ORIGINAL ARTICLE

Journal Section

Probing the fear gauge

Krishnan Raman[1]* | Anqi Wang[1]† | Jianing Li[2]* | Qianli Xu[2]†

[1] 2nd year, Dept of Statistics, Purdue U

[2] 2nd year, Supply Chain and Operations Management, Purdue U

Correspondence
Krishnan Raman
Email: raman23@purdue.edu

The Fear Gauge is the most watched metric on Wall Street. We provide a statistical methodology to probe the efficacy of the fear gauge vis-a-vis the three largest exchange traded funds on the market, SPY, DIA & QQQ.

KEYWORDS

ETF, VIX

1 | INTRODUCTION: ETFs & the VIX

Exchange Traded Funds such as SPY (S&P 500), DIA (Dow Jones Industrials), QQQ (Nasdaq Tech Index), have accumulated over a trillion dollars of assets under management since they were first introduced. Their enormous popularity is driven by a bunch of investing trends - increased availability of passive, hands-off, low-fee funds from Blackrock, State Street & Vanguard, a stable risk profile historically averaging 7% annually over a long time-horizon & consumer trust in broad indices as opposed to individual stock picks. However, ETFs are NOT a safe haven. ETFs do lose money when traded speculatively, especially over shorter time duration. During times of recession, ETFs have lost over 20% at times! As such, investors are interested in keeping tabs on the average ETF's price. However, measuring the volatility (standard deviation) of an ETF over a time horizon is a challenging proposition. An ETF is essentially a weighted sum of stocks. Eg. SPY, is a weighted sum of 500 stocks, DIA has 30 components & QQQ, 100. These component stocks rapidly change price every millisecond during the trading session. It is unclear how to capture these changing price deltas on 100s of timeseries & their standard deviations in a single number.

In 1993, Robert E. Whaley, currently a Management Professor at Vanderbilt U, wrote a seminal treatise in the Journal of Derivatives. Titled "***Derivatives on Market Volatility: Hedging Tools Long Overdue***", the paper outlines a "hedging tool" to systematically compute the volatility of an ETF into a single number. This tool went on to acquire outsize proportions and became the single most traded derivative in market history! Dr. Whaley called it the volatility index aka the VIX.

With the colossal success of the VIX came its critics. Numerous studies dispute the correlation of VIX with the ETF. In a widely cited paper ***"How Good is the VIX as a Predictor of Market Risk?"***, Dr. Clemens Kownatzki, a Professor of Finance at Pepperdine U, alleges "VIX consistently over-estimates actual volatility in normal times...but underestimates volatility in times of market crashes...making it unsuitable for risk-management". Current literature either supports the VIX as a sound measure of the standard deviation of SPY, or fervently believes it is deficient in various statistical metrics.

We probe the efficacy of the VIX in order to answer a few simple queries:

1. Does the VIX statistically correlate with the ETF? Is it a genuine fear index?
2. Can probabilities of market decline be inferred from the movement of the VIX?
3. Does "Market Risk Contagion" exist i.e. are the ETFs themselves strongly correlated, even though the component holdings of each ETF are largely disjoint?
4. Suppose there is a major recession and the VIX peaks. Approximately how long(days/weeks/months) does it take for things to cool down?

2 | METHODS: Contingency Tables

To a statistician, an ETF on a given day is summarized by its closing price. Similarly, the VIX is a single number that summarizes the volatility of the ETF. We obtain voluminous data on the SPY, the QQQ and the DIA, for a period of eighteen years, starting February 2001 until December 2018. Each ETF is associated with its own VIX. Thus, we have six distinct time series for a total of 4500 days. A data snapshot is shown below.

```
> head(read.csv(myfile))
     Date   vix   vxn   vxd spy.PERMNO spy.PRC spy.OPENPRC dia.PERMNO dia.PRC dia.OPENPRC qqq.PERMNO qqq.PRC qqq.OPENPRC
1 2/2/01 21.95 54.89 19.94      84398  134.80      137.40      85765  108.63      109.80      86755   61.55       64.94
2 2/5/01 22.19 55.85 19.98      84398  135.79      134.80      85765  109.80      108.87      86755   61.50       61.19
3 2/6/01 21.98 53.68 19.57      84398  135.39      135.30      85765  109.58      109.30      86755   61.60       61.49
4 2/7/01 21.67 54.41 19.20      84398  134.69      134.72      85765  109.62      109.49      86755   60.60       60.51
5 2/8/01 21.46 54.66 19.23      84398  133.12      134.80      85765  108.73      109.85      86755   58.79       60.90
6 2/9/01 22.03 55.85 19.10      84398  131.84      133.35      85765  108.02      108.50      86755   56.40       58.25
```

Figure 1. Time Series Snapshot

If there existed a contingency table that succinctly captured all that we know about a VIX & its ETF in a simple 2x2 matrix, any association would pop out right away! While there isn't one, we propose a scheme to <u>discretize a continuous variable</u>, essentially moving from continuous price data to ordinal counts. Prices are not a suitable scale to do math in, since they vary in a very wide range.

However, time-windowed returns, such as the Daily Return, defined as

$$Return = \frac{Price[tomorrow] - Price[today]}{Price[today]}$$

are nicely bounded. For instance, the Weekly VIX lies between [-0.5, 2.5], while the weekly ETF loses at-most 24% and gains 17%, lying in [-0.24, 0.17].

A 2x2 contingency table simply counts the number of occasions the VIX & the ETF move in tandem or in opposite directions.

	ETF -ve	ETF +ve
VIX -ve	399	1972
VIX +ve	1526	600

	ETF -ve	ETF +ve
VIX -ve	476	1902
VIX +ve	1453	664

	ETF -ve	ETF +ve
VIX -ve	424	1932
VIX +ve	1492	649

Figure 2. Contingency Tables of VIX-SPY, VXD-DIA, VXN-QQQ

Even before we conduct a formal test, it is obvious that the odds ratio is far from 1. We test for independence of rows & columns using Fisher's Exact Test & Pearson's Chi-square test.

Both tests confirm our intuition:

```
> fisher.test(spy2x2)

        Fisher's Exact Test for Count Data

data:  spy2x2
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.06875146 0.09202854
sample estimates:
odds ratio
0.07961536
> chisq.test(spy2x2)

        Pearson's Chi-squared test with Yates' continuity correction

data:  spy2x2
X-squared = 1380.2, df = 1, p-value < 2.2e-16
```

FIGURE 3. Tests of Independence

Thus, we have a rigorous confirmation over 18 years of data that the VIX & the ETF are strongly inversely correlated. Higher the VIX, lower the ETF & vice-versa.

3 | METHODS: Linear, Multinomial logit & Proportional Odds Models

Inferring market decline probabilities via linear regression with quadratic predictors:

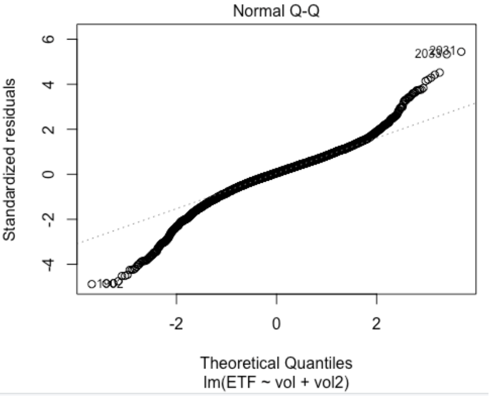
```
> summary(lmod2)

Call:
lm(formula = ETF ~ vol + vol2, data = SPYMdf)

Residuals:
      Min        1Q    Median        3Q       Max
-0.180965 -0.018209  0.002398  0.021303  0.203194

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0082005  0.0005813   14.11   <2e-16 ***
vol         -0.1558977  0.0026536  -58.75   <2e-16 ***
vol2         0.0369804  0.0023269   15.89   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03733 on 4471 degrees of freedom
Multiple R-squared:  0.521,     Adjusted R-squared:  0.5208
F-statistic:  2432 on 2 and 4471 DF,  p-value: < 2.2e-16
```



However, it is clear that the residuals are far from normal. To fit multinomial models, we pick a suitable 8x8 contingency table with weekly returns.

	ETF< -3%	-3% <ETF< -2%	-2% <ETF< -1%	-1% <ETF< 0%	0% <ETF< 1%	1% <ETF< 2%	2% <ETF< 3%	ETF> 3%
VIX< -3%	14	17	65	159	404	516	374	387
-3% <VIX< -2%	4	3	13	31	51	42	9	3
-2% <VIX< -1%	3	4	15	22	45	33	13	8
-1% <VIX< 0%	2	6	12	29	56	24	6	1
0% <VIX< 1%	3	8	19	38	51	31	6	3
1% <VIX< 2%	3	11	14	26	50	25	7	4
2% <VIX< 3%	9	12	10	40	47	13	6	2
VIX> 3%	421	236	304	372	230	78	34	13

The 8x8 contingency table tells us that ETFs & VIXs move inversely on a more granular scale. E.g. there were 421 occasions when the VIX rose over 3% & the ETF had negative 3% or lower returns! On 387 occasions, opposite scenario played out. Response ETF is a categorical variable, but we have 2 choices: we can treat predictor VIX as continuous, or as a factor. Which model is preferable?

```
> modm <- multinom(factor(ETF) ~ vol,weights = freq)
# weights:  24 (14 variable)
initial  value 9351.248613
iter  10 value 8385.006338
iter  20 value 7923.828537
final  value 7923.646098
converged
```

```
> modmf <- multinom(factor(ETF) ~ factor(vol),weights = freq)
# weights:  72 (56 variable)
initial  value 9351.248613
iter  10 value 8154.383644
iter  20 value 7874.286156
iter  30 value 7838.262410
iter  40 value 7833.822729
iter  50 value 7833.403903
final  value 7833.389598
converged
```

```
> pchisq(modm$dev-modmf$dev,modmf$edf-modm$edf,lower=F)
[1] 4.291821e-19
```

By far, the VIX as factor model is vastly better than VIX as continuous variate. We can infer probability of ETF decline as VIX rises, using multinomial model with parameters as below:

```
> summary(modmf)
Call:
multinom(formula = factor(ETF) ~ factor(vol), weights = freq)

Coefficients:
  (Intercept) factor(vol)2 factor(vol)3 factor(vol)4 factor(vol)5 factor(vol)6 factor(vol)7 factor(vol)8
2   0.1934298   -0.4799101   0.09348216   0.89961785    0.7906124  1.107290677   0.09931089   -0.7716594
3   1.5385719   -0.3619464   0.07156182   0.24365553    0.3042507 -0.003193564  -1.43440512   -1.8638244
4   2.4326480   -0.3881684  -0.44111190   0.23296076    0.1028346 -0.277868303  -0.94246623   -2.5564181
5   3.3648479   -0.8221259  -0.65861358  -0.04148503   -0.5349386 -0.555854158  -1.71296674   -3.9690795
6   3.6095466   -1.2611193  -1.21292471  -1.13283936   -1.2767349 -1.493406623  -3.24355173   -5.2955649
7   3.2877100   -2.4809203  -1.82072446  -2.19843681   -2.5963826 -2.443317897  -3.69188691   -5.8053181
8   3.3223758   -3.6074760  -2.35841017  -4.01782323   -3.3238787 -3.037668365  -4.83396985   -6.7999847

Std. Errors:
  (Intercept) factor(vol)2 factor(vol)3 factor(vol)4 factor(vol)5 factor(vol)6 factor(vol)7 factor(vol)8
2   0.3614341    0.8438035    0.8444492    0.8902986    0.7663799    0.7437296    0.5696796    0.3704670
3   0.2949401    0.6428001    0.6972894    0.8162240    0.6870784    0.7004097    0.5459721    0.3043914
4   0.2791109    0.5995895    0.6753068    0.7798691    0.6608167    0.6696750    0.4625625    0.2880403
5   0.2721955    0.5856878    0.6550089    0.7666558    0.6528010    0.6528156    0.4543301    0.2842765
6   0.2712066    0.5888110    0.6607295    0.7816754    0.6620015    0.6675562    0.5114689    0.2979133
7   0.2725599    0.6594630    0.6955185    0.8584372    0.7571327    0.7410011    0.5930221    0.3257607
8   0.2723927    0.8093884    0.7306269    1.2505598    0.8599857    0.8099721    0.8298731    0.3917937

Residual Deviance: 15666.78
AIC: 15778.78
```

Rather than treat VIX & ETF counts as nominals, since there exists a natural ordering from -3% to 3%, a cumulative logit i.e. proportional odds model is the logical candidate:

```
polr(formula = factor(ETF) ~ factor(vol), weights = freq)

Coefficients:
              Value Std. Error t value
factor(vol)2 -1.351     0.1445  -9.343
factor(vol)3 -1.213     0.1545  -7.854
factor(vol)4 -1.611     0.1518 -10.614
factor(vol)5 -1.718     0.1448 -11.867
factor(vol)6 -1.651     0.1549 -10.663
factor(vol)7 -2.074     0.1550 -13.378
factor(vol)8 -3.366     0.0767 -43.891

Intercepts:
    Value    Std. Error t value
1|2  -4.5393   0.0804    -56.4340
2|3  -3.8503   0.0732    -52.6085
3|4  -3.0611   0.0664    -46.0976
4|5  -1.9942   0.0571    -34.9274
5|6  -0.6871   0.0465    -14.7629
6|7   0.4612   0.0455     10.1407
7|8   1.4445   0.0555     26.0334

Residual Deviance: 15752.48
AIC: 15780.48
```

The AICs are virtually identical, and *polr* parameters are parsimonious & simpler to interpret, as its easier to work with monotonically increasing cumulative probabilities.

4 | METHODS: Estimating Market Risk Contagion with cross-correlated contingency tables

A popular misconception is to assume one is sufficiently hedged if funds are distributed among the Dow & the Nasdaq. Since DIA is primarily industrial equities, & QQQ is newfangled tech stocks, the two sets are disjoint. When tech does well, consumer staples lag. However, it is rather easy to show a so-called "regime change" that has created a clear correlation among indices. Indeed, not only are the DIA & QQQ & the SPY massively correlated, knowing the VIX for the SPY successfully predicts the DIA, something it was never intended to do! The 2 contingency tables below narrate this tale.

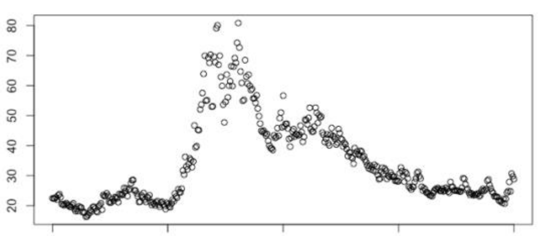
	ETF -ve	ETF +ve
VIX -ve	462	1916
VIX +ve	1462	655

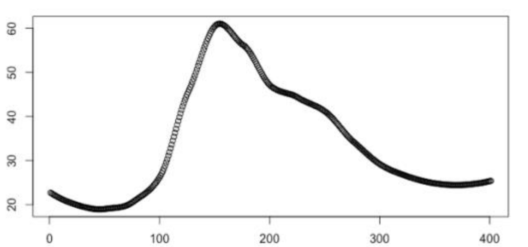
On the left, VIX, the SPY's indicator, predicts DIA! On the right, we show a strong correlation between "disjoint" indices DIA & QQQ

	DIA -ve	DIA +ve
QQQ -ve	2023	355
QQQ +ve	331	1786

5 | METHODS: Estimating return to normality in a crisis via Survival Analysis, Loess.

When a recession is imminent or ongoing, the VIX spikes. These are rather sharp and prolonged, as the market experiences a dramatic & painful downturn. Often, traders have to estimate how long will the turbulence last. A popular technique is to treat the VIX as a sort of mortality curve, and as it "dies off", one estimates how many more "patients" are at risk, until we return to a normal VIX level. We demonstrate with an example pertaining to the so-called Great Recession of 2008 triggered by the collapse in real estate markets. In the USA, it is estimated that the GDP downturn lasted from Q3 2008 to Q2 2009. However, the markets are a leading indicator, so they revert to normalcy in about 200 trading days.



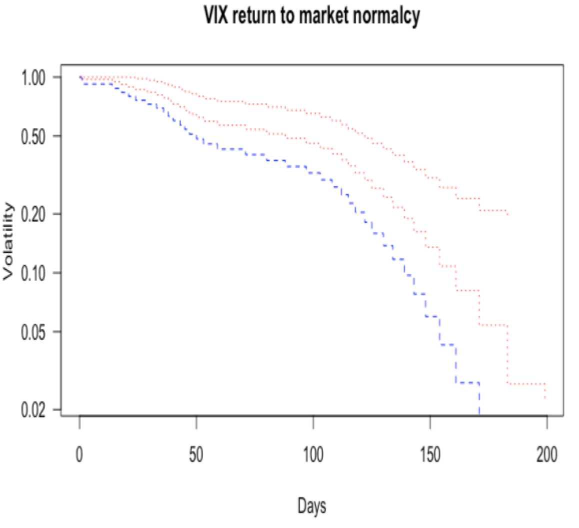


We notice the VIX was treading water in the 20s, and suddenly spiked to 80. The market data is rather noisy, and we experiment with a loess to smooth out the variation.

Now that we have a monotonically decreasing function from the VIX peak, we fit a survival curve to estimate how long the recession lasts.

```
> summary(fit)
Call: survfit(formula = Surv(time, rep(1, 37)) ~ rep("control",
    conf.type = "log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1     37       1   0.9730  0.0267      0.92210        1.000
   14     36       1   0.9459  0.0372      0.87582        1.000
   18     35       1   0.9189  0.0449      0.83504        1.000
   21     34       1   0.8919  0.0510      0.79725        0.998
   24     33       1   0.8649  0.0562      0.76144        0.982
   29     32       1   0.8378  0.0606      0.72710        0.965
   33     31       1   0.8108  0.0644      0.69394        0.947
   36     30       1   0.7838  0.0677      0.66176        0.928
   38     29       1   0.7568  0.0705      0.63041        0.908
   40     28       1   0.7297  0.0730      0.59979        0.888
   43     27       1   0.7027  0.0751      0.56984        0.867
   45     26       1   0.6757  0.0770      0.54049        0.845
   47     25       1   0.6486  0.0785      0.51170        0.822
   50     24       1   0.6216  0.0797      0.48345        0.799
   53     23       1   0.5946  0.0807      0.45569        0.776
   59     22       1   0.5676  0.0814      0.42842        0.752
   71     21       1   0.5405  0.0819      0.40162        0.728
   80     20       1   0.5135  0.0822      0.37527        0.703
   88     19       1   0.4865  0.0822      0.34938        0.677
   97     18       1   0.4595  0.0819      0.32394        0.652
  103     17       1   0.4324  0.0814      0.29895        0.626
  108     16       1   0.4054  0.0807      0.27442        0.599
```



6 | RESULTS

We’ve eschewed traditionally intensive time-series techniques such as GARCH/ARIMA in favor of simpler cumulative logit models to convincingly demonstrate the negative correlation of market volatility, as exhibited by the VIX indicator, with popular market ETFs. Further, we show the risk of market contagion is real, as disjoint ETFs have gotten correlated over the years. Using a novel survival analysis approach, we estimate return to normalcy in events of market crises.

7 | DISCUSSION

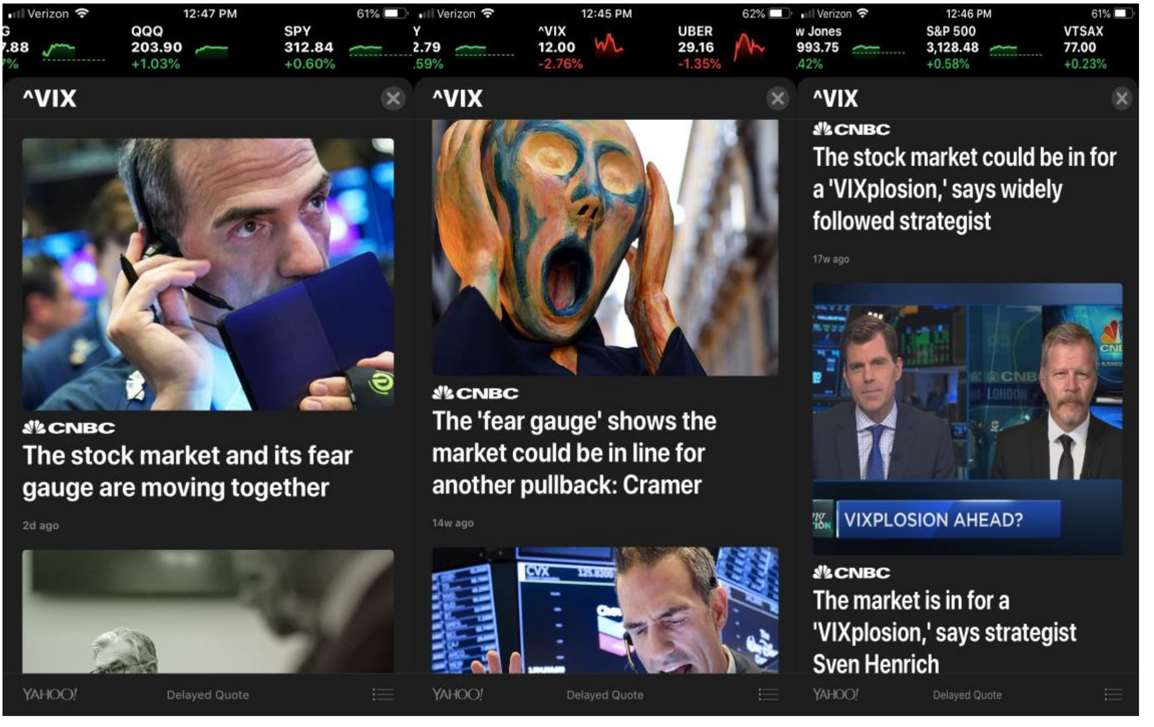
Building a predictive VIX-ETF model turns out to be much harder than foreseen. Despite experimenting with varying time lags, we were unable to build a truly predictive model that would indicate market direction based on past week’s volatility. Concurrent models show much promise. Traders generally use GARCH to build predictive time-series models, since contingency tables are too simplistic in this scenario.

8 | REFERENCES

2001-2018 Market Data was obtained via subscription from the excellent Wharton Research Data Service. In addition to CNBC & Yahoo Finance, we consulted the following papers in our attempt to understand the strengths & deficiencies of the VIX.

- *How Good is the VIX as a Predictor of Market Risk? Clemens Kownatzki*
- *Derivatives on Market Volatility: Hedging Tools Long Overdue: Robert E Whaley*

9 | APPENDIX




```
rm(list=ls())
library(lubridate)
library(readr)
library(ggplot2)
library(nnet)
library(MASS)

# make an nx2 matrix of ETF-vol pair returns, bucketed by time Window
mkReturns<-function(filename, volIndex, underlyingIndex, timeWindow) {
 mat<-data.matrix(read.csv(filename))
 retmat<-matrix(NA,nrow=nrow(mat),ncol=2)
 t = timeWindow
 for (r in 1:(nrow(mat)-t)) {
  c = volIndex
  retmat[r,1] = unname((mat[r+t,c] - mat[r,c])/mat[r,c])
  c = underlyingIndex
  retmat[r,2] = unname((mat[r+t,c] - mat[r,c])/mat[r,c])
 }
 retmat<-na.omit(retmat)
 colnames(retmat) <-c("VIX","ETF" )
 return(retmat)
}

# make a predictive nx2 matrix of ETF-vol pair returns, bucketed by time Window
```

```
# the volatility will lag behind the ETF by past times the timeWindow
mkPredictiveReturns<-function(filename, volIndex, underlyingIndex, pastWindow, futureWindow) {
 mat<-data.matrix(read.csv(filename))
 retmat<-matrix(NA,nrow=nrow(mat),ncol=2)
 t = pastWindow
 t2 = pastWindow + futureWindow
 for (r in 1:(nrow(mat)-t2)) {
  c = volIndex
  retmat[r,1] = unname((mat[r+t,c] - mat[r,c])/mat[r,c])
  c = underlyingIndex
  retmat[r,2] = unname((mat[r+t2,c] - mat[r+t,c])/mat[r+t,c])
 }
 retmat<-na.omit(retmat)
 return(retmat)
}

# border = cutoff at border of the cont table, such as say 20%
# increment = stepping between columns in cont table, such as say 5%
# Makes a contingency table from -border to +border with bucket size = incrementpp
mkContTable<-function(mymatrix, border, increment) {
 # equal number of rows & cols
 rows = 2+ (2*border/increment)
 table<-matrix(0,nrow=rows, ncol=rows)
 # walk thru return matrix, populate contingency table
 for(r in 1:nrow(mymatrix)) {
  e1 = mymatrix[r,1] #vix
  e2 = mymatrix[r,2] #etf
  i1 = ceiling(abs(e1)/increment)
  i2 = ceiling(abs(e2)/increment)
  if (e1 > 0) r1 = min(rows, (rows/2) + i1)
  else r1 = max(1, (rows/2) - i1 + 1)

  if (e2 > 0) c1 = min(rows, (rows/2) + i2)
  else c1 = max(1, (rows/2) - i2 + 1)

  table[r1,c1] = table[r1,c1] + 1
  #print(paste(e1,",",e2,",",i1,",",i2,",",r1,",",c1))
 }
 return(table)
}

# Collapse a large contingency table into a 2x2 table
collapseTable<- function(bigtable) {
 twobytwo = matrix(0,nrow=2,ncol=2)
 r = nrow(bigtable)
 r2 = r/2
 sum = 0
 for(i in 1:r2) {
  for(j in 1:r2) {
   sum = sum + bigtable[i,j]
   #print(paste(i, ",", j,",", bigtable[i,j]))
  }
 }
 twobytwo[1,1] = sum

 sum = 0
 for(i in 1:r2) {
  for(j in (1+r2):r) {
```

```
    sum = sum + bigtable[i,j]
   }
  }
  twobytwo[1,2] = sum

  sum = 0
  for(i in (1+r2):r) {
   for(j in 1:r2) {
    sum = sum + bigtable[i,j]
   }
  }
  twobytwo[2,1] = sum
  twobytwo[2,2] = sum(bigtable) - sum(twobytwo)
  colnames(twobytwo) <- c("ETF -ve", "ETF +ve")
  rownames(twobytwo) <- c("VIX -ve", "VIX +ve")
  return(twobytwo)
}

testTable<-function() {
  mymat = matrix(c(-11,-11,-9,-9,9,9,11,11,9,-3), 5,2,byrow=TRUE)
  table = mkContTable(mymat,10,10)
  print(table)
}


# PLEASE CHANGE THIS PATH TO WHEREVER THE FILE IS LOCATED ON YOUR MACHINE
myfile = "~/Desktop/526/groud.csv"

# read file & convert to ETF-vol pair returns over 7-day & 30-day time windows
SPYWeekly = mkReturns(myfile, 2,6,7)
SPYMonthly = mkReturns(myfile, 2,6,30)

DIAWeekly = mkReturns(myfile, 4,9,7)
DIAMonthly = mkReturns(myfile, 4,9,30)

QQQWeekly = mkReturns(myfile, 3,12,7)
QQQMonthly = mkReturns(myfile, 3,12,30)

#mixed indices
# use vix go predict the dow
VIXDIAWeekly = mkReturns(myfile, 4,6,7)
# use vix to predict nasdaq
VIXQQQWeekly = mkReturns(myfile, 3,6,7)
#BEST Contigency table: USE THE 1% WEEKLY interval with 3% max
vixdia2x2 = collapseTable(mkContTable(VIXDIAWeekly, .03, .01))
fisher.test(vixdia2x2)
vixqqq2x2 = collapseTable(mkContTable(VIXQQQWeekly, .03, .01))
fisher.test(vixdia2x2)

#index vs index
QQQDIAWeekly = mkReturns(myfile, 3,4,7)
qqqdia2x2 = collapseTable(mkContTable(QQQDIAWeekly, .03, .01))
fisher.test(qqqdia2x2)

# Some contingency experiments -
#make contingency tables with 5% intervals, max 20%
#spy5wk = mkContTable(SPYWeekly, .20, .05)
#spy5mon = mkContTable(SPYMonthly, .20, .05)
```

```
#make contingency tables with 10% intervals, max 20%
#spy10wk = mkContTable(SPYWeekly, .20, .10)
#spy10mon = mkContTable(SPYMonthly, .20, .10)

#make contingency tables with 2% intervals, max 10%
#spy2wk = mkContTable(SPYWeekly, .10, .02)
#spy2mon = mkContTable(SPYMonthly, .10, .02)

#BEST Contigency table: USE THE 1% WEEKLY interval with 3% max
spyBest = mkContTable(SPYWeekly, .03, .01)
diaBest = mkContTable(DIAWeekly, .03, .01)
qqqBest = mkContTable(QQQWeekly, .03, .01)

#make 2x2 tables
spy2x2 = collapseTable(spyBest)
dia2x2 = collapseTable(diaBest)
qqq2x2 = collapseTable(qqqBest)

# CONCLUSION 1. VERY STRONG CONCURRENT SIGNAL
# Concurrent 2x2 tables show that
# if VIX decreases during a time window,
# SPY increases during SAME time window

# Conversely, if VIX increases during a time window,
# SPY decreases during SAME time window

# Very strong signal ( 5x ) in both cases

# But can we look into the future ? Make predictive returns ?
# Try to predict tomorrow index based on today's vix
SPYPred1 = mkPredictiveReturns(myfile, 2,6,1,1)
spyPred1Table = collapseTable(mkContTable(SPYPred1, .03, .01))

# Try to predict tomorrow index based on past week vix
SPYPred2 = mkPredictiveReturns(myfile, 2,6,7,1)
spyPred2Table = collapseTable(mkContTable(SPYPred2, .03, .01))

# Try to predict tomorrow index based on past two weeks vix
SPYPred3 = mkPredictiveReturns(myfile, 2,6,14,1)
spyPred3Table = collapseTable(mkContTable(SPYPred3, .03, .01))

# Try to predict weekly index return, based on past month vix
SPYPred4 = mkPredictiveReturns(myfile, 2,6,28,7)
spyPred4Table = collapseTable(mkContTable(SPYPred4, .03, .01))

# Analysis of SPY

# 1. Observe data
df <- read_csv(myfile)
df$Date <- strptime(df$Date,'%m/%d/%Y')
df$weekday <- weekdays(df$Date)
df$month <- month(df$Date)
df$year <- year(df$Date)
SPYdf <- df[,c(1,2,6,14,15,16)]
names(SPYdf) <- c("time","vol","ETF","weekday","month","year")

ggplot(SPYdf, aes(x=as.POSIXct(time))) + geom_line(aes(y=vol,col="vol")) + geom_line(aes(y=ETF,col="ETF")) + xlab("Time") + ylab("")
```

```
# We can see an obvious negative correlation

# 1.2 Use spy2x2 table to test independence
# Use Fisher's exactly test
fisher.test(spy2x2)
# Use Pearson's Chi-squared test
chisq.test(spy2x2)

# CONCLUSION: we conclude that ETF and vol are not independent (both reject H0)

# 2. Try a linear model

SPYMdf <- data.frame(vol = as.numeric(SPYMonthly[,1]),
              ETF = as.numeric(SPYMonthly[,2]))
lmod <- lm(ETF ~ vol,SPYMdf)
summary(lmod)

SPYMdf$vol2 <- SPYMdf$vol^2
lmod2 <- lm(ETF ~ vol + vol2 ,SPYMdf)
summary(lmod2)

anova(lmod,lmod2)
plot(lmod2)

# from the plot, we can find the normality assumption is not true

# 3. Try ordinal multinomial response, using spyBest table (8x8) to fit data
freq <- c(spyBest)
vol <- rep(c(1:8),8)
ETF <- rep(c(1:8),rep(8,8))

# 2.1 Try multinomial model and test whether should we use vol as continuous data
modm <- multinom(factor(ETF) ~ vol,weights = freq)
summary(modm)
c(modm$dev,modm$edf)
modmf <- multinom(factor(ETF) ~ factor(vol),weights = freq)
summary(modmf)
c(modmf$dev,modmf$edf)
pchisq(modm$dev-modmf$dev,modmf$edf-modm$edf,lower=F)

# CONCLUSION: we should not treat vol as continuous

# 2.2 Try proportional model and test whether it is better than  multinomial model
modp <- polr(factor(ETF) ~ vol,weights = freq)
summary(modp)
c(modp$dev,modp$edf)
modpf <- polr(factor(ETF) ~ factor(vol),weights = freq)
summary(modpf)
c(modpf$dev,modpf$edf)
pchisq(modp$dev-modpf$dev,modpf$edf-modp$edf,lower=F)

# similar as before(2.1), should not treat vol as continous

pchisq(modpf$dev-modmf$dev,modmf$edf-modpf$edf,lower=F)

# CONCLUSION: multinomial model fits better than proportional model
```

```
# 2.3 Try monthly data and test whether it is better weekly data

spyBest2 = mkContTable(SPYMonthly, .03, .01)
freq2 <- c(spyBest2)
vol2 <- rep(c(1:8),8)
ETF2 <- rep(c(1:8),rep(8,8))
modmf2 <- multinom(factor(ETF2) ~ factor(vol2),weights = freq2)
summary(modmf2)
c(modmf2$dev,modmf2$edf)
modmf$dev-modmf2$dev

# using monthly data, we can get a lower deviance, so monthly data is better
# test the conclusion of (2.1) & (2.2) for monthly data
modm2 <- multinom(factor(ETF2) ~ vol2,weights = freq2)
summary(modm2)
c(modm2$dev,modm2$edf)
modm$dev-modm2$dev
pchisq(modm2$dev-modmf2$dev,modmf2$edf-modm2$edf,lower=F)
modp2 <- polr(factor(ETF2) ~ vol2,weights = freq2)
summary(modp2)
c(modp2$dev,modp2$edf)
modp$dev-modp2$dev
modpf2 <- polr(factor(ETF2) ~ factor(vol2),weights = freq2)
summary(modpf2)

c(modpf2$dev,modpf2$edf)
modpf$dev-modpf2$dev
pchisq(modp2$dev-modpf2$dev,modpf2$edf-modp2$edf,lower=F)

pchisq(modpf2$dev-modmf2$dev,modmf2$edf-modpf2$edf,lower=F)

# survival analysis
x=1800:2200
y=SPYdf$vol[1800:2200]
plot(y~x)
mod1 = loess(y~x,degree=2,span=0.4)
plot(mod1$fitted)
mod1$fitted
ceiling(mod1$fitted)
time = match(seq(61,25,-1),ceiling(mod1$fitted[149:400]))
fit = survfit(Surv(time, rep(1,37)) ~ rep("control", 37), conf.type="log")
summary(fit)
plot(fit, conf.int=TRUE, lty=3:2, col=c("red","blue"), log=T,
   +    las=1, xlab="Days", ylab="Volatility", main="VIX return to market normalcy",
   +    mark.tim
```