Analysing the Dynamics of the United States' International Oil Trades from the Perspective of Import/Export Volumes

Anantajit Raja

ABSTRACT

In this paper we conduct thorough statistical analyses of weekly import and export volumes of oil into and out of the United States in the 2012-2015 period, on a dataset consisting of barrels of oil loaded onto ships for national and international trades. We do an autocorrelation analysis of the import/export volume time series, a stationarity test and a Gaussianity test for these trade volumes, and estimate the distributions of volumes using Kernel Density Estimation. We also produce a log-log plot of the complementary cumulative distribution function to analyse tail fatness of the distributions of import/export volumes relative to the normal distribution.

After introducing the background for this problem, we begin our investigation through a meticulous preprocessing procedure for our original dataset to obtain weekly volume data for the United States' international trades, followed by insightful statistical analyses of our time series data points complemented by visualisations, and culminating in a detailed presentation of results and a discussion where the implication of our results are explored. We end with concluding remarks, including potential analyses to further enrich this research.

I. INTRODUCTION

In an era where energy security and economic dynamics are more intertwined than ever, understanding the ebb and flow of international oil trades becomes crucial. This analysis delves into the distribution and dynamics of the United States' international oil trades, focusing specifically on import and export volumes.

As a global powerhouse, the U.S. plays a pivotal role in the oil market, not only as a major consumer but increasingly as a significant producer[2]. The intricacies of import/export volumes reveal much about the United States' position in the global oil market. On one hand, imports reflect the country's demand for oil, driven by its vast industrial base, transportation needs, and consumer behavior. On the other, exports represent the U.S.'s growing capacity to supply oil to the world, a testament to its technological advancements in oil extraction and production techniques.

We carry out a comprehensive examination of the United States' international oil trades, emphasising the nuanced properties of import and export volumes.

II. METHODOLOGY & DATA

A. Preprocessing

Our original dataset presents an expansive and intricate portrait of international oil trades, encapsulating over 250,000 distinct datapoints. Each datapoint represents an oil trade, recorded with its specific date of occurrence. The dataset spans a global canvas, featuring trades between a diverse array of origin and destination countries. Accompanying each trade record is detailed information about the corresponding ports of origin and destination, providing a granular view of the global oil trade network. Given the dataset's extensive scope, with trades spanning multiple years and involving numerous countries and ports, the preprocessing stage of our analysis becomes paramount. This stage involves filtering the data to analyse the subset that we are interested in, cleaning the data and handling missing values.

Since we are only interested in the oil trades of the US done internationally, we filter the dataframe for rows (trades) where either the US is only the origin country or the US is only the destination country (we cannot have both at the same time since that allows for domestic trades too). We identify several categories (columns) of our dataset to be extraneous and remove them to streamline the dataset. Notably, columns detailing the names of ports of origin and destination were omitted. This information could offer logistical insights, but the emphasis of our study on the specific countries involved in the trades rendered such details unnecessary. We also exclude from the analysis the columns recording the number of ships involved in each trade and the alternate measure of loaded barrels ('loaded_barrels_new'). At this point we have around 10500 data points.

The presence of duplicates could distort our analysis by artificially inflating trade volumes. Therefore, we rigorously scan our dataset for such duplicates and remove them to maintain the authenticity of our trade records.

We also address the issue of missing values within our dataset. Focusing on the 'loadedbarrels' column, for each group of trades categorised by 'year' and 'week', we calculate two medians: one for trades where the United States is the origin country (exports), and another where the United States is the destination country (imports). This distinction allows us to account for potentially differing trade volumes in exports versus imports, allowing us to use condition-based median imputation[6]. For records with missing 'loadedbarrels' values, we impute these gaps using the previously calculated median corresponding to the trade direction (export or import). This method ensures that the imputed values are reflective of typical trade volumes for that specific time period and direction, preserving the dataset's overall integrity and distribution characteristics. To address any remaining missing values that median imputation does not cover, we employ linear interpolation[5] on the 'loadedbarrels' column as well. Linear interpolation is a method that estimates missing values by connecting two adjacent points with a straight line and using this line to estimate the missing data points. This technique provides a smooth transition between known values, ensuring a more continuous dataset that better reflects the natural fluctuations in oil trade volumes. We remove rows with missing entries in the 'destination_countryname' and 'origin_countryname' columns. Given that these instances are relatively few in number, we expect their exclusion to have minimal impact on the robustness of our analyses. We do this removal to ensure that every trade in our dataset can be associated with both an origin and a destination country.

Finally we aggregate oil volumes, summing up the 'loadedbarrels' for each unique combination of 'year' and 'week', to obtain the United States' oil import and export volumes on a weekly basis, segmented annually. We focus our analysis on the 2012-2015 period, which marked a period of significant technological advancements in oil extraction, particularly with the widespread adoption of fracking and horizontal drilling[4]. We are left with 418 datapoints (209 for imports and 209 for exports) that capture import and export dynamics across the 4 years.

B. Data exploration and visualisation

We can use pandas' 'describe()' function to obtain a surface-level overview of our prepared dataset. It provides us with the necessary foundation for more detailed analyses as the paper progresses.

Imports Statistics	
Loaded Barrels	**Value**
Count	209
Mean	5.70×10^6
Min	8.76×10^4
25%	1.81×10^6
50% (Median)	5.34×10^6
75%	9.06×10^6
Max	1.72×10^7
Standard Deviation	4.13×10^6

Exports Statistics	
Loaded Barrels	**Value**
Count	209
Mean	1.24×10^7
Min	2.16×10^6
25%	9.58×10^6
50% (Median)	1.20×10^7
75%	1.52×10^7
Max	2.91×10^7
Standard Deviation	4.44×10^6

TABLE I: Statistical Summaries of Loaded Barrels for Imports and Exports into/from the U.S. from Jan 2012 to Dec 2015

The mean volume of oil exported from the U.S. (1.24×10^7 barrels) notably surpasses that of the imports (5.70×10^6 barrels), indicating U.S.'s critical role as an oil exporter. The range between the minimum and maximum volumes of oil traded (both imports and exports) reveals the variability in trade volumes over the period. For imports, the range spans from 8.76×10^4 to 1.72×10^7 barrels, whereas for exports, it stretches from 2.16×10^6 to 2.91×10^7 barrels. The broader range for exports emphasises the U.S.'s expanding capability to supply oil internationally, likely influenced by fluctuating global demands and the strategic opening of U.S. oil markets to international buyers. The standard deviation of export volumes (4.44×10^6 barrels) compared to import volumes (4.13×10^6 barrels) suggests slightly higher variability in export volumes. This higher variability in exports could reflect the U.S.'s very responsive adjustments to international market demands.

For the years in question we can graph the volume of oil traded against time to serve as a visual narrative, depicting how import and export volumes have evolved. To illustrate this we display the plots for 2014 (see **Figures 1** and **2**).

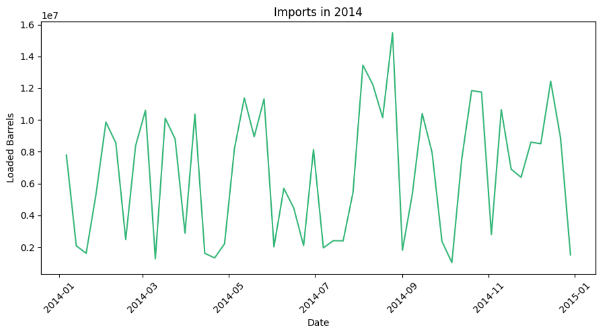

FIG. 1: Plot of weekly oil imports to the U.S. over 2014

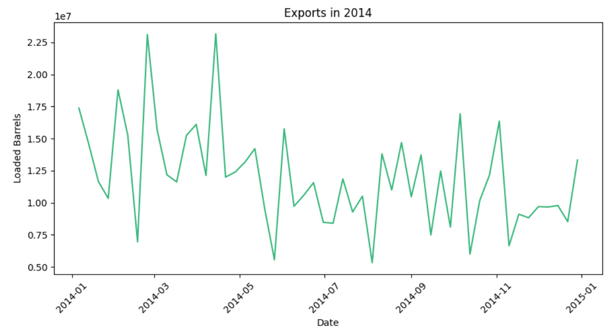

FIG. 2: Plot of weekly oil exports from the U.S. over 2014

The imports plot shows noticeable large fluctuations throughout the year. There are several large peaks followed by low troughs, indicating significant variability in the volume of oil imports week by week. It seems there is a slight downward trend in exports, indicating a slight decrease in trade volume during this period. Both plots could suggest some form of seasonality - we delve into autocorrelation analyses to see if this is the case over the 4 years in question.

C. Statistical Analyses

1. *Autocorrelation analysis*

Autocorrelation plots will enable us to examine whether there is a consistent pattern in the oil trade volumes that repeats over regular intervals, which is indicative of seasonality. This statistical approach will allow us to discern the presence and strength of periodic fluctuations, thereby offering a quantifiable measure of seasonality in the data.

Autocorrelation quantifies the relationship between observations in a time series and previous observations, or lags, of itself. Consider a time series X_t. The autocorrelation for a given lag k assesses the correlation of the series with its own past values, X_{t-k}. Mathematically, it is defined as the ratio of the covariance of X_t with X_{t-k} to the variance of X_t over all T observations:

$$R(k) = \frac{\sum_{t=k+1}^{T}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

Here, $\bar{X}$ denotes the mean value of X_t, and T represents the total number of observations within the time series. Autocorrelation values range from -1 to 1, where 1 indicates perfect positive autocorrelation, -1 indicates perfect negative autocorrelation, and 0 indicates no autocorrelation.

We produce autocorrelation plots for the weekly oil trades in **Figure 3**. At lag 0, autocorrelation is at its maximum by definition, since each data point is perfectly correlated with itself, yielding an autocorrelation coefficient of 1. The substantial decrease in the autocorrelation coefficient as we move to the higher lags suggests a weak dependency between the oil trade volumes in consecutive weeks. This rapid diminution in correlation with increasing lag indicates that the weekly oil trade volumes, both imported to and exported from the United States, are not heavily influenced by their immediate past values. Practically, this implies that the market conditions or trade policies governing oil imports and exports are subject to frequent changes and could be driven by factors that vary on a timescale shorter than our weekly observation interval.

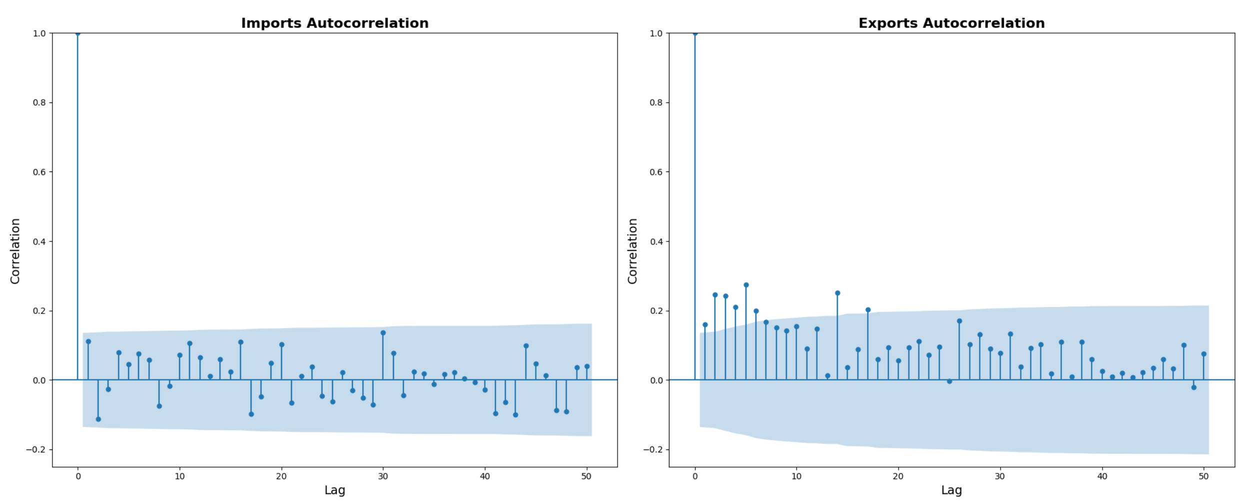

FIG. 3: Autocorrelation function plotted against different lags k from 0-50 for import and export of oil in the U.S. over the whole period from 2012 to 2015, consisting of T = 209 observations each

2. *Stationarity analysis*

Understanding whether our data is stationary helps us characterise its behavior. If the import/export series exhibit non-stationarity, this might indicate the presence of trends or cycles, which could reflect underlying economic phenomena influencing oil trade volumes.

Strong stationarity characterises a time series by the unchanging behavior of its joint probability distributions over time. This means that for any set of observations $(y_1, y_2, \ldots, y_n)$ within the series, their statistical interrelationships remain consistent when shifted across any time lag τ.

In our analysis of U.S. oil import and export volumes, we employ the Augmented Dickey-Fuller (ADF) test[3] to investigate the presence of a unit root, which would suggest non-stationarity in our import/export time series.

Consider a time series y_t expressed through the model:

$$\Delta y_t = a + \rho y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \ldots + \epsilon_t$$

In this equation, Δ denotes the differencing operator, and $a, \rho, c_1, c_2, \ldots$ are the model's constant coefficients. We examine the statistic:

$$\frac{\hat{\rho}}{\mathrm{SE}(\hat{\rho})}$$

where $\hat{\rho}$ is the estimated coefficient for y_{t-1}, and SE represents its standard error. The test evaluates the following hypotheses:

H_0: $\rho = 0$, which indicates the presence of a unit root, suggesting the series may be non-stationary.

H_1: $\rho \neq 0$, which suggests the absence of a unit root, implying the series is stationary.

The test's significance is assessed by comparing the statistic against the test's critical values. A test statistic more negative than the critical value leads to rejection of H_0, signifying stationarity of the series. Conversely, a p-value less than the chosen significance level α also rejects H_0, given the improbability of observing the current data if the series were non-stationary. The outcome of this test, and reasoning behind this outcome, will be presented in the Results and Discussion section, along with all other tests' results.

The presence of a unit root would suggest that the import/export volumes will undergo lasting changes in response to sudden shocks, without returning to its long-term average.

3. *Gaussianity analysis*

Gaussianity implies that the bulk of observations cluster around a central mean, with symmetrical tails extending in both directions. Our Gaussianity analysis will be conducted separately for import and export volumes, allowing us to identify any differences in the distributional properties between these two facets of the oil trade. Insights derived from this analysis will inform the selection of appropriate statistical models and tests for subsequent stages of our study. In this context, we apply the Kolmogorov-Smirnov Distribution (KS) Test[7] to examine whether the empirical evidence supports the notion that the dataset follows a normal distribution.

The KS test is a non-parametric method to evaluate the goodness of fit between a sample's distribution and the reference distribution, determining if $\{y_i\}_{i=1}^{n}$ originates from a predefined CDF $F(y)$. The empirical CDF, $F_n(y)$, is constructed as:

$$F_n(y) = \sum_{i=1}^{n} \theta(y - y_i) \text{ where } \theta(y - y_i) = \begin{cases} 1 & \text{if } y \geq y_i \\ 0 & \text{otherwise.} \end{cases}$$

We test the statistic $D_n = \sup |F(y) - F_n(y)|$, which approaches zero as the sample size n increases, provided that $\{y_i\}_{i=1}^{n}$ truly derives from $F(y)$. Under this framework, the hypotheses are as follows:

H_0: The scaled variable $\sqrt{n}D_n \sim P(K)$ distribution.

H_1: $\sqrt{n}D_n$ diverges from $P(K)$.

We reject H_0 if the empirical statistic $\sqrt{n}D_n$ is greater than the critical value K_α. To assess whether weekly import/export volumes follow a normal distribution, we use the KS test applied using a precise approximation of the normal distribution's CDF provided by a high-quality library function.

4. *The Kernel Density Estimation Method*

Kernel Density Estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable[1]. This feature is particularly beneficial in our context where the distribution of import and export volumes may not conform to any predefined parametric distribution, such as the normal distribution. It allows us to explore the data's structure without imposing strict assumptions about its underlying distribution.

Given a set of n independent and identically distributed (i.i.d.) samples $\{y_i\}_{i=1}^n$ drawn from some distribution with an unknown density f, the KDE of f at a point y is defined as:

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right)$$

where: $\hat{f}(y)$ is the estimate of the density function at point y; $K(\cdot)$ is the kernel, which is a non-negative function integrating to one and h is the bandwidth, a positive parameter that controls the smoothness of the density estimate. h determines the width of the kernel and hence affects the degree of smoothing.

We use the Gaussian kernel given its ability to effectively model the smooth curves that could resemble the distribution of trade volumes. We overplot the histograms representing empirical pdfs (obtained through aggregation into bins of size 10^6) with their KDE for each year to see if and how the dsitributions change over time (see **Figure 4**).

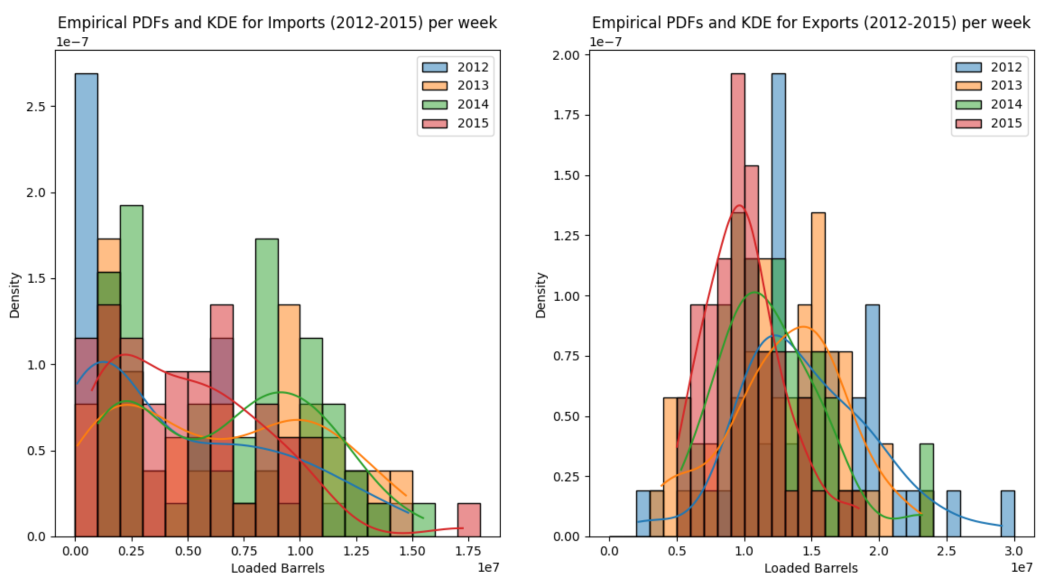

FIG. 4: Empirical PDFs for import and export volumes overplotted with their KDEs for years 2012-2015

Note that the distribution of weekly import volumes is roughly similar across the years whilst the average weekly volume of oil exported seems to decrease year-on-year. We will later confirm this by calculating the relevant statistics. Our visual representation tells us that we may expect to see the ADF test indicating stationarity for imports (the distribution appears to remain roughly the same over the years) and possibly, because of its decreasing mean, non-stationary for exports (though as we will see shortly from the ADF test results, the weekly export volumes are actually stationary - stressing the importance of rigorous hypothesis tests which can go against our intuition).

5. *Complementary cumulative distribution - analysis of tail fatness*

The complementary cumulative distribution function (CCDF) is a powerful tool for analysing the 'fat tails' of a distribution. A fat tail indicates that large values in the data set are more common than would be expected in a normal distribution, which is characterised by relatively thinner tails. The CCDF is calculated as one minus the cumulative distribution function (CDF), which altogether represents the probability that a random variable will take a value greater than or equal to a certain value. By plotting the CCDF on a log-log scale, we can readily discern how the tails of the our distribution compare to the tails of the normal distribution. In such a log-log plot, a normal distribution will typically appear as a concave decreasing curve. The rate of decrease of the CCDF for our distribution relative to this curve will determine how heavy-tailed our distribution is.

Applying this to our context of U.S. oil import and export volumes, using the CCDF could highlight the occurrence of unusually large trading volumes. For instance, if we observe that the CCDF of our import or export data does not decline as rapidly as the normal distribution curve, it would suggest that large fluctuations in trade volume (representing significant trade activities or economic events) occur more often than expected under normal circumstances. We create a log-log plot of the CCDF for our data (**Figure 5**).

For illustrative purposes, here we only display the plots for 2014. We construct the empirical CCDF by sorting the weekly trade volume data and then plotting 1 minus the cumulative proportion of data at each trade volume level against the sorted volumes on a log-log scale. To draw comparisons, we also overlay a theoretical CCDF derived from a normal distribution with the same mean and standard deviation as the empirical data. This is achieved through Monte Carlo simulations, generating a large number of random values from the normal distribution (10,000 in this case) to form a very well-approximated theoretical curve. The theoretical CCDF is plotted similarly by sorting the simulated values and plotting 1 minus their cumulative proportions.

For the import data, the CCDF exhibits a more rapid decline than the theoretical line representing a normal distribution. This sharper falloff indicates that the occurrence of extreme values, or "fat tails", is less frequent than what would be expected if the data were normally distributed. Such a pattern characterises a "thin-tailed" distribution, suggesting that very high or very low import volumes are rarer than average ones and deviations from the mean are less pronounced.

Conversely, the export data's CCDF aligns more closely with the theoretical normal distribution's CCDF. The closeness of the empirical and theoretical curves suggests that the export volumes may fit well within the Gaussian model (though evidence of this will be further strengthened through the results of the KS test), reflecting that high or low export volumes occur with a frequency that one would typically expect in normally distributed data.

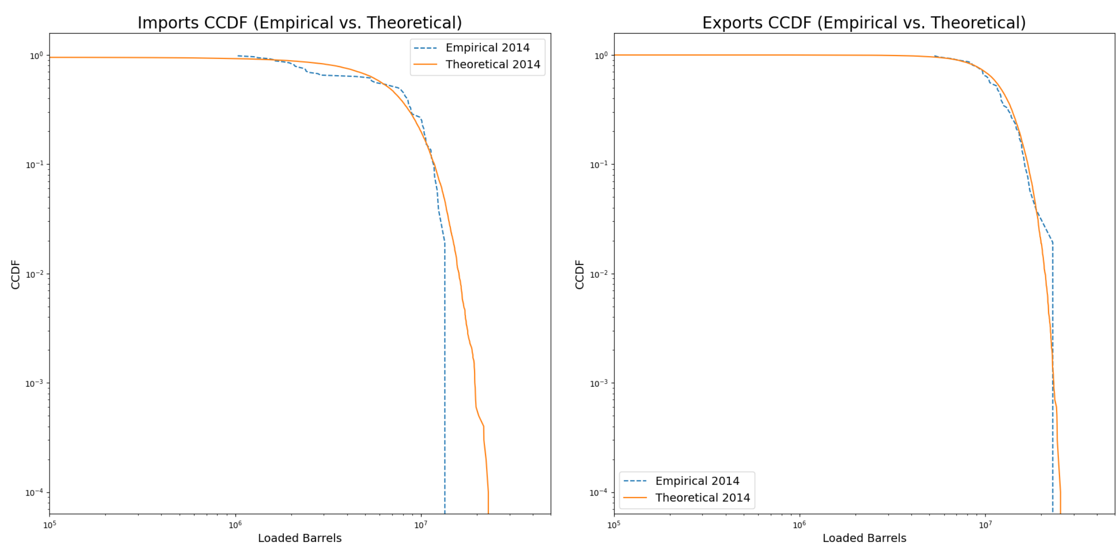

FIG. 5: Log-log plot of the empirical CCDFs for import and export volumes vs theoretical CCDF derived from a normal distribution

III. RESULTS AND DISCUSSION

In this section, we delve into the outcomes of our comprehensive analysis concerning the stationarity, Gaussianity and overall annual statistical characteristics of the weekly U.S. oil import and export volumes. Each of these facets contributes to a holistic understanding of the U.S. international oil trade market dynamics.

A. Stationarity: Results and Analysis

ADF Test for Stationarity	**Measure**	**Imports**	**Exports**
Statistic & p-value	ADF Statistic	-10.858	-3.111
	p-value	~ 0	0.0257
Conclusion: Both series are stationary (reject H_0)			

TABLE II: Summary of Augmented Dickey-Fuller Tests for Weekly Trade Volumes from 2012 to 2015 (3dp)

The results of the ADF tests for U.S. oil trade volumes from 2012 to 2015 provide compelling evidence for the stationarity of weekly import and export volumes. With ADF statistics of -10.858 for imports and -3.111 for exports, coupled with low p-values (close to zero for imports, and 0.0257 for exports), we can confidently reject the null hypothesis of a unit root at the 5% significance level. This implies that both the import and export series are stationary, indicating that they do not exhibit trends or seasonality over time.

This stationarity in trade volumes is likely the result of a combination of regulatory consistency in governing oil trade, the U.S.'s diversified trading partners (multiple different countries) and the adaptability of the U.S. oil supply chain to external pressures, such as fluctuations in global oil prices or changes in demand. These factors collectively mitigate against long-term trends or seasonal effects, ensuring a stable trade environment.

B. Gaussianity: Results and Analysis

Gaussianity (KS test)	**Measure**	**Imports**	**Exports**
Statistic & p-value	Statistic	0.143	0.084
	p-value	~ 0	0.098
Conclusion: Imports - Not Gaussian (reject H_0); Exports - Gaussian (do not reject H_0)			

TABLE III: Summary of Gaussianity Tests for Weekly Trade Volumes from 2012 to 2015 (3dp)

Given that the import and export trade volume series are stationary, we can analyse them further by conducting a Gaussianity test on the series' values.

At the 5% significance level, for weekly import volumes, the test statistic significantly departs from what would be expected under a normal distribution, with a p-value close to zero. This strongly suggests that the import volumes diverge from normality, indicating that the distribution of these volumes exhibits characteristics, such as skewness or kurtosis, that are inconsistent with the Gaussian model. However, at the same significance level, we find that there is not enough evidence to suggest that exports are not normally distributed (i.e. export volumes are indeed Gaussian).

Weekly oil import volumes may follow a non-Gaussian characteristics for many reasons, each one being deeply rooted in the operational and economic aspects of the global oil market. Disruptions in the supply chain, such as geopolitical tensions, can lead to abrupt changes in import volumes, contributing to a distribution that diverges from the normal curve. Furthermore, alterations in trade policies, tariffs, or sanctions can instantly affect oil imports, inducing non-Gaussian behavior. The domestic demand for oil is another big factor: it fluctuates with economic cycles, seasonal energy needs, and shifts towards renewable resources, potentially introducing skewness or kurtosis into the import volume distribution. Oil price volatility, mostly driven by global supply-demand dynamics, also plays a critical role. Such price fluctuations can lead buyers to adjust their import strategies, further distancing the volume distribution from Gaussian norms.

On the other hand, we see that export volumes follow a Gaussian distribution. This can be attributed to U.S. oil production levels being relatively constant, meaning that the volume available for export would likely fluctuate around a stable mean. Exports are also influenced by comprehensive regulatory frameworks that manage the pace and volume of exports, smoothing out potential spikes and drops and bringing the overall distribution closer to normal.

C. General Summary Statistics: Results and Analysis

The mean import volumes show a peak in 2014 at 6.686×10^6, suggesting an increase in oil imports during this year, possibly due to increased domestic demand or lower international oil prices. The lowest means occur in 2012 and 2015, indicating fluctuations in import needs

Year	Mean	Standard Deviation	Skewness	Excess Kurtosis
Weekly Import Volumes				
2012	4.712×10^6	4.338×10^6	0.677	-0.807
2013	6.473×10^6	4.382×10^6	0.172	-1.322
2014	6.686×10^6	4.016×10^6	0.086	-1.221
2015	4.927×10^6	3.516×10^6	0.971	1.328
Weekly Export Volumes				
2012	14.543×10^6	5.023×10^6	0.407	0.889
2013	13.116×10^6	4.468×10^6	-0.258	-0.140
2014	11.957×10^6	3.906×10^6	0.804	1.012
2015	10.071×10^6	2.955×10^6	0.781	0.719

TABLE IV: Statistical Summary of U.S. Oil Imports and Exports from 2012 to 2015

or global market conditions. Export volumes exhibit a gradual decrease in the mean from 14.543×10^6 in 2012 to 10.071×10^6 in 2015. This trend could reflect changing global demand, U.S. production shifts, or changes in U.S. oil export policies.

The standard deviation values, which indicate the variability in trade volumes, are consistently around 4×10^6 for imports. This suggests that the domestic demand for oil in the U.S. may have remained relatively stable over these years. While there are fluctuations, they might not be drastic enough to significantly affect the overall variability of import volumes.

Skewness for both imports and exports fluctuate about a similar range, with a notable spike in 2015 (0.971) for imports and in 2014 (0.804) for exports, implying a shift towards more asymmetric distributions with a longer right tail in those years. These sporadic increases can be attributed to several factors from global oil prices to fluctuations in domestic and international demand.

The negative excess kurtosis values for years 2012-2014 for import volumes suggest that these years had fewer and less extreme outliers than a normal distribution. This could be due to relatively stable global supply conditions and consistent domestic demand, leading to fewer disruptions or spikes in import volumes. A shift is observed in 2015, where the kurtosis turns positive (1.328), possibly due to market volatility, sudden changes in domestic demand, or supply disruptions. The exports' positive kurtosis values are likely due to the higher global demand for U.S. oil in those years. The slight negative value in 2013 can possibly be attributed to a stable year for global oil markets.

IV. CONCLUSION & OUTLOOK

This research paper has provided an in-depth statistical analysis of the weekly oil import and export volumes of the United States.

Through autocorrelation analysis, we have explored the temporal dependencies (or lack thereof) within the import and export volume time series, revealing the complex dynamics of oil trade flows. Our stationarity tests have affirmed the consistency of these trade volumes over time, reflecting a market capable of adapting to various global economic pressures while maintaining a steady course. The Gaussianity test has highlighted the deviation from normal distribution in import volumes. We have also used the log-log plot of the CCDF to show the presence of thin tails in certain years, which suggests a level of predictability and stability in U.S. oil imports. Export volumes exhibit a closer alignment with the Gaussian model, which points towards the influence of external market demands and geopolitical factors on U.S. oil exports.

Looking ahead, there are avenues for further research that could enrich our understanding of international oil trades. Incorporating predictive modeling to forecast future trade volumes and analysing the impact of emerging renewable energy trends on oil trades stand out as promising directions. As the world increasingly moves towards renewable energy sources, understanding how this transition affects traditional oil trade patterns will be crucial for policymakers and industry stakeholders. Comparative studies with other major oil-producing and -consuming nations could also offer valuable global insights. Exploring the relationship between oil prices and trade volumes also presents a compelling research perspective. By examining how fluctuations in oil prices influence import/export volumes, we can gain insights into the price elasticity of oil demand and supply. This could involve statistical analyses to identify correlations or causations between oil prices and trade volumes, potentially using advanced econometric models such as vector autoregression (VAR) or cointegration tests to help understand these complex relationships.

---

[1] Kernel density estimation and its application. In *XLVIII Seminar of Applied Mathematics*, volume 23, page 00037, 2018. Published online 07 November 2018.

[2] Gawdat Bahgat. Oil producers' perspectives on energy security. In *Economics 2013*, chapter 12, pages 258–272. Edward Elgar Publishing, 2013. Restricted access.

[3] Y.-W. Cheung and K. S. Lai. Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280, 1995.

[4] James Feyrer, Erin T. Mansur, and Bruce S. Sacerdote. Geographic dispersion of economic shocks: Evidence from the fracking revolution. *Working Paper*, November 2016. 9.

[5] S. Kay. Some results in linear interpolation theory. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(3):746–749, 1983.

[6] Joseph L. Schafer. Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8(1), 1999.

[7] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, 19(2):279–281, 1948.