CTU-610 Machine-Learning & Neural-Networks
Final-exam project by: Alex Osterneck 11/12/23
Multivariate-time-series WTI-forecasting with VAR(p) and VARMA

Abstract

All terms and formulas in this project are defined in the next section, (prior to the introduction, presentation and conclusion,) under Definition-of-Terms. Most of the terms will be familiar to those with a statistics or R programming background, however they're included here for easy reference or if needed, refresher, and for a general audience with an interest in multivariate-time-series analysis forecasting with Vector-Auto-Regression (VAR / VAR(p)), and VARMA models. VAR and VARMA models can predict macroeconomic variables with high precision. (Athanasopoulos, G. and Vahid, F., 2008.)

The Vector Auto-regression Moving Average (VARMA) is a multi-variate statistical-model which can be used to predict future results based on prior results in time-series data. VARMA is an improvement over Auto-regressive-integrated-moving-average (ARIMA,) because VARMA monitors multiple stationary time-series simultaneously. The easiest way to think of VARMA is: a combination of VAR (Vector Auto-Regression,) and VMA (Vector Moving Average.) VARMA tests causation and can forecast multiple events occurring at the same time. (Mathworks, 2023)

A type of VARMA is the Vector auto-regression model (VAR(m)) model which reveals correlations in multiple time-series variables. Each VAR equation generates a linear-function which includes the variable and it's lag. Obtaining VAR outputs requires correct selection of lag, as well as monitoring heteroscedasticity and stationarity to avoid over-fitting. Determining the optimal hyper-parameters leads to the most accurate forecasts.

This project explores optimization of the coefficient-values in the following Multivariate Vector Auto-Regression formula to maximize the model's performance, where **w(t)** is one of two variables (the variable we wish to forecast,) and **a(t)** is the second of two variables, for which our multivariate statistical-model formulates the value of **w(t)** as the linear-combination of the prior-values of **w(t)** and **a(t)**:

w(t) =

$$\alpha_{\emptyset} + \alpha_1^{w} w(t-1) + \alpha_2^{w} w(t-2) + \ldots \alpha_{p_w}^{w} w(t-p_w) + \alpha_1^{a} a(t-1) + \alpha_2^{a} a(t-2) + \ldots \alpha_{p_a}^{a} a(t-p_a)$$

where, $\alpha_1{}^{w}$ = coefficients of the prior-values of **w(t)**

where, $\alpha_1{}^{a}$ = coefficients of the prior-values of **a(t)**

where, $\alpha_{\emptyset}$ = a time constant, and, **(t-1), (t-2)** the lags, and ...$(t-p_a)$ the p-th lag of $\alpha_{p_a}{}^{a}a$.

In simple terms, the coefficients are multiplied by the lagged-values of the time-series. To do this the VAR model generalizes like Auto-Regressive (AR) models with stationary time-series and lag-order. This implementation with additional data from the time-series can generate better predictions. The two keys to VAR are; number-of-variables to use (**k,**) and number-of-lags to use (**p**) (Hyndman, R.J., Athanasopoulos, G.,2018,) so the number of estimated coefficients in a VAR is:

$\mathbf{k} + \mathbf{p}\mathbf{k}^2$ (or: $1 + \mathbf{p}\mathbf{k}$ per equation)

where, **k** = number-of-variables

where, **p** = number-of-lags

So for example, a VAR with k = 5 variables, and p = 3 lags, has 16 coefficients per equation (3*5 + 1), and, (16*5) for a total of 80 coefficients to be estimated. (Hyndman, R.J., and Athanasopoulos, G.,2018.)

Another concept is the cross-correlations-function (CCF.) Examining this metric can determine how the 'second' time series is impacted by the prior lags of the 'first' time series. In this way, CCF can point to lags in a variable which might forecast another variable. (R Bloggers, 2021.) The definition of CCF is defined in the Definition of Terms.

Normalization, stationarity and ACF/PACF plots are presented, along with MAE and MAPE and for accuracy measurement.

Using MAE to measure our statistical-model accuracy based on the number ***(n)*** intervals, the project uses the formula:

$$\text{MAE} = {}^{1}/\,(n) \sum_{t=1}^{(n)} \left| \mathbf{w(t)} - \mathbf{\hat{w}(t)} \right|$$

where, $\mathbf{w(t)}$ = true

where, $\mathbf{\hat{w}(t)}$ = predicted

where, $\left| \mathbf{w(t)} - \mathbf{\hat{w}(t)} \right|$ = absolute-value of the difference of: true minus predicted

From MAE, the MAPE (Mean Absolute Percentage Error,) formula is the same as the MAE formula, with the additional operation of dividing by the 'true' term: $\left| (\mathbf{w(t)} - \mathbf{\hat{w}(t)}) / \mathbf{w(t)} \right|$

Definition of Terms

ACF / PACF = Auto Correlation Function / Partial Auto Correlation Function. ACF / PACF begin at lag Ø, lag Ø ('lag zero',) is the correlation of the time-series with itself, which equals correlation of 1. The difference between ACF and PACF is, ACF includes indirect correlations, PACF includes *direct* correlations.

Akaike's Information Criterion (AIC) = estimator of prediction error of statistical-model, where **k** is total parameters, lower AIC is best, see also (**SBC**, below): -2 * loglikelihood + 2 * **k** .

Autoregressive-oil-exchange-index (AOX) = proprietary AR index for forecasting WTI price, based on FRB WEI, AR 3-YR WTI futures, US 10-yr bond: **FRB Weekly Economic WEI + AR 3-yr futures WTI-nymex-price / TMUBMUSD10Y**

Causation = relationship between two events or two variables, where the first variable or event increases or decreases as a direct result of the presence or absence of a second variable or event.

Correlation = how two random variables vary together, using two random variables normalized by their respective spreads.

Covariance = how linearly related two random variable are which also indicates direction

Cross Correlation Function (CCF) = identifies lags or leads for two time-series. In this project defined as the set of correlations between **w(t)** + h, and **a(t)**, for h = +/- 1, +/- 2, +/- 3, etc...where a negative-value for h indicates a correlation between **w** at a time before t, and **a** at a time before t. The range of CCF is -1.0 to +1.0

Coefficients = numerical constant(s) placed before, and multiplying, a variable in an equation, in this project:
$\alpha_1^{w} \ldots \alpha_{p_w}^{w}$, $\alpha_1^{a} \ldots \alpha_{p_a}^{a}$, etc...

Differencing = change between consecutive data points in a time series, in this project:
w' (t) = **w(t)** - **w(t-1)** and/or **a' (t)** = **a(t)** - **a(t-1)** used to transform non-stationary data into stationary-data. Required to transform non-stationary-data into stationary for a VAR-model.

Dickey-Fuller test = test for null-hypothesis that a unit-root is present in auto-regressive (AR) time-series model. In this project used to confirm p-value < .05 to confirm stationarity.

Heteroscedasticity = variance of residual is unequal over the range of measured values whereby unequal variance may lead to invalid analysis, so log-transformation is used to morph the distribution closer to the normal-distribution.

Mean Squared Error (MAPE) = averaged squared difference between estimated (predicted,) and actual (true,) value (divided by the estimated(predicted) value, as the expected value of the squared-error-loss. In this project:

$$\text{MAPE} = {}^{1}/_{(n)} \sum_{t=1}^{(n)} \left| \left(w(t) - \hat{w}(t) \right) / w(t) \right|$$

Normalization = transform data to standard-scale, usually between 0 and 1, used with time-series-data when data has varying scales to ensure no single feature dominates the statistical-model forecasts. As it relates to stationarity: mean, variance and auto-correlation are made to remain constant over time. When combined with log-transformation or differencing it will lead to stationarity which is required for VAR models. A common method is: subtract the mean and divide by the standard-deviation for each column.

Schwarz Bayesian Criterion (SBC) = similar to Akaike's Information Criterion, uses penalty-term based on number of parameters in statistical-model, where **k** is total parameters, lower **SBC** is best. *This project uses **SBC*** score, which determines the number of lags to use: -2 * loglikelihood + **k***log(n) .

Seasonality = the mean of the time series is different based on a particular season. Removing seasonality is done to ensure the time-series statistical-properties remain constant over the time range.

Stationarity = time-series plots which don't show trends or seasonality are considered stationary. The test is; constant-mean and constant-variance. See diagram, below, in 'Stationarity, normalization, seasonality' section .

TMUBMUSD10Y = 10-year U.S. Treasury Note, in U.S. Dollars

Unit-root = feature of stochastic-process which can cause problems in statistical-inference of time-series models. A linear-stochastic-process has a unit-root of 1 , which is non-stationary, and therefore signals the data must be transformed through differencing. VAR models must be stationary.

Vector Auto Regression (VAR) = statistical-model which displays relationships between multiple variables and their change over time. Each variable has an equation which models it's change over time, which includes the variables lag (prior values,) the other variables lags, and an error term, $\epsilon(t)$.

Vector Moving Average (VMA) = generalized-version of moving-average model for multivariate stationary time-series.

Introduction

Vector auto-regression is just like uni-variate auto-regression, except vector auto-regression uses matrices, and matrix operations, to replace the scalars and scalar-operations of uni-variate auto-regression. (Sheppard, 2021.) The obvious question is; why do we need multi-variate-regression when we could simply use uni-variate-regressions? The answer is; because multi-variate-regression takes into account the correlations between the variables.

The mathematical-formulas presented in this project, along with the WTI data, and the AOX propriety-index have been converted into programming-code in the R programming-language for statistical-computing and graphics. The source code is translated from the steps and procedures described in the project paper. This source-code is included in the Appendix.

CCF is the cross-correlation-function which, (when implemented in R,) will allow visual examination of two time series and how their relationship is distributed over time. For example a spike at a lag equates to a correlation, and multiple spikes will each explain how the past or present of one series relates to the past or present of another series.

The first step in building a VAR model is to determine whether the multivariate approach is better than treating the signals seperately of a univariate time-series. This examination is done by testing the realtionship between variables using CCF. (Brooms, 2023)

Once the multivariate approach is determined as correct, stationarity is checked, whereby the next steps in building a VAR model are i) split the data into train and test sets, ii) select the VAR order p, fitting the VAR(p) model of order p on the train data, iii) generate predictions, and iv) evaluate model performance using MAPE.

The objective of this project is to explore VAR / VARMA for forecasting and prediction. By way of example, the project paper explores implementing VAR / VARMA to predict the price of West Texas Intermediate Crude (WTI) based on the prior prices of WTI, and the proprietary index, autoregressive-oil-exchange-index (AOX,) which equals: (FRB weekly-economic WEI + AR WTI 3-yr nymex-futures-price) / TMUBMUSD10Y. All data is calendar-year 2022.

Presentation

West-Texas-intermediate-crude (WTI)

(daily closing price, 2022)

cash-spot WTI-nymex-price

w(t) = WTI

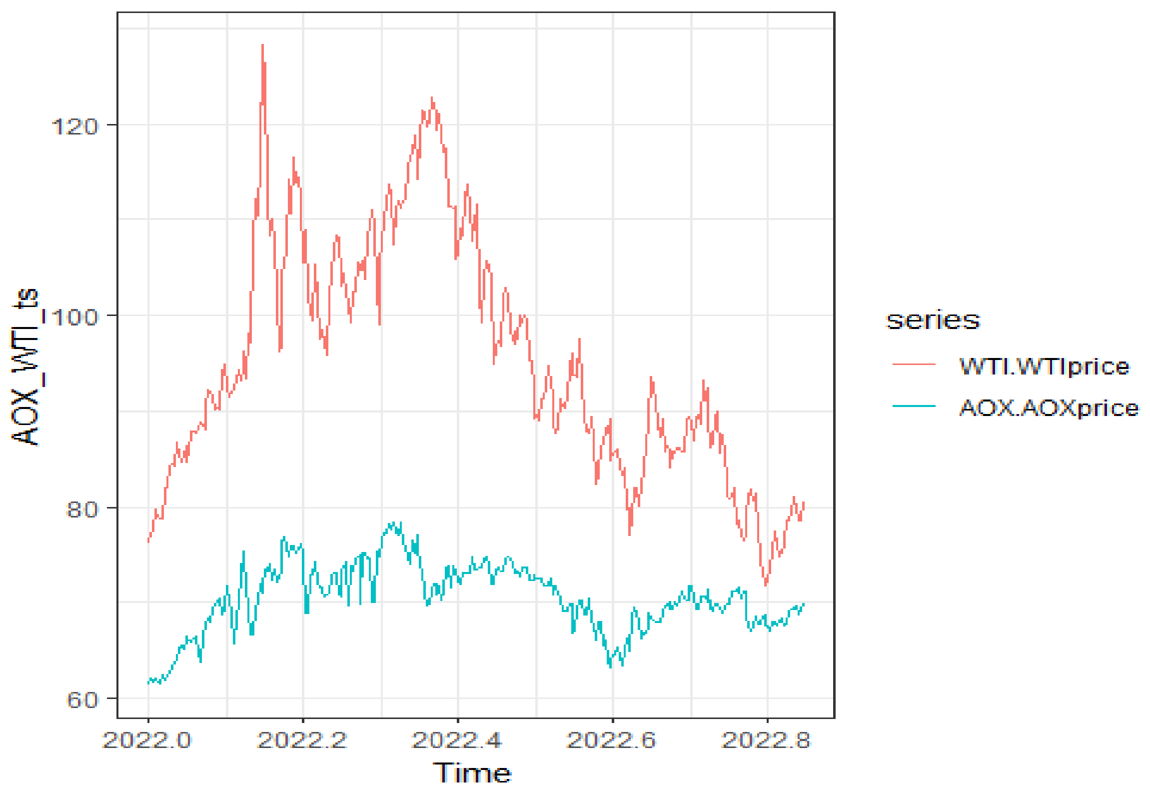
Autoregressive-oil-exchange-index (AOX)

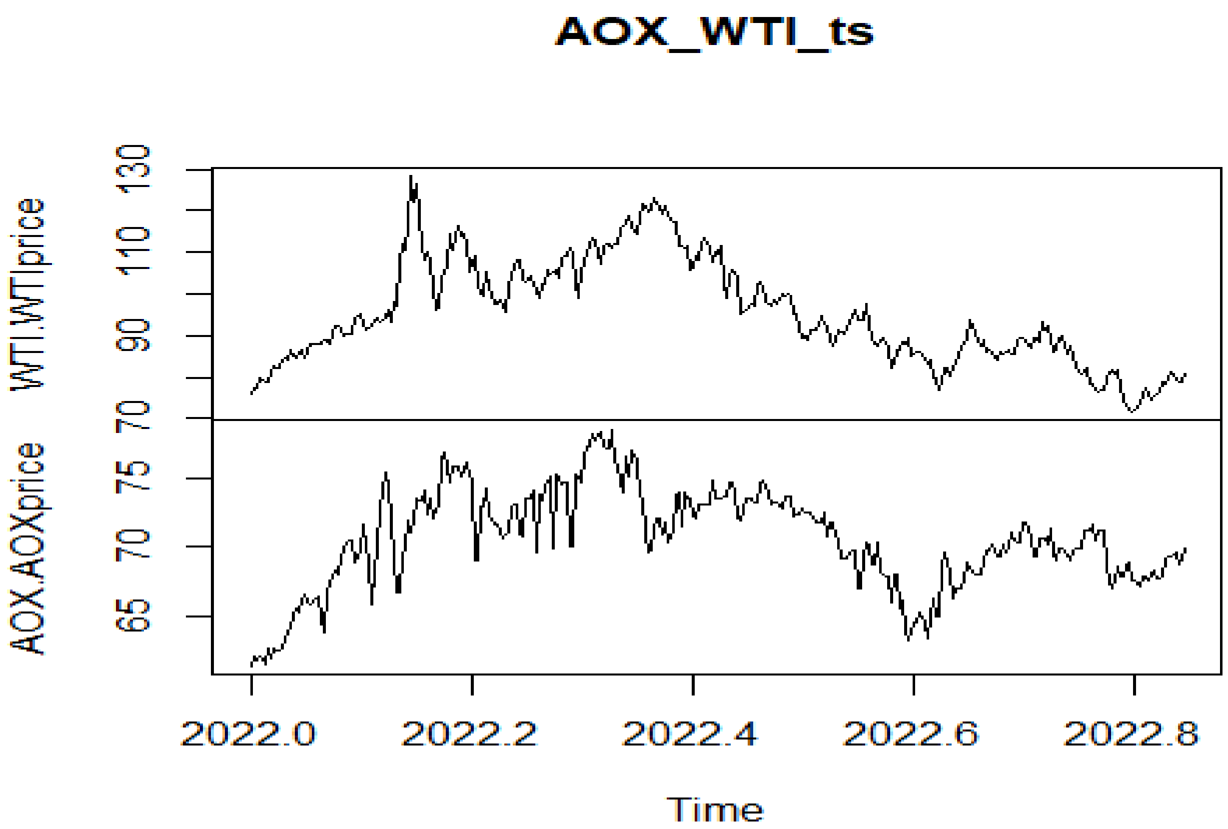
(daily closing price, 2022)

(FRB weekly-economic (WEI) + AR 3-yr futures WTI-nymex-price) / TMUBMUSD10Y

a(t) = AOX

Time-series visualization



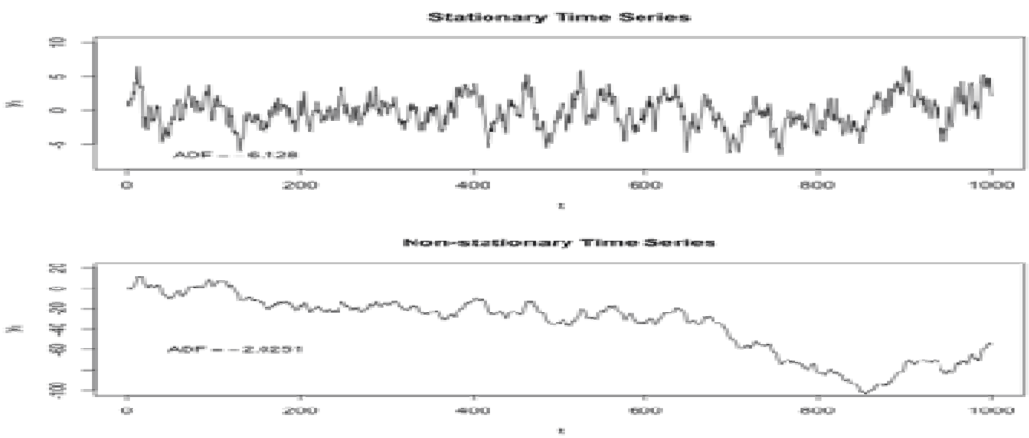


Stationarity, normalization, seasonality

If the time-series data has a unit-root it means there is non-stationarity. If the data has stationarity the p-value of the test will be low (pval < .05) which means the null-hypothesis can be rejected. The way to measure this is using the Augmented-Dickey-Fuller hypothesis test. (Wikipedia, 2023.) The intuition behind the test is that if the series is identified by a unit-root process, then the lag of the series **w(t-1)** will provide no relevant information in predicting the change in **w(t)** besides the one obtained in the lagged changes. Augmented-Dickey-Fuller is a test for statistical-significance via hypothesis testing (with null and alternative-hypotheses,) which generates a p-value to determine whether the time-series is stationary or not. The first step is to determine the presence of a unit-root in the time series, and then to calculate whether z(t) comprises a unit-root:

w(t) = **w(t-1)** + z(t) + Є(t)
where, z(t) = stochastic component
where, Є(t) = stationary error process.

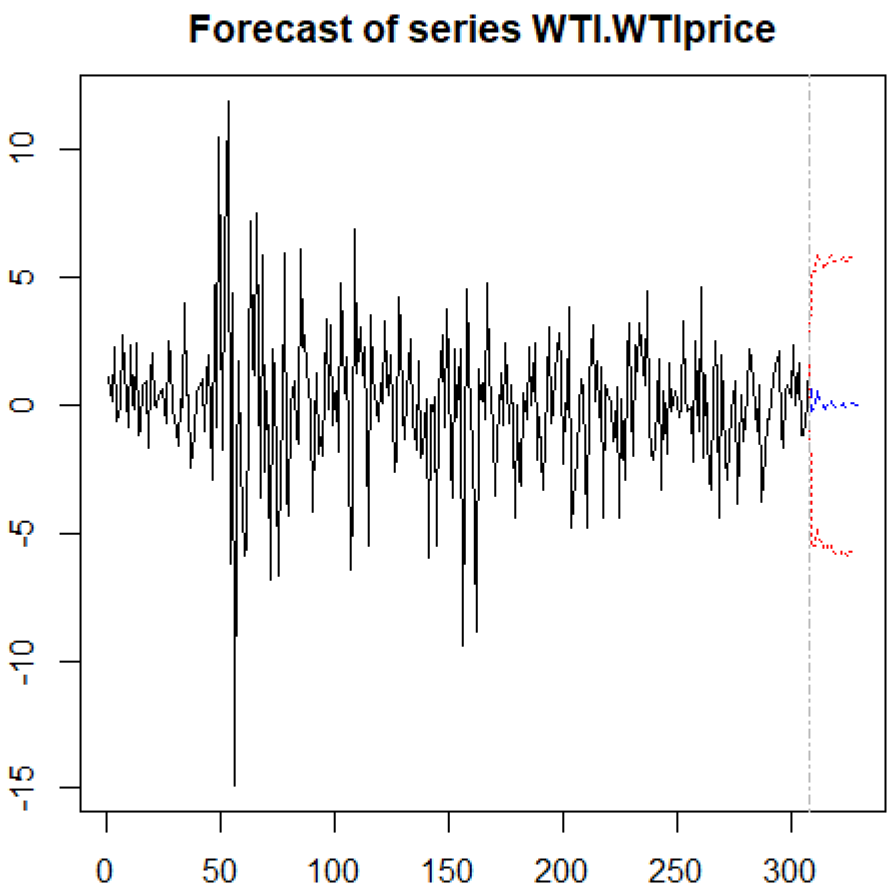
If required, non stationarity can be tranformed by removing seasonality or through normalization. Whether organic or transformed synthetically, the VAR model must be stationary. The following diagram displays the difference between stationary and non-stationary:

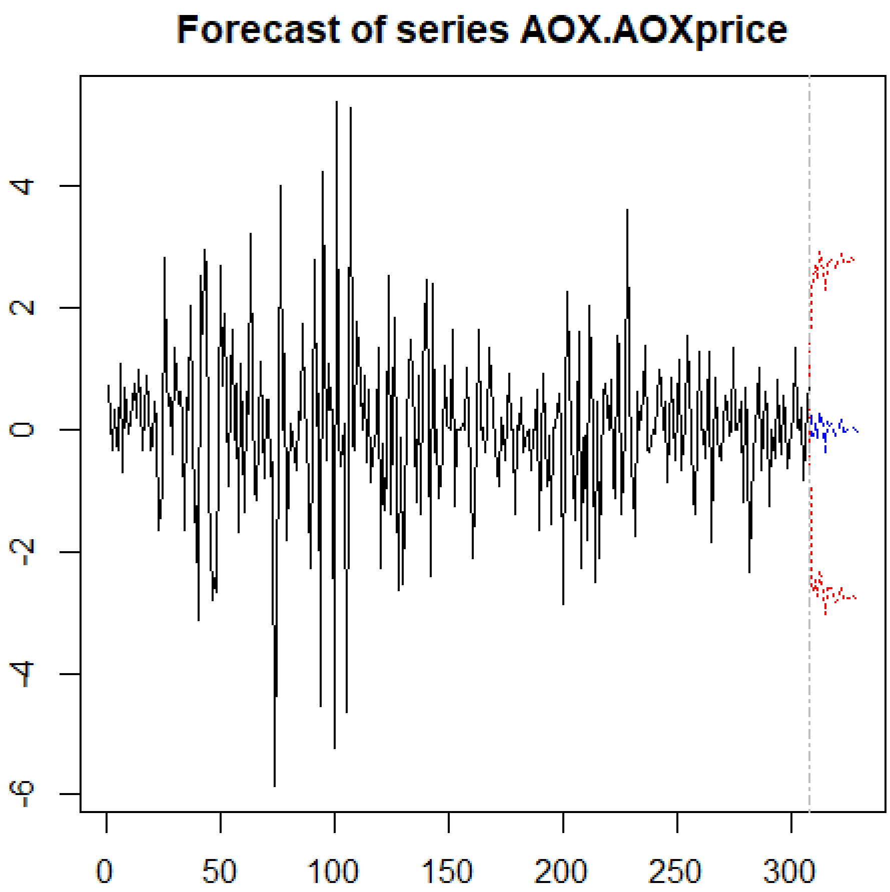


Normalization is simply subtracting the mean and then dividing by the standard-deviation for all columns in the data-set, although before implementing VAR, checking for stationarity is the main requirement. (StatsExchange, 2021.)

Removing seasonality is simply removing the variance as function of time, due to the mean increasing or decreasing with time. This project uses differencing to implement stationarity.

Stationarity forecasts on time-series price-columns (21-day range in red)





VAR model

The easiest way to think of a VAR model is everything depends on everything. (Mohr,2018.) With confirmation of the data being stationary, the VAR model is implemented, first, via AR-terms for **w(t)** (WTI,) and **a(t)** (AOX.) The objective of the project is to explore how VAR / VARMA can be used in forecasting and prediction. In this project, by way of example, to predict WTI based on AOX.

The next step is determine the WTI terms. In an auto-regressive (AR), time-series, the current value is a function of: the prior value(s) correlations, so, the current-value's prediction is calculated as: a linear-function of the previous-value and a constant. To do this use Partial-AutoCorrelation-Function (PACF.)

To get these linear-functions, PACF, in part, calculates the number of past lags to include in the prediction-equation of the model (AR-order of the model.) By building the PACF plot, the optimum features (AR-order,) is determined, as well as the amount of variance that is not explained by the variance in the preceding values, which is done by determining the correlation-coefficient between the values. Each partial auto-correlation is derived as a series of regressions, (Dewangga, S., 2014) for this project, the formula is:

$$\overline{W}(t) = \varnothing_{21}\,\overline{W}(t-1) + \varnothing_{22}\overline{W}(t-2) + \epsilon(t)$$

where, $\mathbf{w(t)} = (w)t$

where, $\overline{W}(t) = W(t) - \overline{W}$ (the original WTI time-series minus the sample-mean)

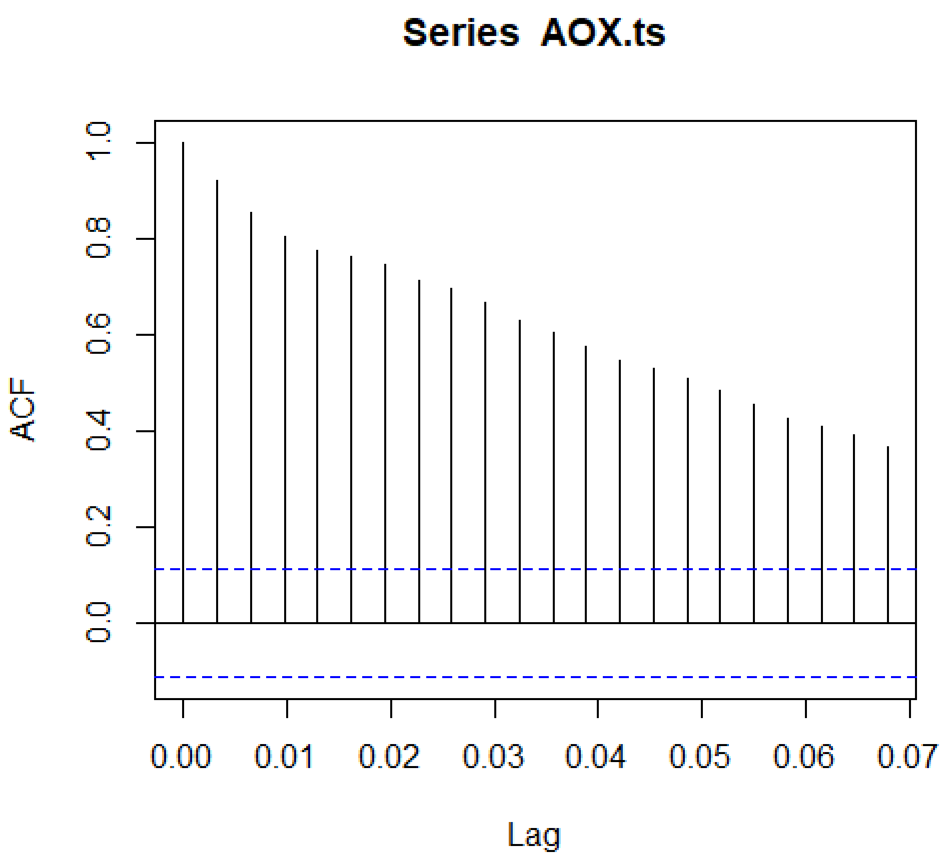
where, $\varnothing_{22} = \varnothing_{22}$ (the value of the partial auto-correlation of order 2)

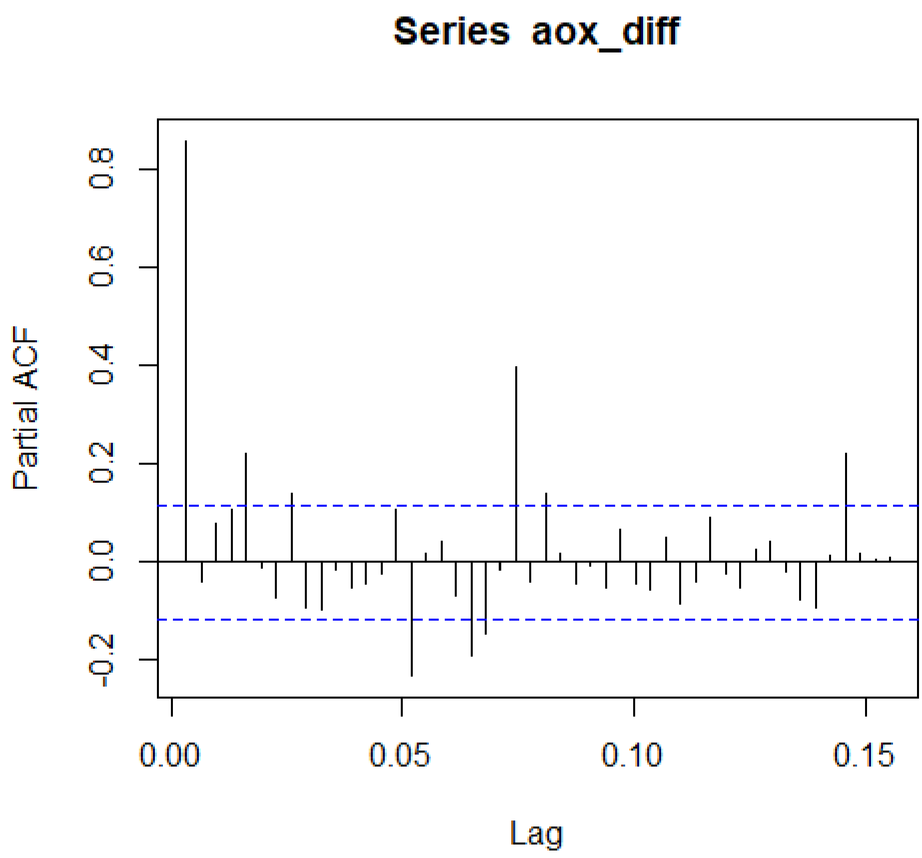
If the variance is significant then the value is added to the prediction-model. The key is the PACF plot removes variations explained by earlier lags so only the relevant features are displayed.

Viewing the PACF plot (next section,) the relevant features which are statistically-significant for what the project is predicting (the price of WTI based on AOX,) are the lines that rise above the horizontal-line area confidence-intervals. The first line is a constant, the second line = lag1, and the third line = lag2.

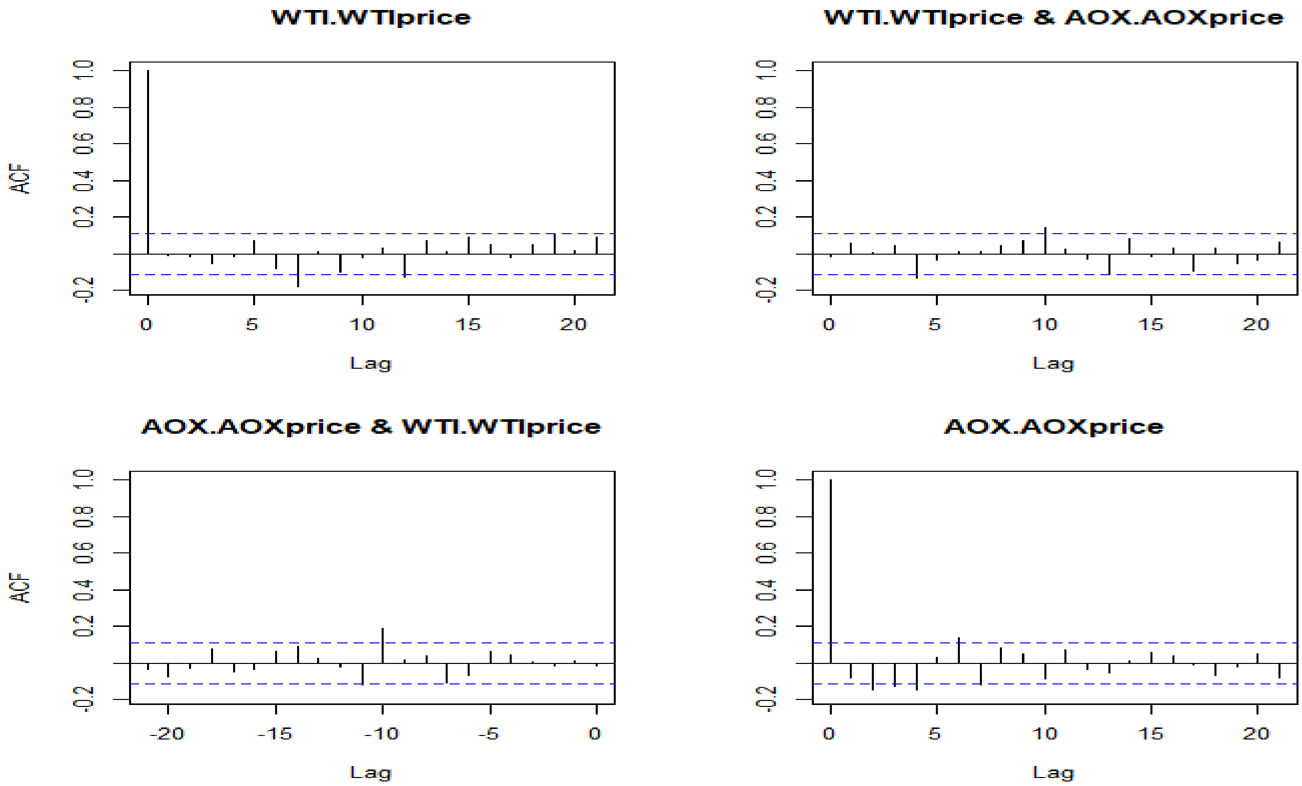
Now that the WTI features are formulated, the correct AR features for the predictor (the AOX index,) are formulated by calculating the correlated p-value of the price of WTI and the AOX index-lags. Viewing the plot (next section,) where correlations < .05 will provide the second part of the solution: determining which AOX-index lag(s) are most important for predicting the price of WTI.

ACF / PACF plots

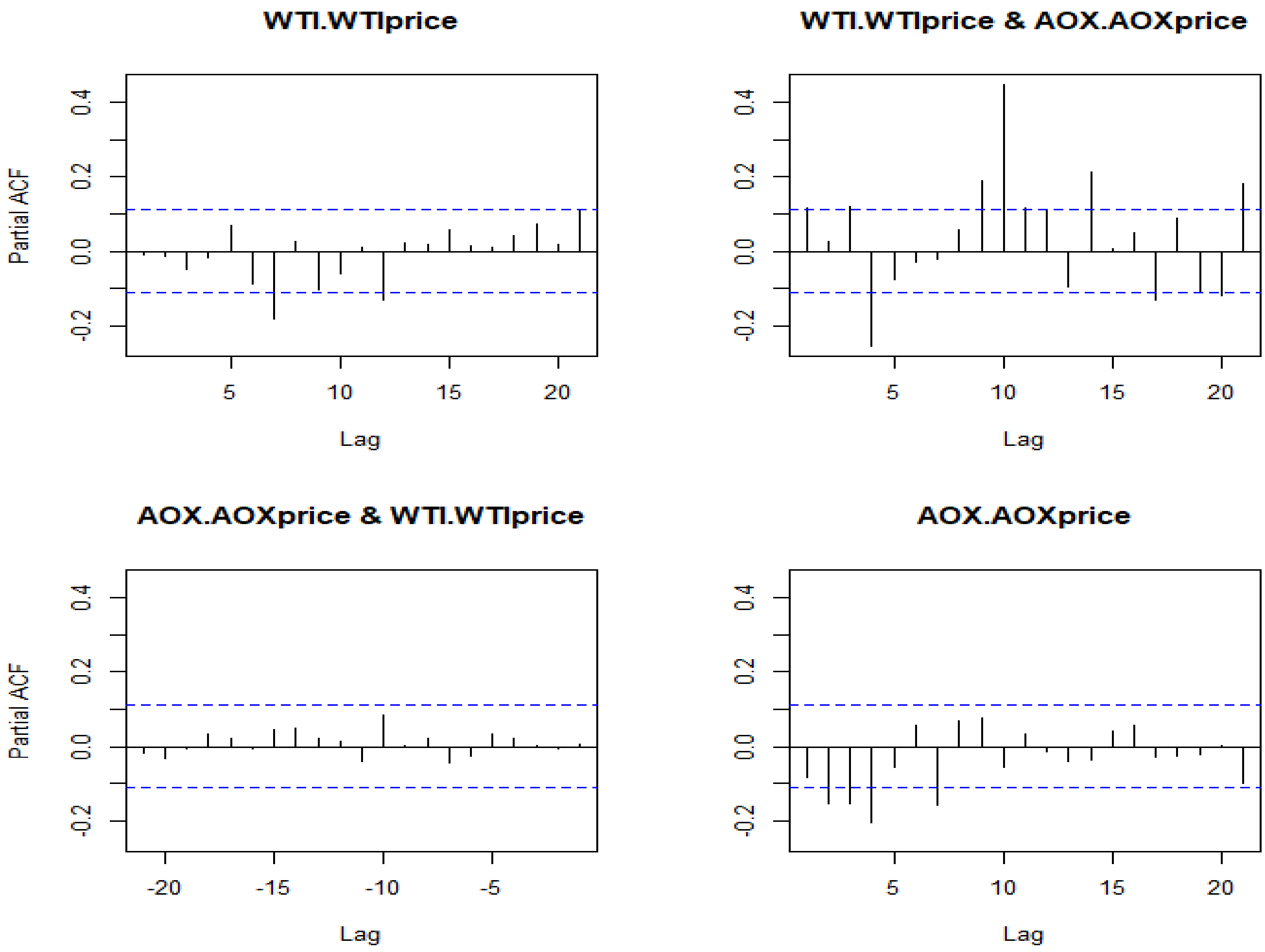




ACF Stationary



PACF Stationary



Conclusion

This project explored optimization of the coefficient-values of the presented Multivariate Vector Auto-Regression formula with the objective of maximization of the model's performance, where **w(t)** is the time-series dependent-variables to be predicted, and **a(t)** is the time-series of the independent-variables used for forecasting, for which the presented multivariate statistical-model formulates the forecast value of **w(t)** as the linear-combination of the prior-values of **w(t)** and **a(t)**.

To summarize, the data was visualized via plotting both the **w(t)** and **a(t)** time-series together and checking for stationarity via the Augmented Dickey-Fuller Test (ADF.) Normalization (subtracting the mean and dividing by the standard-deviation from each column,) and, removal of seasonality (by subtracting the mean and dividing by the standard-deviation for each column,) were preprocessing methods presented. Estimation of auto-lags and cross-lags and fitting the VAR model was done next. Training the model was completed by splitting the data into train and test sets (based on the date chronology,) and fitting the corresponding parameters to the regression model. The VAR model was fitted with both time-series, with focus on **w(t)** which was the time-series to be predicted. Lags with low p-value (p-val < .05) were used to regress the model, and ultimately using the SBC information-criteria, 2 lags were selected, then the model was applied to the test set, where the prediction **ŵ(t)** is a function of the coefficients corresponding to the low p-value terms. Finally, the model was evaluated using MAPE. The MAPE between the AOX proprietary-index and the WTI price was 17%, which is acceptable for forecasting, but of low accuracy, (see below.) However, more importantly, when the AOX proprietary-index was fed into the statistical-formula in conjunction with the correctly-identified lags of WTI, MAPE was 4%, which is considered acceptable *and* of high accuracy.

These procedures, more specifically, were completed by creating the two time-series, transforming them into a multivariate-time-series, visualizing the time series, checking for stationarity, differencing the time series, creating and checking the ACF / PACF and CCF plots, finding the number of lags, referencing the information-criterion's AIC and SBC to determine order, creating the VAR model, extracting estimates and checking the coefficients for significance, calculating impulse-response to interpret the estimated parameter values and determine how **w(t)** responds to a one-time shock in **a(t)**, confirming causality in the price of WTI with the AOX index, (highly causative,) and, forecasting on the VAR model with predictions.

Further study should include longitudinal research into the proprietary AOX index. Due to project time and resource-constraints only the most recent full-year (2022,) data was used, and therefore should not be considered exhaustive. A more vigorous research method than this project should include adjustment of parameters to potentially reduce MAPE below the 17% and 4% thresholds, looking for ways to eliminate redundancy, exploring other lag orders, and including decades of data, at which point consideration for potential commercial use cases might be explored.

Though the AOX index is, in plain terms, a reasonable preliminary predictor of WTI prices, especially for near-term pricing within a 10-day and 21-day window, further examination is required involving multi-year data and more robust lag optimization for greater veracity and tuning.

The Mean Absolute Percentage Error (MAPE,) as defined at the outset of the project was ~4%, for the 10-day forward forecast, which means an average 4% difference between the predicted WTI price and the actual WTI price. For reference, MAPE 5% or below is high ly accurate, MAPE greater than 10% but below 25% is acceptable, but is of low accuracy. And higher than 25% is very low accuracy and considered not acceptable.

Causality between WTI prices and AOX is more robust, with VAR(p) p-value of 0.01688, where a measure below .05 is considered highly causative.

By visual inspection, frequency and amplitude in both time-series are correlated, though certain amplitudes in the proprietary AOX index don't translate to equal amplitude for all peaks in WTI, and same magnitude of impulse in WTI prices. Other than that, in the least, AOX offers a baseline reference for preliminary assessment of near-term (days, weeks or a few months forward pricing) for forecasting WTI prices.

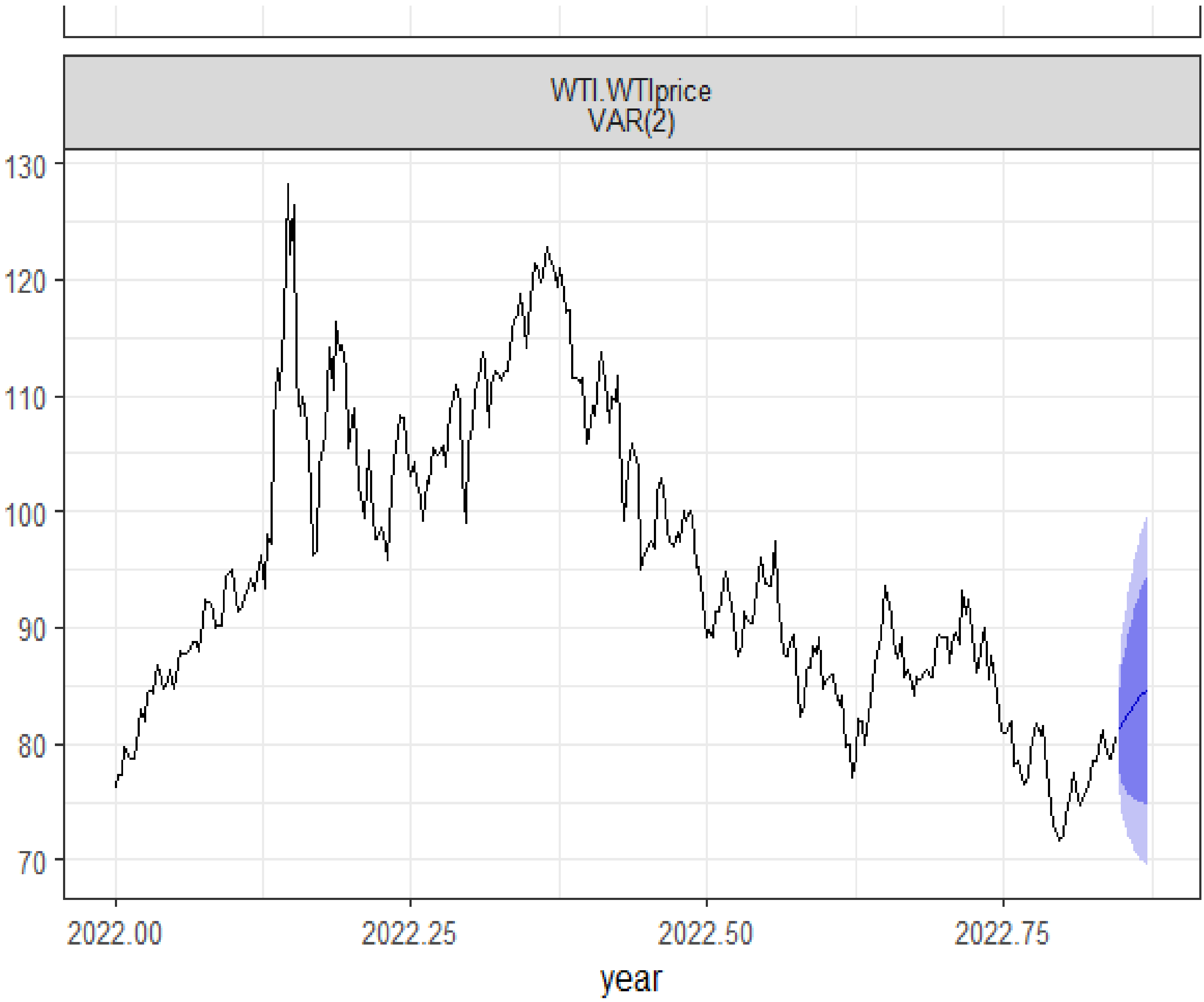
That said, the primary focus of this project-paper was the exploration of multivariate VAR / VARMA models and their potential and proven track-record in macro predictions via coefficient optimization and parameter optimization. Subordinate to the exploration of multivariate time-series analysis was the actual WTI price prediction. The statistical model presented in this project paper, can be inputted with *any* data converted into the correct time-series format.

Further still, professional individuals, private firms and government-institutions, have invested significant resources over many decades, exploring VAR / VARMA models, prediction of macro events and the forecasting of oil prices. Much of the research and results are of very high caliber and granularity.

Conversely, this project paper barely scratches the surface of the depth and potential applications of multivariate time series, for which applications in forecasting for industry, sports, weather, logistics, economics, retail, healthcare and cyber-security are sectors ripe for continued analysis.

The procedures and formulas in this project-paper have been converted and translated into source-code, coded with the R-statistical programming-language, presented in the final section: Appendix.

10-day Forward WTI Price Forecast



Appendix

Source-code

```
setwd("C:/Users/Papar_000/Desktop")
getwd()
#install.packages("reticulate")
library(reticulate)
py_install("pandas")
library(ggplot2)
library(stats)

AOX <- read.csv("AOX.csv")
WTI <- read.csv("WTI.csv")

# create time series (ts) by combining WTI-price & AOX-index
AOX_WTI <- data.frame(WTI$WTIprice,AOX$AOXprice)

# create ts from new combined df
# plot ts to compare and plot
AOX_WTI_ts <- ts(AOX_WTI,frequency=365,start=c(2022,1))
#length(AOX_WTI_ts)
plot(AOX_WTI_ts)

library(tidyverse)
library(MTS)
library(ggfortify)
library(forecast)
library(tseries)

# plot each new ts on same plot prior to pre-processing data
theme_set(theme_bw())
autoplot(AOX_WTI_ts)+theme(plot.title=element_text(hjust=0.5))

AOX_ts <- ts(AOX,frequency=365,start=c(2022,1))
WTI_ts <- ts(WTI,frequency=365,start=c(2022,1))

####normalization procedure included herein for when different data time-series on different scales used with this model
####install.packages('bestNormalize')
####library(bestNormalize)
####bestNormalize(AOX_ts,mode='scale')
####bestNormalize(WTI_ts,mode='scale')

# run Augmented Dickey-Fuller-Test (ADF) on each ts
# need (p-val) < .05, otherwise run ndiffs() to generate diff

AOX.ts <- ts(data=AOX$AOXprice,frequency=309,start=c(2022,1),end=c(2022,309))
adf.test(AOX.ts)

WTI.ts <- ts(data=WTI$WTIprice,frequency=309,start=c(2022,1),end=c(2022,309))
adf.test(WTI.ts)

ndiffs(AOX.ts)
ndiffs(WTI.ts)

# use output from ndiffs() to get new diff-object and re-check p-val

aox_diff <- diff(AOX.ts,differences=2,lag=22)
adf.test(aox_diff)
# plot(aox_diff,type="l",main="AOX index 2022")

wti_diff <- diff(WTI.ts,differences=1,lag=22)
adf.test(wti_diff)  ## see also: ts_adf_test
# plot(wti_diff,type="l",main="WTI price 2022")

# using 21 periods for auto-correlation and cross-correlation plots
acf(AOX.ts,21)
acf(WTI.ts,21)
ccf <- ccf(AOX.ts,WTI.ts,21)

library(devtools)
library(graphics)

# create new stationary object and confirm p-vals < .05 for stationarity
STATIONARY <- diffM(AOX_WTI_ts)
apply(STATIONARY,2,adf.test)

plot.ts(STATIONARY)
autoplot(ts(STATIONARY,start=c(2022,1),frequency=365))
```

```
# plot partial-acf's to determine lag significance
pacf(wti_diff,lag=length(wti_diff)-239,pl=TRUE)
pacf(aox_diff,lag=length(aox_diff)-239,pl=TRUE)
library(vars)
# to find number of lags which yields the most precise results
# in place of order (p), set max leg-length of the model
# view summary to see what AIC recommends for order

var.a <-vars::VAR(STATIONARY,lag.max=21,ic='AIC',type='none')
summary(var.a)

# extract coefficients generated by VAR
invisible(est_coefs <- coef(var.a))

# extract only coefficients for both dependent variables, combine into matrix
est_coefs <-rbind(est_coefs[[1]][,1],est_coefs[[2]][,1])
print(est_coefs)

VARselect(STATIONARY,type='trend',lag.max=10)$selection
VARselect(STATIONARY,type='trend',lag.max=10)

# calculate impulse-response to interpret the estimated parameter values and
# get reaction of response variable to one-time shock in an impulse-variable
# plot displays trajectory of 2nd series response after a shock to 1st series

IRF <- irf(var.a,impulse='AOX.AOXprice', response='WTI.WTIprice', n.ahead=21, ortho=FALSE)
plot(IRF)

# confirm p-val < 0.5 which means a(t) is causative of w(t) for VAR
causality(var.a,cause=c("AOX.AOXprice"))

# run forecast on VAR model to determine what price-range over next 50 days
# reminder: this is on pre-processed stationary data as required for VAR
forecast<-predict(var.a,n.ahead=50)
par(mar=c(2.5,2.5,2.5,2.5))
#print(forecast)
plot(forecast)

# now predict on VAR model 50 days ahead and confirm confidence-interval
pred <- predict(var.a,n.ahead=50,ci=0.95,dumvar=NULL)
#print(pred)
plot(pred,"single")

# install.packages("MLmetrics")
library(MLmetrics)

# extract forecast data
#if first line throws object-error use 2nd line first then enter 1st line
WTI.WTIprice <- forecast$fcst[2];AOX.AOXprice
AOX.AOXprice <- forecast$fcst[1];WTI.WTIprice

A <- WTI.WTIprice$AOX.AOXprice[,1];A
W <- AOX.AOXprice$WTI.WTIprice[,1];W

# get last values to add to cumulative sum for inversion
tail(AOX_WTI_ts)

# get original-scale-time-series via inversion of differencing
# AOX independent variable included for reference
# focus for remainder of code is on dependent variable WTI
#A <- cumsum(A) + 33.58
#par(mar=c(2.5,2.5,1,2.5))
#plot.ts(A)

W <- cumsum(W) + 80.55
par(mar=c(2.5,2.5,1,2.5))
plot.ts(W)

# transform prediction and original-scale-time-series into time-series
#AOXinv <- ts(c(AOX_WTI_ts[,1],A),start=c(2022,1),end=c(2022,50),frequency=365)
#plot(AOXinv)

WTIinv <- ts(c(AOX_WTI_ts[,1],W),start=c(2022,1),end=c(2022,50),frequency=365)
#plot(WTIinv)

#AOXinv_dataframe <- as.data.frame(AOXinv)
#colnames(AOXinv_dataframe) <- c('A')
#tail(AOXinv_dataframe)

WTIinv_dataframe <- as.data.frame(WTIinv)
colnames(WTIinv_dataframe) <- c('W')
head(WTIinv_dataframe)
```

```
library(MLmetrics)
#MAPE(pred,true)

par(mfrow=c(1,2))
acf(STATIONARY,lag=length(wti_diff)-266,pl=TRUE)
pacf(STATIONARY,lag=length(wti_diff)-266,pl=TRUE)

AOX50 <- head(AOX.ts,50)
MAPE(AOX50,W)

#MAPE between 10% and 25% is low accuracy but still in acceptable range.

#now run VAR(p), which is VAR(2) model on combined multivariate ts
#check with serial-test and then compare prediction vs actual

VARselect(AOX_WTI_ts[,1:2],lag.max=8,type="const")[["selection"]]

var2 <- VAR(AOX_WTI_ts[,1:2],p=2,type='const')
var2

serial.test(var2, lags.pt=2,type="PT.asymptotic")

forecast(var2)%>%autoplot()+xlab("year")
head(forecast(var2))

forecast(var2)

fV <-  as.data.frame(forecast(var2))
WTI_forecast <- fV[1:10,  3]

WTI_forecast

WTI_actual <-head(WTI$WTIprice,n=10)
MAPE(WTI_forecast,WTI_actual)
```

References

Athanasopoulos, G., and Vahid, F., (2008, Apr) *VARMA versus VAR for Macroeconomic Forecasting*, Journal for Business & Economic Statistics, https://www.jstor.org/stable/27638977

Brooms, A.C.,(2023,Oct 23) *Stochastic Models and Forecasting, Chapter 9 , Multivariate time series*, University of London, https://www.bbk.ac.uk/ms/brooms/teaching/SMF/SMFl9.pdf

Dewangga, S.(2014, Dec 14th,) *ACF and PACF Formula,* Stats Exchange, https://stats.stackexchange.com/q/129052

Hardy, R., (2021, Apr 19) *Should I scale my dataset before applying VAR?,* Stats.StackExchange. https://stats.stackexchange.com/questions/520462/should-i-scale-my-dataset-before-applyingvar

Hyndman, R.J., and Athanasopoulos, (2018) Forecasting: Principles and Practice , Chapter 11, *Advanced forecasting methods 11.2, Vector Autoregressions*, https://otexts.com/fpp2/VAR.html

Luetkepohl, H. (2005, Jan) *The New Introduction to Multiple Time Series Analysis*, Var Model Selection and checking Model Adequacy, ResearchGate, https://www.researchgate.net/publication/37404453_The_New_Introduction_to_Multiple_Time_Series_Analysis

MathWorks(2023, Oct 12) Vector Autoregression (VAR) Models http://www.mathworks.com/help/econ/introduction-to-vector-autoregressive-var-models.html

Mohr, Frank X. (2018, Aug 31st) *An introduction to Vector Autoregression (VAR)* r-econometrics https://www.r-econometrics.com/timeseries/varintro/

FinnStats in R Bloggers (2021, Aug 11) *How to Calculate Cross-Correlation in R*, R Bloggers https://www.r-bloggers.com/2021/08/how-to-calculate-cross-correlations-in-r/

Sheppard, K., (2021, Feb 2nd) *Financial Econometric Notes*, University of Oxford. https://www.kevinsheppard.com/files/teaching/mfe/notes/financial-econometrics-2020-2021.pdf

How to analyze correlation of multivariate time series. (2019, June 5) In *Stats.StackExchange*. https://stats.stackexchange.com/questions/409407/how-to-analyze-correlation-of-multivariate-time-series

R-stat Console-print 11/12/23 16:25

```
> setwd("C:/Users/Papar_000/Desktop")
> getwd()
[1] "C:/Users/Papar_000/Desktop"
> #install.packages("reticulate")
> library(reticulate)
> py_install("pandas")
+ "C:/Users/papar_000/anaconda3/condabin/conda.bat" "install" "--yes" "--prefix"
 "C:/Users/papar_000/anaconda3" "-c" "conda-forge" "pandas"Collecting package me
tadata (current_repodata.json): ...working... DEBUG:urllib3.connectionpool:Start
ing new HTTPS connection (1): repo.anaconda.com:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): repo.anaconda.co
m:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): repo.anaconda.co
m:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): conda.anaconda.o
rg:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): repo.anaconda.co
m:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): repo.anaconda.co
m:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): repo.anaconda.co
m:443
DEBUG:urllib3.connectionpool:Starting new HTTPS connection (1): conda.anaconda.o
rg:443
DEBUG:urllib3.connectionpool:https://repo.anaconda.com:443 "GET /pkgs/r/win-64/c
urrent_repodata.json HTTP/1.1" 304 0
DEBUG:urllib3.connectionpool:https://repo.anaconda.com:443 "GET /pkgs/msys2/noar
ch/current_repodata.json HTTP/1.1" 304 0
DEBUG:urllib3.connectionpool:https://repo.anaconda.com:443 "GET /pkgs/main/win-6
4/current_repodata.json HTTP/1.1" 304 0
DEBUG:urllib3.connectionpool:https://repo.anaconda.com:443 "GET /pkgs/main/noarc
h/current_repodata.json HTTP/1.1" 304 0
DEBUG:urllib3.connectionpool:https://repo.anaconda.com:443 "GET /pkgs/msys2/win-
64/current_repodata.json HTTP/1.1" 304 0
DEBUG:urllib3.connectionpool:https://repo.anaconda.com:443 "GET /pkgs/r/noarch/c
urrent_repodata.json HTTP/1.1" 304 0
DEBUG:urllib3.connectionpool:https://conda.anaconda.org:443 "GET /conda-forge/wi
n-64/current_repodata.json HTTP/1.1" 200 None
DEBUG:urllib3.connectionpool:https://conda.anaconda.org:443 "GET /conda-forge/no
arch/current_repodata.json HTTP/1.1" 304 0
done
Solving environment: ...working... done

==> WARNING: A newer version of conda exists. <==
  current version: 23.7.2
  latest version: 23.10.0
Please update conda by running

    $ conda update -n base -c defaults conda
Or to minimize the number of packages updated during conda update use
     conda install conda=23.10.0
```

```
# All requested packages already installed.
> library(ggplot2)
> library(stats)
> AOX <- read.csv("AOX.csv")
> WTI <- read.csv("WTI.csv")
> # create time series (ts) by combining WTI-price & AOX-index
> AOX_WTI <- data.frame(WTI$WTIprice,AOX$AOXprice)
> # create ts from new combined df
> # plot ts to compare and plot
> AOX_WTI_ts <- ts(AOX_WTI,frequency=365,start=c(2022,1))
> #length(AOX_WTI_ts)
> plot(AOX_WTI_ts)
> library(tidyverse)
> library(MTS)
> library(ggfortify)
> library(forecast)
> library(tseries)
> # plot each new ts on same plot prior to pre-processing data
> theme_set(theme_bw())
> autoplot(AOX_WTI_ts)+theme(plot.title=element_text(hjust=0.5))
> AOX_ts <- ts(AOX,frequency=365,start=c(2022,1))
> WTI_ts <- ts(WTI,frequency=365,start=c(2022,1))
> ####normalization procedure included herein for when different data time-series
 on different scales used with this model
> ####install.packages('bestNormalize')
> ####library(bestNormalize)
> ####bestNormalize(AOX_ts,mode='scale')
> ####bestNormalize(WTI_ts,mode='scale')
> # run Augmented Dickey-Fuller-Test (ADF) on each ts
> # need (p-val) < .05, otherwise run ndiffs() to generate diff
> AOX.ts <- ts(data=AOX$AOXprice,frequency=309,start=c(2022,1),end=c(2022,309))
> adf.test(AOX.ts)

       Augmented Dickey-Fuller Test

data:  AOX.ts
Dickey-Fuller = -3.0057, Lag order = 6, p-value = 0.1524
alternative hypothesis: stationary
> WTI.ts <- ts(data=WTI$WTIprice,frequency=309,start=c(2022,1),end=c(2022,309))
> adf.test(WTI.ts)
       Augmented Dickey-Fuller Test

data:  WTI.ts
Dickey-Fuller = -2.9181, Lag order = 6, p-value = 0.1893
alternative hypothesis: stationary
> ndiffs(AOX.ts)
[1] 1
 >ndiffs(WTI.ts)
[1]
```

```
> # use output from ndiffs() to get new diff-object and re-check p-val
> aox_diff <- diff(AOX.ts,differences=2,lag=22)
> adf.test(aox_diff)

        Augmented Dickey-Fuller Test

data:  aox_diff
Dickey-Fuller = -3.3976, Lag order = 6, p-value = 0.05531
alternative hypothesis: stationary

> # plot(aox_diff,type="l",main="AOX index 2022")
> wti_diff <- diff(WTI.ts,differences=1,lag=22)
> adf.test(wti_diff)  ## see also: ts_adf_test

        Augmented Dickey-Fuller Test

data:  wti_diff
Dickey-Fuller = -3.7344, Lag order = 6, p-value = 0.02271
alternative hypothesis: stationary

> # plot(wti_diff,type="l",main="WTI price 2022")
> # using 21 periods for auto-correlation and cross-correlation plots
> acf(AOX.ts,21)
> acf(WTI.ts,21)
> ccf <- ccf(AOX.ts,WTI.ts,21)
> library(devtools)
> library(graphics)
> # create new stationary object and confirm p-vals < .05 for stationarity
> STATIONARY <- diffM(AOX_WTI_ts)
> apply(STATIONARY,2,adf.test)

$WTI.WTIprice
        Augmented Dickey-Fuller Test

data:  newX[, i]
Dickey-Fuller = -8.4025, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

$AOX.AOXprice
        Augmented Dickey-Fuller Test

data:  newX[, i]
Dickey-Fuller = -9.143, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
Warning messages:
1: In FUN(newX[, i], ...) : p-value smaller than printed p-value
2: In FUN(newX[, i], ...) : p-value smaller than printed p-value
> plot.ts(STATIONARY)
> autoplot(ts(STATIONARY,start=c(2022,1),frequency=365))
> # plot partial-acf's to determine lag significance
> pacf(wti_diff,lag=length(wti_diff)-239,pl=TRUE)
> pacf(aox_diff,lag=length(aox_diff)-239,pl=TRUE)
> library(vars)
> # to find number of lags which yields the most precise results
```

```
> # in place of order (p), set max leg-length of the model
> # view summary to see what AIC recommends for order
> var.a <-vars::VAR(STATIONARY,lag.max=21,ic='AIC',type='none')
> summary(var.a)

VAR Estimation Results:
=========================
Endogenous variables: WTI.WTIprice, AOX.AOXprice
Deterministic variables: none
Sample size: 298
Log Likelihood: -1198.075
Roots of the characteristic polynomial:
0.8945 0.8945 0.8907 0.8907 0.8819 0.8819 0.8768 0.8768 0.8675 0.8675 0.8652 0.86
37 0.8637 0.8562 0.8562 0.8423 0.8423 0.7763 0.6816 0.6816
Call:
vars::VAR(y = STATIONARY, type = "none", lag.max = 21, ic = "AIC")

Estimation results for equation WTI.WTIprice:
=============================================
WTI.WTIprice = WTI.WTIprice.l1 + AOX.AOXprice.l1 + WTI.WTIprice.l2 + AOX.AOXprice.
l2 + WTI.WTIprice.l3 + AOX.AOXprice.l3 + WTI.WTIprice.l4 + AOX.AOXprice.l4 + WTI.
WTIprice.l5 + AOX.AOXprice.l5 + WTI.WTIprice.l6 + AOX.AOXprice.l6 + WTI.WTIprice.
l7 + AOX.AOXprice.l7 + WTI.WTIprice.l8 + AOX.AOXprice.l8 + WTI.WTIprice.l9 + AOX.
AOXprice.l9 + WTI.WTIprice.l10 + AOX.AOXprice.l10

                 Estimate Std. Error t value Pr(>|t|)
WTI.WTIprice.l1  -0.038605   0.058607  -0.659 0.510626
AOX.AOXprice.l1   0.041243   0.125970   0.327 0.743611
WTI.WTIprice.l2  -0.020709   0.058060  -0.357 0.721605
AOX.AOXprice.l2  -0.020854   0.126830  -0.164 0.869519
WTI.WTIprice.l3  -0.068835   0.058011  -1.187 0.236406
AOX.AOXprice.l3   0.121069   0.128314   0.944 0.346227
WTI.WTIprice.l4  -0.016846   0.056961  -0.296 0.767650
AOX.AOXprice.l4  -0.273577   0.129212  -2.117 0.035125 *
WTI.WTIprice.l5   0.083584   0.056884   1.469 0.142862
AOX.AOXprice.l5   0.005337   0.132189   0.040 0.967824
WTI.WTIprice.l6  -0.072418   0.057051  -1.269 0.205371
AOX.AOXprice.l6   0.117725   0.132139   0.891 0.373741
WTI.WTIprice.l7  -0.204698   0.056902  -3.597 0.000380 ***
AOX.AOXprice.l7   0.097926   0.130153   0.752 0.452451
WTI.WTIprice.l8   0.008029   0.058035   0.138 0.890061
AOX.AOXprice.l8   0.177473   0.128778   1.378 0.169272
WTI.WTIprice.l9  -0.115940   0.058029  -1.998 0.046696 *
AOX.AOXprice.l9   0.251728   0.126902   1.984 0.048279 *
WTI.WTIprice.l10 -0.059472   0.058249  -1.021 0.308142
AOX.AOXprice.l10  0.449614   0.126317   3.559 0.000437 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 2.733 on 278 degrees of freedom
Multiple R-Squared: 0.1332, Adjusted R-squared: 0.07084
F-statistic: 2.136 on 20 and 278 DF,  p-value: 0.003722
```

```
Estimation results for equation AOX.AOXprice:
=============================================
AOX.AOXprice = WTI.WTIprice.l1 + AOX.AOXprice.l1 + WTI.WTIprice.l2 + AOX.AOXprice.
l2 + WTI.WTIprice.l3 + AOX.AOXprice.l3 + WTI.WTIprice.l4 + AOX.AOXprice.l4 + WTI.
WTIprice.l5 + AOX.AOXprice.l5 + WTI.WTIprice.l6 + AOX.AOXprice.l6 + WTI.WTIprice.
l7 + AOX.AOXprice.l7 + WTI.WTIprice.l8 + AOX.AOXprice.l8 + WTI.WTIprice.l9 + AOX.
AOXprice.l9 + WTI.WTIprice.l10 + AOX.AOXprice.l10

                 Estimate Std. Error t value Pr(>|t|)
WTI.WTIprice.l1   0.026748   0.027465   0.974  0.33096
AOX.AOXprice.l1  -0.131485   0.059033  -2.227  0.02673 *
WTI.WTIprice.l2   0.007106   0.027209   0.261  0.79417
AOX.AOXprice.l2  -0.190342   0.059437  -3.202  0.00152 **
WTI.WTIprice.l3   0.010333   0.027186   0.380  0.70418
AOX.AOXprice.l3  -0.192056   0.060132  -3.194  0.00157 **
WTI.WTIprice.l4   0.040128   0.026694   1.503  0.13390
AOX.AOXprice.l4  -0.192437   0.060553  -3.178  0.00165 **
WTI.WTIprice.l5   0.028794   0.026658   1.080  0.28101
AOX.AOXprice.l5  -0.058385   0.061948  -0.942  0.34676
WTI.WTIprice.l6  -0.024767   0.026736  -0.926  0.35506
AOX.AOXprice.l6   0.052113   0.061924   0.842  0.40075
WTI.WTIprice.l7  -0.037763   0.026666  -1.416  0.15785
AOX.AOXprice.l7  -0.143013   0.060994  -2.345  0.01975 *
WTI.WTIprice.l8   0.022278   0.027197   0.819  0.41341
AOX.AOXprice.l8   0.074044   0.060349   1.227  0.22089
WTI.WTIprice.l9   0.003061   0.027194   0.113  0.91047
AOX.AOXprice.l9   0.057042   0.059470   0.959  0.33830
WTI.WTIprice.l10  0.084109   0.027297   3.081  0.00227 **
AOX.AOXprice.l10 -0.054686   0.059196  -0.924  0.35638
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 1.281 on 278 degrees of freedom
Multiple R-Squared: 0.1793, Adjusted R-squared: 0.1203
F-statistic: 3.037 on 20 and 278 DF,  p-value: 2.297e-05



Covariance matrix of residuals:
             WTI.WTIprice AOX.AOXprice
WTI.WTIprice       7.4690      -0.1573
AOX.AOXprice      -0.1573       1.6387

Correlation matrix of residuals:
             WTI.WTIprice AOX.AOXprice
WTI.WTIprice      1.00000     -0.04496
AOX.AOXprice     -0.04496      1.00000

> # extract coefficients generated by VAR
> invisible(est_coefs <- coef(var.a))
> # extract only coefficients for both dependent variables, combine into matrix
> est_coefs <-rbind(est_coefs[[1]][,1],est_coefs[[2]][,1])
> print(est_coefs)
```

```
WTI.WTIprice.l1 AOX.AOXprice.l1 WTI.WTIprice.l2 AOX.AOXprice.l2 WTI.WTIprice.l3
AOX.AOXprice.l3 WTI.WTIprice.l4
[1,]     -0.03860487      0.04124257    -0.020708525     -0.02085365     -0.068834
94       0.1210691     -0.01684571
[2,]      0.02674783     -0.13148534     0.007105585     -0.19034204      0.010332
67      -0.1920556      0.04012824

AOX.AOXprice.l4 WTI.WTIprice.l5 AOX.AOXprice.l5 WTI.WTIprice.l6 AOX.AOXprice.l6
WTI.WTIprice.l7 AOX.AOXprice.l7
[1,]      -0.2735774      0.08358387     0.005337031     -0.07241822       0.117725
3     -0.20469784      0.09792631
[2,]      -0.1924367      0.02879407    -0.058384946     -0.02476682       0.05211
34     -0.03776264     -0.14301289

WTI.WTIprice.l8 AOX.AOXprice.l8 WTI.WTIprice.l9 AOX.AOXprice.l9 WTI.WTIprice.l10
 AOX.AOXprice.l10
[1,]     0.008029295      0.17747317    -0.115939846      0.25172767      -0.05947
204       0.44961387
[2,]     0.022277884      0.07404403     0.003060783      0.05704247       0.084109
27      -0.05468614

> VARselect(STATIONARY,type='trend',lag.max=10)$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
    10      1      1     10

$criteria
               1         2         3         4         5         6         7         8
        9        10
AIC(n)  2.737526  2.740198  2.737430  2.703252  2.715216  2.727216  2.681083  2.6
99245  2.700683  2.644672
HQ(n)   2.767322  2.789860  2.806956  2.792642  2.824471  2.856335  2.830067  2.8
68094  2.889397  2.853251
SC(n)   2.811964  2.864262  2.911119  2.926566  2.988155  3.049781  3.053273  3.1
21061  3.172124  3.165739
FPE(n) 15.448732 15.490156 15.447499 14.928742 15.108880 15.291947 14.603382 14.8
72196 14.895057 14.085422

> # calculate impulse-response to interpret the estimated parameter values and
> # get reaction of response variable to one-time shock in an impulse-variable
> # plot displays trajectory of 2nd series response after a shock to 1st series
> IRF <- irf(var.a,impulse='AOX.AOXprice', response='WTI.WTIprice', n.ahead=21,
ortho=FALSE)> plot(IRF)
> # confirm p-val < 0.5 which means a(t) is causative of w(t) for VAR
> causality(var.a,cause=c("AOX.AOXprice"))

$Granger

        Granger causality H0: AOX.AOXprice do not Granger-cause WTI.WTIprice

data:  VAR object var.a
F-Test = 2.1937, df1 = 10, df2 = 556, p-value = 0.01688
```

```
$Instant

	H0: No instantaneous causality between: AOX.AOXprice and WTI.WTIprice

data:  VAR object var.a
Chi-squared = 0.61325, df = 1, p-value = 0.4336

> # run forecast on VAR model to determine what price-range over next 50 days
> # reminder: this is on pre-processed stationary data as required for VAR
> forecast<-predict(var.a,n.ahead=50)
> par(mar=c(2.5,2.5,2.5,2.5))
> #print(forecast)
> plot(forecast)

# now predict on VAR model 50 days ahead and confirm confidence-interval
> pred <- predict(var.a,n.ahead=50,ci=0.95,dumvar=NULL)
> #print(pred)
> plot(pred,"single")
Hit <Return> to see next plot:

> # install.packages("MLmetrics")
> library(MLmetrics)
> # extract forecast data
> #if first line throws object-error use 2nd line first then enter 1st line
> WTI.WTIprice <- forecast$fcst[2];AOX.AOXprice
$WTI.WTIprice
              fcst     lower    upper       CI
 [1,] -0.2366329594 -5.593628 5.120363 5.356996
 [2,] -0.1440367788 -5.506204 5.218130 5.362167
 [3,] -0.1464820946 -5.509926 5.216962 5.363444
 [4,]  0.5940411946 -4.790592 5.978674 5.384633
 [5,]  0.2167586759 -5.219723 5.653241 5.436482
 [6,] -0.0020065715 -5.456175 5.452161 5.454168
 [7,] -0.1775705320 -5.660432 5.305291 5.482861
 [8,]  0.0237565776 -5.569868 5.617381 5.593624
 [9,]  0.1632381759 -5.446691 5.773168 5.609930
[10,]  0.1783671779 -5.474195 5.830929 5.652562
[11,] -0.0254497451 -5.722305 5.671406 5.696855
[12,] -0.0814317899 -5.782871 5.620008 5.701440
[13,] -0.0404882841 -5.758068 5.677092 5.717580
[14,] -0.0424950083 -5.774401 5.689411 5.731906
[15,]  0.0736811216 -5.658261 5.805624 5.731943
[16,] -0.0707121047 -5.802770 5.661346 5.732058
[17,] -0.1121976153 -5.849088 5.624693 5.736890
[18,]  0.0254082720 -5.717696 5.768512 5.743104
[19,]  0.0493324559 -5.696106 5.794771 5.745438
[20,]  0.0707032424 -5.676225 5.817631 5.746928
[21,] -0.0092322864 -5.756313 5.737848 5.747080
[22,] -0.0413525956 -5.789099 5.706394 5.747746
[23,]  0.0300047219 -5.719519 5.779528 5.749524
[24,]  0.0522367187 -5.697901 5.802374 5.750138
[25,] -0.0141215727 -5.764474 5.736231 5.750353
[26,] -0.0087800353 -5.759470 5.741910 5.750690
[27,] -0.0229350206 -5.774030 5.728160 5.751095
```

```
[28,]  0.0130872664 -5.738619 5.764794 5.751706
[29,]  0.0137618255 -5.737963 5.765487 5.751725
[30,] -0.0120613505 -5.763917 5.739794 5.751855
[31,] -0.0152147096 -5.767174 5.736744 5.751959
[32,] -0.0019867544 -5.754168 5.750195 5.752182
[33,] -0.0015087447 -5.753796 5.750778 5.752287
[34,]  0.0038452789 -5.748479 5.756170 5.752324
[35,] -0.0020796780 -5.754447 5.750288 5.752368
[36,] -0.0043826351 -5.756752 5.747987 5.752370
[37,]  0.0027063270 -5.749681 5.755094 5.752387
[38,]  0.0062875456 -5.746104 5.758679 5.752391
[39,]  0.0043903374 -5.748025 5.756805 5.752415
[40,]  0.0012971624 -5.751134 5.753728 5.752431
[41,] -0.0018521126 -5.754308 5.750603 5.752455
[42,] -0.0005521788 -5.753029 5.751925 5.752477
[43,]  0.0026136544 -5.749864 5.755091 5.752478
[44,]  0.0008640934 -5.751616 5.753344 5.752480
[45,] -0.0031731856 -5.755654 5.749307 5.752480
[46,] -0.0017247774 -5.754206 5.750756 5.752481
[47,] -0.0006307222 -5.753112 5.751851 5.752482
[48,]  0.0006944616 -5.751789 5.753178 5.752483
[49,]  0.0001405308 -5.752348 5.752629 5.752489
[50,] -0.0015883283 -5.754080 5.750903 5.752491

> AOX.AOXprice <- forecast$fcst[1];WTI.WTIprice
$AOX.AOXprice
               fcst     lower    upper       CI
 [1,] -9.831019e-02 -2.608760 2.412139 2.510449
 [2,] -9.413285e-02 -2.631088 2.442822 2.536955
 [3,]  1.064267e-01 -2.467132 2.679986 2.573559
 [4,] -1.295090e-01 -2.728876 2.469858 2.599367
 [5,]  3.059426e-01 -2.314907 2.926792 2.620849
 [6,] -1.990551e-02 -2.643581 2.603770 2.623676
 [7,] -3.654813e-01 -3.019869 2.288906 2.654387
 [8,]  1.188925e-01 -2.560084 2.797869 2.678977
 [9,]  7.192180e-02 -2.617833 2.761677 2.689755
[10,]  1.024938e-01 -2.590884 2.795872 2.693378
[11,]  1.143089e-02 -2.735761 2.758623 2.747192
[12,] -8.887856e-02 -2.842515 2.664758 2.753637
[13,] -3.219124e-02 -2.790056 2.725673 2.757864
[14,]  1.431901e-01 -2.615214 2.901595 2.758404
[15,] -5.081901e-02 -2.810195 2.708557 2.759376
[16,] -2.737002e-02 -2.787138 2.732398 2.759768
[17,] -7.920663e-03 -2.768756 2.752914 2.760835
[18,] -6.904963e-03 -2.772301 2.758491 2.765396
[19,]  4.155585e-02 -2.724611 2.807723 2.766167
[20,]  2.145413e-02 -2.745175 2.788083 2.766629
[21,] -4.503336e-02 -2.812799 2.722732 2.767765
[22,]  4.719017e-03 -2.764787 2.774225 2.769506
[23,]  1.172923e-02 -2.758231 2.781690 2.769961
[24,] -8.974777e-03 -2.778939 2.760990 2.769965
[25,]  1.338764e-02 -2.756733 2.783508 2.770120
[26,] -1.583550e-02 -2.786312 2.754641 2.770477
[27,] -1.405738e-02 -2.784566 2.756451 2.770508
```

```
[28,]  2.173083e-02 -2.749071 2.792533 2.770802
[29,]  2.372937e-03 -2.768512 2.773258 2.770885
[30,] -2.984918e-03 -2.773946 2.767977 2.770961
[31,]  1.858277e-03 -2.769130 2.772846 2.770988
[32,] -8.443908e-03 -2.779590 2.762703 2.771147
[33,]  4.886170e-03 -2.766262 2.776034 2.771148
[34,]  8.665835e-03 -2.762485 2.779817 2.771151
[35,] -9.738358e-03 -2.780911 2.761435 2.771173
[36,]  8.742534e-05 -2.771107 2.771282 2.771195
[37,]  1.478130e-03 -2.769720 2.772676 2.771198
[38,] -6.448676e-04 -2.771872 2.770583 2.771227
[39,]  3.517908e-03 -2.767719 2.774755 2.771237
[40,] -1.864347e-03 -2.773102 2.769373 2.771238
[41,] -3.858315e-03 -2.775096 2.767380 2.771238
[42,]  3.440037e-03 -2.767814 2.774694 2.771254
[43,] -1.489000e-04 -2.771404 2.771106 2.771255
[44,] -8.219251e-04 -2.772078 2.770434 2.771256
[45,]  7.744024e-04 -2.770482 2.772031 2.771256
[46,] -1.467376e-03 -2.772726 2.769792 2.771259
[47,]  6.104127e-04 -2.770649 2.771870 2.771260
[48,]  1.748941e-03 -2.769512 2.773010 2.771261
[49,] -1.067592e-03 -2.772331 2.770195 2.771263
[50,] -1.449233e-04 -2.771408 2.771118 2.771263

> A <- WTI.WTIprice$AOX.AOXprice[,1];A
[1] -9.831019e-02 -9.413285e-02  1.064267e-01 -1.295090e-01  3.059426e-01 -1.9905
51e-02 -3.654813e-01  1.188925e-01
 [9]  7.192180e-02  1.024938e-01  1.143089e-02 -8.887856e-02 -3.219124e-02  1.431
901e-01 -5.081901e-02 -2.737002e-02
[17] -7.920663e-03 -6.904963e-03  4.155585e-02  2.145413e-02 -4.503336e-02  4.719
017e-03  1.172923e-02 -8.974777e-03
[25]  1.338764e-02 -1.583550e-02 -1.405738e-02  2.173083e-02  2.372937e-03 -2.984
918e-03  1.858277e-03 -8.443908e-03
[33]  4.886170e-03  8.665835e-03 -9.738358e-03  8.742534e-05  1.478130e-03 -6.448
676e-04  3.517908e-03 -1.864347e-03
[41] -3.858315e-03  3.440037e-03 -1.489000e-04 -8.219251e-04  7.744024e-04 -1.467
376e-03  6.104127e-04  1.748941e-03
[49] -1.067592e-03 -1.449233e-04

> W <- AOX.AOXprice$WTI.WTIprice[,1];W
[1] -0.2366329594 -0.1440367788 -0.1464820946  0.5940411946  0.2167586759 -0.0020
065715 -0.1775705320  0.0237565776
 [9]  0.1632381759  0.1783671779 -0.0254497451 -0.0814317899 -0.0404882841 -0.042
4950083  0.0736811216 -0.0707121047
[17] -0.1121976153  0.0254082720  0.0493324559  0.0707032424 -0.0092322864 -0.041
3525956  0.0300047219  0.0522367187
[25] -0.0141215727 -0.0087800353 -0.0229350206  0.0130872664  0.0137618255 -0.012
0613505 -0.0152147096 -0.0019867544
[33] -0.0015087447  0.0038452789 -0.0020796780 -0.0043826351  0.0027063270  0.006
2875456  0.0043903374  0.0012971624
[41] -0.0018521126 -0.0005521788  0.0026136544  0.0008640934 -0.0031731856 -0.001
7247774 -0.0006307222  0.0006944616
[49]  0.0001405308 -0.0015883283
```

```
> # get last values to add to cumulative sum for inversion
> tail(AOX_WTI_ts)
Time Series:
Start = c(2022, 304)
End = c(2022, 309)
Frequency = 365
         WTI.WTIprice AOX.AOXprice
2022.830        79.44       69.250
2022.833        81.06       69.250
2022.836        79.84       69.600
2022.838        78.67       68.780
2022.841        78.67       69.212
2022.844        80.55       69.960


> # get original-scale-time-series via inversion of differencing
> # AOX independent variable included for reference
> # focus for remainder of code is on dependent variable WTI
> #A <- cumsum(A) + 33.58
> #par(mar=c(2.5,2.5,1,2.5))
> #plot.ts(A)
> W <- cumsum(W) + 80.55
> par(mar=c(2.5,2.5,1,2.5))
> plot.ts(W)
> # transform prediction and original-scale-time-series into time-series
> #AOXinv <- ts(c(AOX_WTI_ts[,1],A),start=c(2022,1),end=c(2022,50),frequency=365)
> #plot(AOXinv)
> WTIinv <- ts(c(AOX_WTI_ts[,1],W),start=c(2022,1),end=c(2022,50),frequency=365)
> #plot(WTIinv)
> #AOXinv_dataframe <- as.data.frame(AOXinv)
> #colnames(AOXinv_dataframe) <- c('A')
> #tail(AOXinv_dataframe)
> WTIinv_dataframe <- as.data.frame(WTIinv)
> colnames(WTIinv_dataframe) <- c('W')> head(WTIinv_dataframe)
W
1 76.17
2 77.25
3 77.39
4 79.69
5 79.07
6 78.70
> library(MLmetrics)
> #MAPE(pred,true)
> par(mfrow=c(1,2))
> acf(STATIONARY,lag=length(wti_diff)-266,pl=TRUE)
> pacf(STATIONARY,lag=length(wti_diff)-266,pl=TRUE)
> AOX50 <- head(AOX.ts,50)
> MAPE(AOX50,W)
[1] 0.1749586

> #MAPE between 10% and 25% is low accuracy but still in acceptable range.
> #now run VAR(p), which is VAR(2) model on combined multivariate ts
> #check with serial-test and then compare prediction vs actual
```

```
> VARselect(AOX_WTI_ts[,1:2],lag.max=8,type="const")[["selection"]]
AIC(n)  HQ(n)  SC(n) FPE(n) 
     1      1      1      1 
> var2 <- VAR(AOX_WTI_ts[,1:2],p=2,type='const')
> var2

VAR Estimation Results:
======================= 

Estimated coefficients for equation WTI.WTIprice: 
================================================= 
Call:
WTI.WTIprice = WTI.WTIprice.l1 + AOX.AOXprice.l1 + WTI.WTIprice.l2 + AOX.AOXprice.
l2 + const 

WTI.WTIprice.l1 AOX.AOXprice.l1 WTI.WTIprice.l2 AOX.AOXprice.l2           const 
     0.97289436      0.15398720     -0.01335045     -0.07819877     -1.48071800 


Estimated coefficients for equation AOX.AOXprice: 
================================================= 
Call:
AOX.AOXprice = WTI.WTIprice.l1 + AOX.AOXprice.l1 + WTI.WTIprice.l2 + AOX.AOXprice.
l2 + const 

WTI.WTIprice.l1 AOX.AOXprice.l1 WTI.WTIprice.l2 AOX.AOXprice.l2           const 
     0.01454613      0.83980726      0.01211248      0.01719991      7.57276569 

> serial.test(var2, lags.pt=2,type="PT.asymptotic")
       Portmanteau Test (asymptotic)

data:  Residuals of VAR object var2
Chi-squared = 3.4642, df = 0, p-value < 2.2e-16

> forecast(var2)%>%autoplot()+xlab("year")
> head(forecast(var2))
$model

VAR Estimation Results:
======================= 

Estimated coefficients for equation WTI.WTIprice: 
================================================= 
Call:
WTI.WTIprice = WTI.WTIprice.l1 + AOX.AOXprice.l1 + WTI.WTIprice.l2 + AOX.AOXprice.
l2 + const 

WTI.WTIprice.l1 AOX.AOXprice.l1 WTI.WTIprice.l2 AOX.AOXprice.l2           const 
     0.97289436      0.15398720     -0.01335045     -0.07819877     -1.48071800 
```

```
Estimated coefficients for equation AOX.AOXprice:
=================================================
Call:
AOX.AOXprice = WTI.WTIprice.l1 + AOX.AOXprice.l1 + WTI.WTIprice.l2 + AOX.AOXprice.
l2 + const

WTI.WTIprice.l1 AOX.AOXprice.l1 WTI.WTIprice.l2 AOX.AOXprice.l2           const
     0.01454613      0.83980726      0.01211248      0.01719991      7.57276569


$forecast
$forecast$WTI.WTIprice
          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022.8466       81.19629 77.60876 84.78383 75.70963 86.68295
2022.8493       81.69231 76.67876 86.70586 74.02475 89.35987
2022.8521       82.15687 76.11276 88.20097 72.91321 91.40052
2022.8548       82.59227 75.72324 89.46131 72.08699 93.09755
2022.8575       83.00196 75.44057 90.56334 71.43782 94.56610
2022.8603       83.38892 75.22986 91.54798 70.91072 95.86712
2022.8630       83.75571 75.07092 92.44049 70.47348 97.03794
2022.8658       84.10450 74.95101 93.25800 70.10544 98.10357
2022.8685       84.43716 74.86155 94.01278 69.79252 99.08181
2022.8712       84.75529 74.79651 94.71406 69.52466 99.98591

$forecast$AOX.AOXprice
          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022.8466       69.64070 67.96865 71.31275 67.08352 72.19788
2022.8493       69.41759 67.23318 71.60199 66.07683 72.75835
2022.8521       69.23977 66.73734 71.74220 65.41263 73.06691
2022.8548       69.09936 66.37694 71.82178 64.93578 73.26294
2022.8575       68.99035 66.10616 71.87454 64.57936 73.40134
2022.8603       68.90762 65.89884 71.91639 64.30610 73.50914
2022.8630       68.84686 65.73833 71.95538 64.09278 73.60093
2022.8658       68.80443 65.61328 71.99557 63.92399 73.68486
2022.8685       68.77727 65.51563 72.03890 63.78903 73.76550
2022.8712       68.76279 65.43951 72.08607 63.68027 73.84531


$method
WTI.WTIprice AOX.AOXprice
    "VAR(2)"     "VAR(2)"
```

```
> forecast(var2)
WTI.WTIprice
          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022.8466       81.19629 77.60876 84.78383 75.70963 86.68295
2022.8493       81.69231 76.67876 86.70586 74.02475 89.35987
2022.8521       82.15687 76.11276 88.20097 72.91321 91.40052
2022.8548       82.59227 75.72324 89.46131 72.08699 93.09755
2022.8575       83.00196 75.44057 90.56334 71.43782 94.56610
2022.8603       83.38892 75.22986 91.54798 70.91072 95.86712
2022.8630       83.75571 75.07092 92.44049 70.47348 97.03794
2022.8658       84.10450 74.95101 93.25800 70.10544 98.10357
2022.8685       84.43716 74.86155 94.01278 69.79252 99.08181
2022.8712       84.75529 74.79651 94.71406 69.52466 99.98591

AOX.AOXprice
          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022.8466       69.64070 67.96865 71.31275 67.08352 72.19788
2022.8493       69.41759 67.23318 71.60199 66.07683 72.75835
2022.8521       69.23977 66.73734 71.74220 65.41263 73.06691
2022.8548       69.09936 66.37694 71.82178 64.93578 73.26294
2022.8575       68.99035 66.10616 71.87454 64.57936 73.40134
2022.8603       68.90762 65.89884 71.91639 64.30610 73.50914
2022.8630       68.84686 65.73833 71.95538 64.09278 73.60093
2022.8658       68.80443 65.61328 71.99557 63.92399 73.68486
2022.8685       68.77727 65.51563 72.03890 63.78903 73.76550
2022.8712       68.76279 65.43951 72.08607 63.68027 73.84531

> fV <-  as.data.frame(forecast(var2))
> WTI_forecast <- fV[1:10,  3]

> WTI_forecast
[1] 81.19629 81.69231 82.15687 82.59227 83.00196 83.38892 83.75571 84.10450 84.43716 84.7552

> WTI_actual <-head(WTI$WTIprice,n=10)

> MAPE(WTI_forecast,WTI_actual)
[1] 0.0480284
```