ARMA-X Analysis Tutorial

Contents

Setup

Load Libraries & Functions

```r
rm(list=ls())
require(tinytex) #LaTeX
require(ggplot2) #plots
require(AEC) #JP-Renne functions
require(AER) #NW formula
require(forecast) #time series stuff
require(expm) #matrix exponents
require(here) #directory finder
require(stringr) # analysis of strings, important for the detection in tweets
require(dplyr) #data management
require(lubridate) #data dates management
require(zoo) #for lagging
require(jtools) #tables
require(huxtable) #tables
require(lmtest) #reg tests
require(vroom) #for loading data
require(data.table) #for data filtering
require(sysid) #for ARMA-X modeling
require(sandwich) #regression errors
require(stargazer) #nice reg tables
require(tidytext) #text mining
require(textstem) #lemmatization
require(quanteda) #tokenization
require(texreg) #arima tables
require(future.apply) #parallel computation (speed)
require(aTSA) #adf test

getwd()
#setwd("...") -> set wd at base repo folder

#load helper functions
source(here("helperfunctions/data_loaders.R"))
source(here("helperfunctions/date_selector.R"))
source(here("helperfunctions/plotters.R"))
source(here("helperfunctions/quick_arma.R"))
source(here("helperfunctions/r.vol_calculators.R"))
source(here("helperfunctions/truths_cleaning_function.R"))
source(here("helperfunctions/armax_functions.R"))
```

Load Data

Use the full_data script to load our final dataset. Then select the timeframe.

```r
#load final dataset
source(here("helperfunctions/full_data.R"))

#select timeframe
```

```r
data = filter(data,between(timestamp, as.Date('2014-01-01'), as.Date('2025-05-07')))

#first term
#data = filter(data,between(timestamp, as.Date('2017-01-20'), as.Date('2021-01-20')))

#second term
#data = filter(data,between(timestamp, as.Date('2025-01-20'), as.Date('2025-05-07')))
```

Stationarity

Having a p.value of 0.01 for our main variables seems to indicate stationarity.

```r
adf.test(data$SPY_vol)
adf.test(data$VGK_vol)
adf.test(data$ASHR_vol)

adf.test(data$N)
adf.test(data$tariff)
adf.test(data$china)
```

	Model 1
ar1	0.3572***
	(0.0071)
ar2	0.0427***
	(0.0075)
ar3	0.0903***
	(0.0075)
ar4	0.0978***
	(0.0075)
ar5	0.0859***
	(0.0071)
intercept	0.0217***
	(0.0017)
tariff_lag_0	0.0047**
	(0.0015)
tariff_lag_1	0.0201***
	(0.0015)
tariff_lag_2	0.0109***
	(0.0015)
AIC	−44818.4470
AICc	−44818.4359
BIC	−44739.4276
Log Likelihood	22419.2235
Num. obs.	19969

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

Table 1: ARMAX Model Results

Univariate ARMA-X Models

Custom ARMA-X Specification

This first function allows to run a certain specification of an ARMA-X model. We needed to create this function due to the creation of the lags for the exogenous regressor.

```
#armax enables a custom armax specification with p,q,r
res2 = armax(data$SPY_vol, xreg=data$tariff, nb.lags=2,
                    p=5, q=0, d=0, latex=T)
```

Best ARMA-X Given r

This second function uses the base auto.arima function in order to look for the best (lowest AIC) specification given a certain number of x lags. Once again it was necessary to create a custom function for this due to the creation of the lags for the exogenous regressor.

```
#auto.armax selects the lowest AIC value given r (exogenous variable lags)
res1 = auto.armax(data$SPY_vol,xreg=data$tariff,nb.lags=2,
                latex=T,max.p = 6, max.q = 6, max.d=0)
```

	Model 1
ar1	1.7000^{***}
	(0.1313)
ar2	-0.8772^{***}
	(0.1518)
ar3	0.1689^{***}
	(0.0232)
ma1	-1.3999^{***}
	(0.1327)
ma2	0.4605^{***}
	(0.1192)
intercept	0.0217^{***}
	(0.0040)
tariff_lag_0	0.0042^{**}
	(0.0014)
tariff_lag_1	0.0199^{***}
	(0.0015)
tariff_lag_2	0.0112^{***}
	(0.0014)
AIC	-45860.5245
AICc	-45860.5134
BIC	-45781.5051
Log Likelihood	22940.2622
Num. obs.	19969

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

Table 2: ARMAX Model Results

	Model 1
ar1	0.2200***
	(0.0084)
ar2	0.9388***
	(0.0037)
ar3	−0.1837***
	(0.0079)
ma1	0.0870***
	(0.0042)
ma2	−0.8960***
	(0.0042)
intercept	0.0219***
	(0.0042)
tariff_lag_0	0.0035*
	(0.0014)
tariff_lag_1	0.0191***
	(0.0015)
tariff_lag_2	0.0103***
	(0.0015)
tariff_lag_3	−0.0045**
	(0.0014)
AIC	−46020.9547
AICc	−46020.9415
BIC	−45934.0340
Log Likelihood	23021.4774
Num. obs.	19968

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

Table 3: ARMAX selected by AIC

Best ARMA-X

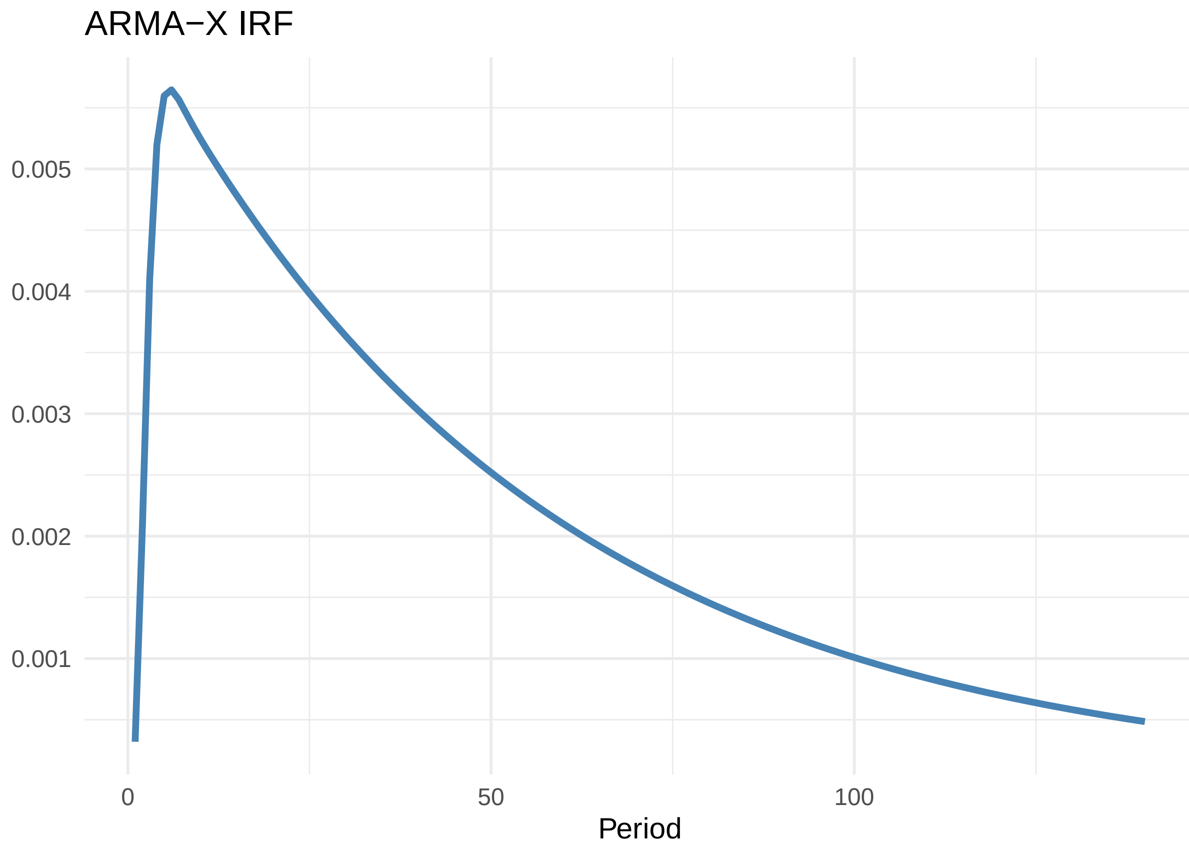
This third function is the most sophisticated. We've had to essentially custom code the auto.arima function in order to loop for p, q and r. It also provides many useful details such as a plot of the AIC value for the different number of lags. Note that the output of this function is a list, so in order get the model output, it is needed to specifically call it (see how the call works for this third model in the IRF section). The latex option in all functions determines whether the output is console-friendly, or output-friendly.

```r
#auto.armax.r selects the lowest AIC checking all 3 p,q,r values
res3 = auto.armax.r(data$SPY_vol, x=data$tariff,
                max_p = 3, max_q = 3, max_r = 3, criterion = "AIC", latex=T)
```

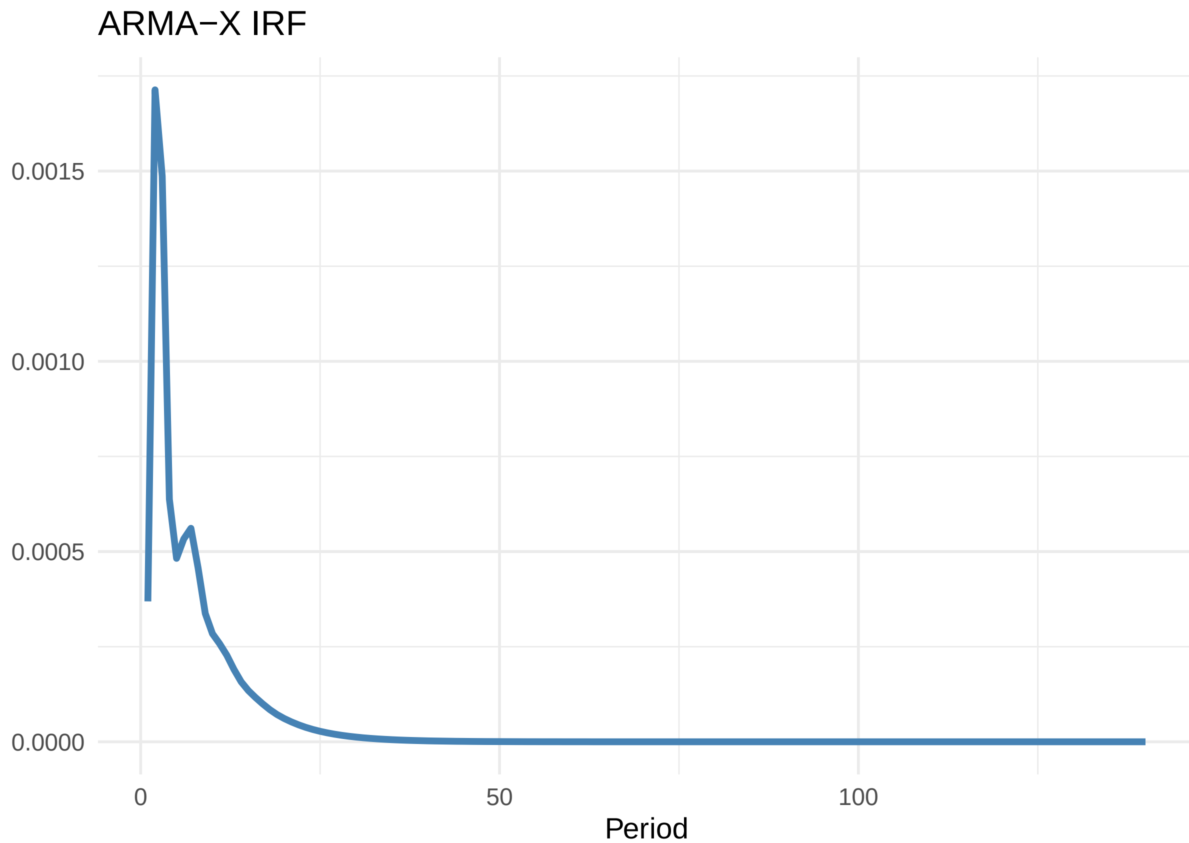
IRF

Finally, we have our personal twist on the IRF plot function from the AEC package. It enables automatic extraction of the phi's and beta's of a fitted ARMA-X model. It then uses the provided sim.arma function where it inputs the beta's as theta's. Finally, it automatically plots it, using a fancy ggplot.

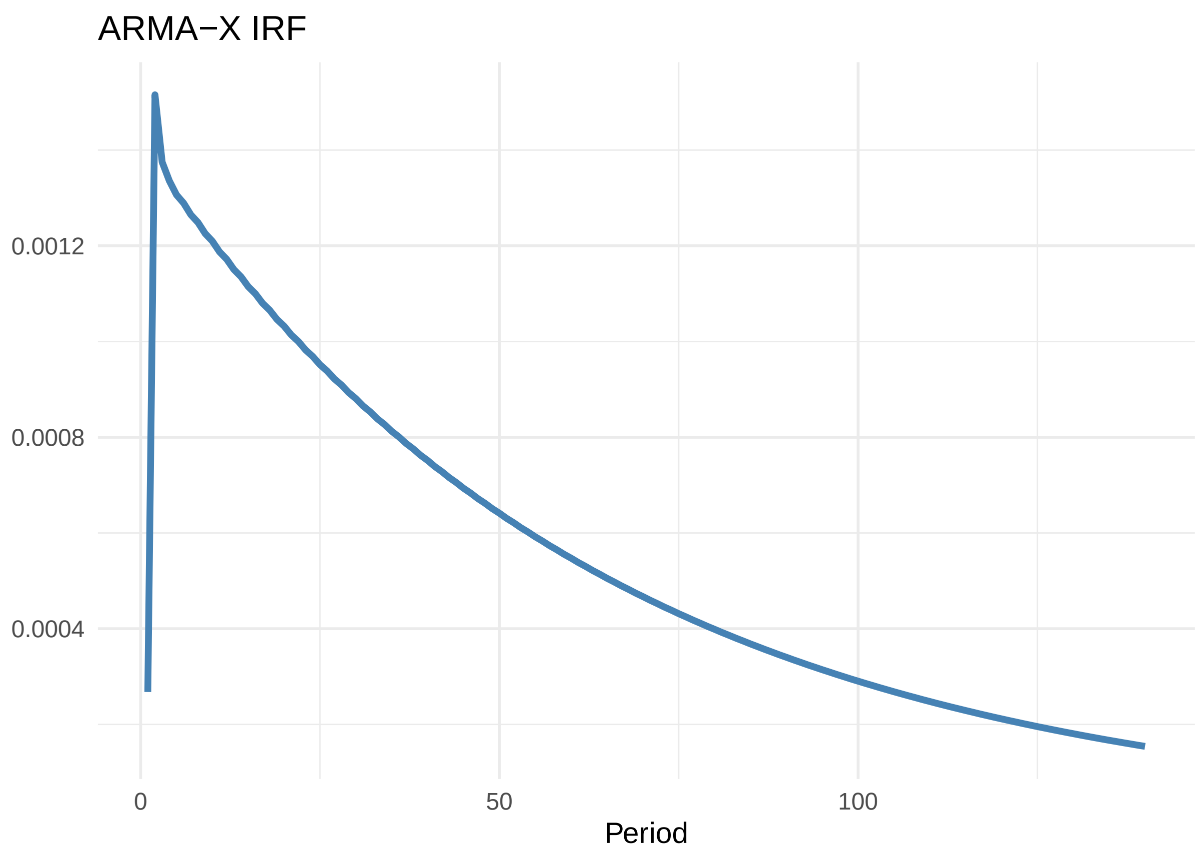
```r
#we want to plot the IRFs of these models
nb.periods = 7 * 20
irf.plot(res1,nb.periods)
```




```
irf.plot(res2,nb.periods)
```




```r
irf.plot(res3$model,nb.periods)
```


ARMA−X IRF
0.0012
0.0008
0.0004
0
50
100
Period