INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY
CAMPUS MONTERREY

INTRODUCTION TO ECONOMETRICS
CD2010.570

EVIDENCE 2 - TIME SERIES MODEL REPORT
Teamwork

TEACHER:
PhD. Erika Peralta Baez.

STUDENTS:
Víctor Benito García Rocha - A01232580.
Bruno Chavez Meza- A01737301.
Alma Ivone Santiago Hernández - A00836636.

July 22rd, 2024.

TABLE OF CONTENTS

EVIDENCE 2 - TIME SERIES MODEL REPORT

CONTEXT

The Ministry of Economy aims to estimate Mexican manufacturing exports based on the peso-dollar exchange rate. For this, it requires using a precise model that passes the tests of normality in the residuals, non stationarity, and serial autocorrelation. Please determine the existence of a cointegration relationship. If it exists, apply a VECM model, if it does not exist, apply a VAR model.

STEPS TO ACHIEVE NORMALITY IN THE RESIDUALS, STATIONARITY, AND NO SERIAL AUTOCORRELATION

1) Libraries

The first step to achieve normality in the residuals, stationarity, and no serial autocorrelation is to load the necessary libraries. We use libraries such as pandas for data manipulation, numpy for numerical calculations, matplotlib and seaborn for data visualization, and statsmodels and scipy for conducting statistical tests and model estimations. These libraries provide the essential tools to perform exploratory data analysis, stationarity and cointegration tests, and to evaluate the residuals of the fitted models.

2) Data exploratory analysis

After importing the data (with an encoding of "latin-1" and ";" as its delimiter), we make sure that the dates were in the proper format but also set them as our index, since it's the way we can apply time series analysis.

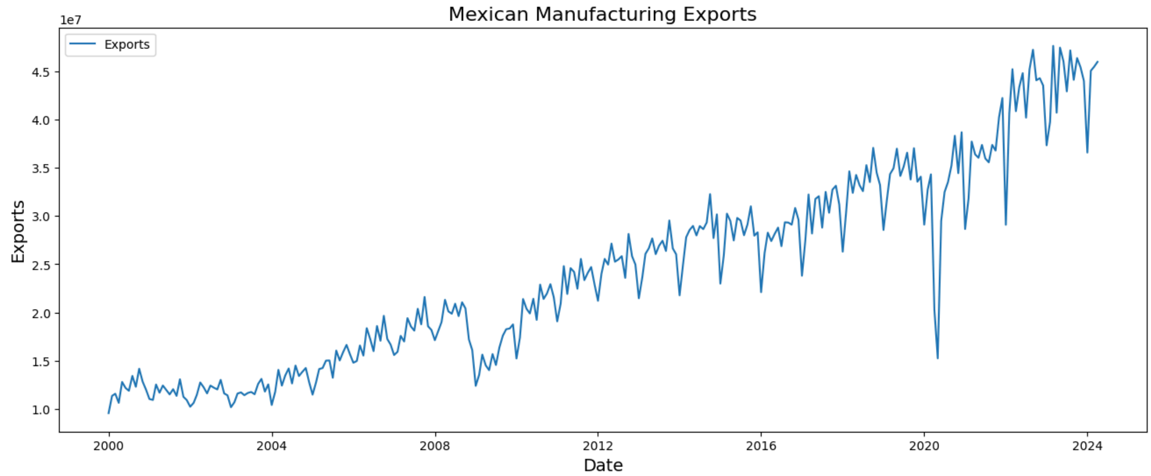
We confirm that there are no null or duplicate values. The dataset has a total of 292 rows of monthly historical data of exports (exp_td) and the MXN-USD exchange rate.

3) General data visualization

Subsequently, we conducted an exploratory data analysis (EDA) to visualize and better understand the time series of Mexican manufacturing exports and the peso-dollar exchange rate. To begin, the time series of exports and the exchange rate were plotted.

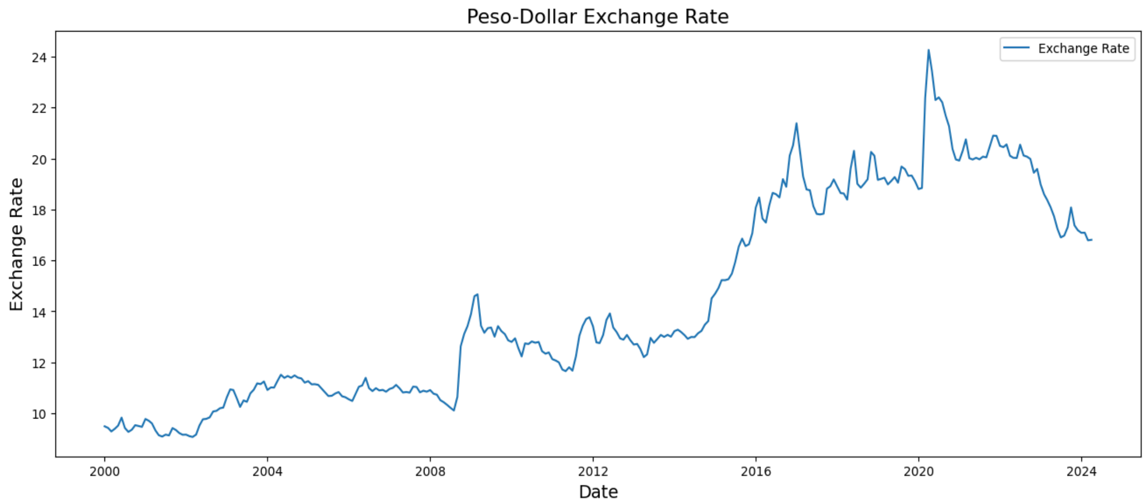
a) Mexican Manufacturing Exports:

The series shows a clear upward trend with seasonal patterns. There is a significant drop around the year 2020, likely due to the COVID-19 pandemic. This event is relevant but is not considered an outlier that should be removed.



b) Peso-Dollar Exchange Rate:

This series also shows an overall upward trend, but with more abrupt changes and less clear seasonality. A sharp peak is observed around the year 2020, again related to the pandemic.

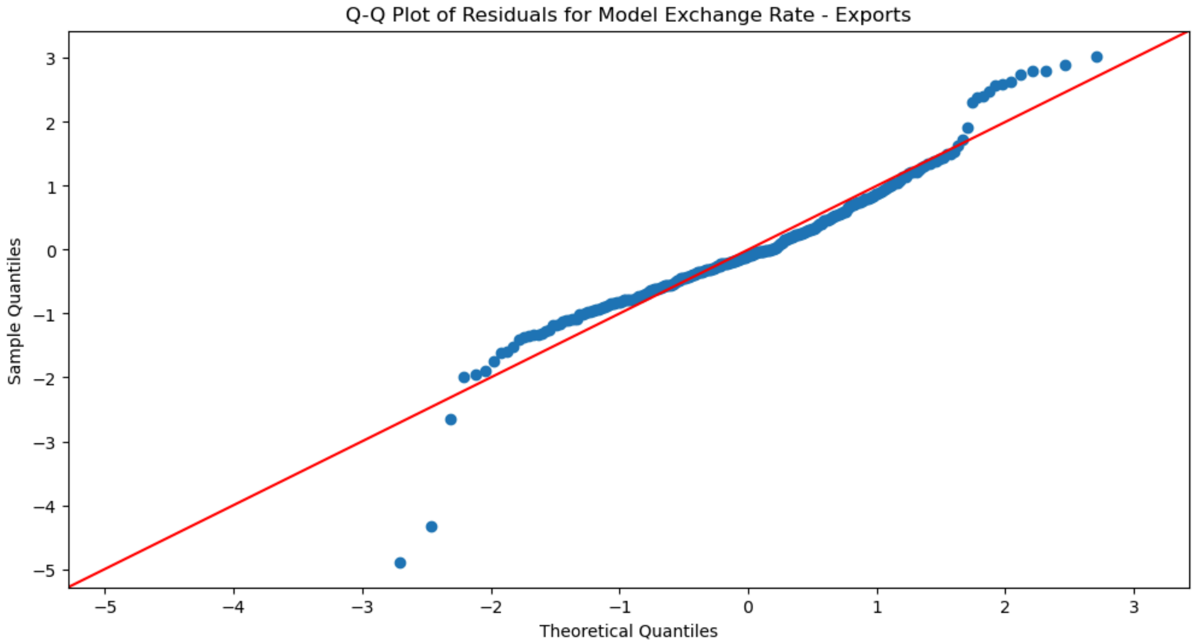


This exploratory data analysis provides us with an initial understanding of the characteristics and behaviors of the time series, which is essential for the selection and adjustment of forecasting models.

4) Normality test

We start using statsmodels's function to add “exchange” our constant, and obtain the residuals of an OLS regression of it with “exp_td”. Now with team we apply a Shapiro-Wilk test on them and obtain **statistic of 0.9451** with a **p-value of 0.0000000057**, which is way less than 0.05, so we reject the null hypothesis that the residuals are normally distributed (so, **<u>we don’t have normally distributed residuals</u>**).

By displaying the Q-Q plots, we can see how the residuals don’t follow the theoretical expected quantiles.

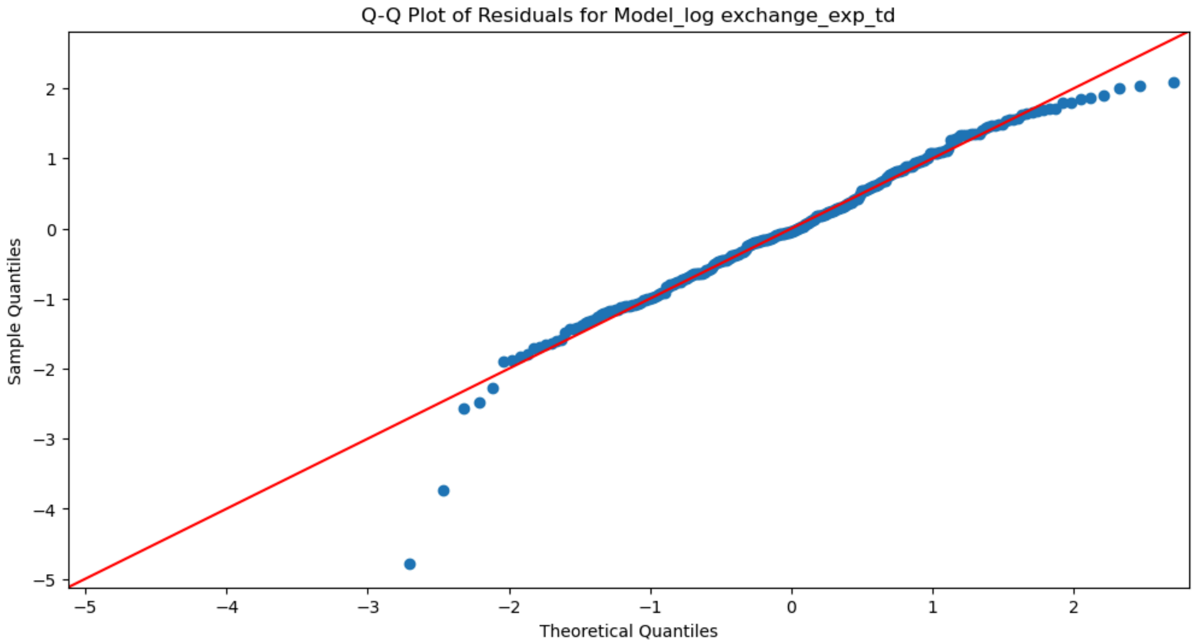


To address this, we can apply 4 things: A logarithmic transformation, a squared root, the Box-Cox, or removing seasonal effects in order to apply robust models such as VAR or VEC.

a) Logarithmic transformation

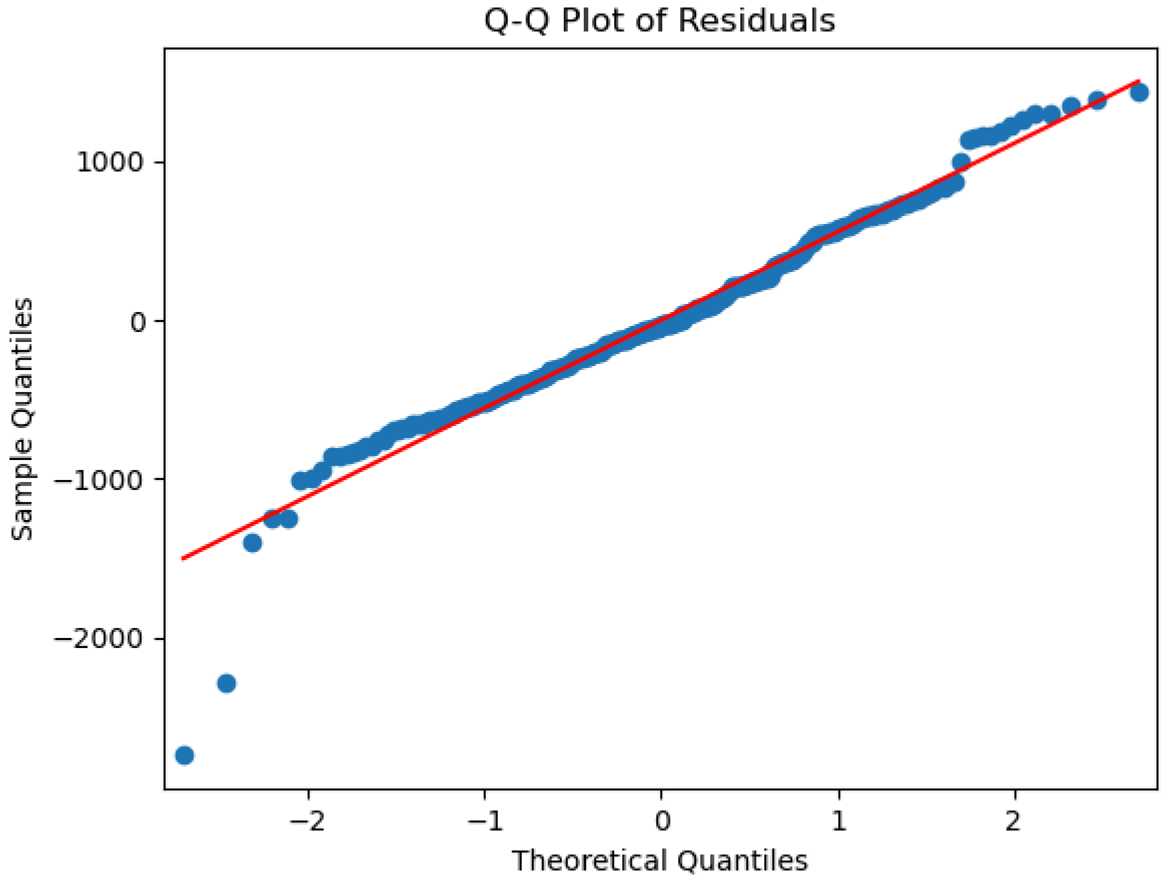
Although we could imply other factors that we will see in other points, part of the decision to use the logs here is because of the big difference in the magnitudes between the MXN-USD exchange rate and the value of the exports. After that we save the new OLS log residuals and the Shapiro-Wilk test gives us a **p-value of 0.000091**, still less than 0.05, but a better result.

Again, we can see it on the Q-Q plot right here:

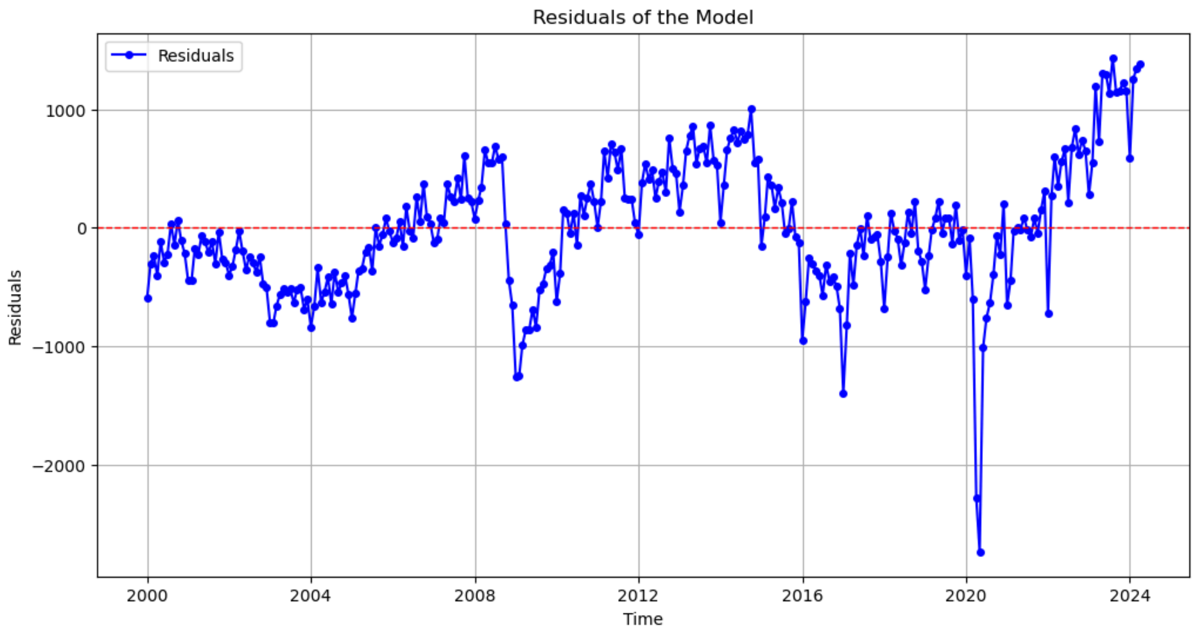


b) Square root transformation

After applying a square root transformation to the original values, get the residuals of a new OLS model and display it's Shapiro-Wilk p-value result, we still encounter less than 0.05 (this time **0.0000079**, one 0 worse than the log one), an additional Jarque-Bera test also confirmed they would not be normally distributed with a drastically low p-value (e-15). Notice from the Q-Q plot that the sample quantiles scale increased.



As expected, we can also identify that a big part of this happens due to the COVID-19 crisis, which specially affected variables such as the ones we analyze, so it's not just a matter of seasonality.



The residuals plot displays the differences between the observed values and the predicted values by the model over time. Ideally, the residuals should be randomly distributed around the zero line, indicating a good fit of the model. However, noticeable patterns and trends suggest that the model may be missing key factors. Significant deviations occur during periods like the 2008 financial crisis and the 2020 pandemic, indicating that the model did not adequately predict these abrupt changes. The increasing variability of residuals over time also suggests possible heteroscedasticity or instability in the model. The current model has limitations in fully capturing the data dynamics and may require additional variables or a more complex model for better accuracy.

c) Regression model of squared root-transformed data

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            sqrt_exp_td   R-squared:                       0.708
Model:                            OLS   Adj. R-squared:                  0.707
Method:                 Least Squares   F-statistic:                     704.7
Date:                Tue, 23 Jul 2024   Prob (F-statistic):           1.33e-79
Time:                        17:31:47   Log-Likelihood:                -2259.4
No. Observations:                 292   AIC:                             4523.
Df Residuals:                     290   BIC:                             4530.
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const        -1473.2585    239.262     -6.158      0.000   -1944.169   -1002.348
sqrt_exchange  1676.0094    63.133     26.547      0.000    1551.752    1800.267
==============================================================================
Omnibus:                       24.483   Durbin-Watson:                   0.293
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               67.125
Skew:                          -0.325   Prob(JB):                     2.66e-15
Kurtosis:                       5.257   Cond. No.                         29.7
==============================================================================
```

The OLS regression results for the model indicate a strong relationship between the square root of the exchange rate (sqrt_exchange) and the square root of exports (sqrt_exp_td), with an R-squared value of 0.708, meaning approximately 70.8% of the variability in sqrt_exp_td can be explained by sqrt_exchange. The coefficient for sqrt_exchange is 1676.0094, which is statistically significant with a t-value of 26.547 and a p-value of 0.000, indicating a strong positive relationship. The constant term (intercept) is -1473.2585, also statistically significant with a t-value of -6.158 and a p-value of 0.000. The F-statistic of 704.7 and its p-value of 1.33e-77 further support the overall significance of the model. The Durbin-Watson statistic is 0.293, suggesting possible positive autocorrelation in the residuals. The Omnibus and Jarque-Bera tests indicate non-normality in the residuals, with p-values of 0.000 and 2.66e-15, respectively. The skewness is -0.325, and the kurtosis is 5.257, which further supports the presence of non-normality in the residuals.

d) Box-Cox transformation

The Box-Cox, which is like an intelligent combination of the last two based on a λ parameter, gave us a Shapiro-Wilk's p-value of the residuals of **0.00028**, which shows how better it actually is to improve residual normality, but still is not good enough to achieve it, as the Jarque-Bera test also confirmed with a p-value of 0.0000000027. We reject the null hypothesis that the residuals are normally distributed.

5) Stationarity and Autocorrelation

a) Application of the Augmented Dickey-Fuller Test (ADF)

At this step, an Augmented Dickey-Fuller (ADF) test was conducted to assess the stationarity of the time series for Mexican manufacturing exports and the peso-dollar exchange rate. Stationarity is an important property for time series models, as it implies that the statistical properties of the process do not change over time.

```
Augmented Dickey-Fuller Test: Exports
ADF test statistic          0.794083
p-value                     0.991540
# lags used                15.000000
# observations            276.000000
critical value (1%)        -3.454267
critical value (5%)        -2.872070
critical value (10%)       -2.572381
```

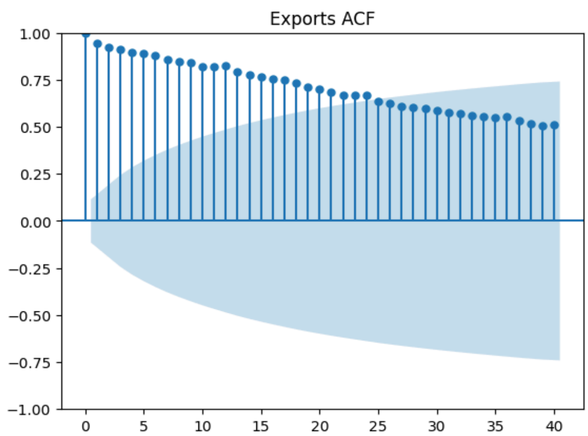
Mexican Manufacturing Exports:
The p-value obtained from the ADF test for Mexican manufacturing exports is 0.991384, which is greater than 0.05. This indicates that we cannot reject the null hypothesis that the time series has a unit root, meaning the series is non-stationary. Additionally, the ADF test statistic is 0.793498, and compared to the critical values (-3.454262, -2.872878, -2.572381), it is higher, reinforcing that the series is non-stationary as it does not fall into the rejection regions of the null hypothesis.

```
Augmented Dickey-Fuller Test: Exchange Rate
ADF test statistic         -1.302771
p-value                     0.627904
# lags used                 2.000000
# observations            289.000000
critical value (1%)        -3.453182
critical value (5%)        -2.871593
critical value (10%)       -2.572127
```

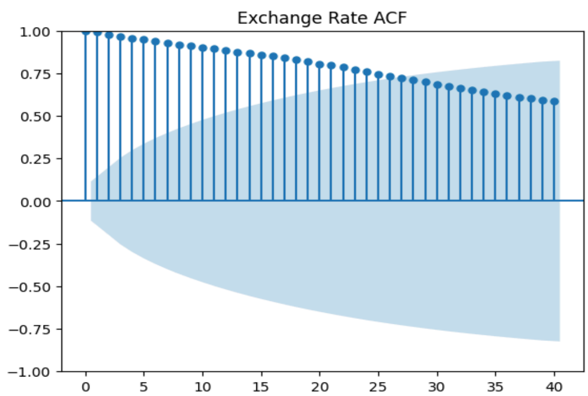
Peso-Dollar Exchange Rate:
The p-value obtained from the ADF test for the peso-dollar exchange rate is 0.627904, which is greater than 0.05. This indicates that we cannot reject the null hypothesis that the time series has a unit root, meaning the series is non-stationary. The ADF test statistic is -1.302771, and compared to the critical values (-3.453182, -2.871593, -2.572127), it is also higher, reinforcing that the series is non-stationary as it does not fall into the rejection regions of the null hypothesis.

6) Check for autocorrelation

Subsequently, autocorrelation analyses were performed for the time series of Mexican manufacturing exports and the peso-dollar exchange rate. Autocorrelation measures the correlation of a time series with its own lags, and the Autocorrelation Function (ACF) allows us to visualize these correlations.



Mexican Manufacturing Exports: The ACF plot for Mexican manufacturing exports shows peaks outside the confidence bands for lags 0 to 24, indicating strong autocorrelation in these lags. From lag 25 onwards, the lines are within the confidence bands, indicating a decrease in autocorrelation. This suggests that the export values are significantly influenced by their previous values in the first 24 lags, indicating that the series is non-stationary.



Peso-Dollar Exchange Rate: The ACF plot for the peso-dollar exchange rate shows peaks outside the confidence bands for lags 0 to 27, indicating strong autocorrelation in these lags. From lag 28 onwards, the lines are within the confidence bands, indicating a decrease in autocorrelation. This suggests that the exchange rate values are strongly influenced by their previous values in the first 27 lags, indicating that the series is non-stationary.

Both series show strong autocorrelation in the initial lags that decays slowly, indicating non-stationarity. This is consistent with the results from the ADF test, which also suggested that the series have a unit root.

7) Check for needed differences to apply an GLS model in time series

date	exp_td	exchange
2000-01-01	9566330	9.4878
2000-02-01	11358821	9.4252
2000-03-01	11581780	9.2876
2000-04-01	10632048	9.3903
2000-05-01	12800667	9.5158
...	...	...
2023-12-01	43993414	17.1860
2024-01-01	36571021	17.0873
2024-02-01	45047576	17.0898

Subsequently, a function called find_differencing was defined to determine how many differences are needed to make the time series of Mexican manufacturing exports and the peso-dollar exchange rate stationary. The function performs differencing on the time series until the Augmented Dickey-Fuller (ADF) test indicates the series is stationary or a maximum of five differences is reached. If the series is too short to continue differencing, the function prints a message and returns the number of differences performed, showing the number of differences applied and the ADF test p-value at each step.

Additionally, a copy of the original DataFrame df was made, and columns that would not be used in the analysis were dropped, keeping only the columns of interest: exp_td and exchange. This was done to prepare the data before applying a Generalized Least Squares (GLS) time series model with the differenced variables.

8) Check the number of differences required

```
Differencing 1 for the current column: ADF p-value = 3.859383374989132e-06
Augmented Dickey-Fuller Test:
ADF test statistic       -5.372966
p-value                   0.000004
# lags used              14.000000
# observations          276.000000
critical value (1%)      -3.454267
critical value (5%)      -2.872070
critical value (10%)     -2.572381
Differencing 2 for the current column: ADF p-value = 1.02344077666341e-15
Augmented Dickey-Fuller Test:
ADF test statistic     -9.315226e+00
p-value                 1.023441e-15
# lags used             1.500000e+01
# observations          2.740000e+02
critical value (1%)    -3.454444e+00
critical value (5%)    -2.872147e+00
critical value (10%)   -2.572422e+00
Differencing 3 for the current column: ADF p-value = 4.750827649966153e-21
Augmented Dickey-Fuller Test:
ADF test statistic     -1.149044e+01
p-value                 4.750828e-21
# lags used             1.600000e+01
# observations          2.720000e+02
critical value (1%)    -3.454622e+00
critical value (5%)    -2.872225e+00
critical value (10%)   -2.572464e+00
Differencing 4 for the current column: ADF p-value = 4.1716116874596985e-23
Augmented Dickey-Fuller Test:
ADF test statistic     -1.241931e+01
p-value                 4.171612e-23
# lags used             1.600000e+01
# observations          2.710000e+02
critical value (1%)    -3.454713e+00
critical value (5%)    -2.872265e+00
critical value (10%)   -2.572485e+00
Differencing 5 for the current column: ADF p-value = 8.091822398879372e-23
Augmented Dickey-Fuller Test:
ADF test statistic     -1.228428e+01
p-value                 8.091822e-23
# lags used             1.600000e+01
# observations          2.700000e+02
critical value (1%)    -3.454804e+00
critical value (5%)    -2.872305e+00
critical value (10%)   -2.572506e+00
```

```
Differencing 1 for the current column: ADF p-value = 3.1163886997637402e-24
Augmented Dickey-Fuller Test:
ADF test statistic     -1.296853e+01
p-value                 3.116389e-24
# lags used             1.000000e+00
# observations          2.890000e+02
critical value (1%)    -3.453182e+00
critical value (5%)    -2.871593e+00
critical value (10%)   -2.572127e+00
Differencing 2 for the current column: ADF p-value = 3.2080672296591947e-12
Augmented Dickey-Fuller Test:
ADF test statistic     -7.946574e+00
p-value                 3.208067e-12
# lags used             1.400000e+01
# observations          2.750000e+02
critical value (1%)    -3.454355e+00
critical value (5%)    -2.872108e+00
critical value (10%)   -2.572401e+00
Differencing 3 for the current column: ADF p-value = 8.221324449555152e-14
Augmented Dickey-Fuller Test:
ADF test statistic     -8.570526e+00
p-value                 8.221324e-14
# lags used             1.600000e+01
# observations          2.720000e+02
critical value (1%)    -3.454622e+00
critical value (5%)    -2.872225e+00
critical value (10%)   -2.572464e+00
Differencing 4 for the current column: ADF p-value = 2.1338308029547186e-18
Augmented Dickey-Fuller Test:
ADF test statistic     -1.038144e+01
p-value                 2.133831e-18
# lags used             1.500000e+01
# observations          2.720000e+02
critical value (1%)    -3.454622e+00
critical value (5%)    -2.872225e+00
critical value (10%)   -2.572464e+00
Differencing 5 for the current column: ADF p-value = 5.199358855246007e-25
Augmented Dickey-Fuller Test:
ADF test statistic     -1.337095e+01
p-value                 5.199359e-25
# lags used             1.500000e+01
# observations          2.710000e+02
critical value (1%)    -3.454713e+00
critical value (5%)    -2.872265e+00
critical value (10%)   -2.572485e+00
```

Subsequently, the number of differences needed to make the time series of Mexican manufacturing exports and the peso-dollar exchange rate stationary was calculated, and the results of the ADF test after each differencing were obtained.

Mexican Manufacturing Exports:
For Mexican manufacturing exports, successive differencing was applied to the time series, and the ADF test was run after each differencing. The results showed that one difference was sufficient to make the series stationary, with a **p-value of 3.859383374989132e-06** after the first differencing, which is less than 0.05, suggesting that the series is stationary. The p-values for additional differencing were even lower: 1.02344077666341e-15, 4.750827649966153e-21, 4.1716116874596985e-23, and 8.091822398879372e-23, confirming stationarity after the first differencing.

Peso-Dollar Exchange Rate:
For the peso-dollar exchange rate, a similar process of differencing and ADF testing was followed. The results indicated that one difference was sufficient to make the series stationary, with a **p-value of 3.1163886997637402e-24** after the first differencing, significantly less than 0.05,

confirming stationarity. The p-values for additional differencing were even lower: 3.2080672296591947e-12, 8.22132449555152e-14, 2.133880829547186e-18, and 5.199358855246007e-25, reaffirming that the series became stationary with a single differencing.

9) Apply differentiation and prepare data for GLS modeling

In these steps, differentiation was applied to the selected columns to prepare the data for Generalized Least Squares (GLS) time series modeling, and the differentiated columns were combined.

```
df_diff = df2.copy()
for col, n_diffs in columns_to_diff.items():
    for _ in range(n_diffs):
        df_diff[col] = df_diff[col].diff()
df_diff = df_diff.dropna()
```

Apply Differentiation

A dictionary columns_to_diff was defined indicating the number of differences needed for each column ('exp_td' and 'exchange'), with 1 difference for both. A copy of the DataFrame df2 was created, named df_diff. For each column in columns_to_diff, the specified differences were applied using a loop. After applying the differentiation, rows with NaN values resulting from the process were dropped, leaving a DataFrame df_diff with differentiated and stationary time series.

```
if columns_no_diff:
    df_non_diff = df[columns_no_diff].iloc[max(columns_to_diff.values()):]
    # Combine differentiated and non-differentiated columns
    df_combined = pd.concat([df_diff[columns_to_diff.keys()], df_non_diff], axis=1).dropna()
else:
    df_combined = df_diff[columns_to_diff.keys()]
```

Align and Combine Columns

It was checked if there were any columns that did not require differentiation, but in this case columns_no_diff was empty, so no action was taken for non-differentiated columns. The differentiated columns were combined into a new DataFrame df_combined. If there had been non-differentiated columns, they would have been aligned and combined with the differentiated ones. Finally, the first rows of the combined DataFrame were printed to verify the results, confirming that the columns exp_td and exchange were correctly differentiated.

```
Combined Data (Differenced where needed):
               exp_td  exchange
date
2000-02-01  1792491.0   -0.0626
2000-03-01   222959.0   -0.1376
2000-04-01  -949732.0    0.1027
2000-05-01  2168619.0    0.1255
2000-06-01  -621261.0    0.3137
```

The first rows of the combined DataFrame show the differentiated data, confirming that the columns exp_td and exchange were correctly differentiated, making the series stationary and ready for GLS modeling.

10) Train-Test Split

```
Train set:
               exp_td  exchange
date
2022-12-01  -759524.0    0.1481
2023-01-01 -6199273.0   -0.6067
2023-02-01  2437184.0   -0.3877
2023-03-01  7870938.0   -0.2237
2023-04-01 -6925383.0   -0.2894
```

In this step, the differenced data was split into training and testing sets. It was established that the last 12 observations would be used for the testing set by defining the variable test_obs with a value of 12. The DataFrame df_combined was then divided into two subsets: the training set (train) containing all observations except the last 12, and the testing set (test) containing the last 12 observations.

```
Test set:
              exp_td  exchange
date
2023-05-01  6742359.0   -0.3482
2023-06-01 -1421434.0   -0.4961
2023-07-01 -3114320.0   -0.3363
2023-08-01  4249812.0    0.0717
2023-09-01 -3051820.0    0.3311
```

To verify that the split was done correctly, the last rows of the training set and the first rows of the testing set were printed. The training set showed observations from December 2022 to April 2023. While the testing set showed observations from May 2023 to September 2023. This verification confirms that the data was correctly separated and that the model can be trained on a broad dataset and tested on the most recent observations.

11) GLS Model

```
X_train = train['exchange']
y_train = train['exp_td']
```

- Adding a constant term for the intercept in the model

```
[249] X_train = sm.add_constant(X_train)
```

- Fit the OLS model on the training set

```
[250] gls_model = sm.GLS(y_train, X_train).fit()
```

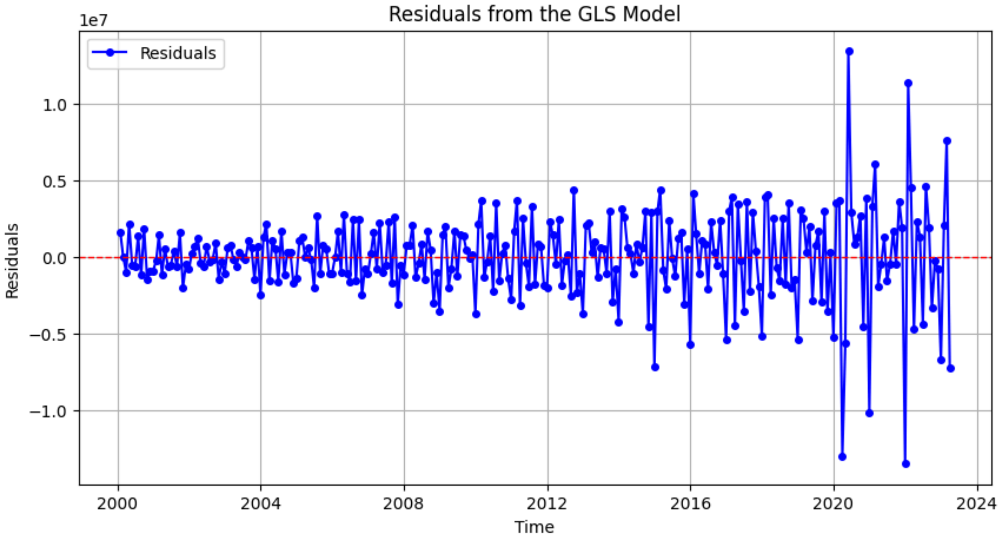
In this step, the necessary variables for the GLS model were defined. X_train was assigned to the exchange rate (exchange) column of the training set, while y_train was set with the manufacturing exports (exp_td) column. To include an intercept term in the model, a constant was added to X_train, allowing the model to fit a constant term that can improve its accuracy by considering a base value independent of the exchange rate. Then, the GLS (Generalized Least Squares) model was fitted on the training set, suitable for differenced time series as it handles autocorrelation in the residuals. Using X_train and y_train, the fit() method was executed to fit the model and obtain the parameters that best explain the relationship between the exchange rate and manufacturing exports.

```
                            GLS Regression Results
==============================================================================
Dep. Variable:                 exp_td   R-squared:                       0.009
Model:                            GLS   Adj. R-squared:                  0.006
Method:                 Least Squares   F-statistic:                     2.605
Date:                Tue, 23 Jul 2024   Prob (F-statistic):              0.108
Time:                        17:33:46   Log-Likelihood:                -4542.2
No. Observations:                 279   AIC:                             9088.
Df Residuals:                     277   BIC:                             9096.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.307e+05   1.71e+05      0.762      0.446   -2.07e+05    4.68e+05
exchange   -6.193e+05   3.84e+05     -1.614      0.108   -1.37e+06    1.36e+05
==============================================================================
Omnibus:                       47.576   Durbin-Watson:                   2.612
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              325.203
Skew:                          -0.414   Prob(JB):                     2.42e-71
Kurtosis:                       8.224   Cond. No.                         2.25
==============================================================================
```

Model Summary:
Finally, the summary of the fitted model was printed to analyze its results. This summary includes various statistical metrics that help evaluate the quality of the model. It was observed that the exchange rate has a negative effect on exports, as indicated by the negative coefficient (-6.193e+05) for the exchange variable, although this relationship is not statistically significant with a p-value of 0.108 (greater than 0.05). The R-squared value is very low (0.009), indicating that the model does not explain much of the variance in exports. Additionally, the p-value for the F-statistic (0.108) is greater than 0.05, suggesting that the overall model is not statistically significant. The residuals are not normally distributed, as indicated by the Jarque-Bera test (JB = 325.203, p-value = 2.42e-71). The Durbin-Watson

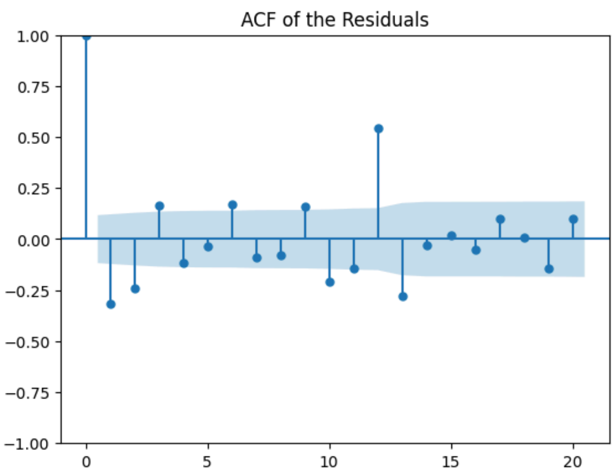
statistic (2.612) suggests possible autocorrelation in the residuals. Overall, these results indicate that the model is not ideal.

Plot of Residuals from the GLS Model:



This plot shows the residuals from the GLS model over time. Residuals are the differences between observed values and the values predicted by the model. The plot demonstrates how the residuals fluctuate around the red line, which represents a residual value of zero. Patterns in the residuals indicate that the model has not captured all the structure present in the data. Notably, significant variations around the year 2020 are related to exceptional events like the COVID-19 pandemic. This residual analysis is crucial for assessing the model's adequacy and understanding its limitations.

Plot of the ACF of the Residuals:



This plot shows the Autocorrelation Function (ACF) of the residuals from the GLS model. The ACF is an essential tool for identifying autocorrelation in the residuals, indicating that the values of the residuals are correlated at different points in time. In this plot, peaks that stand out from the confidence bands indicate significant autocorrelation. Although the autocorrelation decreases over time, it is still evident, suggesting that the model has not completely eliminated the temporal dependencies. Notably, at lag 15, the ACF drops to nearly zero for the first time, indicating a point where the residual autocorrelation is minimal. This residual autocorrelation implies that the model may need further refinement to fully capture the dynamics of the data.

NECESSARY TESTS SUCH AS COINTEGRATION JOHANSEN TEST

12) Stationarity

In this step, the stationarity of the time series for exports (exp_td) and exchange rate (exchange) was checked using the Augmented Dickey-Fuller (ADF) test. The ADF test is used to determine whether a time series is stationary.

```
[255] print('ADF test for Exports:', adf_exp_td)

ADF test for Exports: (0.794082909812392, 0.9915400522312456, 15, 276, {'1%': -3.4542672521624214, '5%': -2.87206958769775, '10%': -2.5723807881747534}, 8837.05502670797)
```

For Exchange Rate: p-value > 0.05, so the series is non-stationary.

```
[256] print('ADF test for Exchange Rate:', adf_exchange)

ADF test for Exchange Rate: (-1.302770796520348, 0.6279037194045802, 2, 289, {'1%': -3.4531816058796143, '5%': -2.871593390724642, '10%': -2.572126815651536}, 320.2178369
```

a) Results for Exports:

In the Augmented Dickey-Fuller (ADF) test conducted for the exports (exp_td) time series, the following results were obtained: an ADF statistic of 0.7941, a p-value of 0.9915, with 15 lags used and 276 observations. The critical values for the 1%, 5%, and 10% significance levels are -3.4543, -2.8721, and -2.5724, respectively. The obtained p-value (0.9915) is significantly greater than 0.05, meaning the null hypothesis that the series has a unit root (is non-stationary) cannot be rejected. This indicates that the export time series is non-stationary, i.e., it has a mean, variance, and autocorrelation that changes over time. The test also shows a test statistic value of 8837.0550, which further confirms the non-stationarity of the series.

b) Results for Exchange Rate:

In the ADF test conducted for the exchange rate (exchange) time series, the results were: an ADF statistic of -1.3028, a p-value of 0.6279, with 2 lags used and 289 observations. The critical values for the 1%, 5%, and 10% significance levels are -3.4532, -2.8716, and -2.5721, respectively. The obtained p-value (0.6279) is also greater than 0.05, meaning the null hypothesis that the series has a unit root (is non-stationary) cannot be rejected. This indicates that the exchange rate time series is also non-stationary. The test also shows a test statistic value of 320.2178, further confirming the non-stationarity of this time series. Non-stationarity in this series implies that its statistical properties change over time, which can impact future analysis and predictions based on this data.

13) Test for cointegration

a) Performing the Johansen Cointegration Test:

In this step, the Johansen cointegration test was conducted for the time series of exports (exp_td) and exchange rate (exchange). This test is used to determine if there is a long-term equilibrium relationship between the two time series, despite them being individually non-stationary. The test was configured with det_order=0, which assumes a non-deterministic trend without a constant, and k_ar_diff=1, which indicates one lagged difference.

```
[258] print(coint_test.lr1)  #

[19.62901952  1.2252661 ]
```

Trace Statistic: The trace statistic results were printed to assess the existence of cointegration. The obtained values were 19.6290 for the first cointegration vector and 1.2253 for the second cointegration vector. The trace statistic compares these values with critical values to determine if we can reject the null hypothesis of no cointegration.

```
[259] print(coint_test.cvt)  # Cri

[[13.4294 15.4943 19.9349]
 [ 2.7055  3.8415  6.6349]]
```

Critical Values: The critical values for the 90%, 95%, and 99% significance levels are provided by the test. For the first vector, the critical values are 13.4294, 15.4943, and 19.9349, respectively. For the second vector, the critical values are 2.7055, 3.8415, and 6.6349, respectively. These values are used for comparison with the trace statistics.

By comparing the trace statistics with the critical values, it is observed that 19.62 is greater than 15.49 at the 95% significance level, while 1.22 is less than 3.84 at the 95% significance level. This means that the null hypothesis is rejected for the first vector, indicating the presence of at least one cointegration vector. However, the null hypothesis is not rejected for the second vector. Therefore, we conclude that there is exactly one cointegration relationship between exports and the exchange rate. This conclusion allows us to proceed with the Vector Error Correction (VEC) model, which will model the long-term relationship and short-term adjustments between the series.

We accept the alternative hypothesis that there is at least one cointegration vector, indicating that there is a long-run equilibrium relationship between exports and the exchange rate. We conclude that there is exactly one cointegration relationship (r=1).

14) Vector Error Correction Model

In this step, the Vector Error Correction Model (VECM) was fitted to the time series of exports (exp_td) and exchange rate (exchange). The VECM is suitable for cointegrated time series as it captures both long-term equilibrium relationships and short-term dynamics between the variables. The model was configured with k_ar_diff=1 to include one lagged difference and coint_rank=1 to indicate the expectation of one cointegration vector. The model was fitted using the fit() method, and a summary of the results was printed.

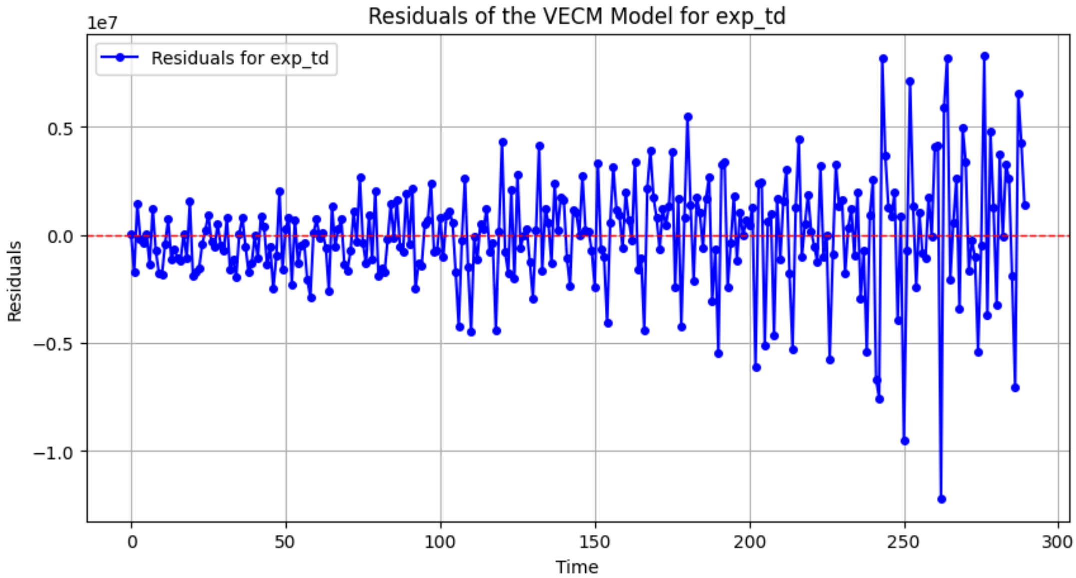
```
Det. terms outside the coint. relation & lagged endog. parameters for equation exp_td
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
L1.exp_td     -0.2977      0.053     -5.613      0.000      -0.402      -0.194
L1.exchange -2.148e+06  3.42e+05     -6.276      0.000   -2.82e+06   -1.48e+06
Det. terms outside the coint. relation & lagged endog. parameters for equation exchange
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
L1.exp_td   8.862e-09   8.85e-09      1.002      0.317   -8.48e-09    2.62e-08
L1.exchange    0.2659      0.057      4.657      0.000       0.154       0.378
               Loading coefficients (alpha) for equation exp_td
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ec1           -0.0981      0.025     -3.876      0.000      -0.148      -0.048
               Loading coefficients (alpha) for equation exchange
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ec1         4.493e-09   4.22e-09      1.064      0.287   -3.78e-09    1.28e-08
          Cointegration relations for loading-coefficients-column 1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
beta.1         1.0000          0          0      0.000       1.000       1.000
beta.2     -1.848e+06   9.91e+04    -18.658      0.000   -2.04e+06   -1.65e+06
==============================================================================
```

Coefficients of Intercept Terms and Lagged Endogenous Parameters: The model summary shows the coefficients for the intercept terms and lagged endogenous parameters. In the equation for exp_td, the coefficient for the first lag of exp_td is -0.2977 with a highly significant p-value (p < 0.000), and the coefficient for the first lag of exchange is -2.148e+06, also with a highly significant p-value (p < 0.000). This indicates that past exports and past exchange rates significantly impact current exports. In the equation for exchange, the coefficient for the first lag of exp_td is 8.862e-09 (not significant) and for the first lag of exchange is 0.2659 with a significant p-value (p < 0.000), indicating that the past exchange rate significantly influences the current exchange rate.

Loading Coefficients (Alpha): The loading coefficients (alpha) indicate the speed at which variables adjust to long-term equilibrium. In the equation for exp_td, the loading coefficient for the first cointegration vector (ec1) is -0.0981 with a highly significant p-value (p < 0.000). This suggests that exports adjust slowly to long-term imbalances. In the equation for exchange, the loading coefficient for ec1 is 4.493e-09 (not significant), suggesting that the exchange rate does not significantly adjust to long-term imbalances in this model.

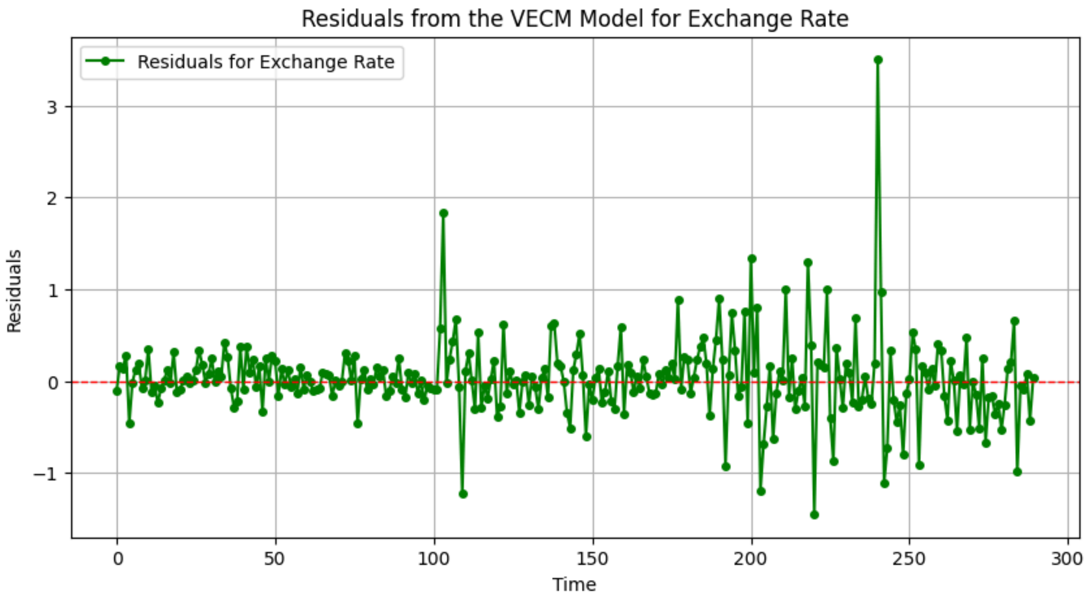
Cointegration Relationships: The cointegration relationships for the loading coefficients show the cointegration vector. The beta coefficient for exp_td is 1.0000 (normalized), and the beta coefficient for exchange is -1.848e+06 with a highly significant p-value (p < 0.000). This confirms that there is a cointegration relationship between exports and the exchange rate, with a long-term negative impact of the exchange rate on exports.

Residuals of the VECM Model for exp_td



The graph shows the residuals of the VECM model for exports (exp_td). The residuals oscillate around zero over time, which is desirable as it indicates that the model does not have a systematic bias. However, some significant spikes can be observed, which may indicate the presence of residual autocorrelation, suggesting that there may be factors not captured by the model. Despite these spikes, most of the residuals fall within an acceptable range, indicating that the VECM model is adequately capturing the dynamics of exports in relation to the exchange rate over the long term and adjusting for deviations in the short term.

Residuals of the VECM Model for the Exchange Rate



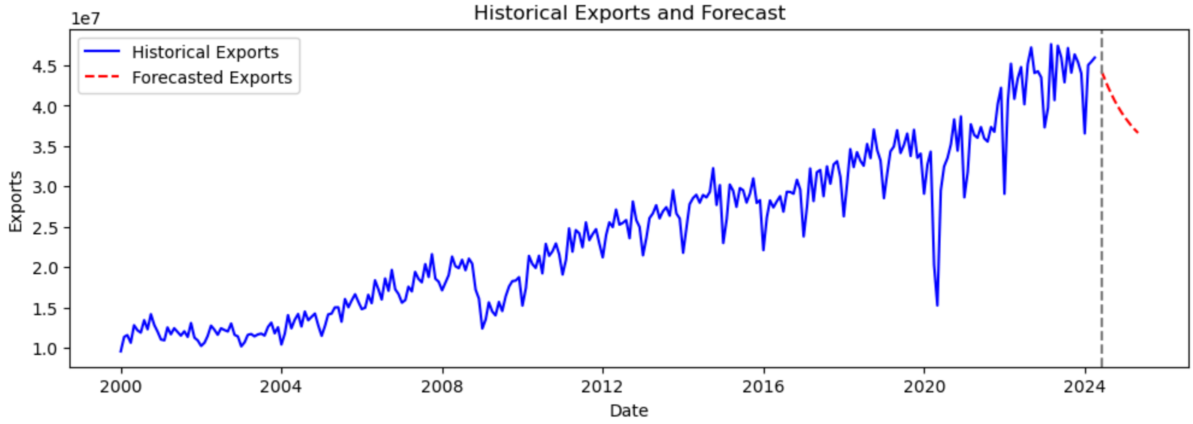
The graph presents the residuals of the VECM model for the exchange rate (exchange). The residuals also oscillate around zero, indicating that the model is not biased. However, some significant spikes are observed, particularly a very high one around the 250-point mark, which could be indicative of exceptional events or the presence of residual autocorrelation. This suggests that while the model captures the long-term relationship between the exchange rate and exports, there may be short-term factors not fully accounted for. Overall, the distribution of residuals around zero is a good sign that the model is reasonably adequate for forecasting the dynamics of the exchange rate.

FORECAST OF EXPORTS

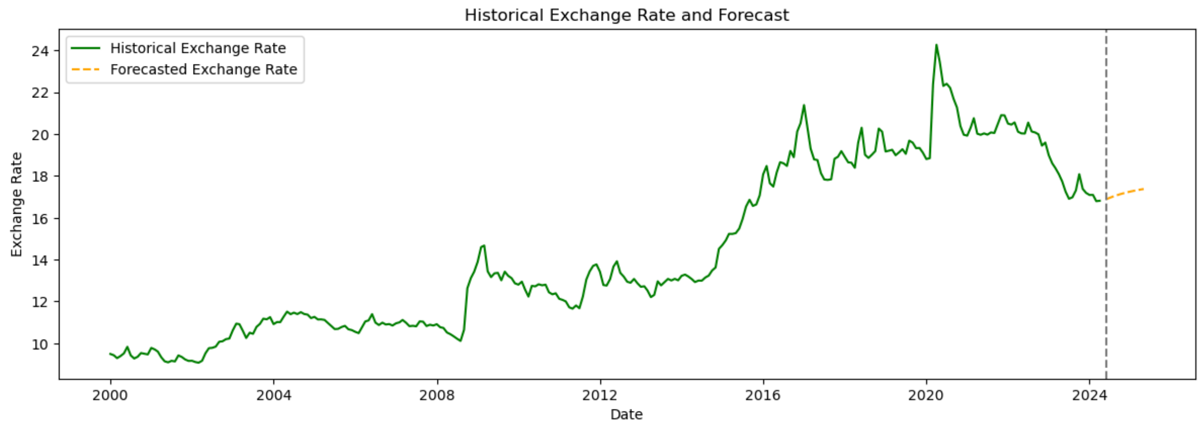
To perform the forecast, 12 steps were defined, corresponding to a full year of future data. Using the previously fitted VECM model, forecasts were generated for manufacturing exports (exp_td_forecast) and the exchange rate (exchange_forecast). The forecast results were stored in a DataFrame for easier reading and analysis. A date index was added to the forecast DataFrame, starting from the end of the historical data (April 30, 2024), generating forecast dates at monthly intervals for the next 12 months.

	exp_td_forecast	exchange_forecast
2024-05-31	4.431725e+07	16.886859
2024-06-30	4.335987e+07	16.951409
2024-07-31	4.232600e+07	17.014138
2024-08-31	4.143161e+07	17.070539
2024-09-30	4.060726e+07	17.122008
2024-10-31	3.986285e+07	17.168654
2024-11-30	3.918649e+07	17.210996
2024-12-31	3.857314e+07	17.249405
2025-01-31	3.801661e+07	17.284253
2025-02-28	3.751174e+07	17.315868
2025-03-31	3.705369e+07	17.344550
2025-04-30	3.663813e+07	17.370571

Manufacturing Exports: The forecasted values for manufacturing exports show a decreasing trend over the next year. In May 2024, exports are expected to be approximately 43.4 million, gradually decreasing to 36.6 million in April 2025. This downward trend suggests a possible contraction in exports, which could be a concern for economic policies and exporting companies. This decrease could be related to various factors such as declining external demand, rising production costs, or loss of competitiveness in the international market.



MXN-USD Exchange Rate: The forecasted exchange rate shows an upward trend. It starts at 16.89 in May 2024 and increases to 17.37 in April 2025. This depreciation of the peso against the dollar implies that the peso might continue losing value in the next year, which could negatively affect export competitiveness and increase import costs. This trend may be influenced by factors such as inflation, monetary policies, political stability, and global economic conditions.



CONCLUSION

At the beginning we identified and proved with the Shapiro-Wilk test that the residuals where not normally distributed, we tried to assess this by making the hypothesis that the Logarithmic, Square-root or Box-Cox transformations could solve it, rejecting all of them due to low p-values.

To identify whether we should apply a Vector AutoRegressive Model (VAR) or Vector Error-Correction Model (VEC), we applied an Augmented Dickey Fuller (ADF) test proving that both variables, exp_td and exchange, where non stationary due to p-values greater than 0.05, which was also seen while plotting the AutoCorrelation Function (ACF) results.

We used a function to find how many differences we would need to solve this matter, where we started to obtain p-values less than 0.05 after the first one and applied it to the dataset, made a train-test split with the last 12 observations (months) and saw that the model was not significant and maintained its original problems, seeing not normally distributed residuals, autocorrelation and non-stationary variables.

This means that our last resource to identify whether to use VAR or VEC was the Johansen Cointegration Test, that made us accept the alternative hypothesis of having at least one cointegration vector (specifically due to exports while considering critical values of 5%, which meant that in order to accurately solve the clear problems of non normality in residuals, non stationary variables and autocorrelation, we need to use the Vector Error-Correction Model.

The results of the model imply manufacturer exports of **44.3M** for the immediate following month and **36.6M** in one year (expected generalized decrease in both short and long term), while for the USD-MXN exchange rate is of **$16.88** for the immediate following month and **$17.37** in one year (expected generalized increment in both short and long term).

Although it is clear that this is our best result with our current abilities, it's clear that whether improving the model or adding more variables could improve it significantly, especially considering that, for example, the current value of the USD-MXN exchange rate is already bigger than the biggest expected (Yahoo! Finance, 2024) due to the recent federal Mexican election results (Strohecker and Campos, 2024) and the current republican winning expectations in the United States (Duran, 2024), which could also decrease even more the expected exports (Palmer, 2024). Either way, we achieve an optimal, theoretically correct solution that actually proves the up or down trend.

RESOURCES

Duran, P. (July 23, 2024). Mexican Peso Rises as Biden Withdraws From US Presidential Race. Mexico Business. https://mexicobusiness.news/finance/news/mexican-peso-rises-biden-withdraws-us-presidential-race.

Palmer, D. (July 02 2024). Trump promised to rebalance trade in North America. The US trade deficit keeps climbing. Politico. https://www.politico.com/news/2024/02/07/donald-trump-trade-deficits-00140101.

Strohecker, K. and Campos, R. (June 03, 2024). Mexican peso, stocks tumble on fears of ruling coalition super-majority in Congress. Reuters. https://www.reuters.com/markets/currencies/mexico-peso-drops-nearly-3-sheinbaum-landslide-raises-reform-worry-2024-06-03/.

Yahoo! Finance. (July 23, 2024). USD/MXN (MXN=X). Yahoo Finance. https://finance.yahoo.com/quote/MXN%3DX/.