ARMA-X Analysis

Contents

Data

Load Base Data

```
# 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"))
```

```
#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

```
#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('2019-01-01'),
                          as.Date('2025-04-10')))
```

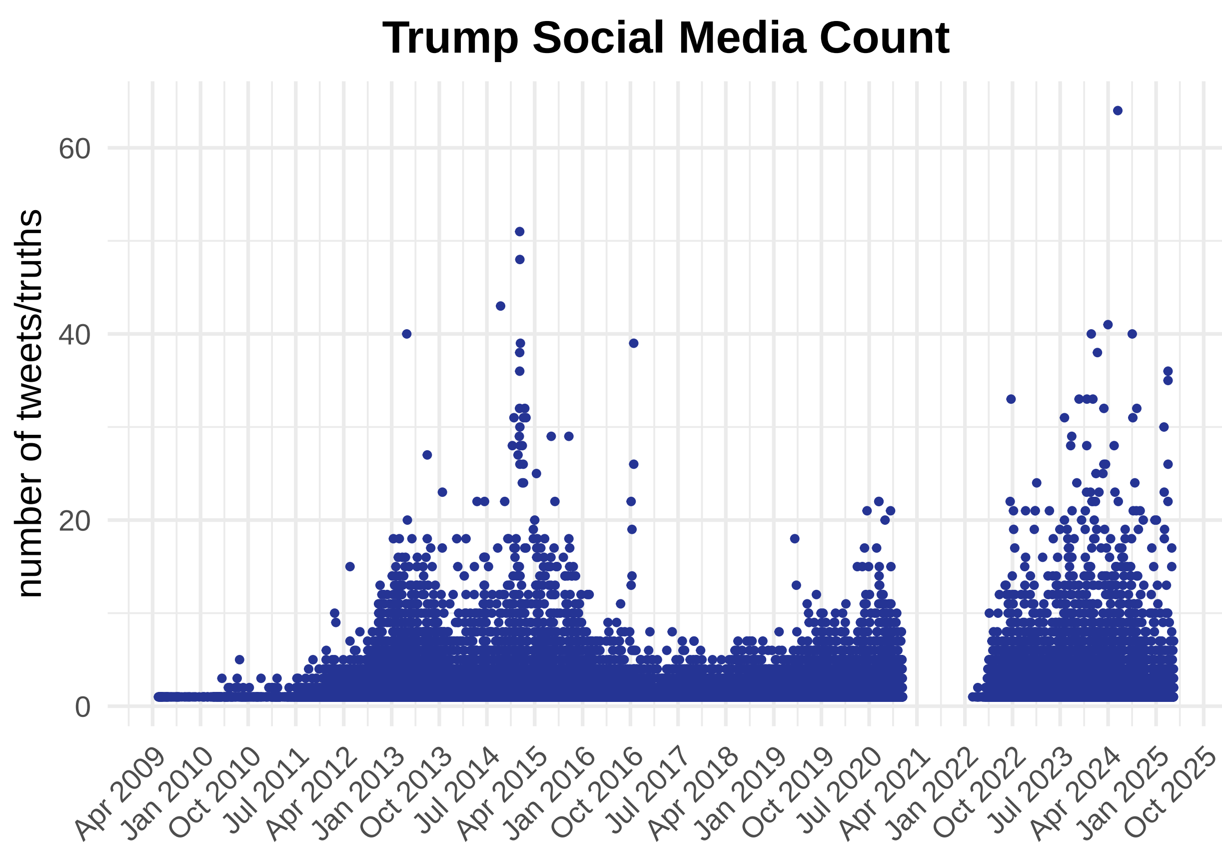
Number of Posts

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

#select time period
tweetcount = filter(tweetcount_alltime,
                    between(timestamp,
                            as.Date('2019-01-01'),
                            as.Date('2025-04-10')))

#plot
ggplot(tweetcount_alltime, aes(x = timestamp, y = N)) +
    geom_point(color = "#253494", size = 1) +
    scale_x_datetime(date_labels = "%b %Y", date_breaks = "9 month") +
    labs(title = "Trump Social Media Count",
         x = NULL,
         y = "number of tweets/truths") +
    theme_minimal(base_size = 14) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1),
          plot.title = element_text(face = "bold", hjust = 0.5))
```

```
## Warning: Removed 1172 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Dummy for Social Media Post

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

#select time period
tweetdummy = filter(tweetdummy_alltime,
                  between(timestamp,
                          as.Date('2019-01-01'),
                          as.Date('2025-04-10')))
```

Number of Tweets Mentioning Tariffs

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

#select time period
tariff = filter(tariff_alltime,
                  between(timestamp,
                          as.Date('2019-01-01'),
                          as.Date('2025-04-10')))
```

Number of Tweets Mentioning Trade

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

#select time period
trade = filter(trade_alltime,
                  between(timestamp,
                          as.Date('2019-01-01'),
                          as.Date('2025-04-10')))
```

Merge

```
#merge our dependant and independant vars
armax_data = left_join(SPY_volatility, tweetcount, by="timestamp")
armax_data = left_join(armax_data, tweetdummy, by="timestamp")
armax_data = left_join(armax_data, tariff, by="timestamp")
armax_data = left_join(armax_data, trade, by="timestamp")

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

ARMA-X Models

Find Number of Lags

```
nb.lags <- 3 #r
count_lags <- embed(armax_data$N, nb.lags + 1)
dummy_lags <- embed(armax_data$dummy, nb.lags + 1)
tariff_lags <- embed(armax_data$total_tariff, nb.lags + 1)
trade_lags <- embed(armax_data$total_trade, nb.lags + 1)
#colnames(count_lags) <- paste0("Lag_", 0:nb.lags)

#align volatility to match count rows (for lag)
vol_aligned <- tail(armax_data$r_vol_h, nrow(count_lags))

#choosing how many lags
# fit an ARMA(0,0,0) model with lm (with r set above)
eq <- lm(vol_aligned ~ count_lags)
eq2 <- lm(vol_aligned ~ dummy_lags)
eq3 <- lm(vol_aligned ~ tariff_lags)
eq4 <- lm(vol_aligned ~ trade_lags)

#compute Newey-West HAC standard errors for count
var.cov.mat1 <- NeweyWest(eq, lag = 7, prewhite = FALSE)
robust_se1 <- sqrt(diag(var.cov.mat1))
#for dummy
var.cov.mat2 <- NeweyWest(eq2, lag = 7, prewhite = FALSE)
robust_se2 <- sqrt(diag(var.cov.mat2))
#for tariff
var.cov.mat3 <- NeweyWest(eq3, lag = 7, prewhite = FALSE)
robust_se3 <- sqrt(diag(var.cov.mat3))
#for trade
var.cov.mat4 <- NeweyWest(eq4, lag = 7, prewhite = FALSE)
robust_se4 <- sqrt(diag(var.cov.mat4))



#output table; significant lags are how many we choose
stargazer(eq, eq, type = "latex",
          column.labels = c("(no HAC)", "(HAC)"), keep.stat = "n",
          se = list(NULL, robust_se1), no.space = TRUE)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, May 01, 2025 - 17:50:37

```
#output table; significant lags are how many we choose
stargazer(eq2, eq2, type = "latex",
          column.labels = c("(no HAC)", "(HAC)"), keep.stat = "n",
          se = list(NULL, robust_se2), no.space = TRUE)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, May 01, 2025 - 17:50:37

Table 1:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
count_lags1	-0.001^{*}	-0.001^{***}
	(0.001)	(0.0002)
count_lags2	-0.001	-0.001^{*}
	(0.001)	(0.0003)
count_lags3	0.0002	0.0002
	(0.001)	(0.0004)
count_lags4	-0.0002	-0.0002
	(0.001)	(0.0002)
Constant	0.036^{***}	0.036^{***}
	(0.001)	(0.002)
Observations	11,036	11,036
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
dummy_lags1	-0.001	-0.001
	(0.003)	(0.003)
dummy_lags2	-0.003	-0.003
	(0.003)	(0.002)
dummy_lags3	0.006^{**}	0.006^{*}
	(0.003)	(0.003)
dummy_lags4	0.004	0.004
	(0.003)	(0.003)
Constant	0.033^{***}	0.033^{***}
	(0.002)	(0.001)
Observations	11,036	11,036
Note:	*p<0.1; **p<0.05; ***p<0.01	

```r
#output table; significant lags are how many we choose
stargazer(eq3, eq3, type = "latex",
          column.labels = c("(no HAC)", "(HAC)"), keep.stat = "n",
          se = list(NULL, robust_se3), no.space = TRUE)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, May 01, 2025 - 17:50:37

Table 3:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
tariff_lags1	0.038	0.038
	(0.024)	(0.026)
tariff_lags2	0.057**	0.057
	(0.024)	(0.045)
tariff_lags3	0.050**	0.050
	(0.024)	(0.033)
tariff_lags4	0.075***	0.075**
	(0.024)	(0.031)
Constant	0.034***	0.034***
	(0.001)	(0.002)
Observations	11,036	11,036
Note:	*p<0.1; **p<0.05; ***p<0.01	

```r
#output table; significant lags are how many we choose
stargazer(eq4, eq4, type = "latex",
          column.labels = c("(no HAC)", "(HAC)"), keep.stat = "n",
          se = list(NULL, robust_se4), no.space = TRUE)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, May 01, 2025 - 17:50:37

Tweet Count on Volatility by hour

```r
#find best armax model and fit
auto.arima(armax_data$r_vol_h, xreg=armax_data$N,
                        max.p = 5, max.q = 5, max.d = 0, ic = "aic")
```

```
## Series: armax_data$r_vol_h
## Regression with ARIMA(5,0,0) errors
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5  intercept    xreg
##       0.3428  0.0325  0.0781  0.1527  0.0987     0.0350  -7e-04
## s.e.  0.0095  0.0101  0.0135  0.0143  0.0128     0.0034   5e-04
```

Table 4:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
trade_lags1	0.006	0.006
	(0.020)	(0.014)
trade_lags2	0.023	0.023
	(0.020)	(0.025)
trade_lags3	0.008	0.008
	(0.020)	(0.015)
trade_lags4	0.016	0.016
	(0.020)	(0.019)
Constant	0.034***	0.034***
	(0.001)	(0.002)
Observations	11,036	11,036
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
##
## sigma^2 = 0.01086:  log likelihood = 9300.97
## AIC=-18585.94   AICc=-18585.93   BIC=-18527.47
```

Tweet Dummy on Volatility by hour

```r
#find best armax model and fit
auto.arima(armax_data$r_vol_h, xreg=armax_data$dummy,
                         max.p = 5, max.q = 5, max.d = 0, ic = "aic")
```

```
## Series: armax_data$r_vol_h
## Regression with ARIMA(5,0,0) errors
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5  intercept     xreg
##       0.3428  0.0325  0.0784  0.1526  0.0985     0.0353  -0.0027
## s.e.  0.0095  0.0101  0.0135  0.0143  0.0128     0.0034   0.0022
##
## sigma^2 = 0.01086:  log likelihood = 9300.66
## AIC=-18585.32   AICc=-18585.31   BIC=-18526.85
```

Tariff Mention on Volatility by hour

```r
#find best armax model and fit
auto.arima(armax_data$r_vol_h, xreg=armax_data$total_tariff,
                         max.p = 5, max.q = 5, max.d = 0, ic = "aic")
```

```
## Series: armax_data$r_vol_h
## Regression with ARIMA(5,0,0) errors
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5  intercept     xreg
##       0.3428  0.0340  0.0767  0.1526  0.0991     0.0345  -0.0376
## s.e.  0.0095  0.0102  0.0136  0.0143  0.0128     0.0034   0.0199
##
## sigma^2 = 0.01086:  log likelihood = 9301.7
## AIC=-18587.4   AICc=-18587.38   BIC=-18528.92
```

Trade Mention on Volatility by hour

```r
#find best armax model and fit
auto.arima(armax_data$r_vol_h, xreg=armax_data$total_trade,
                       max.p = 5, max.q = 5, max.d = 0, ic = "aic")
```

```
## Series: armax_data$r_vol_h
## Regression with ARIMA(5,0,0) errors
##
## Coefficients:
##          ar1     ar2     ar3     ar4     ar5  intercept     xreg
##       0.3430  0.0329  0.0775  0.1525  0.0987     0.0345  -0.0156
## s.e.  0.0095  0.0101  0.0136  0.0143  0.0128     0.0034   0.0161
##
## sigma^2 = 0.01086:  log likelihood = 9300.37
## AIC=-18584.73   AICc=-18584.72   BIC=-18526.26
```