REVISTA MEXICANA DE ECONOMÍA Y FINANZAS
Nueva Época
REMEF
(THE MEXICAN JOURNAL OF ECONOMICS AND FINANCE)

Revista Mexicana de Economía y Finanzas, Nueva Época

Volumen 21 Número 1, enero - marzo 2026, pp. 1-29, e1495

DOI: https://doi.org/10.21919/remef.v21i1.1495

(Received: May 17, 2024; Accepted: October 13, 2025; Published: January 1, 2026)

Asset Representativeness in Mexican Stock Market Sectors: A Principal Component Analysis (2020–2024)

Josué Alan Cantú Esquivel [1] - Instituto Politécnico Nacional, México
Yorka Veruska Arteaga Nagashiro - Instituto Tecnológico Autónomo de México, México
Silvana Teresa Simbrón Arteaga - Tecnológico de Monterrey, México

Abstract

Our work examines the Mexican capital market with the objective of identifying the key assets within each sector. This is accomplished through a principal component analysis (PCA) applied to time series data. The findings elucidate the assets that predominantly contribute to each sector concerning overall variability and illustrate the associations among them according to their correlation vectors. Ultimately, the linear combinations of each sector, referred to as the principal components, are provided, which serve as indicators of their cyclical behavior. These combinations can be utilized as measures of sector dynamism within the market and, in conjunction with other technical analysis tools, may prove valuable as trading signals and for constructing investment portfolios. These insights will be extended for future research applications.

JEL Classification: C10, C15, C22, C32, C46, D81, G11, G15.
Keywords: Principal Component Analysis, Time Series, Stocks, Exchange Market.

Representatividad de los activos en los sectores del mercado bursátil mexicano: un análisis de componentes principales (2020–2024)

Resumen

Nuestra investigación analiza el mercado de capitales mexicano con el objetivo de identificar los activos guía de cada sector. Esto se lleva a cabo mediante un análisis de componentes principales (ACP) en series de tiempo. Los resultados revelan los activos que más contribuyen a cada sector en términos de la variabilidad global en cada uno, además de mostrar la asociación entre estos según sus flechas de correlación. Finalmente, se presentan las combinaciones lineales de cada sector (los componentes principales), que son indicativas del comportamiento cíclico de estos. Tales combinaciones pueden utilizarse como indicadores del dinamismo de los sectores en el mercado, y junto con otras herramientas de análisis técnico, pueden resultar útiles como indicadores de *trading* y portafolios de inversión que extendemos para futuras aplicaciones en líneas de investigación.

Clasificación JEL: C10, C15, C22, C32, C46, D81, G11, G15.
Palabras clave: Análisis de componentes principales, series de tiempo, acciones, bolsa de valores.

[1] Corresponding Author. Email: jcantue@ipn.mx
*No funding source was received for the development of this research.

1. Introduction

Mexico is considered an emerging economy, and its stock market is represented by the IPC BMV/MEXBOL index. This index reflects the general sentiment of the 30-35 most active companies during Mexican market sessions. However, how the index is constructed can make it challenging for economists, portfolio managers, and risk managers to analyze specific stock market dynamics among the 11 business sectors operating in the exchange.

The lack of clarity in monitoring the growth of individual business sectors within the MEXBOL can make it difficult for traders and portfolio managers to identify valuable market prospects when a sector-specific indicator is absent. This limitation can hinder the ability to identify the most valuable assets and pinpoint market opportunities, creating significant obstacles when constructing investment portfolios tailored to specific economic sectors in the current business cycle. According to Portelli & Roncalli (2024), this also applies to analyzing stocks and bonds as investment opportunities. Therefore, it is essential to have a robust and reliable system to track and analyze market movements for each sector and identify leading assets within each sector.

Cantú et al. (2025), as well as Avellaneda & Lee (2008), state that Principal Component Analysis (PCA) can be a reliable method for analyzing the role and contribution of each variable in an information system for trading strategies. This technique can be useful for studying economic activity by sector, as it allows for the identification of which stocks contribute more to the dynamism of the capital market. Doing so can help create precise investment strategies and make informed decisions based on economic activity. Therefore, our research hypothesis suggests that it is feasible to develop an indicator of stock market activity in the Mexican capital market using PCA. In this context, the main objective of our work is to determine, through PCA, the stocks by sector of activity that exert a significant influence on the movement of the capital market, segmented by activity sectors on the Mexican Stock Exchange. We propose that this can be achieved through PCA. To achieve this, we recommend grouping assets by sector and utilizing variance decomposition analysis to construct linear combinations of assets within each sector. This approach aims to address the absence of a benchmark for business sectors. Additionally, we can analyze the assets that have the most impact on the stock market indicator for each sector in the Mexican Exchange using PCA on time-series data. This will help us pinpoint the stocks that have a greater influence on the behavior of each indicator, enabling us to make more informed investment decisions for future research endeavors.

2. Literature review

The IPC BMV/MEXBOL constitutes a significant index. It provides valuable insights into the performance and expectations of the Mexican financial market, rendering it a vital indicator of investor confidence, corporate financing, and the perception of public companies within the Mexican economy. Research studies conducted by Coronado, Martínez & Venegas (2022) and Gil-León et al. (2019) have highlighted the crucial role of the Exchange in a country's economic growth. Their works point out that these markets provide liquidity and capital to companies, which is essential for their expansion and development.

2.1. The significance of the stock market within the Mexican economy

Given that the Mexbol provides a comprehensive view of the behavior of the stock market in Mexico, its relevance to the Mexican economy can be attributed to several key aspects:

1. The IPC BMV is a market sentiment indicator that gauges investors' confidence in the major businesses and economic sectors of the Mexican economy. As per Arango et al. (2019), an increase in this benchmark may indicate confidence in the public enterprises' economic health of the country, while a decline might suggest apprehension among market participants about their economic activities.
2. Bosch (2023) explains that companies can use the Exchange to issue stocks and bonds to obtain financing. Thus, a robust stock market and its benchmarks are essential for corporate funding and economic growth through financing the enterprises in a country. Gavira Durón et al. (2020) further support this notion, suggesting that markets are becoming increasingly competitive, and the pace of these changes is exerting significant pressure on companies. Not only must they strive for success, but they also need to plan and implement strategies to sustain their success in the future. Therefore, corporate sustainability is crucial.
3. Attraction of foreign investment: As noted by Ruiz et al. (2016), a solid and developed stock market can attract foreign investors, encouraging foreign investment to contribute to the economic growth of firms and the internationalization of the economy. Also, Galván et al. (2017) state that foreign investors often consider a country's stock market performance to measure its capital stability, risks, and trade opportunities. In this sense, having a reliable stock market indicator for each economic activity can make Mexico an attractive option for foreign investments, leading to positive outcomes in the economic sectors.
4. A rising stock market is often associated with a growing economy. Santillán et al. (2018) noted that while the Mexbol index mainly reflects the performance of listed companies, its overall movement generally indicates the health and confidence of the economy. Additionally, Bosch (2023) and Cantú et al. (2025) suggest that the growth and development of stock markets offer various benefits for a country's economic progress.
5. Wealth Effect: Langebaek and Ortiz (2007) state that the behavior of the IPC can influence investors' and consumers' wealth perception. Consequently, a bull market can lead to a "wealth effect", where investors feel more financially secure and are more likely to invest and spend in other financial products. This can have positive implications for consumption and investment in the overall economy.
6. Job creation: Financing allows public companies to expand and grow, which can result in both direct and indirect jobs, as noted by Ruiz and Steinwascher (2007). When combined with good corporate governance, a thriving stock market can contribute to the development of the business sector and job creation at a macroeconomic level.

All these points state that it is crucial to have a stock market activity indicator that provides information on companies' financial market performance to promote economic development and help companies raise capital. Rueda (1980) explains that stock markets allow investors to contribute capital to firms for growth and expansion by purchasing stocks or bonds. This process provides liquidity to the market and benefits investors in the medium to long term; as Bogle (2007) explains, investing in stocks and corporate bonds also promotes long-term savings, economic competitiveness, and development in a country.

2.2. PCA and Its Applications in Asset Pricing and Market Analysis

Within the context of the Mexican stock exchange, Principal Component Analysis (PCA) has the capacity to condense information from numerous stocks into a limited number of factors that effectively capture the market mode and sectoral drivers. This approach is consistent with the extraction of common factors in bonds and equities, as demonstrated by Litterman and Scheinkman (1991) and Connor and Korajczyk (1988), as well as with high-dimensional factor models estimated through PCA (Aït-Sahalia & Xiu, 2017). For markets characterized by aggregate shocks and sector rotation, the first principal component offers a synthetic indicator of market dynamics, which proves useful for tracking trends and nowcasting purposes.

Furthermore, Principal Component Analysis (PCA) is instrumental in denoising correlation matrices and differentiating between signal and noise, including cluster analysis. The literature on Random Matrix Theory demonstrates that leading eigenvalues encapsulate common structures and assist in filtering residual noise (Laloux et al., 1999; Plerou et al., 2002). Recent research confirms that principal components represent market regimes and transitions, thereby affirming PCA as a valuable diagnostic and monitoring instrument.

For enhanced representativeness and investment decision-making, Principal Component Analysis (PCA) facilitates the identification of stocks that effectively explain an index or sector, as well as methods for constructing portfolios linked to the first eigenvector. Previous research examines the composition of PC1 within equity indices and cross-market comparisons (Nguyen & Nguyen, 2019), and advocates for principal eigen-portfolios directly derived from principal components (Avellaneda et al., 2022). Additionally, it explores the theoretical properties of PC-based portfolios, which are valuable for index tracking and systematic tilts (Severini, 2022).

Finally, Principal Component Analysis provides insights not only into the common factor but also into residual dimensions characterized by economic content—referred to as "last" components—that encapsulate specific risks or correlation structures. These are valuable for diversification and concentration risk management (Yang, Rea, & Rea, 2017).

Galván, García, and Serna (2017) assert that the Mexican capital market provides a wide range of financial instruments such as stocks, government and corporate bonds, financial derivatives, Exchange-Traded Funds (ETFs), Certificates of Capital Development (CKDs), and Real Estate Investment Trusts (REITs/FIBRAs) that investors can utilize to diversify their portfolios and manage risks more efficiently. This diversification allows investors or entities to better manage their risk appetite by aligning their investment opportunities with their objectives and risk tolerance. This implies that creating an individual indicator for each sector can assist investors in constructing better equity portfolios and making informed buy or sell decisions.

Therefore, it is imperative that, in addition to possessing a generalized sentiment indicator for the Mexican market (such as the Mexbol), specific indicators should be available for each sector within Mexican economic activity. According to Jiménez et al. (2022), this methodology facilitates the development of more successful investment strategies and allows for enhanced monitoring of sectors that are performing well as well as those that may require attention or adjustments. When considering these studies collectively, they support the implementation of PCA within the Mexican stock market to evaluate representativeness, develop sector-specific indicators, and translate co-movements into effective monitoring and allocation strategies.

3. Data and Methodology

This section provides a description of the model implementation, which uses all the national assets available on the Mexican Stock Exchange from January 2, 2020, to April 29, 2024. We decided to include a sector in the study that comprises the 12 listed ETFs representative of the Mexican capital market, in addition to the eleven sectors covered by the S&P 500. However, we excluded the energy sector because there was only one asset available for this (VISTAA.MX). Moreover, there are no national stocks in Mexico that cater to the technology sector.

3.1. Time justification

In a global economy strained by geopolitical conflicts, technological disruptions, reconfigured supply chains, and unpredictable interest-rate cycles, mastering the strategic interpretation of indicators is vital for financial survival. The period from 2020 to 2024 emphasizes these regime shifts: the implementation of the USMCA/T-MEC, the COVID-19 shock followed by normalization, heightened inflation with restrictive monetary policy, and the effects of geopolitical tensions on energy and commodities. Estimating PCA over this timeframe allows us to adjust factors based on current covariances—already reshaped by these shocks—rather than mixing them with less relevant pre-2020 correlations. For Mexico, the period 2020–2024 reflects a strategic relocation of production (nearshoring) toward North America, accompanied by the reallocation of trade and financial flows that impact various sectors differently. Industrial sectors, airports, and transportation benefit from the recovery of mobility and trade; materials are influenced by commodity sensitivity and construction activities; financial sectors are affected by interest rates and liquidity conditions; consumer sectors respond to changes in real income; telecommunications are driven by digitalization; and differences between industrial and office real estate impact real estate investment trusts (FIBRAs). Sector-level principal component analysis (PCA) consolidates these co-movements into interpretable components, with PC1 representing the common *"market mode"* and PC2 indicating internal segmentation. This facilitates the development of macro dashboards—measuring the strength of the common factor—and systemic alerts—detecting shifts in explained variance and the eigenvalue structure.

3.2 Data specification

All information used in the model is daily, based on adjusted closing prices (adjusted for dividends and splits). This approach ensures consistent accounting (no jumps from corporate stocks), maximizes the sample size for stable components (approximately 1,000 trading sessions over four years), and more accurately captures the volatility and dynamics of co-movement across stocks and sectors. PCA is performed on this daily data (correlation matrix), so the results reflect the dependence structure that actual participants face in the Mexican market.

Methodologically, a four to five-year horizon strikes a balance between statistical adequacy and the economic relevance of the current regime as the minimum sufficient sample to capture the contemporary regime of the Mexican equity market. On May 7, 2023, we consulted the official website of the Mexican Stock Exchange[2] to obtain information about available national instruments for the Mexican capital market. Using web scraping techniques, we downloaded information for the past four years (since the consultation date) for the 113 assets listed as available on the website. After data mining and cleaning, it was discovered that most of the information consisted of 1089 daily observations spread across 100 columns. Table 1 provides a detailed breakdown of the number of assets per sector. The description of assets regarding their industry and classification is based on information taken from Bloomberg, Investing, and Yahoo Finance!

3.3 Methodological presentation

The Mexican Stock Exchange offers a diverse range of assets across various economic sectors. Unlike the S&P 500, which has 11 primary sectors[3], the sector classification on the BMV may vary and does not follow a uniform structure like the U.S. market. It's important to note that this classification may vary, and some companies may be classified in more than one sector depending on their main activity. However, the main sectors represented on the Mexican Stock Exchange are presented in Table 1.

Table 1. Number of assets per sector in the Mexican stock Exchange

Sector	Number of Assets
Materials	19
Consumer Staples	15
Financial Services	13
Consumer Discretionary	12
ETF	12
Industrials	8
Real Estate	6
Healthcare	4
Telecommunications	4
Construction Services	3
Transportation and Logistics	3
Energy	1

Source: Own elaboration with information from Bolsa Mexicana de Valores and Yahoo Finance

Mexico is a diverse and growing economy that has traditionally relied on sectors such as mining, agriculture, and manufacturing, resulting in a stock market focused on a significant number of materials and consumer staples companies. These companies cater to the domestic market, which is a vital driver of the economy. Many of them also export globally and benefit from demographic growth. Moreover, usually their stocks are safer and more stable than those of other sectors, making them the primary focus of economic activity in the Mexican financial markets. In the following section, we will explain the process that was applied to each business sector in the data.

2 https://www.bmv.com.mx/en/issuers/issuers-information
3 https://www.sectorspdrs.com/

3.3.1 Principal Component Extraction

When dealing with data, it can be helpful to identify the most significant variables that explain most of the variance in the dataset. This can be achieved through Principal Component Analysis (PCA). According to Jollife and Cadima (2016), PCA is a technique used to reduce the complexity of data management by decreasing the dimensionality of the sample space while minimizing the loss of information. The primary objective of PCA is to determine new variables, known as components, which are linear combinations of the original ones. These components maximize variance and have no correlation with each other.

Mathematically, PCA aims to find a set of vectors a that, when multiplied with matrix X, maximizes the variance/information of the data. Matrix X has dimensions $n{\times}p$, where n is the number of observations and p is the number of variables. The eigenvectors of the covariance matrix of X represent the directions in which the data has the greatest variability, and these are the vectors sought after. In this sense, PCA involves finding the data's eigenvectors, which are then sorted according to the amount of variability they explain.

The first eigenvectors explain most of the variability in the data, while the last ones explain less. Once the eigenvectors are identified, we can choose how many principal components you want to retain based on how much variability we want to maintain in the data. This allows to reduce the dimensionality of the data while retaining as much important information as possible. In this process, note that the component extraction is performed so that the first component is the regression line that captures the most significant variance. The second component should extract as much information as the first component was unable to collect and so on, until the total variability of the system is collected. The goal of this technique is to manage the appropriate number of linear combinations that best represent the variables $X_1, X_2, \dots, X_\rho$. It is considered that the system is adequately explained when these components account for at least 70% of the original variance of the dataset.

Let the linear combinations of these variables in $Z_1, Z_2, \dots, Z_m$, where $m \leq \rho$. We can propose that:

$$Z_m = \sum_{j=1}^{\rho} \phi_{jm} X_j \tag{1}$$

where ϕ_{jm} represents the j variance loading of the m component. The concept of loading refers to the weight each variable has in each component. Note that each loading vector defines the direction in which the data's variance is greatest. A crucial property is that, to avoid inflating the data's variance, the linear combination is normalized such that the sum of the squares of the loadings equals one, as in equation 2:

$$\sum_{j=1}^{\rho} \phi_{j1}^2 = 1 \tag{2}$$

Therefore, this vector defines a line to minimize the distance error between the data and the line originated by this component. To evaluate this performance, the average of the Euclidean distance from the squared error is used as the discriminant of this metric:

$$z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} +, \dots, +\phi_{\rho 1}x_{i\rho} \tag{3}$$

As principal components are arranged hierarchically based on the amount of variance they explain, the loading vector of the first principal component solves the optimization problem aiming to minimize error:

$$\underset{\phi_{11},\dots,\phi_{\rho 1}}{\text{Maximize}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{\rho} \phi_{j1} x_{ij} \right)^2 \middle| \sum_{j=1}^{\rho} \phi_{j1}^2 = 1 \right\} \tag{4}$$

In the case of the second component, it arises from a linear combination of variables that form the second eigenvector, using more information from variables not correlated with the first component. Consequently, the second principal component will be perpendicular to the first one. This process continues in an orderly manner until all linear combinations are exhausted.

3.3.2. Contributions to variance levels

Since the goal of principal components analysis is to grasp the proportion of variance explained by each component, we can express the total variance of a dataset as:

$$\sum_{j=1}^{\rho} Var(X_j) = \sum_{j=1}^{\rho} \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2 \tag{5}$$

Alternatively, we can calculate the variance explained by the m principal component using the equation 6:

$$\frac{1}{n} \sum_{i=1}^{n} z_{im}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{\rho} \phi_{jm} X_{ij} \right)^2 \tag{6}$$

This implies that the proportion of the variance explained by the m principal component is determined by combining the two previous expressions with suitable weights, as shown in equation 7:

$$\frac{\sum_{i=1}^{n} \left(\sum_{j=1}^{\rho} \phi_{jm} X_{ij} \right)^2}{\sum_{j=1}^{\rho} \sum_{i=1}^{n} x_{ij}^2} \tag{7}$$

Taking these factors into account during the analysis, it's important to understand the proportion of variance explained by each component. Since each eigenvalue corresponds to the variance of the component Z_i of the eigenvector $\vec{v_i}$, we have:

$$Var(Z_i) = \lambda_i \tag{8}$$

With this, the proportion of the variability of the system that is explained by the component Z_i is:

$$\frac{\lambda_i}{\sum_{i=1}^{\rho} Var(Z_i)} \tag{9}$$

Adding all eigen-values λ_i yields the total variance of all components that add 1. Finally, note that the optimal number of principal components of n variables and ρ linear combinations is:

$$min(n-1, \qquad \rho) \tag{10}$$

It's worth noting that minimizing redundant variables is crucial when analyzing data. There isn't a fixed method for determining the ideal number of principal components; it relies on the analyst's discretion and the specific problem. Peres-Neto, Jackson, and Somers (2005) recommend trying various methods, around 20 in their study, to find the optimal number of components, but outcomes can differ. In the following section, we proceed with the discussion of results of the PCA in the Mexican stock exchange.

4. Outcomes and Discussion of the results

The first step in the process was to normalize the stock price data to ensure that all variables were on the same scale. This is important because PCA is sensitive to variable scales. After normalizing the data, PCA was applied to determine the contribution of each asset to every sector. Contribution and squared cosine (cos^2) are two measures used in PCA to assess the importance of original variables in forming linear combinations. These concepts are closely related and used to interpret PCA results in different ways: the contribution of a variable to a principal component indicates how much that variable influences the formation of the corresponding principal component. It is calculated as the square of the variable's loading on the principal component, multiplied by 100 to express it as a percentage. In other words, the contribution of a variable is a measure of how much the direction of that variable aligns with the direction of the principal.

4.1.1. PCA Materials

The materials sector has the highest number of variables, with 19 assets. However, the first linear combination does not capture as much variance as in other sectors with fewer stocks, accounting for only around 44% of the total variance. The variables that contribute most to the first dimension are SIMECB.MX, ICHB.MX, GCC.MX, and CMOCTEZ.MX. ORBIA.MX, CYDSASAA.MX, and ALPEKA.MX mainly drive the second dimension, while PE&OLES leads the third dimension.MX and CEMEXCPO.MX. Due to the number of assets, it is natural for them to show the same direction in the correlation circle for the first two dimensions, with arrows pointing in the same direction indicating that the variables are positively correlated. The cyclical behavior is detailed in the appendices.

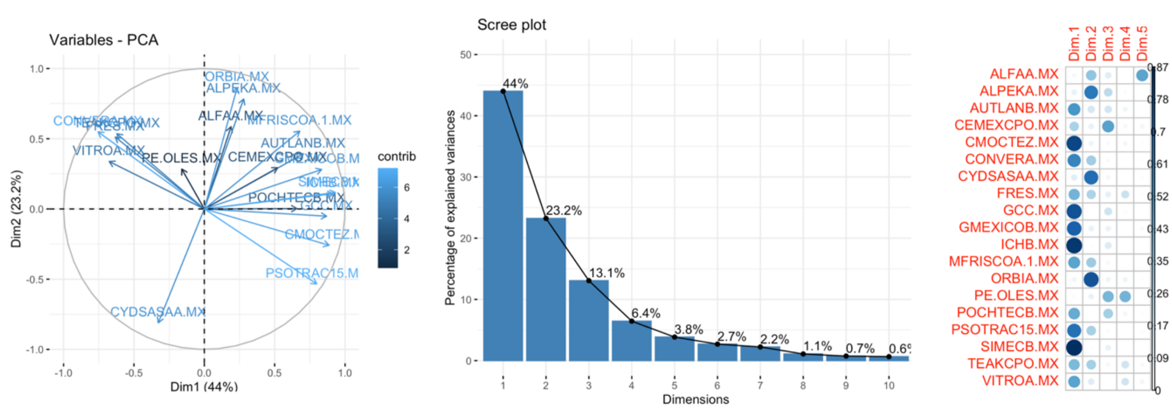


Figure 1. PCA Materials: Correlation Circle, Explained Variance and correlation matrix.
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.2. PCA Consumer Staples

The Consumer Staples sector consists of 15 stocks. Unlike the Materials sector, the first component accounts for a significant percentage of the total variance in the system, approximately 60%. The primary contributing variables to the first component are AC.MX, BFARB.MX, CHDRAUIB.MX, KOFUBL.MX, SORIANAB.MX, and BIMBOA.MX. On the other hand, the second dimension explains nearly 15% of the variance, with the main contributors being CUERVO.MX and LACOMERUBC.MX. From the third component onwards, no variables significantly contribute to the remaining dimensions.

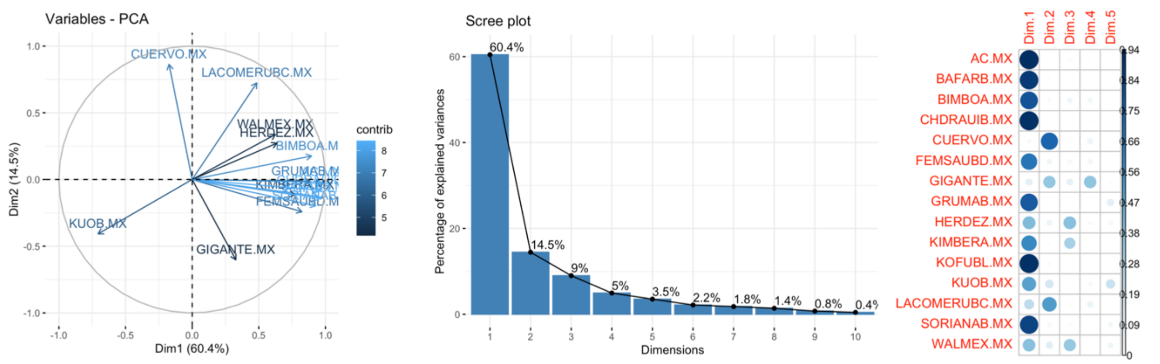


Figure 2. PCA Consumer Staples: Correlation Circle, Explained Variance and squared correlation matrix.
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.3. PCA Financial Services

The Financial Services sector consists of 13 stocks. Like the Consumer Staples sector, the first component captured a significant level of global variance compared to the remaining components. The first dimension is primarily formed by BBAJIO.MX, RA.MX, GFNORTE.MX, GFINBURO.MX, GENTERA.MX, and BBVA.MX. Additionally, the unit circle graph reinforces the correlation of the variables with a rightward direction for the first dimension.

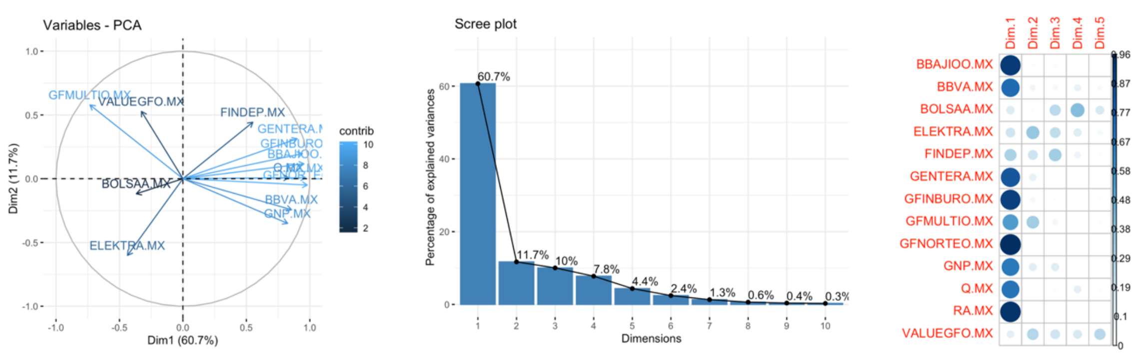


Figure 3. PCA Financial Services: Correlation Circle, Explained Variance and squared correlation matrix.
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.4. PCA Consumer Discretionary

The Consumer Discretionary sector comprises 12 stocks. Within this sector, the opposite orientation of the arrows in the unit circle indicates an inverse correlation primarily between two latent groups of assets. The first principal component accounts for nearly 57% of the system's variance, with the main contributions coming from LAMOSA.MX, VASCONI.MX, LIVEPOL1.MX, LIVEPOLC.1.MX, and ALSEA.MX. Regarding the second dimension, the primary contributions come from HCITY.MX, CMRB.MX, and HOTEL.MX. The representativeness of assets in the remaining dimensions is minimal, as the first two components explain approximately 77% of the total system variance.

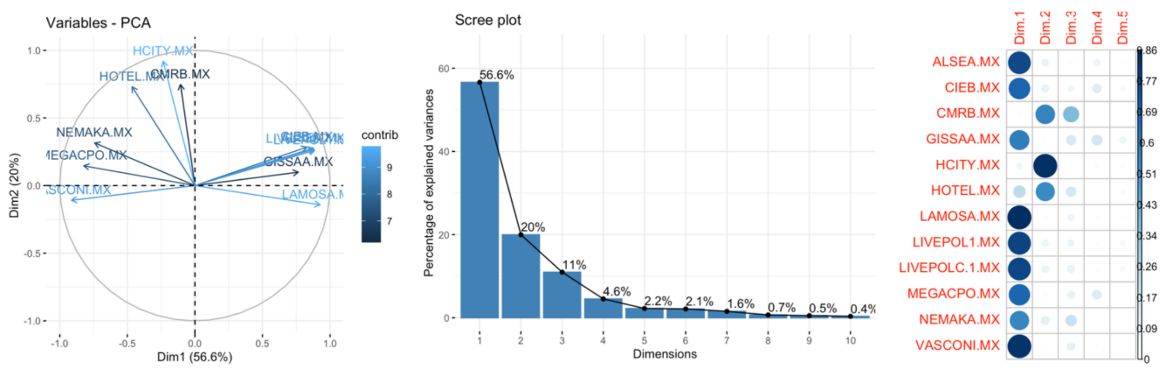


Figure 4. PCA Consumer Discretionary: Correlation Circle, Explained Variance and squared correlation matrix. Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.5. PCA ETF

There was a total of 12 representative ETFs in the Mexican market. Of all sectors, this was one of two that captured the highest level of variance from the first principal component. Like sectors with many assets, this led to the first dimension comprising multiple assets, such as QVGMEX18.MX, NAFTRACISHRS.MX, MEXTRAC09.MX, ESGMEXISHRS.MX, DLRTRAC15.MX, and GBMMODBFF.MX, all with cosine levels ranging from 0.71 to 0.98. On the other hand, for dimension 2, the contribution is mainly from UDITRACISHRS.MX. Subsequent dimensions show little relevant contributions.

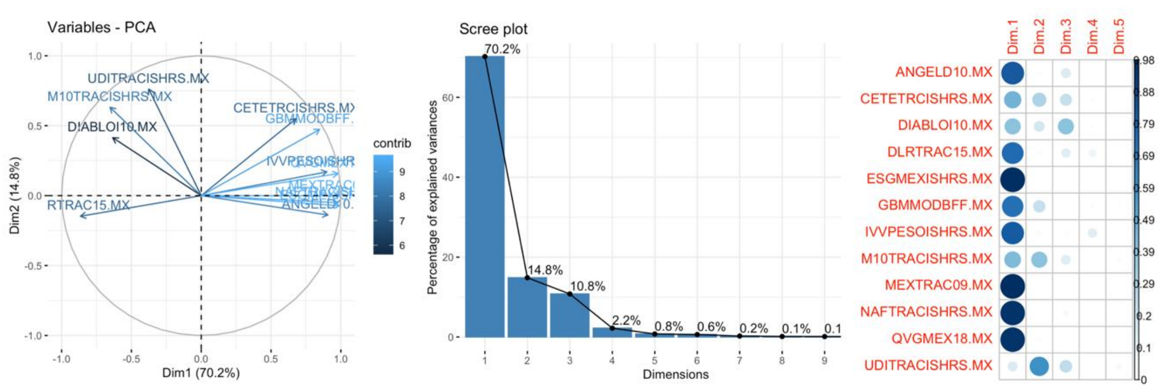


Figure 5. PCA ETF: Correlation Circle, Explained Variance and squared correlation matrix. Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.6. PCA Industrials

From this sector onwards, the number of assets decreased, with only 8 assets classified from the Mexican capital market belonging to this line of business. Despite the reduced number of assets, the first dimension captured almost 60% of the total variance. The main contributions in the first dimension come from five assets: OMAB.MX, ASURB.MX, GCARSOA1.MX, GAPB.MX, and PINFRA.MX, with cosine squared levels ranging from 0.67 to 0.91. On the other hand, for the second dimension, the significant contributors are VOLARA.MX, PINFRAL.MX, and AGUA.MX, with metrics ranging from 0.42 to 0.60.

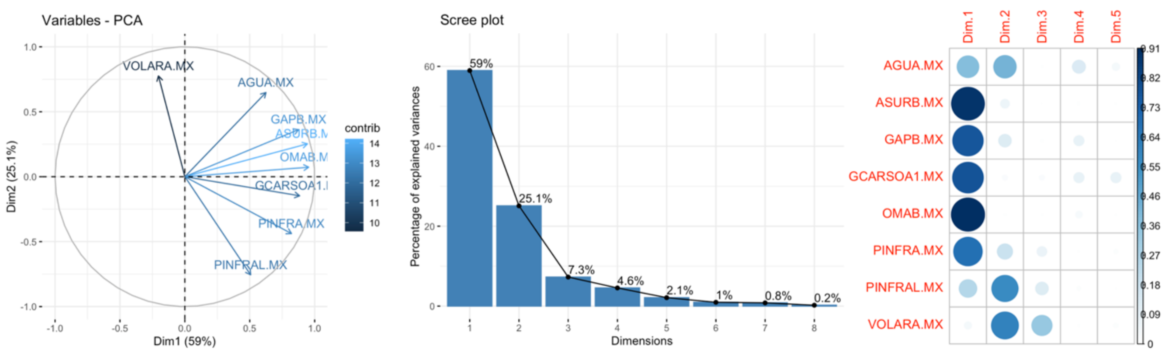


Figure 6. PCA Industrials: Correlation Circle, Explained Variance and correlation matrix .
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.7. PCA Real Estate

In the Mexican capital market, there are only a few companies related to the Real Estate sector. According to Yahoo Finance and Investing, there are only 6 assets that could be worked with this sector, considering how these companies are classified. Despite the small number of companies, the correlation unit circle indicates that two of them (GICSAB.MX and CADUA.MX) show an inverse correlation with the rest. Note that arrows at 90-degree angles to others suggest that variables are uncorrelated. Regarding contributions to variance levels, the first two principal components capture around 80% of the total variance.

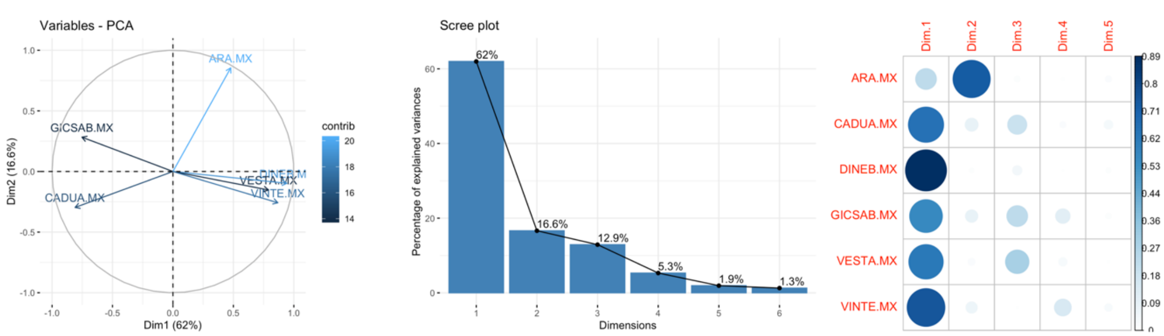


Figure 7. PCA Real Estate: Correlation Circle, Explained Variance and squared correlation matrix.
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.8. PCA Healthcare

It's interesting to note that the Mexican Healthcare sector has only 4 stocks, making its correlation analysis simpler. FRAGUAB.MX and MEDICAB.MX have similar correlations, while LABB.MX shows an inverse correlation to the first two. BEVIDESB.MX seems to have no correlation with the other three variables. The first principal component includes all assets except BEVIDESB.MX, which mainly contributes to the second component. Notably, the first two components capture around 88.4% of the total global variance.

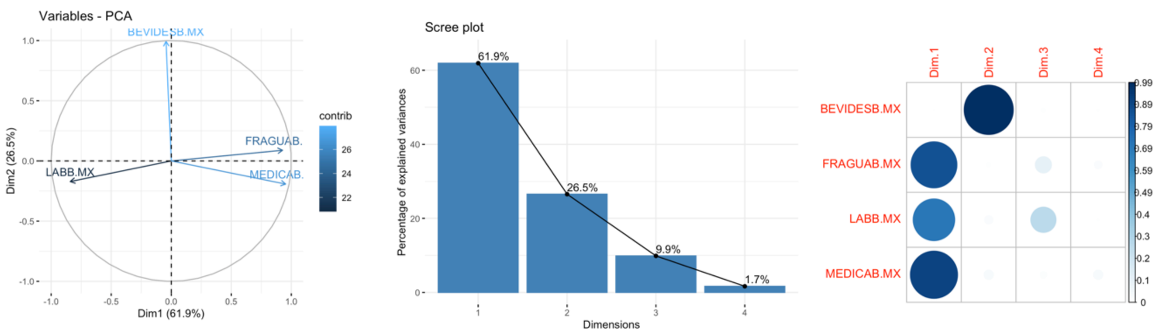


Figure 8. PCA Healthcare: Correlation Circle, Explained Variance and squared correlation matrix
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.9. PCA Telecommunications

This sector also had only 4 assets to model. An interesting finding from the correlation circle is that AMXB.MX was the only asset pointing in the opposite direction to the rest. However, this was not directly reflected in the matrix of contribution levels of each variable, as in the case of Healthcare, since the main contributions for constructing the first component come from AMXB.MX and AXTELCPO.MX, with levels of 0.68 and 0.78, respectively. Although the other two also contribute with levels of 0.40 onwards. On the other hand, for dimension 2, the only variable that contributes significantly is TLEVISACPO.MX, with a level of 0.51. Note that the sum of dimension 1 and 2 accounts for 84.9% of the total variance.

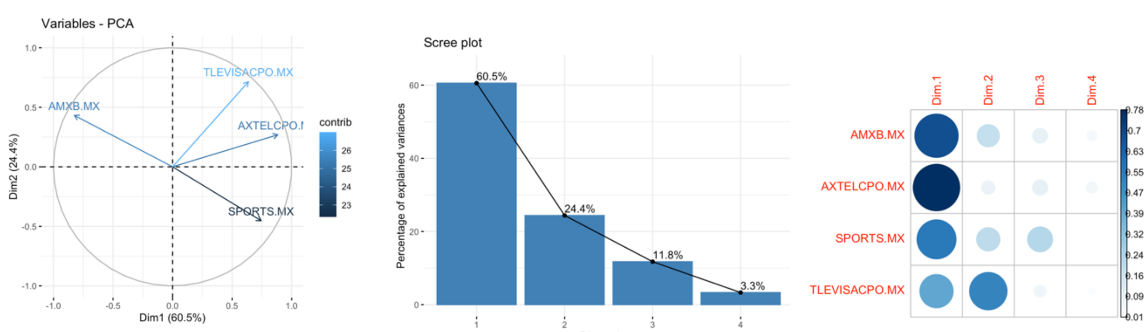


Figure 9. PCA Telecomunications: Correlation Circle, Explained Variance and correlation matrix
Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.10. PCA Construction Services

The Construction Services sector had only three assets for component construction. Since the arrows of the assets point in the same direction but with different inclinations, this explains the correlation among the assets and how, through their linear combinations, they were able to capture a significant percentage, nearly 68%, of the system's global variance in the first dimension. This is reflected in the contribution's matrix. On the other hand, the second dimension is mainly represented by GMD.MX.

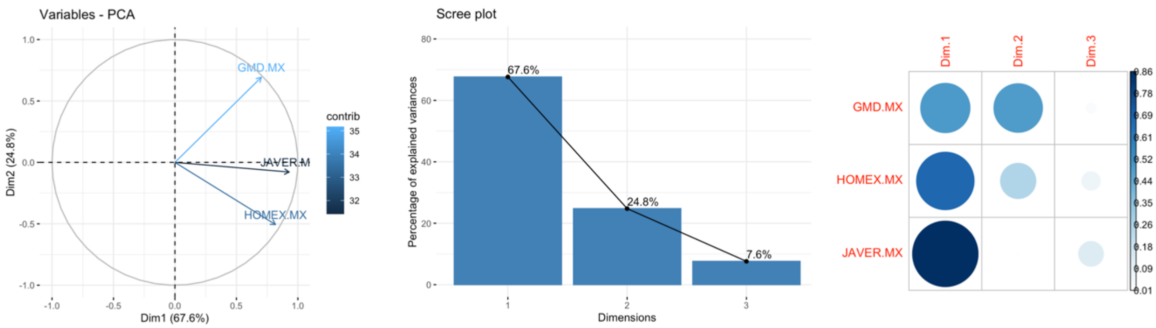


Figure 10. PCA Consumer Services: Correlation Circle, Explained Variance and correlation matrix. Source: Own elaboration with information from Yahoo Finance, R-Studio.

4.1.11. PCA Transportation and Logistics

Finally, the Transportation and Logistics sector had only three assets for component construction. Out of all the databases, this sector captured the most variance in the first component (around 87%), and all the series are well represented in the matrix of cosine squared. The results are shown in Figure 11.

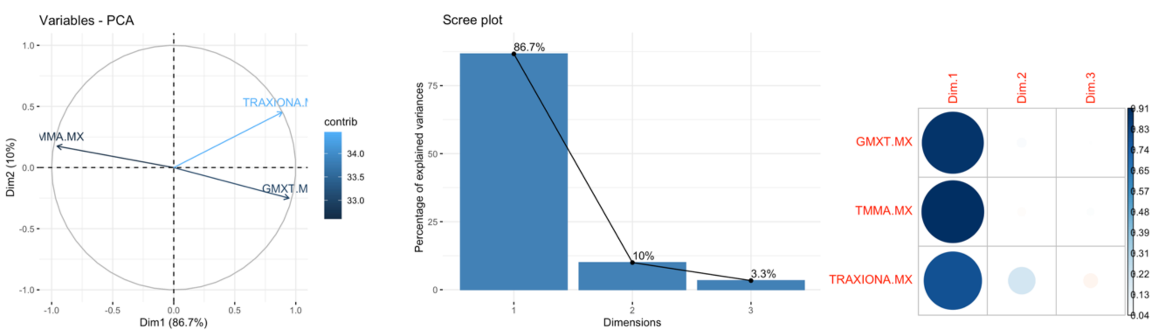


Figure 11. PCA Transportation and Logistics: Correlation Circle, Explained Variance and squared correlation matrix. Source: Own elaboration with information from Yahoo Finance, R-Studio.

The eleven cases modelled illustrate that the PCA can represent a *"statistical benchmark,"* which is grounded in actual co-movements rather than being arbitrary. PCA functions based on linear relationships, whereas alternative methods may capture more intricate or nonlinear interactions.

In emerging markets, such as Mexico, there are fewer issuers and a higher concentration in key sectors, including financials, materials, and consumer sectors. The market also exhibits significant dependence on macroeconomic shocks such as exchange rates, interest rates, and public policy. As a result, PCA identifies a substantially more dominant PC1, explaining between 40–70% of the variance by sector, which aligns with model outcomes.

How do the results vary when applied to markets such as Mexico compared to developed markets? Research conducted by James, N., Menzies, M., & Gottwald, G. A. (2022), as well as Stanciu, C. V., & Spulbar, A. C. (2024), has demonstrated that developed markets (e.g., the U.S. and Europe) typically exhibit high liquidity and a broader spectrum of securities, characterized by a more diversified structure and less sectoral concentration. Consequently, Principal Component Analysis (PCA) frequently reveals that the first component (PC1) accounts for a smaller proportion of the variance—approximately 20–30% in the S&P 500—due to the presence of more independent factors and diminished forced co-movement. PC2 and PC3 tend to capture sector rotations and idiosyncratic shocks.

This suggests that assets within these markets are more correlated and respond collectively to common factors (e.g., inflation, oil prices, fiscal policy), thereby implying that intra-sector diversification is more limited and that "leading assets" possess greater explanatory power compared to those in developed markets.

4.2. Discussion of results

There are multiple methodologies to measure representativeness, such as cluster analysis, factor models, financial network theory, Random Matrix Theory, and modern machine learning techniques (autoencoders, t-SNE). The key difference of PCA is that it provides a linear, simple, and interpretable representation of the co-movement structure, turning a broad set of stocks into dominant factors. In developed markets, the first components usually explain less variance due to greater diversification and liquidity, whereas in emerging markets like Mexico, sectoral concentration and macro shock dependency lead the first component to capture a much larger share of variance, making the “guide assets” of each sector more visible. In developed markets, the first components usually explain less variance due to greater diversification and liquidity, whereas in emerging markets like Mexico, sectoral concentration and macro shock dependency lead the first component to capture a much larger share of variance, making the “guide assets” of each sector more visible.

In our model, the PCA analysis reveals clear, actionable patterns in the Mexican Stock Exchange: Correlation circles for each sector reveal which stocks share the same direction (co-movements), which helps avoid within-sector redundancy by combining stocks with distant loading vectors. The scree plots show that PC1 accounts for roughly 40–70% of the variance in most sectors; therefore, its linear combination synthesizes the core dynamism of each group. The contributions/loadings identify guide stocks (with large absolute loadings), useful to represent the sector in analyses and dashboards.

In the Materials sector (19 stocks), PC1 explains around 44% of the variance—lower than in smaller sectors due to its greater heterogeneity—yet the correlation circle shows most stocks moving in the same direction. Here, PC1 synthesizes the pulse of the construction/investment cycle, and the highest-loading stocks (e.g., leading cement and steel names) act as sector benchmarks.

In Consumer Staples (around 60% in PC1), the common factor is stronger: large retailers and beverage companies share macro shocks (inflation, real wages), so a small set of representative stocks captures much of the sector’s dynamism.

A notable finding appears in Financial Services, where co-movement is very pronounced—consistent with banks and finance companies responding to interest rates and liquidity—so PC1 dominates.

In Consumer Discretionary (around 57% in PC1; around 77% in PC1+PC2), two opposing subgroups emerge (retail/restaurants vs. hotels), which is useful to diversify within the sector without losing exposure.

The Industrials sector (around 60% in PC1) concentrates variance in airports and conglomerates, consistent with common shocks from tourism and external trade.

Across Real Estate, Telecommunications, and Health—sectors with relatively few listings—PC1+PC2 explain at least around 80%, yielding a very compact structure. At the same time, opposite vectors emerge (e.g., a telecom stock or a REIT/FIBRA that behaves differently from peers), pointing to more balanced intra-sector strategies.

Construction shows around 68% in PC1 with high alignment among its three series, while Transport & Logistics reaches around 87% in PC1, reflecting a very strong common driver.

In ETFs, high contributions and $\cos^2$ indicate that PC1 captures the aggregate "market mode," serving as a quick thermometer of country-level sentiment.

Seen together, the results show that: (i) PC1 captures around 40–70% of variance in most sectors, and PC1+PC2 exceeds around 80% when groups are small; (ii) contribution/$\cos^2$ matrices allow the identification of "guide stocks" (large absolute loadings) that represent their sector for monitoring and analysis; (iii) correlation circles reveal co-movements and opposing pairs that help avoid redundancy and build truly diversified intra-sector baskets. This reading fits the reality of the Mexican economy: concentrated sectors (banking, staples, airports) share macro and regulatory factors, while cyclical areas like materials display more dispersion due to differences in exposure to public works, housing, or exports.

Our work demonstrates that PCA is useful and valid for application in the Mexican stock market. First, it reduces dimensionality without losing the group's essential information: one PC1 per sector acts as a synthetic indicator of sector "spirit," making it easy to communicate and monitor. Second, it identifies representative stocks for dashboards and internal benchmarks, avoiding long lists that add little signal. Third, it structures diversification: angles in the correlation circle help combine non-collinear stocks, lowering the risk of duplicated exposures. Fourth, it is replicable and transparent: with public data and clear rules (normalization, extraction, loadings, scores), any analyst can reproduce the results and update them with rolling windows. In short, PCA provides a solid quantitative basis to describe co-movement structure, select guide stocks, and build sector readings aligned with the real drivers of the Mexican economy.

5. Conclusions

Conducting this research enabled us to identify the predominant business sectors within the Mexican stock market based on the available information, as evidenced by the greater number of issuers in these sectors. Moreover, the Principal Component Analysis (PCA) facilitated the quantification of each asset's individual significance within its respective sector. In this context, our work offers substantial contributions, as, to the best of our knowledge, previous studies have analyzed specific sectors; however, none have examined all the available sectors within the Mexican market collectively. The findings suggest employing the scores/loadings of each asset on the contribution to PC1 as indicators for trading, monitoring, and portfolio construction within sectors, particularly when associated with significant loadings on the guiding assets.

An interesting result of the modeling is that, for the sectors, the greater the number of assets each sector had, the more they contributed individually to the first principal component. Moreover, this first component captured significant magnitudes of the system's variance (from 40% to 70%). Meanwhile, for the rest of the components, the individual contributions of each stock were practically negligible: a high value in a dimension indicates that the variable is well represented in that dimension, while a low value indicates the opposite. This is represented in the correlation matrix. We can interpret these values to understand which variables contribute most to each dimension of the PCA and how the variance of the variables is distributed in the dimensional space defined by the PCA. This is reflected in the appendices, where the linear combinations of each dimension seem capable of capturing the oscillatory behavior of the market cycles of the assets.

In sectors with a smaller number of assets (four or fewer), the hierarchical extraction of the components still applies. In these sectors, higher levels of contribution from the assets for the dimensions following the first one begin to be noticed, compared to sectors with a larger number of assets. It's also worth noting that for most sectors, the linear combinations exhibit very uniform oscillatory behaviors, reflecting the cyclical dynamics of each sector. This behavior is evident in the appendices, where the linear combinations of each dimension seem to capture the oscillatory behavior of the market cycles of the assets.

An additional significant point is that our findings are consistent with the research conducted by Heinrich, L., Shivarova, A., & Zurek, M. (2021), as well as Pan, R. K., & Sinha, S. (2007) and Cantú, Mendoza & Arteaga (2024). In developed markets, the primary components generally account for less variance, owing to higher levels of diversification and liquidity. Conversely, our results indicate that in emerging markets such as Mexico, sector concentration and dependence on macroeconomic shocks cause the first component to explain a substantially larger proportion of the variance. This, in turn, accentuates the importance of "guide assets" within each sector.

The selected core components identified through PCA can be helpful in formulating trading strategies in future research. For instance, trading strategies can be developed based on signals derived from the fluctuations in the principal components. It is important to note that PCA is an exploratory analysis tool and not a definitive solution. Therefore, it should be integrated with other methods and considerations, such as fundamental analysis and market sentiment, when devising trading strategies. Our research approach emphasizes the need to evaluate the behavior of assets and their returns concerning sector fluctuations in the Mexican market and their impact on it. Practical implications. (i) Intra-sector selection: prioritizing stocks with high PC1 loadings reduces idiosyncratic noise and captures the sector's pulse; (ii) Intra-sector diversification: combining stocks with non-collinear vectors avoids duplicating exposure; (iii) Tactical monitoring: the PC1 score (the component's time series) can be used as a sector indicator for tracking and reporting, while leaving operational validation (out-of-sample) for future work prior to investment decisions.

The stock market plays a vital role in economic development. It provides financing, encourages investment and savings, promotes transparency in corporate governance, creates employment opportunities, diversifies investment options, and attracts foreign investment. The main function of stock exchanges is to provide companies with a platform to raise funds by issuing stocks and bonds. This enables firms to expand their operations, invest in new projects, and finance research and development activities, thereby contributing to overall economic growth. Therefore, sector-specific and generalized market sentiment indicators are essential. This may help make informed decisions, formulate investment strategies, and effectively allocate resources in equity markets.

References

[1] Aït-Sahalia, Y., & Xiu, D. (2017). Using principal component analysis to estimate a high-dimensional factor model with high-frequency data. Journal of Econometrics, 201(2), 384–399. https://doi.org/10.1016/j.jeconom.2017.08.015

[2] Alam, M. R., & Istiak, K. (2019). Impact of US policy uncertainty on Mexico: Evidence from linear and nonlinear tests. The Quarterly Review of Economics and Finance, 77. https://doi.org/10.1016/j.qref.2019.12.002

[3] Arango Herrera, E., García Fernández, F., & Serna Hinojosa, J. A. (2019). Impacto de la estrategia de diversificación en el desempeño financiero en empresas de la Bolsa Mexicana de Valores. Revista Mexicana de Economía Y Finanzas, 15(1), 135–144. https://doi.org/10.21919/remef.v15i1.384

[4] Avellaneda, M., & Lee, J. H. (2008). Statistical Arbitrage in the U.S. Equities Market. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1153505

[5] Avellaneda, M., Healy, B., Papanicolaou, A., & Papanicolaou, G. (2022). Principal Eigenportfolios for U.S. Equities. SIAM Journal on Financial Mathematics, 13(3), 702–744. https://doi.org/10.1137/20M1383501

[6] Bosch, J. O. (2023, March 21). Interview "Correlación directa: crecimiento económico y desarrollo del mercado de valores" by Adriana Reyes Urrutia. Retrieved from https://www.revista.imef.org.mx/articulo/correlacion-directa-crecimiento-economico-y-desarrollo-del-mercado-de-valores/

[7] Bogle, J. C. (2017). The little book of common sense investing: the only way to guarantee your fair share of stock market returns. Wiley. (Original work published 2007)

[8] Cantú Esquivel, J. A., Cruz Aké, S., & Jiménez Preciado, A. L. (2025). Evaluación de la consistencia de las betas en el modelo de CAPM mediante un análisis de bootstraps con memoria. Revista Mexicana de Economía y Finanzas, Nueva Época, 20(2), 1-21. https://doi.org/10.21919/remef.v20i2.1192

[9] Cantú Esquivel, J. A., Mendoza Rivera, R. J. y Arteaga Nagashiro, Y. V. (2024). Sincronización de fase y Análisis de Componentes Principales para la construcción del ciclo económico de México. Revista Latinoamericana de Investigación Social, 6(1), 22–45. https://revistasinvestigacion.lasalle.mx/index.php/relais/article/view/4062/3663

[10] Cantu-Esquivel, J. A., Mendoza-Rivera, R. J., Razo-De-Anda, J. O., & Borrego-Salcido, C. (2025). Construction of the business cycle by phase synchronization and principal component analysis. Computational Economics. https://doi.org/10.1007/s10614-025-11207-5

[11] Connor, G., & Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. Journal of Financial Economics, 21(2), 255–289. https://doi.org/10.1016/0304-405X(88)90062-1

[12] Coronado, S., Martínez, J. N., & Venegas Martínez, F. (2022). IMPACT OF US ECONOMIC POLICY ON MAJOR LATIN AMERICA'S STOCK MARKETS, 2002-2020. Investigación Económica, 81(320), 62–80. https://www.jstor.org/stable/48658246

[13] Díaz Valencia, G. A. (2017). La influencia de la vivienda en la aversión al riesgo de portafolios familiares. Revista Mexicana de Economía Y Finanzas, 12(3), 89–119. https://doi.org/10.21919/remef.v12i3.98

[14] Galván Vera, A., García Fernández, F., & Serna Hinojosa, J. A. (2017). Estrategia de diversificación empresarial en México: un análisis de valor. Revista Venezolana de Gerencia, 22(79), 368–386. https://www.redalyc.org/journal/290/29055964003/html/

[15] Gavira Durón, N., Martínez Peña, D. G., & Espitia Moreno, I. C. (2020). Determinantes financieras de la Sustentabilidad Corporativa de Empresas que cotizan en el IPC Sustentable de la BMV. Revista Mexicana de Economía Y Finanzas Nueva Época REMEF, 15(2), 277–293. https://www.remef.org.mx/index.php/remef/article/view/485/613

[16] Gil León, J. M., & Silva Pinzón, D. E. (2019). Índice de incertidumbre de política económica (EPU) para Colombia, 2000-2017. Ensayos de Economía, 29(55), 37–56. https://dialnet.unirioja.es/servlet/articulo?codigo=7253165

[17] Heinrich, L., Shivarova, A., & Zurek, M. (2021). Factor investing: alpha concentration versus diversification. Journal of Asset Management, 22(5), 464–487. https://doi.org/10.1057/s41260-021-00226-0

[18] James, N., Menzies, M., & Gottwald, G. A. (2022). On financial market correlation structures and diversification benefits across and within equity sectors. Physica A: Statistical Mechanics and its Applications, 604, Article 127682. https://doi.org/10.1016/j.physa.2022.127682

[19] Jiménez Preciado, A. L., Venegas Martínez, F., & Ramírez García, A. (2022). Stock Portfolio Optimization with Competitive Advantages (MOAT): A Machine Learning Approach. Mathematics, 10(23), 4449. https://doi.org/10.3390/math10234449

[20] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

[21] Laloux, L., Cizeau, P., Bouchaud, J.-P., & Potters, M. (1999). Noise dressing of financial correlation matrices. Physical Review Letters, 83(7), 1467–1470. https://doi.org/10.1103/PhysRevLett.83.1467

[22] Langebaek Rueda, A., & Ortiz E., J. E. (2007). Q de Tobin y Gobierno corporativo de las empresas listadas en bolsa. Borradores de Economia, 447. https://doi.org/10.32468/be.447

[23] Litterman, R. B., & Scheinkman, J. (1991). Common factors affecting bond returns. The Journal of Fixed Income, 1(1), 54–61. https://doi.org/10.3905/jfi.1991.692347

[24] Nguyen, Q., & Nguyen, N. K. K. (2019). Composition of the first principal component of a stock index—A comparison between S&P 500 and VNIndex. Physica A: Statistical Mechanics and its Applications, 536, 120940. https://doi.org/10.1016/j.physa.2019.04.216

[25] Pan, R. K., & Sinha, S. (2007). Collective behavior of stock price movements in an emerging market. arXiv.

[26] Peres Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 49(4), 974–997. https://doi.org/10.1016/j.csda.2004.06.015

[27] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., & Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. Physical Review E, 65(6), 066126. https://doi.org/10.1103/PhysRevE.65.066126

[28] Portelli, L., & Roncalli, T. (2024). Stock-Bond Correlation: Theory & Empirical Results. Ssrn.com. https://doi.org/10.2139/ssrn.4823094

[29] Rueda Cadena, I. (1980). La bolsa de valores, instrumento de concentración del capital. Revista Mexicana de Sociología, 42(4), 1691. https://doi.org/10.2307/3539966

[30] Ruiz Núñez, M. del C., Galán Briseño, L. M., & Ruiz Núñez, G. (2016). Importancia de las bolsas de valores en la economía. Diagnóstico Fácil Empresarial Finanzas Auditoria Contabilidad Impuestos Legal/Diagnóstico Fácil Empresarial Finanzas Auditoria Contabilidad Impuestos Legal, 4(4), 55–64. https://doi.org/10.32870/dfe.vi4.75

[31] Ruiz Porras, A., & Steinwascher, W. (2007, July 3). Gobierno corporativo, diversificación estratégica y desempeño empresarial en México. Mpra.ub.uni-Muenchen.de. https://mpra.ub.uni-muenchen.de/3819/

[32] Santillán Salgado, R. J., Gurrola Ríos, C., Venegas Martínez, F., & Jiménez Preciado, A. L. (2018). La dependencia del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (IPC) con respecto a los principales índices bursátiles latinoamericanos. Contaduría Y Administración, 63(4), 55. https://doi.org/10.22201/fca.24488410e.2018.1256

[33] Souma, W., Takahashi, H., & Aoyama, H. (2021). Characteristics of principal components in stock price dynamics. Frontiers in Physics, 9, 602944. https://doi.org/10.3389/fphy.2021.602944

[34] Severini, T. A. (2022). Some properties of portfolios constructed from principal components of asset returns. Annals of Finance, 18(4), 457–483. https://doi.org/10.1007/s10436-022-00412-z

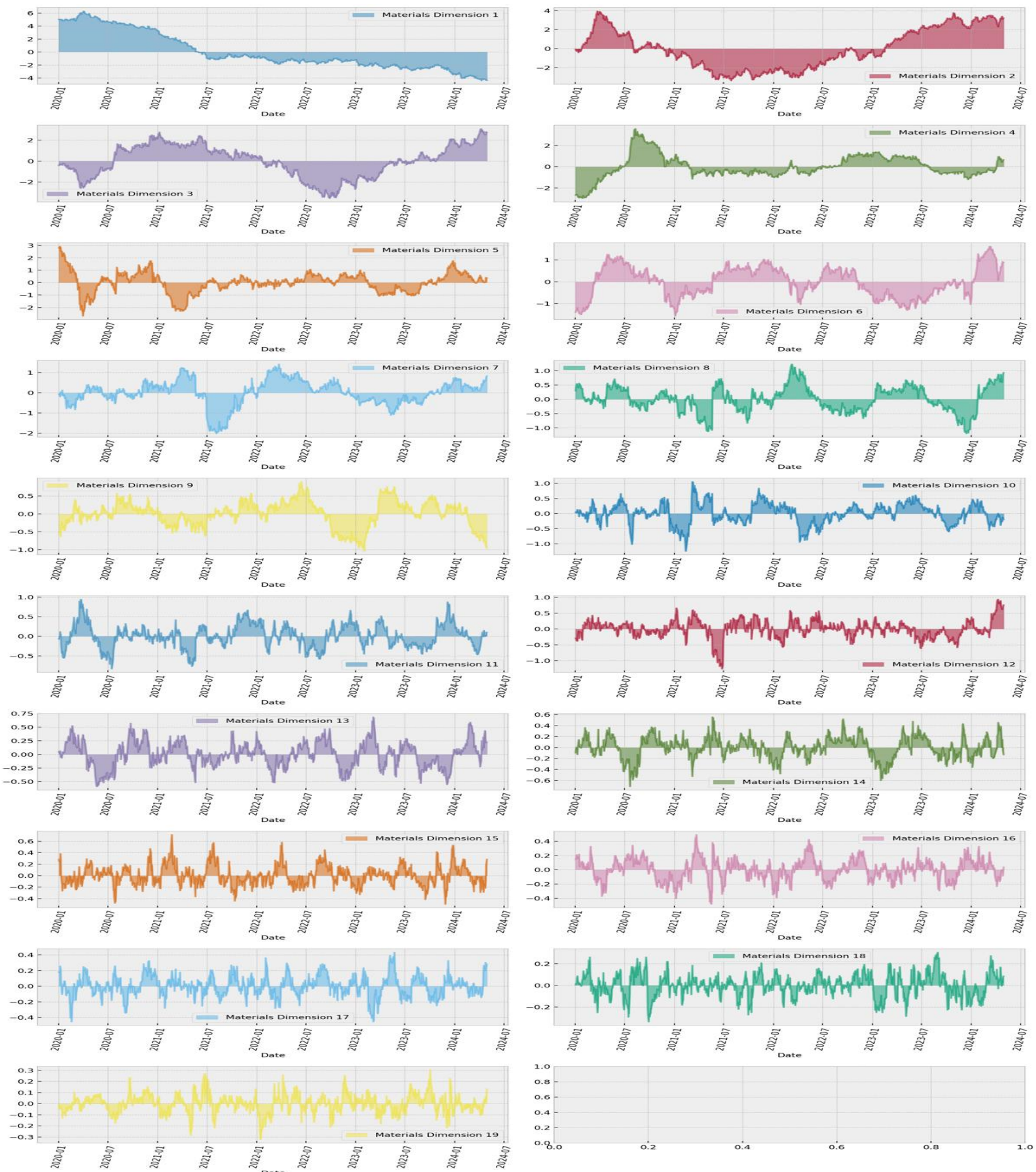
[35] Stanciu, C. V., & Spulbar, A. C. (2024). Financial Integration of the European Union Financial Markets. A PCA Approach. Studies in Business and Economics, 19(3), 241 256. https://doi.org/10.2478/sbe-2024-0054

[36] Yang, L., Rea, W., & Rea, A. (2017). Financial insights from the last few components of a stock market PCA. International Journal of Financial Studies, 5(3), 15. https://doi.org/10.3390/ijfs5030015

4. Appendices

Appendix 1.

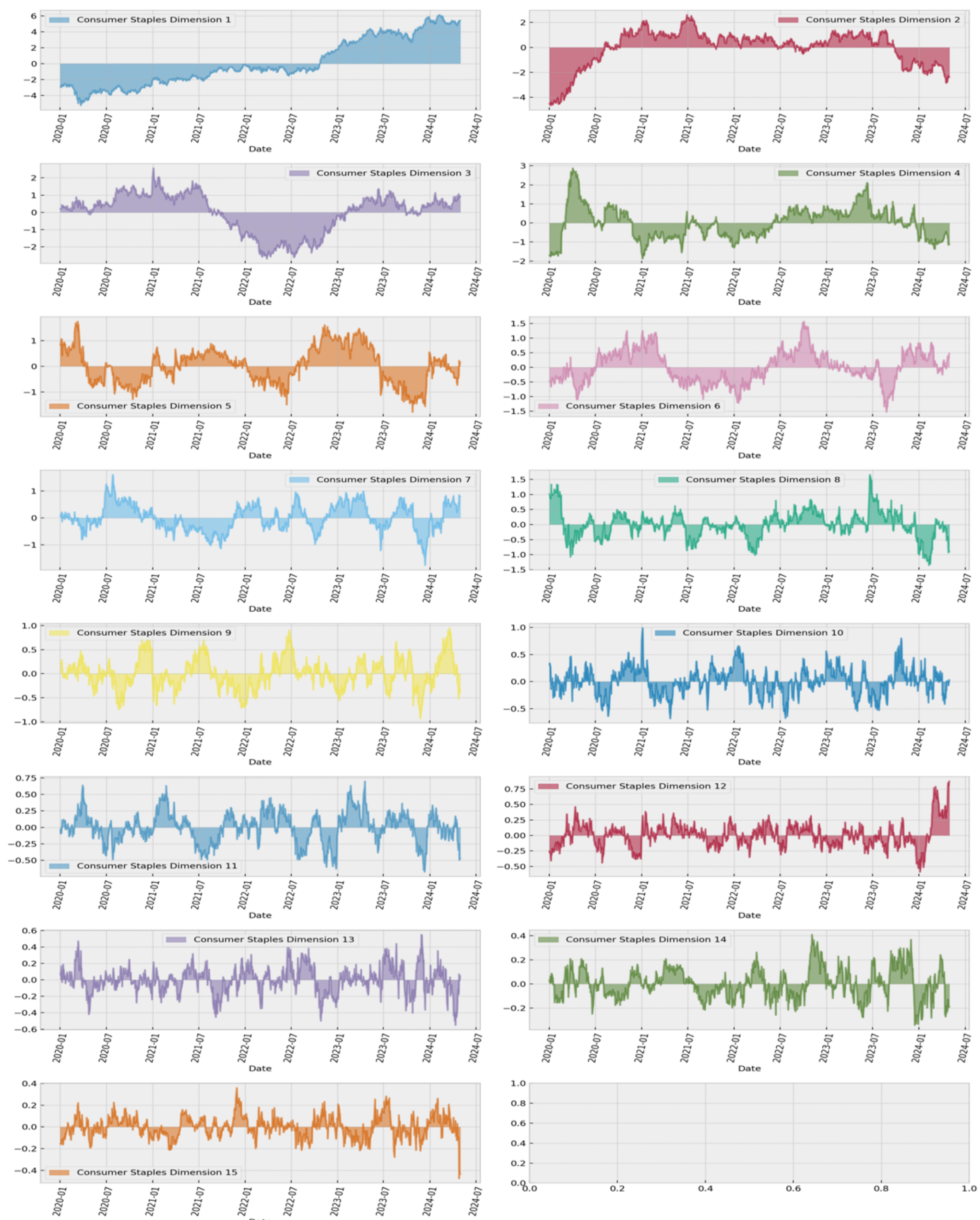
Linear Combinations Materials (19 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 2.

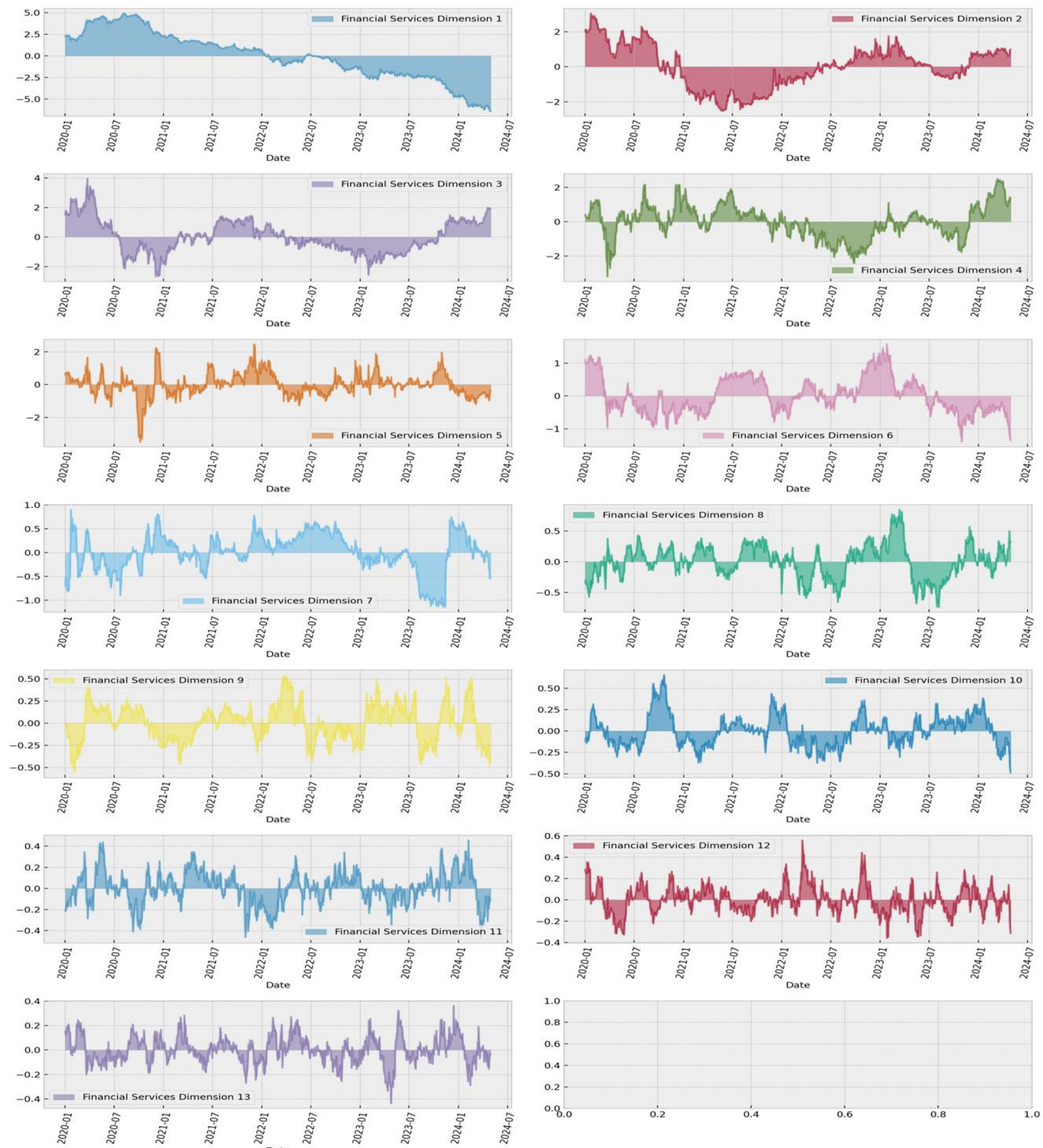
Linear combinations Consumer Staples (15 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 3.

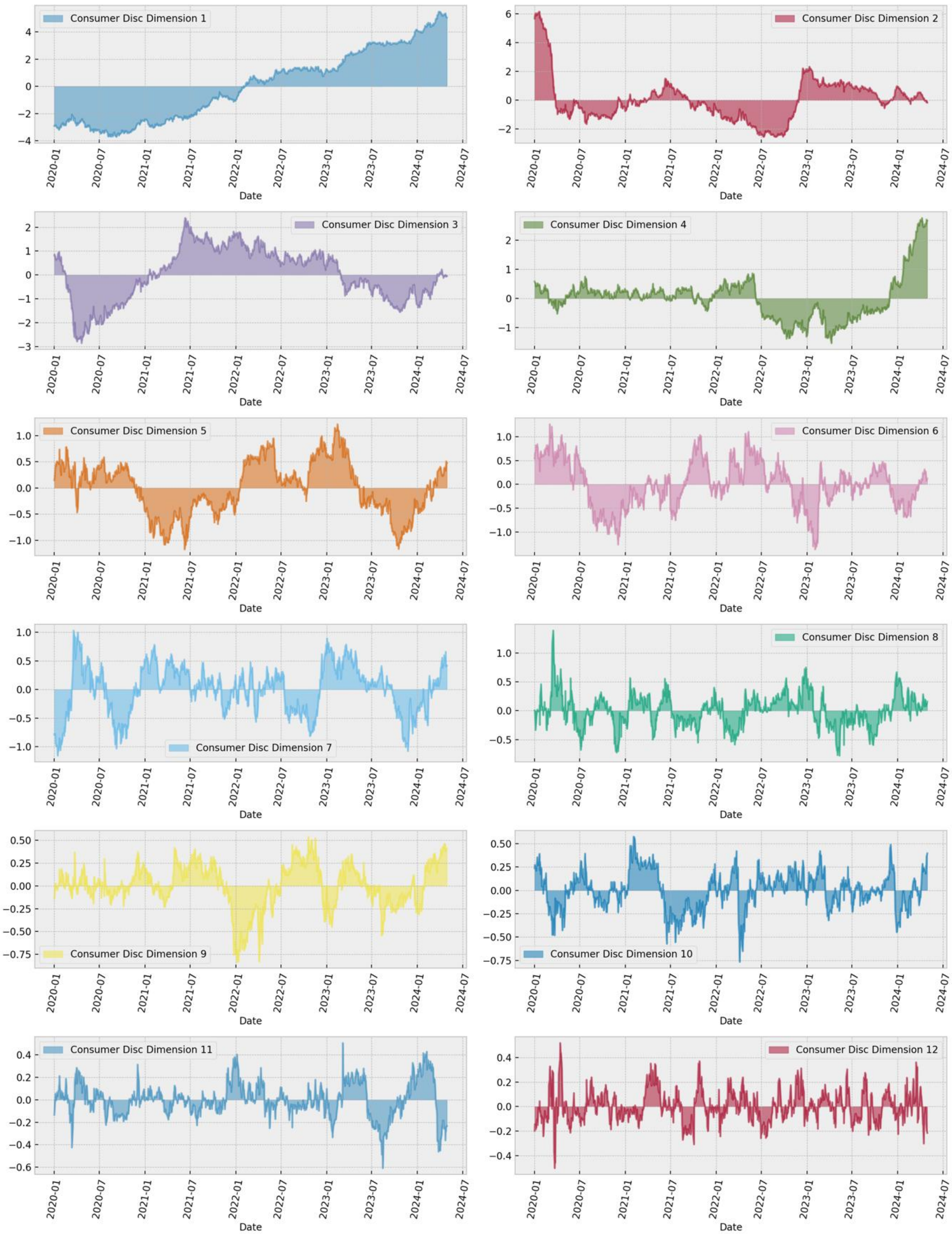
Linear combinations Financial Services (13 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 4.

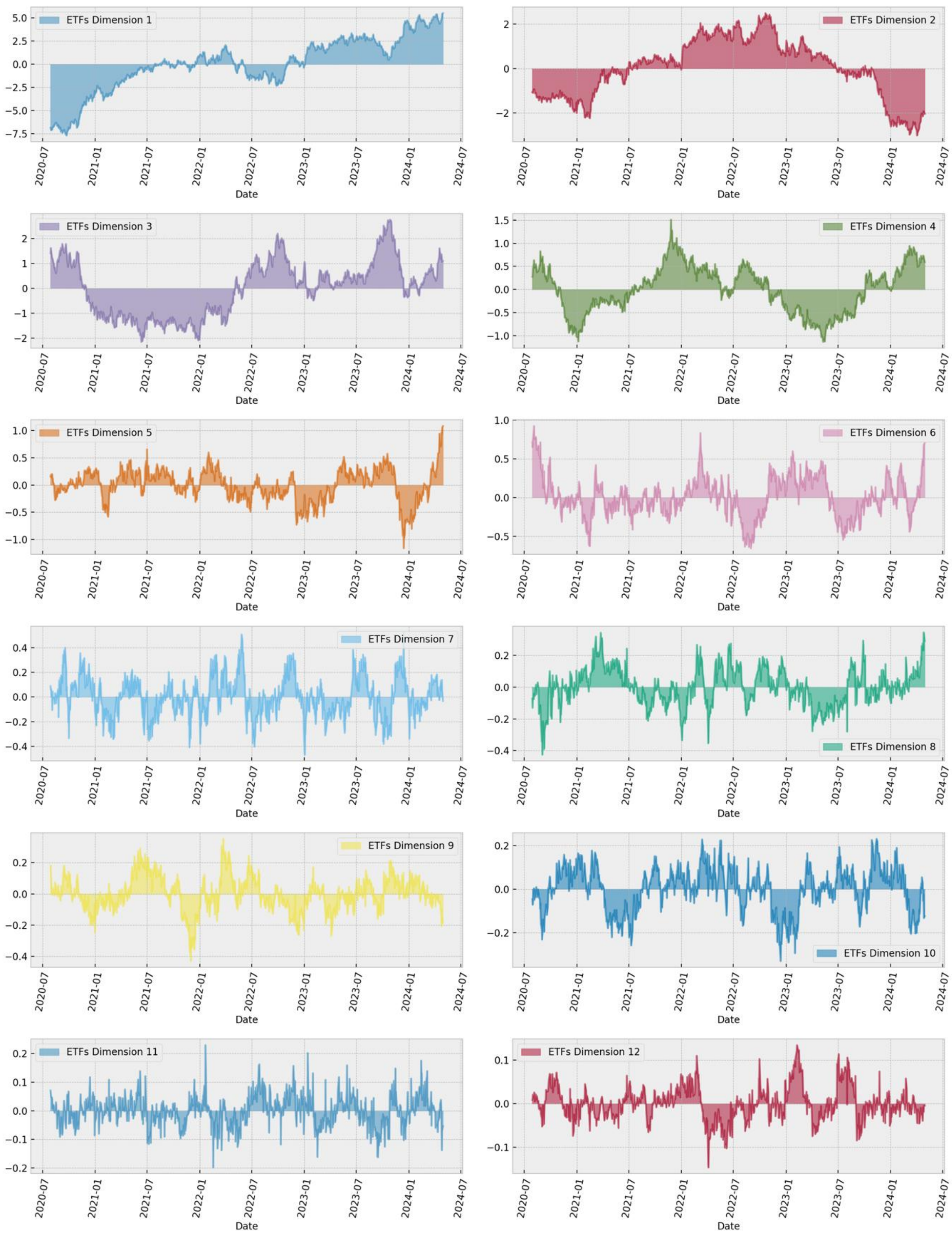
Linear Combinations Consumer Discretionary (12 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 5.

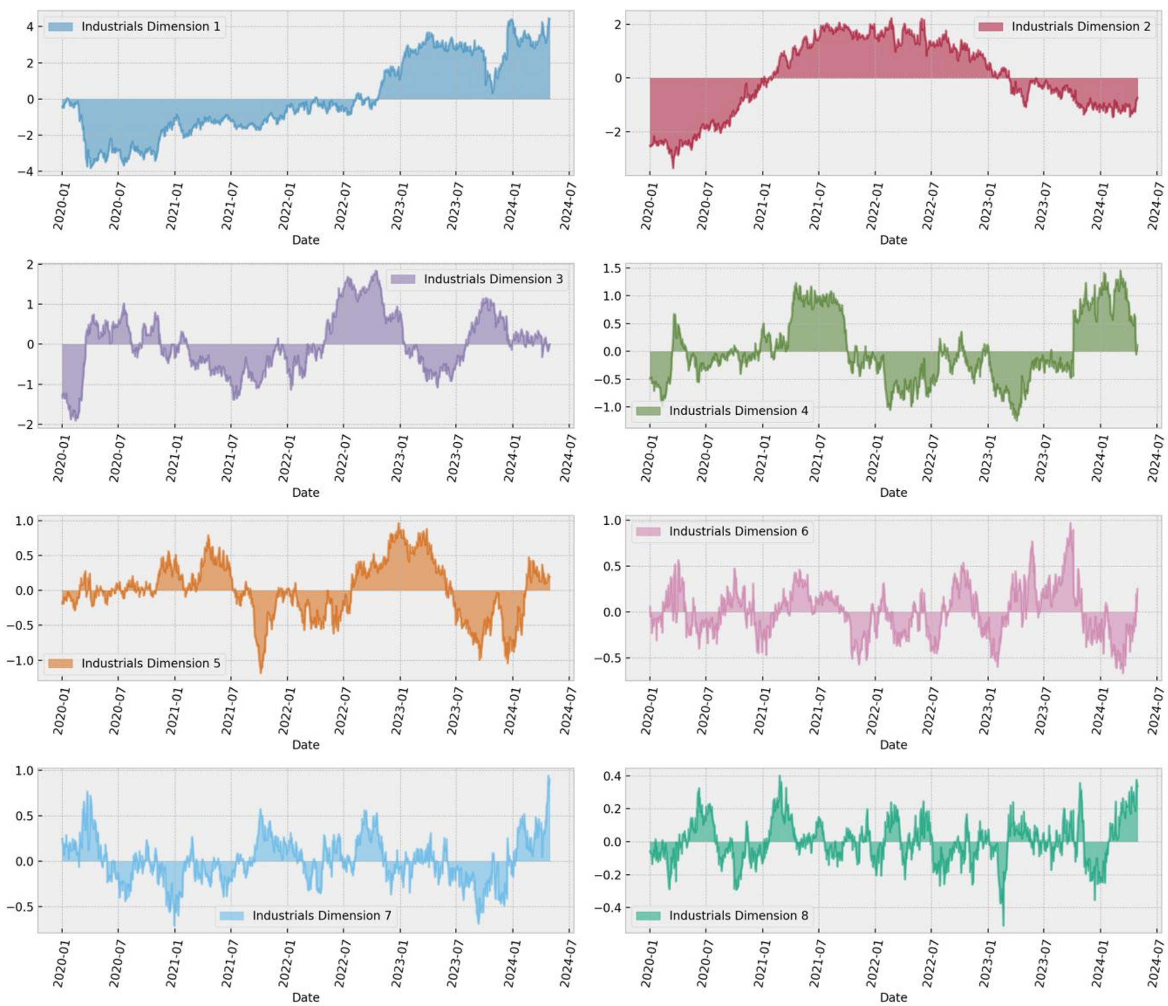
Linear Combinations ETFs (12 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 6.

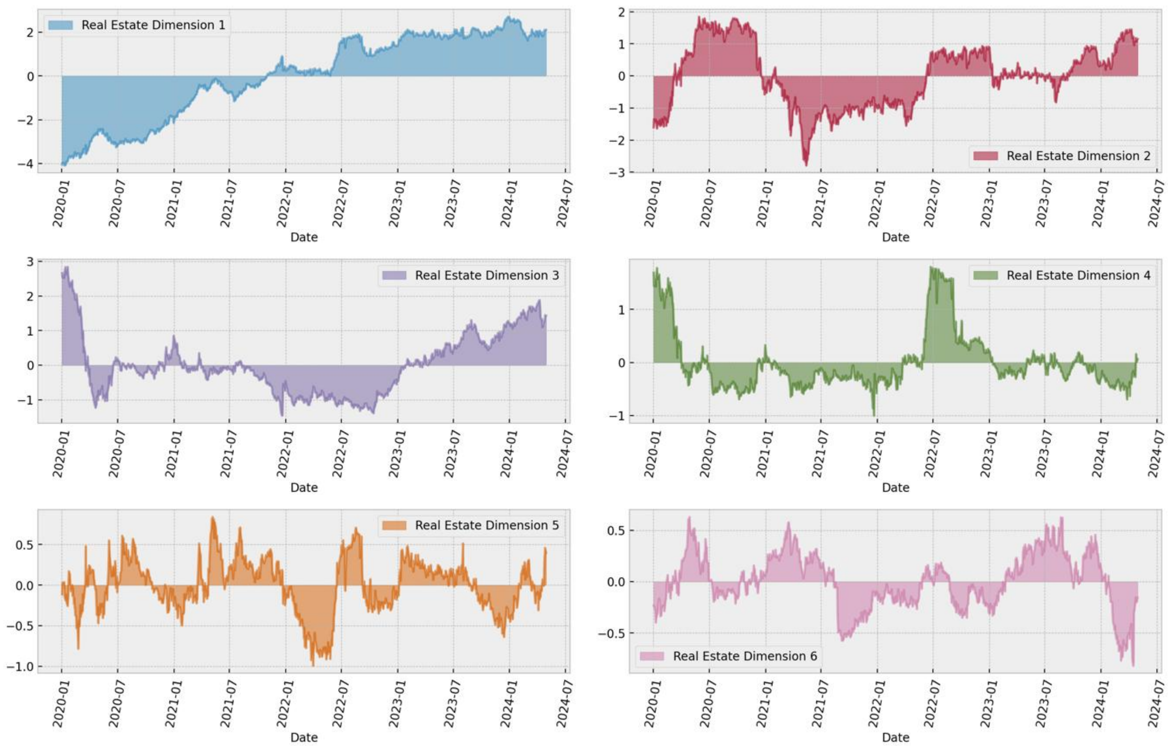
Linear Combinations Industrials (8 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 7.

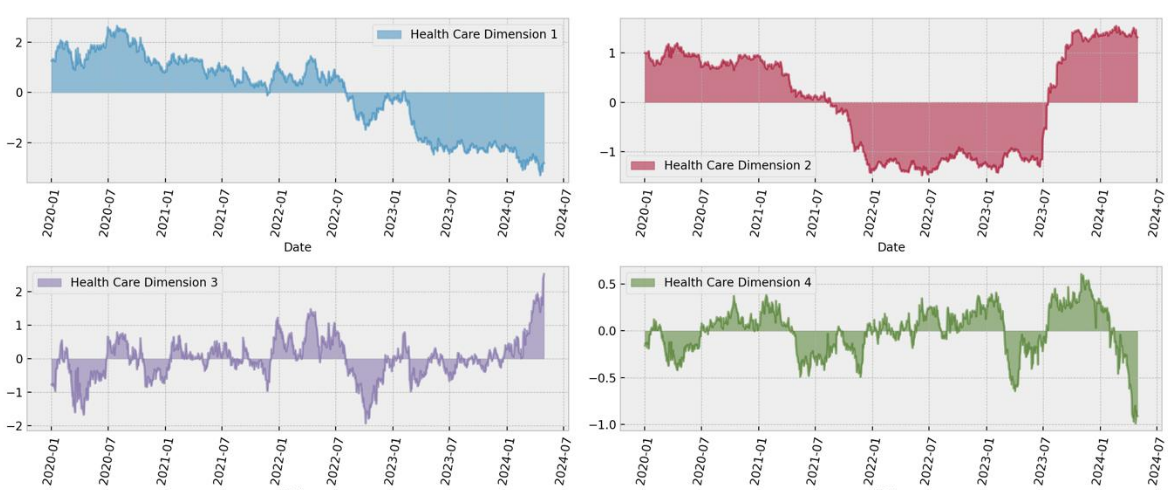
Linear Combinations Real Estate (6 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 8.

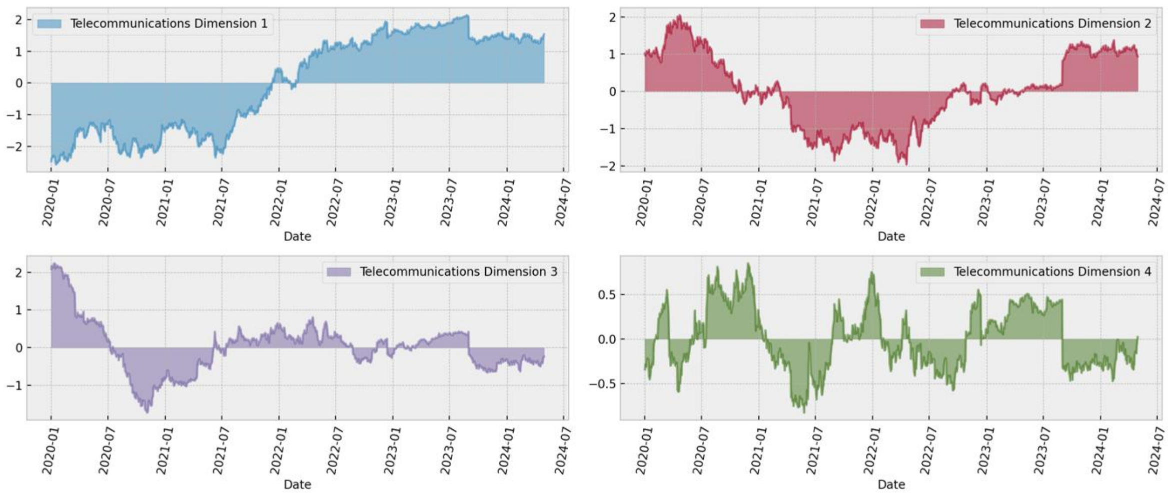
Linear Combinations Health Care (4 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 9.

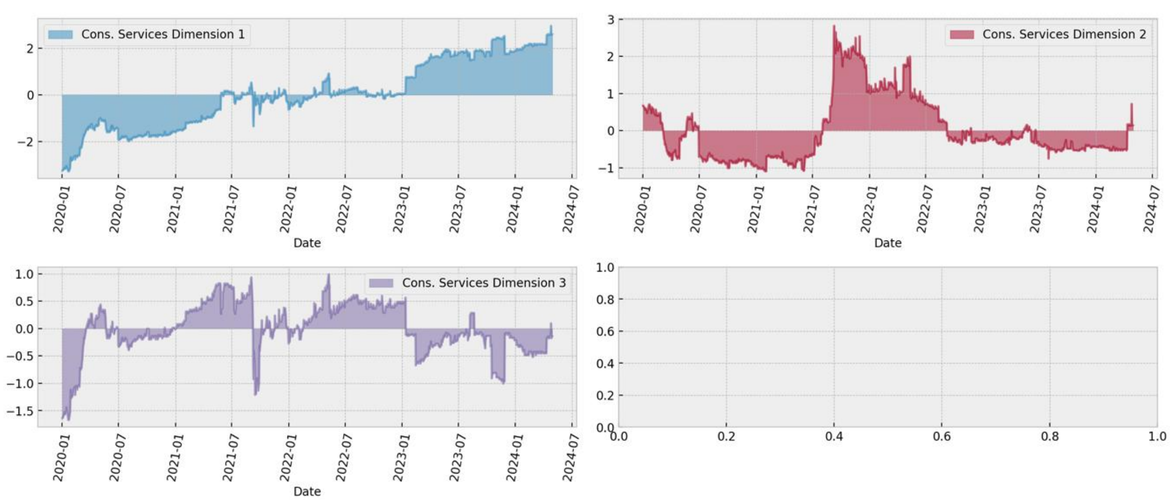
Linear Combinations Telecommunications (4 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 10.

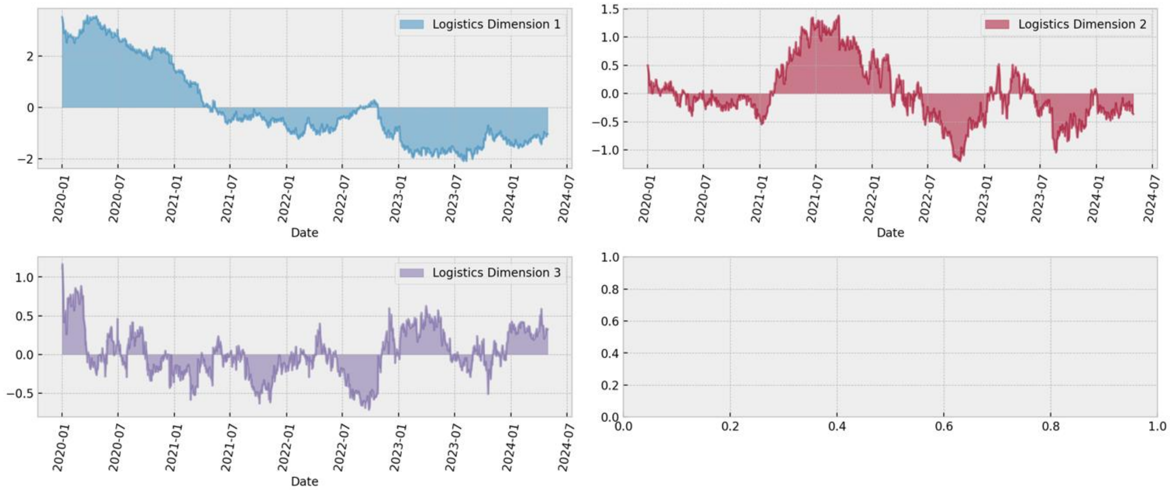
Linear Combinations Construction Services (3 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.

Appendix 11.

Linear Combinations Transportations and Logistics (3 dimensions)



Source: Own elaboration with information from Yahoo Finance, Python.