Longitudinal Features Extraction in International Logistics Performance Index

ALDINA CORREIA[1,*] DIOGO RIBEIRO[2,3]
[1] CIICESI, ESTG, Instituto Politécnico do Porto,
Rua do Curral, Casa do Curral, Margaride, 4610-156, Felgueiras, PORTUGAL.
[2] Department of Data Science and Research, MySense, London, UNITED KINGDOM.
[3] ESMAD, Instituto Politécnico do Porto,
Rua D. Sancho I, n.º 981, 4480-876, Vila do Conde, PORTUGAL.
[*] *Corresponding Author*

Abstract: The importance of the logistics performance of companies, regions and countries to support decision-making is universally recognised, covering the rationalisation of supply chains, the optimisation of inventory management and promoting global collaboration. Efficient logistics integration with innovative technologies is crucial for the prompt delivery of materials and components, increasing the speed and effectiveness of innovation processes and, consequently, the performance of organisations.
The Logistics Performance Index (LPI) is an index created to assesses global logistical performance by measuring factors such as the quality of commercial and transport infrastructure, the ease of customs procedures and the efficiency of customs clearance, among other aspects that influence the transnational flow of goods.
This study examines the robust correlation structure between LPI indicators over several years. Our results confirm the LPI as a longitudinal latent variable, characterised by its indicators, demonstrating excellent internal consistency. This consistency underline the reliability of the LPI for measure global logistics performance of countries. Thus LPI can be recognised as a valuable measure of countries international logistics efficiency, and can be used in practice as a tool for business and politics, guiding strategic decision-making and improving the cost-benefit ratio and competitiveness of organisations.

Key-Words: Logistics Performance, LPI, Logistics Decision-Making, Feature Aggregation, Principal Components Analysis, Longitudinal.

Received: July 9, 2024. Revised: January 27, 2025. Accepted: February 23, 2025. Published: April 4, 2025.

1 Introduction

In the era of globalisation, the speed and efficiency with which companies transport goods across international borders have become fundamental characteristics of their global competitiveness, [1]. This encourages the continued development of performance measures such as the World Bank's Logistics Performance Index (LPI), [2], Global Competitiveness Index (GCI), [3], Global Enabling Trade Index (GETI), [4], and UNCTAD Liner Shipping Connectivity Index, [5]. Access to these measures quickly, automatically and efficiently is crucial, [6], [7].

LPI, measures the performance of countries logistics system and provides a guide for businesses and policymakers interested in global trade and international investments, [8], [9], [10], [11]. The LPI allows countries to assess their current logistics-related strengths and weaknesses, identifying the areas in which they need to improve, and benchmark their performance against global standards, to enhance their international trade capabilities, [2], [12].

Improving the LPI requires a broad and coordinated approach that addresses various factors impacting a logistics system, simplifying and standardising in customs procedures and regulations can significantly reduce the time and cost associated with clearing goods at the border, [2], [6]. Investment in transport infrastructure, the diversification of international transport options, and the optimisation of transportation methods improve the connectivity, reliability, flexibility, and accessibility of logistics networks, [13]. Further investment in human resources, new technologies and innovation is crucial to increase the quality and competence of logistics services. The adoption of digital platforms and tools for better tracking and tracing improves the visibility and security of cargo movements, moreover, reducing variability and uncertainty in logistics operations decreases the probability of failures in delivery, [14].

The relevance of the LPI fundamentally derives from its ability to reflect the nuances of a nation's logistics system – a crucial element for trade effectiveness, economic dynamism and comprehensive social progress, [15], [16]. Nations with high LPI scores often demonstrate more efficient logistics, which not only speeds up the transportation of goods but also strengthens competitiveness and access to markets, [17]. Conversely, low LPI scores can indicate underlying problems, such as high costs, operational delays and significant risks, that can diminish a country's attractiveness in the global trade landscape, [8], [9], [10]. The LPI transcends its function as a simple metric, being crucial for monitoring the effectiveness of logistics reforms and investments, providing policymakers with a tool to assess the impact of strategic changes on national performance, [18].

LPI and his measure ability was very discussed in literature, [13, 19]. It was also considered in business, for example in operational planning for logistics companies [1] based in LPI, in order to include additional factors to increase efficiency, customer satisfaction, and competitiveness, developed a mathematical model including cost, time, customer satisfaction, and environmental impact in the periodic multimodal network.

This paper aims to validate the LPI as a reliable measure of logistical performance by analysing key characteristics and relationships among its indicators. Recognising the LPI as a robust metric allows it to inform strategic decision-making in engineering and other sectors, improving operational cost-effectiveness and competitiveness. Additionally, the paper tracks Portugal's LPI performance from 2007 to 2023, focusing on its changes and implications for logistics strategy. Thus, is intended to reinforme the validity of LPI as a good measure of logistical performance by analysing relationships among its indicators. By recognising the LPI as a feasible and robust metric, it can be employed to inform strategic decision-making across business, logistics, and policy sectors, thereby enhancing operational efficiency and competitiveness on a global scale. For that Principal Component Analysis (PCA) is considered for extracting the essential feature in the data.

The structure considered is divided into four sections. In the first the theme and importance of the LPI index are presented, in the second the main concepts of PCA are highlighted, the third encompasses the data analysis and presents the results obtained, and the fourth consists of the conclusion and future work of the study.

2 Principal Component Analysis

Principal Component Analysis (PCA) is a robust statistical tool utilised extensively in data analysis and machine learning to identify patterns in data and express the data in such a way as to highlight their similarities and differences. Since its establishment, PCA has been widely applied to reduce the dimensionality of large data sets, enhancing interpretability while minimising loss of important information, [20], [21], [22], [23], [24].

PCA provides a comprehensive explanation of the global dataset, retaining the maximum variability in the data and minimising the number of features extracted, which is advantageous for effective analysis.

There are several compelling reasons to consider transforming a dataset into a lower-dimensional space, such as: simplifying data manipulation and reducing computational demands, [25], [26], [27]. However, it is critical to do this transformation systematically, as reducing dimensions can result in a loss of information. It is essential that the chosen algorithm retains the valuable PCA could be introduced from different viewpoints, elucidating why it is advantageous to preserve maximal variability in the data, [28], [29], [30].

The mathematical foundation of PCA is based on the linear algebra concepts of eigenvalues and eigenvectors, [31]. PCA transforms the original variables into a new set of variables, which are linear combinations of the original variables. These new variables, or principal components, are obtained by orthogonal transformation directed along the axes of maximum variance.

The covariance matrix of the initial data is defined as:

$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

The principal components are derived from the eigenvectors of the covariance matrix of the data, scaled by their respective eigenvalues. The first principal component is the direction along which the variance of the data is maximised.

The eigenvalue equation is given by:

$$\Sigma v = \lambda v,$$

where Σ represents the covariance matrix, v denotes an eigenvector matrix, and λ is the corresponding eigenvalue vector.

Dimensionality reduction is achieved by selecting the top k principal components, corresponding to the largest eigenvalues (usually greater than 1), which capture the most significant variance within the dataset. This truncation allows for a lower-dimensional representation of the data, which retains the core characteristics of the original dataset, [32].

The data projection can be expressed as:

$$Y = X \cdot V_k,$$

where X is the original data matrix, V_k is the matrix containing the top k eigenvectors, and Y represents the transformed data with reduced dimensions.

The application of PCA is not limited to dimensionality reduction. It also is as a powerful tool for noise reduction, feature extraction, and data visualisation. By simplifying complexity in high-dimensional data, PCA facilitates a clearer understanding of the underlying data structure, [30].

2.1 Applications of PCA

PCA is extensively applied in several scenarios requiring the reduction of data dimensionality and the extraction of significant patterns from the data, [26]. The utility of PCA extends across several key domains [33], [34, [35], [36]. PCA is particularly valuable in scenarios involving datasets with numerous variables. By identifying and eliminating redundant or less important variables, PCA simplifies the dataset while preserving the most crucial information, thus facilitating more efficient further analyses, [27, 36]. For high-dimensional data, PCA provides a mechanism to reduce dimensions to two or three principal components [37, 38, 39]. This reduction allows for the visual representation of complex structures, making the data more comprehensible and easier to interpret, [29, 40]. In datasets characterised by high levels of noise, PCA enhances data quality by focusing on principal components that capture the core variance of the data, which are less influenced by noise, [30]. In predictive modelling tasks, PCA is employed to derive new features that potentially offer greater insight and less redundancy than the original set of features, [26]. PCA serves as a powerful exploratory tool, helping to uncover hidden variables and detect unusual patterns within the data, which might not be apparent through traditional analysis methods, [27].

Is possible to identify PCA as a valuable tool for dimensionality reduction (extracting the most significant components from complex datasets) and data visualisation (more effective analysis and interpretation of data) across a wide range of fields. PCA applications in Logistics and Supply Chain Management can include performance evaluation, both of employees and of the logistic product or service - allowing a reduction in the number of parameters to be evaluated and a more efficient correction of poor results; the optimisation of logistics processes - allowing to reduce the number of operational metrics, better resource allocation and process optimisation; in risk management - identifying the key components that most significantly impact supply chain stability, helping organisations develop strategies to mitigate potential risks, [41], [42], [43].

In finance and economics, PCA is used to analyse and visualise complex financial data such as portfolio datasets, economic indicator datasets, and others, [44], [45], [46], [47].

In healthcare research, PCA has several important applications, for example, to reduce the dimensionality of imaging data, identify and reduce health data collected on neurodegenerative disorders patients, identify key genetic markers and protein profiles associated with diseases, facilitating more targeted research and personalised medicine, [48], [49], [50].

In environmental science, PCA is used to study the main drivers of algal growth in water reservoirs; in climate change research, identifying the factors causing the greatest variation in climate data; in pollution studies, reducing the dimensionality of pollution metrics, identifying larger sources of pollution, [38], [51], [52].

When it comes to marketing, PCA helps identify distinct market segments by analysing customer data on preferences, purchasing behaviour, and demographics; Knowing the consumer profile also facilitates the development of products according to different profiles, [41], [53].

2.2 Key Assumptions of PCA

PCA serves as an influential statistical method extensively employed to reduce dimensionality in data analysis. Nonetheless, the success of PCA depends on certain assumptions regarding the data it handles. Understanding these assumptions is essential to guarantee that PCA's application yields significant and dependable outcomes, [28], [31], [32], [54], [55].

Key Assumptions of PCA are: (i) Linearity: PCA assumes that the data components have linear relationships among them. This assumption is fundamental because PCA aims to capture the variance through linear combinations of the original features. It implies that the principal components derived from PCA are linear transformations of the original variables. This assumption also means that PCA may not effectively capture complex nonlinear relationships without transformations or adaptations; (ii) Scale: The scale of the data matters in PCA because it directly influences the resulting principal components. Variables with larger variances can dominate the outcome, overshadowing the contributions of other variables. Therefore, it is often necessary to standardise or normalise (or re-scaling) data before applying PCA, so each variable contributes equally to the analysis; (iii) Variance: PCA operates under the assumption that directions in which the variance of the data is maximised are the most important. The technique prioritises these directions to identify the principal components. This assumption may not always hold true, especially in cases where high variance is due to outliers or noise in the data; (iv) Orthogonality: PCA assumes that the principal components are orthogonal, meaning each component is uncorrelated with the others. This orthogonality ensures that the principal components represent independent dimensions of variance, simplifying the interpretation of the data by removing redundant information; (v) Normality: While not strictly required, PCA ideally assumes that the data follows a multivariate normal distribution. This assumption helps in maximising the efficiency and interpretability of PCA, particularly when using PCA results for further statistical inference. The normality assumption is imperative for the application of various statistical tests that may be needed to evaluate the results of PCA; (vi) Size: PCA assumes that there is an adequate sample size to reliably estimate the covariance matrix. A small sample size may lead to over-fitting, where the PCA model captures noise instead of the underlying data structure. Generally, a larger sample size provides a more stable and accurate estimation of covariance, leading to more reliable PCA results.

When applying PCA, it's crucial to assess whether these assumptions hold for the given dataset. Violations of these assumptions can lead to misinterpretations and misleading conclusions. For instance, if the data is not linear, using kernel PCA or another nonlinear method might be more appropriate. If components are not orthogonal due to the presence of multicollinearity, PCA might still reduce dimensionality, but the interpretation of components becomes more complex.

While PCA is a robust and versatile tool in data analysis, careful consideration of its underlying assumptions is essential to utilise its full potential effectively. Ensuring that these assumptions are met or appropriately addressed through pre-processing steps can greatly enhance the insights gained from PCA.

2.3 Validating PCA

To ensure the effectiveness of PCA in specific contexts, several validation steps are recommended, [28], [31], [32], [54], [55]: (i) Optimal number of principal components to retain: Using Scree Plot Analysis – Initiate the validation by examining, for example, the eigenvalues of the covariance matrix through a scree plot. This plot is crucial for suggest the optimal number of principal components to retain, typically identified by the point where the explained variance ceases to decrease significantly; Variance Explained – Quantify the amount of variance each principal component accounts for. The objective is to retain a minimal number of components while capturing a substantial proportion of the data's total variance, with common thresholds ranging from 70% to 90% or Data investigation area – Theoretical investigation in the theme of the data, can suggest the more adequate theoretical number of factor. (ii) Loadings Examination: Analyse the loadings of each principal component to understand the relationships and contributions of the original variables to the components. This analysis can highlight the most significant variables within the dataset. (iv) Cross-Validation: If PCA is applied within predictive modelling, cross-validation should be performed to compare the efficacy of models using both the original and reduced datasets. This step ensures that the reduction process does not omit critical information. (v) Reconstruction Error: The reconstruction error is the discrepancy between the original dataset and its reconstruction, from the selected principal components. A minimal reconstruction error indicates effective information retention by the principal components. (vi) Qualitative Evaluation: Assess the practicality of the PCA-reduced dataset for the intended application. Considerations should include the balance of variable importance, as over-simplification could result in the loss of essential information.

2.4 Statistical Tests for PCA Suitability

Prior to executing PCA, it is important to perform statistical tests to determine if the data is appropriate for such analysis: Bartlett's Test of Sphericity –This test checks if the correlation matrix significantly deviates from an identity matrix. A significant result implies that the conditions are favourable for PCA.; Kaiser-Meyer-Olkin (KMO) Test – This test measures the suitability of data for factor analysis, which also reflects on its appropriateness for PCA. A KMO value above 0.6 is considered adequate, with values approaching 1.0 being optimal. and Measure of Sampling Adequacy (MSA) – Within the KMO test, the MSA score is calculated for each variable to assess individual suitability. Values below 0.5 generally suggest that the variable is not suitable for PCA.

3 Data and Results

The LPI, [56], is an aggregate index for a welfare evaluation of countries' logistics performance worldwide based on multiple variables. The LPI is derived from a survey of international freight forwarders and express carriers around the world and is an overall impression of countries' performance in six key dimensions, [14], [57]:

1. Customs Clearance Process: The efficiency of customs and border clearance processes.
2. Infrastructure Quality: The quality of trade and transport-related infrastructure, including ports, railroads, and roads.

3. Ease of Arranging (International) Shipments: The ease of arranging competitively priced shipments.
4. Logistics Competence and Quality of Services: The competence and quality of logistics services, such as transport operators and customs brokers.
5. Timeliness of Shipments: The frequency with which shipments reach the consignee within the scheduled or expected delivery time.
6. Tracking and Tracing: The ability to track and trace consignments.

The present study considers LPI indicators over several years, being therefore a longitudinal approach, using the software R. The excel data file, the results and R code are available in [58].

This analysis is based on the values of the LPI, [17], indicators in the years 2007, 2010, 2012, 2014, 2016, 2018 and 2023 (Figure 1, Appendix).

The scores for each of these aspects of logistics performance are weighted averages of country scores and hard data that capture logistic efficiency by quantifying the ease with which goods cross borders and circulate in the country, that is, to allow goods to flow into, and to transit within, the country in a speedy and cost-efficient manner. In Table 2, in Appendices, were presented the descriptive statistics of LPI indicators, for all the years. In Table 1 were presented the descriptive statistics of LPI indicators, for each one of the years.

Generally across all six dimensions, the descriptive measures suggest an upward trend over the years 2007 to 2023. This indicates a general trend towards improvement in logistics performance in all countries over the years. Improvements in customs, infrastructure and service quality playing a significant role. The relatively higher scores for Timeliness and Tracking and Tracing indicate that countries have been particularly successful in ensuring that shipments reach their destinations on time and can be monitored effectively.

Having into account the descriptive statistics of LPI indicator in Table 1 (Appendix A) a more detailed analysis for each one of the years, can be done:

- Customs scores have increased from an average of 2.556 in 2007 to 2.800 in 2023, with the standard deviation relatively stable, indicating consistent improvement in more countries.
- Infrastructure scores started with an average of 2.581 in 2007, infrastructure scores have also improved, reaching an average of 2.922 in 2023. The standard deviation suggests moderate variability between countries' infrastructure quality.
- International Ships scores range from 2.718 in 2007 to 2.925 in 2023, with relatively low standard deviations indicating less variability in performance compared to other indicators.
- The indicator Logistics Competence and Quality showed a significant improvement, with average scores increasing from 2.706 in 2007 to 3.029 in 2023. The consistent increase highlights the highlights of the quality and competence of the logistics service.
- Timeliness had the highest average scores across all dimensions, starting with 3.170 in 2007 and reaching 3. 242 in 2023, demonstrating that xxx are increasing on time.
- Tracking and Tracing scores have improved from an average of 2.729 in 2007 to 3.051 in 2023, with standard deviations indicating moderate consistency in tracking capabilities across different countries.

To identify the variables that most contribute to the countries' logistical performance, techniques for selecting or extracting key indicators can be used. In both cases, the aim is to retain the indicators that maximise the variance extracted from the original data.

In [25], to ensure that the extracted characteristics were relevant and informative in relation to the 2023 Logistics Performance Index indicators, extraction techniques were used, in particular Exploratory Factor Analysis (EFA), using PCA in the JASP, JASP, software.

As the EFA carried out in [25], allowed the identification of a single factor, in this work a Principal Component Analysis (PCA) was carried out over the available years, considering one component.

Assumptions for this techniques (EFA and PCA) are: (i) sample dimension – 5 to 10 observations by variable; (ii) normality of the variables; (iii) linearity of the variables; (iv) homocedasticity.

(i) Sample dimension - The sample comprises between 139 and 160 countries (Figure 1), and there is no missing values. Therefore, the sample size is adequate for the analysis of the six indicators.

(ii) Normality Assessment - Both the Mardia Test and the Energy Test were used, under the null hypothesis of multivariate normality. In some year Mardia's Test presented p values greater than 0.05, therefore, a multivariate normal distribution can be considered, but in most years none of the tests allows to confirm this assumption.

(iii) Linearity – Pearson's Correlations between Indicators are significant at the 5% significance level, in addition, they are all high, exceeding the value of 0.8, suggesting a strong interdependence between them (multicollinearity), as illustrated in Figure 2 (Appendix A). Consequently, the extraction of factors from these characteristics is justified and imperative, in order to exclude potential redundancies and increase the robustness of the analysis.

(iv) Homoscedasticity – The standard deviation values of the indicators are close to and below 1, which suggests the existence of homoscedasticity and a relatively consistent distribution of data around the respective means for each indicator, over the years.

Once the necessary conditions have been verified, PCA can be implemented to extract the principal component that maximises the retention of variance of the observed variables.

The Kaiser-Meyer-Olkin (KMO) Measure of Sample Adequacy (MSA) or Overal MSA and the MSA for each variable indicates the suitability of the variable for PCA.

KMO ranges between 0.92 and 0.94, over the years and the individual MSA of each variable between 0.88 and 0.98. This all individual MSAs are greater than the 0.5 threshold, indicating that they are adequate for the analysis. Additionally, the overall MSA or KMO values suggests excellent suitability for applying PCA, [59].

In determining the number of factors to retain by focusing on PCA eigenvalues exceeding 1, [60], a single factor was extracted over the years. The percentage of variance retained of these indicators over the years is between 87.8 and 94.3 (Table 4, in Appendices). This means that all indicators contribute to the same latent variable, indicating a high level of correlation between them.

Within this factor, the indicator *Logistics Competence and Quality* is always that one with highest loading. The behaviour of the loading's values across the time can be observed in Table 4 and Figure 3, in Appendices. Loading's values are between the values 0.86 and 0.98.

The *Infrastructure* indicator presented the second highest value since 2007, but was surpassed by *Tracking and Tracing* in 2023.

The *Customs* indicator has shown almost constant behaviour over the years, generally appearing in third or fourth position, in relation to loading values.

The indicators with lower values were *Timeliness* and *International Shipments*, particularly in 2010.

Since 2016, the loading's of all indicators have been decreasing.

The reliability of factors in the context of factor analysis refers to the consistency or stability of the measurement of underlying constructs represented by those factors. There are several ways to assess the reliability of factors, being the commonly used measure the Cronbach's Alpha, a frequentist scale Reliability Statistics. It assesses how well the items within a factor consistently measure the same underlying construct. A higher Cronbach's alpha value (typically above 0.70) indicates greater reliability.

The estimate Cronbach's α of the factor obtained across the time is between 0.970 and 0.987, indicating excellent reliability, [59]. This provides evidence of the excellent internal consistency of the latent variable.

Cronbach's α of the extracted component obtained over time under analysis is between 0.970 and 0.987, indicating excellent reliability, [59]. This provides evidence of the excellent internal consistency of the latent variable over the years.

4 Discussion

Having into account our results is possible sate that the LPI is an important tool for evaluating the logistical performance of countries around the world, capturing their logistic efficiency and quality. This study used PCA to study the existence of patterns and trends in the correlation structure of LPI indicators over several years, providing insights into the evolution of logistics performance in different countries.

This idea is mostly consensual in the literature, but some authors do not recognise the LPI as a good indicator, for example in [2], the authors presented a study that aimed to improve the LPI, claiming that it is based on a global survey of logistics experts, which may be biased towards a subjective view of the logistics systems of different countries . So they developed a measure, which they believe is more objective and comprehensive, based on international statistical data, to accurately assess logistics systems in 159 countries.

Our analysis allows us to conclude that the LPI indicators are highly correlated, as indicated by the high KMO values (ranging from 0.92 to 0.94 over the years). These KMO values reflect an excellent degree of sample adequacy, confirming that PCA is an appropriate technique for this data set, Table 3, in Appendices.

PCA results show that most of the variance of the LPI indicators over the years (between 87.8% and 94.4%) can be explained by a single principal component, highlighting the strong coherence between the six LPI dimensions and their collective influence on performance general logistics, Table 4 and Figure 3.

The indicator *Logistics Competence and Quality* presented the highest loadings in all years. This result highlights the critical role that logistics competence and quality play in determining a country's overall logistics performance. The high loads suggest that strengthening logistical skills, service quality and operational competence can be fundamental strategies for countries that want to improve their LPI scores. Investment in training, quality assurance and logistics capacity building could therefore produce significant benefits. The results obtained by the authors in [13], already showed that improvements in any of the LPI components can lead to significant growth in a country's trade flows.

The *Infrastructure* and *Tracking and Tracing* indicators also showed substantial contributions to the principal component. The *Infrastructure* indicator was the second largest contribution since 2007, but was surpassed in 2023 by the indicator *Tracking and Tracing*, that is, location and tracking capabilities. This change may point to the importance of monitoring and managing logistics processes in real time. As the logistics industry evolves, digital infrastructure and advanced tracking systems become increasingly crucial. Countries can benefit, if they decide to invest in smart logistics solutions, digital infrastructure and real-time tracking technologies. The countries and companies can, for example, follow the suggestions on the role of Smart Logistics in the development of Smart Cities, [61], [62], [63].

Regarding the loadings for *Customs* indicator, the results show that it maintained a relatively stable loading, generally appearing in third or fourth position. This consistency highlights that although customs efficiency is important, there were no dramatic changes in relation to other logistics performance indicators during the observed period. This suggests that improvements in customs processes alone may not be sufficient for significant gains in global logistics performance. Countries must adopt a holistic approach, improving customs processes along with other critical logistical dimensions. the potential differences observed when using different variables. For example, a data envelopment analysis (DEA-LPI) was proposed in [19], including in addition to LPI indicators indicators about income and geographical area. Their findings suggest that the logistics performance depends largely on income and geographical area.

Indicators of *Timeliness* and *International Shipments* showed relatively lower loadings, especially observed in 2010. This may imply measures to ensure on-time deliveries and manage international shipments effectively. Factors such as geopolitical issues, economic fluctuations and political changes can have an impact on these areas. Improving the punctuality and efficiency of shipments may require specific interventions, such as optimising logistics networks, improving coordination and implementing best practices in supply chain management.

These results are in line with the obtained in [64]. In this work the authors considered a decision tree regression analysis using JASP software revealed that logistics service quality significantly impacts the LPI, followed by trade- and transport-related infrastructure, the ability to track and trace, customs and border management, timeliness of shipments, and availability of competitively priced international shipments.

5 Conclusion and Future Work

In this study it is concluded that there is a strong correlation between all LPI indicators over the years, suggesting a good definition of the LPI as latent variable. This correlation structure remains almost unchanged over the time, with the main indicator always being the same over the years. In fact, the positioning of the loadings of all indicators is at values above 0.85, therefore all of them are very close to 1.

The LPI also presents a notable internal consistency over the years, meaning the high level of correlation between these indicators, which implies a coherent and unified measurement of the latent variable, reinforcing the reliability and coherence of the LPI as a robust tool for evaluating logistics performance.

This vision contributes to an understanding of logistics performance and can be generalised to organisations as a reference measure on a national and global scale, guiding the decision-making and strategic decisions with regard to the logistics performance of organisations, regions and countries.

By using LPI as a benchmark measure, organisations can more accurately assess and compare their logistics capabilities to industry standards and best practices. This approach enables policymakers, government officials and business leaders to make informed strategic decisions that improve logistics infrastructure, optimise supply chains and improve overall performance.

For example, national governments can use the LPI to identify areas that require investment or policy intervention to improve their logistics systems, thereby promoting economic growth and improving competitiveness in global markets.

At the organisational level, companies can leverage LPI data to evaluate their logistics operations, identify gaps in performance, and develop specific strategies to resolve inefficiencies. This allows organisations to streamline operations, reduce costs and improve service delivery, ultimately leading to increased customer satisfaction and competitive advantage.

Furthermore, the global applicability of the LPI provides a common framework for international collaboration and harmonisation of logistics standards, facilitating more harmonious trade and stronger economic partnerships. By aligning logistics performance metrics across countries, it becomes easier to identify best practices and share knowledge, driving improvements across the global logistics landscape.

In conclusion, the results of this study not only validate the LPI as an effective tool for measuring logistics performance, but also highlight its practical implications for strategic decision making.

A good suggestion for future work is to develop a logistics performance index for logistics companies and extending these concepts to the industrial level. This would be very useful for these companies.

Another suggestion is to consider a Longitudinal approach to PCA, using for example the R package, [65], and also a model based clustering and dimensionality reduction of mixed data, [66], approach.

Declaration

The authors wrote, reviewed and edited the content as needed and they have not utilised artificial intelligence (AI) tools. The authors take full responsibility for the content of the publication.

References:

[1] Eyakem Fikru and Gobu Eticha. Freight transport service on international trade with its implication for global competitiveness. *Studies in Humanities and Education*, 1(1):21–39, 2020.

[2] Ruslan Beysenbaev and Yerkin Dus. Proposals for improving the logistics performance index. *The Asian Journal of Shipping and Logistics*, 36(1):34–42, 2020.

[3] World Economic Forum. The global competitiveness report 2020. `https://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2020.pdf`, 2020. Accessed: 2024-10-16.

[4] World Economic Forum. The global enabling trade report 2018. `https://www.weforum.org/reports/the-global-enabling-trade-report-2018`, 2018. Accessed: 2024-10-16.

[5] UNCTAD. Review of maritime transport 2020. United Nations Conference on Trade and Development, `https://unctad.org/webflyer/review-maritime-transport-2020`, 2020. Accessed: 2024-10-16.

[6] Bulent Babayigit, Fatih Gürbüz, and Burak Denizhan. Logistics performance index estimating with artificial intelligence. *International Journal of Shipping and Transport Logistics*, 16(3-4):360–371, 2023.

[7] Ben Shepherd and Thanachart Sriklay. Extending and understanding: an application of machine learning to the world bank's logistics performance index. *International Journal of Physical Distribution & Logistics Management*, 53(9):985–1014, 2023.

[8] Jean-François Arvis, Lauri Ojala, Ben Shepherd, Dina Ulybina, and Christina Wiederer. *Connecting to Compete 2023: Trade Logistics in the Global Economy – The Logistics Performance Index and Its Indicators*. World Bank Group, Washington, DC, 2023.

[9] Jean-François Arvis, Lauri Ojala, Christina Wiederer, Ben Shepherd, Anasuya Rajand, Karlygash Dairabayeva, and Tuomas Kiiski. *Connecting to Compete 2018: Trade Logistics in the Global Economy – The Logistics Performance Index and Its Indicators*. World Bank Group, Washington, DC, 2018.

[10] Jean-François Arvis, Daniel Saslavsky, Lauri Ojala, Ben Shepherd, Christina Busch, Anasuya Raj, and Tapio Naula. *Connecting to Compete 2016: Trade Logistics in the Global Economy – The Logistics Performance Index and Its Indicators*. World Bank Group, Washington, DC, 2016.

[11] Mehmet E. Civelek, Nazif Uca, and Mustafa Çemberci. The mediator effect of logistics performance index on the relation between global competitiveness index and gross domestic product. *European Scientific Journal*, 11(13), 2015.

[12] Mustafa Polat, Kemal Kara, and Aziz Z. Acar. Competitiveness based logistics performance index: An empirical analysis in organisation for economic co-operation and development countries. *Competition and Regulation in Network Industries*, 24(2-3):97–119, 2023.

[13] Luis Martí, Rodolfo Puertas, and Luis García. The importance of the logistics performance index in international trade. *Applied Economics*, 46(24):2982–2992, 2014.

[14] J. Arvis, L. Ojala, C. Wiederer, B. Shepherd, A. Rajand, K. Dairabayeva, and T. Kiiski. Connecting to compete 2018: Trade logistics in the global economy – the logistics performance index and its indicators. Technical report, Washington, DC: World Bank Group, 2018.

[15] Jafar Rezaei, Wouter S. van Roekel, and Lorant Tavasszy. Measuring the relative importance of the logistics performance index indicators using best worst method. *Transport Policy*, 68:158–169, 2018.

[16] Muhammad Sharif, Bo Yiaw, Anuar Ismail, Umi Khand, Ghassan Hassan, and Ahmed Shahrn. Leveraging technology for sustainable logistics: Logistics performance index in inland ports with eco strategy and sustainable practices. *International Journal of Intelligent Systems and Applications in Engineering*, 12(1):666–675, 2024.

[17] World Bank. International lpi global ranking. `http://lpi.worldbank.org/international/global/ranking`, 2023. Accessed: 2024-10-16.

[18] World Bank. About | logistics performance index. `http://lpi.worldbank.org/about`. Accessed: 2024-10-16.

[19] Luisa Martí, Juan Carlos Martín, and Rosa Puertas. A dea-logistics performance index. *Journal of applied economics*, 20(1):169–192, 2017.

[20] Ian T Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 2016.

[21] João Marôco. *Análise Estatística com o SPSS Statistics*. ReportNumber, 8 edition, 2021.

[22] Adam H Monahan. Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system. *Journal of Climate*, 13(4):821–835, 2000.

[23] Yoshio Takane and Michael A Hunter. Constrained principal component analysis: a comprehensive theory. *Applicable Algebra in Engineering, Communication and Computing*, 12(5):391–419, 2001.

[24] Michael W. Watkins. Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3):219–246, 2018.

[25] Aldina Correia and David Mendes. International logistics performance feature extraction insights and portugal's global positioning. In Jose Machado, Filomena Soares, Justyna Trojanowska, Vitalii Ivanov, Katarzyna Antosz, Dagmar Cagáňová, Vijaya Kumar Manupati, and Alejandro Pereira, editors, *Innovations in Industrial Engineering III*, pages 300–311, Cham, 2024. Springer Nature Switzerland.

[26] Xue Wang and Kuldip K Paliwal. Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*, 36(10):2429–2439, 2003.

[27] Jin Wu, Jianhui Wang, and Li Liu. Feature extraction via kpca for classification of gait patterns. *Human Movement Science*, 26(2):393–411, 2007.

[28] Rasmus Bro and Age K Smilde. Principal component analysis. *Analytical Methods*, 2014.

[29] Bahaa M Sabbar Hasan and Adnan M Abdulazeez. A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining*, 2(1):34–45, 2021.

[30] Gordana Ivosev, Lucas Burton, and Ron Bonner. Dimensionality reduction and visualization in principal component analysis. *Analytical Chemistry*, 80(17):6717–6725, 2008.

[31] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, 1987.

[32] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[33] Ana Teixeira, Marisa R Ferreira, Aldina Correia, and Vanda Lima. Students' perceptions of corporate social responsibility: Evidences from a portuguese higher education institution. *International Review on Public and Nonprofit Marketing*, 15(2):235–252, 2018.

[34] Vitor Braga, Aldina Correia, Alexandra Braga, and Sofia Lemos. The innovation and internationalisation processes of family businesses. *Review of International Business and Strategy*, 27(2):231–247, 2017.

[35] Ana Rita Gonçalves Lopes, Aldina Isabel de Azevedo Correia, and Alexandra Maria da Silva Braga. Motivations for and barriers to innovation in non-profit organizations: The case of nursing homes in northern portugal. *International Journal of Innovation Studies*, 8(1):25–44, 2024.

[36] Diana Machado, Vítor Braga, Aldina Correia, Alexandra Braga, and Carina Silva. How female entrepreneurship may boost business innovation and internationalization. *Global Business and Organizational Excellence*, 42(5):91–110, 2023.

[37] Carlos Filipe Gomes, Vítor Braga, and Aldina Correia. Government support for new and growing firms: Evidence from gem data. *Juventud, Emprendimiento y Universidad: Un reto a tres bandas*, page 40.

[38] Eliana Costa e Silva, Isabel Cristina Lopes, Aldina Correia, and A. Manuela Gonçalves. Application of factor analysis to the water quality in reservoirs. In *AIP Conference Proceedings*, volume 1836(1), pages 020074–1–5. AIP Publishing, 2017.

[39] Aldina Correia, Ana Machado, Alexandra Braga, Vıtor Braga, and M Fátima Almeida. Marketing innovation using cis portuguese dataset. *Int. J. Ecol. Stat*, 5:27–33, 2017.

[40] Isabel Cristina Lopes, Aldina Correia, A Manuela Gonçalves, et al. Hierarchical clusters of phytoplankton variables in dammed water bodies. In *AIP Conference Proceedings*, volume 1836 (1). AIP Publishing, 2017.

[41] Yenny Alejandra Aguirre-Álvarez, Mauricio Montoya-Peláez, Raúl Trujillo-Cabezas, and Camilo Esteban Solano Hurtado. Competency development in industrial engineering and related academic programs regarding the optimization of production and logistics processes: The case of medellín. *Trilogía Ciencia Tecnología Sociedad*, 15(31):1–24, 2023.

[42] Zixue Guo and Yi Zhang. The third-party logistics performance evaluation based on the ahp-pca model. In *2010 International Conference on E-Product E-Service and E-Entertainment*, pages 1–4. IEEE, 2010.

[43] Jing Wang, Christopher LE Swartz, Brandon Corbett, and Kai Huang. Supply chain monitoring using principal component analysis. *Industrial & Engineering Chemistry Research*, 59(27):12487–12503, 2020.

[44] Marco Avellaneda. Hierarchical pca and applications to portfolio management. *Revista mexicana de economía y finanzas*, 15(1):1–16, 2020.

[45] Makrand Wagale and Ajit Pratap Singh. Socio-economic impacts of low-volume roads using a mixed-method approach of pca and fuzzy-topsis. *International Review for Spatial Planning and Sustainable Development*, 9(2):112–133, 2021.

[46] Xiaoqun Wang. On the effects of dimension reduction techniques on some high-dimensional problems in finance. *Operations Research*, 54(6):1063–1078, 2006.

[47] Ana Borges, Aldina Correia, Eliana Costa e Silva, and Glória Carvalho. The dynamics between structural conditions and entrepreneurship in europe: Feature extraction and system gmm approaches. *Mathematics*, 10(8):1349, 2022.

[48] Josefa Díaz Álvarez, Jordi A Matias-Guiu, María Nieves Cabrera-Martín, José L Risco-Martín, and José L Ayala. An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders. *BMC bioinformatics*, 20:1–12, 2019.

[49] Ramesh Kumar Lama, Jeonghwan Gwak, Jeong-Seon Park, and Sang-Woong Lee. Diagnosis of alzheimer's disease based on structural mri images using a regularized extreme learning machine and pca features. *Journal of healthcare engineering*, 2017(1):5485080, 2017.

[50] Bo Wang and Michael A Kennedy. Principal components analysis of protein sequence clusters. *Journal of structural and functional genomics*, 15:1–11, 2014.

[51] Susana Fernández, Tomas Cotos-Yanez, Javier Roca-Pardiñas, and Celestino Ordóñez. Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in northwest spain. *Geoderma*, 311:120–129, 2018.

[52] Jacquelynne R King and Donald A Jackson. Variable selection in large environmental data sets using principal components analysis. *Environmetrics: The official journal of the International Environmetrics Society*, 10(1):67–77, 1999.

[53] Eric T Bradlow. Exploring repeated measures data sets for key features using principal components analysis. *International Journal of Research in Marketing*, 19(2):167–179, 2002.

[54] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

[55] Martin J McKeown and Terrence J Sejnowski. Independent component analysis of fmri data: examining the assumptions. *Human Brain Mapping*, 6(5-6):368–372, 1998.

[56] World Bank. Logistics performance index homepage. https://lpi.worldbank.org/international/global, 2024. Accessed: 2024-01-30.

[57] J. Arvis, D. Saslavsky, L. Ojala, B. Shepherd, C. Busch, A. Raj, and T. Naula. Connecting to compete 2016: Trade logistics in the global economy – the logistics performance index and its indicators. Technical report, Washington, DC: World Bank Group, 2016.

[58] Diogo Ribeiro. International logistics performance. https://github.com/DiogoRibeiro7/international_logistics_performance/, 2024. Accessed: 2024-10-16.

[59] Joseph F. Hair, William Black, Barry Babin, and Rolph E. Anderson. *Multivariate Data Analysis*. Pearson, New Jersey, 7 edition, 2010.

[60] John S. Carroll. The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, 14(1):88–96, 1978.

[61] Gülçin Büyüközkan and Öykü Ilıcak. Smart urban logistics: Literature review and future directions. *Socio-Economic Planning Sciences*, 81:101197, 2022.

[62] Gaspare D'Amico, Katarzyna Szopik-Depczyńska, Izabela Dembińska, and Giuseppe Ioppolo. Smart and sustainable logistics of port cities: A framework for comprehending enabling factors, domains and goals. *Sustainable Cities and Society*, 69:102801, 2021.

[63] Jerzy Korczak and Kinga Kijewska. Smart logistics in the development of smart cities. *Transportation Research Procedia*, 39:201–211, 2019.

[64] Nguyen Thi Ngoc Diep, Tran Quang Canh, and Nguyen Ngọc Thach. How to improve the logistics performance index. In *Partial Identification in Econometrics and Related Topics*, pages 481–494. Springer, 2024.

[65] Anders Hagen Jarmund, Torfinn Støve Madssen, and Guro F Giskeødegård. Alasca: an r package for longitudinal and cross-sectional analysis of multivariate data by asca-based methods. *Frontiers in Molecular Biosciences*, 9:962431, 2022.

[66] Monia Ranalli and Roberto Rocci. A model-based approach to simultaneous clustering and dimensional reduction of ordinal data. *psychometrika*, 82:1007–1034, 2017.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
This work has been supported by national funds through FCT - Fundação para a Ciência e Tecnologia through project UIDB/04728/2020.

Conflicts of Interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.

A Appendices

Table 1: Descriptive Statistics of LPI indicators

		Median	Mean	Std. Deviation	Minimum	Maximum
Customs Score	2007	2.375	2.556	0.618	1.300	3.992
	2010	2.381	2.594	0.617	1.333	4.038
	2012	2.510	2.655	0.577	1.665	4.099
	2014	2.582	2.726	0.595	1.500	4.208
	2016	2.593	2.705	0.636	1.111	4.179
	2018	2.577	2.673	0.578	1.571	4.092
	2023	2.700	2.800	0.625	1.500	4.200
Infrastructure Score	2007	2.333	2.581	0.719	1.100	4.290
	2010	2.436	2.637	0.731	1.348	4.336
	2012	2.602	2.765	0.670	1.272	4.258
	2014	2.571	2.766	0.662	1.500	4.323
	2016	2.572	2.755	0.720	1.238	4.439
	2018	2.547	2.723	0.674	1.556	4.374
	2023	2.700	2.922	0.721	1.700	4.600
International Shipments Score	2007	2.604	2.718	0.602	1.222	4.049
	2010	2.832	2.846	0.469	1.333	3.861
	2012	2.760	2.824	0.512	1.571	4.175
	2014	2.811	2.864	0.492	1.705	3.818
	2016	2.765	2.866	0.574	1.363	4.235
	2018	2.748	2.830	0.515	1.804	3.995
	2023	2.900	2.925	0.524	1.700	4.100
Logistics Competence and Quality Score	2007	2.556	2.706	0.668	1.250	4.250
	2010	2.586	2.758	0.636	1.333	4.316
	2012	2.730	2.818	0.591	1.429	4.144
	2014	2.736	2.854	0.583	1.750	4.192
	2016	2.673	2.823	0.645	1.394	4.279
	2018	2.699	2.815	0.610	1.883	4.311
	2023	2.900	3.029	0.646	1.800	4.400
Timeliness Score	2007	3.013	3.170	0.654	1.375	4.529
	2010	3.389	3.409	0.575	1.378	4.580
	2012	3.186	3.258	0.555	1.665	4.394
	2014	3.156	3.253	0.587	1.875	4.706
	2016	3.226	3.269	0.620	2.024	4.796
	2018	3.172	3.236	0.576	2.037	4.410
	2023	3.200	3.242	0.565	2.100	4.300
Tracking and Tracing Score	2007	2.577	2.729	0.696	1.000	4.248
	2010	2.794	2.918	0.650	1.167	4.273
	2012	2.770	2.877	0.614	1.571	4.144
	2014	2.835	2.899	0.582	1.750	4.168
	2016	2.712	2.863	0.700	1.514	4.378
	2018	2.784	2.901	0.613	1.636	4.323
	2023	3.000	3.051	0.675	1.600	4.400

Table 2: Descriptive Statistics of LPI indicators - all years

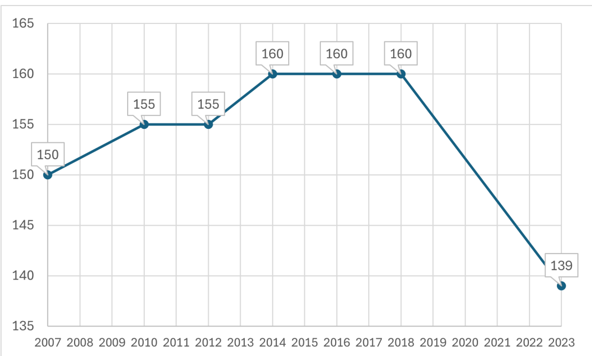
	Median	Mean	Std. Deviation	Minimum	Maximum
Customs Score	2.513	2.672	0.609	1.111	4.208
Infrastructure Score	2.541	2.734	0.704	1.100	4.600
International Shipments Score	2.787	2.839	0.530	1.222	4.235
Logistics Competence and Quality Score	2.700	2.827	0.631	1.250	4.400
Timeliness Score	3.192	3.263	0.593	1.375	4.796
Tracking and Tracing Score	2.785	2.890	0.652	1.000	4.400

Table 3: Kaiser-Meyer-Olkin and MSA

Indicator (Score)	**2007**	**2010**	**2012**	**2014**	**2016**	**2018**	**2023**
Customs	0.91	0.92	0.93	0.89	0.95	0.95	0.91
Infrastructure	0.90	0.88	0.94	0.91	0.93	0.92	0.90
International Shipments	0.96	0.98	0.96	0.97	0.96	0.96	0.97
Logistics Competence and Quality	0.93	0.89	0.91	0.91	0.93	0.91	0.93
Timeliness	0.97	0.98	0.96	0.94	0.94	0.96	0.94
Tracking and Tracing	0.93	0.94	0.94	0.93	0.94	0.95	0.92
KMO - Overall MSA	0.93	0.92	0.94	0.92	0.94	0.94	0.93
Number of Countries	150	155	155	160	160	160	139

Table 4: PCA Loadings, Explaned Variance and Reliability

Indicator (Score)	**2007**	**2010**	**2012**	**2014**	**2016**	**2018**	**2023**
Customs	0.965	0.960	0.960	0.943	0.968	0.958	0.958
Infrastructure	0.972	0.968	0.970	0.973	0.976	0.972	0.964
International Shipments	0.958	0.858	0.948	0.931	0.966	0.932	0.915
Logistics Competence and Quality	0.975	0.973	0.981	0.978	0.982	0.980	0.973
Timeliness	0.932	0.907	0.940	0.935	0.960	0.954	0.939
Tracking and Tracing	0.968	0.951	0.970	0.957	0.976	0.966	0.967
Explaned Variance (%) - 1 factor	92.5	87.8	90.8	92.4	94.3	92.3	90.8
Reliability - Cronbach' α	0.983	0.970	0.982	0.979	0.987	0.982	0.978



Year	**Frequency**	**Percentage**
2007	150	13.902
2010	155	14.365
2012	155	14.365
2014	160	14.829
2016	160	14.829
2018	160	14.829
2023	139	12.882
Total	1079	100

Figure 1: Number of countries across the years

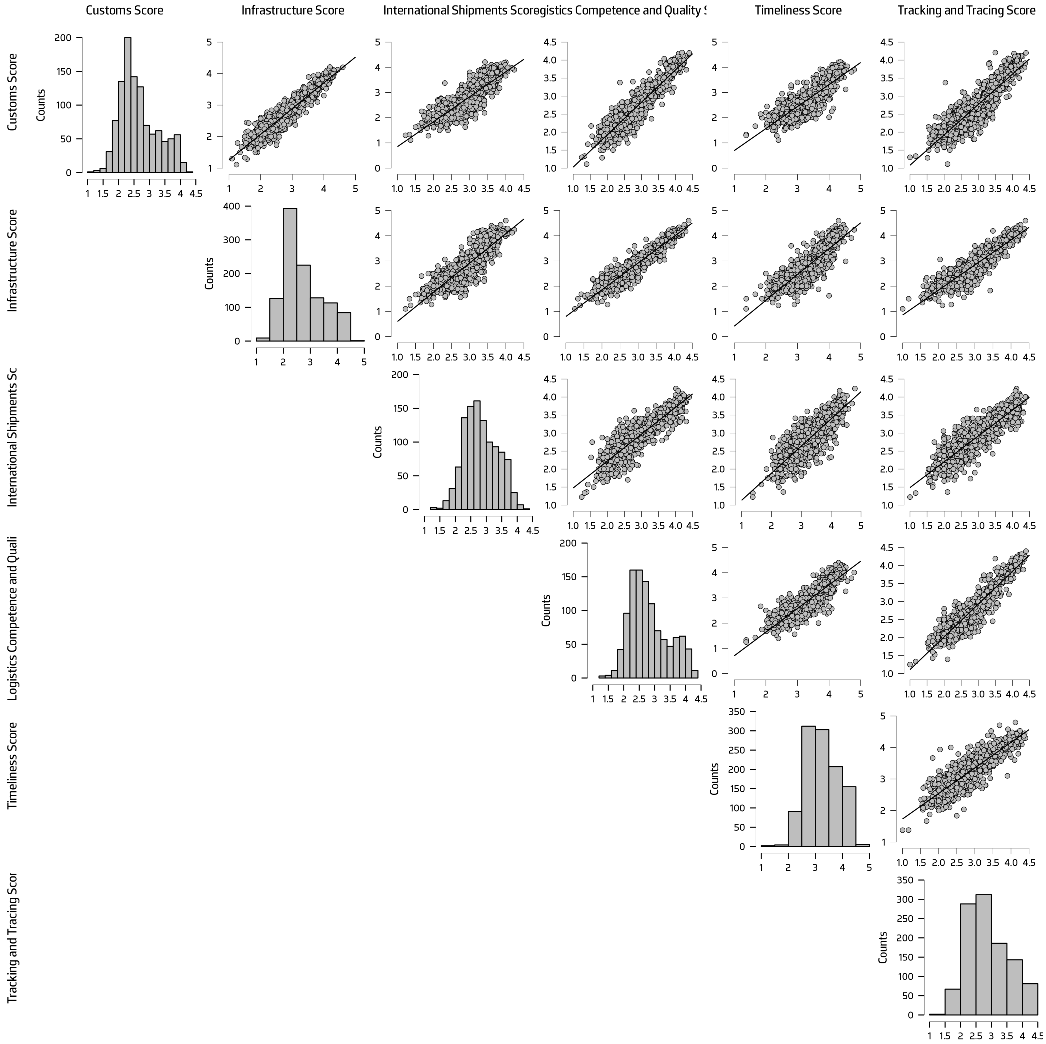


Figure 2: Correlation plot

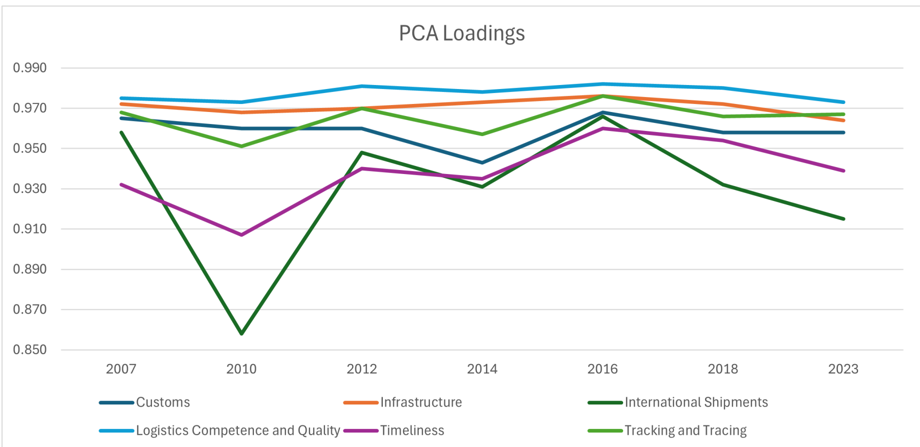


Figure 3: PCA Loadings across the years