VAR Analysis

Contents

Data

Load Base Data

```r
# 1. Load Political Social Media

#contains posts from Twitter & TruthSocial
social <- read.csv(here("data/mothership", "social.csv"))

social_hourly <- read.csv(here("data/mothership", "socialhourly.csv"))


# 2. Load Financial

#S&P500
SPY <- read.csv(here("data/mothership", "SPY.csv"))

#STOXX50
VGK <- read.csv(here("data/mothership", "VGK.csv"))

#CSI 300 (China)
ASHR <- read.csv(here("data/mothership", "ASHR.CSV"))
```

```r
#make posixct
SPY$timestamp = as.POSIXct(SPY$timestamp,format = "%Y-%m-%d %H:%M:%S")
VGK$timestamp = as.POSIXct(VGK$timestamp,format = "%Y-%m-%d %H:%M:%S")
ASHR$timestamp = as.POSIXct(ASHR$timestamp,format = "%Y-%m-%d %H:%M:%S")
social$timestamp = as.POSIXct(social$timestamp,format = "%Y-%m-%d %H:%M:%S")
social_hourly$timestamp = as.POSIXct(social_hourly$timestamp,format = "%Y-%m-%d %H:%M:%S")
```

Volatility

```r
#find hourly volatility
#NOTE: this ignores tweets made outside trading hours!!
SPY_volatility_alltime = dplyr::select(SPY,timestamp,r_vol_h)

#aggregating per hour
SPY_volatility_alltime = SPY_volatility_alltime %>%
        mutate(timestamp = floor_date(timestamp, unit = "hour")) %>%
        distinct(timestamp, .keep_all = TRUE)

#select time period
SPY_volatility = filter(SPY_volatility_alltime,
                between(timestamp,
                        as.Date('2014-01-01'),
                        as.Date('2025-04-10')))
```

```r
#find hourly volatility
#NOTE: this ignores tweets made outside trading hours!!
VGK_volatility_alltime = dplyr::select(VGK,timestamp,r_vol_h)
```

```r
#aggregating per hour
VGK_volatility_alltime = VGK_volatility_alltime %>%
        mutate(timestamp = floor_date(timestamp, unit = "hour")) %>%
        distinct(timestamp, .keep_all = TRUE)

#select time period
VGK_volatility = filter(VGK_volatility_alltime,
              between(timestamp,
                    as.Date('2014-01-01'),
                    as.Date('2025-04-10')))
```

```r
#find hourly volatility
#NOTE: this ignores tweets made outside trading hours!!
ASHR_volatility_alltime = dplyr::select(ASHR,timestamp,r_vol_h)

#aggregating per hour
ASHR_volatility_alltime = ASHR_volatility_alltime %>%
        mutate(timestamp = floor_date(timestamp, unit = "hour")) %>%
        distinct(timestamp, .keep_all = TRUE)

#select time period
ASHR_volatility = filter(ASHR_volatility_alltime,
              between(timestamp,
                    as.Date('2014-01-01'),
                    as.Date('2025-04-10')))
```

Number of Posts

```r
#find count
tweetcount_alltime = dplyr::select(social_hourly,timestamp,N)

#select time period
tweetcount = filter(tweetcount_alltime,
              between(timestamp,
                    as.Date('2014-01-01'),
                    as.Date('2025-04-10')))
```

Dummy for Social Media Post

```r
#find dummy
tweetdummy_alltime = dplyr::select(social_hourly,timestamp,dummy)

#select time period
tweetdummy = filter(tweetdummy_alltime,
              between(timestamp,
                    as.Date('2014-01-01'),
                    as.Date('2025-04-10')))
```

Number of Tweets Mentioning Tariffs

```r
#find count
tariff_alltime = dplyr::select(social_hourly,timestamp,total_tariff)

#select time period
tariff = filter(tariff_alltime,
                between(timestamp,
                        as.Date('2014-01-01'),
                        as.Date('2025-04-10')))
```

Number of Tweets Mentioning Trade

```r
#find count
trade_alltime = dplyr::select(social_hourly,timestamp,total_trade)

#select time period
trade = filter(trade_alltime,
               between(timestamp,
                       as.Date('2014-01-01'),
                       as.Date('2025-04-10')))
```

Proportion of Positive

```r
#find count
positive_alltime = dplyr::select(social_hourly,timestamp,prop_positive)

#select time period
positive = filter(positive_alltime,
                  between(timestamp,
                          as.Date('2014-01-01'),
                          as.Date('2025-04-10')))
```

Proportion of Negative

```r
#find count
negative_alltime = dplyr::select(social_hourly,timestamp,prop_negative)

#select time period
negative = filter(negative_alltime,
                  between(timestamp,
                          as.Date('2014-01-01'),
                          as.Date('2025-04-10')))
```

Merge

```r
#merge our dependant and independant vars
var_data = left_join(SPY_volatility, VGK_volatility, by="timestamp")
var_data = left_join(var_data, ASHR_volatility, by="timestamp")
var_data = left_join(var_data, tweetdummy, by="timestamp")
var_data = left_join(var_data, tweetcount, by="timestamp")
var_data = left_join(var_data, tariff, by="timestamp")
var_data = left_join(var_data, trade, by="timestamp")
var_data = left_join(var_data, positive, by="timestamp")
var_data = left_join(var_data, negative, by="timestamp")

#rename volatility columns
names(var_data)[2] <- "SPY_vol"
names(var_data)[3] <- "VGK_vol"
names(var_data)[4] <- "ASHR_vol"

#convert NA to zeroes
var_data$N[is.na(var_data$N)] = 0
var_data$dummy[is.na(var_data$dummy)] = 0
var_data$total_tariff[is.na(var_data$total_tariff)] = 0
var_data$total_trade[is.na(var_data$total_trade)] = 0
var_data$prop_positive[is.na(var_data$prop_positive)] = 0
var_data$prop_negative[is.na(var_data$prop_negative)] = 0
```

S&P500 VAR Models

Find Number of Lags

Tweet Count on Volatility by hour

Tweet Dummy on Volatility by hour

Tariff Mention on Volatility by hour

Positive Vibe on Volatility by hour

European Market VAR Models

Find Number of Lags

Tweet Count on Volatility by hour

Tweet Dummy on Volatility by hour

Tariff Mention on Volatility by hour

Negative Vibe on Volatility by hour

Chinese Market VAR Models

Find Number of Lags

Tweet Count on Volatility by hour

Tweet Dummy on Volatility by hour

Tariff Mention on Volatility by hour

Positive Vibe on Volatility by hour

Negative Vibe on Volatility by hour