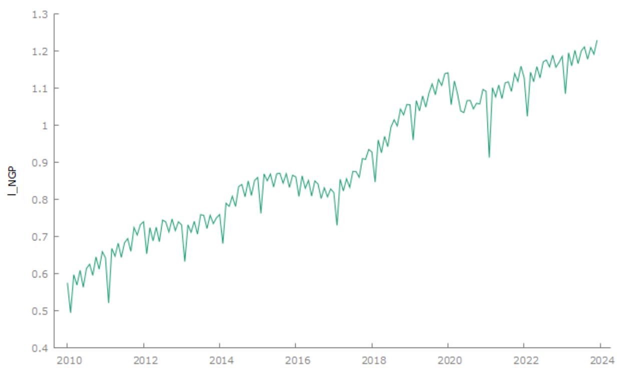
Time Series for Finance & Economics - Final Project
Ziad El Harairi, Juan D. Saavedra, Juan D. Ocampo

1.
The dataset selected for this analysis was the production of Natural Gas (Dry) in the US, the dataset contains monthly observations from January 1973 to December 2023, for the current project only observations from January 2010 to December 2023 are taken into consideration.



1. Time series plot of Natural Gas Production (01/2010-12/2023)

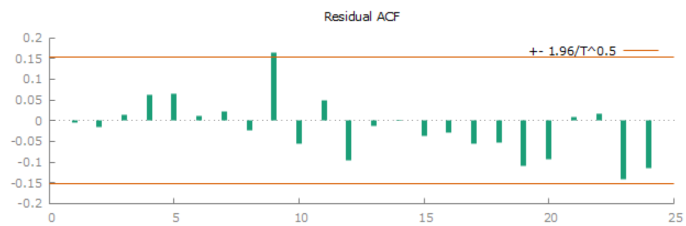
We identified several key trends and patterns. Firstly, there is a clear upward trajectory in production over time, indicating a general increase in output. This upward trend may be influenced by various factors such as technological advancements, shifts in market demand, or changes in regulatory policies affecting extraction and distribution. Secondly, we have detected a notable seasonal pattern in the data, with production levels consistently dipping during the month of January. This recurring decrease suggests a possible seasonal effect, which could be attributed to factors such as fluctuations in consumer demand during the winter season. Moreover, within each month, we observe a cyclical pattern of intra-monthly fluctuations of production levels, characterized by regular peaks and troughs.

2.

For the ARIMA model selection, we started by transforming logarithmically the variable, creating an initial **AR (1)**, the model result in an **r2=0.9459** and a **BIC=-558**,when checking the residual correlograms, autocorrelation was spotted on several lags.

For the second iteration a model **AR (1)** with seasonal dummies was considered improving the model to an **r2=0.9923** and **BIC=-841**, while the correlogram show less autocorrelation in multiple lags, there was still some at lag 1 and 9.

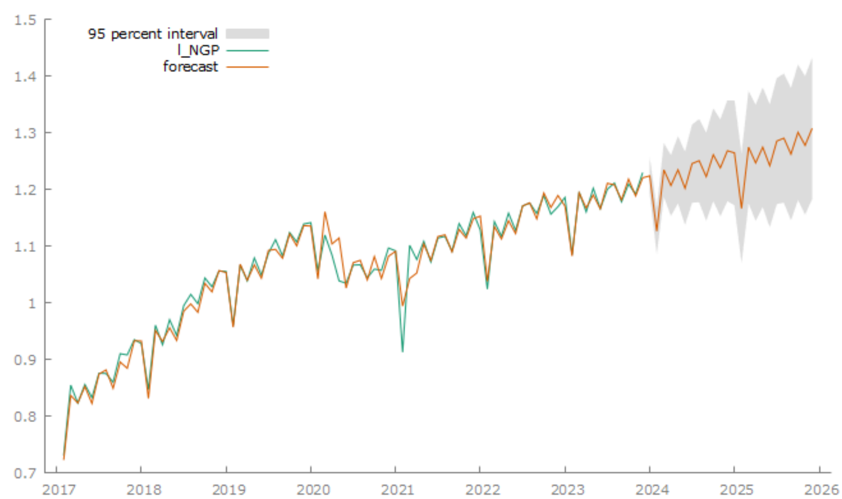
An script was used for further explore multiple orders in both the Autoregressive and moving average part, the top performer in terms of BIC was the model **AR (1)** but as seen before showed autocorrelation, the second one was **AR (2)** which further improved the model having only weak autocorrelation at lag 9, the modulus for the **AR 1** was of **1.0643** while being above **1** still was close, so a unit root test(**ADF**) was performed to check for stationarity which confirmed the non-stationarity, according to this an ARIMA model was considered which the best performant being an **ARIMA(1,1,0)** with a **r2=0.9924 BIC=-839.95**, thanks to the differencing the modulus of the AR part had a value of **3.7431**, while the model still exhibit a weak autocorrelation at lag 9 it was accepted as in order to remove it the model would need a high order parameter.



3.

The model was retrained with data from January 2010 to December 2021 and the 24 last months were forecasted and the error calculated **ME=-0.016044** , **MAE= 0.016390**, **RMSE= 0.018679**, both ME and MAE are in the same order of magnitude and as ME is negative indicate we're systematically underestimating, this would need further investigation to determine the bias source.

4.



As the variable is non-stationary and we are using an ARIMA mode, the confidence interval (shaded area) tends to increase with time.

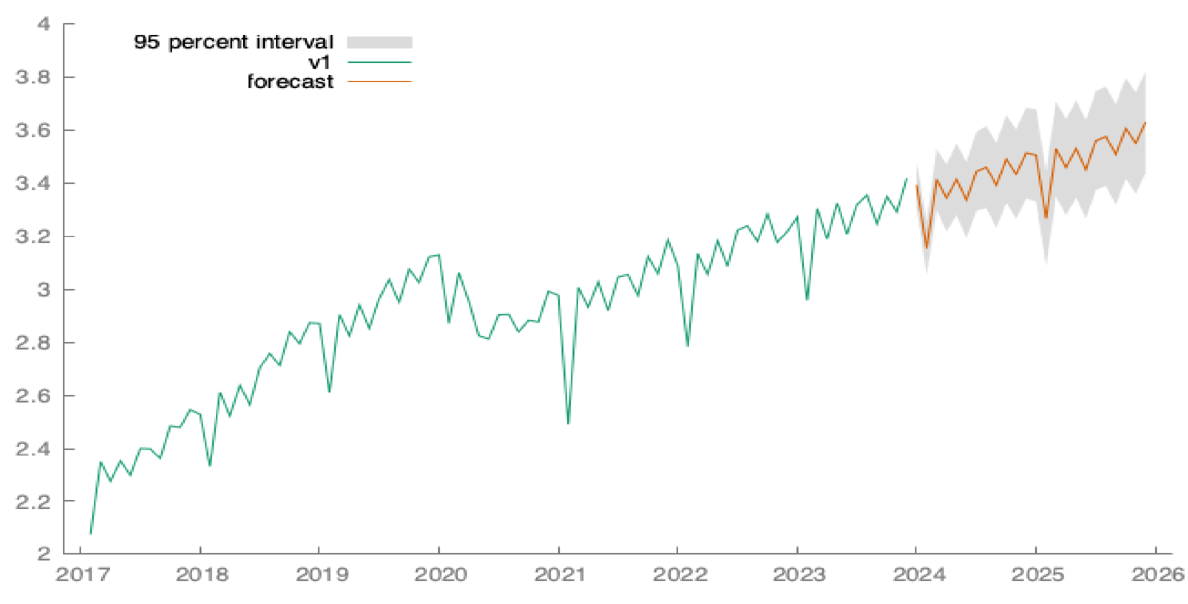
5.

Finally, we employed the VAR model to examine and predict three additional variables within the identical timeframe, spanning from January 2010 to December 2023: the consumption of natural gas, the quantity imported, and the quantity exported.

Initially, we began by selecting the appropriate VAR lag for the exogenous variables, including constant and trend. It's evident that different criteria yield varying lag selections, yet we chose the LAG with the highest BIC values, namely VAR(2), subsequently examining and preserving the residuals (See Appendix 1.3). Upon plotting the correlogram of these residuals, we observed correlation. Additionally, through trial and error, we determined VAR (2) is still the optimal choice, demonstrating superior values compared to adjustments in lag numbers, inclusive of constant, trend, and seasonal dummies. However, upon scrutinizing the correlograms of all residuals with and without seasonality, we encountered consistent ACF and PACF plots.

This coherent progression aligns with the anticipated correlation among the diverse variables, as evidenced in the correlogram (See Appendix 1.4). Attempting to eliminate this correlation seems futile. Consequently, we can assert the presence of a relationship among the production, consumption, import, and export of natural gas in the US.

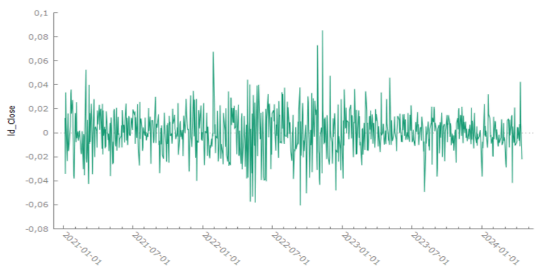
Finally for the forecast of VAR(2) we get this result:



When we compare our ARIMA(1,1,0) that we implemented earlier we see that VAR forecasting is beneficial when there are multiple variables interacting with each other. They can capture the dynamic relationships between variables and often perform well in capturing short-term dependencies. ARIMA model is useful more for a single variable and can better capture long-term trends and seasonality. However they might not perform as well when there are complex interactions between multiple variables.

Problem 2

The selected series is the Apple stock price from January 1st, 2021, to April 17th, 2024. After loading the data, we applied a log difference transformation, plotted the outcome and generated descriptive statistics:

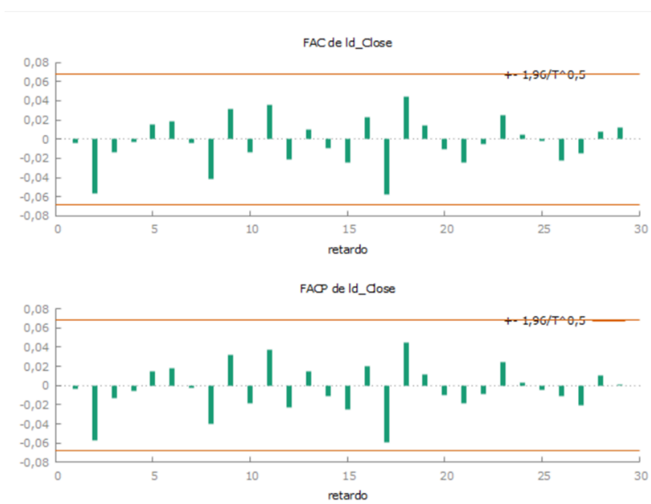


Estadísticos principales, usando las observaciones 2021-01-04 - 2024-04-15
para la variable 'ld_Close' (824 observaciones válidas)

Media	0,00035014
Mediana	0,00050987
Mínimo	-0,060472
Máximo	0,085236
Desviación típica	0,017201
C.V.	49,127
Asimetría	0,020715
Exc. de curtosis	1,6357
Percentil del 5%	-0,030103
Percentil del 95%	0,027548
Rango intercuartílico	0,019055
Observaciones ausentes	1

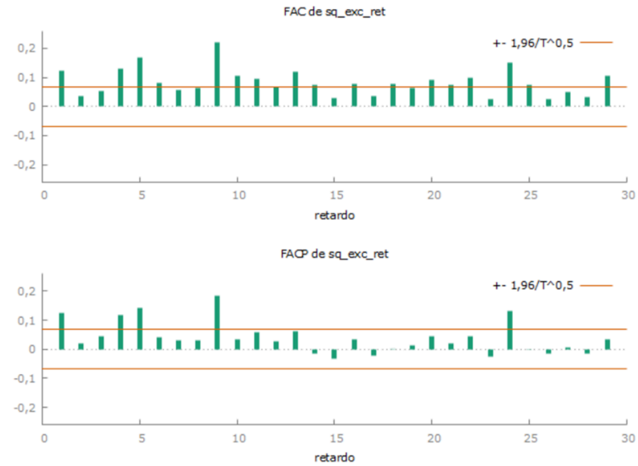
It is possible to observe that the mean is close to zero, that the standard deviation is 1.7%, and that the coefficient of variation is high, 49.12.

Proceed to analyze the correlogram:



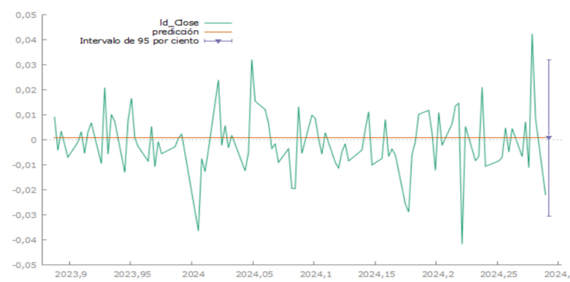
It is not possible to identify autocorrelations. Therefore, it is possible to assume that the time series is uncorrelated. For further analysis, we calculate the squared excess return:

$$exc\_return = ln\_close - mean(ln\_close)$$



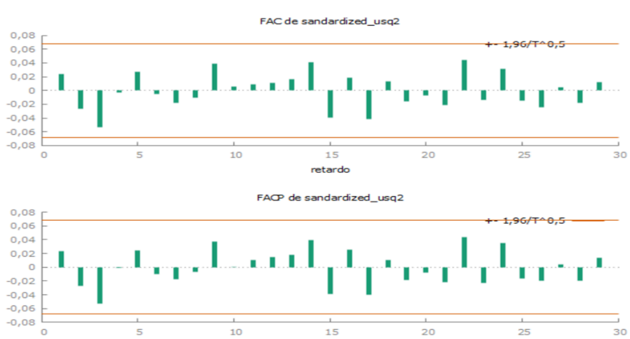
It seems that volatility may have a level of predictability. By applying the GARCH(1,1) model, we get the following statistics and predict one step in the future:

```
Evaluaciones de la función: 96
Evaluaciones del gradiente: 22

Modelo 1: GARCH, usando las observaciones 2021-01-05:2024-04-15 (T = 8:
Variable dependiente: ld_Close
Desviaciones típicas basadas en el Hessiano

             coeficiente   Desv. típica     z      valor p
  --------------------------------------------------------
  const      0,000770034   0,000541065     1,423   0,1547

  alpha(0)   3,95069e-06   1,93289e-06     2,044   0,0410   **
  alpha(1)   0,0425619     0,0113710       3,743   0,0002   ***
  beta(1)    0,943095      0,0154128      61,19    0,0000   ***

Media de la vble. dep.  0,000350   D.T. de la vble. dep.   0,017201
Log-verosimilitud       2222,838   Criterio de Akaike     -4435,676
Criterio de Schwarz    -4412,105   Crit. de Hannan-Quinn  -4426,634

Varianza incondicional del error = 0,000275439
Contraste de razón de verosimilitudes para los términos (G)ARCH:
  Chi-cuadrado(2) = 87,6407 [9,31248e-020]
```



Ro (Volatility) = 0.0159
Unconditional standard error = 0.016596355
Volatility is close to the unconditional standard error

Finally, we review the model by predicting the error variance, computing the standardized squared error residuals, and plotting its correlogram:



Where we can see no autocorrelation.

Appendix

Appendix 1.3

```
VAR system, maximum lag order 8

The asterisks below indicate the best (that is, minimized) values
of the respective information criteria, AIC = Akaike criterion,
BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags         loglik    p(LR)        AIC          BIC          HQC

   1     1426.25130            -17.528141   -17.066865   -17.340833
   2     1508.08343  0.00000   -18.351043   -17.582249*  -18.038862
   3     1529.81862  0.00024   -18.422733   -17.346422   -17.985680
   4     1543.26059  0.04279   -18.390757   -17.006929   -17.828833
   5     1563.04003  0.00090   -18.438000   -16.746655   -17.751204
   6     1604.65881  0.00000   -18.758235   -16.759372   -17.946566
   7     1641.90088  0.00000   -19.023761   -16.717381   -18.087220
   8     1679.86877  0.00000   -19.298360*  -16.684462   -18.236946*
```

Appendix 1.4