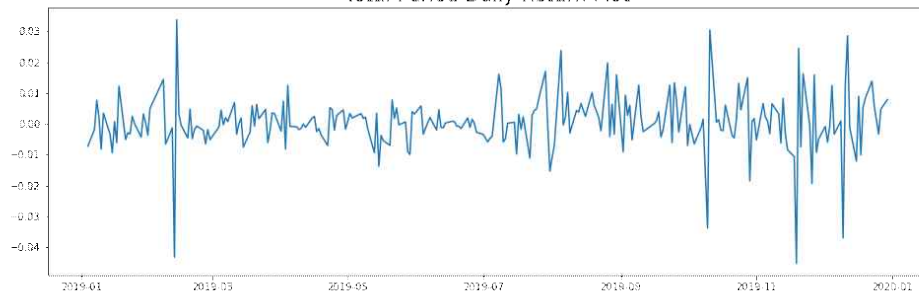
	Group0	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8	Group9		
0	NAVER	DGB금융지주	메리츠금융지주	SK이노베이션	현대글로비스	고려아연	KCC	씨젠	한화	한화솔루션		
1	서흥	세방전지	영원무역	S-Oil	이마트	롯데케미칼	SK네트웍스	진원생명과학	대한유화	한국금융지주		
2			코리안리	BGF		롯데정밀화학		유나이티드제약		NH투자증권		
3						한진		넥슨지티		키움증권		
4										다우데이터		
			Group0	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8	Group9
,다엔씨미디어,일진다이아,				6.54E-05								
,다엔씨미디어,미코,				7.77E-05								
,RFHIC,안랩,					0.028807							
,아이센스,위닉스,						0.005957						
,알서포트,송원산업,							0.042159					
,인텔리안테크,삼양식품,								0.037038				
,현대퓨처넷,이베스트투자증권,											0.01537	
,현대퓨처넷,인탑스,											0.02668	
,현대퓨처넷,유비쿼스홀딩스,											0.02858	
,영원무역홀딩스,이베스트투자증권,											0.037848	
,영원무역홀딩스,풀무원,											0.044762	

## ■ 결과 분석 - PnL

Cumulative return plot

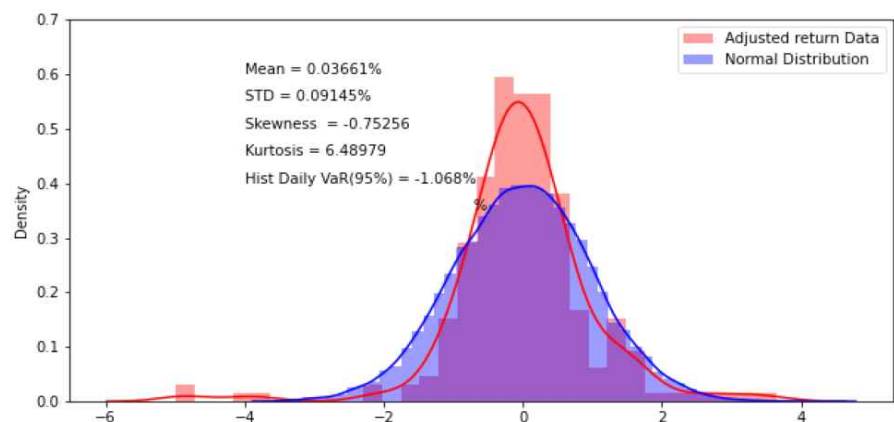


Total Period Daily Return Plot



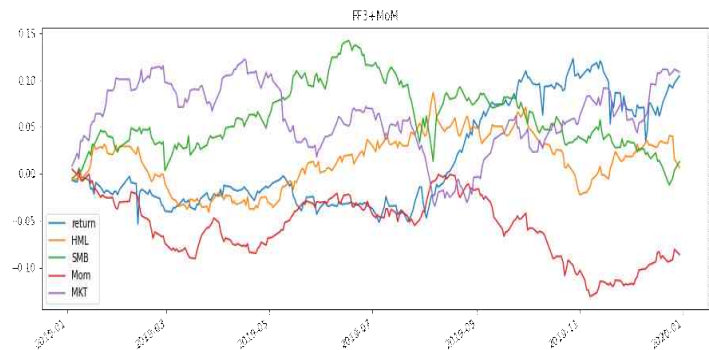
2019.01.01.~2020.01.01.기간 거래비용을 고려한 누적 수익률을 확인결과 2019.02.12.기간에 약-5%로 가장 큰 손실, 2019.10.28.에 약 12%로 가장 큰 수익을 나타내고 최종 누적 수익률은 10%로 마감했다. 상반기에는 손실 상태로 마감했지만 하반기부터 급격하게 상승하는 것을 볼 수 있다.

■ Excess Return Distribution(CD금리(2019년) : 1.6% 적용)



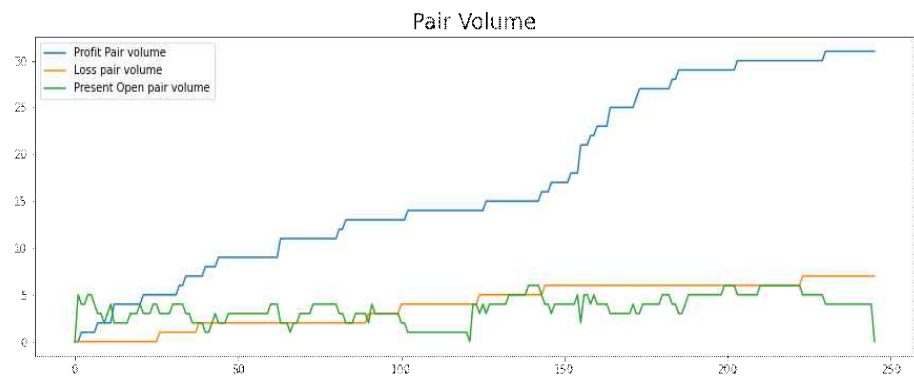
Daily adj return	0.0366%
Annual adj return	9.22%
Daily adj Std	0.0914%
Annual adj Std	1.45%
Sharpe Ratio	0.63
T-stat(u=0)	4.151(0.00)
T-stat(u=rf)	0.47(0.63)

■ Alpha Test



OLS Regression Results									
Dep. Variable:	return	R-squared:	0.038						
Model:	OLS	Adj. R-squared:	0.021						
Method:	Least Squares	F-statistic:	2.327						
Date:	Fri, 17 Dec 2021	Prob (F-statistic):	0.0570						
Time:	20:37:13	Log-Likelihood:	801.13						
No. Observations:	243	AIC:	-1592.						
Df. Residuals:	238	BIC:	-1575.						
Df. Model:	4								
Covariance Type:	nonrobust								
	coef	std. err.	t	P> t	[0.025	0.975]			
const	0.0005	0.001	0.842	0.401	-0.001	0.002			
HML	0.0120	0.104	0.115	0.908	-0.193	0.217			
SMB	-0.2019	0.083	-2.429	0.016	-0.366	-0.038			
Mom	0.1666	0.122	1.365	0.174	-0.074	0.407			
MKT	0.0190	0.079	0.241	0.810	-0.136	0.174			
Diagnostics									
Durbin-Watson:	70.785	2.462							
Prob(Durbinus):	0.000	547.248							
Skew:	-0.896	1.47e-119							
Kurtosis:	10.130	Cond. No.							
		236.							

■ Position Plot



19년 한 해 동안 정상적으로 닫힌 페어의 수와 손절 기준을 넘은 페어의 수에 대한 그래프이다. 한 해 동안 정상적으로 페어가 열리고 닫힌 수는 31번이고 손절 기준을 넘어 손실을 보면서 페어가 닫힌 횟수는 7번임을 확인할 수 있다. 페어가 한번 열리고 이익을 보면서 닫히는 경우의 평균적인 기간은 약 8.52일(영업일 기준) 임을 확인했다.

## ■총 평

통계적 차익거래 기법을 이용하여 시장 중립 전략 중 하나인 페어 트레이딩을 2018/01/01~2019/07/01(약 18개월)을 훈련기간으로 정하고 2019/01/01~2019/12/31(약 12개월)을 투자기간로 하여 투자 시뮬레이션을 진행하였다. 상반기 구간에서는 손실이 났지만 하반기 구간 부터는 급격한 회복세를 보이기 시작했다. 그로인해 19년(12개월)동안 거래비용을 고려한 누적 수익률은 약 10%로 나타났다. 연 평균 수익률과 표준편차는 약 9%, 1.45%로 나타났음을 확인할 수 있다. 또한 샤프지수는 0.64로 높은 수치로 받아들이기는 어렵다.

여러 가지 통계적 기법을 통해 분석을 해보았을 때, 수익률을 0과 무위험 이자율과의 차이에 대해서 t-test 진행 결과 0과는 다르지만 무위험이자율과는 다르지 않다라는 통계적 수치가 나타나는 것을 확인할 수 있다. 또한 수익률과 KOSPI 수익률을 비교했을 시, KOSPI의 수익률보다는 조금 못미치는 결과가 나왔고 FF3 factor + MoM과의 regression을 통해 alpha에 대한 검증을 해보았을 때, SMB와 MoM이 가장 잘 설명한다고 볼 수 있고 alpha의 존재는 유의하지 않았음을 확인할 수 있다.