Quantifying Regional Causalities and Lags Between Crude and Distillates

Jack Thompson

Objective and Introduction

Crude oil is the primary input for refined products such as gasoline, diesel, kerosene (jet fuel), and liquefied petroleum gas (LPG). While it is well understood that fluctuations in crude oil prices influence refined product prices, the strength and timing of this transmission vary significantly across regions and seasons due to supply chain dynamics and shifting demand patterns.

While the long-term price relationship between crude oil and refined products is well-established through their production linkages, the short-term price transmission mechanisms are more nuanced and less understood. This study aims to quantify these dynamic, short-term influences by employing Granger causality to assess predictive relationships and lead-lag analysis to capture temporal dynamics in price transmission. By comparing regional and seasonal variations, we seek to uncover structural differences in how quickly and strongly refined products respond to movements in crude benchmarks.

Research Questions:

- To what extent do crude oil prices Granger-cause short-term price changes in their refined counterparts?
- How do lead-lag dynamics vary across regions (e.g., Gulf Coast vs. East Coast)?
- Which refined products respond the most quickly, or slowly, to shifts in crude oil prices?

The analysis focuses on the relationship between two crude oil benchmarks; WTI (West Texas Intermediate) and Brent, and key refined products, including diesel, gasoline, jet fuel, and propane. By examining multiple U.S. markets, this research provides insights into both global (Brent) and domestic (WTI) drivers of price transmission in the distillates market, particularly in the context of day-to-day price formation and short-term forecasting.

Contents

1 Data Preparation

1.1 Data Sources

The Federal Reserve Economic Data[2] (FRED) system provides a wide range of economic indicators, including daily prices for crude oil and refined petroleum products. These prices are sourced from various entities such as the Energy Information Administration (EIA), the Intercontinental Exchange (ICE), and the CME* Group. However, FRED itself does not directly price commodities but aggregates and republishes data from these sources. Below is an outline of the pricing methodology for crude oil and distillates available in the FRED database.

1.1.1 Crude Oil Pricing in FRED

FRED provides benchmark crude oil prices for both **West Texas Intermediate (WTI)** and **Brent** crude:

- **WTI Crude Oil Price (DCOILWTICO)**: This is based on front-month futures contracts traded on the CME Group and represents pricing at Cushing, Oklahoma.
- **Brent Crude Oil Price (DCOILBRENTEU)**: This reflects front-month futures contracts from the ICE Exchange, representing North Sea crude pricing.

Since there is no true spot market for crude oil, these futures-based prices serve as the closest approximation to real-time market values.

1.1.2 Distillate Pricing in FRED

FRED also provides pricing for refined petroleum products, including diesel, gasoline, and jet fuel. These prices are generally survey-based and represent wholesale or retail prices from EIA reports. The datasets include:

- **Diesel Prices**:
  - **Los Angeles (DDFUELLA)**
  - **New York Harbor (DDFUELNYH)**
  - **U.S. Gulf Coast (DDFUELUSGULF)**
- **Gasoline Prices**:
  - **New York Harbor (DGASNYH)**
  - **U.S. Gulf Coast (DGASUSGULF)**
- **Jet Fuel Prices**:
  - **U.S. Gulf Coast (DJFUELUSGULF)**
- **Propane Prices**:
  - **Mont Belvieu, Texas (DPROPANEMBTX)**

*The CME Group operates multiple exchanges, including the Chicago Mercantile Exchange (CME), the Chicago Board of Trade (CBOT), the New York Mercantile Exchange (NYMEX), and the Commodity Exchange Inc. (COMEX).

Even though distillate prices are available daily, they are likely based on survey data, spot market transactions, and limited futures contracts, rather than purely futures-driven pricing like crude oil. This is a key distinction when analyzing the relationship between WTI/Brent and U.S. distillate pricing.

1.2 Imputing Missing Values

For this paper I am using the last 5 years of daily price data for each of the products. Starting from 01/01/2020 until 01/01/2025. Naturally, there will be breaks in these data due to various factors, such as government-imposed trading halts, national holidays, or other market disruptions. However, the missing days do not align across all products, creating a challenge when comparing time series data against one another. Inconsistent gaps across datasets pose a more complex problem than if they were uniformly distributed across all time series. If the missing values were synchronized, we could simply exclude those time points from all series without introducing bias. However, since the gaps occur at different times for different products, we need a method to impute missing values without introducing distortions or biases into our models.

Column	NaN Count	% Missing
DCOILBRENTEU	40	3.1
DCOILWTICO	53	4.1
DDFUELLA	59	4.5
DDFUELNYH	54	4.1
DDFUELUSGULF	54	4.1
DGASNYH	54	4.1
DGASUSGULF	54	4.1
DJFUELUSGULF	54	4.1
DPROPANEMBTX	54	4.1

Table 1: Total Missing Values in Price Data per Product

Fortunately, in relation the the total number of days recorded (1305), the proportion of missing values is very small for each time series.

To better understand the nature of these missing values, we visualize their distribution over time. This allows us to determine whether they are clustered into large contiguous gaps or randomly scattered throughout the dataset.

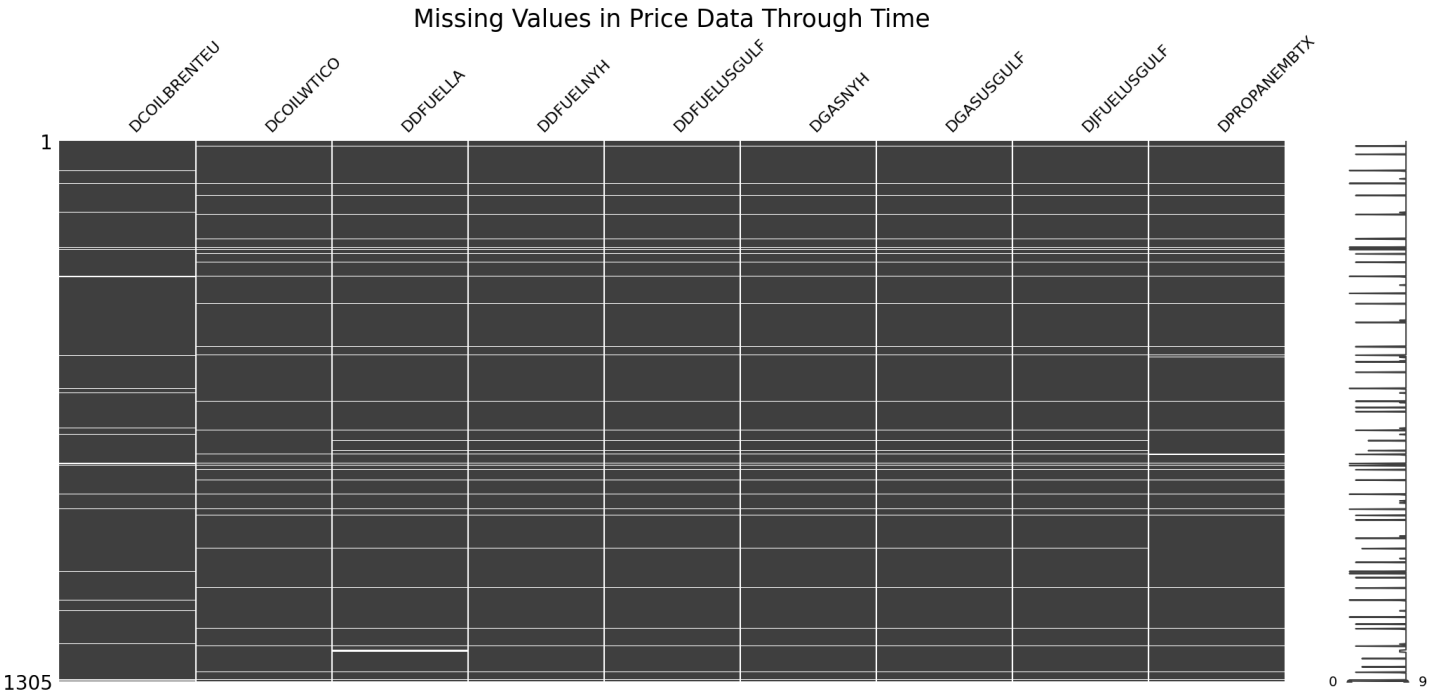

Figure 1: Missing Values in Price Data Through Time per Product (y-axis = time & x-axis = product class)

As seen in Figure 1, the missing data exhibits sporadic behavior, with most products showing minor clustering near the beginning and middle of the time series—except for Brent, which appears to have more evenly scattered gaps.

Since there are no prolonged periods of missing data, we should not lose significant information through imputation. To address these gaps, I will apply three different imputation techniques and evaluate them using specific metrics to determine the most suitable approach for further analysis.

1.2.1 Moving-Average Method

The Moving Average (MA) method with centered window size 3 provides balanced local smoothing for missing value imputation. This approach uses a symmetric neighborhood of 1 period before and after each missing value (plus the current position), making it particularly responsive for high-frequency commodity price data.

Precise Window Formulation For a time series $\{S_t\}_{t=1}^{T}$ with a missing value at position t, the imputed value $\hat{S}_t$ is estimated as:

$$\hat{S}_t = \frac{1}{|\mathcal{W}_t|} \sum_{i \in \mathcal{W}_t} S_i$$

where the window $\mathcal{W}_t$ is defined as:

$$\mathcal{W}_t = \{i \in \{t-1, t, t+1\} \cap \{1, \ldots, T\} \mid S_i \text{ is not missing}\}$$

Algorithm 1 Moving Average (MA-3) Imputation

Input: Time series $\{S_t\}_{t=1}^{T}$ with missing values
for each t where S_t is missing **do**
  Construct window: $\mathcal{W} \leftarrow \{t-1, t, t+1\} \cap \{1, \ldots, T\}$
  Extract valid observations: $\mathcal{W}_{\text{valid}} \leftarrow \{i \in \mathcal{W} \mid S_i \text{ is not missing}\}$
  if $\mathcal{W}_{\text{valid}} \neq \emptyset$ **then** ▷ Ensure at least one valid value
    Compute imputed value: $\hat{S}_t \leftarrow \frac{1}{|\mathcal{W}_{\text{valid}}|} \sum_{i \in \mathcal{W}_{\text{valid}}} S_i$
  end if
end for
Output: Imputed time series $\{\hat{S}_t\}_{t=1}^{T}$

Key Characteristics

- **Symmetric treatment**: Equal weight to past and future observations
- **Adaptive padding**: Automatically shrinks window at series edges
- **Volatility preservation**: Maintains price jumps better than larger windows

Advantages for Oil Prices

- Handles typical 1-3 day gaps in commodity pricing data
- Smooths minor fluctuations while preserving major trends
- More stable than single-point methods during periods of:
  - Exchange rate fluctuations
  - Temporary supply disruptions
  - Reporting artifacts

Limitations

- **Edge attenuation**: Weakens signal at price turning points
- **3-day memory**: Can slightly delay response to shocks
- **Non-causal**: Requires future values (may not suit real-time systems)

The choice of window size 3 represents an optimal balance for daily oil price data, providing sufficient smoothing without excessive lag - as demonstrated in the comparative results in Section 1.2.4.

1.2.2 Linear Interpolation

Linear interpolation provides a more sophisticated approach for gap filling by estimating missing values through straight-line connections between existing data points. This method is particularly effective for continuous time series data with relatively small gaps.

Mathematical Foundation For a time series S_t with missing values at $t \in T_{\text{missing}}$, the linear interpolation estimate $\hat{S}_t$ between two known points (t_1, S_{t_1}) and (t_2, S_{t_2}) is given by:

$$\hat{S}_t = S_{t_1} + \frac{t - t_1}{t_2 - t_1}(S_{t_2} - S_{t_1}) \quad \text{for} \quad t_1 < t < t_2$$

where t_1 and t_2 are the nearest non-missing time points before and after t respectively.

Algorithm 2 Linear Interpolation Imputation

Input: Time series S with missing values
Identify all gaps $G = \{(t_{\text{start}}, t_{\text{end}})\}$
for each gap $(t_a, t_b) \in G$ **do**
    $S_{\text{slope}} \leftarrow \frac{S_{t_b} - S_{t_a}}{t_b - t_a}$
    for $t \in (t_a, t_b)$ **do**
        $\hat{S}_t \leftarrow S_{t_a} + S_{\text{slope}} \cdot (t - t_a)$
    end for
end for
Output: Imputed series $\hat{S}$

Advantages

- **Preserves local trends**: Captures directional movements between observed points
- **Smooth transitions**: Avoids discontinuities in the imputed series
- **Computationally efficient**: $O(n)$ complexity for n missing values
- **Parameter-free**: No tuning required for basic implementation

Limitations

- **Assumes linearity**: May overshoot during non-linear price movements
- **Sensitive to gap size**: Accuracy decreases with larger gaps
- **Edge case handling**: Cannot interpolate leading/trailing missing values

Application to Oil Price Data For the Brent crude oil dataset, linear interpolation works particularly well because:

- Short-term price movements often follow quasi-linear trends
- The typical gap length in our dataset (1-3 periods) is suitable for linear estimation
- It preserves the overall shape of price trajectories better than constant-fill methods

The method's performance can be formally evaluated using the metrics from Section 1.2.4, where it typically outperforms simple filling methods while remaining computationally inexpensive.

1.2.3 Forwards & Backwards Fill Method

The forwards and backwards fill (LOCF-NOCB) method is a simple yet robust approach for imputing missing values in time series data. The technique combines two fundamental operations:

1. **Last Observation Carried Forward (LOCF)**: Replaces missing values with the most recent available observation
2. **Next Observation Carried Backward (NOCB)**: Fills remaining gaps using the next available observation

The implementation consists of the following steps:

Algorithm 3 Forwards-Backwards Fill Imputation

1: Input: Time series S with missing values
2: $S_{\text{forward}} \leftarrow \text{ffill}(S)$ ▷ Forward fill operation
3: $S_{\text{complete}} \leftarrow \text{bfill}(S_{\text{forward}})$ ▷ Backward fill operation
4: Output: Imputed series S_{complete}

Advantages

- **Computationally efficient**: Only requires two passes (forward + backward) through the data.
- **No parameter tuning**: Works without selecting window sizes or model assumptions.
- **Preserves observed values**: Does not smooth or alter existing data points.

Limitations

- **Lags behind sharp changes**: Cannot adjust to sudden shifts during gaps—only reflects values at gap edges.
- **Flat imputations**: Creates artificial plateaus in missing regions, distorting trends.
- **Ignored dynamics**: Fails to account for underlying patterns (e.g., momentum, seasonality).

Mathematical Formulation For a time series S_t with gaps at $t \in T_{\text{missing}}$, the imputed values $\hat{S}_t$ are:

$$\hat{S}_t = \begin{cases} S_{t_{\text{prev}}} & \text{if } t_{\text{prev}} = \max\{t' < t \mid t' \notin T_{\text{missing}}\} \\ S_{t_{\text{next}}} & \text{if } t_{\text{next}} = \min\{t' > t \mid t' \notin T_{\text{missing}}\} \end{cases}$$

where t_{prev} and t_{next} are the nearest non-missing time points before and after t respectively.

This method was particularly suitable for our Brent crude oil price dataset due to its ability to handle short gaps while preserving the sharp price movements characteristic of commodity markets.

1.2.4 Results

For the comparison, I extracted the largest contiguous subset of data with no pre-existing gaps. This subset consisted of Brent Oil prices from the period 27/08/2024 to 24/12/2024, which I refer to as *Original Data*.

To evaluate imputation methods, I artificially introduced missing values by randomly removing approximately 5,10,15% of the data points. This resulted in a gapped dataset with a known ground truth, allowing for an accurate assessment of imputation performance.

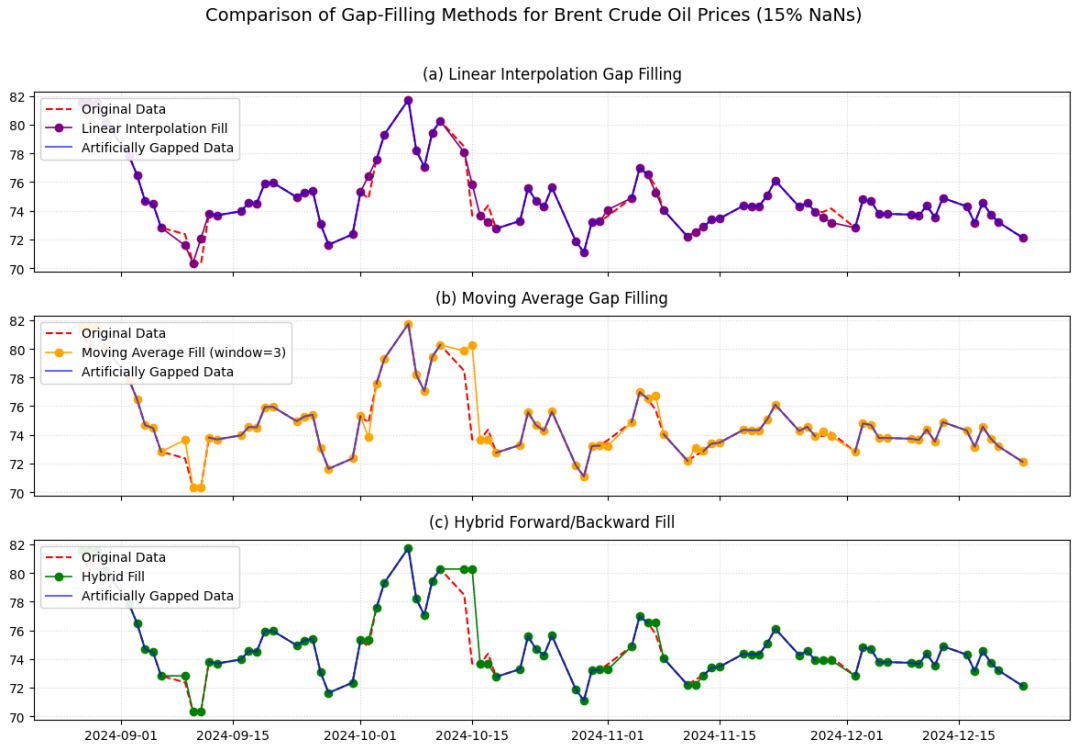

Figure 2: Imputation Methods Performance Visual Comparisons. Purple: Linear Interpolation, Yellow: Moving Average, Green: Forward & Backward Fill. Red-line: artificially removed original data.

Visually, we can see that the Moving-Average & Hybrid Fill methods seem to struggle with suddent movements in price, showing a lagged response to extreme fluctuations. This is was a limitation we foresaw with both. The MA-model has this limitation because it only fills based on historical data, which is not a restriction for our purpose since we are filling in gaps in already recorded data, not filling it in live.

Additionally, we can also identify the false plateu-ing behaviour we also noted as a possible limitation of the hybrid filling method.

We can see from Figure 2 that our linear interpolation method may be the best suited imputation algorithm to fit. But before we decide this we need to look at some quantitative results.

Evaluation Metrics The imputation methods were assessed using four common metrics for quantifying the errors of a prediction. In our case the $\hat{y}_i$ represent the filled $\hat{S}_t$ series data annotation used to define our models earlier.

Metric	Formula	Interpretation
MSE (Mean Squared Error)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	• Measures average squared difference • Sensitive to large errors (punishes outliers heavily) • Harder to interpret due to squared units
RMSE (Root MSE)	$\sqrt{\text{MSE}}$	• Square root of MSE • Still emphasizes large errors
MAE (Mean Absolute Error)	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	• Average absolute difference • More robust to outliers than MSE/RMSE
MAPE (Mean Absolute Percentage Error)	$\frac{100\%}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert$	• Percentage version of MAE • Relative error measure (scale-indep.) • Problematic when $y_i \approx 0$

Table 2: Imputation Performance Metrics and their formula

These results allow us to make a more data-driven decision regarding the choice of imputation method. By calculating each metric for each method across 5%, 10%, and 15% missing data, we observe that the Linear Interpolation method consistently outperforms the others, providing the most accurate imputations.

From Table 1, we see that our dataset contains approximately 4.1% missing data, distributed in spuriously placed gaps. Therefore, the results from the 5% missing data scenario should be fairly representative when applying Linear Interpolation to our actual missing data. Based on this, we will proceed with Linear Interpolation before conducting further statistical analysis.

Method	Missing %	MSE	RMSE	MAE	MAPE
Moving Average	5%	0.1891	0.4348	0.0745	0.10%
	10%	0.2369	0.4867	0.1309	0.17%
	15%	0.6166	0.7852	0.1771	0.24%
Linear Interpolation	5%	0.0623	0.2495	0.0457	0.06%
	10%	0.0938	0.3063	0.0828	0.11%
	15%	0.1908	0.4368	0.1372	0.18%
Forwards/Backwards Fill	5%	0.2174	0.4663	0.0779	0.10%
	10%	0.2519	0.5019	0.1173	0.15%
	15%	0.5966	0.7724	0.1565	0.21%

Table 3: Imputation Methods Comparison Across Different Missing Data Percentages

1.3 Stationarity

Before we can get to applying our tests to quantify any sort of statistical causalities, our data must first be stationary.

Now, price data itself is almost never stationary. However, if we difference the data, we get the returns. The returns can very often be stationary themselves, while containing all the same information as the price.

To difference each of our time series we will be using the log returns. The log return r_t at time t is calculated as:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln(S_t) - \ln(S_{t-1}) \tag{1}$$

where:

- S_t is the price at time t
- S_{t-1} is the price at the previous time period
- $\ln(\cdot)$ denotes the natural logarithm

The consequence of this is we lose 1 datapoint in the very beginning of our entry since for our initial start time, $t = 0$, $S_{0-1} = S_{-1}$ is not defined.

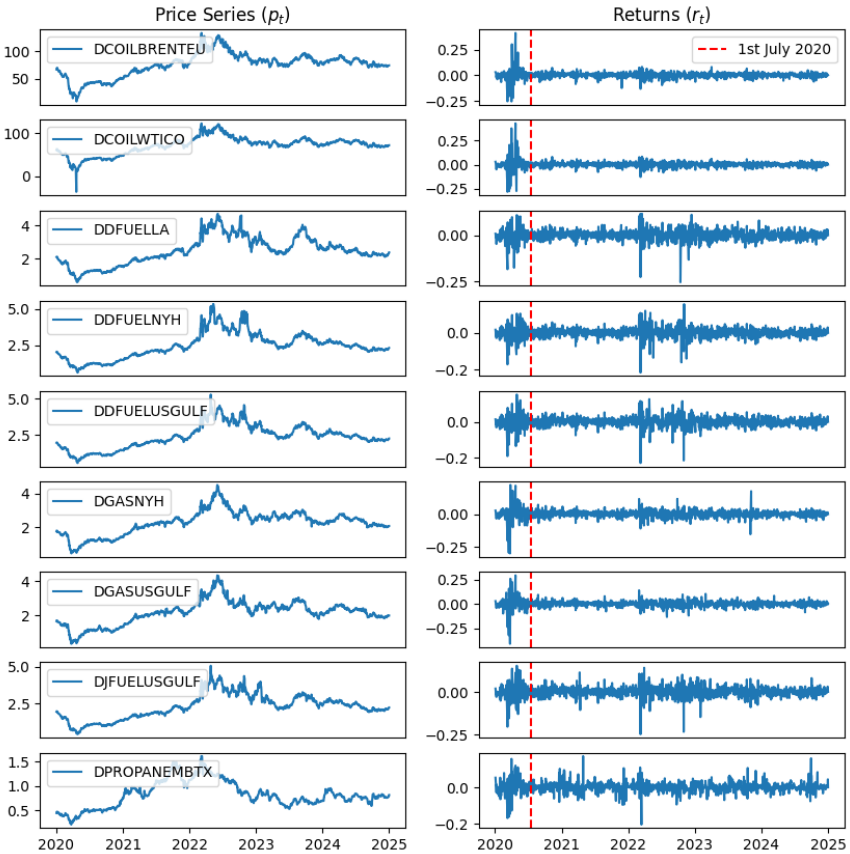

Figure 3: Comparison plot of a price time series vs the log-returns, and the start point of our new trimmed dataset in red dashed lines at 07/01/2020

Every product exhibits a clear cluster of high variance in early 2020, coinciding with the COVID-19-induced oil crisis. During this period, WTI crude prices even turned negative. These extreme market conditions created price dynamics that were fundamentally different from normal market operations. Including this period would distort our analysis of typical crude-distillate price relationships and potentially bias our lag structure estimations. By beginning our analysis from July 2020, we ensure our results reflect the structural relationships between crude and refined products under more representative market conditions.

Once we trim this data out, we perform statistical tests to check for stationarity. Two highly popularised tests are the ADF and the KPSS tests.

1.3.1 Augmented Dickey-Fuller (ADF) Test

The ADF test evaluates the null hypothesis that a unit root is present (non-stationarity) against the alternative of stationarity.

- **Model Specification**:
$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + ... + a_s x_{t-s} + \epsilon_t \tag{2}$$
- **Test Statistic**:
$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \epsilon_t \tag{3}$$
where $\gamma = \rho - 1$ from the AR(1) model
- **Null Hypothesis (H_0)**: The characteristic equation has a root equal to 1
- **Alternative (H_1)**: All roots of characteristic equation are less than 1 in absolute value
- **Interpretation**:
  - Reject H_0 when p-value $< \alpha$ (typically 0.05)
  - If any root > 1, the process is explosive
  - If m roots $= 1$ and others < 1, the m^{th} difference is stationary

1.3.2 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

The KPSS test complements the ADF test by reversing the hypotheses, testing stationarity as the null hypothesis.

- **Model Specification**:
$$y_t = r_t + ct + \epsilon_t \tag{4}$$
where $r_t = r_{t-1} + \gamma_t$ is a random walk with γ_t being i.i.d.
- **Test Statistic**:
$$KPSS = \frac{1}{T^2} \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}^2} \tag{5}$$
where $S_t = \sum_{j=1}^{t} \hat{\epsilon}_j$ is the partial sum of residuals and $\hat{\sigma}^2$ is the long-run variance estimator
- **Null Hypothesis (H_0)**: Series is level/trend stationary ($\gamma_t = 0$)
- **Alternative (H_1)**: Series contains a unit root ($\gamma_t \neq 0$)
- **Interpretation**:
  - Reject H_0 when p-value $< \alpha$ (typically 0.05)
  - Used complementarily with ADF test for robust stationarity analysis
  - High p-values (e.g., >0.05) suggest stationarity

ADF Result	**KPSS Result**	**Conclusion**
Reject H_0	Fail to reject H_0	Stationary
Fail to reject H_0	Reject H_0	Non-stationary

Table 4: Test Guide

1.3.3 Results

Product	ADF p-value	ADF H_0 Rejected (5%)	KPSS p-value	KPSS H_0 Rejected (5%)	**Conclusion**
DCOILBRENTEU	<0.0001	True	>0.1	False	Stationary
DCOILWTICO	<0.0001	True	>0.1	False	Stationary
DDFUELLA	<0.0001	True	>0.1	False	Stationary
DDFUELNYH	<0.0001	True	>0.1	False	Stationary
DDFUELUSGULF	<0.0001	True	>0.1	False	Stationary
DGASNYH	<0.0001	True	>0.1	False	Stationary
DGASUSGULF	<0.0001	True	>0.1	False	Stationary
DJFUELUSGULF	<0.0001	True	>0.1	False	Stationary
DPROPANEMBTX	<0.0001	True	>0.1	False	Stationary

Table 5: KPSS and ADF Test Results, where p-values >0.1 have been clipped and p-values <0.0001 also clipped. Showcasing that all the returns of the crude and distillates are stationary

Ok now our data has been processed, and verified to be stationary. We can move on to the answering the research questions.

2 Weiner-Granger Causality

Figure 3 clearly shows a strong linear correlation between the various price series. Although this context is not directly provided in the data, it is well understood that distillate prices are heavily influenced by the type of crude oil used as input. In the United States, West Texas Intermediate (WTI) is the primary benchmark crude, whereas Brent (from the North Sea) plays a more dominant role in Europe, as well as parts of Asia and Africa.

But linear correlation itself does not provide enough insight. What is really useful to know, is causality. This gives us a sense of direction behind the correlation.

While **true causality** is (as far as we know) statistically impossible to infer through data alone. What we can quantify is **predictive causality**.

This is where **Weiner-Granger Causality** comes into play, detecting the lagged relationships between two time-series for example. Granger himself described the idea as, "temporally related" [4]. Rather than testing whether X causes Y, the Granger causality test evaluates whether X forecasts Y [5].

The foundational idea behind Weiner-Granger causality originated with Wiener, who in 1956 proposed that if the prediction of a time series Y can be improved by incorporating the past values of another series X, then X may be said to cause Y in a predictive sense [7]. However, Wiener's formulation was abstract and grounded in continuous-time stochastic processes, limiting its practical application. Granger later formalised this notion in a testable econometric framework, introducing the now widely-used Granger causality test [3]. By applying Wiener's idea using discrete-time vector autoregressive (VAR) models and hypothesis testing, Granger made predictive causality accessible to test and quantify across a variety of fields.

2.1 Methodology

Before starting, the Granger tests assumes no unit-roots (i.e stationary process). We confirmed this for all our time-series earlier in section 1.3, through ADF and KPSS tests.

The foundation of the tests start by formulating two autoregressive models:

$$Y_{t+h} = \beta_0 + \sum_{s=1}^{n} \beta_{1,s} Y_{t-\lambda_s} + \epsilon_1 \tag{6}$$

$$Y_{t+h} = \beta_2 + \sum_{s=1}^{n} \beta_{3,s} Y_{t-\lambda_s} + \sum_{s=1}^{m} \beta_{4,s} X_{t-\tau_s} + \epsilon_2 \tag{7}$$

The $\lambda' s$ are the lags of the Y time series, and the $\tau' s$ are the lags of the time series, X, that we are attempting to detect the lagged relationship between.

We compare our two regression models, (6) and (7), to assess predictive causality:

- **Equation (6)** (Restricted model): This predicts future values of Y (at horizon h) using only its own lagged values $\{Y_{t-\lambda_s}\}_{s=1}^{n}$, where:
  - β_0: Intercept term
  - $\beta_{1,s}$: Coefficients for lagged Y
  - ϵ_1: IID error term
- **Equation (7)** (Unrestricted model): This extends Equation (6) by incorporating lagged values of X ($\{X_{t-\tau_s}\}_{s=1}^{m}$), where:
  - $\beta_{3,s}$: Coefficients for lagged Y (analogous to $\beta_{1,s}$)
  - $\beta_{4,s}$: Coefficients for lagged X
  - ϵ_2: IID error term

(both regressions are fit using ordinary least squares (OLS) method)

The key comparison tests whether the inclusion of X's history (Model 7) *significantly improves* the prediction of Y over Model (6). This is the essence of Granger causality testing.

The "goodness" of the regressions can be quantified and compared using the residual error terms ($\epsilon_{1\&2}$).

More specifically, the variance of the errors. We can quantify the improvement between the predictions of two models by the following,

$$F_{X \to Y} = \ln(\frac{\sigma_{\epsilon 1}^2}{\sigma_{\epsilon 2}^2}) \tag{8}$$

where,

- $\sigma_{\epsilon_i}^2 = Var(\epsilon_i)$ is the variance of errors ϵ_i which can be defined as $\epsilon_i = Y_{t+h}^{true} - \hat{Y_{t+h}}$

We can interpret the results of $F_{X \to Y}$ according to the following logic:

Value Range	Interpretation	Implied Conclusion
$F_{X\to Y} > 0$	Model with X has smaller prediction errors ($\sigma^2_{\epsilon_2} < \sigma^2_{\epsilon_1}$)	X improves Y's predictions
$F_{X\to Y} = 0$	No reduction in error variance ($\sigma^2_{\epsilon_1} = \sigma^2_{\epsilon_2}$)	X provides no predictive value
$F_{X\to Y} < 0$	Model with X performs worse ($\sigma^2_{\epsilon_2} > \sigma^2_{\epsilon_1}$)	X adds noise or misspecification

Table 6: Interpretation of Variance Ratio Metric $F_{X\to Y} = \ln(\frac{\sigma^2_{\epsilon_1}}{\sigma^2_{\epsilon_2}})$

Rather than employing traditional hypothesis testing approaches that yield binary conclusions about causality, our methodology quantifies the predictive improvement through a direct comparison of model performances. The metric $F_{X\to Y}$ provides a continuous measure of predictive power:

$$F_{X\to Y} = \ln(\frac{\sigma^2_{\epsilon 1}}{\sigma^2_{\epsilon 2}}) \tag{9}$$

This ratio of error variances offers three interpretable outcomes described in Table 6.

This approach is particularly appropriate for our analysis of crude oil and distillate prices, where our interest lies not in establishing the existence of a relationship, but in quantifying the strength and nature of price transmission mechanisms across different time horizons.

To determine the optimal lag lengths ($\lambda' s$ and $\tau' s$), we employ information criteria that balance model fit against complexity. The Akaike Information Criterion (AIC), developed by Hirotugu Akaike[1], and the Bayesian Information Criterion (BIC), proposed by Gideon Schwarz[6], are two fundamental approaches:

$$\text{AIC} = 2n - 2\ln(\sup P(\hat{x}|M,\theta)) \tag{10}$$

$$\text{BIC} = n\ln q - 2\ln(\sup P(\hat{x}|M,\theta)) \tag{11}$$

where n is the number of parameters, q is the sample size, and $P(\hat{x}|M,\theta)$ is the likelihood function.

These criteria differ primarily in their penalty terms:

- **AIC** uses $(2n)$: Favors more complex models, capturing subtle dynamics
- **BIC** uses $(n\ln q)$: Favors parsimony, focusing on stronger relationships

For our analysis of crude oil and distillate price relationships, we employ AIC due to the complex nature of energy market dynamics. While the causal relationship between these prices is established, the transmission mechanisms involve multiple factors:

- Price adjustment delays across market conditions
- Storage and transportation lags
- Contract settlement periods
- Refinery adjustment times

AIC's tolerance for longer lag structures aligns with our objective of uncovering nuanced price transmission dynamics, including seasonal patterns and market friction effects, rather than merely establishing causality.

2.2 Empirical Results

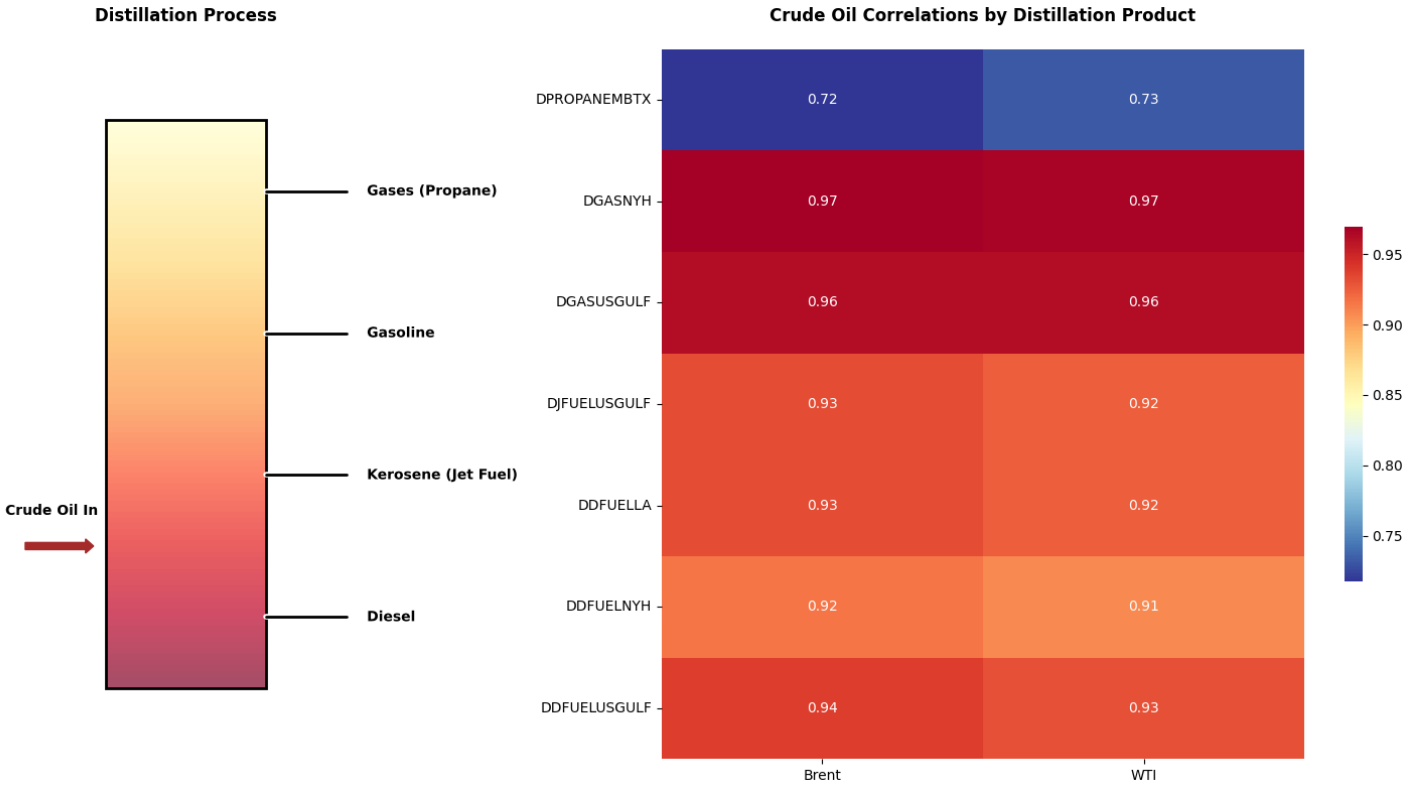

Figure 4: Relationship between crude oil distillation process and price correlations. The left panel illustrates the fractional distillation tower, where crude oil is heated and separated into different petroleum products based on their boiling points, with lighter products rising to the top. The right panel shows the correlations between crude oil benchmarks (Brent and WTI) and their refined products, ordered by their respective distillation fractions. The correlation heatmap reveals that despite the physical production sequence, price relationships vary across products and regions, suggesting that market factors beyond the production process influence price formation.

The correlation analysis reveals several key patterns in the crude-product price relationships. First, gasoline prices show the strongest correlations with both crude benchmarks (0.96-0.97), suggesting particularly tight price integration. Interestingly, propane exhibits the weakest correlations (0.72-0.73) despite being the lightest fraction, indicating that its prices may be influenced by factors beyond crude oil, such as natural gas markets and seasonal heating demand.

Diesel and jet fuel correlations (0.91-0.94) fall between these extremes, with minimal variation across regions. The similarity between Brent and WTI correlations across all products (typically differing by only 0.01) suggests that global crude price movements, rather than regional benchmark differences, dominate long-term price relationships.

However, these strong correlations only tell part of the story. While they confirm the fundamental price relationships inherent in the production process, they don't capture the dynamic aspects of price formation or short-term adjustments. Our subsequent Granger causality analysis will examine these nuanced temporal relationships, potentially revealing differences in how quickly and strongly different products respond to crude price changes across regions.

2.2.1 Diesel

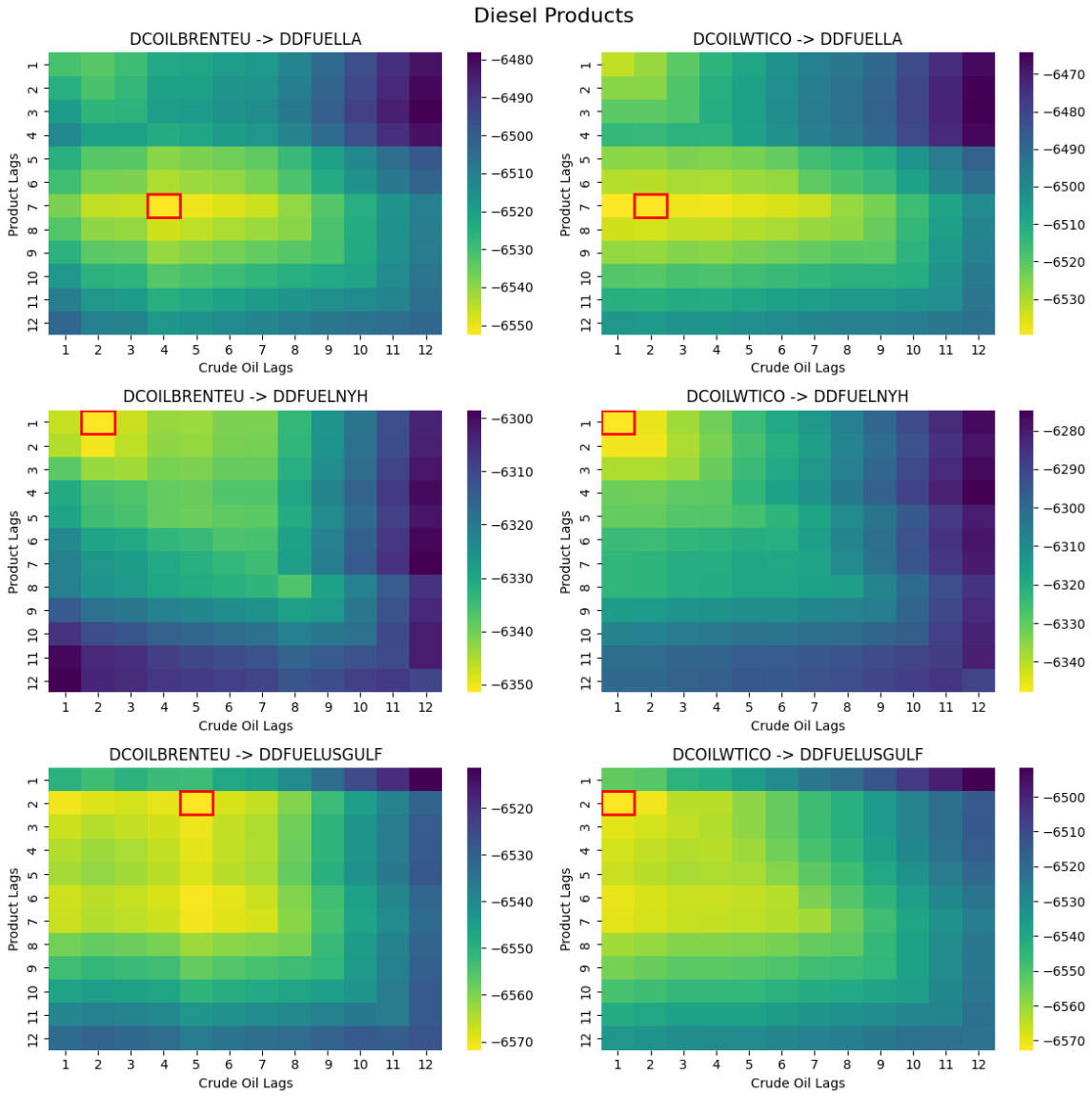

Figure 5: AIC grid-search across lags for Diesel Products in LA, NY and Gulf Coast. x-axis (Crude Oil Lags) & y-axis (Distillate Product Lags). Lighter Yellow signifies minimal AIC, darker Blue signifies maximal AIC. Red outlined square is optimal X&Y lag for minimising AIC

The heatmaps reveal optimal lag structures that vary both by region and crude benchmark. For diesel prices, the AIC surfaces suggest more complex dynamics compared to other products, particularly for LA diesel which requires longer lag structures (7,4) for Brent and (7,2) for WTI. This indicates a longer price discovery process in the West Coast market. The quantitative results confirm Brent crude's stronger predictive power across all diesel markets, with variance reductions ranging from 1.39% to 2.56%, compared to WTI's 0.54% to 0.73%. The US Gulf diesel market shows the strongest Brent influence (F-stat = 0.03, variance reduction 2.56%) despite being geographically closer to WTI production, while requiring fewer own-price lags (2,5) than LA. This suggests that despite WTI's physical proximity, Brent crude remains the more influential benchmark for US diesel price formation. However, the high R-squared values in the restricted models (all > 0.99) indicate that diesel prices are primarily driven by their own recent history, with crude prices providing marginal but statistically meaningful improvements in predictive power.

Table 7: Comparison of Crude Oil Impact on Diesel Prices by Region

Region	Crude	Optimal (Y,X)	Variance Reduction	F-stat	R-squared With	R-squared Without
LA	Brent	(7,4)	2.02%	0.02	0.9911	0.9909
	WTI	(7,2)	0.73%	0.01	0.9910	0.9909
NY Harbor	Brent	(1,2)	1.39%	0.01	0.9907	0.9906
	WTI	(1,1)	0.54%	0.01	0.9906	0.9906
US Gulf	Brent	(2,5)	2.56%	0.03	0.9911	0.9909
	WTI	(2,1)	0.71%	0.01	0.9909	0.9909

2.2.2 Gasoline

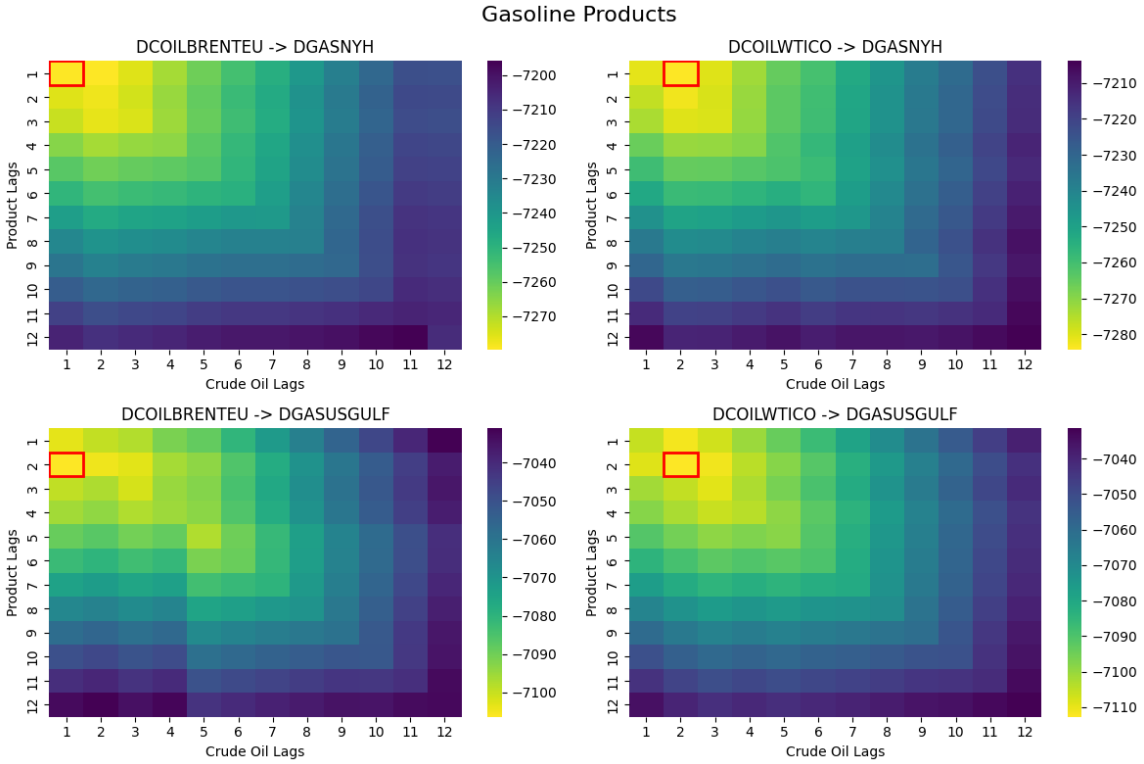

Figure 6: AIC grid-search across lags for Gasoline Products in NY, and Gulf Coast. x-axis (Crude Oil Lags) & y-axis (Distillate Product Lags). Lighter Yellow signifies minimal AIC, darker Blue signifies maximal AIC. Red outlined square is optimal X&Y lag for minimising AIC

The gasoline heatmaps demonstrate notably simpler lag structures compared to diesel, with optimal combinations rarely exceeding two lags for either crude benchmark. This suggests a more efficient and rapid price transmission mechanism in gasoline markets. Interestingly, while diesel showed stronger Brent influence, gasoline exhibits stronger WTI causality across regions. The NY Harbor gasoline market shows the most pronounced WTI influence (variance reduction 0.99%, F-stat = 0.01) compared to minimal Brent impact (0.02%, F-stat = 0.00), despite its geographical proximity to Atlantic basin crude imports. Similarly, US Gulf gasoline shows stronger WTI predictive power (0.69% variance reduction) versus Brent (0.05%). The consistently short lag structures (1-2 lags) and high restricted model R-squared values (>0.991) indicate that gasoline prices are highly efficient at incorporating market information, with crude prices providing

only modest improvements in predictive power. This efficiency might reflect gasoline's higher trading volumes and more standardized specifications compared to diesel markets.

Table 8: Comparison of Crude Oil Impact on Gasoline Prices by Region

Region	Crude	Optimal (Y,X)	Variance Reduction	F-stat	R-squared With	R-squared Without
NY Harbor	Brent	(1,1)	0.02%	0.00	0.9925	0.9925
	WTI	(1,2)	0.99%	0.01	0.9926	0.9925
US Gulf	Brent	(2,1)	0.05%	0.00	0.9912	0.9912
	WTI	(2,2)	0.69%	0.01	0.9912	0.9912

2.2.3 Jet Fuel & Propane

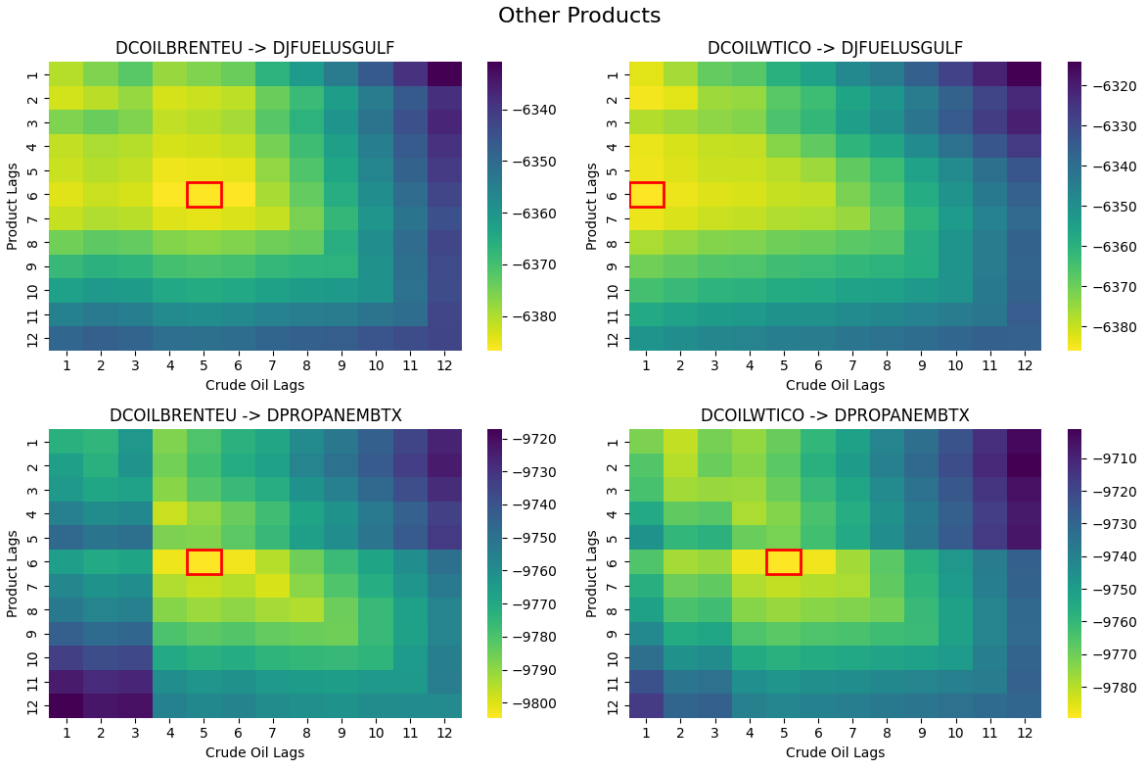

Figure 7: AIC grid-search across lags for Jet Fuel and Propane Products in NY, and Gulf Coast. x-axis (Crude Oil Lags) & y-axis (Distillate Product Lags). Lighter Yellow signifies minimal AIC, darker Blue signifies maximal AIC. Red outlines square is optimal X&Y lag for minimising AIC

Propane and jet fuel exhibit distinct patterns from both diesel and gasoline markets. Propane shows the strongest crude oil influence among all products studied, with Brent crude delivering the highest variance reduction (3.64%, F-stat = 0.04) and WTI following closely (2.51%, F-stat = 0.03). The heatmaps reveal complex lag structures for both products, with propane consistently requiring longer lag combinations (6,5) for both crude benchmarks, suggesting a more intricate price discovery process. Jet fuel demonstrates an interesting asymmetric response: while it requires multiple lags of its own price (6), it shows different crude lag requirements between Brent (5 lags) and WTI (1 lag). This asymmetry, combined with stronger Brent influence (1.29% variance reduction vs WTI's 0.63%), might reflect jet fuel's global trading nature and its closer relationship with international crude benchmarks. Despite these stronger crude relationships,

both products maintain the characteristic high restricted model R-squared values (> 0.99), confirming the dominant role of own-price history in short-term price formation.

Table 9: Comparison of Crude Oil Impact on Jet Fuel and Propane

Product	Crude	Optimal (Y,X)	Variance Reduction	F-stat	R-squared With	R-squared Without
Jet Fuel	Brent	(6,5)	1.29%	0.01	0.9901	0.9900
	WTI	(6,1)	0.63%	0.01	0.9900	0.9900
Propane	Brent	(6,5)	3.64%	0.04	0.9932	0.9929
	WTI	(6,5)	2.51%	0.03	0.9931	0.9929

3 Future Directions and Open Challenges

This analysis reveals several key areas for future research and highlights specific challenges that need to be addressed.

3.1 Statistical Limitations

Due to the high linear correlation between crude and distillate prices, the R^2 values from the restricted models were already significantly high (>0.99). This limits our ability to measure short-term causalities from crude to distillates, as reflected in the relatively small F-statistics. However, by comparing these metrics across products and regions, we can still gain meaningful insights into the relative strengths and patterns of price transmission. Future research might explore alternative methodologies that can better isolate and quantify these short-term relationships despite the strong underlying correlations.

3.2 Seasonal Data Analysis

Limited by sample size, extending to a longer period of 10-15 years would allow for more robust analysis of seasonal patterns, especially for products like propane which show strong seasonal demand characteristics. A larger dataset would help separate actual seasonal effects from structural market changes, and might show how price transmission between crude and refined products has evolves throughout the seasons like Summer, Winter and transitional periods.

3.3 Trading Applications

Our findings on lag structures point to potential trading opportunities, particularly in products showing longer or more complex price transmission patterns. The lag structures we found for propane (6,5) and the different responses of US Gulf jet fuel to Brent (6,5) versus WTI (6,1) could be used to develop systematic trading strategies. Future work could focus on testing these strategies while accounting for real market factors like transaction costs and liquidity. We could also explore pairs trading between products that show different sensitivities to crude oil prices, and test if these strategies work differently in various market conditions.

3.4 Market Fundamentals

Getting access to regional supply and demand data would help us better understand price mechanics. How local demand changes flow through the product supply chain, would give us better insight into how these

markets actually work. This deeper look at market fundamentals would add to our statistical findings and might explain why we see different lag structures across regions and products.

Conclusion

Our research questions have been answered through quantitative analysis of price transmission dynamics between crude and refined products. We conclude that crude oil does Granger-cause short-term price changes in refined products, though this relationship varies significantly by product and region. The strongest and most efficient price transmission occurs in gasoline markets, with simple lag structures of 1-2 days, while propane and jet fuel exhibit more complex dynamics requiring up to 6 lags. Regional heterogeneity is most pronounced in diesel markets, where LA diesel demonstrates notably longer lag structures (7,4) compared to NY Harbor (1,2), indicating significant geographical differences in price discovery processes.

The analysis revealed differential predictive capacities between crude benchmarks: Brent crude exhibited superior explanatory power for the diesel and jet fuel markets, producing variance reductions in Diesel of up to 2.56% compared to the average influence of WTI of 0.66%, while WTI demonstrated a stronger predictive influence on gasoline prices with variance reductions of 0.99% compared to Brent's reductions of 0.02 & 0.05% which is essentially no added predictive power. These findings highlight the distinct roles of international and domestic crude benchmarks in the formation of refined product prices in the US.

References

[1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[2] Federal Reserve Bank of St. Louis. Commodities - daily prices, 2024.

[3] Clive W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[4] Clive W. J. Granger and Paul Newbold. *Forecasting Economic Time Series*. Academic Press, New York, 1977.

[5] James D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994. Available online (PDF).

[6] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[7] Norbert Wiener. The theory of prediction. In E. F. Beckenbach, editor, *Modern Mathematics for Engineers*, pages 165–190. McGraw-Hill, New York, 1956.