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('2013-01-01'),
                          as.Date('2025-04-10')))
```

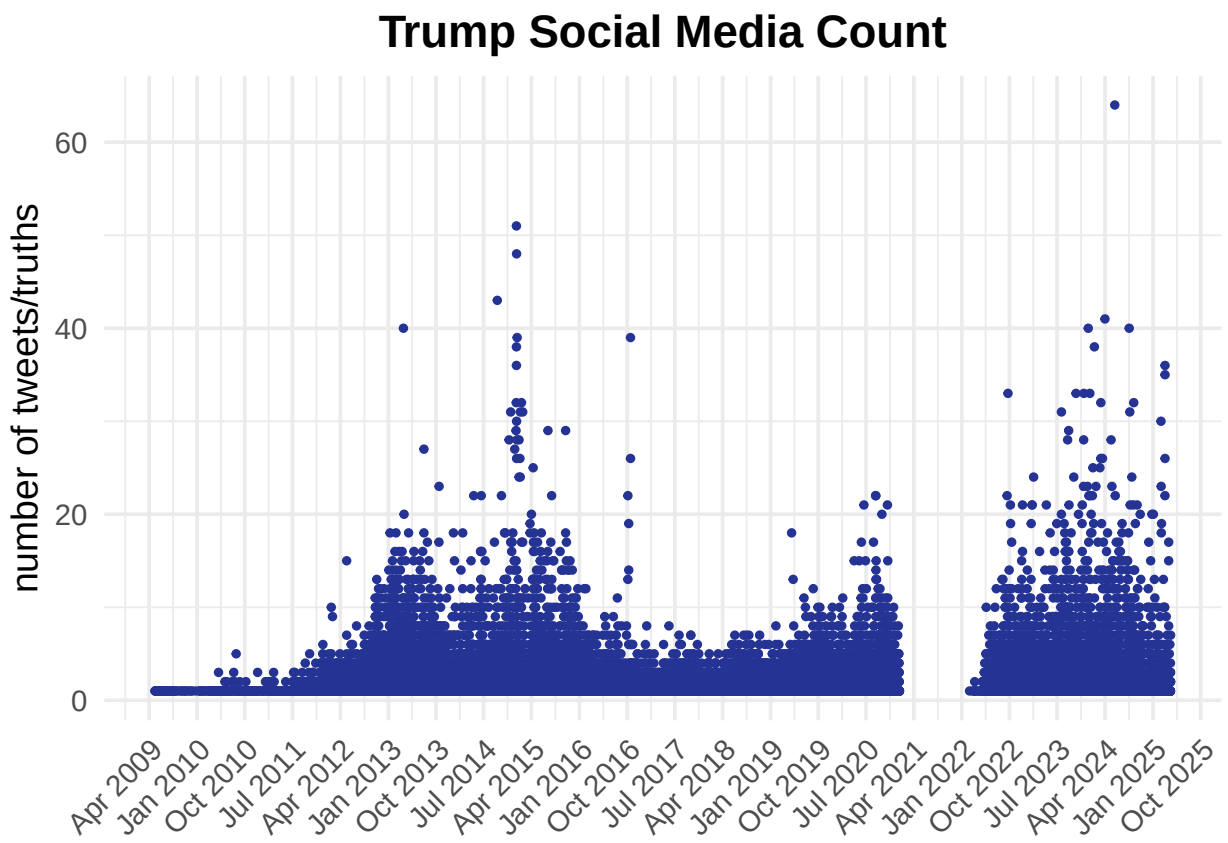
Number of Posts

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

#select time period
tweetcount = filter(tweetcount_alltime,
                    between(timestamp,
                            as.Date('2013-01-01'),
                            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('2013-01-01'),
                            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('2013-01-01'),
                            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('2013-01-01'),
                            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

```
#build number of lags we will test
nb.lags <- 12 #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("_", 0:nb.lags)
colnames(dummy_lags) <- paste0("_", 0:nb.lags)
colnames(tariff_lags) <- paste0("_", 0:nb.lags)
colnames(trade_lags) <- paste0("_", 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 = 16, prewhite = FALSE)
robust_se1 <- sqrt(diag(var.cov.mat1))
#for dummy
var.cov.mat2 <- NeweyWest(eq2, lag = 16, prewhite = FALSE)
robust_se2 <- sqrt(diag(var.cov.mat2))
#for tariff
var.cov.mat3 <- NeweyWest(eq3, lag = 16, prewhite = FALSE)
robust_se3 <- sqrt(diag(var.cov.mat3))
#for trade
var.cov.mat4 <- NeweyWest(eq4, lag = 16, 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: Fri, May 02, 2025 - 11:07:17

```
#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 1:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
count_lags_0	−0.0005	−0.0005***
	(0.0003)	(0.0001)
count_lags_1	−0.0002	−0.0002
	(0.0003)	(0.0002)
count_lags_2	0.0003	0.0003
	(0.0003)	(0.0002)
count_lags_3	0.00005	0.00005
	(0.0003)	(0.0001)
count_lags_4	0.0003	0.0003
	(0.0003)	(0.0003)
count_lags_5	−0.0002	−0.0002*
	(0.0003)	(0.0001)
count_lags_6	−0.0002	−0.0002
	(0.0003)	(0.0002)
count_lags_7	−0.0001	−0.0001
	(0.0003)	(0.0004)
count_lags_8	−0.00003	−0.00003
	(0.0003)	(0.0001)
count_lags_9	0.001*	0.001**
	(0.0003)	(0.0003)
count_lags_10	0.0001	0.0001
	(0.0003)	(0.0001)
count_lags_11	−0.0001	−0.0001
	(0.0003)	(0.0001)
count_lags_12	−0.0002	−0.0002
	(0.0003)	(0.0002)
Constant	0.020***	0.020***
	(0.001)	(0.001)
Observations	21,597	21,597
Note:	*p<0.1; **p<0.05; ***p<0.01	

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, May 02, 2025 - 11:07:17

Table 2:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
dummy_lags_0	−0.001	−0.001
	(0.001)	(0.002)
dummy_lags_1	−0.001	−0.001
	(0.001)	(0.001)
dummy_lags_2	0.003**	0.003**
	(0.001)	(0.001)
dummy_lags_3	0.002	0.002*
	(0.001)	(0.001)
dummy_lags_4	0.002*	0.002
	(0.001)	(0.002)
dummy_lags_5	−0.001	−0.001
	(0.001)	(0.001)
dummy_lags_6	−0.001	−0.001
	(0.001)	(0.001)
dummy_lags_7	−0.002	−0.002
	(0.001)	(0.001)
dummy_lags_8	0.0004	0.0004
	(0.001)	(0.001)
dummy_lags_9	0.004***	0.004**
	(0.001)	(0.002)
dummy_lags_10	0.0004	0.0004
	(0.001)	(0.001)
dummy_lags_11	−0.001	−0.001
	(0.001)	(0.001)
dummy_lags_12	−0.001	−0.001
	(0.001)	(0.001)
Constant	0.019***	0.019***
	(0.001)	(0.002)
Observations	21,597	21,597
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
#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: Fri, May 02, 2025 - 11:07:18

```
#output table; significant lags are how many we choose
stargazer(eq4, eq4, type = "latex",
```

Table 3:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
tariff_lags_0	0.034**	0.034
	(0.016)	(0.022)
tariff_lags_1	0.041***	0.041
	(0.016)	(0.038)
tariff_lags_2	0.042***	0.042
	(0.016)	(0.028)
tariff_lags_3	0.053***	0.053**
	(0.016)	(0.023)
tariff_lags_4	0.063***	0.063**
	(0.016)	(0.029)
tariff_lags_5	0.083***	0.083**
	(0.016)	(0.041)
tariff_lags_6	0.047***	0.047**
	(0.016)	(0.022)
tariff_lags_7	0.041***	0.041*
	(0.016)	(0.022)
tariff_lags_8	0.036**	0.036**
	(0.016)	(0.015)
tariff_lags_9	0.306***	0.306
	(0.016)	(0.254)
tariff_lags_10	0.096***	0.096*
	(0.016)	(0.051)
tariff_lags_11	0.110***	0.110*
	(0.016)	(0.056)
tariff_lags_12	0.039**	0.039**
	(0.016)	(0.017)
Constant	0.019***	0.019***
	(0.001)	(0.001)
Observations	21,597	21,597
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
          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: Fri, May 02, 2025 - 11:07:18

Table 4:

	Dependent variable:	
	vol_aligned	
	(no HAC)	(HAC)
	(1)	(2)
trade_lags_0	−0.002	−0.002
	(0.008)	(0.004)
trade_lags_1	0.003	0.003
	(0.008)	(0.007)
trade_lags_2	−0.0003	−0.0003
	(0.008)	(0.004)
trade_lags_3	0.002	0.002
	(0.008)	(0.006)
trade_lags_4	−0.0003	−0.0003
	(0.008)	(0.005)
trade_lags_5	−0.001	−0.001
	(0.008)	(0.005)
trade_lags_6	−0.003	−0.003
	(0.008)	(0.004)
trade_lags_7	−0.003	−0.003
	(0.008)	(0.003)
trade_lags_8	0.001	0.001
	(0.008)	(0.006)
trade_lags_9	0.002	0.002
	(0.008)	(0.005)
trade_lags_10	0.002	0.002
	(0.008)	(0.006)
trade_lags_11	0.007	0.007
	(0.008)	(0.010)
trade_lags_12	−0.00001	−0.00001
	(0.008)	(0.004)
Constant	0.020***	0.020***
	(0.001)	(0.001)
Observations	21,597	21,597
Note:	*p<0.1; **p<0.05; ***p<0.01	

Tweet Count on Volatility by hour

```
#use selected number of lags
nb.lags <- 0
count_lags <- embed(armax_data$N, nb.lags + 1)
```

	Model 1
ar1	0.3482***
	(0.0068)
ar2	0.0349***
	(0.0072)
ar3	0.0832***
	(0.0096)
ar4	0.1546***
	(0.0102)
ar5	0.1057***
	(0.0091)
intercept	0.0206***
	(0.0019)
count_lag_0	−0.0005*
	(0.0003)
AIC	−50660.1539
AICc	−50660.1472
BIC	−50596.3070
Log Likelihood	25338.0770
Num. obs.	21609

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: ARMAX Model Results

```
colnames(count_lags) <- paste0("count_lag_", 0:nb.lags)

#variables
y = tail(armax_data$r_vol_h, nrow(count_lags))
x = count_lags


#find best armax model and fit
tab1 = auto.arima(y, xreg = x, seasonal = FALSE, max.p = 5, max.q = 5, max.d = 0,
              stepwise = FALSE, approximation = FALSE, trace = FALSE)

#screenreg(tab1, digits=4)

texreg(tab1, caption = "ARMAX Model Results", label = "tab:armax", digits=4)
```

Tweet Dummy on Volatility by hour

```
#use selected number of lags
nb.lags <- 3
dummy_lags <- embed(armax_data$dummy, nb.lags + 1)
colnames(dummy_lags) <- paste0("dummy_lag_", 0:nb.lags)

#variables
y = tail(armax_data$r_vol_h, nrow(dummy_lags))
x = dummy_lags
```

	Model 1
ar1	0.3481***
	(0.0068)
ar2	0.0347***
	(0.0072)
ar3	0.0841***
	(0.0096)
ar4	0.1546***
	(0.0102)
ar5	0.1053***
	(0.0091)
intercept	0.0193***
	(0.0020)
dummy_lag_0	−0.0015
	(0.0011)
dummy_lag_1	−0.0008
	(0.0011)
dummy_lag_2	0.0036**
	(0.0011)
dummy_lag_3	0.0019
	(0.0011)
AIC	−50656.2791
AICc	−50656.2669
BIC	−50568.4911
Log Likelihood	25339.1395
Num. obs.	21606

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 6: ARMAX Model Results

```
#find best armax model and fit
tab2 = auto.arima(y, xreg = x, seasonal = FALSE, max.p = 5, max.q = 5, max.d = 0,
              stepwise = FALSE, approximation = FALSE, trace = FALSE)

#screenreg(tab2, digits=4)

texreg(tab2, caption = "ARMAX Model Results", label = "tab:armax", digits=4)
```

Tariff Mention on Volatility by hour

```
#use selected number of lags
nb.lags <- 3
tariff_lags <- embed(armax_data$total_tariff, nb.lags + 1)
colnames(tariff_lags) <- paste0("tariff_lag_", 0:nb.lags)

#variables
y = tail(armax_data$r_vol_h, nrow(tariff_lags))
x = tariff_lags
```

	Model 1
ar1	0.3477***
	(0.0068)
ar2	0.0362***
	(0.0073)
ar3	0.0838***
	(0.0097)
ar4	0.1573***
	(0.0103)
ar5	0.1047***
	(0.0092)
intercept	0.0204***
	(0.0019)
tariff_lag_0	−0.0374**
	(0.0132)
tariff_lag_1	−0.0258
	(0.0138)
tariff_lag_2	−0.0459***
	(0.0138)
tariff_lag_3	−0.0229
	(0.0132)
AIC	−50658.8660
AICc	−50658.8537
BIC	−50571.0780
Log Likelihood	25340.4330
Num. obs.	21606

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 7: ARMAX Model Results

```
#find best armax model and fit
tab3 = auto.arima(y, xreg = x, seasonal = FALSE, max.p = 5, max.q = 5, max.d = 0,
              stepwise = FALSE, approximation = FALSE, trace = FALSE)

#screenreg(tab3, digits=4)

texreg(tab3, caption = "ARMAX Model Results", label = "tab:armax", digits=4)
```

Trade Mention on Volatility by hour

```
#use selected number of lags
nb.lags <- 3
trade_lags <- embed(armax_data$total_trade, nb.lags + 1)
colnames(trade_lags) <- paste0("trade_lag_", 0:nb.lags)

#variables
y = tail(armax_data$r_vol_h, nrow(trade_lags))
x = trade_lags
```

	Model 1
ar1	0.3483***
	(0.0068)
ar2	0.0349***
	(0.0072)
ar3	0.0829***
	(0.0097)
ar4	0.1549***
	(0.0102)
ar5	0.1059***
	(0.0091)
intercept	0.0204***
	(0.0019)
trade_lag_0	−0.0060
	(0.0066)
trade_lag_1	−0.0011
	(0.0068)
trade_lag_2	−0.0086
	(0.0068)
trade_lag_3	−0.0062
	(0.0066)
AIC	−50644.2169
AICc	−50644.2047
BIC	−50556.4289
Log Likelihood	25333.1085
Num. obs.	21606

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: ARMAX Model Results

```
#find best armax model and fit
tab4 = auto.arima(y, xreg = x, seasonal = FALSE, max.p = 5, max.q = 5, max.d = 0,
              stepwise = FALSE, approximation = FALSE, trace = FALSE)

#screenreg(tab4, digits=4)

texreg(tab4, caption = "ARMAX Model Results", label = "tab:armax", digits=4)
```