SMA & HULL TRADING

Week 2

Today's Outline

1. **Data Distribution**
2. **S-Volume Analysis**
3. **S-Score Metrics**
4. **Stationarity**

Data Factors

| Factor Name	Description
Raw-S	Unweighted Sentiment Estimate
Raw-S-Mean	20 Day Moving Average of Raw-S
Raw-Volatility	20 Day Moving Standard Deviation of Raw-S
Raw-Score	Normalized Value of Raw-S
S	Exponentially Weighted Sentiment Estimate
S-Mean	20 Day Moving Average of S
S-Volatility	20 Day Moving Standard Deviation of S
S-Score	Normalized Value of S. This Is SMA's S-Score
S-Volume	Indicative Tweet Volume Used To Compute The Sentiment Estimate
SV-Mean	20 Day Moving Average of S-Volume
SV-Volatility	20 Day Moving Standard Deviation of S-Volume
SV-Score	Normalized Value of S-Volume
S-Dispersion	Measurement of The Tweet Source Diversity Contributing To A Sentiment Estimate
S-Buzz	Measurement of Unusual Volume Activity
S-Delta	Change In S-Score Over A Look back Period

Data Selection

Number	Industry	Tickers	Weight in SPY
1	Technology	XLK	26.2%
2	Health Care	XLV	15%
3	Financial	XLF	13.46%
4	Customer Discretionary	XLY	13.12%
5	Industrial	XLI	9.72%
6	Customer Staples	XLP	6.69%
7	Energy	XLE	6.01%
8	Utilities	XLU	2.78%
9	Real Estate	VNQ	2.61%
10	Materials	GDX	2.44%
11	Telecommunication services	VOX	1.97%

1. Data Distribution

Raw-S

Raw-S-Mean

S-dispersion
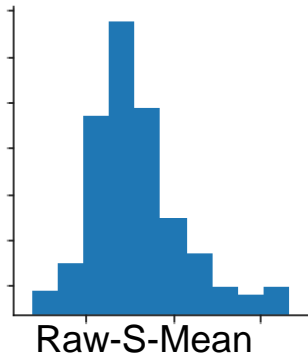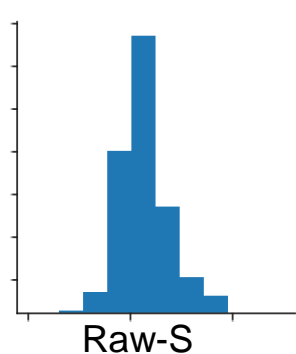
S-volatility

- Basically 4 types of distributions
- Raw-S-Mean is right-skewed: there is some data that is relatively too high. May consider a skewed distribution.
- S-dispersion is heavily left-skewed
- S-volatility is closer to a normal distribution, however there are some outliers which we will discuss on the next slide
- Raw-S has fat tails and many extreme values.
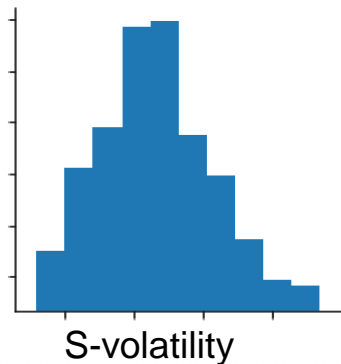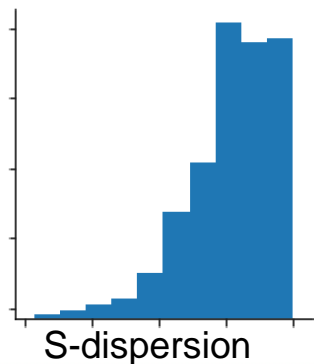
1. Data Distribution
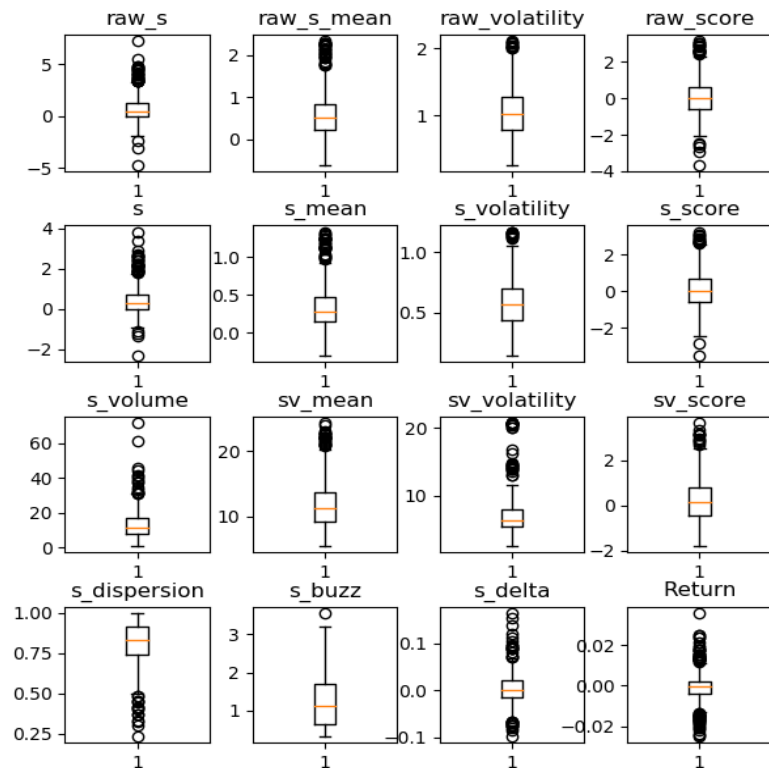
- Raw-S, as we saw earlier, has fat tails and a lot of extreme values.
- Raw-S must have high kurtosis, not normally distributed.
- As a result, may consider it log-normal or Leptokurtic

Raw-S
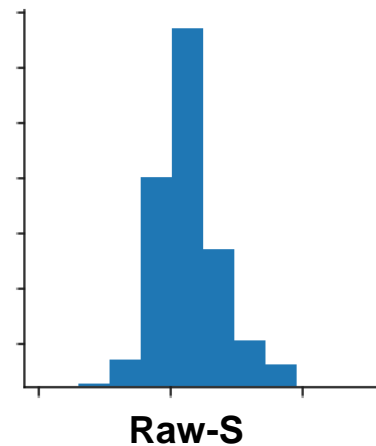
2. S-Volume Analysis

Top 10			<= 1.0			Avg. S-Vol (all data)	2.528
SPY	542.529		AAIT	1.000			
QQQ	151.662		ABCS	1.000		**Top 10 Avg. S-Vol**	751.06
IWM	104.438		AGOL	1.000		**Total**	3506.96
GLD	92.104		AXIT	1.000		*% of Total*	21.42%
USO	79.901		AZIA	1.000			
DIA	73.21		BGU	1.000		**Top 25 Avg. S-Vol**	1136.68
IBB	62.836		BONO	1.000		**Total**	3506.96
TLT	55.129		BRAF	1.000		*% of Total*	32.41%
GDX	45.574		BRIS	1.000			
UVXY	45.411		BRXX	1.000		**Lowest 166 Avg. S-Vol**	166.000
VXX	40.798		BRZS	1.000		**Total**	3506.96
OIL	38.527		BSCF	1.000		*% of Total*	4.73%
NUGT	34.452		BSJF	1.000			
XLF	33.288		BSR	1.000		**Top 25 (incl. SPY)**	1679.21
GBTC	31.501		BVT	1.000		**Total**	4049.48
XBI	30.323		CFT	1.000		*% of Total*	41.47%

- Average S-Volume 2015 -2017
- Top 25 Tickers (not shown): XLE, SLV, UWTI, EEM, USD, FXI, UUP, UNG, DUST

2. S-Volume Analysis
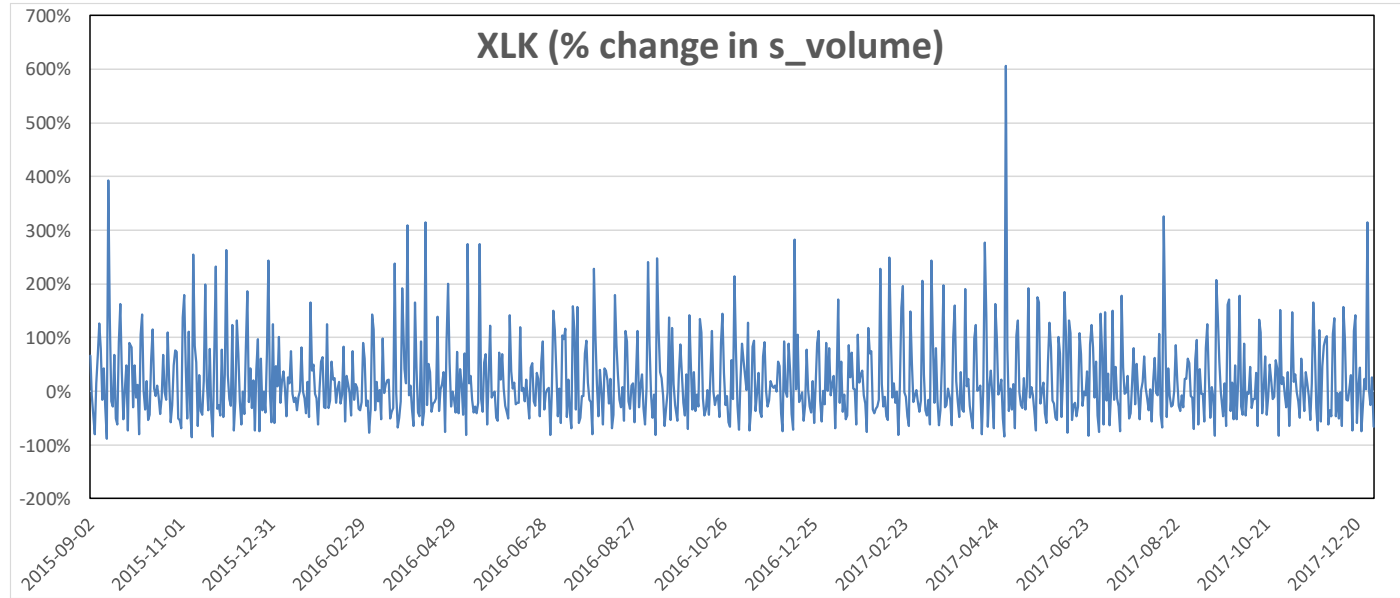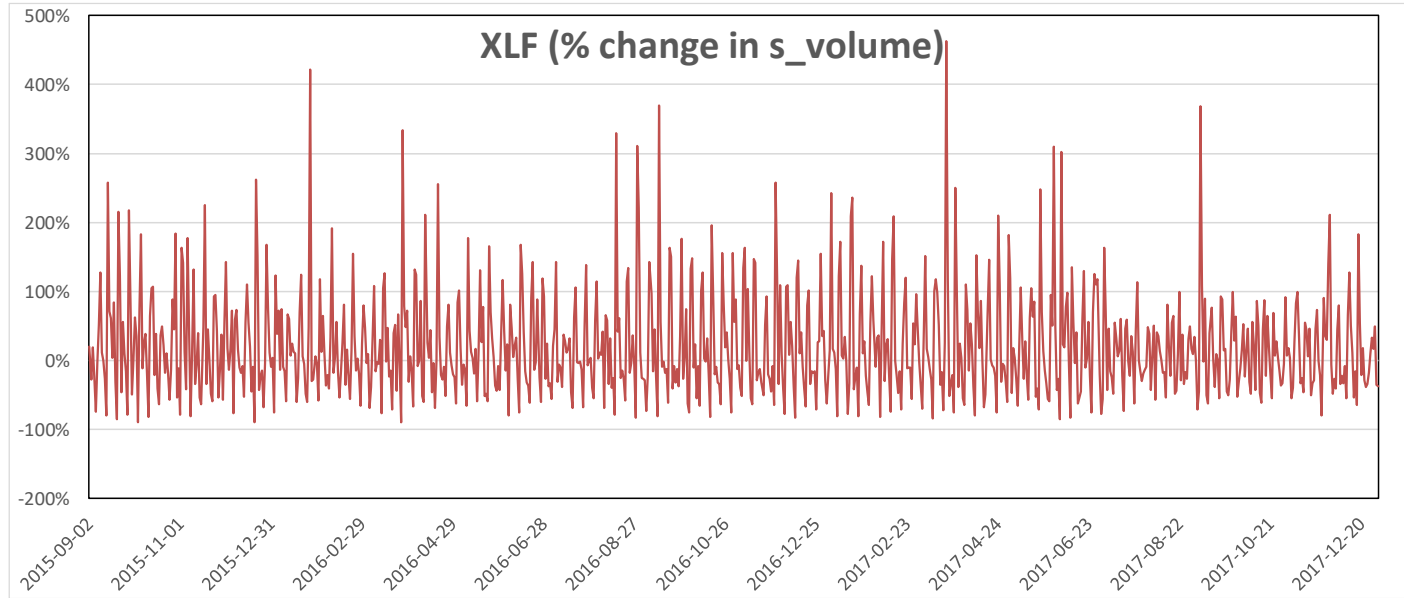
XLK (% change in s_volume)

2. S-Volume Analysis
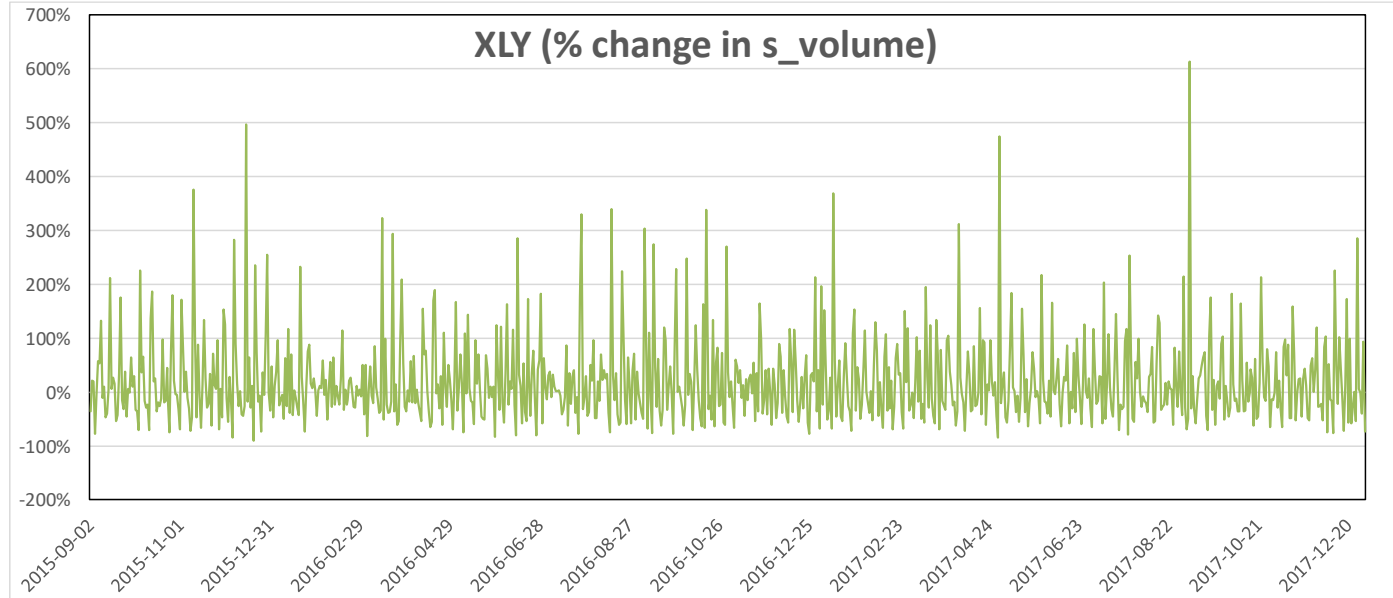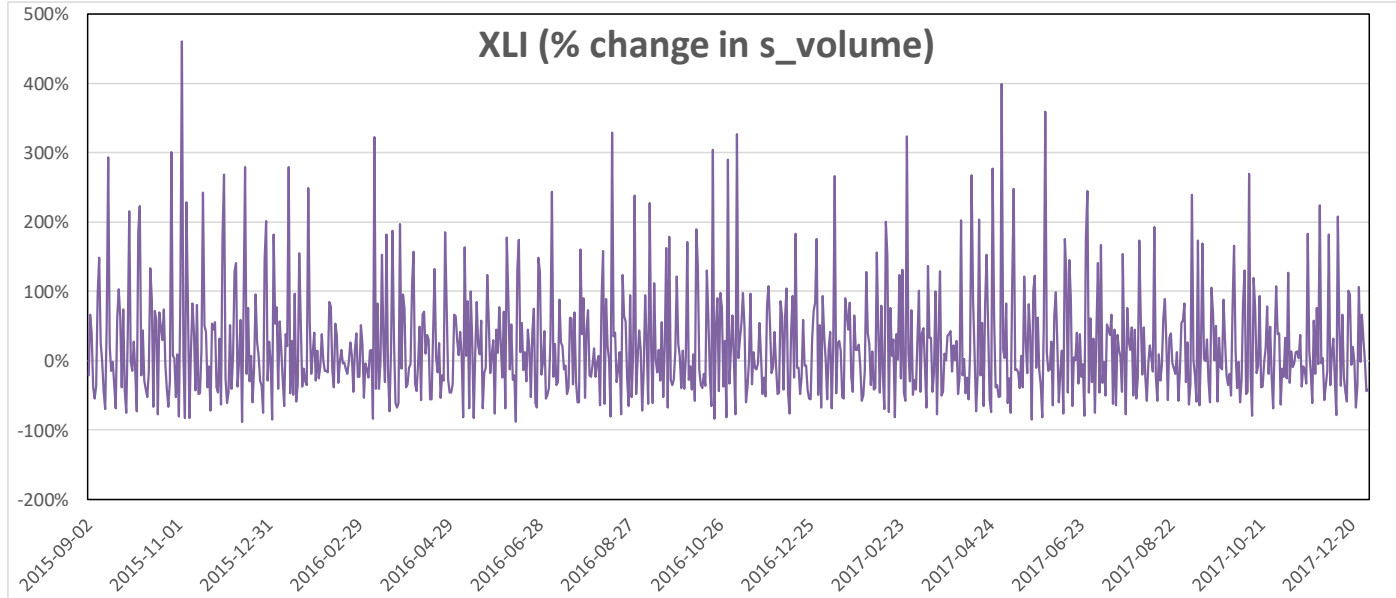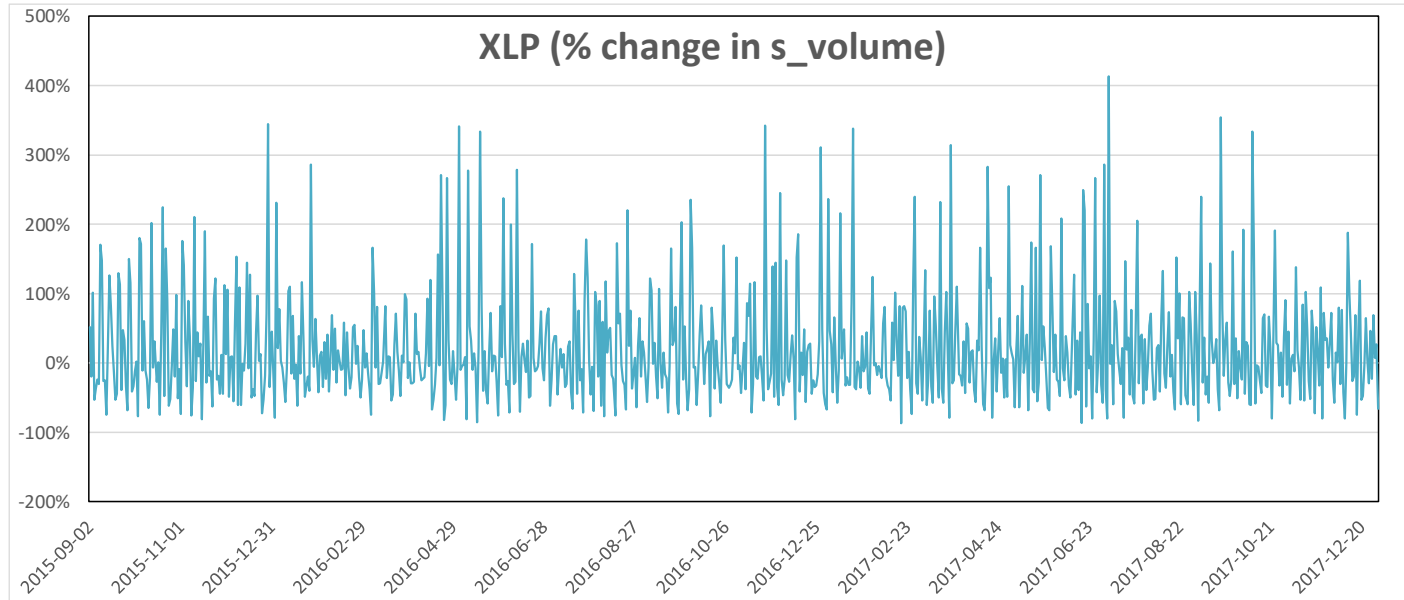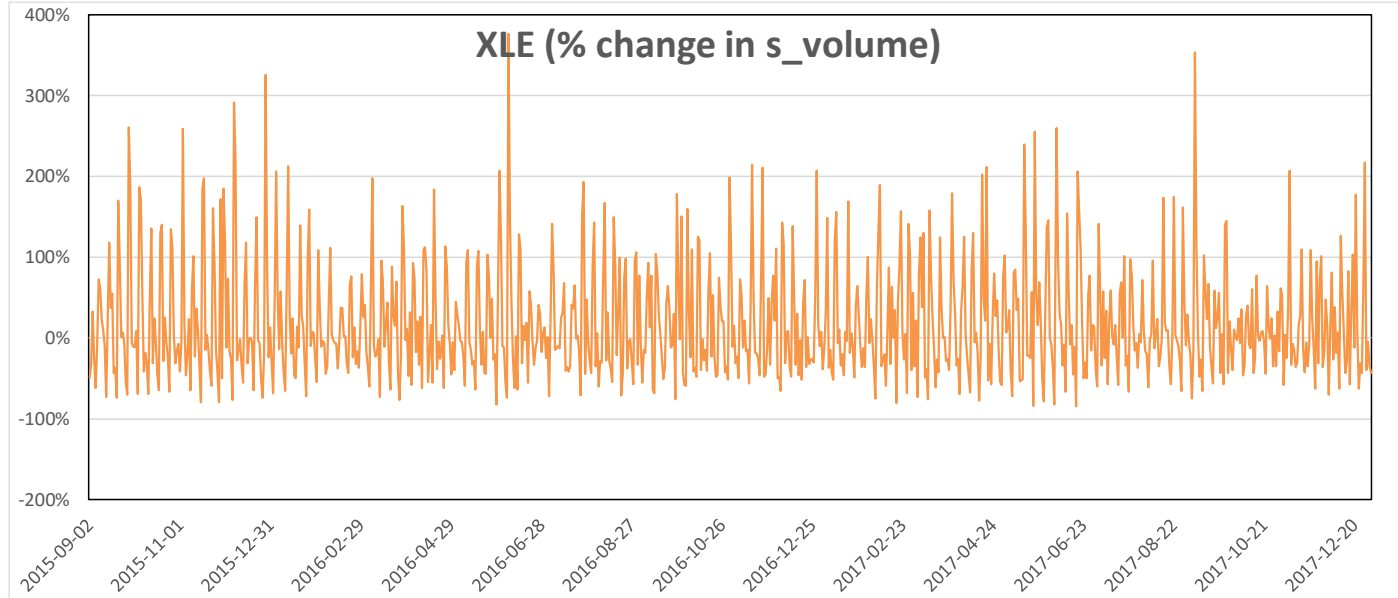

XLF (% change in s_volume)

2. S-Volume Analysis

2. S-Volume Analysis

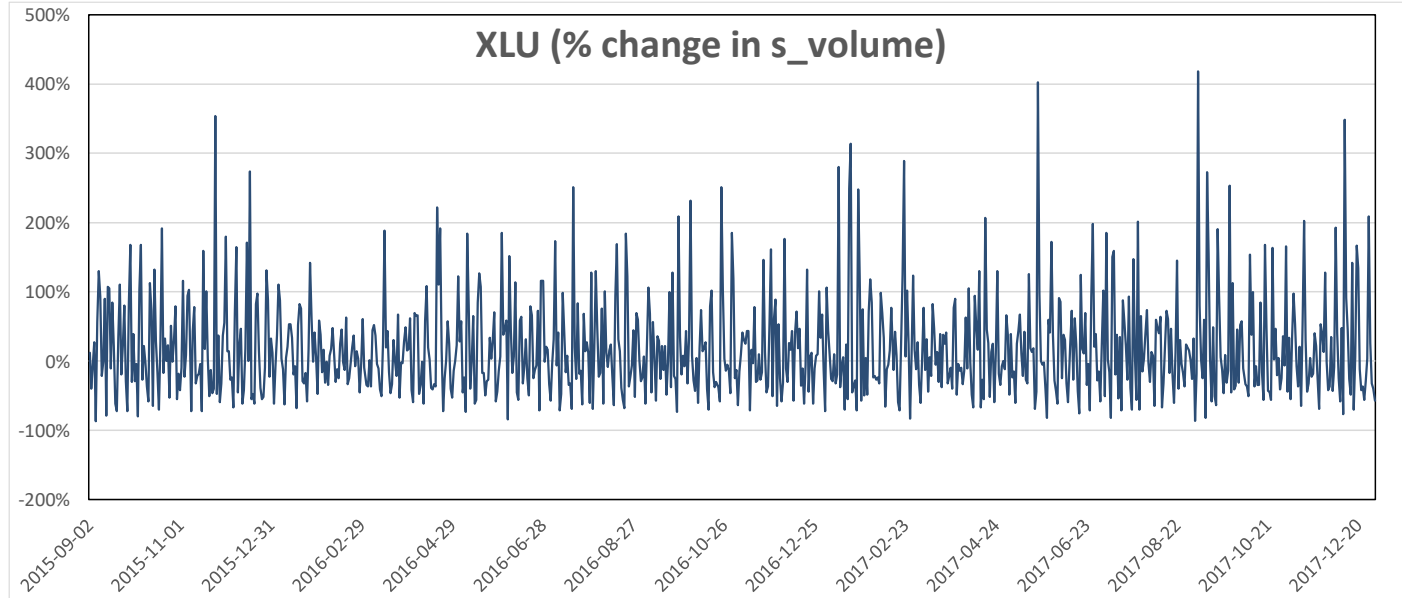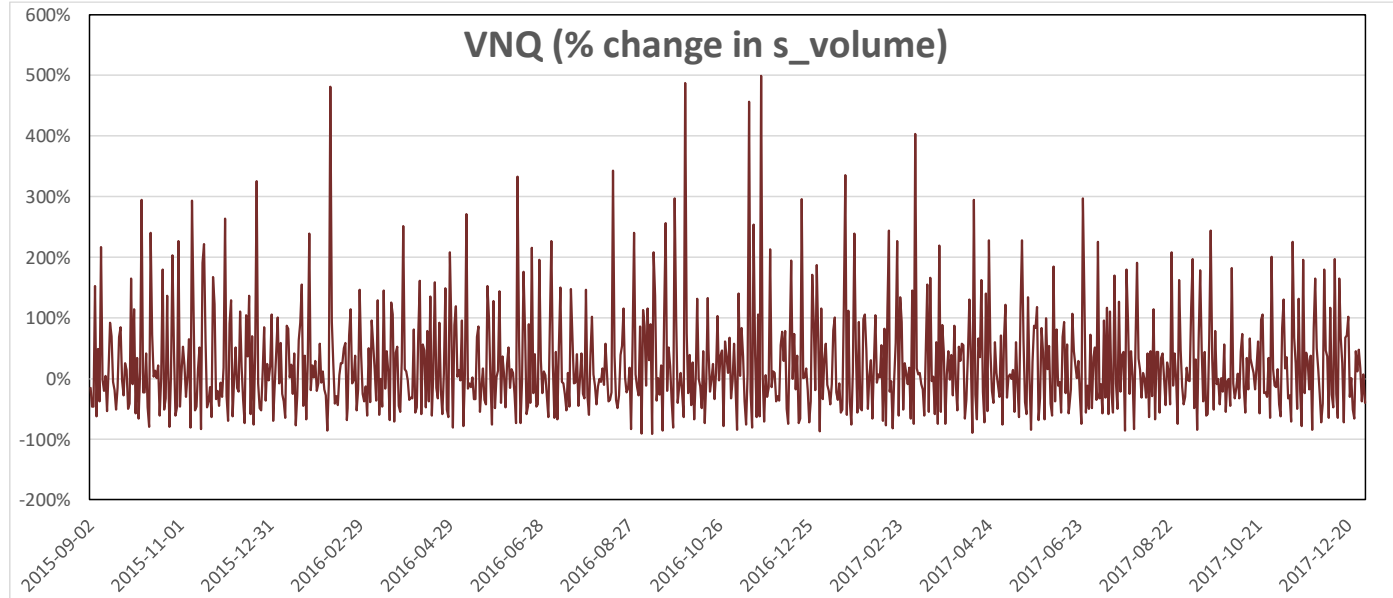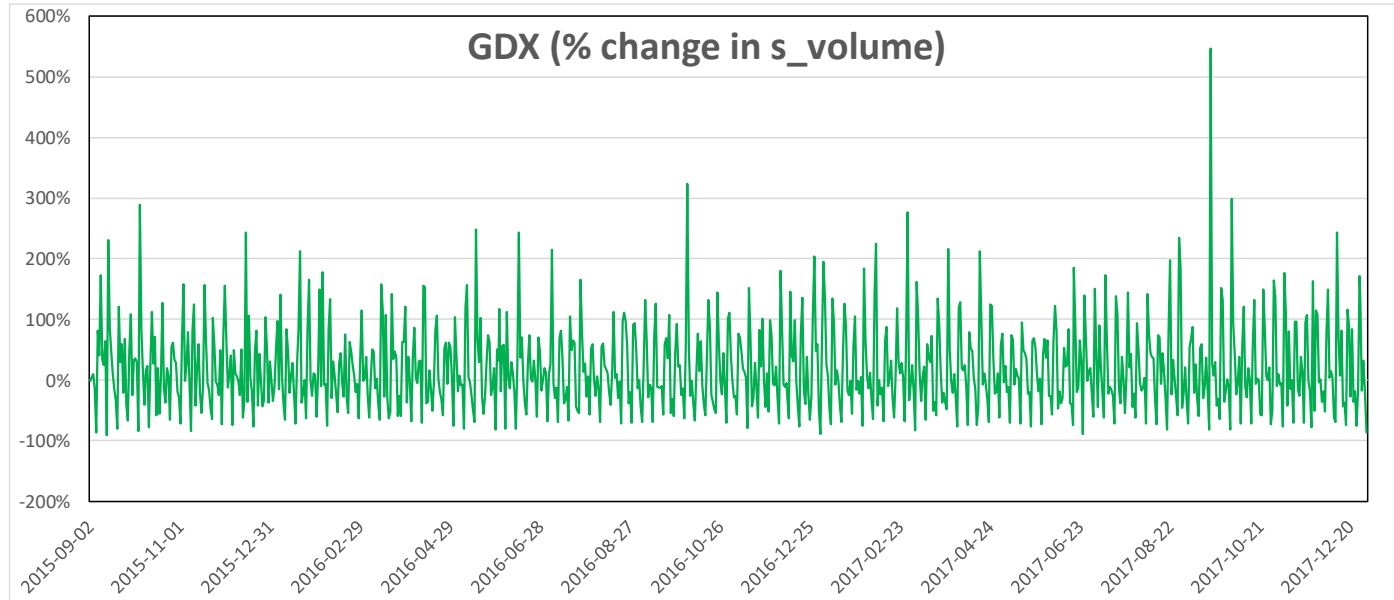# 2. S-Volume Analysis

2. S-Volume Analysis

XLE (% change in s_volume)

2. S-Volume Analysis

XLU (% change in s_volume)

2. S-Volume Analysis

VNQ (% change in s_volume)

2. S-Volume Analysis

GDX (% change in s_volume)

3. S-Score Metrics

```
df_open_score.head(100)
```

	s-volatility	s-score	s-buzz	s-delta	Date	Time
35	15.706	0.349	1.670	0.031	2015-09-01	08:55:00
129	15.578	-0.458	1.715	0.041	2015-09-02	08:55:00
225	16.194	1.335	0.584	0.000	2015-09-03	08:55:00
321	16.127	0.405	1.086	-0.011	2015-09-04	08:55:00
417	16.355	-0.492	0.644	-0.007	2015-09-05	08:55:00
...	...	...	...	...	...	...
9115	9.208	1.241	1.770	-0.068	2015-12-05	08:55:00
9210	9.077	-0.942	0.381	0.065	2015-12-06	08:55:00
9306	9.243	-1.008	0.660	-0.025	2015-12-07	08:55:00
9402	9.400	-0.730	2.033	-0.007	2015-12-08	08:55:00
9498	8.564	0.886	1.532	0.060	2015-12-09	08:55:00

- S-Score is the normalized value of S-Volume
- S-Volatility: 20-day moving average of "S" (exponentially-weighted sentiment estimate)
- S-Buzz is the measurement of unusual volume activity

3. S-Score(Close) – Close to Close
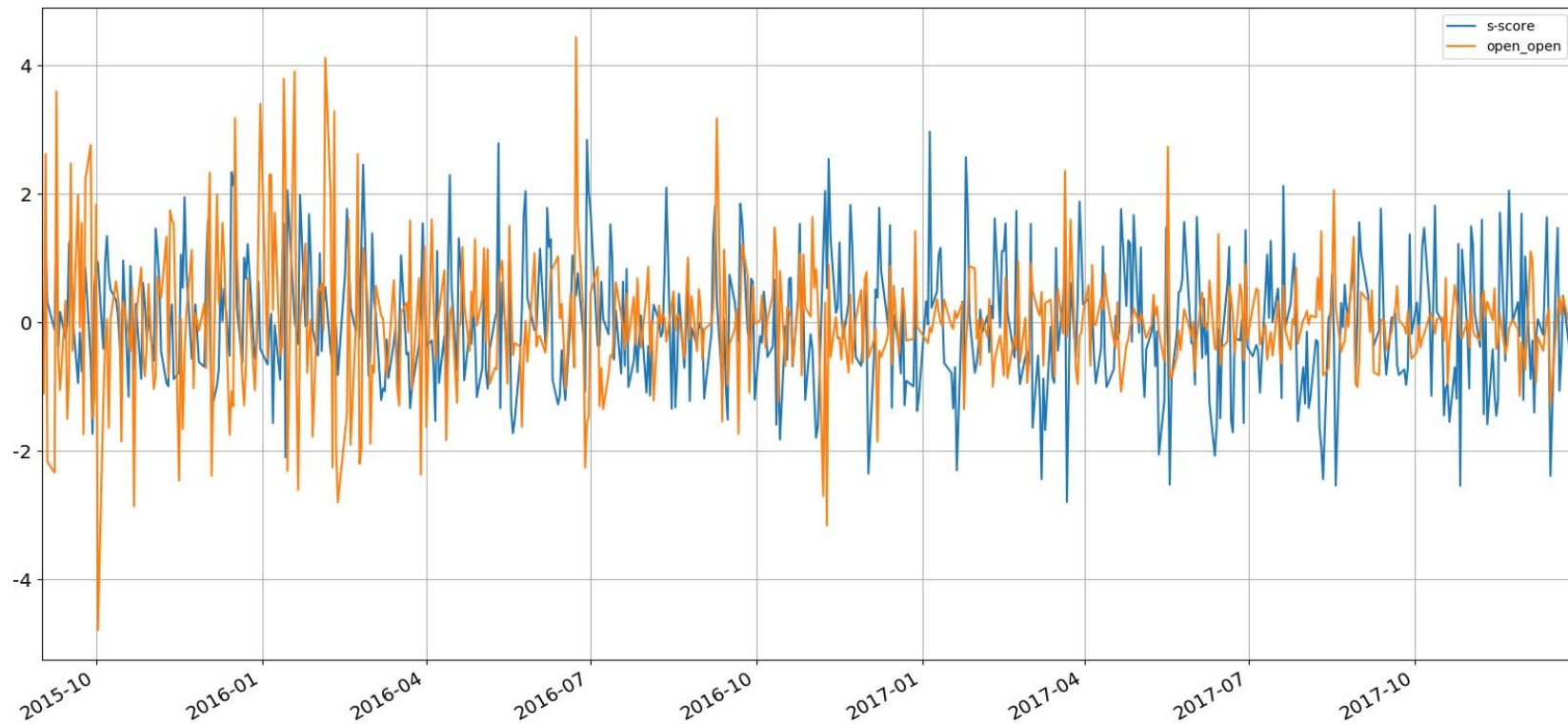
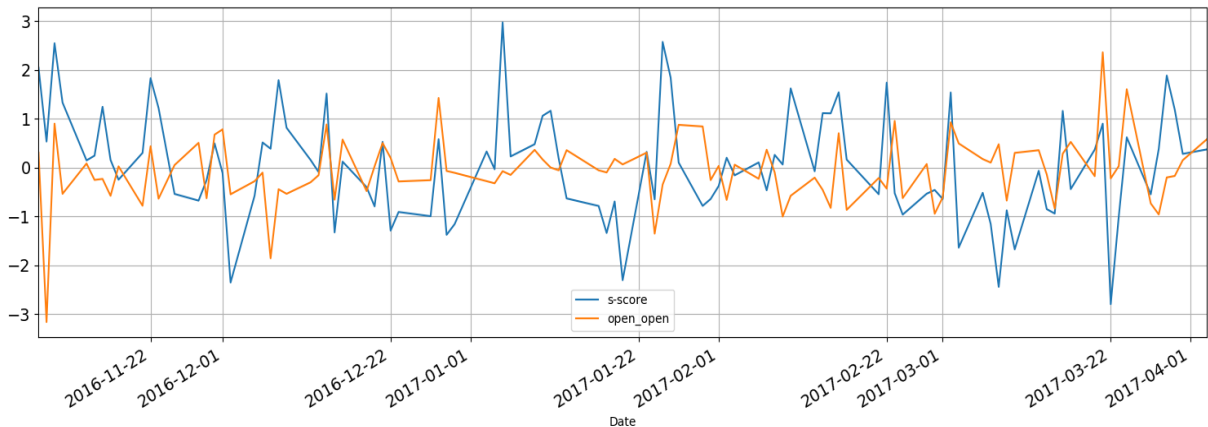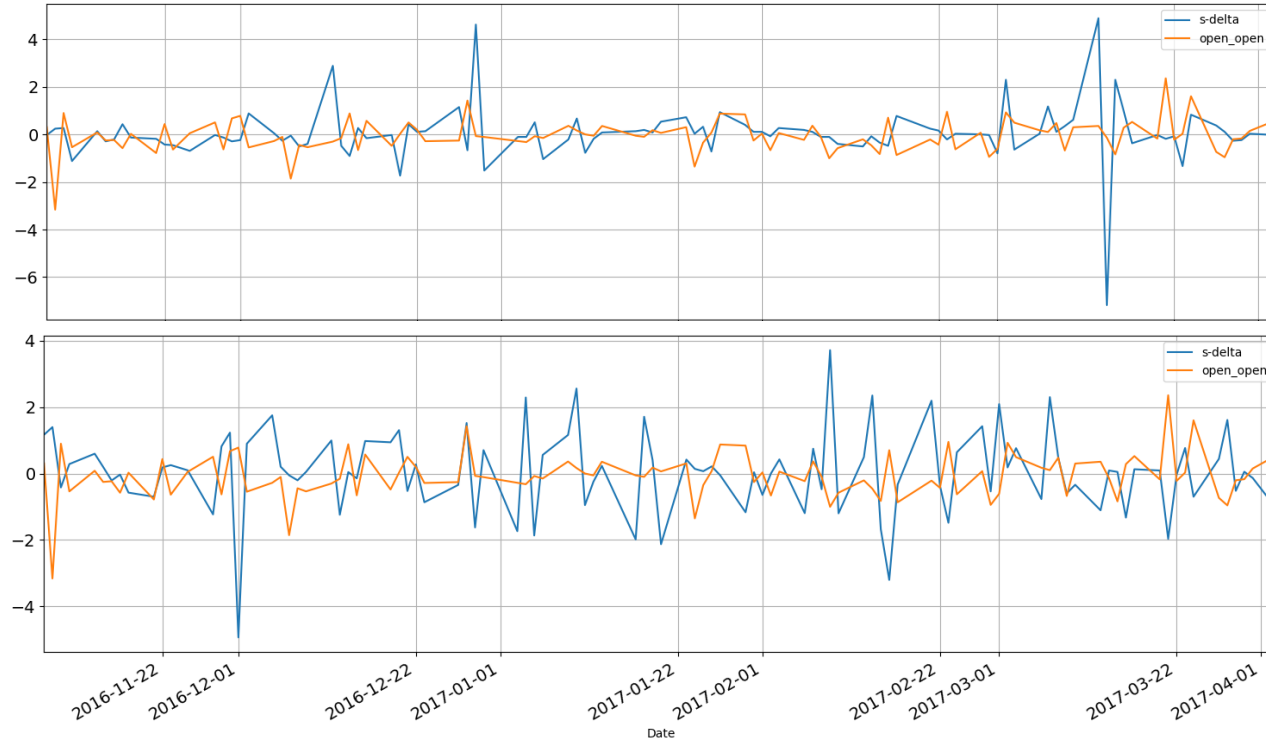# 3. Lets zoom in…
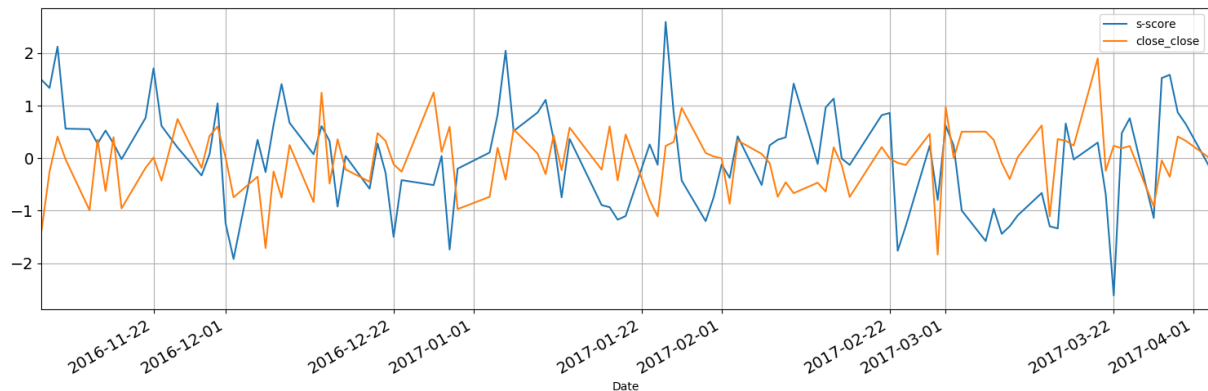
- Notice that some peaks in the S-Score is followed by a peak in the returns for both open-open and close-close

3. S-Delta: Noisy?

- S-Delta measures the change in S-Score over a lookback period
- Does not seem to give any glaring indicators or helpful information

4. Stationarity check

```
Results of Augment Dickey-Fuller Test:
Test Statistic                    -7.216111e+00
p-value                            2.171054e-10
#Lags Used                         0.000000e+00
Number of Observations Used        8.400000e+01
Critical Value (1%)               -3.510712e+00
Critical Value (5%)               -2.896616e+00
Critical Value (10%)              -2.585482e+00
dtype: float64
```

- Use ADF test Method.
- The T-stats should be smaller than 0.05 at 95% confidence level.
- in this case it is so small that we can say it's definitely stationary.
- p-value should be as small as possible as well, which is true.

4.Stationarity check

	raw_s	raw_s_mean	raw_volatil	raw_score	s	s_mean	s_volatility	s_score	s_volume	sv_mean	sv_volatilit	sv_score	s_dispersio	s_buzz	s_delta	Return
XLK	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
XLV	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
XLF	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
XLY	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
XLI	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
XLP	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
XLE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
XLU	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
VNQ	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
GDX	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
VOX	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
SPY	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE

- Use ADF test on every data-set and every factor. This is the result.

Conclusion

- Proved most of data is stationary so we can potentially use time-series models
- Depending on importance of S-volume, **finish**