



Case report

Topology of currencies: Persistent homology for FX Co-movements: A Comparative clustering study

Pattravadee de Favereau de Jeneret¹ and Ioannis Diamantis^{2,*}

¹ Maastricht University, School of Business and Economics, P.O.Box 616, 6200 MD, Maastricht, The Netherlands

² Department of Data Analytics and Digitalisation, Maastricht University, School of Business and Economics, P.O.Box 616, 6200 MD, Maastricht, The Netherlands

* **Correspondence:** Email: i.diamantis@maastrichtuniversity.nl.

Abstract: This study investigates whether Topological Data Analysis (TDA) can provide additional insights beyond traditional statistical methods in clustering currency behaviors. We focus on the foreign exchange (FX) market, which is a complex system often exhibiting non-linear and high-dimensional dynamics that classical techniques may not fully capture. We compare clustering results based on TDA-derived features versus classical statistical features using monthly logarithmic returns of 13 major currency exchange rates (all against the euro). Two widely-used clustering algorithms, *k*-means and Hierarchical clustering, are applied on both types of features, and cluster quality is evaluated via the Silhouette score and the Calinski-Harabasz index. Our findings show that TDA-based feature clustering produces more compact and well-separated clusters than clustering on traditional statistical features, particularly achieving substantially higher Calinski-Harabasz scores. However, all clustering approaches yield modest Silhouette scores, underscoring the inherent difficulty of grouping FX time series. The differing cluster compositions under TDA vs. classical features suggest that TDA captures structural patterns in currency co-movements that conventional methods might overlook. These results highlight TDA as a valuable complementary tool for analysing financial time series, with potential applications in risk management where understanding structural co-movements is crucial.

Keywords: topological data analysis; persistent homology; currency co-movements; clustering; Foreign Exchange (FX) market; clustering methods; wasserstein distance

JEL Codes: C38, C63, F31, G15

1. Introduction

In April 2022, trading in the over-the-counter foreign exchange (FX) market averaged more than \$7.5 trillion per day, representing an increase of 14% compared to three years earlier (European Central Bank, 2023). Given this large scale and the FX market's central role in global finance, even small fluctuations can have far-reaching effects across the world. For example, the 2008 global financial crisis highlighted how shocks in one currency can quickly spread to others. Understanding currency co-movements is therefore crucial for risk management, portfolio diversification, and mitigating systemic instability. Yet, currencies rarely move in isolation: their behavior reflects a complex combination of macroeconomic, political, and structural factors.

Reference rates play a central role in this system, serving as benchmarks for other financial instruments such as loans and exchange rates. They provide a common mechanism that links international interest payments and valuations to market conditions (ECB, 2023). Consequently, reference rates offer valuable information about currency co-movements, which can enhance systematic risk assessment.

Traditional statistical methods, such as correlation, covariance, and clustering, have been widely applied to uncover relationships between currency movements. These approaches are efficient and interpretable, but they rely on assumptions of linearity, low dimensionality, and stationarity that may not hold in the dynamic, interconnected FX market. As a result, they often fail to capture non-linear or higher-order dependencies, potentially underestimating systemic risk and misguiding policy or investment decisions.

Topological Data Analysis (TDA), a recent framework grounded in algebraic topology, provides a complementary approach. TDA characterizes the “shape” of data by studying how geometric and topological features evolve across scales (Shultz, 2023). Unlike traditional statistical tools, TDA does not depend on coordinate systems or linear assumptions, and it is robust to noise, making it well suited for analysing high-dimensional, non-linear, and complex systems such as the FX market.

Building on this foundation, the present study investigates the extent to which TDA can add value to classical statistical methods in the analysis of foreign-exchange dynamics. Specifically, we perform a comparative clustering study based on classical statistical features and TDA-derived features, evaluating whether TDA uncovers additional structures that conventional techniques may overlook.

Hence, the central research question is:

How does clustering based on TDA-derived features compare in quality to clustering based on classical statistical features?

We note that the focus of this paper is on benchmarking clustering structure induced by classical versus topological representations; accordingly, the study is descriptive and does not attempt causal identification. For clarity, the analysis is designed as a comparative benchmark rather than a hypothesis-testing or econometric modeling exercise. All experiments are conducted under a single, fixed experimental configuration to ensure comparability between classical statistical and topological representations. Therefore, a regime-specific analysis, formal statistical inference, and exhaustive robustness testing are left for future work, as they would require a fundamentally different study design.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on classical statistical and topological approaches to currency co-movements; Section 3 describes the

data selection and preprocessing; Section 4 presents an exploratory analysis of the dataset; Section 5 outlines both statistical and TDA-based clustering procedures; Section 6 discusses the empirical results, limitations, and directions for future research. Finally, Section 7 concludes the paper.

2. Literature review

2.1. Statistical methods in FX studies

Numerous methods have been used to analyze currency co-movements and interrelationships, including correlation, covariance, and clustering. These approaches remain popular in finance due to their mathematical efficiency and interpretability. They can also be combined to provide a more comprehensive picture, with correlation and covariance often serving as foundational measures.

Mai et al. (2017) employed network-based correlation analysis to study currency co-movements in the FX market, finding that currencies often cluster by geographic proximity. Thus, their paper exemplifies both the strengths of correlation analysis, such as the derivation of regional modules, and its weaknesses, notably its dependence on linear relationships. Similarly, Drozd et al. (2007) used correlation matrices to explore collective behavior and clustering among 60 world currencies. Their findings indicated that results were strongly dependent on the choice of base currency: when the euro or US dollar served as the base, the derived co-movements were more diverse and revealed subtler interdependencies. Together, these studies show that correlation analysis can uncover meaningful patterns but remains limited by its linear assumptions.

Other studies have instead relied on covariance as a foundation. Andersen et al. (2003) used realized covariance to measure how pairs of asset returns move together, providing a valuable basis for improved volatility modeling and forecasting, an essential task in financial applications.

Building on these classical measures, clustering has become an increasingly common tool for analysing time-series behavior in financial markets. Paparrizos et al. (2024) offer a comprehensive review of clustering methods, highlighting the relative strengths of approaches such as hierarchical clustering and emphasizing the importance of selecting appropriate distance metrics. Traditional distance measures, they argue, often fail when directly applied to complex time-series data.

In financial applications, the typical input for clustering analyses is logarithmic returns, which are preferred for their stationarity and additive properties. For instance, Verma et al. (2018) used log-returns to create log-volatility features, which were then clustered to capture empirical market dynamics. Similarly, Mantegna (1999) applied hierarchical tree analysis to price time series, showing that clusters tended to group companies operating within the same industry or sub-industry sectors, revealing meaningful economic structure through statistical similarity.

2.2. Topological data analysis

In recent years, Topological Data Analysis (TDA) has emerged as a versatile framework for uncovering the underlying *shape* of data, offering tools that extend beyond traditional statistical measures. Garcia (2022) provides an accessible introduction to persistent homology, demonstrating its capacity to extract structure from complex, high-dimensional data in applications ranging from natural-language processing to handwritten-digit recognition.

Lum et al. (2013) illustrated TDA's broad applicability by extracting topological features across

diverse domains, including cancer research, political science, and sports analytics, highlighting TDA's unique ability to detect subgroups and hidden relationships. El-Yaagoubi et al. (2023) further advanced this perspective by applying persistent homology to dependence networks in multivariate time series, with specific applications to brain connectivity. Their results showed that TDA can identify hidden cyclic structures through persistence landscapes, a property that naturally translates to financial data, where multivariate dependencies are common.

Within economics, TDA adoption has been limited but promising. Schultz (2023) positioned the method as a complementary tool to econometrics, particularly well suited to non-linear and high-dimensional datasets where traditional models struggle. Among the potential applications discussed were early-warning indicators of market instability, an area where topology can reveal structural signals obscured by noise.

Aguilar et al. (2020) used TDA to detect structural changes in US stock-market data, showing that persistent homology can capture early signs of instability and critical transitions. They cautioned, however, that the method remains computationally intensive and requires careful parameterisation of persistence landscapes. Majumdar et al. (2020) demonstrated TDA's potential for clustering by combining it with self-organising maps (SOMs) and random forests on stock-price series. Their hybrid approach successfully distinguished simulated from real financial processes, underscoring TDA's ability to capture structures beyond volatility or correlation alone.

Recent comparative studies have extended these insights. Hobbelhagen et al. (2024) compared TDA with Symbolic Aggregate Approximation (SAX) on European stock markets, finding that while SAX offered greater scalability and interpretability for broad market trends, TDA provided finer insight into local stock movements. Building on this, Bereta et al. (2025) added the extended SAX (eSAX) approach and applied all three methods to consumer-behavior time series. Their findings confirmed the complementarity of symbolic and topological methods: symbolic approaches were computationally efficient and interpretable, whereas TDA avoided "catch-all" clusters despite noisy data.

2.3. Research gap and contribution

Both statistical methods, such as covariance, correlation, and clustering, and TDA have yielded valuable insights in financial research. TDA, in particular, has proven effective in extracting meaningful structure from complex and noisy datasets, identifying early-warning signals of structural shifts, and modeling volatility. Despite this progress, applications of TDA within finance remain largely concentrated on stock-market data, with studies on the FX market still scarce. Moreover, most existing comparisons, such as those in Hobbelhagen et al. (2024) and Bereta et al. (2025), focus on symbolic representations rather than direct contrasts between TDA and conventional clustering approaches like k -means or hierarchical clustering.

Accordingly, this study addresses these gaps by offering a direct comparison of statistical and TDA-based clustering, specifically, k -means and hierarchical methods, within the context of the FX market. The results aim to reveal whether TDA captures non-linear dependencies and structural co-movements that traditional statistical techniques may overlook.

Importantly, this comparison is not intended to position TDA as a wholesale replacement for classical statistical methods. In noisy and highly interconnected systems such as foreign exchange markets, demonstrating clear methodological superiority is inherently challenging. Instead, the contribution of this study lies in providing a transparent benchmark for the practical performance of TDA in FX clustering.

Our results show that TDA-based features offer an independent, shape-based perspective that largely supports and refines conclusions drawn from traditional approaches, while modestly improving cluster compactness and separation. By making this distinction explicit, the paper clarifies both the strengths and the current limitations of TDA in applied financial settings and sets a clear bar for future methodological advances.

3. Data preparation

3.1. Data selection

The data used in this study were collected from the European Central Bank (ECB) database. Multiple datasets were retrieved, each representing a time series of reference rates, defined as “an exchange rate that is not intended to be used in any market transactions” (ECB, 2023), with the euro serving as the base currency across all series.

Using reference rates rather than traded exchange rates reduces the influence of central bank interventions and emphasizes factors such as global trade, inflation, monetary policy, and market activity. The ECB derives these reference rates primarily from actual market transactions in liquid markets with high trading volumes and frequent trades. For less liquid markets, the rates rely on firm bid/ask quotes, USD cross-rate calculations, and discretionary adjustments (ECB, 2023).

The use of a uniform base currency ensures a consistent and accurate comparison of relationships and movements among the different currencies. For instance, if volatility rises in some pairs while remaining stable in others, it may indicate regional or geopolitical shocks; conversely, a simultaneous rise across all pairs suggests systemic disruption centred around the euro area.

The ECB provides public access to 30 reference rates globally, including many European currencies not part of the euro area. For this study, a diverse set of currencies was selected to include both advanced and emerging economies:

AUD, BRL, CHF, CNY, GBP, INR, JPY, KRW, RUB, THB, TRY, USD, ZAR, and EUR.

The aggregated and cleaned dataset spans from 13 January 2000 to 1 March 2022, comprising 14 columns (excluding dates) and 5 993 observations. Each column represents a currency’s reference rate time series, with EUR remaining constant as the numeraire. A second dataset was also constructed excluding the Russian ruble (RUB), extending the coverage to 27 November 2024 to capture more recent global events. These two datasets were later used in the exploratory analysis to better understand temporal behavior and to contextualise clustering results.

Several currencies were selected because of their membership in the BRICS bloc (Brazil, Russia, India, China, South Africa), providing a robust environment for studying reference-rate behavior under diverse macroeconomic regimes. To complement these, advanced economies (AUD, CHF, GBP, JPY, USD) were included, as they often act as global financial anchors and dominate international reserves. Changes in their rates can propagate globally, though the precise transmission dynamics require granular analysis. Finally, emerging-market currencies such as THB and KRW, which have shown relative stability, and TRY, which has undergone prolonged inflationary episodes since the late 2010s, were included to enrich the contrast between stability and volatility.

3.2. Data manipulation

Before conducting the statistical and topological analyses, several preprocessing steps were necessary to ensure the comparability and stability of the currency time series. Foreign exchange reference rates exhibit heteroskedasticity, nonstationarity, and large variations in scale across currencies. To address these issues and enable meaningful clustering, the raw price levels were transformed into logarithmic returns, and the resulting series were standardized. These manipulations reduce the influence of differing volatility levels and align all currencies onto a common scale, ensuring that subsequent clustering and topological features reflect structural co-movements rather than differences in magnitude or scale.

All preprocessing steps, including return construction, monthly aggregation, standardization, and windowing, are applied uniformly across all currencies and are fully specified to ensure reproducibility. No rolling-window or regime-adaptive preprocessing is employed, as the objective is to compare representations under a controlled and fixed pipeline rather than to optimise performance across market states.

3.2.1. Logarithmic returns

Raw price levels, whether scaled or not, tend to cluster according to long-term trends rather than behavioral similarity (Mattera et al., 2025). To capture relative dynamics more accurately, the main analyses (Section 5) use monthly logarithmic returns. The euro was excluded since its constant value as the numeraire adds no information on relative movements and could generate a trivial, economically meaningless cluster. Moreover, in econometrics, returns typically satisfy the stationarity assumptions required by most time-series models better than price levels. The resulting dataset contains 13 columns (currencies) and 266 monthly observations, with varying numbers of daily data points per month aggregated accordingly.

3.2.2. Standardization

Data scaling is a critical preprocessing step for both classification and distance-based clustering. De Amorim et al. (2022) and Wongoutong (2024) showed that scaling can substantially affect model performance, with standardization (Z-score normalization) often outperforming min-max normalisation.

Standardization centers each feature at zero and rescales it so that its standard deviation equals one, ensuring equal weighting across currencies. This transformation preserves the shape of the data because it is linear, maintaining both order and relative distances. Unlike normalization, it is unbounded and therefore more robust to outliers, an essential property for volatile financial time series.

This procedure was applied to the monthly log-returns dataset. Following Berthold et al. (2016), standardization ensures that clustering is based on co-movement patterns rather than differences in magnitude or volatility. In this setting, Euclidean distances between standardized series approximate correlation distances, resulting in clusters that group currencies with similar directional movements rather than comparable variance alone.

4. Exploratory data analysis

4.1. Time series analysis

Figure 1 presents the standardized movements of all currencies over the respective sample periods. These plots enable the visual identification of economic and geopolitical patterns. A general depreciation of the euro is visible between 2009 and 2012, when most currencies trend downward simultaneously. In contrast, when different currencies deviate in direction, this suggests that localized factors dominate rather than systemic eurozone effects.

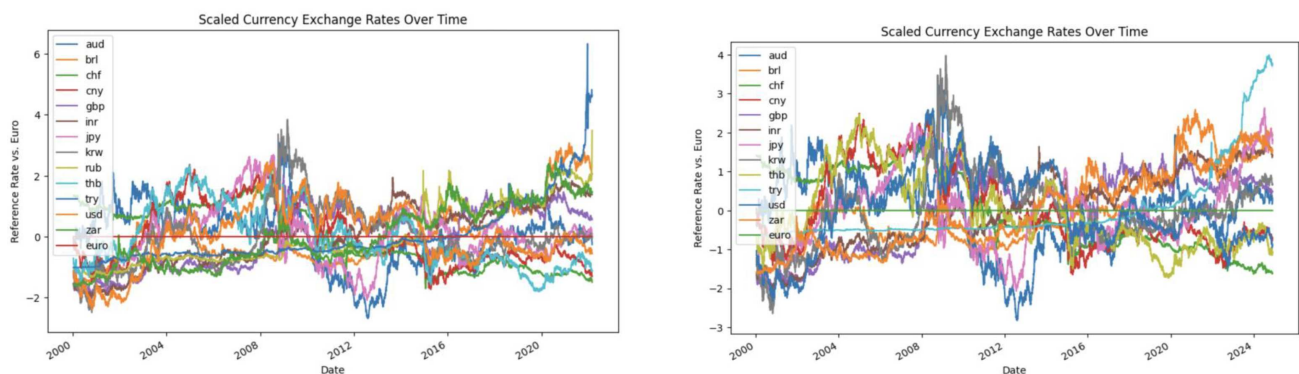


Figure 1. Comparison of standardized currency movements across two time periods. On the left 2000–2022, and on the right 2000–2024.

An interesting feature is that the vertical scale (standardized z-scores) contracts slightly in the longer dataset ending in 2024. This is most likely driven by the extreme inflation of the Turkish lira (TRY) combined with the scaling procedure: since standardization rescales each feature to unit variance, the inclusion of more high-valued data points increases the pre-scaled standard deviation, reducing the resulting z-scores and thereby compressing the apparent spread.

4.2. Variance

To complement the time-series plots, Figure 2 displays the variance of each currency's log-returns for both sample periods. The earlier dataset (2000–2022) and the extended dataset (2000–2024) show that most currencies exhibit only modest shifts in variance. Currencies such as the AUD, CNY, GBP, KRW, THB, and USD experienced slight decreases in volatility, whereas the BRL, CHF, INR, JPY, TRY, and ZAR displayed higher variance.

The TRY stands out, showing a more than ninefold increase in variance over just two and a half years, signaling a prolonged phase of extreme volatility and macroeconomic instability. While variance captures overall dispersion, it does not reveal time-dependent relationships; hence, further temporal decomposition was performed.

4.3. Time series decomposition

Although variances summarise dispersion, they ignore temporal ordering. To investigate structural patterns over time, each series was decomposed into trend, seasonal, and residual components using the Seasonal-Trend decomposition via Loess (STL) method (Theodosiou, 2011). This approach extends

aud	0.023585	aud	0.021533
brl	1.384665	brl	1.701752
chf	0.042873	chf	0.050991
cny	1.333121	cny	1.304023
gbp	0.009807	gbp	0.009387
inr	184.030733	inr	212.672634
jpy	247.536238	jpy	294.504029
krw	27749.352821	krw	25380.316768
rub	380.217517	thb	22.647294
thb	23.267871	try	66.218514
try	7.047286	usd	0.024463
usd	0.025341	zar	16.832439
zar	13.022319		

Figure 2. Comparison of variance of currencies across the same two periods. On the left 2000–2022, and on the right 2000–2024.

classical additive and multiplicative models, which assume fixed seasonal patterns and smooth trends, by employing locally weighted regression (LOESS) to fit evolving local structures. Therefore, STL decomposition offers robustness to outliers and adaptability to non-stationary economic data.

Figure 3 shows the decomposition of the Brazilian real (BRL). The long-term trend reveals cyclical appreciation and depreciation phases, punctuated by large structural shifts associated with political and global economic events, such as the 2008 financial crisis and the COVID-19 pandemic. Residual spikes are observed around 2002, 2008, 2016, 2020, and 2022, aligning with major domestic disruptions including the corruption scandal and presidential impeachment of 2016 (Watts, 2016) and the 2022 election-related uncertainty (Italian Institute for International Political Studies, 2022).

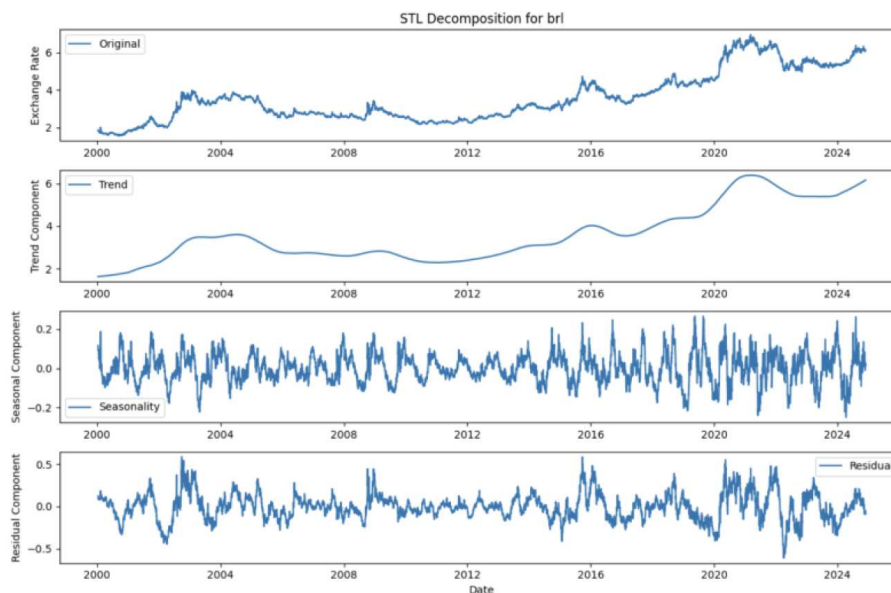


Figure 3. STL decomposition of the Brazilian real (BRL).

Figure 4 presents the Turkish lira (TRY), whose decomposition displays an abrupt structural shift around 2018 following an extended period of relative stability. This sharp change corresponds to

unorthodox monetary policies, including rate cuts that accelerated inflation and undermined investor confidence (Pierini, 2021). The seasonal component exhibits disrupted cycles, further emphasizing the economy's structural transformation.

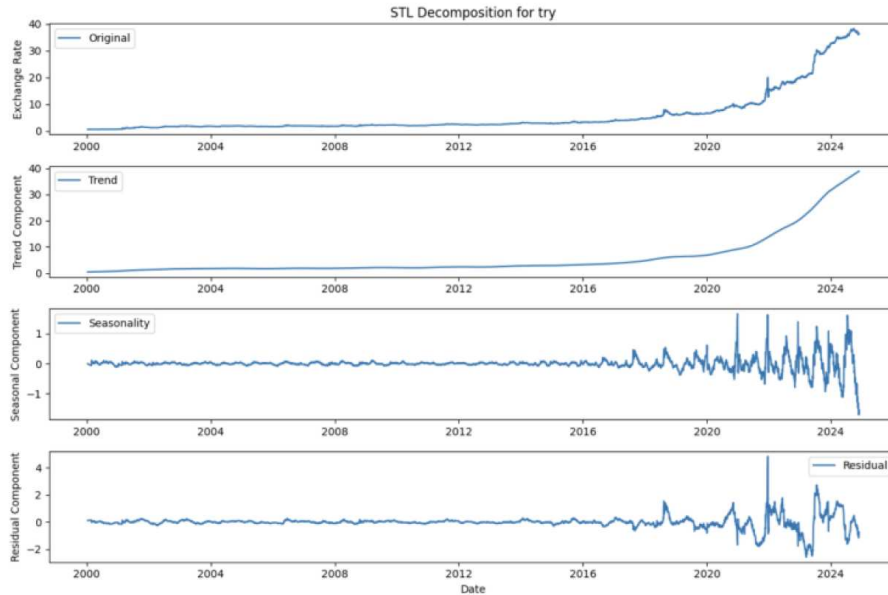


Figure 4. STL decomposition of the Turkish lira (TRY).

Finally, Figure 5 displays the decomposition for the US dollar (USD), which acts as a baseline of relative stability. Its trend shows moderate fluctuations, consistent with the dollar's role as a global safe-haven currency. Unlike emerging-market counterparts, its residual component is small and relatively stable through time.

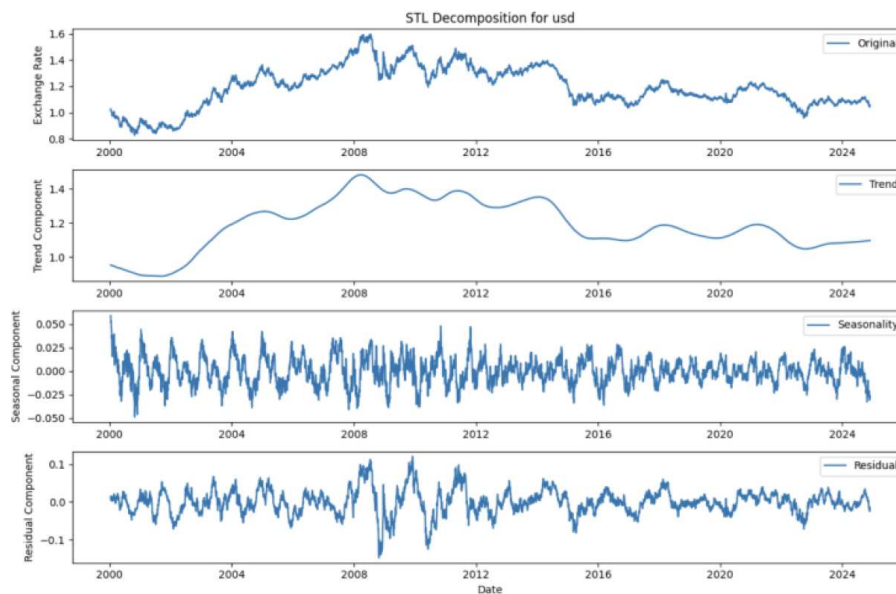


Figure 5. STL decomposition of the US dollar (USD).

5. Methods and analysis

This section presents the analytical framework adopted to examine currency co-movements through both classical statistical and topological approaches. The analysis proceeds in two main stages. First, traditional econometric tools are applied to the standardized time series to establish a statistical benchmark based on covariance, correlation, and clustering. Second, TDA is introduced to extract structural and nonlinear features from the same data using persistent homology and Wasserstein distances. By comparing the clustering outcomes obtained from these two fundamentally different representations, the study aims to evaluate whether TDA can reveal latent relationships in the FX market that are not captured by linear statistical measures.

5.1. Classical analysis

The classical component of the analysis establishes a benchmark based on well-known statistical measures of dependence and similarity. It focuses on covariance and correlation structures among currency returns and applies standard clustering algorithms, namely, k -means and hierarchical agglomerative clustering, to these statistical features. This provides a linear, interpretable baseline against which the added value of topological methods can later be assessed.

5.1.1. Covariance matrix

Covariance quantifies joint variability in both direction and magnitude, providing a first view of co-movements that can inform clustering. We compute a 13×13 covariance matrix on monthly log-returns (see Figure 6). As expected, self-pairs equal one. The largest off-diagonal covariance is for CNY-USD (≈ 0.96), implying closely aligned movements; this is consistent with China's management of the yuan relative to the dollar (Anstey, 2024). From a euro-based perspective, EUR-USD and EUR-CNY offer limited diversification if their fluctuations are highly similar.

CNY also covaries strongly with THB (0.75) and INR (0.72), potentially reflecting regional trade linkages; USD shows a similar pattern. The highest *average* covariance across pairs is for THB (0.474), followed by INR (0.471), suggesting broad alignment with other series (more global forces, fewer idiosyncratic shocks). At the other end, CHF-TRY is the only negative pair (-0.08), pointing to diversification potential. CHF generally exhibits low covariances (e.g., with RUB 0.02; AUD 0.05; BRL 0.05; ZAR 0.09), consistent with an independent policy stance and safe-haven role from a European vantage.

Covariance alone, however, conflates magnitude with direction and ignores temporal lead-lag. While standardizing variables makes the covariance numerically comparable to Pearson correlation, alternative correlation measures can reveal distinct aspects of dependence.

Remark 5.1 (On covariance versus correlation). Because all time series were standardized to zero mean and unit variance before computing the covariance matrix, the resulting values are numerically equivalent to Pearson correlation coefficients. Throughout this paper, the term “covariance” is used in a broad sense to emphasize joint variation, but readers should note that, under standardization, the two measures coincide up to a constant scaling factor. This does not affect any inference or clustering results but is important for conceptual precision.

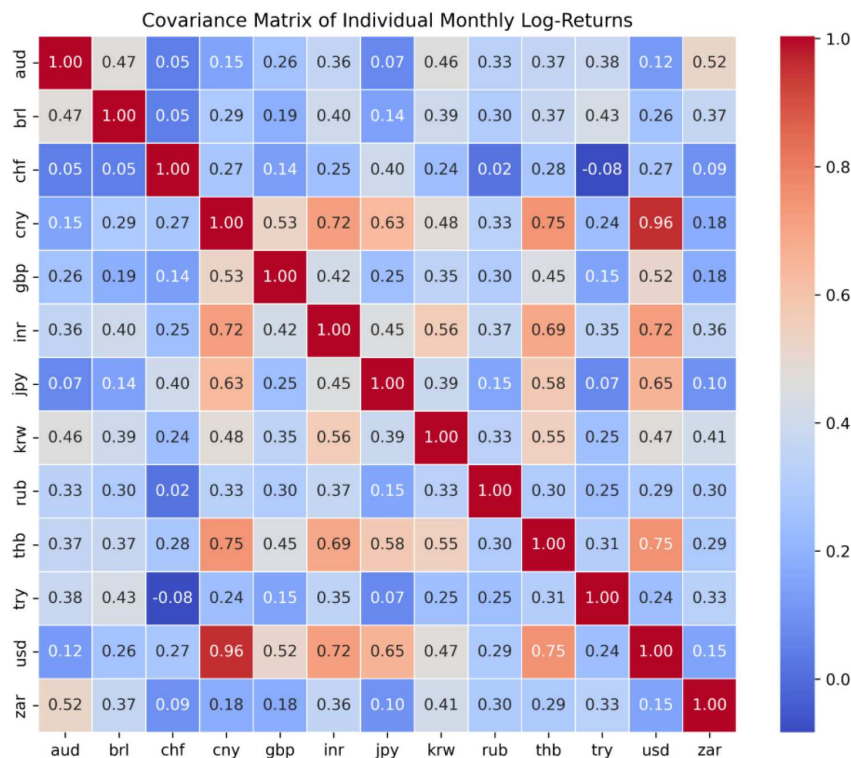


Figure 6. Covariance heatmap of monthly log-returns (13 currencies).

5.1.2. Correlation matrices

Pearson correlation assumes linearity, continuity, approximate normality, and contemporaneous relationships (Schober et al., 2018). We therefore also examine Spearman correlation (monotone but potentially nonlinear associations (De Winter et al., 2016) and *cross-correlation* with lags up to one month, which can capture delayed responses (Conlon et al., 2008). The cross-correlation matrix, illustrated in Figure 7, is most appropriate for monthly time series in this setting.

Most pairs behave similarly across covariance/correlation (see for example CNY-USD and CHF-TRY), but notable exceptions exist. AUD-JPY, for example, shows 0.07 covariance versus -0.12 cross-correlation, indicating that lagged responses can flip the apparent relationship: movements are not fully synchronized, and one series' shifts can be followed by the other moving in the opposite direction.

Remark 5.2 (On the interpretation of cross-correlations). The reported cross-correlation coefficients correspond to the maximum absolute correlation within a one-month lag window. While this approach captures short-term lead-lag effects, it can inflate dependence estimates when multiple testing across lags is unaccounted for. No formal significance testing was applied here, as the focus was exploratory; future analyses could assess lag significance via Bartlett's formula and adjust for multiple comparisons to obtain confidence bounds for the strongest correlations.

5.1.3. Clustering setup and metrics

We cluster currencies to detect naturally occurring groups based on behavioral similarity. We use two unsupervised algorithms: k -means and agglomerative hierarchical clustering. Cluster quality is evaluated with:

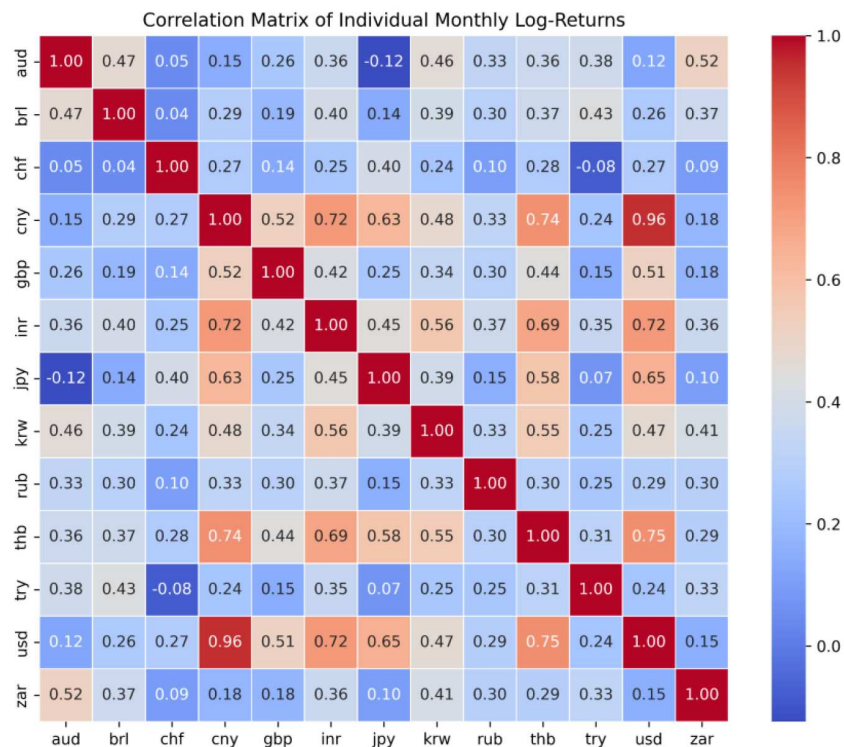


Figure 7. Cross-correlation heatmap of monthly log-returns (max lag 1 month).

- (i) the average Silhouette coefficient (range $[-1, 1]$; larger is better (Rousseeuw, 1987)) and
- (ii) the Calinski-Harabasz (CH) index, which balances between-cluster separation and within-cluster cohesion (Wang et al., 2021). CH is comparative rather than absolute.

***k*-means on statistical features.** Each currency is represented by its standardized monthly return vector. We select $k = 3$ via the elbow method (as in Bereta et. al.). For an illustration see Figure 8.

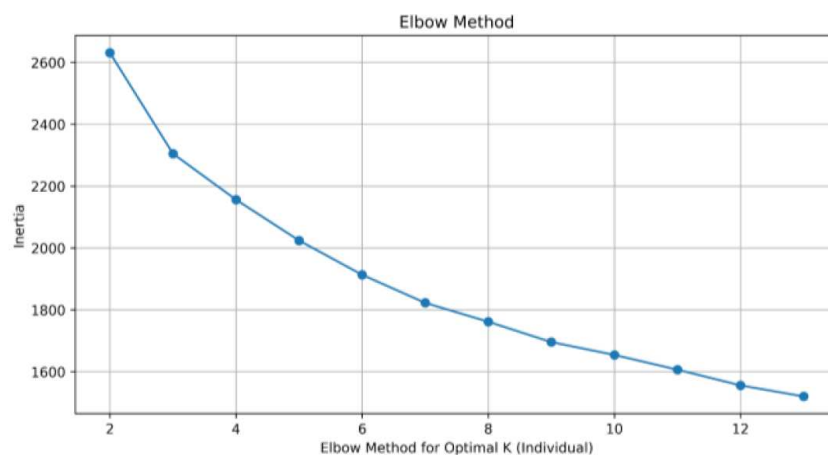


Figure 8. Elbow method for selecting k on statistical features.

Cluster 1 is a GBP singleton, consistent with idiosyncratic shocks (e.g., Brexit) and structural breaks relative to the euro. Cluster 2 mixes advanced (CHF, JPY, USD) and relatively stable emerging currencies (CNY, INR, KRW, THB), plausibly reflecting steadier distributions and, for some, managed regimes.

Cluster 3 (AUD, BRL, RUB, TRY, ZAR) exhibits lower correlations to others and trend components showing persistent depreciation/appreciation, often tied to domestic instability and commodity exposure. Scores: Silhouette 0.110; CH 2.657 (see table 1).

Table 1. Statistical k -means clusters (standardized return histories).

Cluster	Currencies
1	GBP
2	CHF, CNY, INR, JPY, KRW, THB, USD
3	AUD, BRL, RUB, TRY, ZAR

Hierarchical clustering on statistical features. We use complete linkage with Euclidean distance. The resulting dendrogram illustrated in Figure 9, differs from k -means: GBP is no longer a singleton; RUB appears boundary-like (late merge). INR/KRW/THB remain near CNY/USD, consistent with regional linkages.

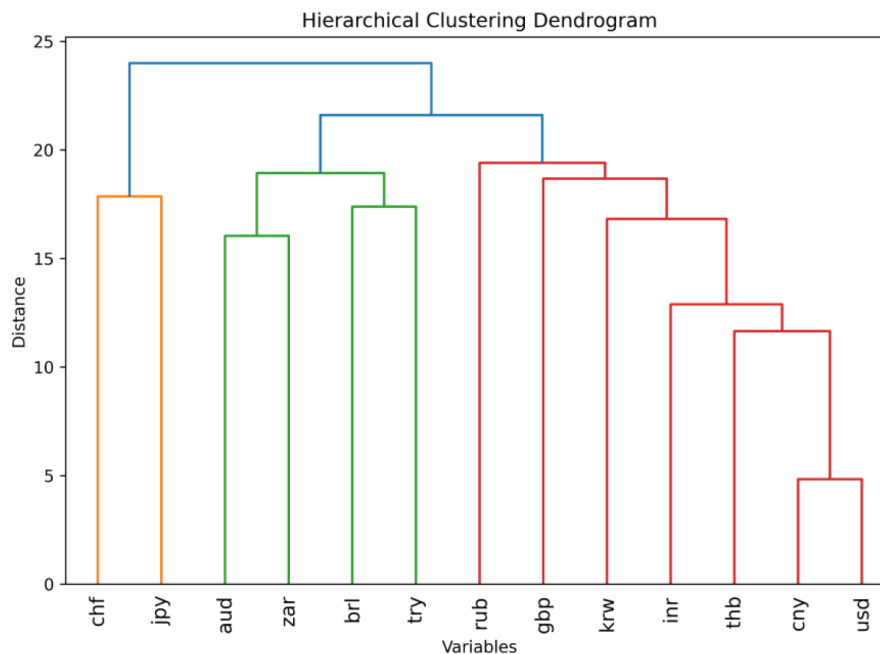


Figure 9. Hierarchical clustering.

CHF and JPY form a tight pair despite their cross-correlation sensitivity, aligning with their shared safe-haven role. Scores: Silhouette 0.111; CH 2.942.

5.2. Topological Data Analysis (TDA)

TDA characterizes the *shape* of data via persistent homology, tracking when topological features (components H_0 , loops H_1 , higher-order cavities) appear and disappear across a filtration scale. In our context, shapes can signal cycles, structural breaks, or regime switches not captured by linear contemporaneous measures.

5.2.1. Sliding-Window embedding and point clouds

Each univariate time series is mapped into \mathbb{R}^d using a sliding window with window length d and delay τ (Perea and Harer, 2013). Smaller windows capture local variation; larger windows can capture full cycles but risk noise in high dimensions. Following diagnostics (false nearest neighbours/autocorrelation, cf. Truong, 2017), most currencies favour $(d, \tau) = (4, 1)$; we adopt these for consistency across series and the pairwise set. This yields 13 point clouds (one per currency). Topological summaries are stacked and projected for visualization via Principal Component Analysis (PCA) (Figure 10).

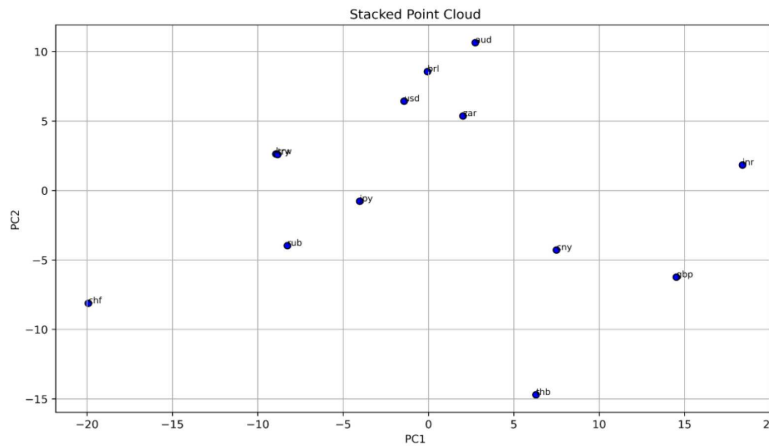


Figure 10. Two-dimensional visualization (PCA) of stacked topological summaries across currencies.

Remark 5.3. The embedding dimension and delay parameters are fixed across currencies to avoid introducing currency-specific tuning effects. While alternative choices or adaptive calibration strategies are possible, a single global configuration is deliberately adopted to preserve methodological comparability in this benchmark study.

5.2.2. Mathematical background: Persistent homology for embedded time series

Let $x_1, \dots, x_T \in \mathbb{R}$ denote a univariate time series (here: monthly log-returns for a currency). A sliding-window (delay) embedding with window length $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ produces a point cloud

$$X = \left\{ \mathbf{v}_t = (x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \in \mathbb{R}^d : t = (d-1)\tau + 1, \dots, T \right\},$$

equipped with a metric $d(\cdot, \cdot)$ (we use Euclidean). Intuitively, X encodes local temporal patterns as geometric structure in \mathbb{R}^d .

Vietoris-Rips filtration. For $\varepsilon \geq 0$, the Vietoris-Rips complex $\text{VR}_\varepsilon(X)$ is the abstract simplicial complex whose k -simplices are $(k+1)$ -point subsets of X with all pairwise distances $\leq \varepsilon$; increasing ε grows the complex:

$$\text{VR}_\varepsilon(X) \subseteq \text{VR}_{\varepsilon'}(X) \quad \text{for } \varepsilon \leq \varepsilon'.$$

The nested family $\{\text{VR}_\varepsilon(X)\}_{\varepsilon \geq 0}$ is a *filtration* that tracks how connectivity and higher-order relations appear as we thicken points into simplices (Edelsbrunner and Harer, 2010). For an illustration see Figure 11.

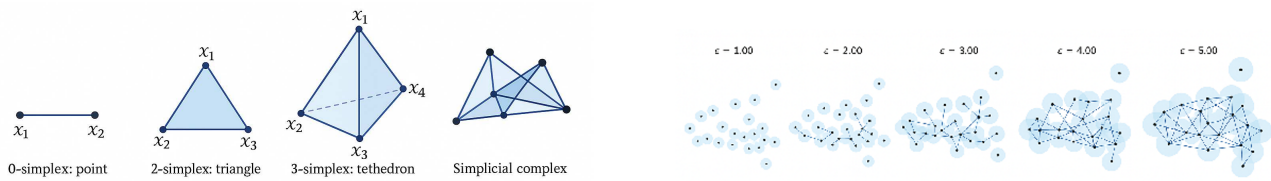


Figure 11. Topological building blocks and filtration process.

Homology and persistent homology. For each ε , the k -th homology group $H_k(\text{VR}_\varepsilon(X))$ (with coefficients in a field, e.g. \mathbb{Z}_2) is the quotient

$$H_k(\text{VR}_\varepsilon(X)) = \ker \partial_k / \text{im } \partial_{k+1},$$

where ∂_k are the simplicial boundary maps of the chain complex associated with $\text{VR}_\varepsilon(X)$. Elements of H_0 are connected components; elements of H_1 are loops (1-dimensional cycles), and higher H_k capture higher-dimensional voids.

Filtration inclusions induce linear maps on homology,

$$\iota_{\varepsilon \leq \varepsilon'} : H_k(\text{VR}_\varepsilon(X)) \longrightarrow H_k(\text{VR}_{\varepsilon'}(X)),$$

so $\{H_k(\text{VR}_\varepsilon(X)), \iota_{\varepsilon \leq \varepsilon'}\}$ forms a *persistence module*. Under mild finiteness assumptions (finite metric space X), this decomposes into interval modules; equivalently, each topological feature has a *birth* scale b and *death* scale d (Edelsbrunner and Harer, 2010). This yields:

- the *barcode*: a multiset of intervals $[b_i, d_i)$, and
- the *persistence diagram* $D_k(X)$: a multiset of points (b_i, d_i) in the birth–death plane.

Features farther from the diagonal ($d - b$ larger) are more *persistent* and typically regarded as more structurally meaningful.

Distances and stability. To compare shapes, we use metrics between diagrams, notably the p -Wasserstein distance W_p (or the bottleneck distance as $p \rightarrow \infty$), which computes the optimal matching cost between birth–death pairs (with the diagonal as a sink for unmatched mass). Crucially, persistence diagrams are *stable* with respect to perturbations of the underlying distances: small changes in X (e.g. noise, minor resampling) lead to small changes in $D_k(X)$ under W_p (Edelsbrunner and Harer, 2010). This robustness is especially desirable when analysing financial time series, which are inherently noisy and non-stationary.

Why this matters for FX dynamics. The delay-embedded cloud X turns *temporal behavior* into *geometry*. Persistent homology then interprets that geometry as topological structure:

- H_0 (components): long-lived components indicate *persistent segmentation* of local patterns (e.g. regime shifts or structural breaks where return patches stay distinct over a wide scale ε). Rapid H_0 merging suggests homogeneous local dynamics.
- H_1 (loops): persistent loops encode *recurrent* or *cyclic* behavior in the embedded patterns (e.g. seasonality, oscillatory mean-reversion, business-cycle effects). Short-lived loops signal weak or noisy cycles; long-lived loops indicate strong, stable cyclical structure.

In summary, persistent homology captures aspects of currency behavior that classical contemporaneous correlation cannot: (i) *shape* and recurrence of local trajectories, (ii) *scale* at which structures appear/disappear, and (iii) *robustness* of those structures to noise. In our application, pairwise Wasserstein distances between diagrams quantify how similar two currencies are in their *structural dynamics*, not merely in linear co-movements. Those distances are then used for clustering (directly for hierarchical clustering; via Euclidean embedding for k -means), providing a complementary partition to standard statistical features.

This mathematical formulation provides the foundation for the computational procedure described next, where we construct the persistence modules explicitly from the embedded currency trajectories using Vietoris-Rips filtrations.

5.2.3. Complexes and filtrations

We build Vietoris-Rips complexes on pairwise distances, growing simplices as the filtration threshold ε increases.

Alternative complexes (such as alpha or Delaunay–Rips) can more faithfully capture the geometric structure of the underlying point cloud because they rely on exact proximity relations rather than pairwise thresholds. However, these constructions become computationally demanding and numerically less stable in higher dimensions, where simplicial growth scales combinatorially (Edelsbrunner and Harer, 2010; Mishra and Motta, 2023).

5.2.4. Persistence diagrams and barcodes

Persistence diagrams plot (birth, death) of features; points far from the diagonal are more persistent, and thus, more structurally meaningful (see Figure 12). Barcodes provide an equivalent interval view (see Figure 13).

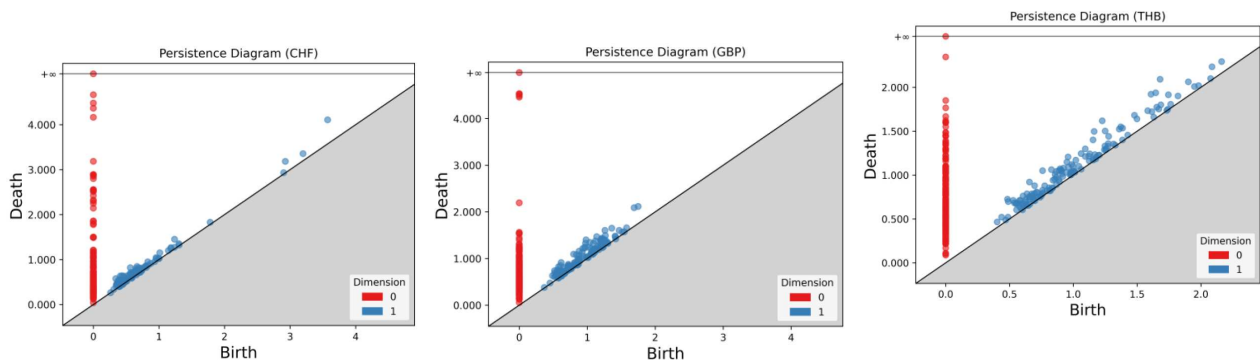


Figure 12. Persistence diagrams (H_0 red, H_1 blue) for CHF, GBP and THB.

GBP's H_0 deaths cluster near the diagonal (rapid merging), with a notable fragmentation episode (structural breaks/outliers). CHF shows an intermediate H_0 profile and multiple persistent H_1 loops (strong cyclicity). THB's features are short-lived and close to the diagonal, consistent with homogeneous/managed dynamics.

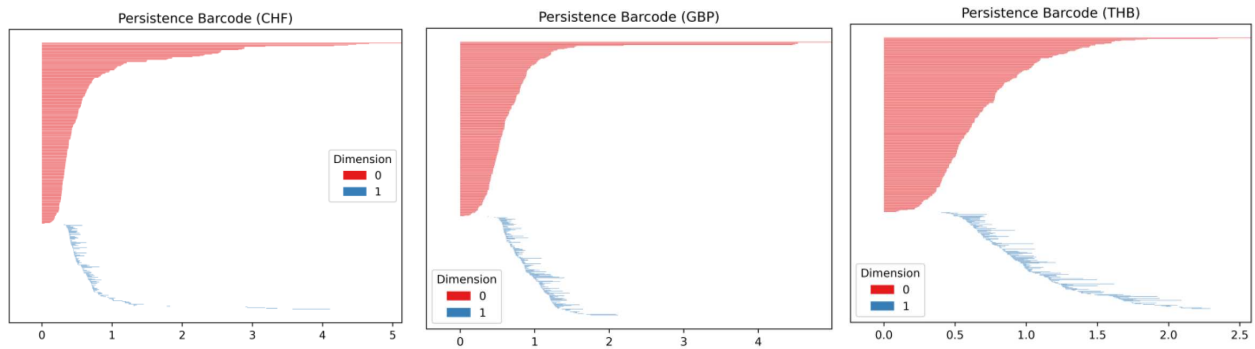


Figure 13. Persistence barcodes for CHF, GBP and THB.

5.2.5. Persistence landscapes and betti curves

Landscapes map persistence intervals to piecewise-linear “tents”, yielding vector features usable in statistical/ML models. We use three layers $\lambda_1, \lambda_2, \lambda_3$, a common choice for financial data (see Figure 14).

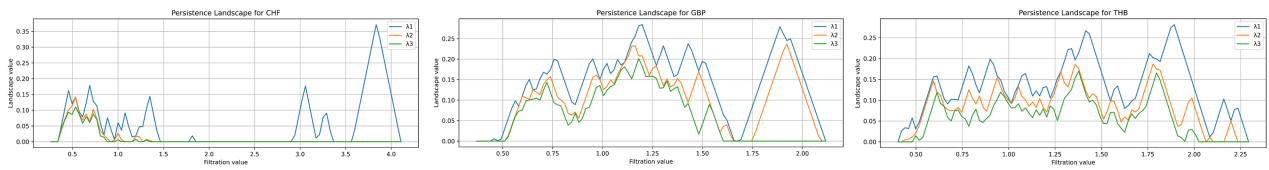


Figure 14. Persistence landscapes ($\lambda_1, \lambda_2, \lambda_3$) of CHF, GBP and THB.

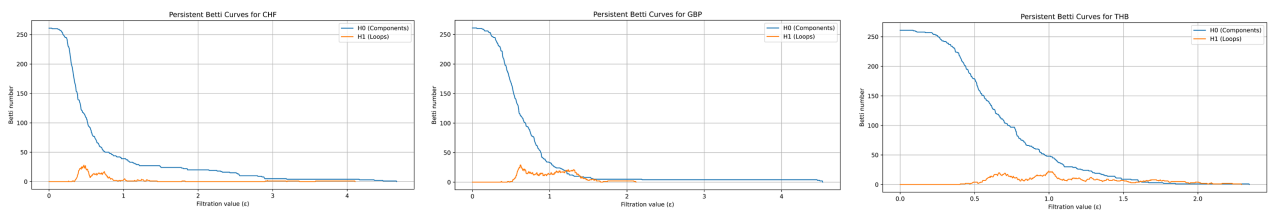


Figure 15. Betti curves (H_0 and H_1) for CHF, GBP and THB.

For CHF, a few dominant peaks suggest sparsity in highly persistent features (with higher peak at larger ε). GBP and THB show more medium-level peaks and lower maxima, indicating multiple moderately persistent structures.

Betti curves count the number of active features at each ε . CHF's H_0 drops steeply, then flattens; H_1 concentrates for $\varepsilon \in [0.5, 1.5]$ and vanishes near $\varepsilon \approx 4$. GBP's H_0 plateau arrives sooner; H_1 persists longer initially but dies near $\varepsilon \approx 2.1$. THB's H_0 and H_1 vanish around the same scale ($\varepsilon \approx 2.6$), with gradual ramps. For an illustration see Figure 15.

5.2.6. Distances and TDA clustering

Vectorising landscapes introduces Euclidean geometry that can distort comparisons between diagrams with different cardinalities. We therefore compute *Wasserstein* distances between persistence diagrams (Songdechakraiwut et al., 2021), matching birth-death pairs to measure minimal transport cost between topological structures. The resulting 13×13 distance matrix serves as input to clustering. We emphasize that in this paper, the role of persistent homology is operational: it defines a shape-based dissimilarity between currency return dynamics via Wasserstein distances between persistence diagrams of delay-embedded trajectories. The analysis is comparative rather than causal, and the term ‘topological insight’ refers specifically to structure captured by persistence-based similarity that is not encoded by contemporaneous linear dependence measures.

***k*-means on TDA.** The resulting TDA-based feature matrices were used directly for clustering under the same conditions as their statistical counterparts to ensure comparability. The evaluation metrics for all four clustering models are summarised in Table 2.

Table 2. TDA-derived *k*-means clusters.

Cluster	Currencies
1	GBP, INR
2	AUD, BRL, CNY, JPY, KRW, THB, TRY, USD, ZAR
3	CHF, RUB

Unlike the statistical case, no singleton emerges: GBP groups with INR, implying similar *shape*-based persistence despite distinct statistical profiles. Most currencies fall into Cluster 2 (moderate persistence cycles); Cluster 3 (CHF, RUB) features few but dominant persistent structures (CHF shown earlier; RUB in the appendix). Scores: Silhouette 0.191; CH 4.850.

Remark 5.4 (On distance geometry and embedding accuracy). The *k*-means clustering on TDA features required embedding the Wasserstein distance matrix into a Euclidean space via multidimensional scaling (MDS). Although the five-dimensional embedding preserved over 90% of pairwise variance in preliminary tests, some distortion of inter-currency distances is inevitable. A fully metric-preserving alternative would involve *k*-medoids clustering applied directly to the Wasserstein matrix, bypassing MDS but at higher computational cost. Future work will include such metric-based clustering to confirm the robustness of the present Euclidean embedding results.

Hierarchical clustering on TDA (direct Wasserstein). Hierarchical clustering (complete linkage) is applied directly to the Wasserstein matrix, avoiding embedding error.

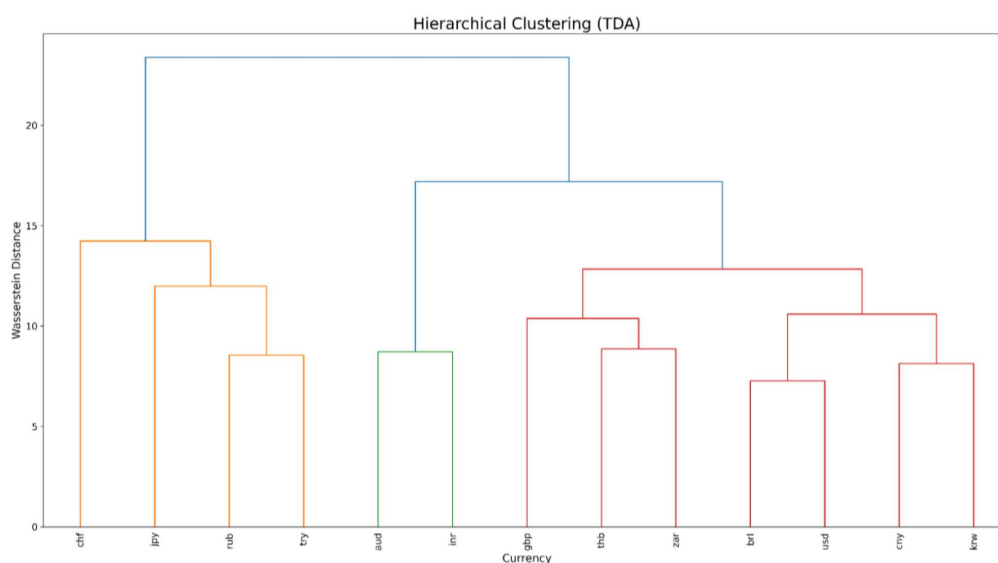


Figure 16. Hierarchical clustering on TDA features.

Here, clusters are more evenly distributed (see Figure 16). INR now groups with AUD (not GBP); CHF-RUB remain paired and are joined by TRY and JPY (RUB-TRY especially close). The contrast with TDA k -means highlights sensitivity to the Euclidean embedding step. Compared with statistical hierarchical clustering, linkage distances are more evenly spread, suggesting a more integrated structure under topology. Scores: Silhouette 0.182; CH 5.905.

5.2.7. Sensitivity analysis for embedding and filtration parameters

A well-known challenge in applying Topological Data Analysis to time series lies in the choice of embedding and filtration parameters. To assess whether our results are driven by a particular parameter configuration, we conducted targeted checks across a small range of embedding dimensions d , time delays τ , and filtration ranges ε_{\max} . These checks confirmed that the qualitative clustering structure and relative relationships between currencies remain stable under moderate parameter perturbations.

We emphasize that the objective of the present study is comparative benchmarking rather than exhaustive hyperparameter optimization. A full-scale robustness analysis, such as systematic evaluation across a broad parameter grid, bootstrap resampling of return series and consensus clustering, would substantially increase computational cost and is therefore left for future work. The use of a single, consistent parameter configuration across all currencies ensures comparability between statistical and topological representations, which is central to the study's design.

Remark 5.5 (On the economic interpretation of clusters). The clustering results primarily capture structural similarity in temporal behavior rather than explicit macroeconomic causality. Any economic interpretation is therefore illustrative rather than identificatory. Nevertheless, the identified groups align with well-established economic narratives: one cluster aggregates “safe-haven” currencies (CHF, JPY, USD); another groups emerging or policy-managed currencies (CNY, INR, KRW, THB); and a third

combines commodity-linked or more volatile economies (AUD, BRL, ZAR, TRY). These correspondences suggest that persistence-based similarity in the time-series domain can reflect underlying economic regimes, even in the absence of explicitly modeled macroeconomic variables.

Remark 5.6. A well-known challenge in applying Topological Data Analysis to time series lies in the choice of embedding and filtration parameters. To assess whether the results are driven by a particular configuration, targeted checks were conducted across a limited range of embedding dimensions, time delays, and filtration scales, confirming that the qualitative clustering structure remains stable under moderate perturbations. Accordingly, the present analysis evaluates relative clustering behavior under controlled perturbations rather than claiming exhaustive robustness or optimality of any specific parameter configuration.

6. Discussion

6.1. Key findings

The results reported below should be interpreted as evidence of complementary structure captured by persistence-based similarity, rather than as proof of statistical superiority or universal applicability of topological methods in FX analysis.

Table 3 aggregates the evaluation metrics across the four clustering specifications. In both cases (statistical features vs. TDA features), we report the average Silhouette coefficient and the Calinski-Harabasz (CH) index.

Table 3. Evaluation metrics across clustering models.

Model	Avg. Silhouette	Calinski-Harabasz
Statistical k -means	0.110	2.657
Statistical hierarchical	0.111	2.942
TDA-based k -means	0.191	4.850
TDA-based hierarchical	0.182	5.905

Overall, TDA-derived features improve clustering quality relative to classical statistical features. Both TDA models outperform their statistical counterparts on both metrics, with a more pronounced gain in the CH index, indicating tighter clusters that are better separated in the TDA feature space.

That said, Silhouette values remain modest (0.110–0.191). This is not unexpected: clustering FX reference-rate returns is intrinsically difficult due to overlapping regimes, cross-market spillovers, and heterogeneous reactions to macro shocks. The modest Silhouette values therefore reflect the problem’s inherent complexity rather than methodological failure.

Differences between k -means and hierarchical clustering become more pronounced in the TDA setting. This underscores (i) sensitivity to the geometry of the input space, in particular, the Euclidean embedding of Wasserstein distances for k -means, and (ii) the importance of distance/linkage choices. In practice, these results argue for reporting multiple specifications and emphasizing robust, recurring group structures rather than a single “one-true” partition.

These findings do not suggest that Topological Data Analysis supplants classical statistical methods in FX analysis. Rather, TDA provides an independent and structurally grounded perspective that

confirms and refines patterns identified by traditional approaches, highlighting where linear similarity aligns with deeper shape-based dynamics. In this sense, the present study establishes a clear benchmark for what can currently be expected from TDA in noisy financial settings and underscores the broader challenge of demonstrating genuine methodological superiority in applied contexts.

In sum, TDA features uncover complementary structure beyond that captured by contemporaneous statistical similarity. Because statistical and topological feature spaces encode fundamentally different aspects of dependence (directional similarity versus shape persistence), they should be viewed as *complements*, not substitutes, in empirical currency analysis.

Remark 6.1 (On robustness and sample size). The number of currencies analyzed ($n = 13$) is relatively small for clustering analyses, and the evaluation metrics (Silhouette and Calinski-Harabasz) can vary under small perturbations of the data. A full robustness check, such as bootstrap resampling of return series, recomputation of TDA features, and consensus clustering across replications, would provide additional evidence on cluster stability but was computationally prohibitive within the current framework. Nevertheless, exploratory tests using perturbed subsamples yielded qualitatively similar partitions, suggesting that the reported clusters are not artefacts of the particular sample. Future work may formally assess stability through resampling or the gap statistic, particularly as higher-frequency data become tractable for TDA computation.

6.2. Limitations

While the results are promising, several limitations must be acknowledged. First, the computation of persistent homology is resource-intensive, especially for higher-dimensional embeddings and large datasets. This computational cost limits the feasible granularity of inputs, such as sampling frequency or window length, and constrains the extent of experimentation across alternative parameter settings. Second, the analysis is inherently sensitive to parameter selection. The optimal sliding-window dimension and time delay (d, τ) can vary across currencies, yet a single global configuration was used here to maintain comparability. Although this ensures methodological consistency, it may not capture the most informative local dynamics for every time series.

A third limitation concerns the preprocessing choices applied to the data. The FX time series required resampling, aggregation to monthly frequency, and standardization, each of which can subtly affect the geometry of the embedded point clouds and, consequently, the resulting persistence diagrams and clusters. Finally, interpretation remains a significant challenge. While statistical features such as correlations or covariances have well-established economic meanings, the translation of topological structures, particularly the H_0 and H_1 features, into economic mechanisms is still an open research area. Bridging this gap between topological signatures and economic interpretation is crucial for advancing TDA as a practical analytical tool in finance.

6.3. Future prospects

The findings of this study open several promising avenues for future research. First, applying Topological Data Analysis to higher-frequency data, such as daily or weekly exchange rate series, could uncover short-term dependencies and regime switches that monthly aggregates tend to smooth out. Such extensions would provide a more granular view of how topological features evolve in response to rapid market events, including policy announcements or financial shocks.

A second direction involves experimenting with alternative filtrations. While the Vietoris-Rips complex was chosen here for its balance between generality and computational feasibility, other constructions such as alpha or Delaunay-Rips complexes may yield improved geometric fidelity, albeit at a higher computational cost. Similarly, exploring multi-parameter filtrations could help disentangle the overlapping temporal and structural dynamics inherent in financial data.

From a computational perspective, future studies would benefit from scalable pipelines that leverage approximate persistent homology, GPU acceleration, or sketching-based algorithms. These tools could enable richer sensitivity analyses across a broader range of window sizes, delays, and filtration parameters, thereby improving the robustness of results.

Beyond methodological refinements, future work should extend the framework to other asset classes, including commodities and cryptocurrencies. These assets exhibit high-frequency volatility, sudden structural shifts, and complex interdependencies that may not be adequately captured by conventional correlation-based methods. The application of TDA to cryptocurrency time series could therefore reveal distinctive topological signatures associated with market sentiment, liquidity regimes, and speculative phases. This extension would also allow for comparative studies between decentralised and centralised monetary systems, offering valuable insights into the geometry of digital asset co-movements.

Finally, a fruitful potential direction of this research lies in coupling topological features with economic variables, such as trade exposure, interest rate differentials, or policy shocks, to strengthen interpretability and develop causal narratives linking geometric structures to underlying economic mechanisms.

6.4. Practical implications

From a practical standpoint, the findings of this study offer potential value for both market participants and policymakers. For portfolio managers and risk managers, TDA-based clustering offers an alternative way to group currencies for diversification and stress testing, particularly in environments where linear correlations shift rapidly or break down. Currencies that share similar persistence-based profiles may respond to global shocks in comparable ways even when contemporaneous correlations appear weak.

For central banks and regulators, the ability of TDA to identify persistent structural dependencies suggests a complementary tool for monitoring systemic vulnerability and potential contagion across currency blocs. A natural extension of this work is to study time-varying topological distances over rolling windows and to relate changes in persistence structure to established spillover or contagion metrics.

Overall, the integration of topological features into standard econometric toolkits has the potential to strengthen FX risk assessment frameworks and to contribute to a deeper understanding of global monetary interdependence.

7. Conclusions

The main objective of this study was to investigate whether Topological Data Analysis (TDA) can reveal additional structure and insights into the co-movements of currencies in the foreign exchange (FX) market beyond what is captured by traditional statistical methods. To this end, two established clustering algorithms, k -means and hierarchical clustering, were applied to both traditional statistical features (such as correlations and covariances) and to TDA-derived features (based on persistent homology). This

resulted in four clustering models, whose performance was evaluated through the Silhouette coefficient and the Calinski-Harabasz index.

Empirically, the findings show that TDA-based clustering models outperform their statistical counterparts on both metrics, with the improvement being particularly evident in the Calinski-Harabasz index. This suggests that clusters derived from topological features are more compact and well separated, indicating that TDA captures additional structural dependencies that are not readily visible in conventional linear approaches. However, the Silhouette coefficients remain modest even for TDA-based models, reflecting the inherent complexity of the FX reference rate dynamics, where overlapping regimes and cross-market influences make clear-cut partitions difficult.

The differences between the two clustering algorithms become more pronounced in the TDA context, further illustrating that topological embeddings and distance metrics can strongly influence clustering outcomes. This sensitivity underscores both the power and the fragility of TDA: while it offers a new lens for revealing latent geometry in time series data, it also requires careful calibration of parameters such as window size, delay, and filtration thresholds. Moreover, the computational demands of persistent homology limited the granularity of the analysis, and the lack of universally optimal parameters constrained comparability across currencies.

From a methodological standpoint, this work extends the use of persistent homology and topological summaries into the domain of FX analysis, a field where applications of TDA remain scarce. From an empirical standpoint, it demonstrates that the shape-based representation of currency dynamics provides complementary information to that extracted from linear measures, enhancing our understanding of global currency interdependence.

In a broader sense, this study highlights the potential of TDA as a complementary tool in financial analytics. Its robustness to noise, invariance to coordinate systems, and ability to capture higher-order nonlinear dependencies make it particularly promising for modeling the evolving and complex structure of international financial markets. Future research should further explore high-frequency and multiscale FX data, integrate topological features with economic fundamentals, and develop scalable computational pipelines to unlock the full potential of topological methods in finance.

Author contributions

P.F.J. performed the data collection, coding, and initial empirical analysis. I.D. developed the conceptual and methodological framework and supervised the research. Both authors contributed to the interpretation of the results and to writing and revising the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

- Aguilar A, Ensor K (2020) Topology data analysis using mean persistence landscapes in financial crashes. *J Mathematical Financ* 10: 648–678. <https://doi.org/10.4236/jmf.2020.104038>
- Andersen TG, Bollerslev T, Diebold FX, et al. (2003) Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. <https://doi.org/10.1111/1468-0262.00418>
- Anstey C (2024) Bloomberg New Economy: The promise and peril of China’s strong yuan policy. Bloomberg. Available from: <https://www.bloomberg.com/news/newsletters/2024-02-03/bloomberg-new-economy-the-promise-and-peril-of-china-s-strong-yuan-policy>.
- Bereta P, Diamantis I (2025) The shape of consumer behavior: A symbolic and topological analysis of time series. *arXiv*. <https://doi.org/10.48550/arXiv.2506.19759>
- Berthold MR, Höppner F (2016) On clustering time series using Euclidean distance and Pearson correlation. *arXiv*. <https://doi.org/10.48550/arXiv.1601.02213>
- Conlon T, Ruskin H, Crane M (2008) Cross-correlation dynamics in financial time series. *Physica A* 388: 705–714. <https://doi.org/10.1016/j.physa.2008.10.047>
- Das P, Barman S (2025) Perspective chapter: An overview of time series decomposition and its applications. In: *Applied and Theoretical Econometrics and Financial Crises*. <https://doi.org/10.5772/intechopen.1009268>
- De Winter JCF, Gosling SD, Potter J (2016) Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychol Methods* 21: 273–290. <https://doi.org/10.1037/met0000079>
- De Amorim LB, Cavalcanti GD, Cruz RM (2022) The choice of scaling technique matters for classification performance. *Appl Soft Comput* 133: 109924. <https://doi.org/10.1016/j.asoc.2022.109924>
- Drożdż S, Górski AZ, Kwapien J (2007) World currency exchange rate cross-correlations. *Eur Phys J B* 58: 499–502. <https://doi.org/10.1140/epjb/e2007-00246-8>
- Edelsbrunner H, Harer J (2010) Computational topology: An introduction. Springer. https://doi.org/10.1007/978-3-540-33259-6_7
- El-Yaagoubi AB, Chung MK, Ombao H (2023) Topological data analysis for multivariate time series data. *Entropy* 25: 1509. <https://doi.org/10.3390/e25111509>
- European Central Bank (2023) Framework for the euro foreign exchange reference rates. Available from: <https://www.ecb.europa.eu/stats/exchange/eurofxref/html/index.en.html>.
- Garcia JS (2022) Applications of topological data analysis to natural language processing and computer vision. Master’s thesis, Colorado State University.
- Hobbelhagen F, Diamantis I (2024) A comparative study of symbolic aggregate approximation and topological data analysis. *Quant Financ Econ* 8: 705–732. <https://doi.org/10.3934/QFE.2024027>
- Italian Institute for International Political Studies (2022) Trust, uncertainty and disinformation in the Brazilian elections. Available from: <https://www.ispionline.it/en/publication/trust-uncertainty-and-disinformation-brazilian-elections-36310>.

- Lum PY, Singh G, Lehman A, et al. (2013) Extracting insights from the shape of complex data using topology. *Sci Rep* 3: 1236. <https://doi.org/10.1038/srep01236>
- Mai Y, Chen H, Zou J, et al. (2017) Currency co-movement and network correlation structure of the foreign exchange market. *Physica A* 492: 65–74. <https://doi.org/10.1016/j.physa.2017.09.068>
- Majumdar S, Laha AK (2020) Clustering and classification of time series using topological data analysis with applications to finance. *Expert Syst Appl* 162: 113868. <https://doi.org/10.1016/j.eswa.2020.113868>
- Mantegna R (1999) Hierarchical structure in financial markets. *Eur Phys J B* 11: 193–197. <https://doi.org/10.1007/s100510050929>
- Mattera R, Scepi G, Kaur P (2025) Time series clustering for high-dimensional portfolio selection: A comparative study. *Soft Comput* 29: 4219–4231. <https://doi.org/10.1007/s00500-025-10656-2>
- Mishra A, Motta FC (2023) Stability and machine learning applications of persistent homology using the Delaunay-Rips complex. *Front Appl Math Stat* 9. <https://doi.org/10.3389/fams.2023.1179301>
- Müllner D (2011) Modern hierarchical, agglomerative clustering algorithms. *arXiv*. <https://doi.org/10.48550/arXiv.1109.2378>
- Munkres JR (2018) Elements of algebraic topology. CRC Press. <https://doi.org/10.1201/9780429493911>
- Murtagh F, Legendre P (2014) Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *J Classif* 31: 274–295. <https://doi.org/10.1007/s00357-014-9161-z>
- Bank for International Settlements (2022) OTC foreign exchange turnover in April 2022. Available from: https://www.bis.org/statistics/rpfx22_fx.htm.
- Paparrizos J, Yang F, Li H (2024) Bridging the gap: A decade review of time-series clustering methods. *arXiv*. <https://doi.org/10.48550/arXiv.2412.20582>
- Penn State Eberly College of Science (n.d.) Stat 555: Clustering. <https://online.stat.psu.edu/stat555/node/88>
- Perea J, Harer J (2013) Sliding windows and persistence: An application of topological methods to signal analysis. *arXiv*. <https://doi.org/10.48550/arXiv.1307.6188>
- Pierini M (2021) Why is Turkey's president cutting interest rates, spurring inflation and lowering the value of the lira? Carnegie Endowment for International Peace. Available from: <https://carnegieendowment.org/diwan/2021/12/why-is-turkeys-president-cutting-interest-rates-spurring-inflation-and-lowering-the-value-of-the-lira?lang=en>.
- Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20: 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Schober P, Boer C, Schwarte LA (2018) Correlation coefficients: Appropriate use and interpretation. *Anesth Analg* 126: 1763–1768. <https://doi.org/10.1213/ane.0000000000002864>
- Shultz C (2023) Applications of topological data analysis in economics. Preprint. <https://doi.org/10.2139/ssrn.4378151>
- Songdechakraiwut T, Krause BM, Banks MI, et al. (2021) Fast topological clustering with Wasserstein distance. *arXiv*. <https://doi.org/10.48550/arXiv.2112.00101>

-
- Theodosiou M (2011) Forecasting monthly and quarterly time series using STL decomposition. *Int J Forecast* 27: 1178–1195. <https://doi.org/10.1016/j.ijforecast.2010.11.002>
- Truong P (2017) An exploration of topological properties of high-frequency one-dimensional financial time series data using TDA. Master's thesis, KTH Royal Institute of Technology. Available from: <https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1169943>.
- Verma A, Buonocore RJ, Di Matteo T (2018) A cluster-driven log-volatility factor model: A deepening on the source of the volatility clustering. *Quant Financ* 19: 981–996. <https://doi.org/10.1080/14697688.2018.1535183>
- Vijaya N, Sharma S, Batra N (2019) Comparative study of single linkage, complete linkage, and Ward method of agglomerative clustering. In: *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, 568–573. <https://doi.org/10.1109/comitcon.2019.8862232>
- Wang Y, Xu Y, Gao T (2021) Evaluation method of wind turbine group classification based on Calinski-Harabasz. In: *Proceedings of the IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, 2630–2635. <https://doi.org/10.1109/ei252483.2021.9713300>
- Watts J (2016) Brazil's Dilma Rousseff impeached by Senate in crushing defeat. *The Guardian*. Available from: <https://www.theguardian.com/world/2016/aug/31/dilma-rousseff-impeached-president-brazilian-senate-michel-temer>.
- Wongoutong C (2024) The impact of neglecting feature scaling in k-means clustering. *PLoS One* 19: e0310839. <https://doi.org/10.1371/journal.pone.0310839>
- Yim O, Ramdeen KT (2015) Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data. *Quant Methods Psychol* 11: 8–21. <https://doi.org/10.20982/tqmp.11.1.p008>

Appendix A. Reproducibility and Computational Framework

To ensure full reproducibility, all data analyzed in this study were obtained from the publicly accessible ECB Statistical Data Warehouse. Specifically, the *Euro foreign-exchange reference rates* datasets were used, where each currency is represented as a separate daily time series with the euro (EUR) serving as the base currency.

The raw datasets were concatenated into a unified time series panel comprised of 13 currencies against the euro. Missing entries were checked and linearly interpolated when required to maintain continuity. To facilitate comparability, the data were resampled to a monthly frequency and transformed into logarithmic returns:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right),$$

where P_t denotes the reference rate at month t . The euro itself was excluded from the dataset due to its fixed numeraire role.

All computations were executed in Python 3.12.12 using open-source scientific packages, ensuring methodological transparency and replicability. The core libraries employed were:

- `pandas`, `numpy`, `scipy`, and `statsmodels` for data manipulation and statistical analysis,
- `scikit-learn` for clustering algorithms and performance evaluation,
- `gudhi`, `ripser`, and `persim` for topological data analysis and persistent homology computation,
- `tqdm` for pipeline monitoring and `mplfinance` and `matplotlib` for graphical rendering.

The entire workflow was designed to rely exclusively on publicly available data and open-source software, thus facilitating full reproducibility.

Appendix B. Persistence Diagrams of the Currencies

This appendix presents the complete set of persistence diagrams for all 13 currencies analyzed in the study, complementing the representative examples (CHF, GBP, THB) shown in Section 5.2.4. Each diagram illustrated Figure 17 summarizes the birth and death of homological features (connected components and loops) obtained through persistent homology of the time-delay embedded point clouds, computed via the Vietoris-Rips filtration.

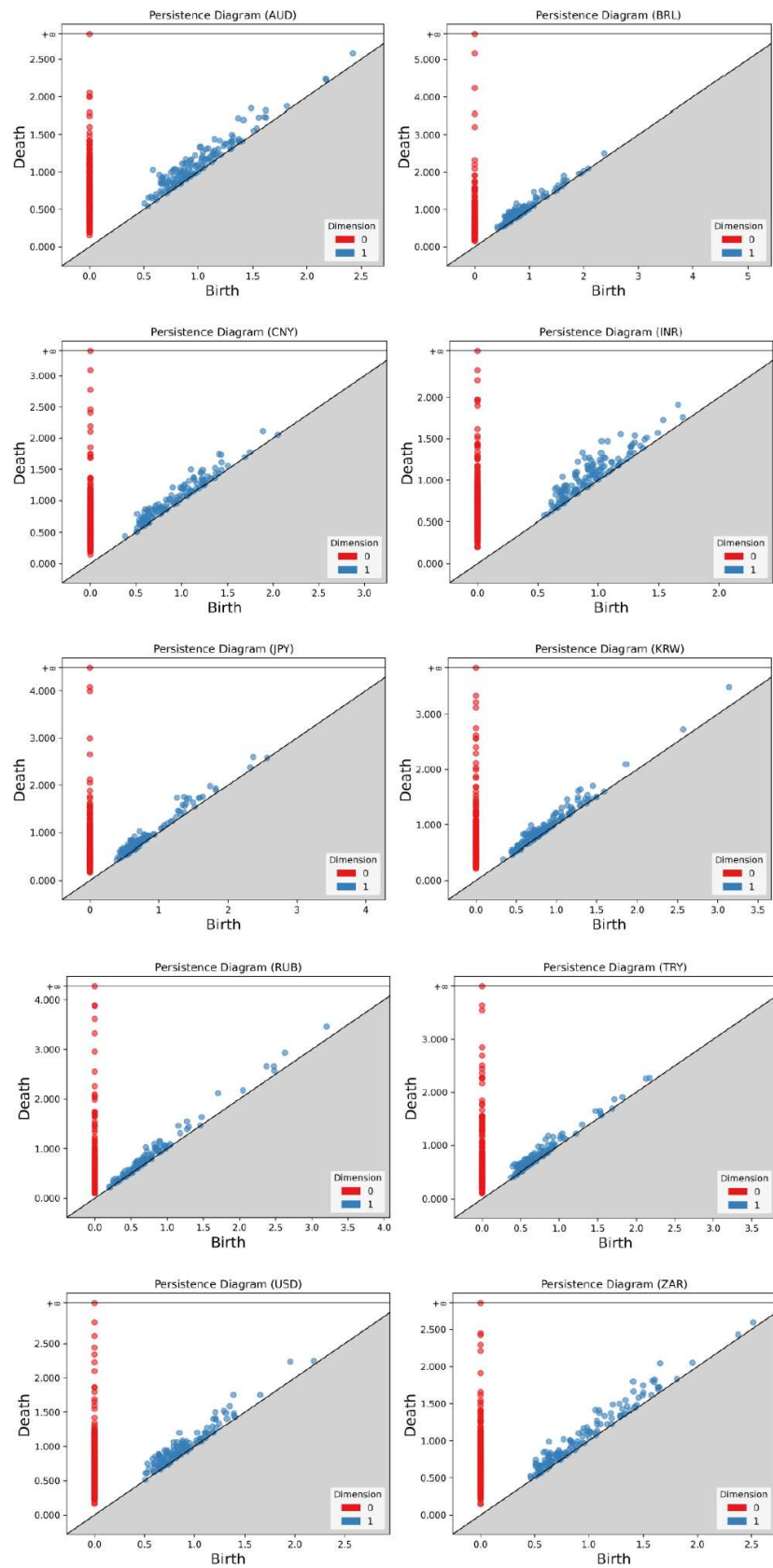


Figure 17. Persistence diagrams of the currencies.

All persistence diagrams were computed using a uniform filtration range to ensure comparability across currencies. Points above the diagonal represent persistent topological features, with red markers denoting dimension 0 (connected components) and blue markers denoting dimension 1 (loops).

These diagrams provide the topological foundation for the Wasserstein distance matrix used in the TDA-based clustering analyses (Section 5.2.4). The complete collection of diagrams is included as supplementary material to this paper.



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)