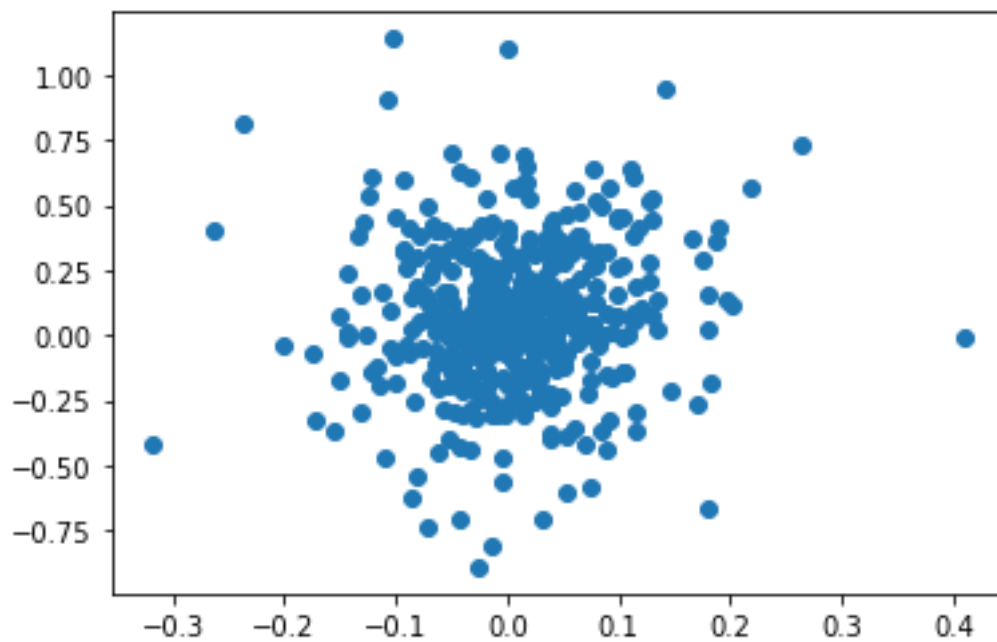


1. KOSPI의 경우

아래 표에서 log는 변동률을, lag_11은 미래 11개월 변동률을 의미한다.

	year	month	log	lag_11
count	477.000000	477.000000	477.000000	477.000000
mean	2001.377358	6.471698	0.006611	0.071039
std	11.487304	3.447343	0.073195	0.276152
min	1982.000000	1.000000	-0.318104	-0.891025
25%	1991.000000	3.000000	-0.031712	-0.069282
50%	2001.000000	6.000000	0.006421	0.041052
75%	2011.000000	9.000000	0.042852	0.243722
max	2021.000000	12.000000	0.410616	1.139919

1달이 미래 11개월을 예측할 수 있는지 여부에 대한 산점도



상관관계는 없어보인다.

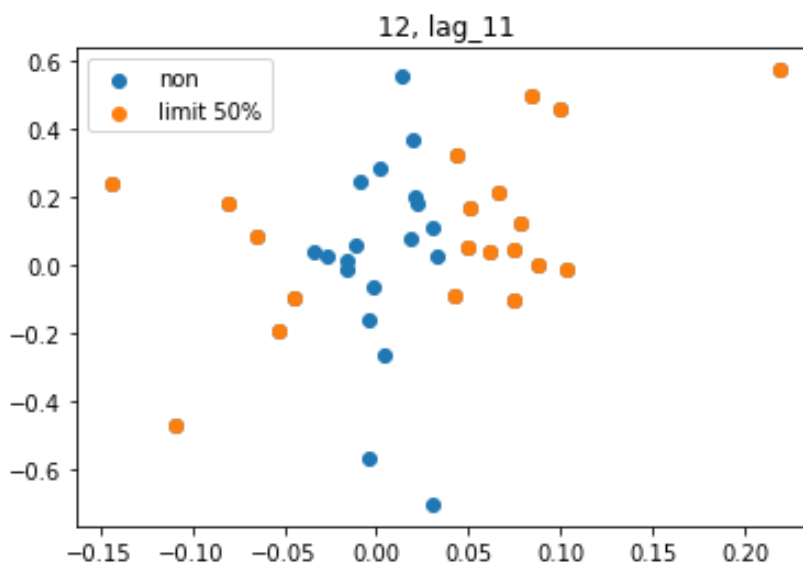
```
def pt_R(R):
    f = pd.DataFrame(index = np.arange(1,13), columns = ['상관계수', '유의성', '변동성 상위 %d percent 상관계수' %R, '유의성'])
    for i in range(1,13):
        k = grp_df.loc[grp_df['month'] == i].reset_index(drop=True)
        kx = k.loc[abs(k['log']) >= np.percentile(abs(k['log']), 100 - R)]
        a = np.corrcoef(k['log'], k['lag_11'])[0,1]
        b = bool(stats.pearsonr(k['log'], k['lag_11'])[1] < 0.05)
        cx = kx['log']
        dx = kx['lag_11']
        c = np.corrcoef(cx, dx)[0,1]
        d = bool(stats.pearsonr(cx, dx)[1] < 0.05)
        f.loc[i] = [a,b,c,d]
    print(f)
```

#변동성 상위 n%를 추출할 때 표준편차 등을 고려하여 기준을 세우고 싶었으나 정규성을 만족하지 못하여 실패..

	상관계수	유의성	변동성 상위 50 percent 상관계수	유의성
1	0.149925	False	0.241894	False
2	-0.053334	False	-0.024236	False
3	-0.011925	False	-0.154162	False
4	0.040076	False	0.103915	False
5	-0.100306	False	-0.170119	False
6	-0.01593	False	-0.041187	False
7	0.209683	False	0.154601	False
8	-0.116755	False	-0.111501	False
9	0.012447	False	-0.066133	False
10	0.253494	False	0.30497	False
11	0.190635	False	0.211455	False
12	0.337892	True	0.489975	True

1월의 경우, 변동성 수준을 조절하더라도 상관계수가 유의하다($\alpha=0.05$)고 볼 수 없었다.

그러나 12월의 경우, 변동성 수준을 제한하지 않더라도 미래 11개월 지수 방향과의 상관관계가 유의하였고, 변동성 상위 50%로 제한하는 경우에도 상관관계가 0.49 정도로 유의하였다. 다만 10%p 단위로 조절할 때, 12월과 미래 11개월 간 상관성이 유의하다고 나타나지는 않았다.



노란 점은 12월과 미래 11기간 간 산점도 중 변동성 상위 50%만을 가리킴.

t.test

국고채 1년 데이터

```
rate = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/rate.csv', parse_dates = ['변환'])
rate.columns = ['date', 'rate']
rate['rate'] = rate['rate']/100 * (11/12) # %변환 및 11개월 반영
rate['year'] = rate['date'].dt.year
rate['month'] = rate['date'].dt.month
```

```
new = pd.merge(grp_df, rate, left_on = ['year', 'month'], right_on = ['year', 'month'], how = 'inner').drop('date', axis = 1) #국채금리 자료부족..
new['updown'] = 0
for i in range(0, len(new)):
    if new['lag_11'].iloc[i] > new['rate'].iloc[i]:
        new['updown'].iloc[i] = 1
    else:
        new['updown'].iloc[i] = -1
new['indi'] = new['log'] * new['updown']
new['indi'] = np.ceil(new['indi'])
```

	year	month	log	lag_11	rate	updown	indi
0	2000	2	-0.130527	-0.293129	0.078192	-1	1.0
1	2000	3	0.038553	-0.398278	0.076542	-1	-0.0
2	2000	4	-0.171315	-0.326707	0.076450	-1	1.0
3	2000	5	0.008907	-0.237151	0.076450	-1	-0.0
4	2000	6	0.115174	-0.293797	0.075533	-1	-0.0
...
255	2021	5	0.017652	-0.172958	0.005601	-1	-0.0
256	2021	6	0.028541	-0.204900	0.007315	-1	-0.0
257	2021	7	-0.029040	-0.316875	0.008342	-1	1.0
258	2021	8	-0.000953	-0.266223	0.009176	-1	1.0
259	2021	9	-0.041630	-0.216245	0.009836	-1	1.0

미래 11개월 지수 변동이 1년(11개월) 국채 금리보다 높다면 updown은 +1이고 아니라면 -1이다.

과거 1개월의 변동률과 updown의 부호가 같다면 indi는 1이고 그렇지 않다면 0이다.

따라서 indi의 평균이 0.5와 같다면 과거 1개월은 미래 11개월 지수 향방 예측에 무의미하다는 것이고 그렇지 않다면 유의미하다는 것이다,

```
stats.ttest_1samp(new['indi'],0.5)
```

```
Ttest_1sampResult(statistic=0.6194382162118393, pvalue=0.5361722251474946)
```

p.value>0.05이므로 과거 1개월이 미래 11개월 지수 향방 예측에 도움을 주지 않는다.

유사하게,

```
for i in range(1,13):  
    print(i, '월 p.value는', stats.ttest_1samp(new.loc[new['month'] == i]['indi'], 0.5)[1])
```

```
1 월 p.value는 0.28596363317570905  
2 월 p.value는 0.6799528986653569  
3 월 p.value는 0.6799528986653569  
4 월 p.value는 1.0  
5 월 p.value는 0.40637117235267894  
6 월 p.value는 0.6799528986653572  
7 월 p.value는 0.20801346649528196  
8 월 p.value는 0.40637117235267894  
9 월 p.value는 0.6799528986653572  
10 월 p.value는 0.28596363317570905  
11 월 p.value는 0.28596363317570905  
12 월 p.value는 0.8333286370363984
```

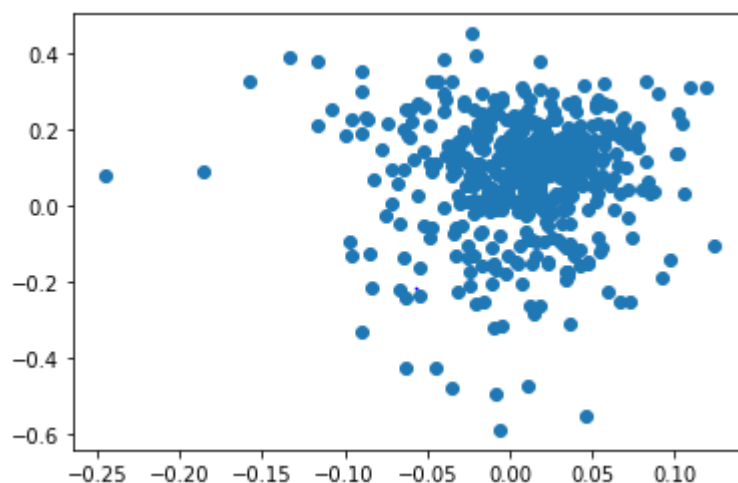
모두 p.value > 0.05이므로 어떤 달도 미래 11개월 지수 예측 향방에 도움을 주지 않는다고 할 수 있다.

2. S&P 500

```
print(pd.describe_percentiles(log_returns, [0, 25, 50, 75, 100]))
```

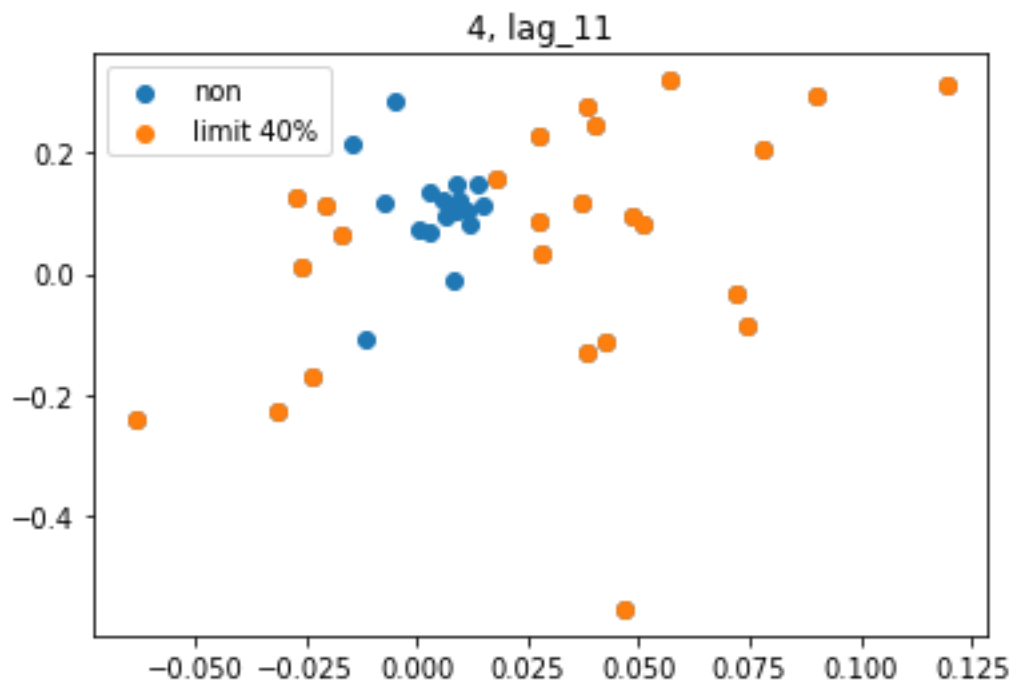
	year	month	log	lag_11
count	505.000000	505.000000	505.000000	505.000000
mean	2000.542574	6.489109	0.007394	0.076967
std	12.158619	3.460718	0.043793	0.151472
min	1980.000000	1.000000	-0.245295	-0.587437
25%	1990.000000	3.000000	-0.016497	0.013231
50%	2001.000000	6.000000	0.011510	0.098313
75%	2011.000000	9.000000	0.035100	0.169757
max	2022.000000	12.000000	0.123729	0.450796

1달이 미래 11개월을 예측할 수 있는지 여부에 대한 산점도



	상관계수	유의성	변동성	상위 40 percent	상관계수	유의성
1	0.157741	False		0.285696	False	
2	0.006548	False		-0.037355	False	
3	-0.221429	False		-0.381709	False	
4	0.271181	False		0.514025	True	
5	0.010874	False		-0.164288	False	
6	0.170521	False		0.212208	False	
7	-0.006137	False		-0.296235	False	
8	-0.102698	False		-0.054576	False	
9	0.048928	False		0.090274	False	
10	0.006205	False		0.030807	False	
11	0.124687	False		0.157032	False	
12	-0.077232	False		-0.286303	False	

4월의 경우, 변동성을 상위 40%로 제한하는 경우 유일하게 과거 4월의 변동성과 미래 11개월 간 상관관계가 존재하고 유의하였다.



```
stats.ttest_1samp(new['indi'],0.5)
```

Ttest_1sampResult(statistic=0.6194382162118393, pvalue=0.5361722251474946)

매 과거 1개월은 향후 11개월의 S&P 지수 향방을 예측할 수 없다.

1 월 p.value는 0.4063711723526793
 2 월 p.value는 0.08807448930807442
 3 월 p.value는 0.20801346649528196
 4 월 p.value는 0.20801346649528193
 5 월 p.value는 1.0
 6 월 p.value는 0.08807448930807442
 7 월 p.value는 0.08807448930807442
 8 월 p.value는 1.0
 9 월 p.value는 0.40637117235267894
 10 월 p.value는 0.6799528986653569
 11 월 p.value는 0.029328876961785502
 12 월 p.value는 0.40637117235267894

각 달로 분리해도 결과는 같다.

3. 결론

Kospi와 S&P500에 나타나는 특수한 경우(12월, 4월)의 상관관계의 확인에는 더 많은 데이터가 필

요한 것 같다.

미래 11개월의 지수 향방을 상승과 하락으로 범주화한 경우에도 과거 1개월의 데이터만으로는 그를 예측할 수 없었다. 따라서 미신은 틀린 것 같다.

다만 S&P의 경우에는 미국 국채 금리가 아닌 한국 금리를 사용하였다는 점에서 문제가 있을 수 있다.

한편, lag11과 국채 금리를 단순 비교하였다는 점, 변동률이 국채금리와 유사한 경우는 고려하지 않았다는 점, 변동률 제한시 데이터 수가 부족하게 된다는 점 등의 문제가 있을 수 있다.