

Working Papers

One Great Pool or Many? Measuring Global Oil Market Integration Using a Dynamic Time Warping-Hierarchical Cluster Approach

Marc Gronwald, Luyao Zhu

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
- from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

One Great Pool or Many? Measuring Global Oil Market Integration using a Dynamic Time Warping-Hierarchical Cluster Approach

Marc Gronwald*

Luyao Zhu[†]

May 2026

Abstract

This paper measures global crude oil market integration using dynamic time warping combined with hierarchical cluster analysis. Using ten global crude oil prices (2000-2026), we find that markets cluster primarily by geography: West African, Middle Eastern, and U.S. markets each form distinct regional clusters: Brent occupies an intermediate position. A rolling window analysis reveals an increase in integration over time, but a remarkably stable regional structure. A subsample analysis shows that during the turbulent 2020-2024 period, integration increases but the clear geographical pattern weakens. Validation metrics such as silhouette width and cophenetic correlation confirm robust clustering across all periods.

Keywords: Oil market integration, hierarchical clustering, dynamic time warping, global crude oil markets, one great pool

JEL-Classification: Q43, C38, F15, C32

*International Business School Suzhou, Xi'an Jiaotong-Liverpool University; CESifo and ifo Institute. Email: marc.gronwald@xjtlu.edu.cn.

[†]International Business School Suzhou, Xi'an Jiaotong-Liverpool University

1 INTRODUCTION

This paper measures global crude oil market integration using dynamic time warping (DTW) in combination with hierarchical cluster analysis. The question of whether the world oil market constitutes “one great pool” dates back to Adelman (1984). This classic formulation captures a central tension in the literature: is the global crude oil market truly globalised, with prices showing high degree of co-movement across regions, or does it remain fragmented into regional markets shaped by local infrastructure, policy, and transportation costs? For decades, this question has motivated a large body of empirical research which seeks to characterise the degree and evolution of oil market integration. For a long time, the dominant approach in the literature has been the pairwise analysis of crude oil price series. Papers such as Plante and Strickler (2021) and Bravo Caro et al. (2020) epitomise this tradition. Plante and Strickler (2021) examine structural breaks in the differentials between various crude oil price benchmarks; they interpret declining means and volatilities of these differentials over time as evidence of increasing integration. Bravo Caro et al. (2020), in addition, focus on the relationship between Brent and WTI prices by employing methods that assess long-run linkages between the two series. While these pairwise approaches have generated valuable insights, they are inherently limited by their bivariate nature: they consider markets two at a time, leaving open the question of how integration patterns generalise across the full set of global crude oil markets.

A methodological advance was proposed by Gronwald and Jin (2024), who introduced a more sophisticated approach that, while still pairwise, yields visually interpretable results. The Thick Pen method they apply allows for a more nuanced characterisation of how price relationships evolve over time, which offers a clearer picture of integration dynamics than traditional binary classifications. More recently, a number of studies have moved beyond pairwise frameworks by employing multivariate models that analyse multiple markets simultaneously. Examples include the work of Cui and Maghyreh (2023), Chatziantoniou et al. (2023), and others, who fit all markets into a rich VAR structure. These multivariate approaches hold the promise of capturing the full interconnectedness of global oil markets in a unified framework by recognising that integration is not merely a collection of bilateral relationships but a system-wide phenomenon. Despite this shift toward multivariate analysis, pairwise approaches remain popular. Recent contributions such as Shi et al. (2025) and Wu and Han (2025) continue to rely on pairwise frameworks, which underscores that the transition to fully multivariate methods is certainly not yet complete.

This study offers a distinct path forward. Rather than relying on pairwise tests or conventional multivariate time-series models, we propose a novel approach that combines hierarchical cluster analysis with dynamic time warping. This methodology allows us to capture similarities in price movements across a large number of crude oil markets without imposing restrictive parametric assumptions. By grouping markets according to the shape and timing

of their price dynamics, we are able to assess the structure of global oil market integration in a way that is both data-driven and visually interpretable. In doing so, we move beyond the binary logic of cointegration and the limitations of pairwise comparisons. We offer a more comprehensive picture of how regional markets align with- or diverge from - the notion of a single, integrated global pool. While Mastroeni et al. (2021) apply DTW to study the decoupling between WTI and Brent, the present study generalizes this approach to a system-wide analysis of global crude oil markets.

The three key findings that emerge from our analysis can be summarised as follows. First, global crude oil markets cluster primarily by geography: West African, Middle Eastern, and U.S. markets form distinct regional groups. Second, the rolling window analysis reveals an increase in integration over time, but also a remarkably stable regional structure. Third, during turbulent periods (2020-2024), integration increases, but the clear geographical pattern weakens, giving way to more fluid and idiosyncratic relationships.

Our approach is essentially about allowing the data to speak - in the tradition of the “empirics first” approach, consistent with recent developments in the marketing literature; see Golder et al. (2023). This perspective is not new to economics; it aligns with established traditions in reduced-form econometrics and exploratory data analysis (EDA). Such approaches are likely to become increasingly common in economics, driven by the ongoing data revolution.

The remainder of the paper is organised as follows. Section 2 presents and

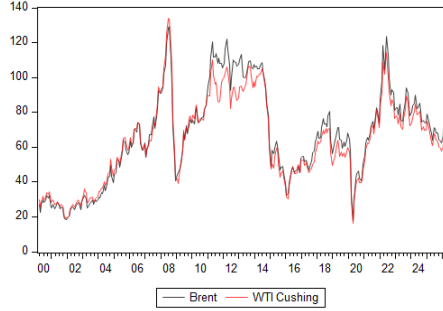
discusses the data used in this study. Section 3 outlines the methodological framework, followed by the presentation of the empirical results in Section 4. Section 5 concludes with some final remarks.

2 DATA

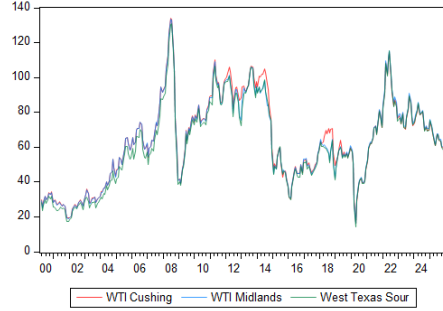
This paper uses monthly crude oil price data from January 2000 to March 2026. The sample includes ten regional crude oil markets: Algerian Saharan (Algeria), Bonny Light (Nigeria), Brent (UK), Dubai (UAE), Oman (Oman), Saudi Heavy (Saudi Arabia), Tapis (Malaysia), West Texas Sour (USA), WTI Cushing (USA), WTI Midland (USA). All prices are nominal spot prices expressed in US dollars per barrel. The data are sourced from Bloomberg and DataStream. The selection of markets covers the world’s major crude oil producing regions: North America (three U.S. grades), the Middle East (Dubai, Oman, Saudi Heavy), Europe (Brent), West Africa (Bonny Light, Algerian Saharan), and Southeast Asia (Tapis). This geographic diversity allows for a comprehensive assessment of global oil market integration. The sample period includes several notable episodes of oil market stress: the price surge preceding the 2008 Global Financial Crisis, the Brent-WTI spread period from 2010 to 2014, the price collapse of 2014-2015, the Covid period 2020-2022, the price surge following Russia’s invasion of Ukraine in 2022, and the market developments through early 2026.

The data is displayed in Figure 1. Figure 1a shows the most famous pair, Brent and WTI Cushing. The most apparent feature is the well-known spread

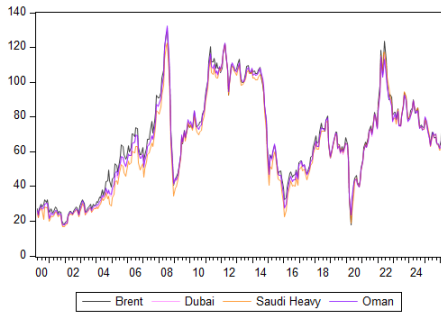
Figure 1: One great pool



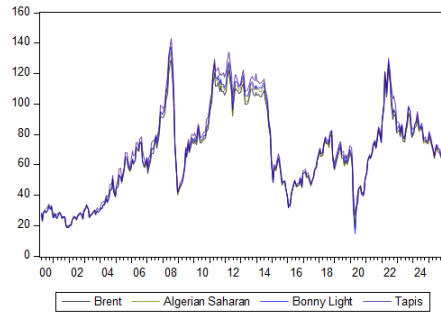
(a) Brent and WTI Cushing



(b) US American markets



(c) Middle East



(d) Rest of the World

observed between 2010 and 2014. However, a closer look reveals additional patterns. Prior to 2010, the two global benchmarks actually moved closer together than they did after 2016. Apparent divergences also appear around 2018 and again after 2022. Figure 1b shows the three U.S. markets. The pattern here is almost the reverse. Some divergence is visible prior to 2008, but afterward the three series move closer together overall - except for 2014 and 2018, when WTI Cushing temporarily moved away from WTI Midland and West Texas Sour. Figure 1c presents the Middle Eastern price series

alongside Brent. Brent has been chosen as the benchmark because the three Middle Eastern series follow it closely during the 2010-2014 period. Overall, these prices move together very strongly, with only slight divergence apparent prior to 2008. Figure 1d shows Brent alongside Tapis and the two West African series (Bonny Light and Algerian Saharan). The pattern is broadly similar to the Middle Eastern case, though there is somewhat more divergence during the 2010-2014 period.

3 METHODOLOGY

3.1 DATA PREPROCESSING AND NORMALIZATION

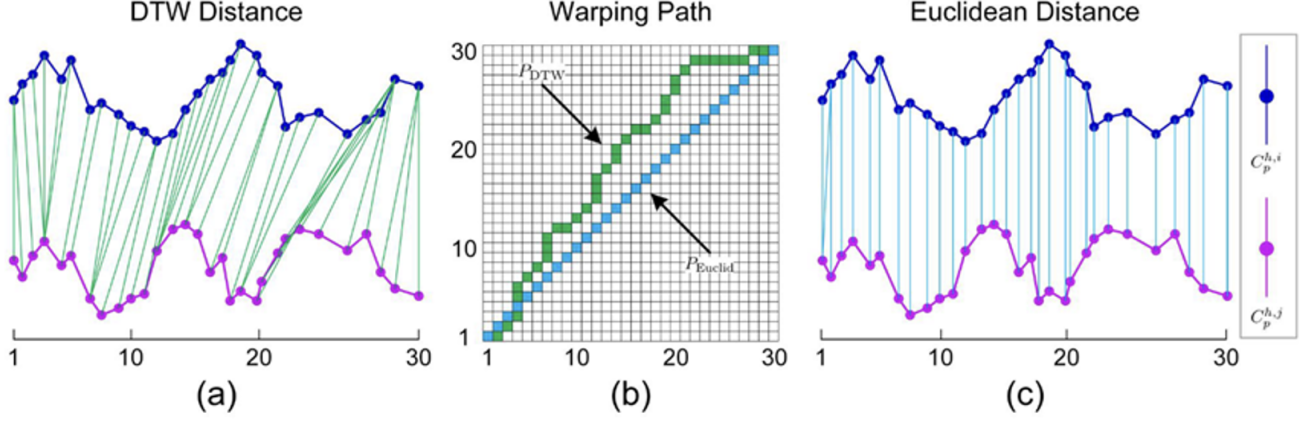
Before implementing DTW, all time series are z-normalized to ensure that similarity measures capture structural features rather than differences in absolute price levels; see, e.g., Lima and Souza (2023). Each series is transformed to have a mean of zero and a standard deviation of one. For any reference series X and query series Y , the normalized values are calculated as:

$$\tilde{x}_i = \frac{x_i - \mu_X}{\sigma_X}, \quad \tilde{y}_j = \frac{y_j - \mu_Y}{\sigma_Y}$$

where μ_X and σ_X are the mean and standard deviation of X , and similarly for Y . After normalization, the algorithm focuses on structural dynamics rather than absolute price levels.

3.2 FIND THE OPTIMAL WARPING PATH (P)

Figure 2: Optimal Warping Paths.



After preprocessing, we find the optimal warping path. Unlike Euclidean distance, which requires a strict one-to-one correspondence between observations at the same time indices, DTW allows for local stretching and compression along the time axis. This is illustrated in Figure 2, Panel (a); taken from Li (2019). Panel (c) shows how Euclidean distance forces a rigid alignment, while DTW flexibly matches points to achieve a more intuitive similarity measure. Panel (b) illustrates differences in the warping paths obtained from applying these two methods.

We first create an $N \times M$ cost matrix, where each entry represents the local distance between observations. The point-by-point Euclidean distance is given by:

$$c_{ij} = \|\tilde{x}_i - \tilde{y}_j\|^2$$

We then extract the optimal alignment pathway between the two time series by minimizing the cumulative distance. This allows the sequences to be stretched and compressed, as mentioned above. The cumulative distance is computed recursively as:

$$\gamma(i, j) = c_{ij} + \min\{\gamma(i-1, j), \gamma(i, j-1), \gamma(i-1, j-1)\}$$

The optimal warping path $\mathbf{P} = (p_1, p_2, \dots, p_K)$ is then plotted, where each point $p_k = (i_k, j_k)$ represents the alignment between the two time series. The result is shown as the green line in Figure 2, Panel (b).

To ensure that the warping path remains economically interpretable, it must satisfy three constraints:

- **Boundary conditions:** The path starts at the initial indices of the two sequences and ends at the last indices. Mathematically, $p_1 = (1, 1)$ and $p_K = (N, M)$.
- **Monotonicity:** The indices of the path must rise monotonically, preserving the temporal order of the sequences: $i_{k+1} \geq i_k$ and $j_{k+1} \geq j_k$.
- **Continuity:** The path moves only one step at a time, allowing i and j to increase by at most 1: $i_{k+1} - i_k \leq 1$ and $j_{k+1} - j_k \leq 1$.

3.3 ALIGNMENT INDICES

To quantify the degree of pathway deviation in DTW, we employ two indices: the Warping Index (WI) and Warping Path Distance (WPD). WI measures overall distortion between two time series, while WPD captures local deviations within a particular alignment window.

3.3.1 WARPING INDEX (WI)

Introduced by Mastroeni et al. (2021), the Warping Index measures the area between the optimal warping path and the 45-degree diagonal. Before computation, the Sakoe-Chiba constraint limits the amount of warping deviation: $|i_k - j_k| \leq r$, where r is the window size (Sakoe and Chiba, 1978). The Warping Index is defined as:

$$WI = \frac{1}{K} \sum_{k=1}^K |i_k - j_k|$$

A WI distance matrix $\mathbf{D}_{WI}$ summarizes the long-term alignment pattern across all markets:

$$\mathbf{D}_{WI} = [d_{pq}], \quad d_{pq} = WI(X_p, X_q)$$

3.3.2 WARPING PATH DISTANCE (WPD)

To further investigate patterns of market integration over time, we introduce the Warping Path Distance:

$$WPD = \frac{1}{K} \sum_{k=1}^K |i_k - j_k|$$

A lower WPD indicates stronger integration, while a higher WPD indicates weaker co-movement.

3.4 HIERARCHICAL CLUSTERING

To move beyond pairwise comparisons and assess the system-wide structure of global crude oil markets, we apply hierarchical clustering to the distance matrix derived from the WPD measures. Hierarchical clustering is an agglomerative method that builds a hierarchy of clusters by successively merging the most similar markets.

The algorithm proceeds as follows. Initially, each market forms its own cluster. At each step, the two closest clusters are merged based on a linkage criterion. We use Ward’s minimum variance criterion (Ward.D2), which merges the pair of clusters that minimizes the increase in total within-cluster variance. This criterion is particularly suitable for use with DTW-based distance metrics because it respects the non-linear nature of the underlying similarity measures.

The result is a dendrogram, a tree-like diagram that records the sequence of cluster mergers. The vertical axis represents the dissimilarity at which clusters merge, allowing us to read the grouping structure at any chosen level of aggregation. To determine the quality of the clustering solution, we

compute two validation metrics.

The cophenetic correlation coefficient measures how faithfully the dendrogram preserves the original pairwise distances:

$$\rho = \frac{\sum_{i < j} (d_{ij} - \bar{d})(t_{ij} - \bar{t})}{\sqrt{\sum_{i < j} (d_{ij} - \bar{d})^2 \sum_{i < j} (t_{ij} - \bar{t})^2}}$$

where d_{ij} is the original distance between markets i and j , and t_{ij} is the cophenetic distance the height at which the two markets are first joined in the dendrogram. Values of ρ close to 1 indicate that the dendrogram is an excellent representation of the original distance structure.

The silhouette width measures both within-cluster cohesion and between-cluster separation. For a given market i assigned to cluster C , define:

$$a_i = \frac{1}{|C| - 1} \sum_{j \in C, j \neq i} d_{ij}, \quad b_i = \min_{C' \neq C} \frac{1}{|C'|} \sum_{j \in C'} d_{ij}$$

where a_i is the average distance to other markets in the same cluster, and b_i is the minimum average distance to markets in any other cluster. The silhouette width for market i is then:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

The average silhouette width across all markets provides an overall assessment of cluster quality. Values above 0.5 indicate reasonably well-defined clusters, while values above 0.7 indicate strong clustering structure.

3.5 ROLLING WINDOW CLUSTERING

Market integration is not static. To capture its evolution over time, we implement a rolling window framework. A fixed window size of $m = 60$ months (five years) with a step size of $s = 1$ month is used. For each window w , we recompute the pairwise WPDs and construct the corresponding dissimilarity matrix $\mathbf{D}_w$ of size $m \times m$, where m is the number of markets.

For each distance matrix $\mathbf{D}_w$, we apply agglomerative hierarchical clustering using Ward's criterion as described above. With the number of clusters set to $k = 2$ (and later $k = 3$), we obtain a sequence of membership vectors $\mathbf{C}_w$.

To evaluate temporal stability, we compute the Adjusted Rand Index (ARI) between cluster assignments in consecutive windows. The ARI measures agreement between two clusterings while correcting for chance:

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}$$

where n_{ij} is the number of markets jointly assigned to the same cluster in adjacent windows. An ARI value close to 1 indicates high structural persistence, while a significant decline signals a potential structural break.

3.6 SUBSAMPLE ANALYSIS

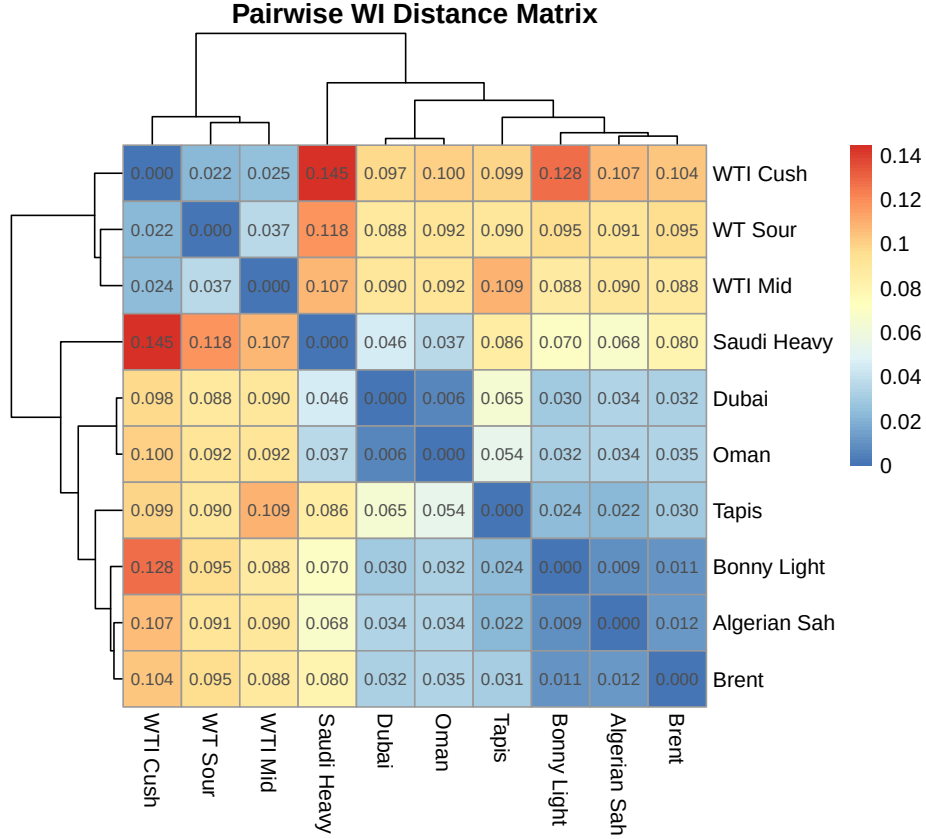
In addition to the rolling window approach, we conduct a subsample analysis based on well-known idiosyncratic episodes in the crude oil market. The

sample is divided into three periods of equal length: 2005-2009 (pre-GFC), 2010-2014 (Brent-WTI spread), and 2020-2024 (COVID and Ukraine war). For each subsample, we compute heatmaps, dendrograms, and the validation metrics (ρ and average silhouette width) to compare integration patterns across distinct market environments.

4 RESULTS

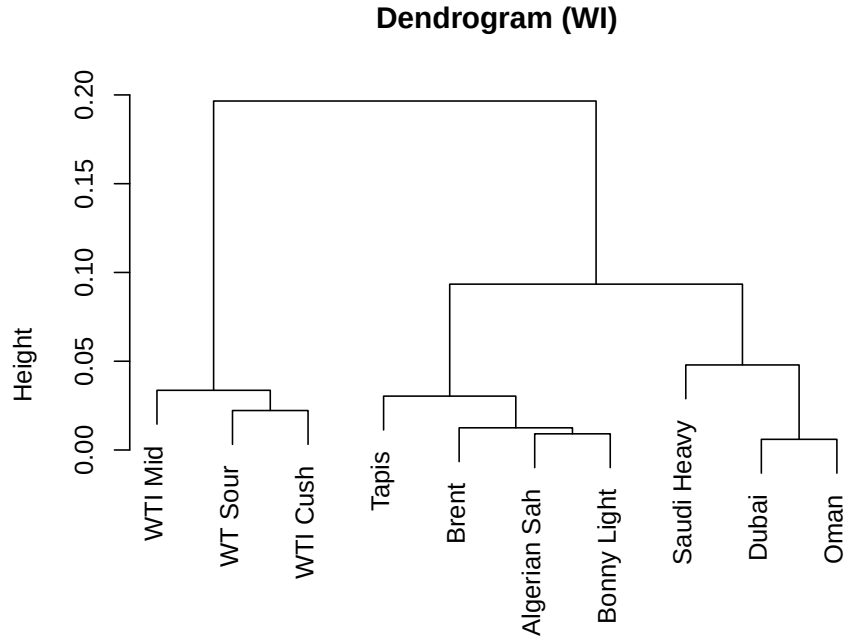
The presentation of results begins with the pairwise Warping Path Distance (WPD) matrix, shown in Figure 3. As explained above, we quantify similarity based on the synchronization of price movements across regional oil markets, where lower WPD values reflect a larger similarity, stronger co-movement, and thus, greater market integration. The WPD represent a quantitative attempt to capture price similarity, which in this context is interpreted as greater market integration. WPD measures synchronization by aligning temporal mismatches - such as lags or differing response speeds - between price series. In the matrix, larger index values indicate greater dissimilarity between two price series. Correspondingly, in the heatmap, blue represents high similarity (low dissimilarity), orange represents moderate similarity, and red represents high dissimilarity. The results are broadly consistent with expectations, yet they reveal several interesting details. A clear geographic pattern emerges from the matrix: the highest price similarity (blue) has been found among the West African markets (Algerian Saharan and Bonny Light), followed by the Middle Eastern markets (Dubai, Oman, and Saudi

Figure 3: Heatmap.



Heavy) and the U.S. markets (WTI Cushing, WTI Midland, and West Texas Sour). Worth noting is that both the African and the Middle Eastern markets exhibit stronger price similarity to one another than the U.S. markets do among themselves. UK Brent, the second global benchmark alongside WTI Cushing, sits in an intermediate position. It shows considerable price similarity to many other regional markets, including those in West Africa and the Middle East, as well as Tapis and Saudi Heavy, but is considerably less similar to the U.S. markets. This latter pattern is largely attributable to

Figure 4: Dendrogram.



the well-documented price spread between Brent and WTI observed between 2010 and 2014, which will be examined in greater detail later. The lowest price similarity (red) occurs between the three U.S. series on one side and Tapis and Saudi Heavy on the other.

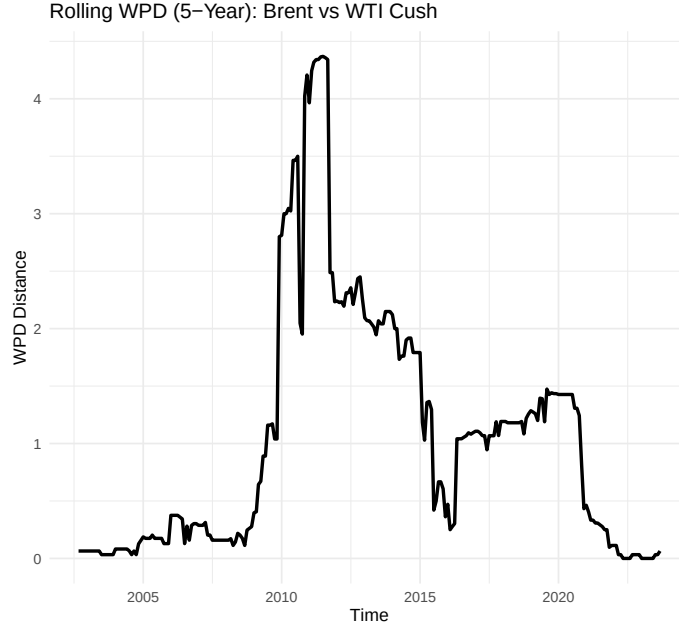
This similarity matrix is converted into a dendrogram, which reveals the hierarchical clustering structure among regional oil markets. The vertical axis represents the levels of dissimilarity at which clusters merge. In other words, it illustrates the progressive aggregation of markets into distinct regional

groupings based on similarities in price behavior. The resulting tree structure delineates clear primary clusters corresponding to major geographic regions, while subsequent branching within these clusters captures more nuanced subgroupings that reflect varying degrees of integration and interconnectedness among individual markets.

This structure largely reflects the patterns observed in the similarity matrix. Moving up along the dissimilarity axis, Dubai and Oman form the first cluster, followed by Saudi Heavy. A separate cluster emerges which consists of Brent, Tapis, and the two West African series (Bonny Light and Algerian Saharan). On the opposite side, the three U.S. markets form a distinct cluster. Two findings stand out. First, Brent remains separate from WTI Cushing until a much higher level of dissimilarity before eventually merging with its U.S. counterpart. Second, the Middle Eastern markets remain isolated; they only join the cluster containing one of the global benchmarks at a substantially higher level of dissimilarity.

To evaluate how well the hierarchical clustering captures the true relational patterns among regional oil markets, we use the cophenetic correlation coefficient, which measures the faithfulness with which the dendrogram preserves the original pairwise distances. The resulting value of 0.92 indicates a very strong correspondence, which confirms that the clustering solution reliably reflects the underlying market structures and that the groupings identified are robust representations of actual similarities rather than artifacts of the clustering procedure. To further assess the quality of the clus-

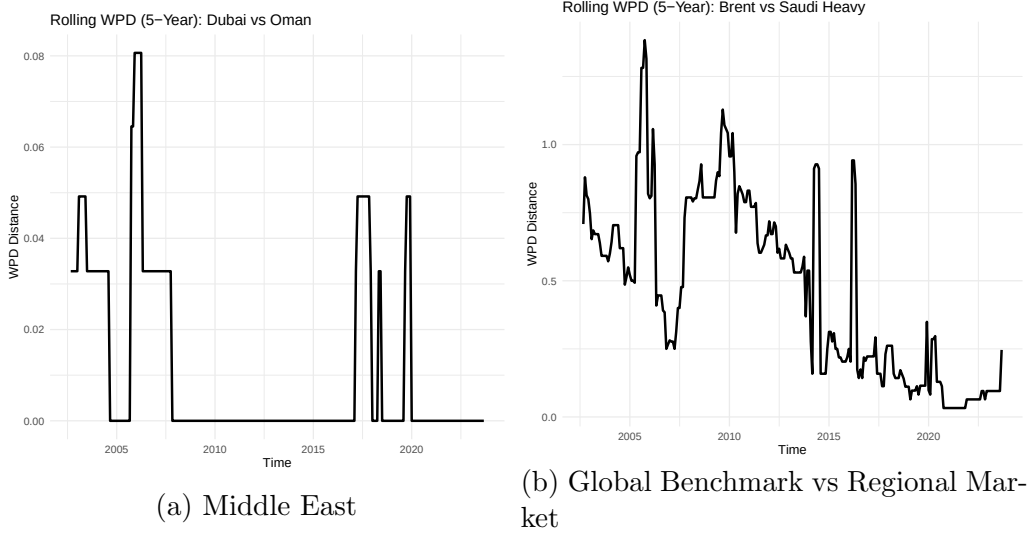
Figure 5: Rolling Window Analysis: Brent vs WTI Cushing.



tering solution, we computed the average silhouette width, which measures both cohesion within clusters and separation between them. The resulting value of 0.65 indicates that the regional oil market groupings are reasonably well-defined; most markets show strong internal consistency while remaining sufficiently distinct from markets in other regions. Taken together, these metrics provide complementary evidence that the hierarchical clustering captures meaningful and robust regional structures within the global oil market.

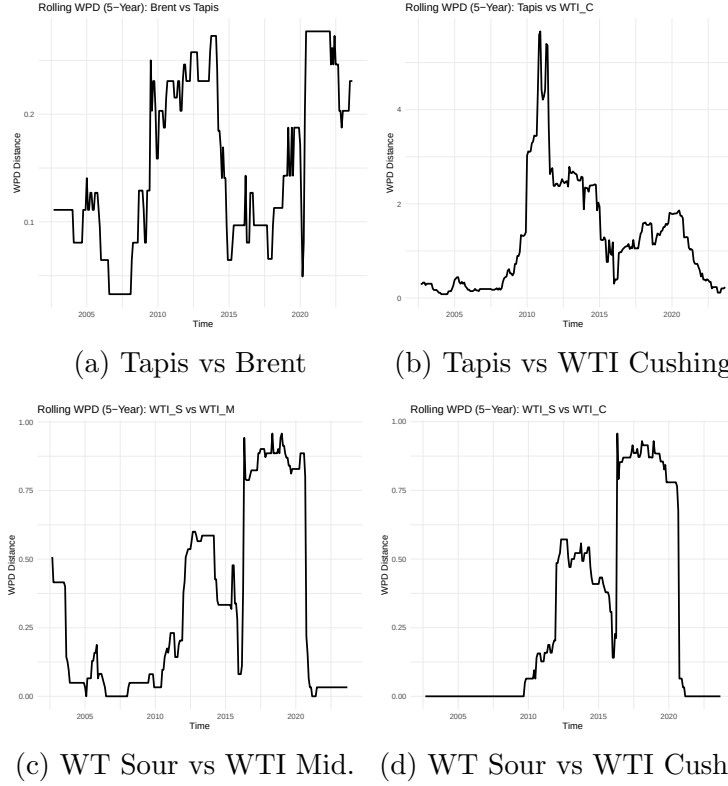
The second stage employs a rolling window framework based on the Warping Path Distance (WPD), which offers a distinct advantage over static clustering approaches as it allows for capturing the dynamic nature of market

Figure 6: Rolling Window Analysis (continued)



integration over time. Unlike traditional methods that assume stable relationships, the rolling window approach allows us to track how the similarity structure among regional oil markets evolves continuously, and, thus, reveals periods of convergence, divergence, or structural shifts that would otherwise remain hidden in a time-averaged analysis. WPD is particularly well-suited for this task because it accounts for temporal misalignments - such as lagged price transmissions or asynchronous market responses - that are common in global energy markets, thereby providing a more accurate measure of integration than conventional distance metrics. By sliding a fixed window across the time series and recomputing pairwise dissimilarities at each step, we construct a temporal sequence of dissimilarity matrices that illuminates how market linkages strengthen, weaken, or reconfigure in response to geopolitical

Figure 7: Roling Window Analysis (continued).



events, supply shocks, or evolving trading patterns. This dynamic perspective transforms the analysis from a static snapshot of regional groupings into a nuanced narrative of how oil market integration unfolds through time.

Figure 5 presents the five-year rolling WPD for the most prominent pair: Brent and WTI Cushing. This pair is well-known for the substantial price spread observed between 2010 and 2014, during which prices diverged dramatically, as previously discussed. The dissimilarity measure captures this episode clearly: it rises sharply at the onset, subsequently declines, and then

drops to near zero by 2015. This pattern confirms that the analysis produces credible results. Following this period, a stable phase with moderate dissimilarity persists between 2015 and 2020. Most notably, however, dissimilarity is very low both prior to 2009 and after 2020 - these periods correspond to the dramatic oil price surge which precedes the Global Financial Crisis and the volatile price movements triggered by the COVID-19 pandemic and, subsequently, the geopolitical tensions surrounding the Ukraine war.

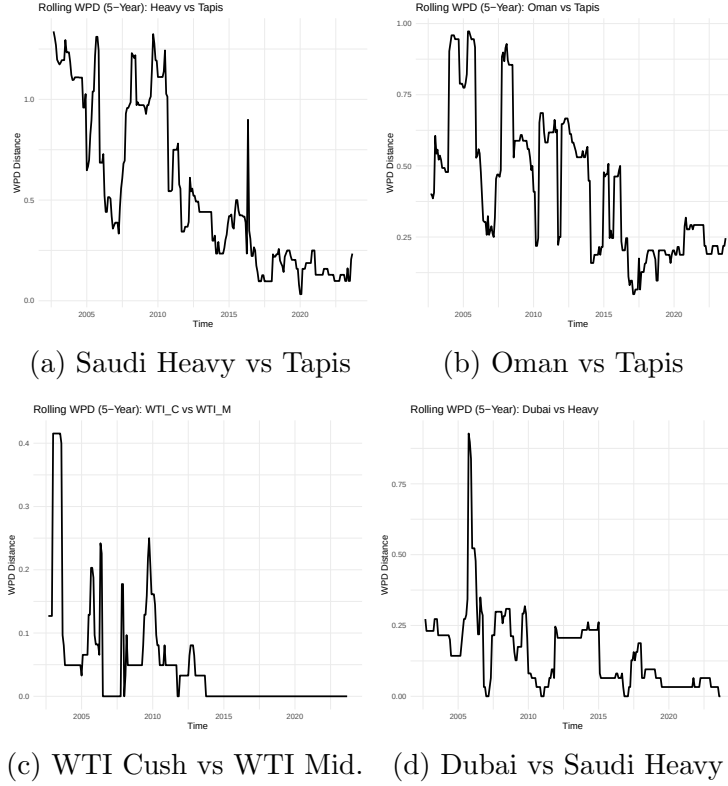
We now turn to two additional pairs of interest: the Middle Eastern markets Dubai and Oman, which, as discussed above, form a cohesive regional cluster, and the pairing of Brent - a global benchmark - with Saudi Heavy, which represents Middle Eastern crude. These results are presented in Figure 6. As expected, the two Middle Eastern markets exhibit strong similarity, which is reflected in consistently low dissimilarity scores throughout the sample period (Figure 6a). For the latter pairing - Brent and Saudi Heavy - there is a notable degree of dissimilarity in the early part of the sample, which gradually diminishes over time (Figure 6b). This pattern illustrates how crude oil markets become increasingly integrated over time; it captures the process by which a regional crude series becomes more aligned with a global benchmark.

Figure 7 presents additional results. Turning first to Tapis, its price relationship with Brent (Figure 7a) remains highly similar throughout the sample; however, dissimilarity increases during the 2010-2015 WTI-Brent spread period and again after 2020. Overall, the level of dissimilarity stays low,

only slightly above the WTI-Brent similarity observed before 2009 and after 2020. A different picture emerges for Tapis versus WTI Cushing (Figure 7b). During the spread period, Tapis follows Brent and becomes highly dissimilar to WTI Cushing. This suggests that regional prices must choose a global benchmark to follow - in Tapis' case, Brent. This pattern is inconsistent with a fully unified global oil market. Turning to the U.S. markets (Figures 7c and 7d), West Texas Sour becomes less similar to WTI Midland and WTI Cushing over time. A certain level of dissimilarity has already been present during 2010-2015, this divergence intensifies between 2016 and 2020, which shows that even within U.S. markets, integration can temporarily weaken. However, all three U.S. series revert to high similarity during the post-2020 crisis period, a pattern of heightened similarity during extreme episodes also observed for the WTI Cushing - Brent relationship.

We now turn to Figure 8, which reveals a highly interesting pattern. Figures 9a and 9b examine the evolution of integration among regional markets that are not global benchmarks: Saudi Heavy, Tapis, and Oman. These pairs exhibit some degree of price dissimilarity initially, which gradually declines over time. This pattern illustrates the path toward a more unified global oil market - a textbook finding. We next consider two pairs from within the same region: WTI Cushing and WTI Midland within the United States (Figure 9c), and Dubai and Saudi Heavy within the Middle East (Figure 9d). The patterns are strikingly similar: both pairs exhibit low levels of price dissimilarity throughout, and in both cases, dissimilarity declines over

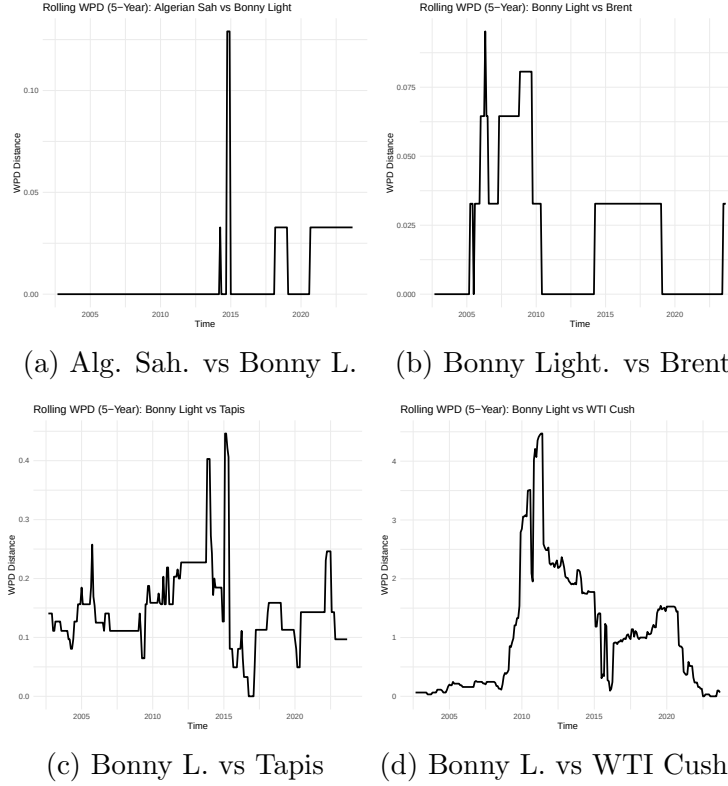
Figure 8: Rolling Window Analysis (continued)



time.

Figure 9 focusses on the two West African series, Bonny Light and Algerian Saharan. Among all regional pairs, they exhibit the highest degree of price similarity, second only to the Dubai-Oman pair. Their trajectories track each other extremely closely throughout most of the sample. The dissimilarity range is slightly wider than that of the Dubai-Oman pair, mainly due to a spike just before 2015, but during the first half of the sample they move in near lockstep. Regarding their relationship with global benchmarks,

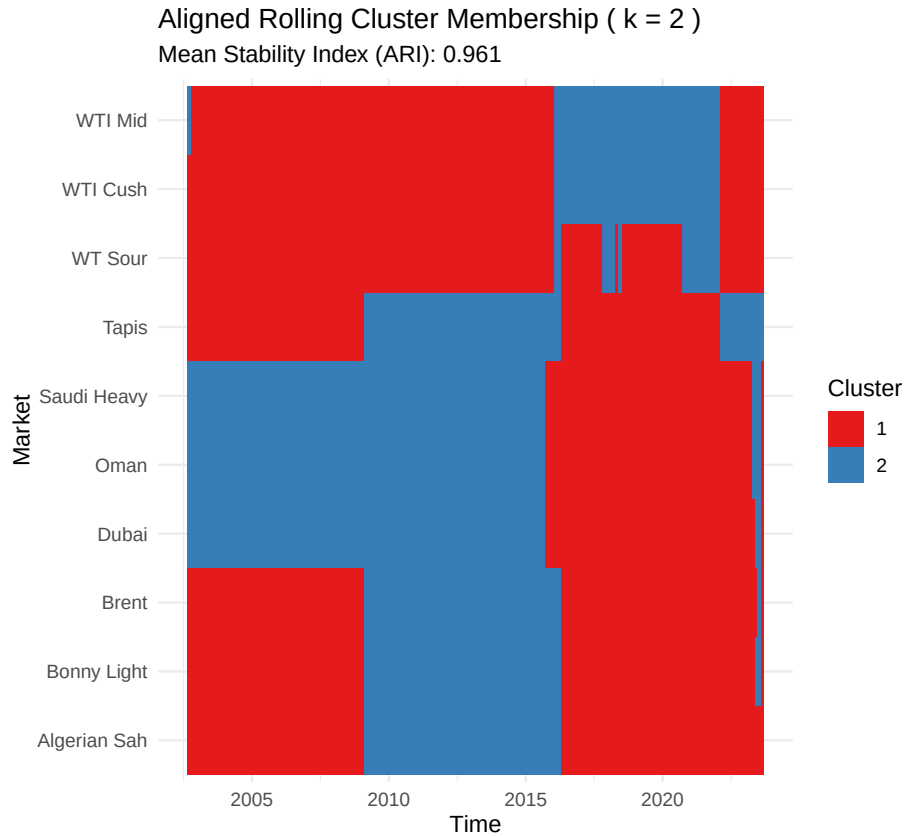
Figure 9: Rolling Window Analysis (continued)



Bonny Light - a market that is of more than just regional relevance and approaches near-benchmark status - is consistently more similar to Brent than to WTI Cushing. Dissimilarity with Brent is largest before 2010, after which the two series become quite similar. With WTI Cushing, the pattern is markedly different: dissimilarity spikes dramatically during the Brent - WTI spread period (2010-2014) and remains elevated even after 2015. Only post-2020 do Bonny Light and WTI Cushing become more similar. Finally, the similarity between Bonny Light and Tapis is high throughout the sample.

Some fluctuation exists, but there is no clear downward trend. This suggests that, unlike the gradual convergence observed for other pairs, Bonny Light and Tapis have maintained a stable and relatively high level of similarity over time. One would not expect further convergence, as the two series reflect quite different geographical regions. The remaining pairs are presented in the Appendix.

Figure 10: Cluster Stability Analysis



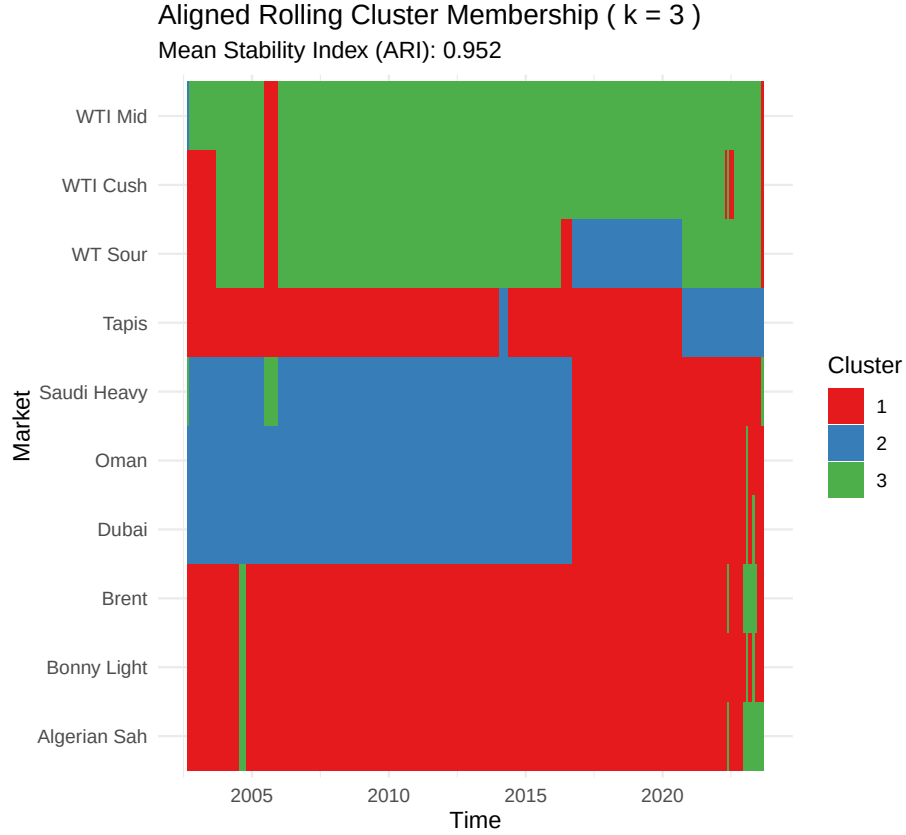
The rolling window approach not only allows us to track how the price relationship between pairs of series evolves over time; it also enables a re-

peated cluster analysis, as each window yields a dissimilarity matrix (see Section 3 for details). We select a fixed number of clusters, initially $k = 2$. The two clusters roughly reflect geographical regions; however, not always as expected. Cluster 2 initially contains the Middle Eastern markets, while Cluster 1 contains the remaining markets: Brent, the American markets, the West African markets, and Tapis. This configuration shifts around 2009; following the oil price collapse after the peak preceding the Global Financial Crisis. From 2009 to 2016, which includes also the period of the WTI-Brent spread, Tapis, Brent and the West African markets move to Cluster 2, leaving only the American markets in Cluster 1. Post 2016, there is only US markets vs the rest of the world, and at times West Texas Sour becomes more similar to the Rest of the World than American markets and moves to Cluster 1.

After 2022, Brent and the Middle Eastern markets switch into the American cluster; only Tapis remains in Cluster 2. This reflects that during a period characterised by a large extent of oil market stress caused by the Ukraine war, a truly globalised oil market emerged. Only in the very last few months towards the end of the sample, the Middle Eastern markets leave this global cluster and join Tapis. Recall that cluster membership is determined by the relative price similarity across markets. As previously noted, West Texas Sour exhibits signs of price dissimilarity relative to the other American markets; this is reflected here in its departure from the American cluster to join the other - again, strictly in terms of price similarity.

To assess the stability of the regional clustering structure over time, we

Figure 11: Cluster Stability Analysis (continued).



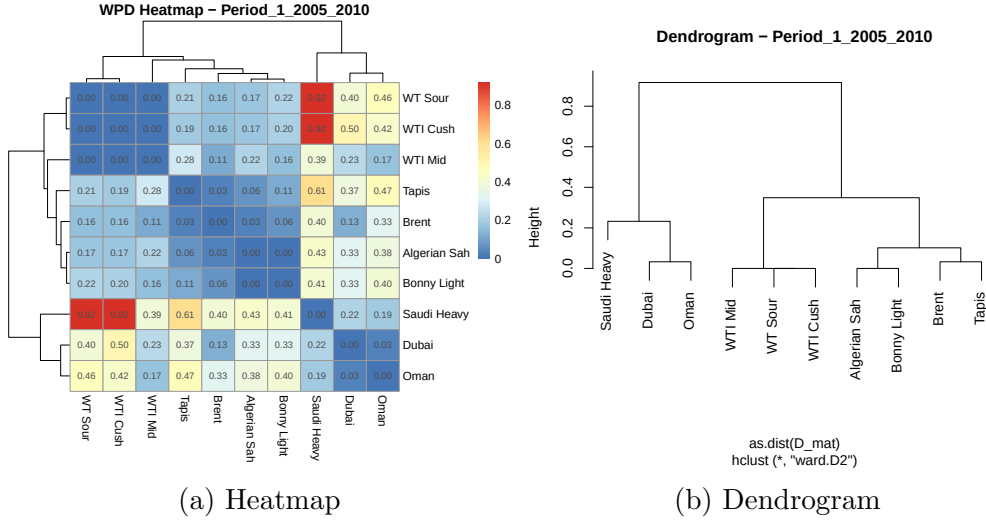
employed the Adjusted Rand Index (ARI), a metric that measures the similarity between two clusterings while correcting for agreement expected by chance, with values ranging from 0 (no better than random) to 1 (perfect agreement). Comparing the cluster assignments across consecutive rolling windows yielded an ARI of 0.961, which indicates an exceptionally high degree of temporal stability in the regional clusters. This finding demonstrates that despite the dynamic evolution of market integration captured by the Warping Path Distance, the underlying regional structure remains remark-

ably persistent.

Choosing $k = 2$ forces the global oil market into a binary global structure. To allow for more flexibility and acknowledge global heterogeneity, we also set $k = 3$. Also for this parameter setting, the results reveal an interesting geographical pattern. From the beginning of the sample until approximately 2016, one cluster contains only the three American markets, with very few brief exceptions. Another cluster (Cluster 2) contains only the three Middle Eastern markets. The remaining markets - Tapis, Brent, and the two West African series - are grouped together in Cluster 1. This structure changes quite drastically after 2016. The Middle Eastern cluster dissolves; the three Middle Eastern series join the Rest-of-the-World cluster. Post-2016, the pattern is clear: there is an American cluster and a Rest-of-the-World cluster. With $k = 3$ exogenously set, a third cluster must be created. This residual cluster, however, does not have a straightforward geographical interpretation. Instead, it consistently contains the price series that is least similar to the others - first West Texas Sour (2016-2020), then Tapis (toward the end of the sample). This is a caveat of the fixed $k = 3$ assumption, but it also reveals an interesting finding: when forced into three clusters, the algorithm isolates the most idiosyncratic series rather than forming a meaningful third region.

The preceding analysis highlights the strong influence of the Brent-WTI spread period on the results. To further investigate this, we conduct a simple subsample analysis based on well-known idiosyncratic episodes in the crude

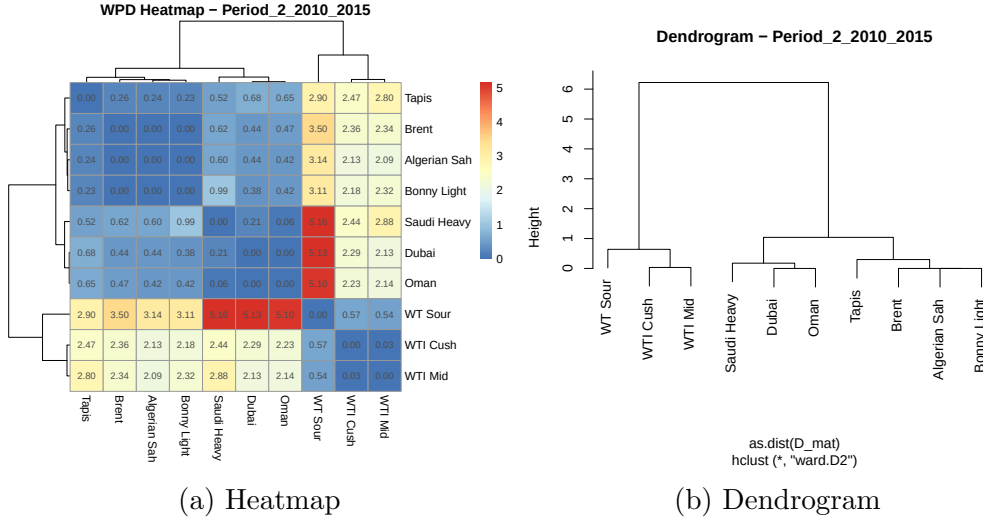
Figure 12: 2005-2010



oil market. We divide the sample into three distinct periods. The first, pre-2010, captures both the relentless price increase leading up to the 2008 peak and the subsequent crash during the Global Financial Crisis. The second period, 2010-2014, covers the Brent-WTI spread period. The third period, 2020-2024, includes not only the price collapse related to the outbreak of COVID-19, the subsequent recovery, but also the price surge following Russia's invasion of Ukraine. This approach allows us to assess whether the clustering structure differs across these qualitatively distinct episodes.

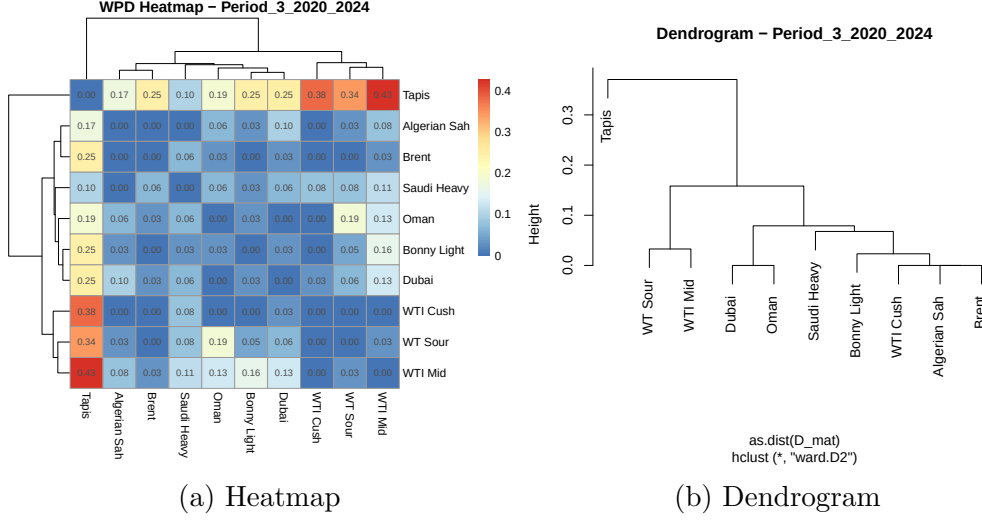
Figure 12 summarizes the results for the 2005-2010 period. The heatmap (Figure 12a) reveals the highest levels of price similarity within the same geographic regions: the American markets group together, as do the West African and Middle Eastern markets. At the opposite end, Saudi Heavy is highly dissimilar to the American markets, particularly to West Texas Sour

Figure 13: 2010-2015



and WTI Cushing. Tapis also shows relatively high dissimilarity to other markets, especially those in the Middle East. The American and Middle Eastern markets are quite dissimilar to one another as well. Brent, one of the global benchmarks, is overall relatively similar to other markets, with the closest relationships to Tapis and the West African markets. However, the degree of dissimilarity is larger than the within-region similarity observed within the three geographic areas. This pattern is clearly reflected in the dendrogram (Figure 12b). Moving up along the dissimilarity axis, three clusters initially emerge: the Middle East, the American markets, and a Rest-of-the-World cluster. Within the Rest-of-the-World cluster, the West African markets are closer to each other than to the other members. The American cluster then merges with the Rest-of-the-World cluster, leaving only the Middle Eastern markets isolated. Only at a high level of dissimilarity

Figure 14: 2020-2024



do all markets merge into a global cluster.

Figure 13a presents the heatmap for the 2010-2015 period. Unlike previous results, the colors appear to be very regular. Three main groups still emerge: the American markets (especially the two WTIs), the Middle East, and a group comprising West Africa plus Brent. The strongest dissimilarity occurs between West Texas Sour and the three Middle Eastern markets. As the subsamples contain the same number of observations, the dissimilarity values are directly comparable to those from the 2005-2010 period: They are substantially larger, which is attributable to the WTI-Brent spread. Overall, the American markets are fairly dissimilar to all other markets, but not as dissimilar as West Texas Sour is to Saudi Heavy. This clearly illustrates that when two global benchmarks diverge, prices must choose which benchmark to follow. The dendrogram for this period (Figure 13b) reveals an initial

Middle Eastern cluster and an American cluster, with West Texas Sour initially separate from the other two U.S. series. Brent and the West African markets form one cluster, later joined by Tapis. A Rest-of-the-World cluster emerges that contains all markets except the American series. Only at an extreme level of dissimilarity does a global cluster form.

The 2020-2024 period (Figure 14) looks very different from both previously discussed subsamples. The heatmap in Figure ?? is much bluer, which means that similarities are much higher in general. The lowest dissimilarities again appear within American markets. In addition, many markets are very similar to Brent. Others are less similar to American markets, with Tapis standing out; Oman and Bonny Light also show notable dissimilarity to American markets. As overall similarity is very high, the cluster membership becomes less clear. The dendrogram (Figure 14b) shows the following: Moving up along the dissimilarity axis, the first cluster contains the two global benchmarks plus Algerian Saharan. Dubai and Oman form a cluster that later joins this one. Before that, however, Saudi Heavy and Bonny Light - geographically far apart - cluster together and also join. A separate cluster contains the two more regional American markets, with Tapis joining only at a much higher dissimilarity level. Note that this level remains considerably lower than in the 2005-2010 subsample. Sample sizes are identical across periods, allowing direct comparison.

This finding carries an important implication. During periods of market turbulence, global oil market integration increases, but the clear geographical

pattern observed in calmer periods becomes less pronounced. In other words, when markets are under stress, regional distinctions blur as prices move more synchronously across the globe. However, this heightened integration does not follow a neat regional logic. Instead, idiosyncratic relationships - such as the pairing of geographically distant markets like Saudi Heavy and Bonny Light - emerge alongside traditional regional clusters. The geographical signal weakens; it is replaced by a more fluid and less predictable structure. This suggests that while crises temporarily unify global oil markets, they also disrupt the orderly regional segmentation that characterizes more tranquil periods.

The quality of the clustering solutions is assessed using the silhouette width and the cophenetic correlation coefficient. The former measures both within-cluster cohesion and between-cluster separation, while the latter indicates how faithfully the dendrogram preserves the original pairwise distances. For the 2005-2009 subsample, the silhouette width is 0.6825 and the cophenetic correlation is 0.7716. The silhouette value exceeds the commonly accepted threshold of 0.5, which indicates that the clusters are reasonably well-defined, but not extremely strong. The cophenetic correlation suggests a moderately good representation of the distance matrix in the dendrogram. The 2010-2014 period shows the strongest results, with a silhouette width of 0.8734 and a cophenetic correlation of 0.8818. These values are substantially higher than those for the other two periods, which reflects the exceptionally clear regional clustering pattern observed during the Brent-WTI spread

episode. The high silhouette width indicates that markets within each cluster are very similar to one another and well-separated from other clusters. The high cophenetic correlation confirms that the dendrogram accurately captures the underlying similarity structure. For the 2020-2024 subsample, the silhouette width is 0.6929 and the cophenetic correlation is 0.8237. The silhouette value, while lower than in 2010-2014, still indicates reasonably well-defined clusters. The cophenetic correlation remains relatively high, suggesting that the hierarchical structure is well captured despite the more fluid integration pattern observed in this turbulent period. Overall, these metrics confirm that the clustering solutions are robust across all periods. The 2010-2014 period exhibits the clearest structure, while the 2005-2009 and 2020-2024 periods, while somewhat less distinct, still produce meaningful and reliable clusterings.

5 CONCLUSION

This paper introduces a novel approach to measure global crude oil market integration by combining dynamic time warping with hierarchical cluster analysis. Unlike pairwise tests or conventional multivariate models, the method captures system-wide patterns of comovement without imposing restrictive parametric assumptions. It handles temporal misalignments naturally and produces visually interpretable results.

The findings are credible yet revealing. Global crude oil markets cluster primarily by geography. West African markets (Bonny Light and Algerian

Saharan) form the tightest regional cluster, followed by the Middle Eastern markets (Dubai, Oman, and Saudi Heavy) and then the U.S. markets (WTI Cushing, WTI Midland, and West Texas Sour). Brent and Tapis occupy intermediate positions. A rolling window analysis shows that integration has increased over time, particularly between Brent and Saudi Heavy, but the regional structure remains remarkably stable. The Brent-WTI spread episode of 2010-2014 is captured cleanly as a period of heightened dissimilarity. The subsample analysis reveals that during the turbulent 2020-2024 period, integration increases but the clear geographical pattern weakens, giving way to more fluid and idiosyncratic relationships. This pattern one might not have expected to detect in advance.

Taken together, the evidence suggests several regional pools rather than a single fully integrated market. The method proves well-suited to this research question: it is flexible, data-driven, and reveals structure that pairwise approaches miss. More broadly, this study embraces an “empirics first” approach - a perspective that has gained traction in marketing (Golder et al., 2023) and is increasingly relevant in economics. The ongoing data revolution makes vast amounts of new information available, which creates both opportunities and demands for methods that let the data speak without strong theoretical priors. In times of heightened uncertainty, purely data-driven approaches are essential: they can detect patterns that may be difficult to anticipate, and offer insights that theory-driven methods might overlook. The framework developed here is well-suited for application beyond oil markets,

including other commodity markets, financial assets, or any setting where understanding the structure of comovement across multiple time series is of interest.

REFERENCES

- Adelman, M. A. (1984). International Oil Agreements. *The Energy Journal*, 5(3), 1–10. <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol15-No3-1>
- Bravo Caro, J. M., Golpe, A. A., Iglesias, J., and Vides, J. C. (2020). A new way of measuring the WTI–Brent spread: Globalization, shock persistence and common trends. *Energy Economics*, 85, 104546. <https://doi.org/10.1016/j.eneco.2019.104546>
- Chatziantoniou, I., Gabauer, D., and Gupta, R. (2023). Integration and risk transmission in the market for crude oil: New evidence from a time-varying parameter frequency connectedness approach. *Resources Policy*, 84, 103729.
- Cui, J. and Maghyreh, A. (2023). Time-frequency dependence and connectedness among global oil markets: Fresh evidence from higher-order moment perspective. *Journal of Commodity Markets*, 30, 100323.
- Gronwald, M. and Jin, X. (2024). Measuring world oil market integration with a Thick Pen. *Energy Economics*, 130, 107315. <https://doi.org/10.1016/j.eneco.2024.107315>

- Lima, F. T. and Souza, V. M. A. (2023). A Large Comparison of Normalization Methods on Time Series. *Big Data Research*, 34, 100407. <https://doi.org/10.1016/j.bdr.2023.100407>
- Mastroeni, L., Mazzocchi, A., Quaresima, G., and Vellucci, P. (2021). Decoupling and recoupling in the crude oil price benchmarks: An investigation of similarity patterns. *Energy Economics*, 94, 105036. <https://doi.org/10.1016/j.eneco.2020.105036>
- Plante, M. and Strickler, G. (2021). Closer to one great pool? Evidence from structural breaks in oil price differentials. *The Energy Journal*, 42(2), 1–30. <https://doi.org/10.5547/01956574.42.2.mpla>
- Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>
- Shi, H., Wang, Y., and Zheng, X. (2025). Extreme comovement and risk spillovers in crude oil prices: A tale of two events. *Journal of Futures Markets*. <https://doi.org/10.1002/fut.70059>
- Wu, S. and Han, W. (2025). Asian geopolitical risks: A key driver behind WTI-Brent spread market volatility. *Finance Research Letters*, 77, 107090.
- Golder, P. N., Dekimpe, M. G., An, J. T., van Heerde, H. J., Kim, D. S. U., and Alba, J. W. (2023). Learning from Data: An Empirics-First Approach

to Relevant Knowledge Generation. *Journal of Marketing*, 87(3), 319–336.

<https://doi.org/10.1177/00222429221129200>