UNIVERSITÄT HOHENHEIM

Navigating the indicator jungle: A roadmap to analyze Common Agricultural Policy indicators

Anne-Katrin Gorn, Ayoub Barissoul and Christine Wieck

Hohenheimer Agrarökonomischer Arbeitsbericht Nr. 41

Institut für Agrarpolitik und Landwirtschaftliche Marktlehre

Universität Hohenheim, Stuttgart

September, 2024

UNIVERSITY OF HOHENHEIM

Veröffentlichung des Institutes für
Agrarpolitik und Landwirtschaftliche Marktlehre der Universität Hohenheim

ISSN 1615-0473

Herausgeber: Institut für Agrarpolitik und Landwirtschaftliche Marktlehre
Universität Hohenheim (420)
70593 Stuttgart
https://i420.uni-hohenheim.de/

Gesamtherstellung: Institut für Agrarpolitik und Landwirtschaftliche Marktlehre
Universität Hohenheim (420)
70593 Stuttgart
Zugriff unter: https://i420.uni-hohenheim.de/arbeitsberichte

Contents

List of Figures

List of Tables

List of Appendix

Abbreviations

CAP	European Common Agricultural Policy
OECD	Organisation for Economic Cooperation and Development
PCA	Principal Component Analysis
FA	Factor Analysis
C-α	Cronbach's coefficient alpha
EU	European Union
RDP	Rural Development Programme (European Commission NaN)
LEADER	Liaison Entre Actions de Développement de L`Économie Rurale (European Commission NaN)
MICE	Multivariate imputation by chained equations
NA	Not Available
NaN	Not a Number
%	percent

1. Introduction[1]

By aligning the European Common Agricultural Policy (CAP) with the objectives of the Green Deal and the Farm to Fork Strategy, the European Union aims to create a more sustainable and resilient agricultural sector that can effectively address the challenges of climate change, biodiversity loss, and food security while ensuring the livelihoods of farmers and the well-being of rural communities (European Commission. Directorate General for Agriculture and Rural Development 2024c).

To ensure the CAP achieves its intended objectives effectively, it is essential to conduct thorough evaluations and revisions (European Commission - Directorate-General for Agriculture and Rural Development 2024b). In 2015, the European Commission formulated a Common Monitoring and Evaluation framework serving as a roadmap for assessing the performance of the CAP and identifying areas for improvement (European Commission. Directorate General for Agriculture and Rural Development 2015). The monitoring process employs various policy indicators to gauge the progression of agricultural markets, rural development, and the use of finances (European Commission. Directorate General for Agriculture and Rural Development 2015). Through statistical analyses, the outcomes of the CAP are then evaluated.

The CAP monitoring framework includes a wide range of policy indicators, covering output, results, context, and impact. Output indicators reflect the activities directly implemented through interventions, while context indicators measure overarching contextual trends (European Commission. Directorate General for Agriculture and Rural Development 2015). Impact indicators cover the outcomes of the policy apart from the direct outcomes of the interventions, while result indicators measure the immediate impacts of specific CAP actions (European Commission. Directorate General for Agriculture and Rural Development 2015). Each indicator category covers several indicators. The result indicators, for instance, entail 44 indicators alone (European Commission. Directorate General for Agriculture and Rural Development 2015). With so many indicators, the dataset becomes complex and analyses become intensive.

In environmental policy sciences, the complexities of analysing policy instruments and indicators are known and described (Herman and Shenk 2021). Techniques like statistical pattern analysis and grouping indicators into "umbrella" or composite indicators are recommended to simplify analysis (Herman and Shenk 2021). In the following, both words are used as synonyms. These composite indicators are several single indicators (when grouped

[1] This Hohenheimer Arbeitsbericht is based on a master thesis submitted by Anne-Katrin Gorn.

together then called sub-indicators) grouped together to reduce the number of indicators (Saisana and Tarantola 2002). These methods, prevalent in biological conservation and ecological sciences, are less common in agricultural economics and political sciences.

The first research question of this study explores which methods from multivariate statistics or data mining are suitable for grouping policy indicators together (e.g. calculation of different types of correlation coefficients, factor analysis with eigenvalue decomposition). In keeping up with scientific method development, the aim is to develop a pre-analysis plan for constructing umbrella indicators and evaluate, if this is a useful tool for this type of analysis.

Given the potential overlap of information in CAP indicators as suggested by their descriptions in the monitoring framework (European Commission. Directorate General for Agriculture and Rural Development 2015, see Appendix), investigates whether similar composite indicator methods could be applied to CAP indicators. Specifically, the goal is to identify overarching umbrella indicators within the CAP framework by empirically applying the statistical methods identified in the earlier step.

Composite indicators have gained more importance over the last decades for e.g. comparing country performances regarding competitiveness or assessing sustainability (Freudenberg 2003; Rogge 2012; Nardo et al. 2008; Fusco 2015). Several studies detail constructing composite policy indicators. Zhou and Zhang (2018) outlined steps for creating a composite indicator for sustainability in China. Freudenberg (2003) with the Organisation for Economic Cooperation and Development (OECD) discussed building composite indicators for global country performance comparisons. Additionally, Nardo et al. (2008) with the OECD published a comprehensive handbook, Saisana and Tarantola (2002) created a state-of-the-art guideline on constructing composite indicators, and Sébastien and Bauler (2013) explored composite indicators' significance within EU institutions. Despite the breadth of these methodologies, their application to CAP indicators remains unexplored.

This master's thesis addresses this gap and contributes to existing research by applying these methods to CAP indicators. The methodology for this study is outlined in chapter 2, which presents a comprehensive review of current literature on statistical methods for constructing composite indicators. Chapter 3 summarizes the key findings from this literature review and presents a roadmap for constructing composite indicators. Chapter 4 details the statistical application of these methods to CAP result indicator data, focusing on the identification and creation of specific umbrella indicators to simplify the dataset while preserving essential information. Chapter 5 contextualizes this statistical application by examining trade-offs, potential biases, and information loss, providing practical guidelines for future research and policy analysis. Finally, chapter 6 offers a brief summary of the study's contributions and an outlook for further research.

2. Literature review

To gain a comprehensive overview of the current literature on the construction and analysis of policy indicators, an extensive systematic literature review was conducted using Scopus and Google Scholar. Boolean search operators were employed, specifically (("policy indicators") AND ("multivariate statistics" OR "correlation" OR "factor analysis" OR "regression trees" OR "statistical pattern analysis" OR "umbrella indicators" OR "composite indicators" OR "multivariate techniques")), to refine and focus the search results. The review was done with the first 77 most relevant results. The systematic literature search and selection process is illustrated in figure 1. In addition to the initial search results, references from identified papers were reviewed using the snowball approach to ensure a comprehensive examination of the topic. This literature review helped identify suitable methods for constructing umbrella indicators, summarized in the following section.

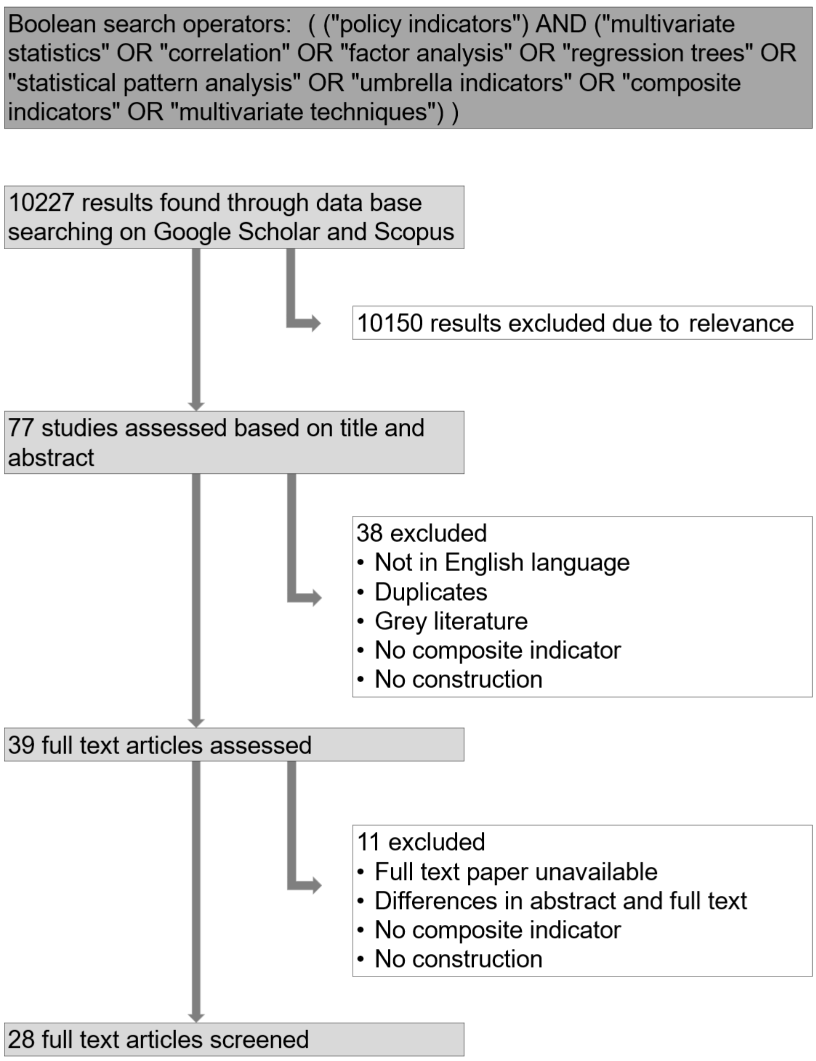


Figure 1: PRISMA flow diagramme of systematic literature search and selection process

3. Roadmap for constructing umbrella indicators

The literature review revealed a relatively consistent structure in the development of umbrella indicators, with the essential elements remaining largely the same across the viewed literature. The findings are summarized and organized in table 1 below, answering this work's first research question. The primary references that informed the creation of the roadmap table include the works of Nardo et al. (2005), Nardo et al. (2008), Freudenberg (2003), Saisana and Tarantola (2002) and Zhou and Zhang (2018).

Table 1: Roadmap for constructing a composite indicator as result of the literature review (based on Nardo et al. (2005), Nardo et al. (2008), Freudenberg (2003), Saisana and Tarantola (2002) and Zhou and Zhang (2018))

Step	Explanation/goal	References for method
1. Selection2	Which indicators are relevant/ Which indicators cover similar aspects?	(Saisana and Tarantola 2002; Botta and Koźluk 2014)
2. Identification	What structure has my data? What is the relation between my sub-indicators?	(Saisana 2004; Saisana and Tarantola 2002; Nardo et al. 2008)
3. Normalization	Ensuring comparability of indicators	(Zhou and Zhang 2018; Freudenberg 2003)
4. Imputing missing data	Filling in missing data	(Nardo et al. 2008)
5. Weighting	Examining the weight of each sub-indicator for the composite indicator	(Saisana and Tarantola 2002; Zhou and Zhang 2018; Nardo et al. 2005)
6. Aggregation	Aggregating several sub-indicators according to their weights to one composite indicator	(Nardo et al. 2005; Saisana and Tarantola 2002; Nardo et al. 2008; Zhou and Zhang 2018; Munda 2012)
7. Uncertainty/ Sensitivity analysis	How much bias or uncertainty is established through creation of one composite	(Nardo et al. 2008; Saisana et al. 2005)

3.1. Selection

In constructing a composite indicator, several sub-indicators are combined to form an umbrella indicator (Saisana and Tarantola 2002). The critical question is which sub-indicators should be grouped together, a decision that depends on the specific aspect being measured and the relevant framework (Saisana and Tarantola 2002; Nardo et al. 2008). Saisana and Tarantola (2002) argue that there is no entirely objective method for selecting sub-indicators. The primary challenge lies in the potential for uncertainty or bias, given that this step is not fully objective (Nardo et al. 2008).

3.2. Identification

After selecting the relevant indicators to describe the phenomenon, it is crucial to identify the data structure, statistical dimensions, and relationships among these indicators (Saisana and Tarantola 2002; Nardo et al. 2008). According to Saisana and Tarantola (2002), this step is fundamental for the subsequent stages of analysis. Multivariate statistics play a key role in this phase, offering a comprehensive understanding of the dataset, its structure, and the interrelationships among the sub-indicators (Nardo et al. 2008; Nardo et al. 2005; Saisana and Tarantola 2002; Hausner et al. 2017).

Nardo et al. (2008) identified two primary dimensions within the dataset: the indicators and the countries. Different analytical methods are employed to explore the structure of each dimension (Nardo et al. 2008).

For the analysis of sub-indicators, multivariate statistical methods such as Principal Component Analysis (PCA) and Factor Analysis (FA) are employed (Nardo et al. 2008). These methods facilitate the identification of dimensions within the data, uncovering similarities and assessing the weights of the indicators (Saisana and Tarantola 2002).

Principal Component Analysis: PCA is a multivariate statistical method used to reduce the complexity of a dataset while preserving as much of the original variance as possible (Saisana and Tarantola 2002; Nardo et al. 2008; Manly 1994). It achieves this by transforming correlated indicators into new, uncorrelated variables known as principal components (Nardo et al. 2008) Each principal component is a linear combination of the original indicators, determined by eigenvectors, which represent the direction of the indicators in the new component, and eigenvalues, which quantify the amount of variance each eigenvector captures (Saisana and Tarantola 2002; Keita 2018).

PCA utilizes a covariance or correlation matrix to identify these linear combinations, which convey the same information as the original indicators but without the redundancy introduced by their correlation (Nardo et al. 2005; Manly 1994). The more correlated the indicators are, the fewer dimensions are needed to capture the variance, resulting in a smaller number of principal components being retained (Nardo et al. 2005; Manly 1994) This reduction in dimensions helps to uncover the underlying structure of the data and provides valuable insights into the interactions and correlations between indicators (Nardo et al. 2008).

Factor Analysis: FA achieves a similar result through a different approach (Saisana and Tarantola 2002). Instead of a correlation matrix, FA uses a statistical model (Nardo et al. 2008; Saisana and Tarantola 2002).

Cronbach's coefficient alpha (C-α): Another method mentioned in the literature is C-α, which estimates the degree of correlation among indicators and assesses the extent to which the indicators cover the same information (Nardo et al. 2008; Saisana and Tarantola 2002). This coefficient α ranges from zero to one, with a low α indicating multiple dimensions in the dataset, and a high α suggesting that the indicators measure the same information (Saisana and Tarantola 2002).

For the second dimension, the countries, cluster analysis is frequently used in the construction of composite indicators (Nardo et al. 2005).

Cluster analysis: This analysis provides an overview of similarities of the different countries between the indicators (Nardo et al. 2005). By identifying patterns and clusters among countries, cluster analysis informs the weighting and aggregation methods for subsequent steps (Nardo et al. 2008). Various clustering methods, including hierarchical and non-hierarchical approaches such as k-means clustering, are available (Nardo et al. 2005).

3.3. Normalization

Given that different indicators measure various aspects, they are often expressed in different units (Zhou and Zhang 2018; Nardo et al. 2008). To combine these indicators, it is essential to bring them to a common unit or scale, which is accomplished through normalization methods (Nardo et al. 2008; Zhou and Zhang 2018). According to Hausner et al. (2017) normalization serves the same purpose as PCA, which is to reduce the complexity of the data. Zhou and Zhang (2018) mention the three most common types:

Z-Score: This method normalizes values through z-standardization (Zhou and Zhang 2018). By subtracting the mean and dividing by the standard deviation, values are converted to a common standardized scale (Zhou and Zhang 2018). This is the most widely used method, as it assumes a normal distribution with a mean of zero (Freudenberg 2003).

Distance to reference: This method normalizes values by comparing them to a reference point, which could be a historical value or another benchmark (Zhou and Zhang 2018; Nardo et al. 2008). In the context of panel data, according to Zhou and Zhang (2018) and Cherchye et al. (2007) suggest that using the first year as a reference point enables consistent comparisons over time.

Re-scaling: This method re-scales values to a dimensionless scale ranging from zero to one, based on the global maximum and minimum (Zhou and Zhang 2018). The formula involves subtracting the minimum value from the actual value and dividing it by the range (the difference between the maximum and minimum values) (Zhou and Zhang 2018).

3.4. Imputing of missing data

For constructing reliable composite indicators, reliable data is essential (Saisana and Tarantola 2002; Nardo et al. 2008). Missing data presents a significant challenge in this context (Nardo et al. 2008; Freudenberg 2003). There are three primary methods for addressing missing data: the deletion of the entire observation, single imputation, and the multiple imputation (Nardo et al. 2008). All three methods may introduce some degree of uncertainty into the dataset, which must be accounted for in the subsequent analysis (Nardo et al. 2008; Nardo et al. 2005).

Deletion of the case: Deleting cases with missing values is typically done only when the amount of missing data is relatively small (Pigott 2001). Deleting cases, as well as imputing values using mean or median, can alter the structure and variance of the data (Nardo et al. 2005).

Single imputation: As noted by Nardo et al. (2008), single imputation can be performed using techniques such as regression or mean substitution. When dealing with data from different time periods, time series analysis can be employed to impute missing data, providing precise estimates based on the existing data (Saisana and Tarantola 2002).

Multiple Imputation: In preserving the original structure and variance of the dataset, multiple imputation is the most accurate method, as single imputation often underestimates variance (Nardo et al. 2008). This approach can be implemented using techniques such as Markov chains (Nardo et al. 2005; Nardo et al. 2008). Several complete datasets are estimated and generated, and a combination of these is used to impute the missing values (Pigott 2001; Nardo et al. 2005).

Saisana and Tarantola (2002) and Nardo et al. (2008) suggest performing this step after the selection of indicators. However, other studies, such as Nardo et al. (2005) recommend conducting this step following the examination of the data structure and the normalization of indicators.

3.5. Weighting

When combining sub-indicators into a composite indicator, a key question arises: how much each sub-indicator should contribute to the overall indicator. This contribution is determined by assigning weights to each sub-indicator, and there are several methods for doing so (Nardo et al. 2005).

Equal weighting: The simplest and most common method is equal weighting, where each sub-indicator is assigned the same weight (Nardo et al. 2008). However, Nardo et al. (2008) cautioned that this technique risks double counting when collinear indicators are included in the composite indicator, as both indicators may influence the composite while representing the same dimension of information. Despite this, Nardo et al. (2005) argued that equal weighting can be advantageous when sub-indicators are in the same units and are scaled. Additionally, equal weighting of correlated sub-indicators allows for different aspects of a dimension to be represented in the composite indicator (Nardo et al. 2005).

PCA/FA: Multivariate statistical methods such as PCA or FA not only examine the relationships between sub-indicators but can also be used to define their weights (Saisana and Tarantola 2002; Hausner et al. 2017). According to Nardo et al. (2008), a correlation between sub-indicators is necessary for using these methods for weight estimation.

The literature also mentions several other weighting methods, such as expert opinion, benefit of the doubt, and multiple regression models (Nardo et al. 2005; Saisana and Tarantola 2002). Numerous options are available, and it is crucial to select the most appropriate method for each specific construction process (Nardo et al. 2008).

3.6. Aggregation

The process of selecting an appropriate function to combine the chosen sub-indicators and their respective weights into a composite indicator is known as aggregation (Zhou and Zhang 2018). Several methods are discussed in the literature:

Linear aggregation: Also referred to as simple additive weighting, this method involves creating a linear combination of the sub-indicators and their respective weights (Zhou and Zhang 2018). The sum of these linear combinations constitutes the composite indicator (Zhou and Zhang 2018). When all sub-indicators are measured in the same unit and standardized, this method is both straightforward and effective (Nardo et al. 2005).

Geometric Aggregation: Also known as the weighted product, this method involves multiplying the sub-indicators, with their weights raised to the power of the respective indicators (Zhou and Zhang 2018). Geometric aggregation is only used when sub-indicators are strictly positive and on different scales (Nardo et al. 2005). While not commonly employed in constructing composite indicators, it presents significant opportunities according to Zhou and Zhang (2018).

Multi-criteria method: When numerous diverse aspects are combined to create a composite indicator, such as in environmental indices, neither linear nor geometric aggregation may effectively capture all dimensions (Nardo et al. 2005). In such cases, the multi-criteria approach is more suitable, ensuring that each aspect is adequately represented in the composite indicator (Nardo et al. 2005).

3.7. Uncertainty and Sensitivity Analysis

When combining multiple sub-indicators into a single composite indicator, assessing the robustness of this composite is crucial (Saisana and Tarantola 2002; Saisana et al. 2005). Several studies have explored uncertainty and sensitivity analysis for composite indicators (Saisana et al. 2005; Nardo et al. 2008).

Uncertainty Analysis: Uncertainty analysis evaluates how variations in inputs affect the output (Saisana et al. 2005). For composite indicators, inputs include the methods used in their construction, while outputs refer to the composite indicator values, country rankings, and/or rank differences (Saisana et al. 2005; Nardo et al. 2005). According to Saisana et al. (2005), sources of uncertainty may arise from the selection of sub-indicators, data selection, data editing, normalization, weighting, and aggregation.

Sensitivity Analysis: Sensitivity analysis aims to assess the contribution of each input to the overall variance of the output (Saisana et al. 2005; Nardo et al. 2005). This involves altering inputs to observe the resulting changes in the output (Saisana et al. 2005). Numerous variance-based techniques are discussed in the literature, including the Monte Carlo approach and the Sobol method (Nardo et al. 2005). Nardo et al. (2008) suggest one approach is to conduct a sensitivity analysis to identify which sources of uncertainty have the greatest influence on for example determining the relative ranking of two entities.

The proposed roadmap for constructing umbrella indicators using statistical methods will be tested in the next chapter to assess its suitability for this analysis and to answer the second research question: whether these methods can be applied to the CAP framework to identify umbrella indicators.

4. Statistical Application

4.1. CAP result indicator dataset

The European Commission's data portal for the Common Agricultural Policy offers several datasets related to indicator data (European Commission - Directorate-General for Agriculture and Rural Development 2024a). In this master's thesis, two available datasets - one covering indicator data from 2010 to 2023 and another from the current CAP period of 2023 to 2029 - were examined to determine which is most suitable for applying the methods for constructing composite indicators found in the literature review.

For the 2023-2029 dataset, because the current CAP period only began in 2023, real empirical data is available only for 2023, while the data for the subsequent years up to 2029 consists of target values. To ensure the use of actual empirical data, the dataset containing annual empirical data of CAP indicators from 2010 to 2023 has been selected for this study.

The dataset from 2010 to 2023 includes data on various monitoring indicators, including impact, result, output, and context indicators (European Commission. Directorate General for Agriculture and Rural Development 2015). In this work, the primary focus is on result indicators, which measure the immediate impacts of specific CAP actions (European Commission. Directorate General for Agriculture and Rural Development 2015). The result indicators were extracted from the data portal into an Excel file, and the statistical analysis was conducted using the statistical software RStudio version 4.2.1 (Posit team 2024). Packages used within this software were dplyr, tidyverse, tidyr, corr, ggcorrplot, FactoMiner, factoextra, mice, zoo, ggplot2, reshape2, nlme, cluster, psych, dendextend and ggdendro.

Since the CAP framework divides actions into Pillar 1 and Pillar 2, the result indicators are categorized accordingly. The dataset includes data corresponding to both pillars for the period from 2010 to 2023, with this study focusing specifically on the result indicators of Pillar 2.

According to the European Commission's monitoring framework, there are 25 result indicators for Pillar 2 (European Commission. Directorate General for Agriculture and Rural Development 2015). The labels of these result indicators follow the structure R.X_PII, where X represents the specific number of the indicator as described in the monitoring framework (European Commission. Directorate General for Agriculture and Rural Development 2015). Detailed definitions of each result indicator are provided in the appendix.

The dataset comprises panel data for each result indicator from the 27 member states of the European Union (EU), covering the years 2010 to 2023. Additionally, it includes data for the EU as a whole for each indicator. The dataset consists of 23 columns in total, including a column for the indicator, sub-indicators (if applicable), a code identifying each indicator and its sub-indicators, the data source, the measurement unit, the member states, and columns for each year from 2010 to 2023. Upon examining the dataset, it becomes evident that the indicators were measured over different time periods. For the result indicators in Pillar 2, data is available for the years 2015 to 2022, with complete data provided for all indicators for each year.

However, the dataset obtained from the EU Data Explorer was incomplete for several indicators. The missing indicators include R.02_PII, R.13_PII, R.14_PII, R.15_PII, R.18_PII, and R.19_PII. According to the European Commission. Directorate General for Agriculture and Rural Development (2015), R.02_PII measures the change in agricultural output on supported farms per annual work unit. R.13_PII reflects the increase in water use efficiency in agriculture within Rural Development Programmes (RDP) supported projects, while R.14_PII measures the increase in energy use efficiency in agriculture and food processing within these projects (European Commission. Directorate General for Agriculture and Rural Development 2015). R.15_PII captures the renewable energy produced from supported projects, R.18_PII measures the reduction in methane and nitrous oxide emissions, and R.19_PII measures the reduction in ammonia emissions (European Commission. Directorate General for Agriculture and Rural Development 2015). Consequently, the dataset contains data for 19 result indicators.

Except for R.21_PII and R.24_PII, which measure jobs created in supported projects and jobs created in supported LEADER projects respectively, all indicators in the dataset are measured as percentages (European Commission. Directorate General for Agriculture and Rural Development 2015; European Commission NaN).

For data manipulation, the original dataset was filtered to include only member states, years, and the relevant indicators. The data pivot was adjusted so that the year became a column, with separate columns for each indicator. The member state column was retained, while the value for the European Union as a whole was removed since it represents the sum of the individual member states' values. Thus, the dataset now contains 21 columns, including member states, years, and indicators, with 216 observations per variable, corresponding to values for each of the 27 member states across the eight years.

The first step in the statistical application involved conducting descriptive statistics on the dataset and its variables. Table 2 provides an overview of all result indicators. The minimum value for all indicators is either zero or very close to zero, with 0.003 for indicator R.03_PII. The maximum values, however, vary significantly. For thirteen of the indicators, the maximum values are below 100 (table 2). The other six indicators have maximum values exceeding 100. For indicators R.21_PII and R.24_PII, this is logical as their units are numerical counts.

However, for the other indicators, which are measured as percentages, values exceeding 100 are inconsistent with their intended definition. A percentage value above 100, however, could be interpreted as an increase beyond the initial value.

In analyzing the dataset, the prevalence of zero values was notable (see Appendix). Indicators R.01_PII, R.07_PII, R.08_PII, R.19_PII, and R.22_PII each had between 10 and 20 zeros among the 216 observations. Indicators R.06_PII, R.09_PII, R.11_PII, R.17_PII, R.20_PII, and R.21_PII had between 80 and 156 zeros, with R.09_PII having the highest count at 156. The remaining indicators with zero counts could be identified as those with missing values.

Table 2 shows that six of the 19 indicators have more than 20,000 % missing values. In this context, zeros were considered as actual values, while missing values were represented as NAs, indicating the absence of data. Notably, indicators R.12_PII and R.25_PII had more than half of their values missing. Additionally, table 2 reveals that the means for each indicator are relatively low compared to their maximum values. This discrepancy is further highlighted by the fact that the standard deviations are generally higher than the means. This divergence can be attributed to the high number of zero values and the significant amount of missing data in the dataset.

Table 2: Summary statistics of the 19 CAP result indicators

Indicator	Statistics across 27 member states and 2015-2022				
	Min	Max	Mean	Std	% of NAs
R.01_PII	0.000	53.000	4.738	8.677	0.000
R.03_PII	0.003	11.825	2.199	2.356	22.222
R.04_PII	0.000	58.458	2.092	6.969	12.040
R.05_PII	0.000	125.029	6.523	22.604	39.350
R.06_PII	0.000	12.141	0.763	1.533	0.000
R.07_PII	0.000	98.437	21.933	24.066	0.000
R.08_PII	0.000	88.473	18.113	23.568	0.000
R.09_PII	0.000	2.229	0.160	0.377	0.000
R.10_PII	0.000	91.471	18.351	22.616	0.000
R.11_PII	0.000	2.306	0.201	0.446	0.000
R.12_PII	0.000	6.607	0.829	1.406	56.021
R.16_PII	0.000	22.125	2.318	4.586	32.410
R.17_PII	0.000	50.555	3.081	8.976	0.000
R.20_PII	0.000	14.155	1.065	2.720	0.000
R.21_PII	0.000	7070.000	514.769	1233.846	0.000
R.22_PII	0.000	105.750	60.206	28.778	0.000
R.23_PII	0.000	120.743	20.076	27.207	27.781
R.24_PII	0.000	14640.000	812.546	1854.448	0.000
R.25_PII	0.000	32.165	2.919	6.740	56.480

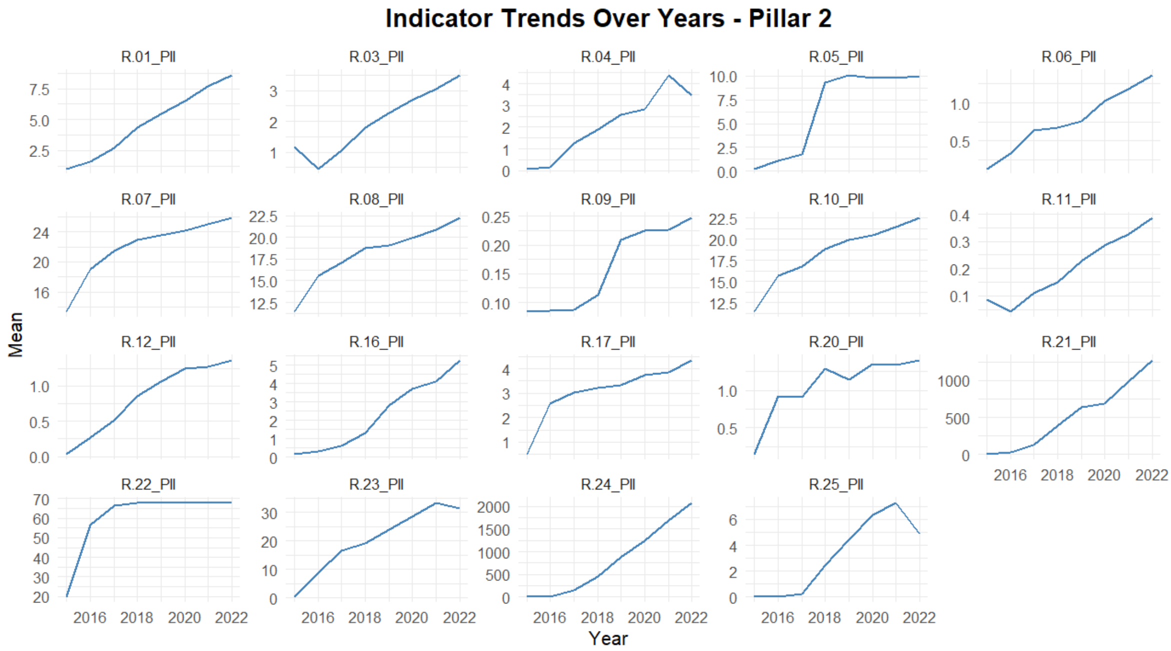


Figure 2: Trend visualization of the CAP result indicators between 2015-2022

Figure 2 illustrates the trends of the indicators over time and across member states. Overall, each indicator demonstrates a positive trend, reflecting an upward trajectory in their respective measurements. This trend is consistent across the years, indicating a general improvement or increase in the values of these CAP indicators.

Examining the data by year, 2015 had the highest number of missing values, with 81 (see Appendix), followed by 2016 with 73 missing values. The subsequent years generally had fewer missing values, averaging around 60 per year. The year with the lowest number of missing values was 2022, with 58. The distribution of zeros across the years is similar (see Appendix).

Examining the data across member states, Bulgaria and Cyprus show the highest total numbers of zeros, with 70 and 66, respectively (see Appendix). In contrast, Italy and Germany have the lowest totals, with 16 and 17 zeros, respectively. Regarding missing values (Not Available, NAs), Italy, Spain, and nine other member states have either no NAs or very few, while the remaining states have up to 41 missing values, with Czechia having the highest count (see Appendix).

4.2. Building a CAP umbrella indicator

4.2.1. Selection

Following the roadmap for constructing a composite indicator in table 1, the first step involves identifying and selecting relevant indicators that accurately describe a specific phenomenon (Nardo et al. 2008). Unlike other contexts where composite indicators might be created from scratch, the CAP result indicators to select from are already provided in the dataset. Here, the objective is to identify indicators that represent and describe similar aspects of the CAP, enabling their combination into a unified indicator for further facilitated analysis as it was suggested by Galeotti et al. (2020).

Unlike other approaches to constructing composite indicators that might involve subjective criteria, this method relies solely on statistical analysis to identify relevant indicators (Saisana and Tarantola 2002). In this case, the selection of indicators is based on the data structure rather than subjective judgment.

The analysis following the steps of the pre-analysis plan in chapter 3 was initially complicated by the presence of missing values. Consequently, the roadmap for constructing the composite indicator, as outlined in the literature review (table 1), required adjustments to accommodate the statistical application. The first step in this adjusted process was the imputation of missing values.

4.2.2. Imputing missing values

Since the interest was in all member states and years and the amount of missing data for some indicators too high, the deletion of all cases with missing values, as proposed by Nardo et al. (2008) was not feasible. While mean imputation was considered and applied as recommended by several sources, the complexity and extent of missing data necessitated a more sophisticated approach (Nardo et al. 2008; Becker 2022).

Therefore, the decision was made to use multivariate imputation by chained equations (MICE) (van Buuren and Oudshoorn 2000). This method is endorsed by the R package COINr, developed by the European Commission, as a suitable approach for **multiple imputation** (Becker 2022; Becker et al. 2022; van Buuren and Oudshoorn 2000; Kleinke et al. 2011). MICE is an iterative procedure that imputes missing values by using an imputation function for each indicator column with missing data, employing other indicators as inputs for the model, thus preserving the data structure (van Buuren and Oudshoorn 2000). The algorithm generates multiple datasets with imputed values through predictive mean matching (PMM), and these solutions are then combined to produce a final imputed value (van Buuren and Oudshoorn 2000; Nardo et al. 2005; Pigott 2001).

As shown in the summary statistics in Table 2, imputation was performed only for the indicators R.03_PII, R.04_PII, R.05_PII, R.12_PII, R.16_PII, R.23_PII, and R.25_PII.

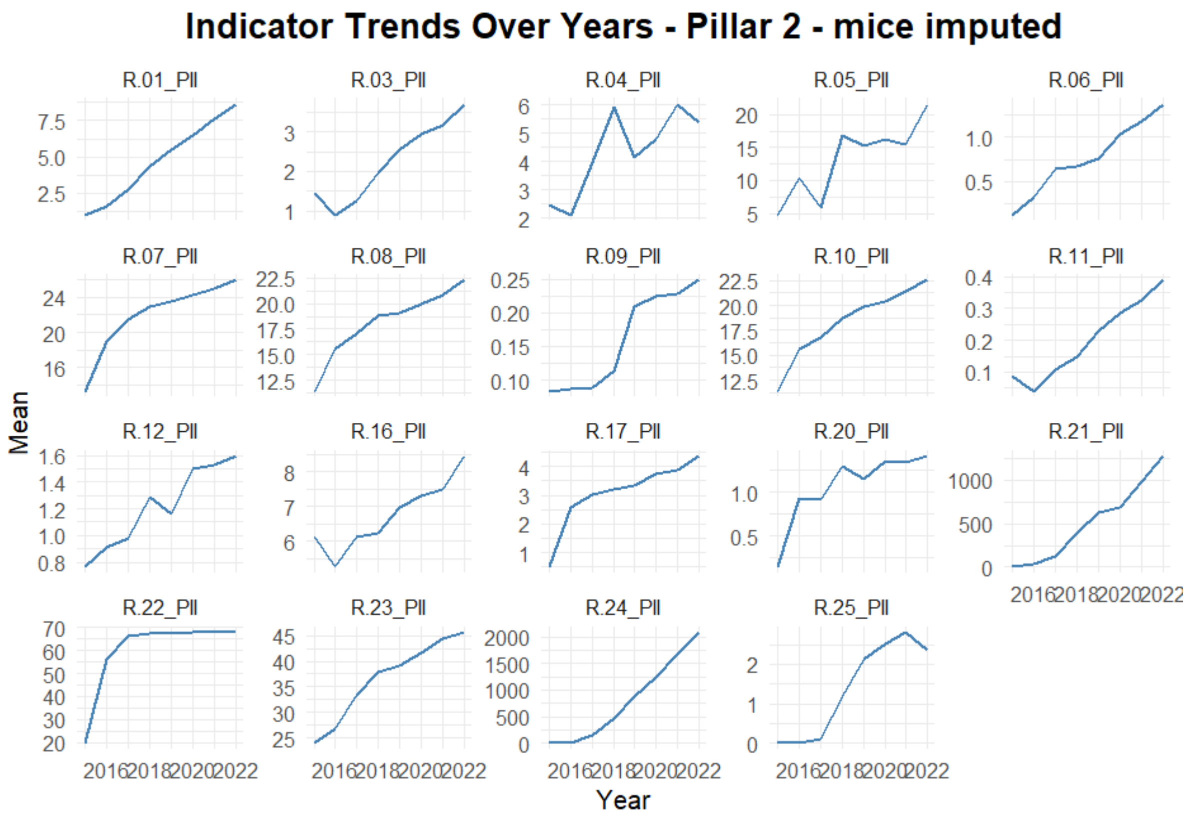


Figure 3: Trend visualization of the CAP result indicators between 2015-2022 after imputing missing values via multiple imputation by chained equations (MICE, van Buuren and Oudshoorn 2000)

Figure 3 illustrates the trends of these sub-indicators over the years following imputation by MICE. The graphs for the imputed sub-indicators generally exhibit a similar positive trend to that observed in the original dataset. However, the graphs for R.04_PII, R.05_PII, R.12_PII, and R.16_PII show slight deviations, and a less smooth trend compared to the original dataset. Despite these minor differences, the trends in the imputed data closely resemble those in the original dataset, and the summary statistics indicate only slight deviations (see Appendix). Therefore, the analysis proceeded with the dataset imputed using MICE.

The most sophisticated approach for imputing missing data would have been to utilize time series analysis, which likely would have provided a more accurate estimation by accounting for temporal patterns and trends in the data (Kleinke et al. 2011; Bashir and Wei 2018; Suo et al. 2019). However, due to time constraints, this method was not implemented.

4.2.3. Normalization

According to the original roadmap outlined in table 1, the next step following the identification of relevant indicators would be the examination of the data structure. However, an examination of the dataset reveals that the value ranges of the indicators differ (see table 2). The imputation of missing values using MICE and predictive mean matching did not alter these value ranges (see Appendix). Additionally, indicators R.21_PII and R.24_PII are measured in total numbers rather than percentages, unlike the other indicators (European Commission. Directorate General for Agriculture and Rural Development 2015). To ensure comparability among the indicators and to prevent high values of specific indicators from disproportionately influencing subsequent analyses, normalization is required (Freudenberg 2003). Consequently, instead of first performing the identification as suggested by Nardo et al. (2008), the original roadmap (table 1) was adjusted to include normalization immediately following the imputation of missing values.

To address the varying scales and to facilitate accurate comparison among all indicators**, z-score standardization** was applied. This method, recommended by Zhou and Zhang (2018) and others, is widely used for normalization (Freudenberg 2003). Z-score standardization transforms each indicator to a common scale with a mean of zero and a standard deviation of one by subtracting the mean value and dividing by the standard deviation (Nardo et al. 2008; Freudenberg 2003). This method is particularly suitable in cases with extreme values, as indicated by the large scale differences in Table 2. Compared to min-max normalization, z-score standardization mitigates the impact of extreme values on the composite indicator (Nardo et al. 2005). Although ranking, as proposed by Nardo et al. (2008) would have been a simpler option and would also address outliers, it would result in a loss of information about variance between countries (Freudenberg 2003). Since variance and covariance are crucial for constructing the composite indicator in the subsequent steps, z-score standardization was deemed the more appropriate method (Nardo et al. 2008).

Since the CAP result indicator data comprises panel data across member states and years, and given the interest in year-to-year differences, the indicator values were normalized using z-score standardization for each year separately. This approach facilitates year-wise comparison of the indicator values in subsequent analysis steps.

A challenge encountered was that for indicators such as R.21_PII and R.24_PII, all values for the year 2015 were zero. In the year-wise normalization for these indicators, the mean, the deviation of values from the mean, and especially the standard deviation were all zero. This resulted in NaN (Not a Number) values due to the mathematical issue of dividing by zero. To address this, the columns with constant values for these specific years were treated as constant and retained as zero in the dataset.

At the end of this step, the data was completed, with each indicator normalized separately for each year.

4.2.4. Identification

In this step, unlike in the typical creation of composite indicators, the selection of sub-indicators from the dataset was based on the findings from the identification process. Essentially, the selection and identification steps were combined here.

The initial approach involved conducting a **correlation analysis** of the indicators and performing **PCA** for all years to gain a comprehensive understanding of the dataset. This analysis was then repeated separately for each year to observe how the data structure evolved over time. The underlying assumption, as suggested by Saisana et al. (2005), was that highly correlated indicators are likely to convey similar information, making them suitable candidates for grouping together into a composite indicator.

Figure 4 presents the correlation heatmap of the CAP result indicators. In this heatmap, a perfect positive correlation of one is represented by dark red, zero correlation by white, and a perfect negative correlation of minus one by dark blue. The heatmap reveals that most indicators tend to be positively correlated, with fewer instances of negative correlations (see Figure 4).

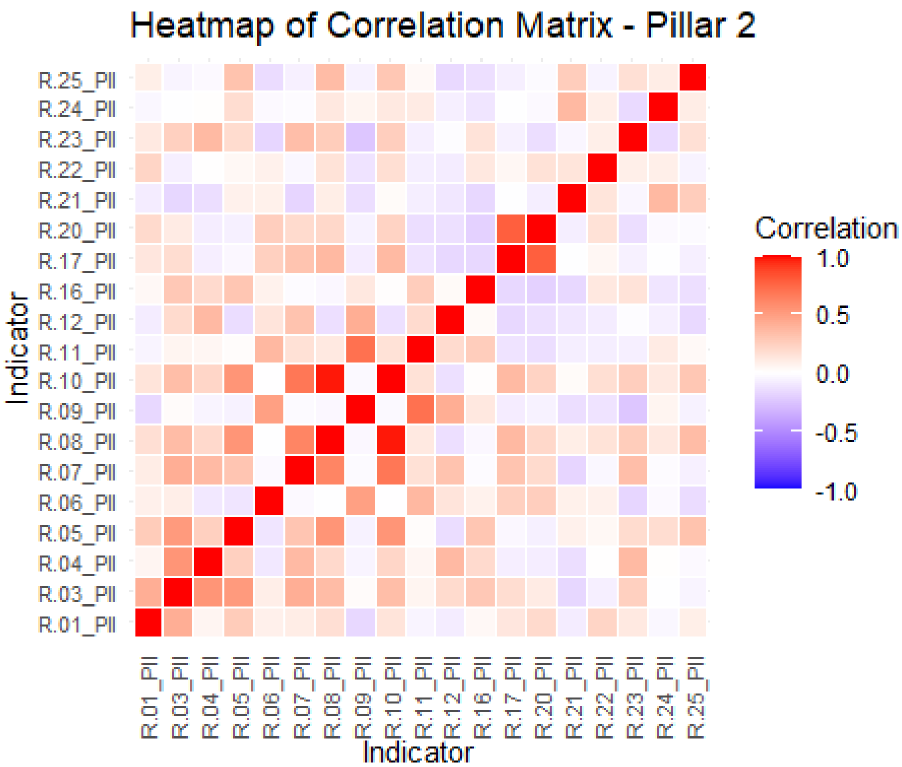


Figure 4: Correlation Heatmap of the whole CAP result indicator dataset with strong positive correlation indicated by red and strong negative correlation indicated by blue squares

The diagonal elements in a correlation matrix represent the correlation of each indicator with itself, which is always equal to 1. Indicators highly correlated with a correlation larger than 0.9 were indicators R.08_PII and R.10_PII.

The next method used to further analyze the data structure was PCA, as recommended by Saisana and Tarantola (2002) and Nardo et al. (2008). The goal of PCA was to build upon the insights gained from the correlation analysis, allowing for a deeper understanding of the relationships between indicators and the extent to which they describe similar aspects of the CAP framework. However, the efficacy of using PCA had to be assessed beforehand through correlation analysis. If the indicators were not correlated, PCA would not provide additional valuable information (Saisana and Tarantola 2002). Therefore, ensuring that there were meaningful correlations between the indicators was a prerequisite for successfully applying PCA.

The PCA results, as shown in Figure 5, confirm the multidimensional nature of the dataset, which is consistent with its status as panel data and the variety of distinct aspects captured by the indicators.

The scree plot in Figure 5 graphically displays the eigenvalues, representing the proportion of variance explained by each principal component (Keita 2018). It reveals that the first seven principal components account for only 74.5 % of the variance, while ten dimensions are needed to explain 87.2 % of the total variance. This demonstrates the multidimensionality of the dataset, as ten dimensions are required to explain more than 80 % of the variance (Saisana et al. 2005).

The biplot in figure 5 visually illustrates the direction and strength of each indicator's contribution to the first two principal components (Keita 2018). Each vector corresponds to an indicator, with its direction reflecting whether it has a positive or negative impact on the first two principal components (Keita 2018). The colour of the vectors, as shown in the legend, signifies the strength of their contributions to these components (Keita 2018). Close vectors suggest that the indicators they represent are highly correlated and likely cover similar information. For example, indicators R.10_PII and R.08_PII are positioned very closely, with similar direction and colour, indicating that they capture nearly the same underlying aspects.

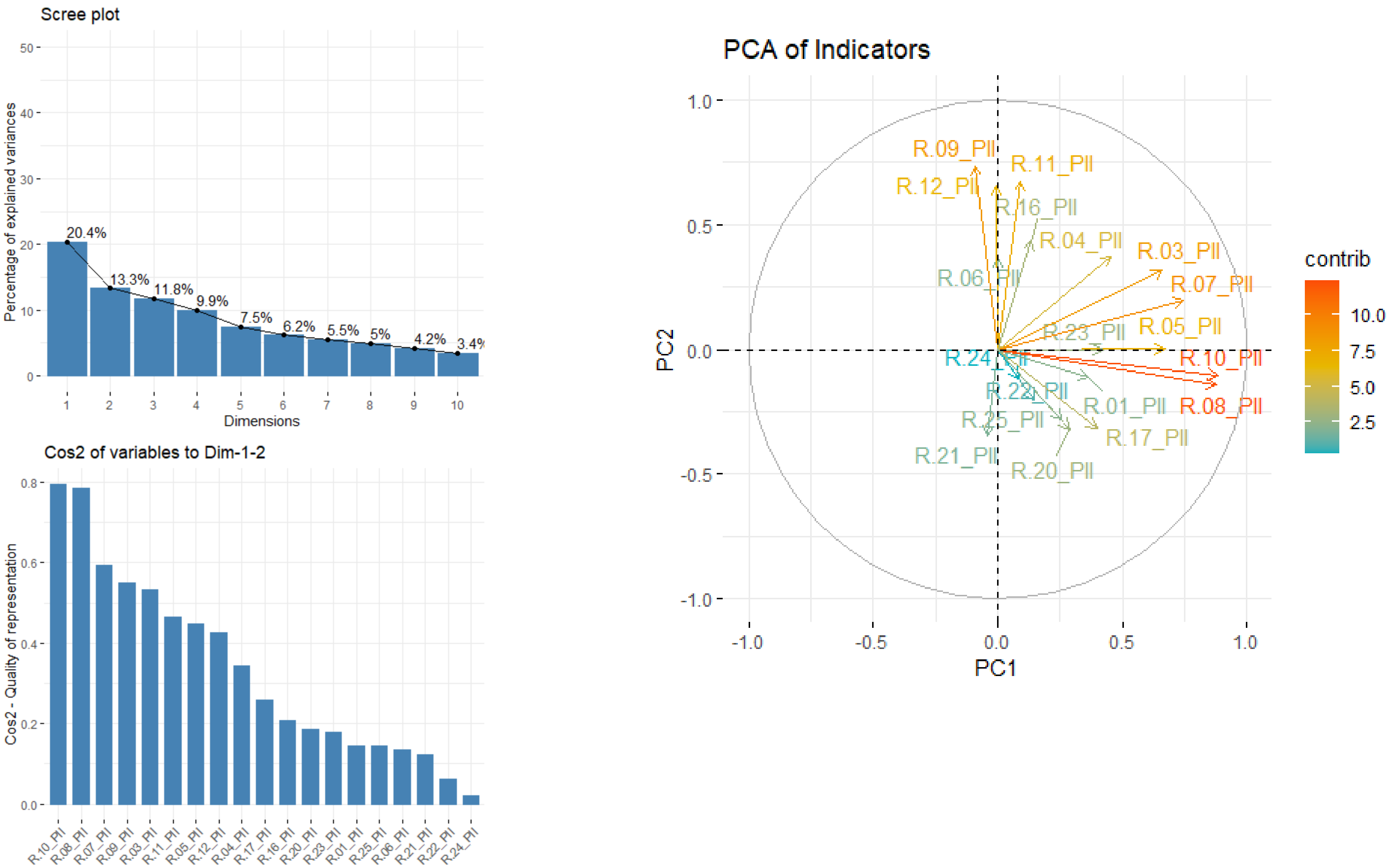


Figure 5: PCA results of the whole CAP result indicator dataset including Scree Plot, Cos2 plot and Biplot (PCA of indicators)

The Cos2 plot in figure 5 shows the quality of the contribution of the result indicators to the first principal component (Keita 2018). Cos2, or squared cosine, measures how well each original variable is represented or projected onto the principal component (Keita 2018). High Cos2 values indicate a strong contribution of the variable to the principal component, meaning the variable aligns well with that component (Keita 2018). From this plot, it is evident that indicators R.10_PII and R.08_PII exhibit the highest quality of contribution, signifying their strong alignment and influence on the first principal component.

Given the high multidimensionality of the data revealed by the PCA on the entire dataset, further correlation analysis and PCA were conducted separately for each year. This approach was taken to examine how the data structure might change over time and to assess whether the dimensionality of the data could be reduced.

The year-wise correlation analysis provided valuable insights into the relationships between indicators across different years. The correlation heatmaps for each year, found in the Appendix, demonstrate that the correlation structure of the indicators varies over time, as indicated by the changes in color on the heatmaps. Generally, the indicators tend to be more positively correlated with one another rather than negatively correlated. This positive correlation aligns with the trend analysis shown in figures 2 and 3, where all indicators exhibit a positive trend over time. This consistent upward trend across indicators likely contributes to their positive correlations, as they are all moving in the same direction. For instance, the correlation heatmap for 2021, shown in Figure 6, exemplifies this trend.

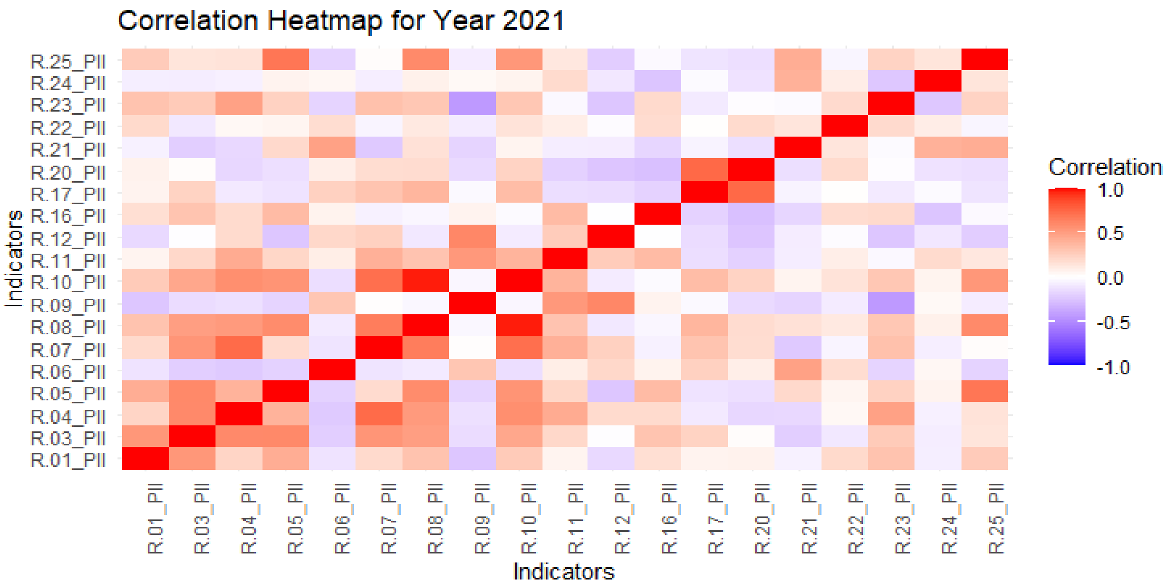


Figure 6: Correlation heatmap for the CAP result indicators for the year 2021 (the correlation heatmaps for the rest of the years between 2015 and 2022 can be found in the Appendix)

However, it is important to note that for the year 2015, no correlation could be estimated for indicators R.21_PII and R.24_PII, due to the presence of only zero values for both indicators in that year.

While the pairs of highly correlated indicators change from year to year, a consistent pattern was observed: indicators R.08_PII and R.10_PII are highly correlated in every year analyzed (as also highlighted in Figures 4 and 6). This consistent correlation suggests a strong relationship between these two indicators across all observed periods.

The next step involved conducting PCA for each year individually to identify similar dimensions within the data. This year-by-year analysis sought to reveal shifts in the data structure over time and to detect consistent patterns or variations (see Appendix). The results showed that the scree plots for each year remained similar to the overall scree plot for all years (as shown in Figure 5). This indicates that the dataset's multidimensionality persists across the years.

The biplots for each year demonstrated that the indicators contributing to the first two principal components varied between years (see Appendix). However, indicators R.08_PII and R.10_PII consistently remained highly similar across all years, as seen in the overall plot in Figure 5. The biplot for 2015, presented in Figure 6, shows that this year had the highest number of indicators contributing significantly to the first two principal components. Not only did R.08_PII and R.10_PII demonstrate strong contributions to the first dimension, but also indicators R.09_PII, R.06_PII, and R.11_PII exhibited strong impacts, with similar vector directions indicating common information captured by these indicators.

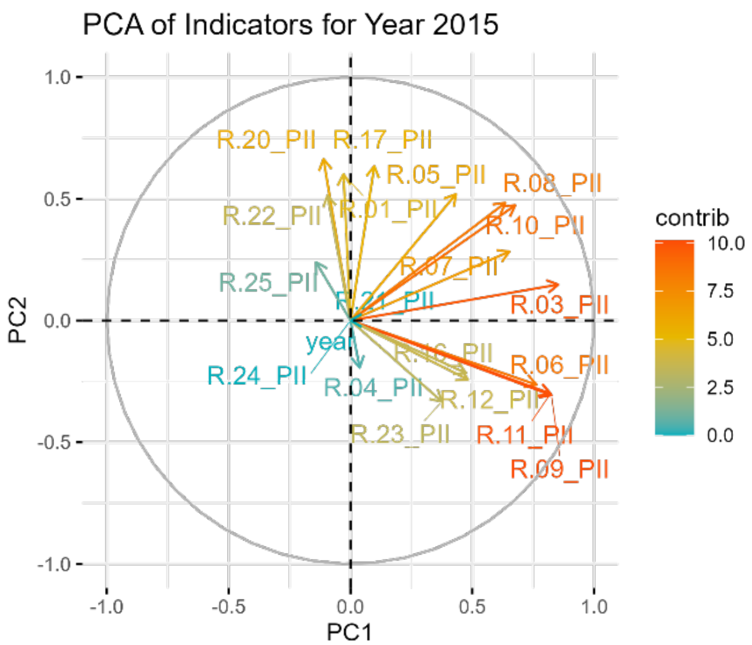


Figure 7: Biplot of CAP result indicator dataset for year 2015 with vectors representing each indicator and the colour as indicated by the legend on the right the indicators contribution to the respective principal components one on the x-axis and two on the y-axis

Looking at the Cos2 plots for each year, it becomes clear that only in 2015 do R.10_PII and R.08_PII not show the highest quality of contribution to the first principal component (see Appendix). This anomaly highlights that, across most years, these two indicators consistently capture the same information, reinforcing their strong relationship throughout the analysis.

Given the presence of different dimensions and varying indicator contributions across years, a hierarchical cluster analysis was performed for each year to group highly correlated indicators. Indicators with correlations above 0.9 were clustered together. These clusters varied across different years, indicating the dynamic relationships within the dataset. Some indicators were strongly correlated in one year but not in another, reflecting evolving interactions over time within the CAP result indicators.

Highly correlated indicator clusters are summarized in table 3. A correlation threshold of 0.9 was chosen to define high correlation as done by Becker (2022). This threshold was selected to avoid an excessive number of indicators and to prevent overly crowded groupings, which would have been the case with a lower threshold, such as 0.7. Additionally, using a higher threshold of 0.9 helps ensure that the indicators within each group cover similar aspects, thereby minimizing information loss and reducing the uncertainty and variance associated with combining indicators.

Table 3: Highly correlated indicator groups per year in CAP result indicator dataset

Year	Indicator group	Correlation
2015	R.06_PII_R.09_PII_R.11_PII	0.91,0.91,1
2015	R.08_PII_R.10_PII	0.98
2016	R.08_PII_R.10_PII	0.98
2017	R.08_PII_R.10_PII	0.98
2017	R.09_PII_R.11_PII	0.97
2017	R.17_PII_R.20_PII	0.95
2018	R.08_PII_R.10_PII	0.99
2019	R.08_PII_R.10_PII	0.98
2019	R.17_PII_R.20_PII	0.9
2020	R.08_PII_R.10_PII	0.98
2021	R.08_PII_R.10_PII	0.97
2022	R.08_PII_R.10_PII	0.97

As shown in table 3, indicators R.08_PII and R.10_PII, which were identified as highly correlated across all years in the PCA biplots, consistently demonstrate strong correlation. In 2015, a notable group comprising more than two indicators was identified, including R.06_PII, R.09_PII, and R.11_PII. This finding aligns with the graphical analysis of the biplot for 2015, depicted in Figure 6, which also shows these indicators exhibiting strong correlations and contributing significantly to the principal components for that year.

These indicator groups will serve as the sub-indicators for the condensed indicators in the subsequent steps of the analysis.

PCA is indeed based on correlation (Saisana and Tarantola 2002). However, it’s important to note that a high correlation between indicators doesn’t necessarily imply that they capture the same information or aspect of the phenomenon being measured (Saisana and Tarantola 2002). Therefore, it is crucial to not rely solely on statistical correlation when deciding to combine indicators into a composite measure (Saisana and Tarantola 2002). A careful examination of the descriptions and definitions of these indicators is necessary to ensure that they actually cover similar aspects by definition, rather than just being statistically correlated. This step is essential to validate the meaningfulness of combining these indicators, ensuring that the resulting composite indicator is conceptually sound and accurately represents the intended aspects of the CAP framework.

For example, the consistently high correlation between indicators R.08_PII and R.10_PII is supported by their descriptions taken form the monitoring framework (European Commission. Directorate General for Agriculture and Rural Development 2015). R.08_PII measures the percentage of agricultural land under management contracts to improve water management, while R.10_PII measures the percentage of agricultural land under management contracts to improve soil management and/or prevent soil erosion (European Commission. Directo2rate General for Agriculture and Rural Development 2015). Since both aspects, water management and soil management, are closely related, it is logical that these indicators show high correlation (Council et al. 1993).

Similarly, indicators R.06_PII, R.09_PII, and R.11_PII, which all pertain to forestry land, are highly correlated in 2015 (European Commission. Directorate General for Agriculture and Rural Development 2015). R.06_PII measures the percentage of forest or other wooded areas under management contracts supporting biodiversity, R.09_PII covers the percentage of forestry land under management contracts to improve water management, and R.11_PII measures the percentage of forestry land under management contracts to improve soil management and/or prevent soil erosion (European Commission. Directorate General for Agriculture and Rural Development 2015). This correlation is expected as all these indicators relate to the management and conservation of forestry land.

Indicators R.17_PII and R.20_PII also reflect related aspects: R.17_PII measures the percentage of agricultural land under management contracts targeting the reduction of GHG and/or ammonia emissions, while R.20_PII measures the percentage of agricultural and forest land under management contracts contributing to carbon sequestration or conservation (European Commission. Directorate General for Agriculture and Rural Development 2015). Given their similar focus on emissions and carbon management, a high correlation would be expected.

It remains unclear why some indicator pairs, such as R.17_PII and R.20_PII, do not exhibit consistent high correlations across all years. Similarly, the high correlation between R.09_PII and R.11_PII in 2017, but not with R.06_PII as seen in 2015, is also not fully understood.

4.2.5. Weighting

For each of these clusters, PCA was performed to determine their loadings for the first principal component (Nardo et al. 2008; Saisana and Tarantola 2002). In other contexts of composite indicator creation, these loadings could be used as weights (Saisana and Tarantola 2002). As Freudenberg (2003) states, weights should be assigned to each sub-indicator in a manner that aligns with the underlying framework. In this case, the goal was to combine indicators that represent similar aspects into a single composite indicator. Since the selected indicators are assumed to cover the same or similar information, each indicator should contribute equally to the final composite indicator. Consequently, the loadings were normalized to sum to 1, ensuring **equal weighting** for each sub-indicator.

Using the loadings as weights would result in the composite indicator reflecting more information than the individual contributions of each sub-indicator, thereby making the composite indicator's values larger than the simple aggregation of the individual indicators.

Table 4: Highly correlated indicator groups per year in CAP result indicator dataset with their loadings on first principal component (PC1) and their normalized loadings

Year	Sub-indicator group	Loadings PC1	Normalized Loadings	Correlation
2015	R.06_PII_R.09_PII_R.11_PII	0.96,0.99,0.99	0.33,0.34,0.34	0.91,0.91,1
2015	R.08_PII_R.10_PII	0.99,0.99	0.5,0.5	0.98
2016	R.08_PII_R.10_PII	1,1	0.5,0.5	0.98
2017	R.08_PII_R.10_PII	1,1	0.5,0.5	0.98
2017	R.09_PII_R.11_PII	0.99,0.99	0.5,0.5	0.97
2017	R.17_PII_R.20_PII	0.99,0.99	0.5,0.5	0.95
2018	R.08_PII_R.10_PII	1,1	0.5,0.5	0.99
2019	R.08_PII_R.10_PII	1,1	0.5,0.5	0.98
2019	R.17_PII_R.20_PII	0.97,0.97	0.5,0.5	0.9
2020	R.08_PII_R.10_PII	0.99,0.99	0.5,0.5	0.98
2021	R.08_PII_R.10_PII	0.99,0.99	0.5,0.5	0.97
2022	R.08_PII_R.10_PII	0.99,0.99	0.5,0.5	0.97

4.2.6. Aggregation

To construct the composite indicator values, each indicator's value was multiplied by its normalized loading, and then these weighted values were summed to produce the composite indicator. This approach employs **linear aggregation**. Given that the indicators in each group cover similar information, the composite indicator represents a balanced combination of the respective indicators.

Since different groups of indicators exhibit high correlations in different years, separate composite indicators were created for each year based on these correlations. Specifically, a composite indicator was developed for each group of indicators in the years they showed high correlation, as detailed in tables 3 and 4. For the indicators R.08_PII and R.10_PII, a composite indicator was created for each year, reflecting their consistently high correlation across years, as shown in tables 3 and 4.

Since the composite indicator for R.08_PII and R.10_PII was created for all years, it is of particular interest. Figure 8 presents a comparison between the composite indicator and its sub-indicators.

As shown in figure 8, the composite indicator values (blue line) consistently fall between the lines representing its sub-indicators, R.08_PII (red line) and R.10_PII (green line). This suggests that the construction of the composite indicator has been successful, as it reflects a balanced integration of the two sub-indicators. The blue line of the composite indicator consistently lies between the two sub-indicator lines (red and green), providing strong evidence of the effective application of statistical methods and the successful construction of the composite indicator.

Figure 8: Comparison of composite indicator composite_R.08_PII_R.10_PII (blue line) and its sub-indicators R.08_PII (red line) and R.10_PII (blue line) over the years for each member state

4.2.7. Uncertainty and Sensitivity Analysis

When constructing composite indicators by combining multiple sub-indicators, it is essential to evaluate their robustness (Saisana and Tarantola 2002; Saisana et al. 2005). In this study, the uncertainty analysis focuses on the composite indicators involving sub-indicators R.08_PII and R.10_PII, as this is the most coherent composite indicator across years. Several studies have outlined methods for conducting uncertainty and sensitivity analyses for composite indicators (Saisana et al. 2005; Nardo et al. 2008; Nardo et al. 2005). The follwing is based on the procedures mentioned in there. These provide the foundation for this analysis.

Uncertainty can arise from multiple sources in the process of constructing composite indicators Saisana et al. (2005). However, given the dataset and analysis performed here, two key sources of uncertainty, dataset provision and indicator selection, are not relevant. Instead, the focus is on data imputation, normalization, and aggregation. These steps are crucial in this application, as they directly affect the robustness of the composite indicator. The weighting step is not considered a source of uncertainty, as it is designed to align with the framework's assumptions. Aggregation, however, remains a potential source of uncertainty (Saisana et al. 2005).

The methods considered for addressing imputation, normalization, and aggregation are summarized in Table 5. These methods represent the primary sources of uncertainty and were adjusted for this particular application, based on the procedures outlined by Saisana et al. (2005).

Table 5: Sources of uncertainty adjusted for this statistical application (based on Saisana et al. (2005))

Number	Possible sources of uncertainty		
	Imputation	Normalization	Aggregation
1	MICE	z-score	linear
2	Mean	min-max	geometric
3	Median	raw value	-

Each method is assigned a number, and each combination of methods represents a different composite indicator construction. For example, the method combination "MICE, z-score, linear" is denoted as 1-1-1.

For all 18 possible method combinations, the process of identification, equal weighting of sub-indicators, and aggregation was repeated. The reference composite indicator is the one constructed using the combination "MICE, z-score, linear" (1-1-1).

For each method combination in table 5, the composite indicator values were computed for each year, and the member states were ranked accordingly. This ranking process follows the uncertainty analysis approach for composite indicators described by Saisana et al. (2005).

To assess the robustness of the indicator method combinations, the average rank differences of composite indicators for R.08_PII and R.10_PII were compared across all method combinations relative to the reference indicator (method combination 1-1-1). This analysis includes the mean, along with the corresponding 5th and 95th percentiles for each member state, as shown in Figure 9. For most countries, the average rank shift hovers around zero. However, in some countries, such as Spain, Romania, Italy, and Finland, the mean average rank shift is slightly above one, indicating that the ranks tend to be higher in the method combinations compared to the reference indicator. Conversely, in countries like Belgium and Bulgaria, the ranks tend to shift to lower levels. The variation in average rank shifts differs significantly between countries. For instance, in Finland, ranks vary dramatically, spanning from very low to very high across the different method combinations. In contrast, in countries like Cyprus, the rank shifts are relatively minor.

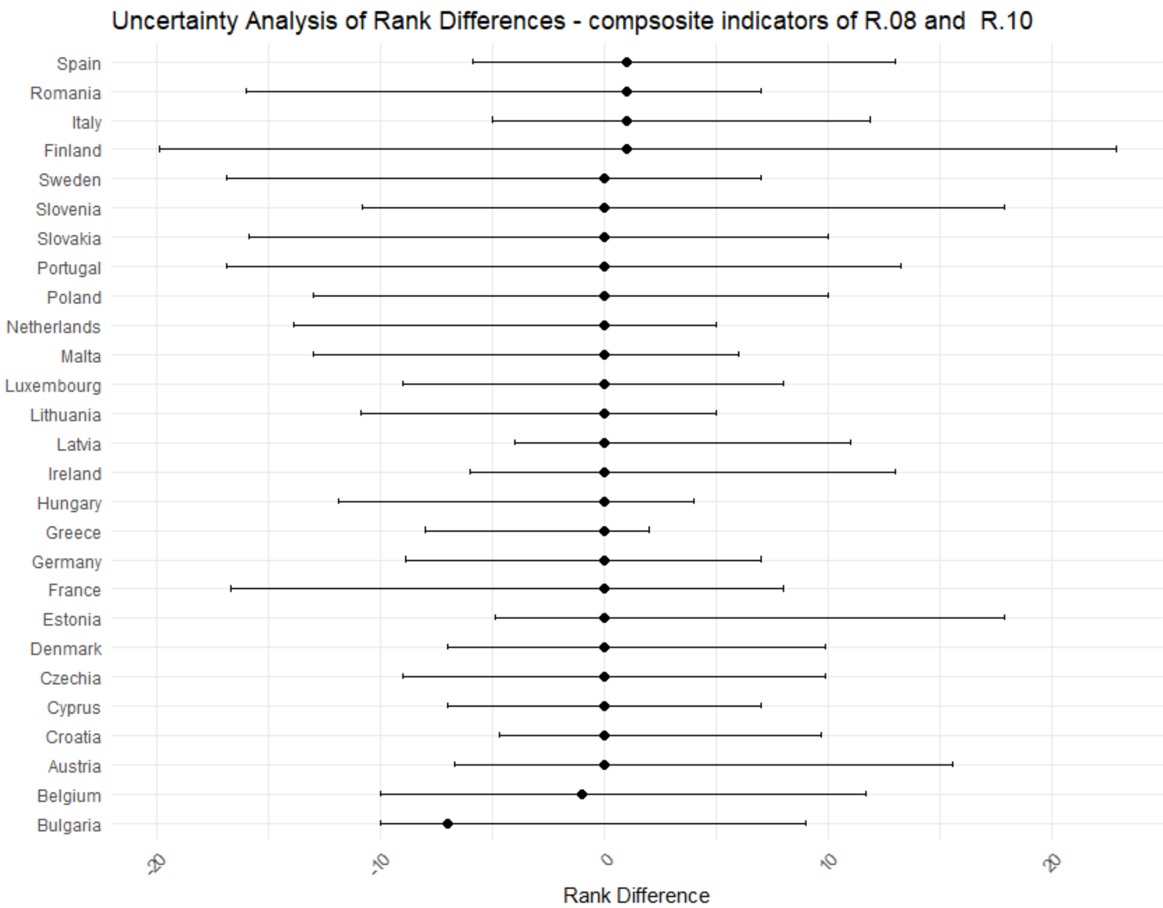


Figure 9: Uncertainty indicated by average rank differences of composite indicators of R.08_PII and R.10_PII constructed with all method combinations in comparison of the reference indicator with method combination 1-1-1 including the mean, and the corresponding 5th and 95th percentiles for each member state

In the graphs depicting the average rank shifts of composite indicators for sub-indicators R.08_PII and R.10_PII, presented in the appendix, the average rank shifts of the sub-indicators relative to the 1-1-1 method combination are shown for each member state. These graphs reveal that the shifting patterns of the sub-indicators compared to 1-1-1 closely mirror the shifting patterns of their corresponding composite indicator. This suggests a strong relationship between the rank shifts of the sub-indicators and the overall structure of their composite indicator.

Figure 10 illustrates the distribution of rank differences between the sub-indicators and the method combinations compared to the reference indicator, presented as box plots. The reference indicator, when compared to the sub-indicators, shows whiskers extending to a maximum rank difference of 27, indicating that member states display substantial rank variation across the reference indicator (1-1-1).

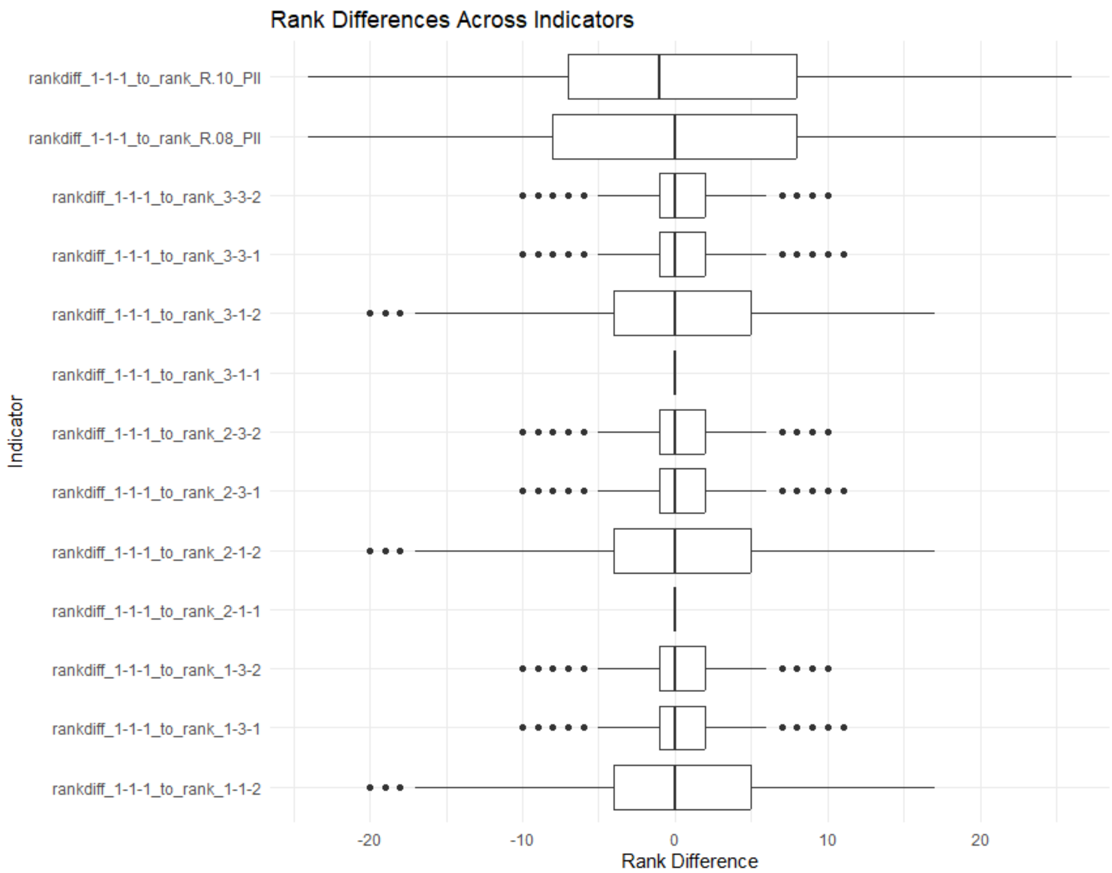


Figure 10: Uncertainty indicated by boxplots of differences between composite indicators of the sub-indicators R.08_PII and R.10_PII constructed with all method combinations and the sub-indicators in comparison of the reference indicator with method combination 1-1-1

However, the interquartile range (the box representing the 25th to 75th percentiles) is relatively narrow and centered around zero for both sub-indicators, with most rank differences falling within approximately ±5 or 10. This suggests that the reference indicator effectively incorporates the sub-indicators and their respective rankings.

In general, the mean rank shifts across all indicators are centered around zero. Method combinations such as 3-3-2, 3-3-1, 2-3-2, 2-3-1, 1-3-2, and 1-3-1 exhibit narrower boxes, indicating more consistent rank shifts when compared to the reference indicator. This suggests that these method combinations are more closely aligned with the reference indicator. Conversely, method combinations like 3-1-2, 2-1-2, and 1-1-2, which use z-score normalization and geometric aggregation, exhibit wider boxes and whiskers ranging from -17 to +17. This indicates larger differences from the reference indicator. However, their interquartile ranges are relatively small, typically between -4 and 5, suggesting that while the overall variation is greater, most of the rank differences still remain within this narrower range. Furthermore, the method combinations 2-1-1 and 3-1-1, which use mean and median normalization alongside z-score and linear aggregation, show no box or whiskers, only the mean of zero rank shift. This indicates that these combinations are identical to the reference indicators. This similarity between different imputation methods arises because the uncertainty analysis only considers the composite indicators constructed from sub-indicators R.08_PII and R.10_PII. Since these sub-indicators do not have missing values (see Table 2), no imputation was necessary, and thus the method combinations do not differ based on imputation.

From figure 10, we can conclude that method combinations involving z-score normalization and geometric aggregation exhibit the greatest variation in rank differences compared to the reference indicator (1-1-1). This is expected, as geometric aggregation involves multiplying the weights by the power of each sub-indicator's value. Method combinations that use raw value normalization (x-3-x) show only slight variations in rank differences relative to the reference indicator, as evidenced by the narrow boxes, typically ranging from -2 to +3. These indicators appear to be very similar to the reference indicator.

For the sensitivity analysis, the approach suggested by Nardo et al. (2008) was employed to identify which sources of uncertainty exert the greatest influence on determining the relative rankings of entities. To illustrate this, two member states, Bulgaria and Estonia, were selected. The composite indicator values and rank differences for each method combination were compared to the reference indicator over the years, allowing us to assess the impact of the various sources of uncertainty on the relative positioning of these two countries. (see figures 11 and 12).

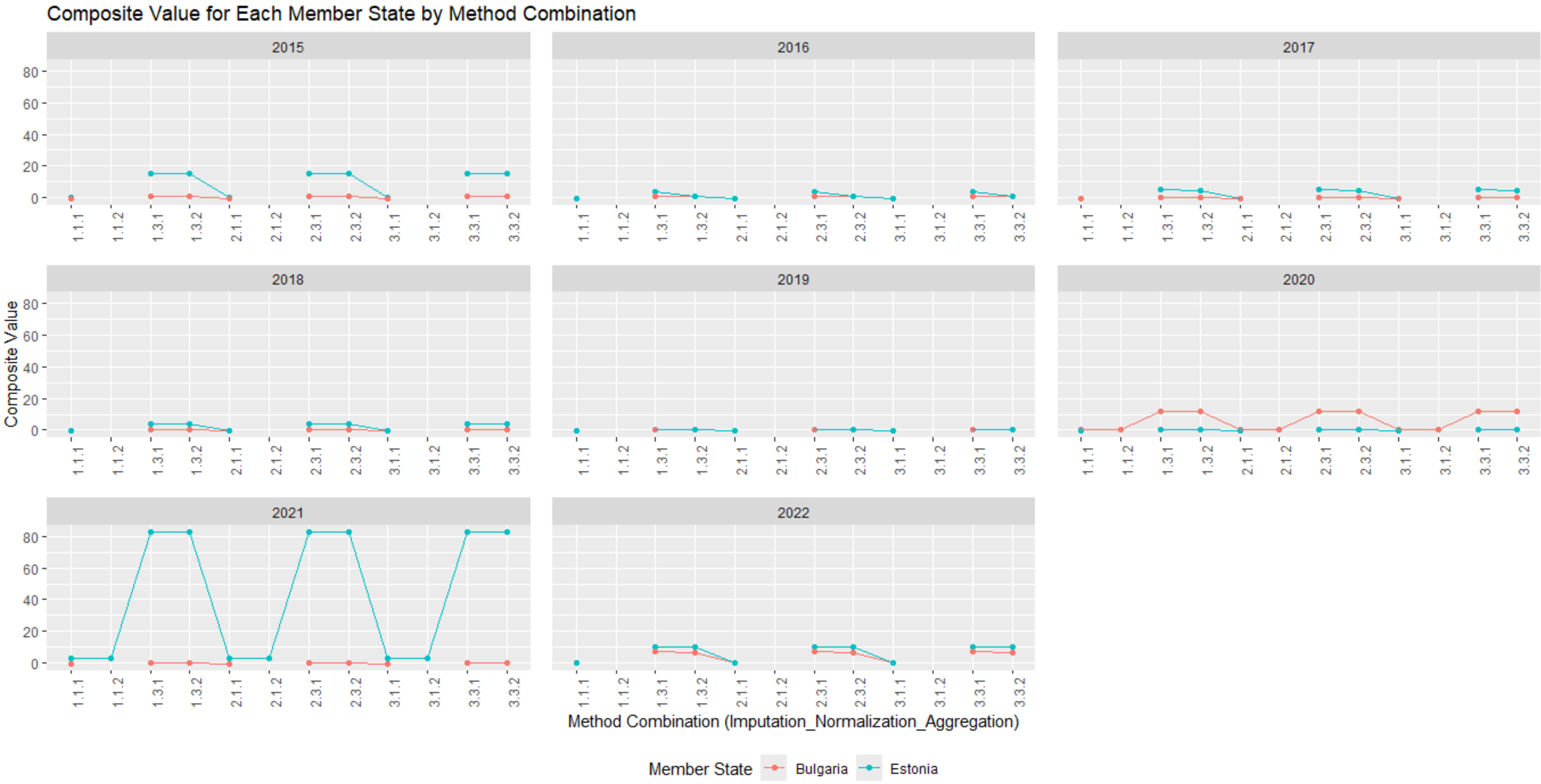


Figure 11: Sensitivity Analysis of composite indicators of sub-indicators R.08_PII and R.10_PII with the help of composite indicator values for each method combination compared to refence indicator for Bulgaria (red line) and Estonia (blue line) over the years 2015-2022

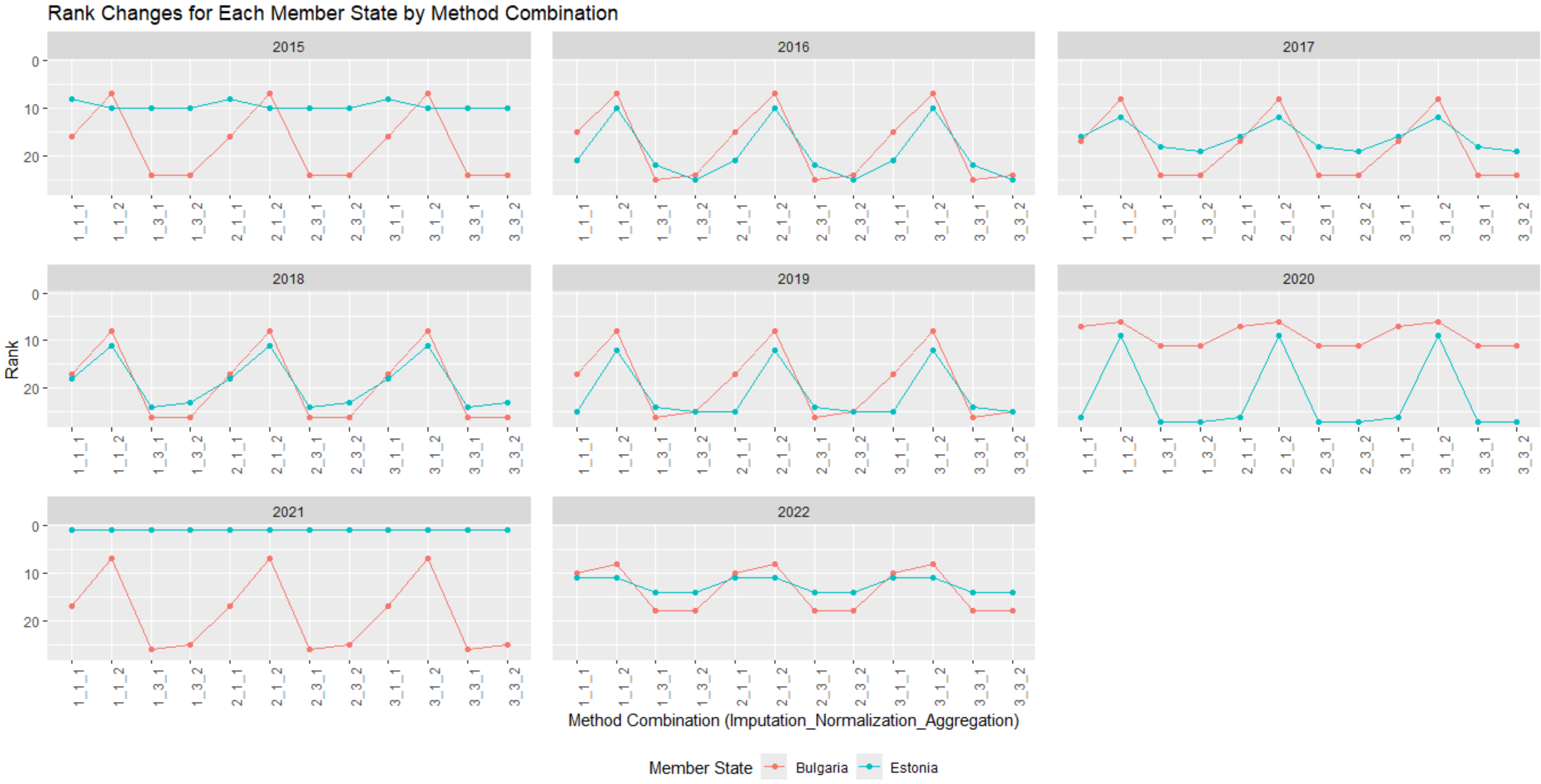


Figure 12: Sensitivity Analysis of composite indicators of sub-indicators R.08_PII and R.10_PII with the help of rank differences for each method combination compared to refence indicator for Bulgaria (red line) and Estonia (blue line) over the years 2015-2022

As shown in figure 11, the composite indicator values are highest in most years for method combinations that involve raw values (x-3-x). This trend corresponds with the rank differences visualized in figure 12, which also demonstrate larger deviations from the reference indicator for these same method combinations. This phenomenon occurs because raw values tend to be higher compared to the z-score standardized values used in the reference indicator, leading to larger differences between these method combinations and the reference indicator.

This trend exhibits some variation over the years. In certain years, the variation in composite values is more pronounced than in others, with the year 2021 standing out due to particularly significant fluctuations. In almost all years, Estonia’s values are consistently higher than those of Bulgaria. An exception occurs in 2020, where Bulgaria's values surpass Estonia's. Nevertheless, the method combinations using raw values continue to exhibit the highest values during this period as well.

Examining the rank changes of Bulgaria and Estonia across the various method combinations, as shown in figure 12, reveals some notable patterns. Estonia's ranks remain relatively stable, particularly in the years 2015 and 2022. In contrast, Bulgaria shows significant rank variability across most years, except for 2020, where the rank differences are less pronounced. A striking observation is that geometric aggregation consistently introduces substantial rank differences, as compared to the linear aggregation method (see Appendix). This observation aligns with the results from the uncertainty analysis, which highlighted that the aggregation method, particularly geometric aggregation, plays a critical role in influencing the variability of results. Similarly, normalization methods, such as using raw values, also contribute to increased uncertainty (see Appendix).

In conclusion, the greatest sources of uncertainty in the construction of the composite indicator stem from the aggregation and normalization methods. Specifically, using raw values combined with geometric aggregation introduces the highest levels of uncertainty, as evidenced by the rank fluctuations in the sensitivity analysis.

However, it is important to note that this analysis was limited to the composite indicators constructed from the sub-indicators R.08_PII and R.10_PII, which had no missing values. As a result, there was no variation introduced by the imputation methods, potentially leading to some bias in the findings. It would be valuable to extend this analysis in future work to other sub-indicators where missing values have been imputed, as this could provide further insights into the impact of imputation on the robustness of the composite indicators. This would help in understanding how different methods of handling missing data influence the reliability and stability of the composite measures.

4.3. Adjusted roadmap for constructing umbrella indicators

In response to the challenges encountered during the statistical application of methods to the CAP result indicator dataset, the initial roadmap outlined in table 1 has been adjusted. The revised roadmap, presented in table 6 below, reflects these modifications, which were made to address the specific statistical requirements of the dataset.

Table 6: Adjusted roadmap for constructing a composite indicator as response to challenges encountered during the statistical application (based on Nardo et al. (2005), Nardo et al. (2008), Freudenberg (2003), Saisana and Tarantola (2002) and Zhou and Zhang (2018))

Step		Explanation/goal	References for method
1.	Imputing missing data	Filling in missing data	(Nardo et al. 2008)
2.	Normalization	Ensuring comparability of indicators	(Zhou and Zhang 2018; Freudenberg 2003)
3.	Identification / 4. Selection	What structure has my data? What is the relation between my sub-indicators? Which indicators are relevant/ Which indicators cover similar aspects?	(Saisana 2004; Saisana and Tarantola 2002; Nardo et al. 2008; Botta and Koźluk 2014)
5.	Weighting	Examining the weight of each sub-indicator for the composite indicator	(Saisana and Tarantola 2002; Zhou and Zhang 2018; Nardo et al. 2005)
6.	Aggregation	Aggregating several sub-indicators according to their weights to one composite indicator	(Nardo et al. 2005; Saisana and Tarantola 2002; Nardo et al. 2008; Zhou and Zhang 2018; Munda 2012)
7.	Uncertainty/ Sensitivity analysis	How much bias or uncertainty is established through creation of one composite	(Nardo et al. 2008; Saisana et al. 2005)

Evaluating the pre-analysis roadmap in table 1 and its methods confirms their overall suitability. However, the structure required rearrangement, as shown in the revised roadmap, which addresses the first research question concerning the evaluation of the methods.

5. Discussion

As mentioned in the introduction, composite indicators have been constructed in various scientific fields for diverse reasons and purposes, gaining increasing importance over recent decades for applications such as comparing country performances in competitiveness or assessing sustainability, often serving as a basis for political actions (Freudenberg 2003; Rogge 2012; Nardo et al. 2008; Fusco 2015).

Composite indicators, as noted by Nardo et al. (2005), Nardo et al. (2008) and Saisana and Tarantola (2002) offer the advantage of summarizing complex and multifaceted issues, making them easily understandable and interpretable for both decision-makers and the public. By reducing large volumes of information to a specific, condensed form, composite indicators enhance comprehension, streamline econometric analysis, and provide decision-makers and the public with a clear overview (Nardo et al. 2005; Saisana and Tarantola 2002). Furthermore, by maintaining this condensed format, additional information can be incorporated without overwhelming the audience (Nardo et al. 2008). Additionally, graphical representations of composite indicators can effectively illustrate changes over time or the performance of specific countries on key issues (Saisana and Tarantola 2002; Nardo et al. 2005; Nardo et al. 2008).

This master's thesis investigates methods for constructing umbrella indicators and presents a literature-based roadmap for their application, addressing the first research question. It demonstrates that these methods can be successfully applied not only in other scientific fields but also to CAP indicators, which relates to the second research question. This has been validated through the statistical application of these methods to the result indicators of Pillar 2 of the Common Agricultural Policy (CAP). Specifically, Figure 8 illustrates how composite indicator methods effectively combine information from multiple sub-indicators into a single, condensed indicator.

However, the aggregation of sub-indicators into composite or umbrella indicators presents risks and challenges, which have been extensively discussed in the literature (Nardo et al. 2005; Nardo et al. 2008; Saisana and Tarantola 2002). When used as a basis for policy recommendations, composite indicators can be misleading if the construction process is not rigorously executed (Nardo et al. 2008; Saisana and Tarantola 2002; Nardo et al. 2005). In this context, transparency is crucial (Nardo et al. 2008).

The aim of this work was not to construct composite indicators for policy advice, but rather to test whether the methods found in the literature could be applied to the CAP dataset. Even in this context, as outlined in the uncertainty and sensitivity analysis chapter, combining sub-indicators with the methods used to construct composite indicators introduces a range of uncertainties and potential biases, as demonstrated in this statistical application (Saisana et al. 2005; Nardo et al. 2005).

As Saisana et al. (2005) point out, bias is often introduced from the very first step - the selection of indicators. This bias typically arises from the use of subjective criteria as suggested by Saisana and Tarantola (2002) and and value judgments, as noted by Garnåsjordet et al. (2012). However, unlike many other approaches to constructing composite indicators, this statistical application relies solely on statistical analysis to identify relevant indicators. In this case, the selection is guided by the data structure rather than subjective judgment, making the process more objective and less prone to bias (Saisana and Tarantola 2002). Consequently, the need for a fitting framework for the sub-indicators, as discussed by Saisana and Tarantola (2002) and Nardo et al. (2005), is still relevant in terms of ensuring that the indicators genuinely measure similar aspects. While the statistical analysis may identify correlations, the descriptions of the indicators must be carefully examined to confirm that they indeed capture similar things. This validation step is crucial to ensure that the combined indicators are conceptually aligned and not just statistically correlated, as done above (Saisana and Tarantola 2002).

In this approach, the selection of relevant indicators was based on correlation analysis, PCA, and hierarchical clustering. Indicators with correlations above 0.9 were identified as highly correlated, as indicated by Becker (2022), providing an initial basis for relevance. However, one could argue that alternative thresholds, such as 0.8 or 0.7, might also have been appropriate for identifying related indicators. Hierarchical clustering was then applied, and the dendrogram was cut at a height of 0.1 to form clusters of highly correlated indicators. A composite indicator was constructed from these clusters, though, as with the correlation threshold, different cutting heights might have been suitable, potentially resulting in different groupings and outcomes. However, it is necessary to verify the assumptions underlying the statistical procedures before applying them, which was not addressed in this instance and could be improved (Nardo et al. 2005).

Missing data is a significant challenge in many statistical applications, particularly in the construction of composite indicators (Pigott 2001; Freudenberg 2003; Nardo et al. 2008). The choice of method for imputing missing values can introduce bias and uncertainty into the dataset, potentially affecting the robustness of the composite indicators (Nardo et al. 2008; Nardo et al. 2005). In this study, the Multiple Imputation by Chained Equations (MICE) method was used to impute missing data for the entire dataset, with the goal of preserving its underlying structure (van Buuren and Oudshoorn 2000). An alternative approach that could further maintain the structure of the dataset would be to apply the MICE method separately for each year, allowing for year-specific patterns and structures to be better preserved.

Furthermore, as Saisana and Tarantola (2002) suggest, time series analsis could be employed for a more accurate estimation of missing value. Suo et al. (2019), Bashir and Wei (2018) and Kleinke et al. (2011) present intriguing approaches that could be considered for this purpose.

The absence of entire indicators in the dataset, specifically for R.02_PII, R.13_PII, R.14_PII, R.15_PII, R.18_PII, and R.19_PII, could introduce bias into the statistical analysis. As mentioned by Saisana and Tarantola (2002), the quality of the data forms the foundation for the construction of composite indicators. In this statistical application, missing indicators might have been part of relevant clusters, and additional composite indicators could have been constructed if data for these indicators were available. Therefore, the missing indicators potentially limit the depth and accuracy of this analysis, underscoring the importance of complete and high-quality data (Saisana and Tarantola 2002; Freudenberg 2003).

In addition to addressing missing values and indicators, it is also important to determine how to handle zeros – as missing values or actual zeros (Martín-Fernández 2003). In this case, empty data cells were treated as missing values, while zeros were recorded as zeros.

The large number of zeros in the dataset created challenges for the application of statistical procedures. As a result, during the identification step, constant or zero-valued columns were excluded to avoid errors in the statistical processes. However, this exclusion may have introduced bias into the analysis, ultimately leading to the inclusion of only 12 out of 19 possible method combination composite indicators in the uncertainty analysis. Addressing this issue would enhance the precision of the analysis, and could potentially lead to the identification of different indicators for constructing composite indicators.

In the normalization step, for some indicators and years, the values were treated as constants to avoid the generation of NaN due to the issue of division by zero. For example, in this application, zero values were assigned to indicators R.21_PII and R.24_PII in the year 2015 to address this problem. However, this method might introduce potential bias, as other constant values in a column could also be reduced to zero by the R code, which might distort the analysis. Adjustments should be made to account for these cases, ensuring that constant values are handled more carefully to preserve data integrity.

Regarding weighting and aggregation methods, there are numerous approaches available (Zhou and Zhang 2018; Saisana and Tarantola 2002; Nardo et al. 2005; Nardo et al. 2008). Given the objective of constructing an umbrella indicator to simplify the dataset by reducing the number of indicators, as discussed by Saisana and Tarantola (2002), equal weighting is deemed appropriate (Nardo et al. 2005). To enhance this approach, incorporating expert opinions, as suggested in the literature, could provide a more refined weighting methodology (Saisana and Tarantola 2002). In this manner, the linear aggregation method was selected.

However, numerous studies outline various aggregation methods tailored to specific situations (Munda 2012; Zhou and Zhang 2018; Nardo et al. 2008).

For the purposes of simplicity and illustration, the uncertainty and sensitivity analysis was constrained to three methods: imputation, normalization, and aggregation. There may be opportunities to explore additional method combinations. Additionally, due to the high frequency of zeros, only 12 out of 19 possible method combinations were applied to the sub-indicators R.08_PII and R.10_PII. Consequently, not all method combinations could be compared with the reference indicator, which could have provided valuable insights. Furthermore, since these two sub-indicators initially had no missing values, no imputation was performed, making the comparison of imputation methods redundant. Expanding the uncertainty and sensitivity analysis to include additional composite indicators could yield further insights (Saisana et al. 2005).

Overall, a more comprehensive analysis would provide additional insights and a deeper understanding of the robustness of the statistical application discussed. Based on the results above, the initial approach appears to have been successful, demonstrating that the condensation of indicators covering similar information into umbrella indicators is feasible without significant loss of information (Saisana and Tarantola 2002; Galeotti et al. 2020). As long as transparency is maintained, composite indicators serve as an effective tool to facilitate further analyses while preserving the data structure and relevant information (Nardo et al. 2008).

In the context of composite indicators, significant research has been conducted by the European Commission's Joint Research Centre (Becker 2022; European Commission 2024a). This Centre even maintains a dedicated website for composite indicators (European Commission 2024b). One of their contributions is the development of the R package COINr (Becker et al. 2022). This was created by researchers, including authors such as Michaela Saisana, whose methods were utilized in this work. This tool consolidates all necessary steps and methods for composite indicator construction into a single package (Becker et al. 2022; Becker 2022). Becker (2022) provides a detailed explanation of these methods. Applying the statistical techniques discussed earlier using this package could offer valuable insights and reveal alternative approaches or findings.

6. Conclusion and outlook

Overall, methods from other scientific fields, as found in the literature, can indeed be applied to CAP indicators within agricultural sciences. However, the roadmap for creating composite indicators, initially developed based on the literature, must be adapted to suit the specific statistical requirements of the dataset at hand. As long as transparency is maintained, composite indicators can serve as an effective tool for facilitating further analysis while preserving the data structure and relevant information (Nardo et al. 2008). Further research incorporating more extensive statistical applications on CAP indicators, including those beyond result indicators, would provide deeper insights into more complex construction processes and enable more thorough robustness testing.

7. Publication bibliography

Bashir, Faraj; Wei, Hua-Liang (2018): Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. In *Neurocomputing* 276, pp. 23–30. DOI: 10.1016/j.neucom.2017.03.097.

Becker, William (2022): Composite Indicator Development and Analysis in R with COINr. Available online at https://bluefoxr.github.io/COINrDoc/, updated on 7/7/2022, checked on 8/15/2024.

Becker, William; Caperna, Giulio; Del Sorbo, Maria; Norlén, Hedvig; Papadimitriou, Eleni; Saisana, Michaela (2022): COINr: An R package for developing composite indicators. In *JOSS* 7 (78), p. 4567. DOI: 10.21105/joss.04567.

Botta, E.; Koźluk, T. (2014): Measuring environmental policy stringency in OECD countries: a composite index approach. OECD Economics Department Working Papers No. 1177: OECD Publishing. Available online at https://www.oecd-ilibrary.org/economics/measuring-environmental-policy-stringency-in-oecd-countries_5jxrjnc45gvg-en.

Cherchye, Laurens; Knox Lovell, C. A.; Moesen, Wim; van Puyenbroeck, Tom (2007): One market, one number? A composite indicator assessment of EU internal market dynamics. In *European Economic Review* 51 (3), pp. 749–779. DOI: 10.1016/j.euroecorev.2006.03.011.

Council, National Research; Agriculture, Board on; Policy, Committee on Long-Range Soil and Water Conservation (1993): Soil and water quality. An agenda for agriculture. Washington, D.C: National Academy Press.

European Commission (NaN): LEADER/CLLD - The European Network for Rural Development (ENRD) - European Commission. Available online at https://ec.europa.eu/enrd/leader-clld_en.html, updated on 3/19/2021, checked on 8/29/2024.

European Commission (2024a): About the Competence Centre on Composite Indicators and Scoreboards | Knowledge for policy. Available online at https://knowledge4policy.ec.europa.eu/composite-indicators/about_en, updated on 8/26/2024, checked on 8/26/2024.

European Commission (2024b): Composite Indicators. Available online at https://knowledge4policy.ec.europa.eu/composite-indicators_en, updated on 8/26/2024, checked on 8/26/2024.

European Commission - Directorate-General for Agriculture and Rural Development (2024a): Agri-food data portal | CAP Indicators. Available online at https://agridata.ec.europa.eu/extensions/DataPortal/cap_indicators.html, updated on 6/20/2024, checked on 7/4/2024.

European Commission - Directorate-General for Agriculture and Rural Development (2024b): CMEF. Available online at https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cmef_en, updated on 8/14/2024, checked on 8/14/2024.

European Commission. Directorate General for Agriculture and Rural Development (2015): The monitoring and evaluation framework for the common agricultural policy 2014–2020: Publications Office.

European Commission. Directorate General for Agriculture and Rural Development (2024c): CAP at a glance. Available online at https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-glance_en, updated on 6/17/2024, checked on 6/17/2024.

Förderung, Infodienst (NaN): LEADER. Available online at https://foerderung.landwirtschaft-bw.de/,Lde/Startseite/Foerderwegweiser/LEADER, updated on 8/27/2024, checked on 8/27/2024.

Freudenberg, Michael (2003): Composite Indicators of Country Performance A Critical Assessment. OECD Science, Technology and Industry Working Papers: OECD (2003/16). Available online at https://www.oecd-ilibrary.org/science-and-technology/composite-indicators-of-country-performance_405566708255.

Fusco, Elisa (2015): Enhancing non-compensatory composite indicators: A directional proposal. In *European Journal of Operational Research* 242 (2), pp. 620–630. DOI: 10.1016/j.ejor.2014.10.017.

Galeotti, Marzio; Salini, Silvia; Verdolini, Elena (2020): Measuring environmental policy stringency: Approaches, validity, and impact on environmental innovation and energy efficiency. In *Energy Policy* 136, p. 111052. DOI: 10.1016/j.enpol.2019.111052.

Garnåsjordet, Per Arild; Aslaksen, Iulie; Giampietro, Mario; Funtowicz, Silvio; Ericson, Torgeir (2012): Sustainable Development Indicators: From Statistics to Policy. In *Env Pol Gov* 22 (5), pp. 322–336. DOI: 10.1002/eet.1597.

Hausner, Vera H.; Engen, Sigrid; Bludd, Ellen K.; Yoccoz, Nigel G. (2017): Policy indicators for use in impact evaluations of protected area networks. In *Ecological Indicators* 75, pp. 192–202. DOI: 10.1016/j.ecolind.2016.12.026.

Herman, Kyle S.; Shenk, Justin (2021): Pattern Discovery for climate and environmental policy indicators. In *Environmental Science & Policy* 120, pp. 89–98. DOI: 10.1016/j.envsci.2021.02.003.

Keita, Zoumana (2018): Principal Component Analysis in R Tutorial. In *DataCamp*, 8/9/2018. Available online at https://www.datacamp.com/tutorial/pca-analysis-r?dc_referrer=https%3A%2F%2Fwww.google.com%2F, checked on 8/20/2024.

Kleinke, Kristian; Stemmler, Mark; Reinecke, Jost; Lösel, Friedrich (2011): Efficient ways to impute incomplete panel data. In *AStA Adv Stat Anal* 95 (4), pp. 351–373. DOI: 10.1007/s10182-011-0179-9.

Manly, Bryan F. J. (1994): Multivariate statistical methods. A primer. Third edition. Boca Raton, FL: Chapman and Hall/CRC, an imprint of Taylor and Francis (Chapman & Hall Book).

Martín-Fernández, J. A. (2003): Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. In *Mathematical Geology* 35 (3), pp. 253–278. DOI: 10.1023/A:1023866030544.

Munda, Giuseppe (2012): Choosing Aggregation Rules for Composite Indicators. In *Soc Indic Res* 109 (3), pp. 337–354. DOI: 10.1007/s11205-011-9911-9.

Nardo, Michela; Saisana, Michaela; Saltelli, Andrea; Tarantola, Stefano (2005): Tools for Composite Indicators Building. Available online at https://publications.jrc.ec.europa.eu/repository/handle/JRC31473.

Nardo, Michela; Saisana, Michaela; Saltelli, Andrea; Tarantola, Stefano; Hoffman, Anders; Giovannini, Enrico (2008): Handbook on constructing composite indicators: methodology and user guide. OECD Statistics Working Papers. Paris, France: OECD Publishing (2005/03). Available online at https://www.oecd-ilibrary.org/economics/handbook-on-constructing-composite-indicators_533411815016.

Pigott, Therese D. (2001): A Review of Methods for Missing Data. In *Educational Research and Evaluation* 7 (4), pp. 353–383. DOI: 10.1076/edre.7.4.353.8937.

Posit team (2024): RStudio. Integrated Development Environment for R. Boston, MA. Available online at http://www.posit.co/.

Rogge, Nicky (2012): Undesirable specialization in the construction of composite policy indicators: The Environmental Performance Index. In *Ecological Indicators* 23, pp. 143–154. DOI: 10.1016/j.ecolind.2012.03.020.

Saisana, M. (2004): Composite indicators: a review. Second Workshop on Composite Indicators of Country Performance. OECD. Paris, 2004. Available online at https://www.researchgate.net/profile/michaela-saisana/publication/267986167_composite_indicators_-_a_review/links/554b77e10cf21ed213594143/composite-indicators-a-review.pdf.

Saisana, M.; Saltelli, A.; Tarantola, S. (2005): Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. In *Journal of the Royal Statistical Society Series A: Statistics in Society* 168 (2), pp. 307–323. DOI: 10.1111/j.1467-985X.2005.00350.x.

Saisana, Michaela; Tarantola, Stefano (2002): State-of-the-art report on current methodologies and practices for composite indicator development. 214th ed. Ispra, Italy: European Commission, Joint Research Centre, Institute for the Protection and …

Sébastien, Léa; Bauler, Tom (2013): Use and influence of composite indicators for sustainable development at the EU-level. In *Ecological Indicators* 35, pp. 3–12. DOI: 10.1016/j.ecolind.2013.04.014.

Suo, Qiuling; Yao, Liuyi; Xun, Guangxu; Sun, Jianhui; Zhang, Aidong (2019): Recurrent Imputation for Multivariate Time Series with Missing Values. In : 2019 IEEE International Conference on Healthcare Informatics (ICHI): IEEE.

van Buuren, S.; Oudshoorn, C.G.M. (2000): Multivariate imputation by chained equations. MICE V1.0 User's manual. 1-68. TNO Prevention and Health. Available online at https://publications.tno.nl/publication/34618573/amelnr/buuren-2000-multivariate.pdf.

Zhou, P.; Zhang, L. P. (2018): Composite Indicators for Sustainability Assessment: Methodological Developments. In Ruizhi Pang, Bai Xuejie, C. A. Knox Lovell (Eds.): Energy, environment and transitional green growth in China. Singapore: Springer, pp. 15–36. Available online at https://link.springer.com/chapter/10.1007/978-981-10-7919-1_2.

Appendix

A1. Result Indicator description of Pillar 2 taken from the Monitoring Framework (European Commission. Directorate General for Agriculture and Rural Development 2015)

Indicator	Indicator description
R.01_PII	percentage of agricultural holdings with RDP support for investments in restructuring or modernisation (focus area 2A)
R.02_PII	Change in Agricultural output on supported farms/AWU (Annual Work Unit) (focus area 2A)
R.03_PII	percentage of agricultural holdings with RDP supported business development plan/investments for young farmers (focus area 2B
R.04_PII	percentage of agricultural holdings receiving support for participating in quality schemes, local markets and short supply circuits, and producer groups/organisations (focus area 3A)
R.05_PII	percentage of farms participating in risk management schemes (focus area 3B)
R.06_PII	percentage forest or other wooded area under management contracts supporting biodiversity (focus area 4A)
R.07_PII	percentage agricultural land under management contracts supporting biodiversity and/or landscapes (focus area 4A)
R.08_PII	percentage of agricultural land under management contracts to improve water management (focus area 4B)
R.09_PII	percentage of forestry land under management contracts to improve water management (focus area 4B)
R.10_PII	percentage of agricultural land under management contracts to improve soil management and/or prevent soil erosion (focus area 4C)
R.11_PII	percentage of forestry land under management contracts to improve soil management and/or prevent soil erosion (focus area 4C)
R.12_PII	percentage of irrigated land switching to more efficient irrigation systems (focus area 5A)
R.13_PII	Increase in efficiency of water use in agriculture in RDP supported projects (focus area 5A) (*)
R.14_PII	Increase in efficiency of energy use in agriculture and food-processing in RDP supported projects (focus area 5B) (*)
R.15_PII	Renewable energy produced from supported projects (focus area 5C) (*)
R.16_PII	percentage of LU (Live-stock Unit) concerned by investments in live-stock management in view of reducing GHG (Green House Gas) and/or ammonia emissions (focus area 5D)
R.17_PII	percentage of agricultural land under management contracts targeting reduction of GHG and/or ammonia emissions (focus area 5D)
R.18_PII	Reduced emissions of methane and nitrous oxide (focus area 5D) (*)
R.19_PII	Reduced ammonia emissions (focus area 5D) (*)
R.20_PII	percentage of agricultural and forest land under management contracts contributing to carbon sequestration or conservation (focus area 5E)
R.21_PII	Jobs created in supported projects (focus area 6A)
R.22_PII	percentage of rural population covered by local development strategies (focus area 6B)

R.23_PII	percentage of rural population benefiting from improved services/infrastructures (focus area 6B)
R.24_PII	Jobs created in supported projects (Leader) (focus area 6B)
R.25_PII	percentage of rural population benefiting from new or improved services/infrastructures (Information and Communication Technology - ICT) (focus area 6C)

A2. Amount of zeros and NAs per member state over all years and indicators

Number	Member state	Total amount of	
		zeros	NAs
1	Austria	32	19
2	Belgium	37	19
3	Bulgaria	70	9
4	Croatia	32	18
5	Cyprus	66	9
6	Czechia	37	42
7	Denmark	28	40
8	Estonia	27	36
9	Finland	51	15
10	France	27	8
11	Germany	17	7
12	Greece	60	1
13	Hungary	41	9
14	Ireland	44	35
15	Italy	16	0
16	Latvia	35	18
17	Lithuania	38	17
18	Luxembourg	64	40
19	Malta	81	24
20	Netherlands	63	48
21	Poland	41	26
22	Portugal	22	10
23	Romania	35	9
24	Slovakia	49	22
25	Slovenia	57	33
26	Spain	21	0
27	Sweden	48	18

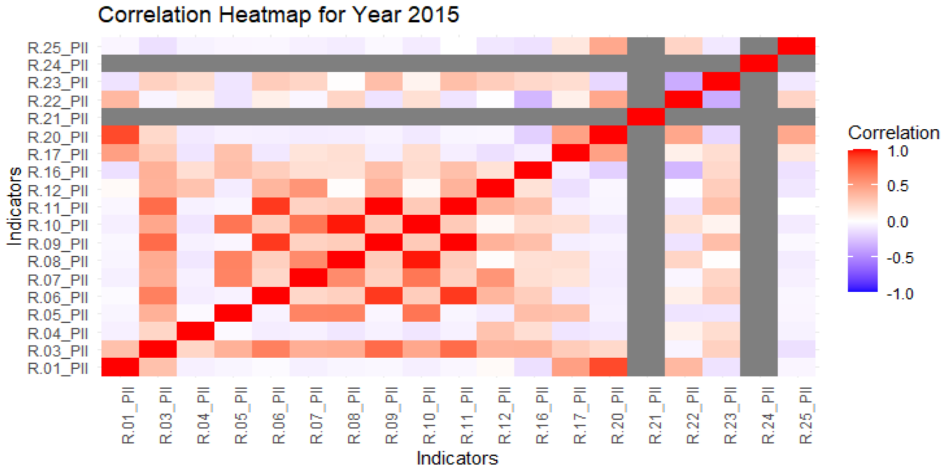
A3. Total amount of zeros and NAs per year

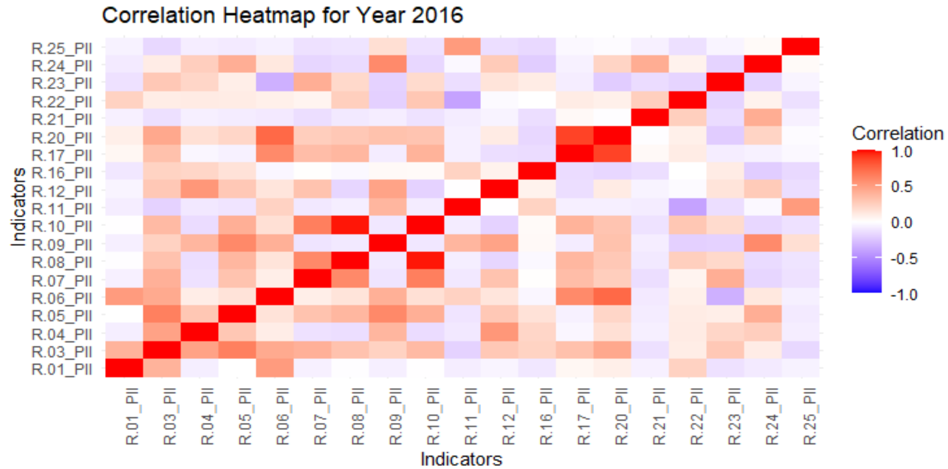
Year	Total amount of zeros	Total amount of NAs
2015	327	81
2016	199	73
2017	149	62
2018	107	63
2019	94	66
2020	89	66
2021	88	63
2022	86	58

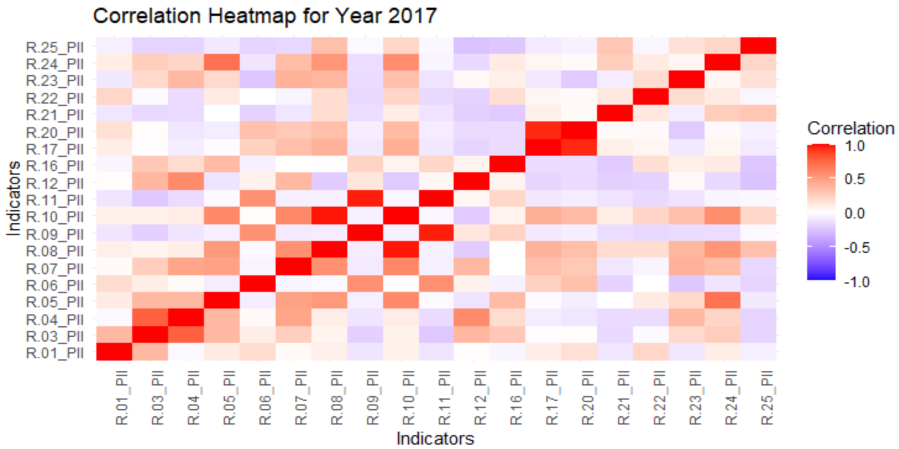
A4. Summary statistics for imputed data via MICE

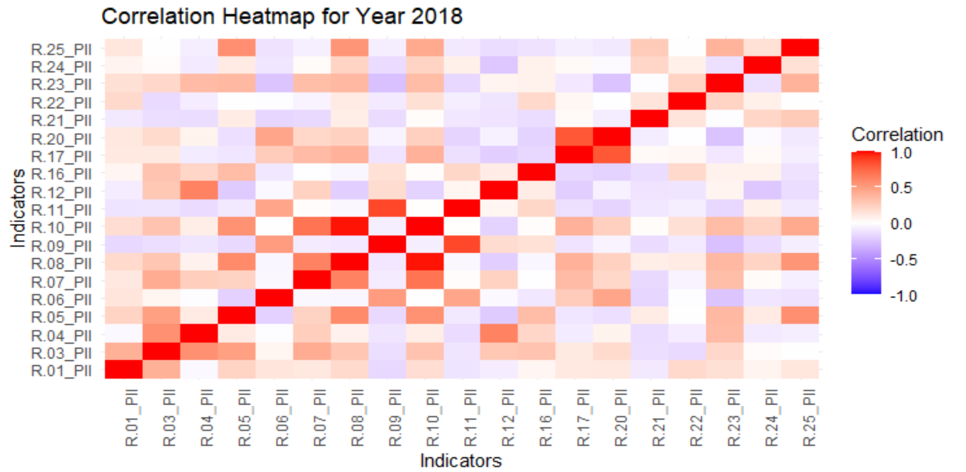
Indicator	Statistics across 27 member states and eight years			
	Min	Max	Mean	SD
R.01_PII	0	53.000	4.738	8.677
R.03_PII	0.003	11.825	2.241	2.387
R.04_PII	0.000	58.458	4.352	10.499
R.05_PII	0.000	125.029	13.284	33.806
R.06_PII	0.000	12.141	0.763	1.533
R.07_PII	0.000	98.437	21.933	24.066
R.08_PII	0.000	88.473	18.113	23.568
R.09_PII	0.000	2.229	0.160	0.377
R.10_PII	0.000	91.471	18.351	22.616
R.11_PII	0.000	2.306	0.201	0.446
R.12_PII	0.000	6.607	1.215	1.732
R.16_PII	0.000	22.125	6.749	8.969
R.17_PII	0.000	50.555	3.081	8.976
R.20_PII	0.000	14.155	1.065	2.720
R.21_PII	0.000	7070.000	514.769	1233.846
R.22_PII	0.000	105.750	60.206	28.778
R.23_PII	0.000	120.743	36.591	39.073
R.24_PII	0.000	14640.000	812.546	1854.448
R.25_PII	0	32.165	2.919	6.740

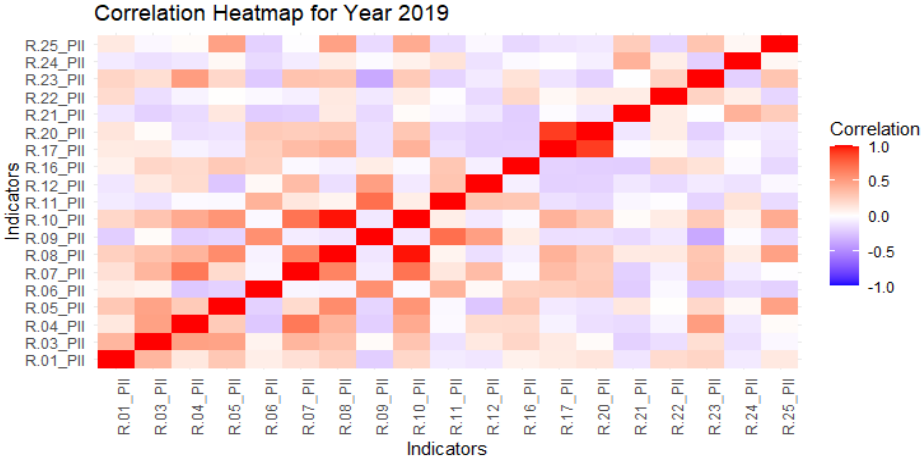
A5. Year-wise Correlation Heatmap of indicators

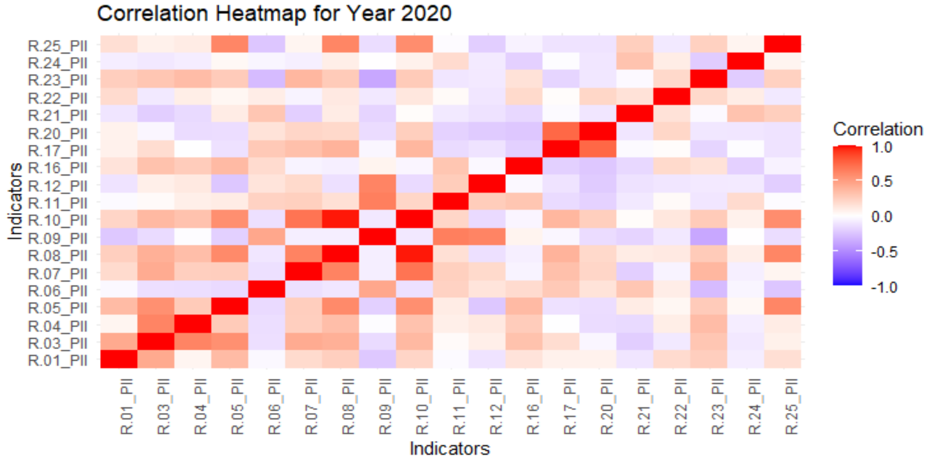


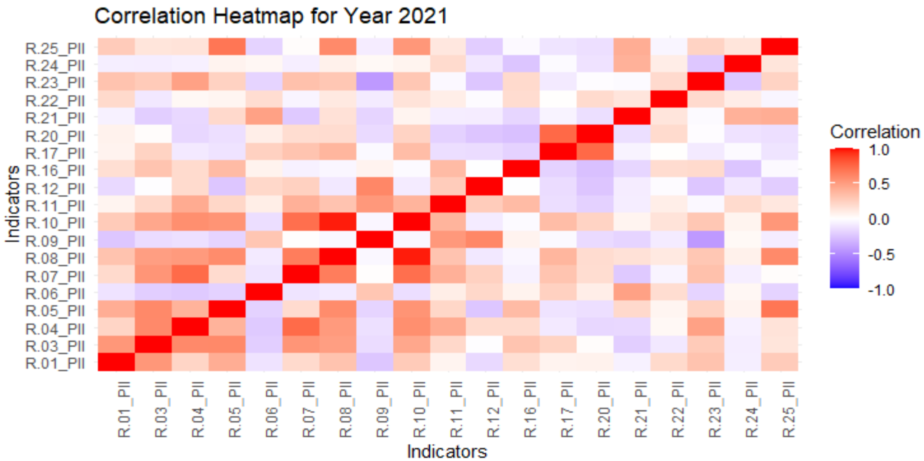


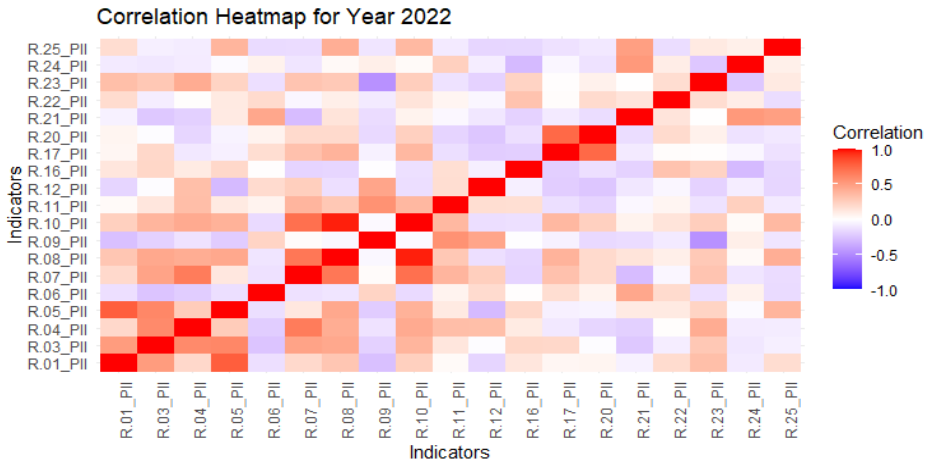




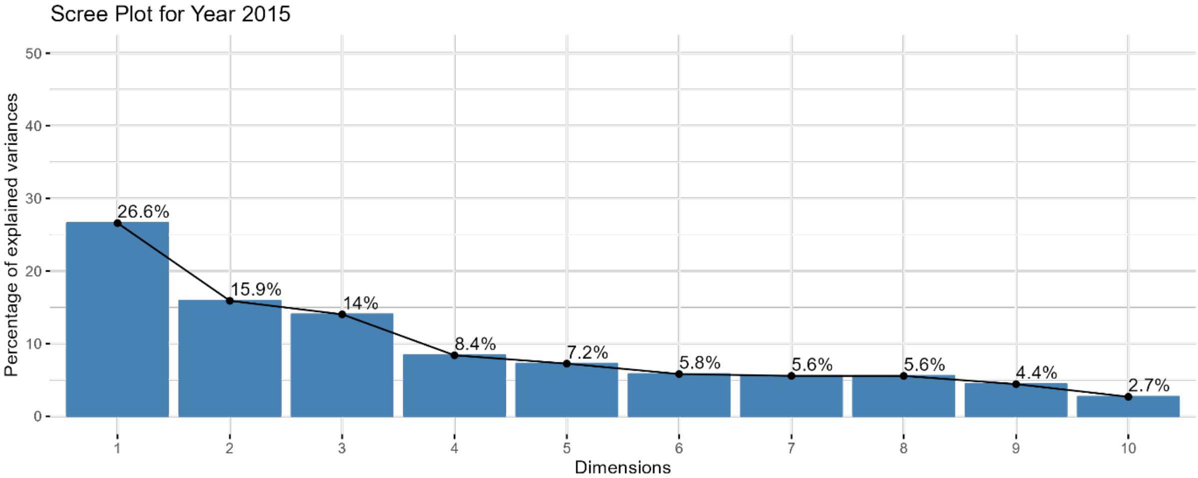

Correlation Heatmap for Year 2019
R.25_PII
R.24_PII
R.23_PII
R.22_PII
R.21_PII
R.20_PII
R.17_PII
R.16_PII
R.12_PII
R.11_PII
R.10_PII
R.09_PII
R.08_PII
R.07_PII
R.06_PII
R.05_PII
R.04_PII
R.03_PII
R.01_PII
Indicators
R.01_PII R.03_PII R.04_PII R.05_PII R.06_PII R.07_PII R.08_PII R.09_PII R.10_PII R.11_PII R.12_PII R.16_PII R.17_PII R.20_PII R.21_PII R.22_PII R.23_PII R.24_PII R.25_PII
Indicators
Correlation
1.0
0.5
0.0
-0.5
-1.0

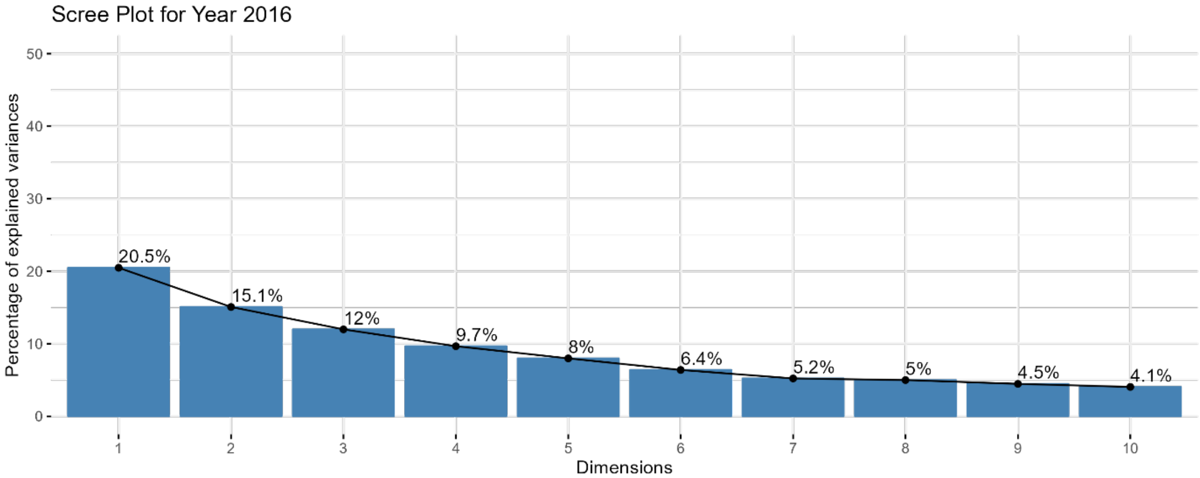

Correlation Heatmap for Year 2020
R.25_PII
R.24_PII
R.23_PII
R.22_PII
R.21_PII
R.20_PII
R.17_PII
R.16_PII
R.12_PII
R.11_PII
R.10_PII
R.09_PII
R.08_PII
R.07_PII
R.06_PII
R.05_PII
R.04_PII
R.03_PII
R.01_PII
Indicators
R.01_PII R.03_PII R.04_PII R.05_PII R.06_PII R.07_PII R.08_PII R.09_PII R.10_PII R.11_PII R.12_PII R.16_PII R.17_PII R.20_PII R.21_PII R.22_PII R.23_PII R.24_PII R.25_PII
Indicators
Correlation
1.0
0.5
0.0
-0.5
-1.0

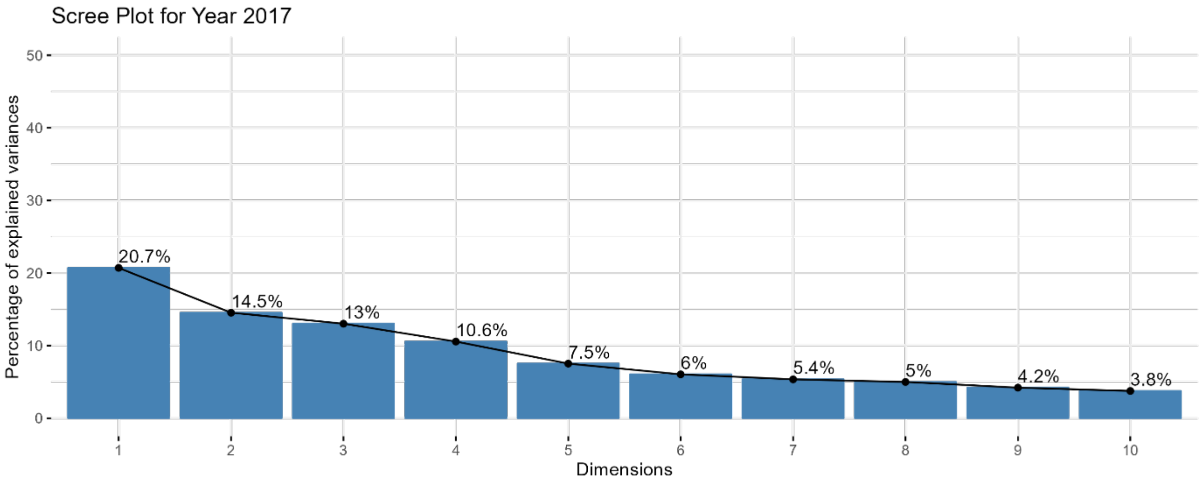

Correlation Heatmap for Year 2021
R.25_PII
R.24_PII
R.23_PII
R.22_PII
R.21_PII
R.20_PII
R.17_PII
R.16_PII
R.12_PII
R.11_PII
R.10_PII
R.09_PII
R.08_PII
R.07_PII
R.06_PII
R.05_PII
R.04_PII
R.03_PII
R.01_PII
Indicators
R.01_PII R.03_PII R.04_PII R.05_PII R.06_PII R.07_PII R.08_PII R.09_PII R.10_PII R.11_PII R.12_PII R.16_PII R.17_PII R.20_PII R.21_PII R.22_PII R.23_PII R.24_PII R.25_PII
Indicators
Correlation
1.0
0.5
0.0
-0.5
-1.0


Correlation Heatmap for Year 2022
R.25_PII
R.24_PII
R.23_PII
R.22_PII
R.21_PII
R.20_PII
R.17_PII
R.16_PII
R.12_PII
R.11_PII
R.10_PII
R.09_PII
R.08_PII
R.07_PII
R.06_PII
R.05_PII
R.04_PII
R.03_PII
R.01_PII
Indicators
R.01_PII R.03_PII R.04_PII R.05_PII R.06_PII R.07_PII R.08_PII R.09_PII R.10_PII R.11_PII R.12_PII R.16_PII R.17_PII R.20_PII R.21_PII R.22_PII R.23_PII R.24_PII R.25_PII
Indicators
Correlation
1.0
0.5
0.0
-0.5
-1.0

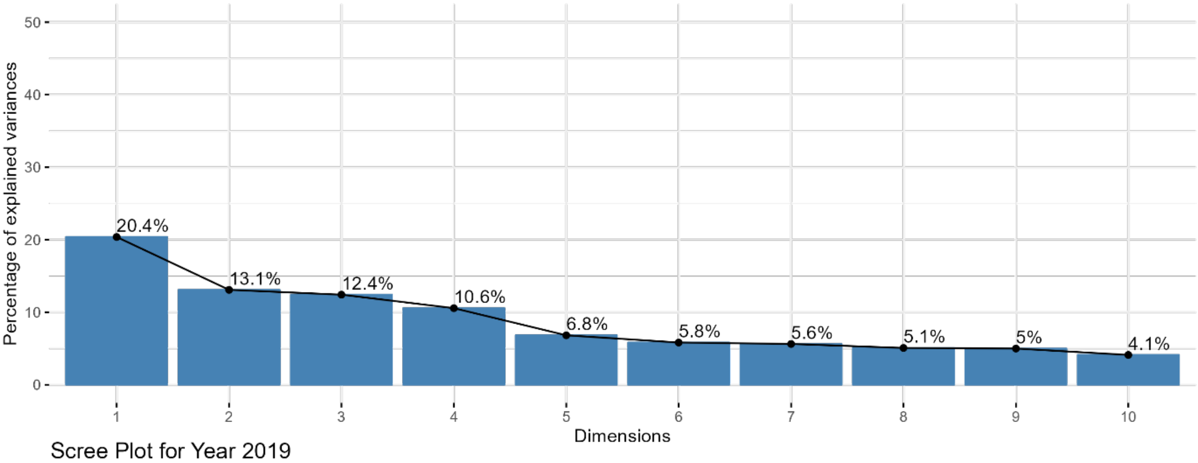
A6. PCA result plots for each year - Scree plots



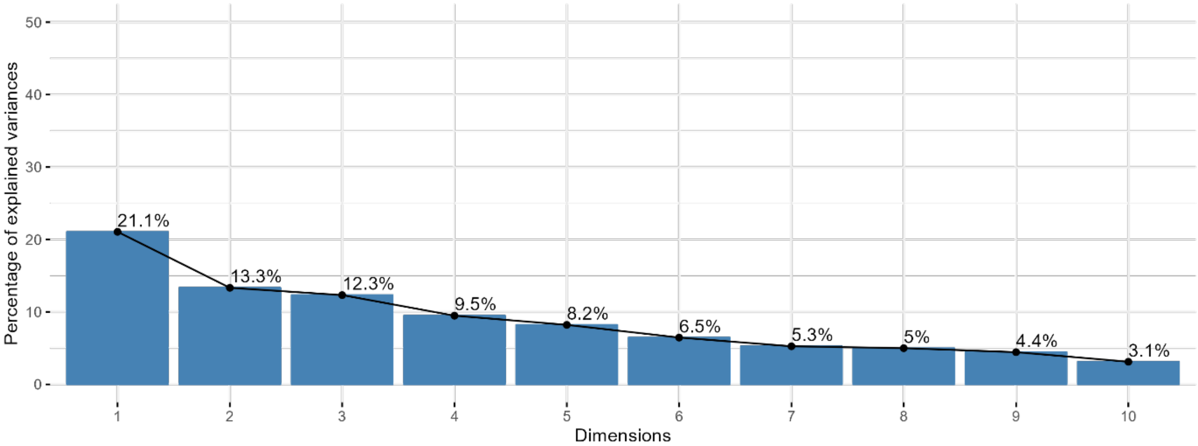




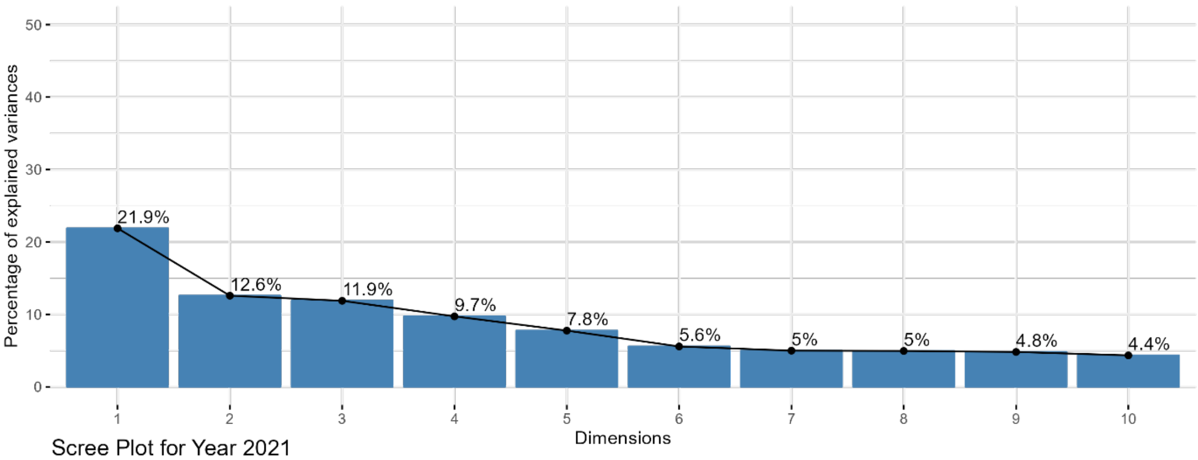
Scree Plot for Year 2018



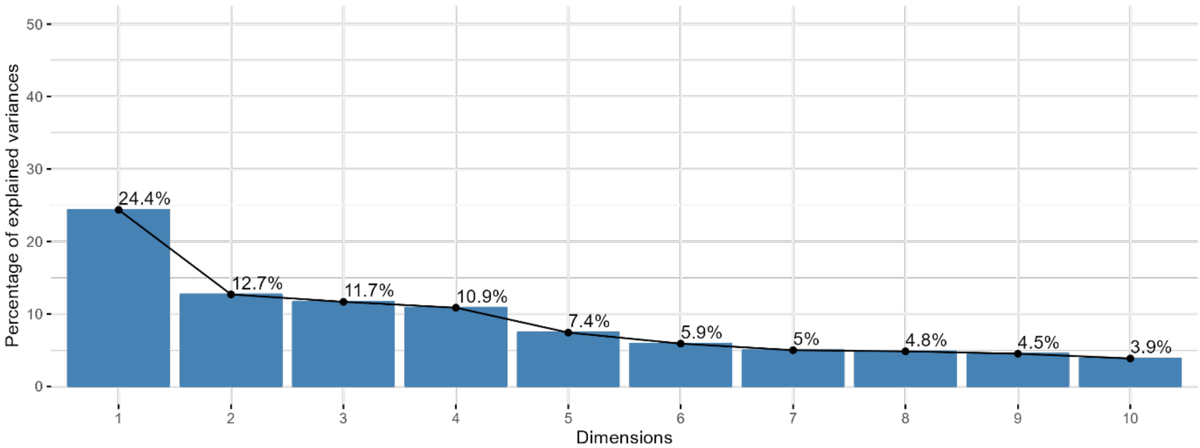
Scree Plot for Year 2019

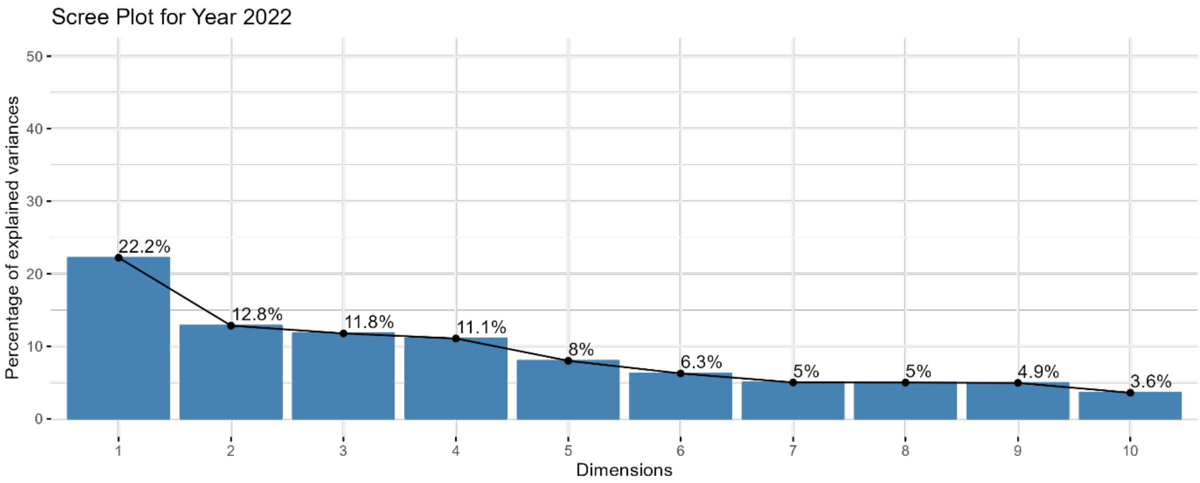


Scree Plot for Year 2020

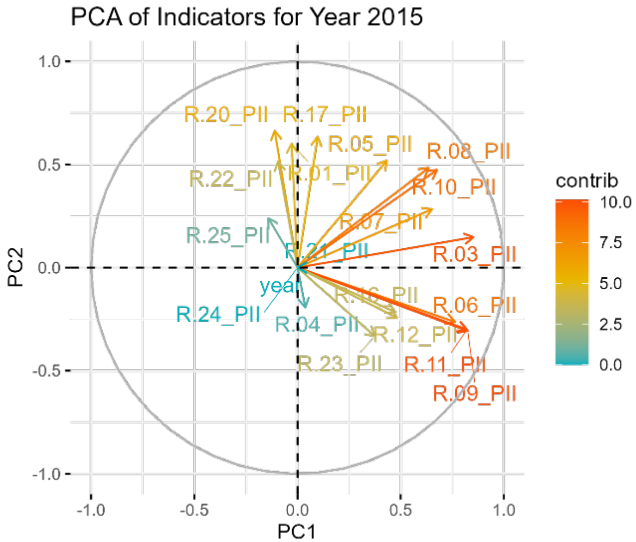


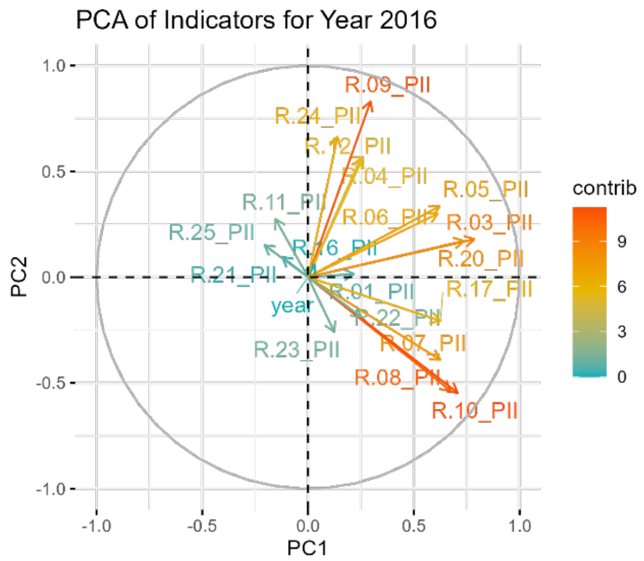
Scree Plot for Year 2021

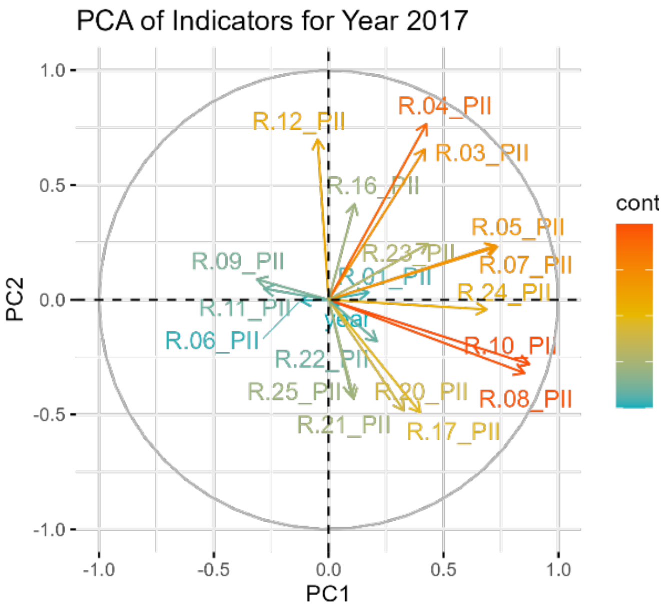


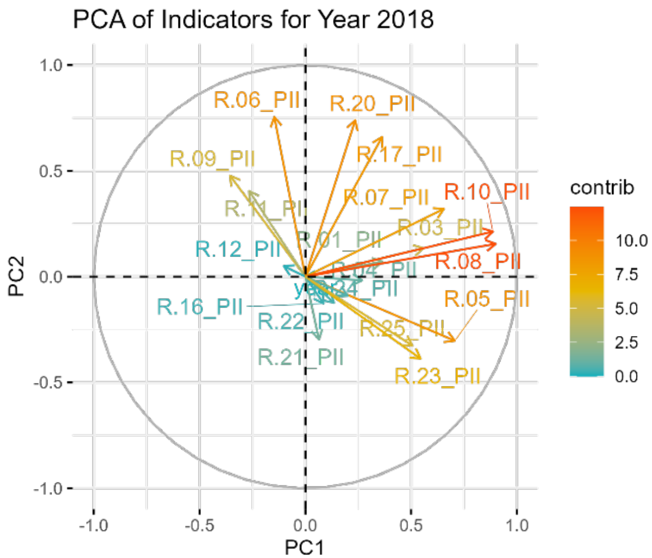

Scree Plot for Year 2022
Percentage of explained variances
0
10
20
30
40
50
22.2%
12.8%
11.8%
11.1%
8%
6.3%
5%
5%
4.9%
3.6%
1
2
3
4
5
6
7
8
9
10
Dimensions

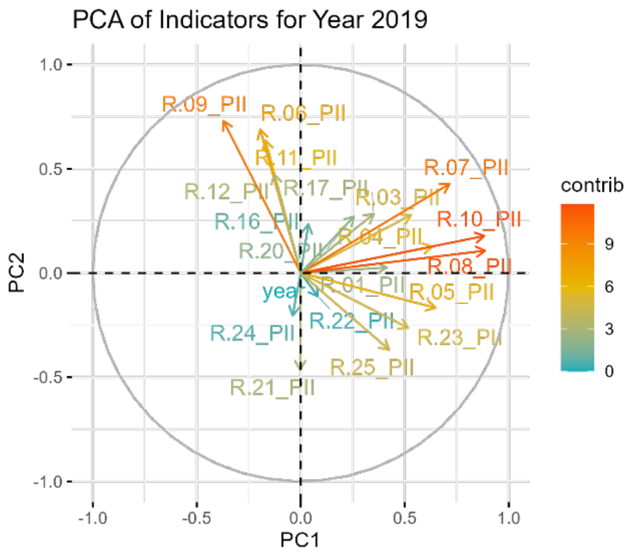
A7. PCA results for each year - Biplots

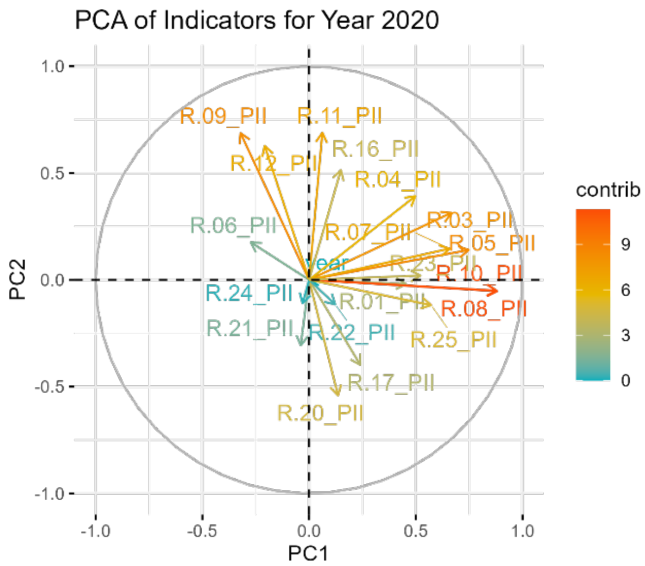


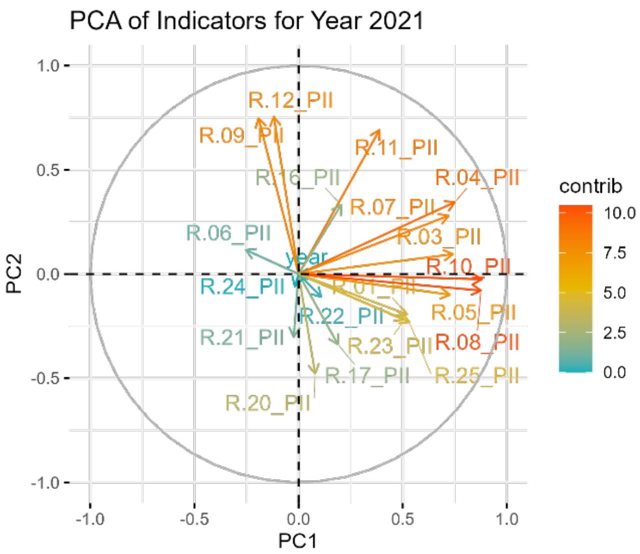


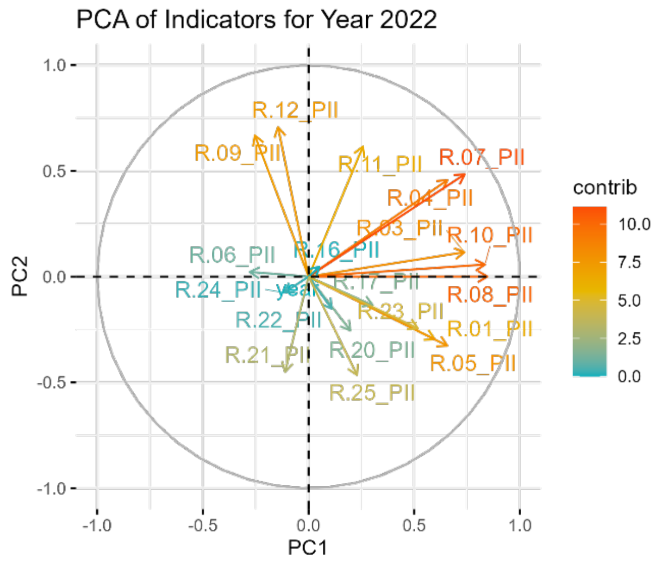





PCA of Indicators for Year 2019
PC1
PC2
contrib
R.09_PII
R.06_PII
R.11_PII
R.12_PII
R.17_PII
R.03_PII
R.07_PII
R.16_PII
R.10_PII
R.04_PII
R.20_PII
R.08_PII
R.01_PII
R.05_PII
R.22_PII
R.24_PII
R.23_PII
R.25_PII
R.21_PII

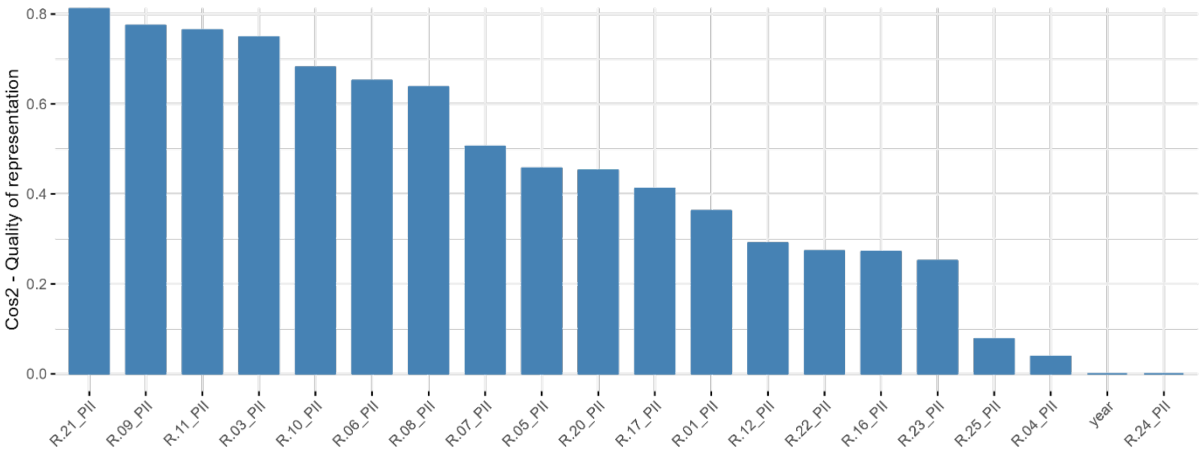

PCA of Indicators for Year 2020
PC1
PC2
contrib
R.09_PII
R.11_PII
R.12_PII
R.16_PII
R.04_PII
R.06_PII
R.07_PII
R.03_PII
R.05_PII
year
R.23_PII
R.10_PII
R.24_PII
R.01_PII
R.08_PII
R.21_PII
R.22_PII
R.25_PII
R.17_PII
R.20_PII


PCA of Indicators for Year 2021
PC1
PC2
contrib
R.12_PII
R.09_PII
R.11_PII
R.16_PII
R.04_PII
R.07_PII
R.06_PII
R.03_PII
year
R.10_PII
R.24_PII
R.01_PII
R.22_PII
R.05_PII
R.21_PII
R.23_PII
R.08_PII
R.17_PII
R.25_PII
R.20_PII

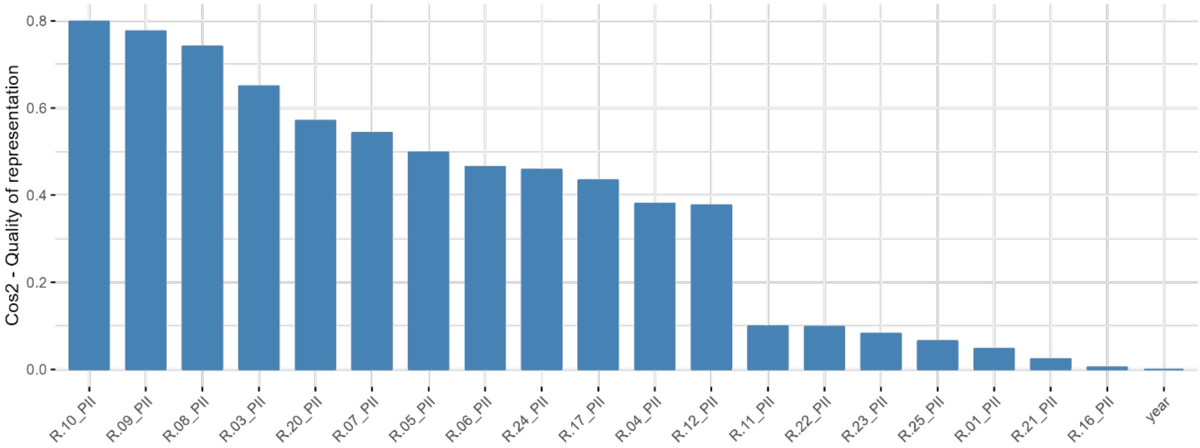

PCA of Indicators for Year 2022
PC1
PC2
contrib
R.12_PII
R.09_PII
R.11_PII
R.07_PII
R.04_PII
R.03_PII
R.10_PII
R.06_PII
R.16_PII
R.24_PII
year
R.17_PII
R.08_PII
R.22_PII
R.23_PII
R.01_PII
R.21_PII
R.20_PII
R.05_PII
R.25_PII

A8. PCA results for each year – Cos2 plots

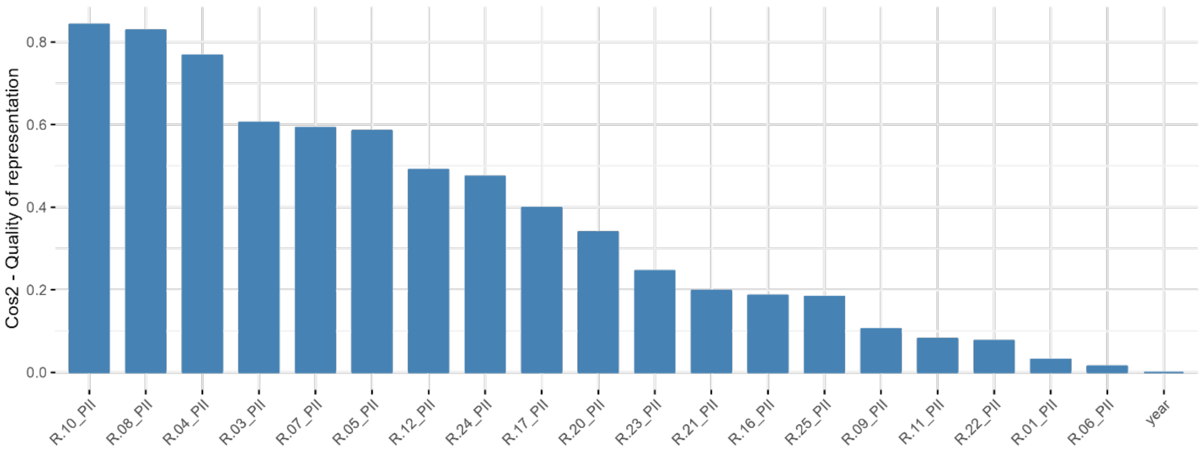
Cos2 for Year 2015



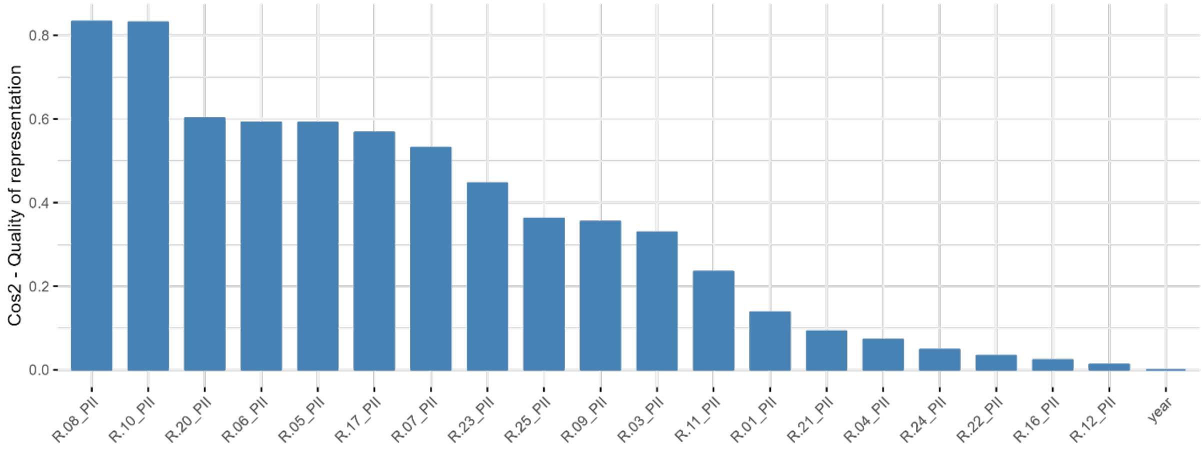
Cos2 for Year 2016



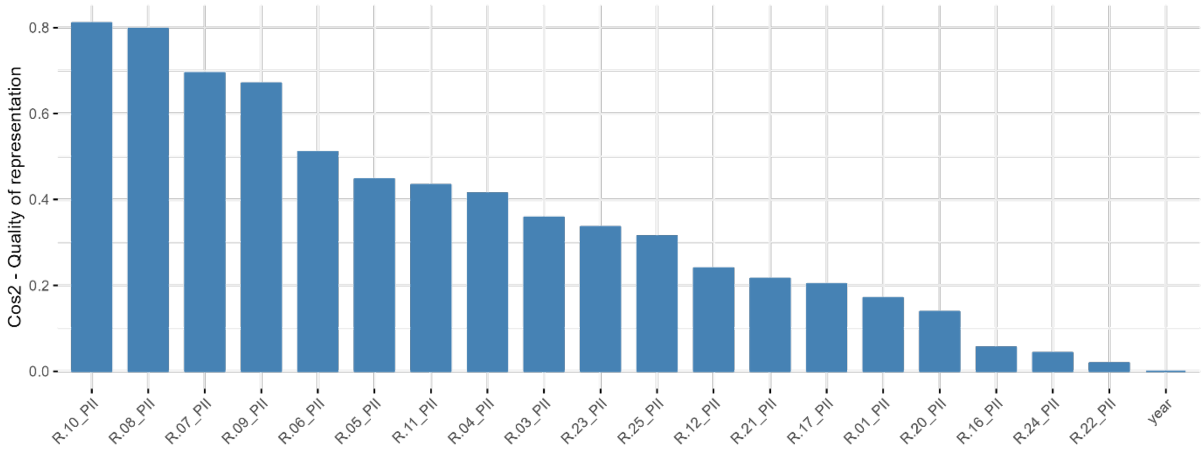
Cos2 for Year 2017



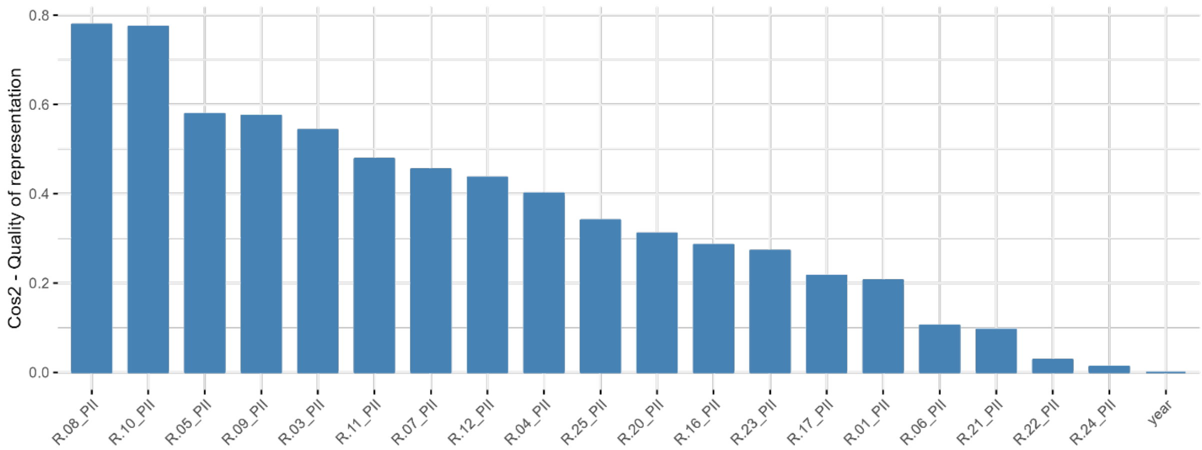
Cos2 for Year 2018

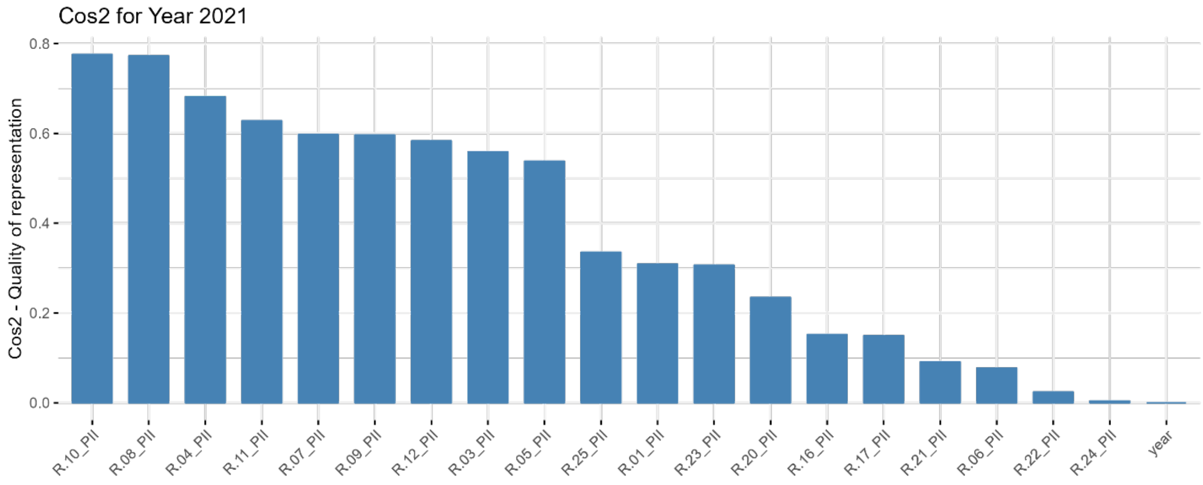


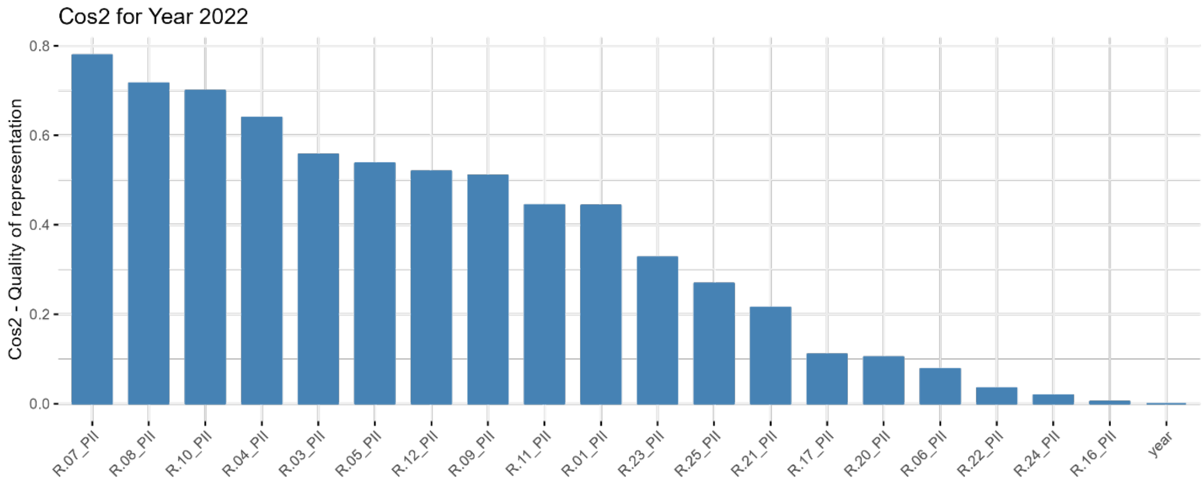
Cos2 for Year 2019



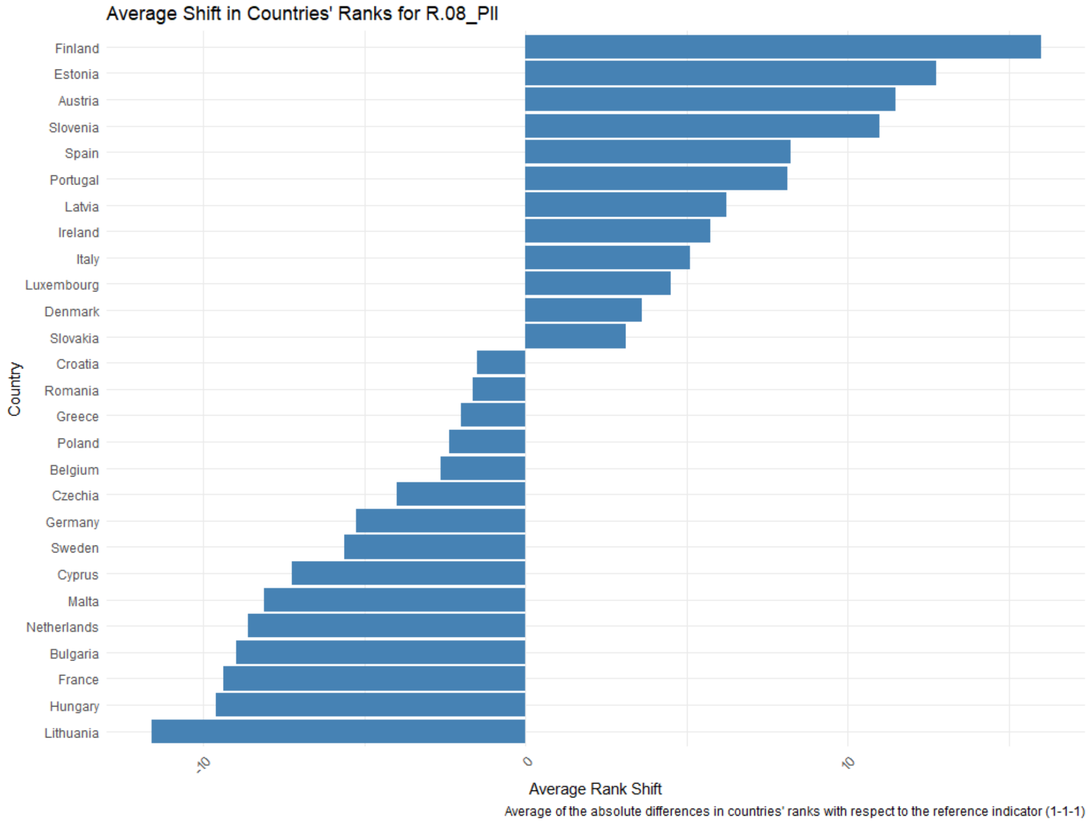
Cos2 for Year 2020

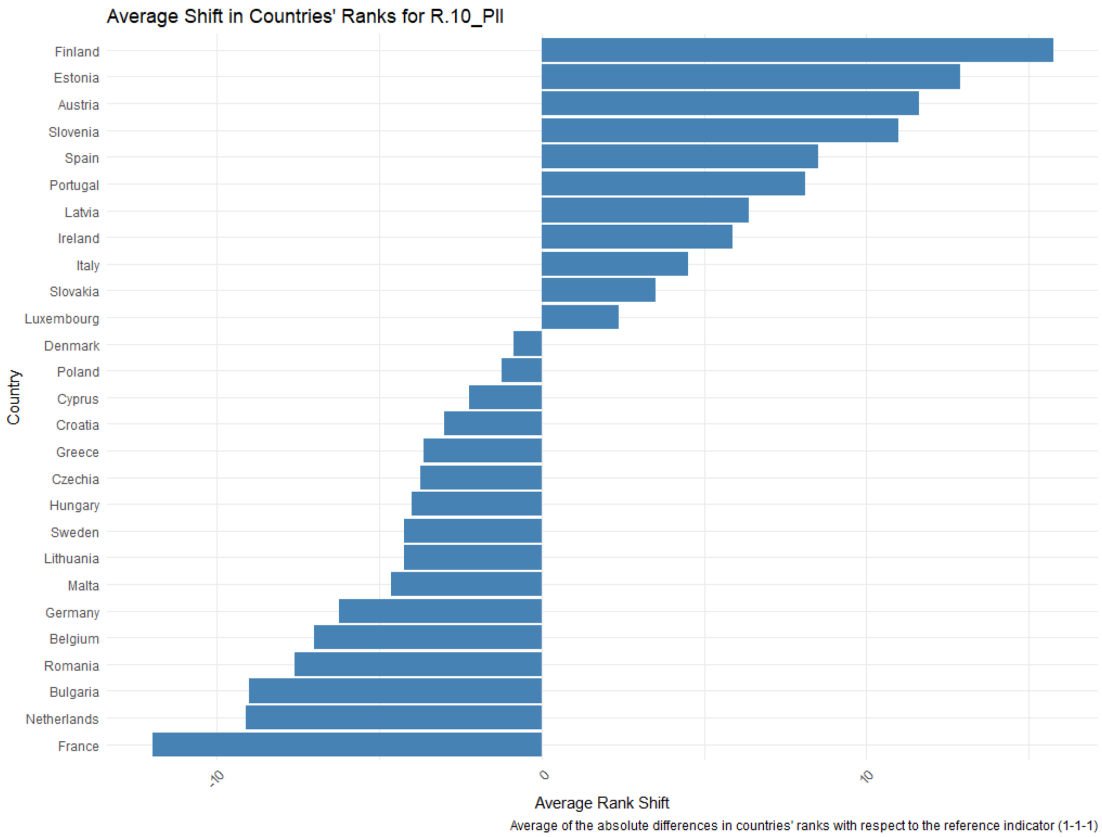


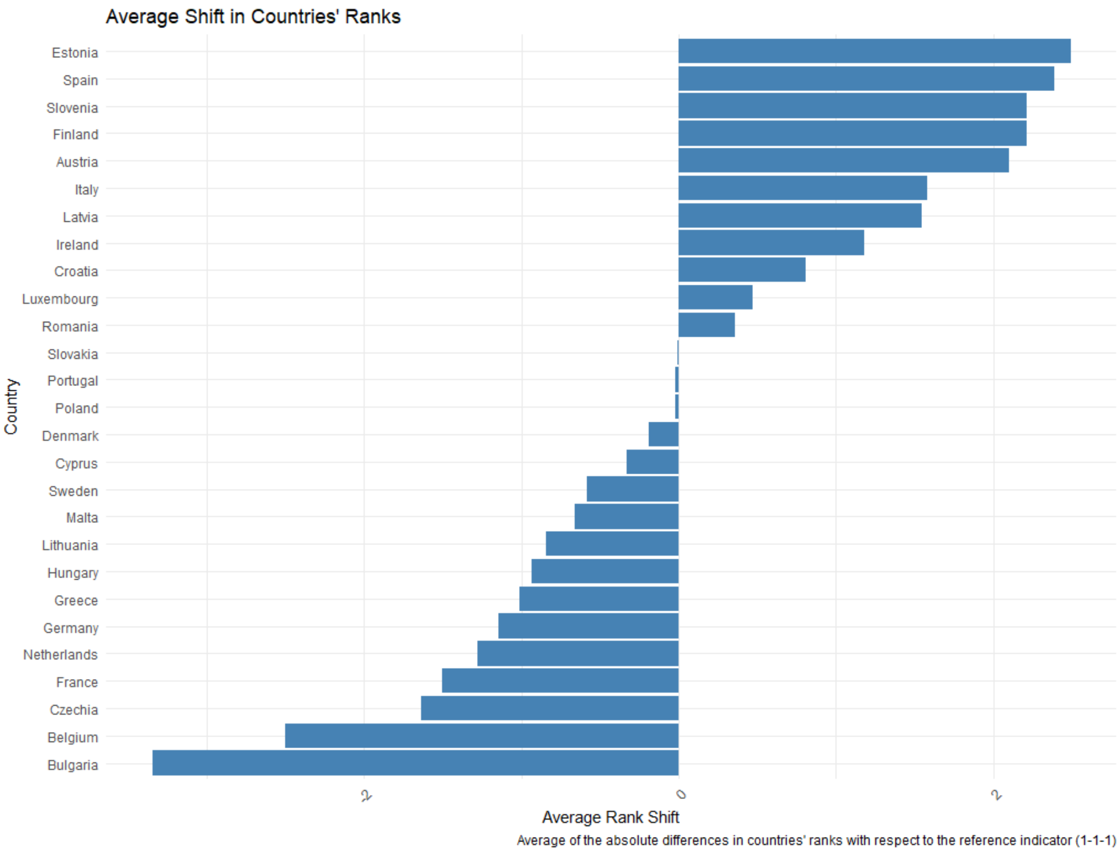

Cos2 for Year 2021
Cos2 - Quality of representation
0.0
0.2
0.4
0.6
0.8
R.10_PII
R.08_PII
R.04_PII
R.11_PII
R.07_PII
R.09_PII
R.12_PII
R.03_PII
R.05_PII
R.25_PII
R.01_PII
R.23_PII
R.20_PII
R.16_PII
R.17_PII
R.21_PII
R.06_PII
R.22_PII
R.24_PII
year

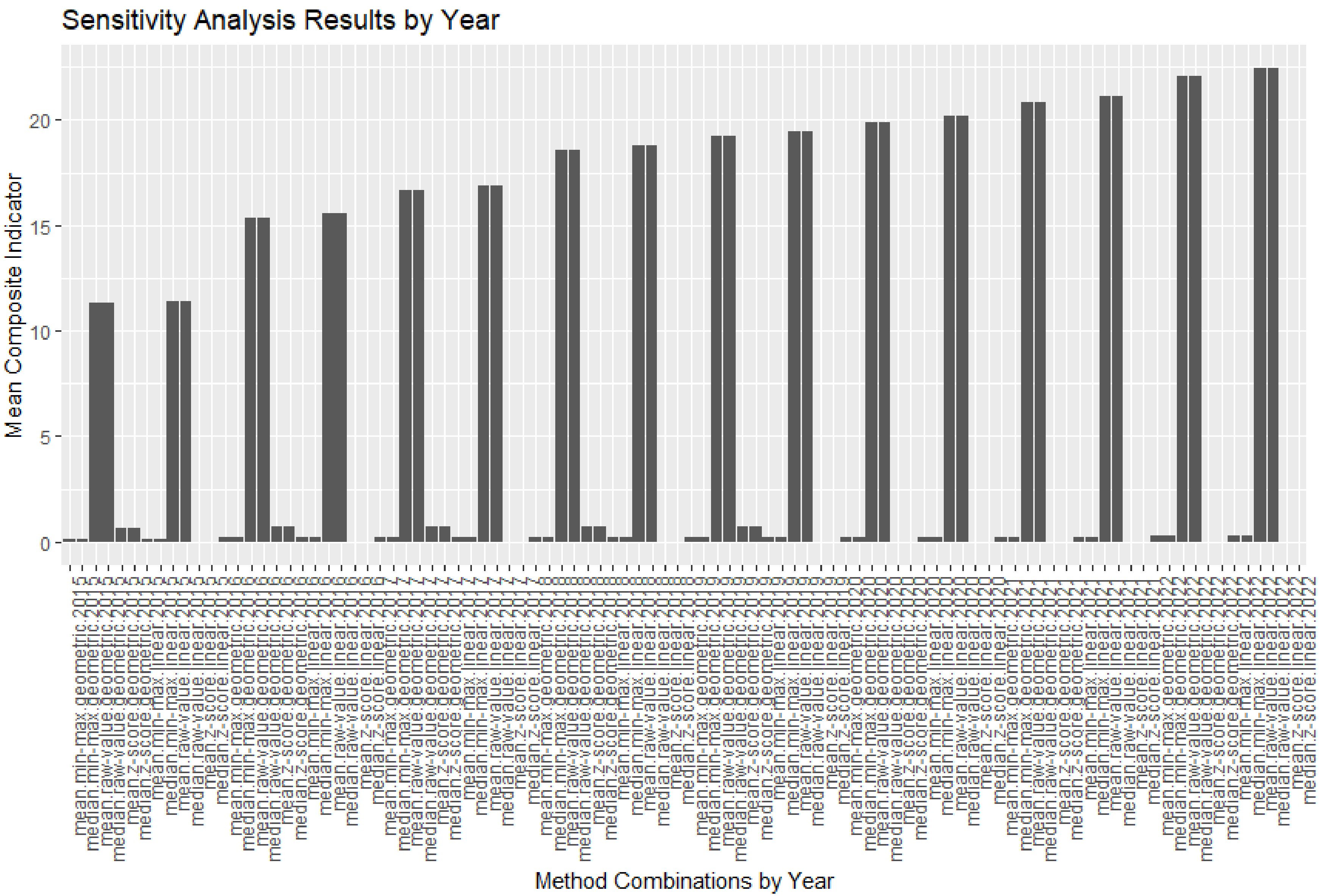

Cos2 for Year 2022
Cos2 - Quality of representation
0.0
0.2
0.4
0.6
0.8
R.07_PII
R.08_PII
R.10_PII
R.04_PII
R.03_PII
R.05_PII
R.12_PII
R.09_PII
R.11_PII
R.01_PII
R.23_PII
R.25_PII
R.21_PII
R.17_PII
R.20_PII
R.06_PII
R.22_PII
R.24_PII
R.16_PII
year

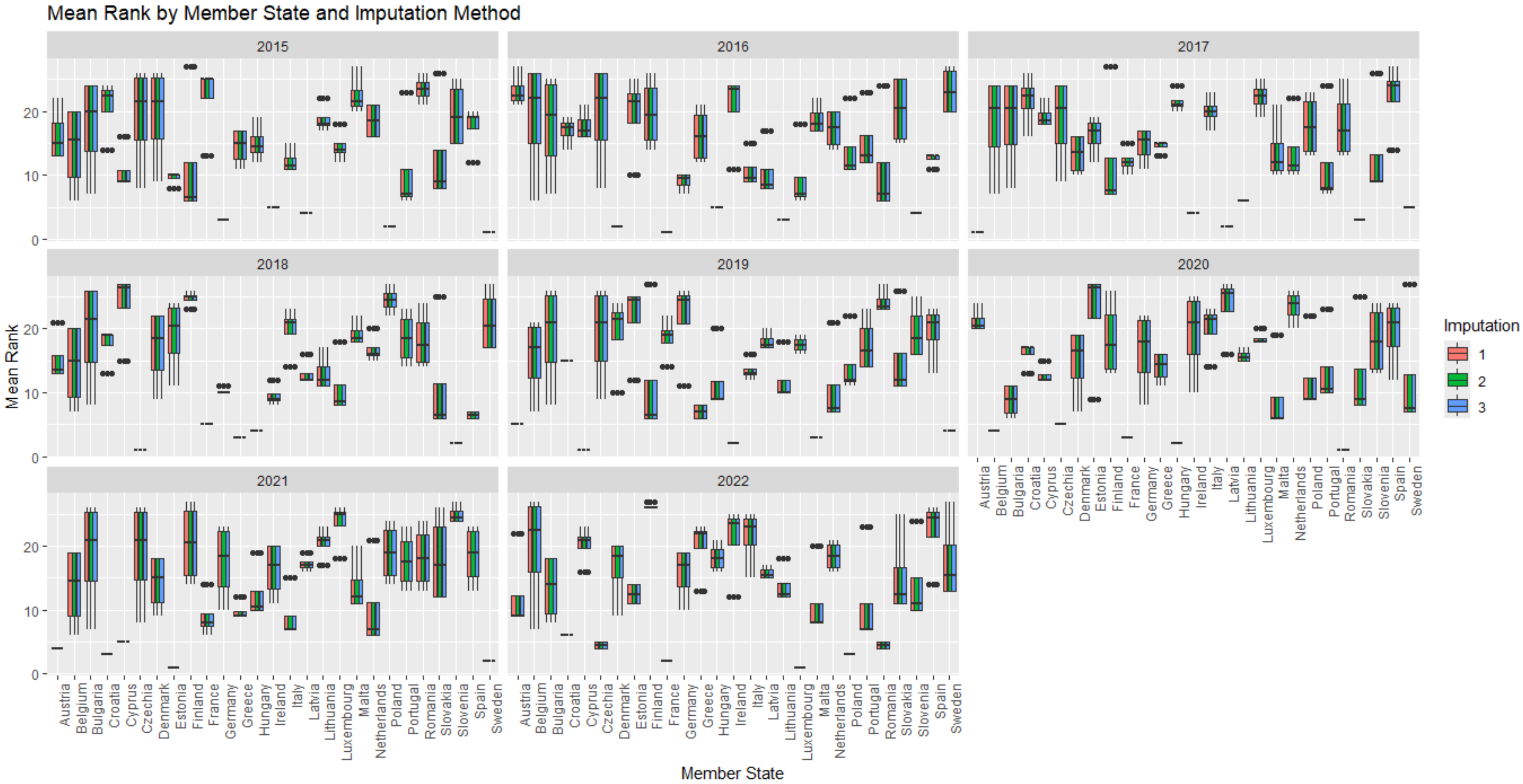
A9. Uncertainty and Sensitivity Analysis of comp_R.08_PII_R.10_PII



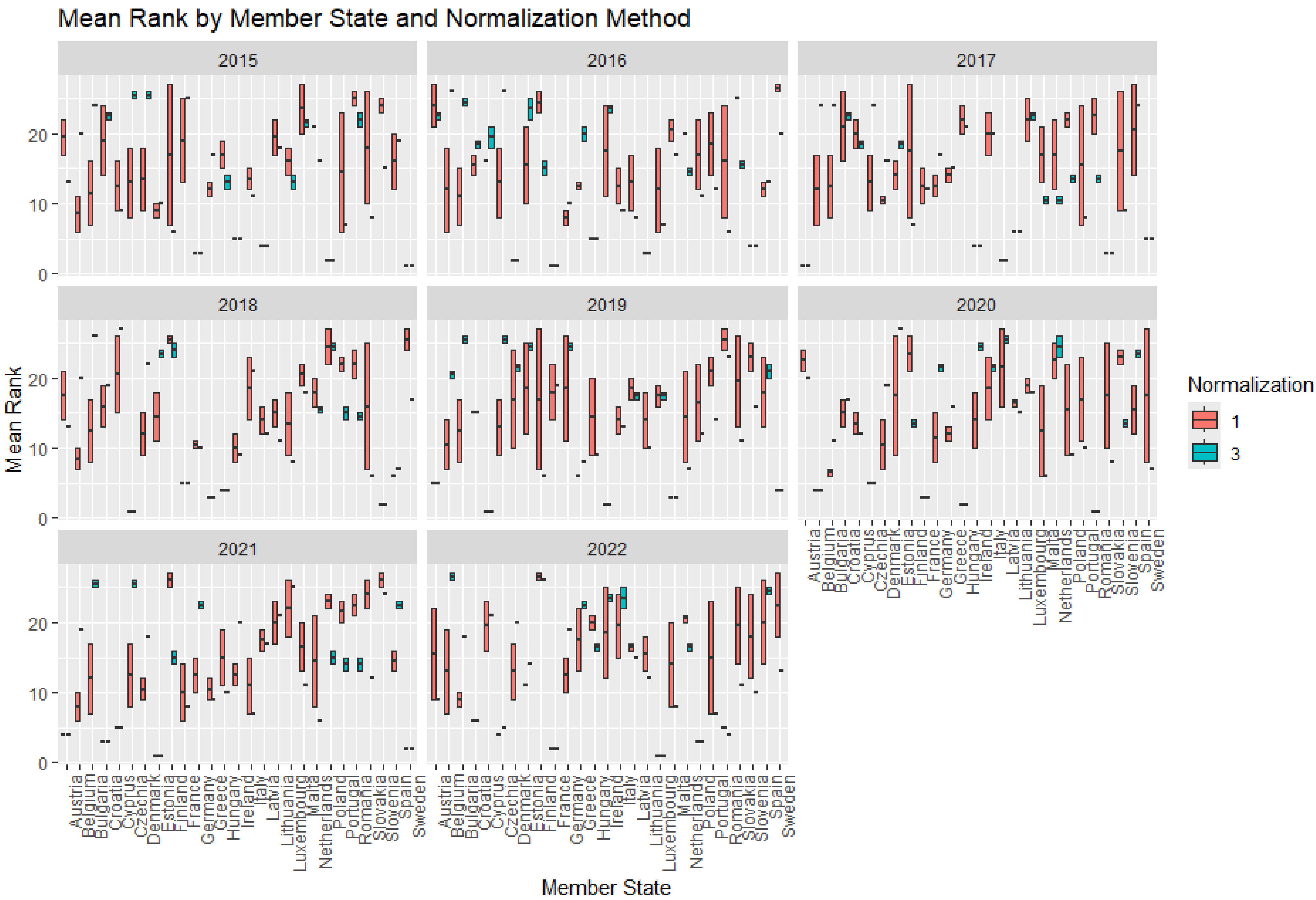


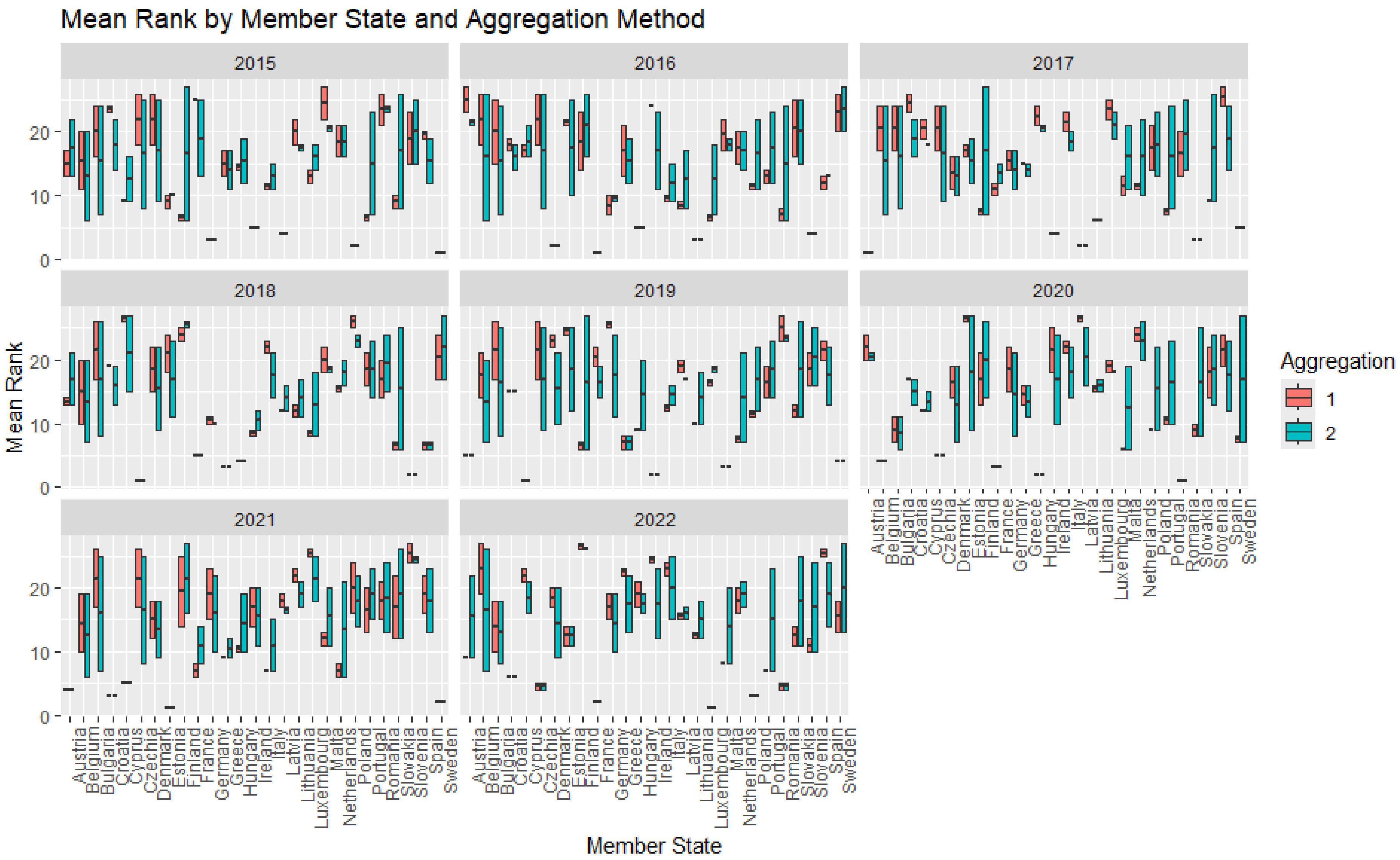

Average Shift in Countries' Ranks
Country
Estonia
Spain
Slovenia
Finland
Austria
Italy
Latvia
Ireland
Croatia
Luxembourg
Romania
Slovakia
Portugal
Poland
Denmark
Cyprus
Sweden
Malta
Lithuania
Hungary
Greece
Germany
Netherlands
France
Czechia
Belgium
Bulgaria
Average Rank Shift
Average of the absolute differences in countries' ranks with respect to the reference indicator (1-1-1)


Sensitivity Analysis Results by Year
Mean Composite Indicator
0
5
10
15
20
Method Combinations by Year



Mean Rank by Member State and Imputation Method
2015
2016
2017
2018
2019
2020
2021
2022
Mean Rank
0
10
20
Member State
Imputation
1
2
3
Austria
Belgium
Bulgaria
Croatia
Cyprus
Czechia
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
Netherlands
Poland
Portugal
Romania
Slovakia
Slovenia
Spain
Sweden


Mean Rank by Member State and Normalization Method
2015
2016
2017
2018
2019
2020
2021
2022
Mean Rank
Member State
Normalization
1
3
0
10
20
Austria
Belgium
Bulgaria
Croatia
Cyprus
Czechia
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
Netherlands
Poland
Portugal
Romania
Slovakia
Slovenia
Spain
Sweden


Mean Rank by Member State and Aggregation Method
2015
2016
2017
2018
2019
2020
2021
2022
Mean Rank
Member State
Aggregation
1
2
0
10
20
Austria
Belgium
Bulgaria
Croatia
Cyprus
Czechia
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
Netherlands
Poland
Portugal
Romania
Slovakia
Slovenia
Spain
Sweden

Hohenheimer Agrarökonomische Arbeitsberichte

Becker, T.; Benner, E (2000): Zur Problematik der Herkunftsangabe im regionalen Marketing
Arbeitsbericht Nr. 1

Chaipan, C. (2000): The Euro and its Impact on ASEAN Economies
Arbeitsbericht Nr. 2

Henze, A. (2000): Marktwirtschaft - Wirtschaftliche Freiheit, motiviertes Handeln und Wettbewerb: Que llen des Wohlstands
Arbeitsbericht Nr. 3

Benner, E. (2000): Zur effizienten Herkunftsangabe im europäischen Binnenmarkt
Arbeitsbericht Nr. 4

Vorgrimler, D. (2000): Wettbewerbstheorie und stagnierende Märkte
Arbeitsbericht Nr. 5

Beerbaum, S. (2001): Grundzüge einer internationalen Zusammenarbeit im Klimaschutz aus ökonomischer Sicht
Arbeitsbericht Nr. 6

Vorgrimler, D.; Wübben, D. (2001): Prognose der Entwicklung des Agrartechnikmarktes - Eine Expertenbefragung nach der Delphi-Methode
Arbeitsbericht Nr. 7

Tesch, I. (2003): Informationsbedarf und Informationsbeschaffung von Konsumenten bei Lebensmitteln pflanzlicher Herkunft - Eine empirische Untersuchung anhand von Fokus-Gruppen -
Arbeitsbericht Nr. 8

Benner, D. (2004): Quality Ambiguity and the Market Mechanism for Credence Goods
Arbeitsbericht Nr. 9

Benner, E., Kliebisch, C. (2004): Regio-Marketing-Strategien des Lebensmitteleinzelhandels
Arbeitsbericht Nr. 10

Benner, E., Heidecke, S.-J. (2005): Grundpreisaufschläge bei Groß- und Familienpackungen - eine empirische Untersuchung im deutschen und französischen Lebensmitteleinzelhandel -
Arbeitsbericht Nr. 11

Becker, T. (2006): Zur Bedeutung geschützter Herkunftsangaben.
Arbeitsbericht Nr. 12, 1. und 2. Auflage

Elsäßer, A., Benner, E., Becker, T. (2006): Marketing auf Wochenmärkten
Arbeitsbericht Nr. 13

Becker, T. (2006): Die CMA auf dem Prüfstand
Arbeitsbericht Nr. 14

Staus, A. (2007): An Ordinal Regression Model using Dealer Satisfaction Data
Arbeitsbericht Nr. 15

Kliebisch, C., Rügge, M. (2007): Alte und neue Wege des Gemeinschaftsmarketings für Agrarprodukte und Lebensmittel
Arbeitsbericht Nr. 16

Staus, A. (2008): Standard and Shuffled Halton Sequences in a Mixed Logit Model
Arbeitsbericht Nr. 17

Staus, A., Becker, T. (2009): Die Zufriedenheit der Landmaschinenhändler mit den Herstellern
Arbeitsbericht Nr. 18

Becker, T., Heinze, K. (2011): Gesellschaftliches Management von Verbraucherbeschwerden: Funktion und Finanzierung
Arbeitsbericht Nr. 19

Siddig,K., Flaig, D., Luckmann, J., Grethe, H. (2011): A 2004 Social Accounting Matrix for Israel. Documentation of an Economy-Wide Database with a Focus on Agriculture, the Labour Market, and Income Distribution
Working Paper No. 20

Bücheler, G. (2011): Biokraftstoff-Zertifizierungssysteme ISCC und REDcert: Darstellung, Vergleich und kritische Diskussion
Working Paper No. 21

Gebhardt, B. (2012): Akzeptanz und Erfolg kleinräumiger Systeme der Lebensmittelversorgung im urbanen Umfeld am Beispiel Stuttgart - Empirische Untersuchungen von Verbrauchern und Unternehmen
Working Paper No. 22

Luckmann, J., McDonald, S. (2014): Stage_W: An Applied General Equilibrium Model With Multiple Types of Water
Working Paper No. 23

Hauck, M., Becker, T. (2015): Evaluierung des Qualitätszeichens Baden-Württemberg (QZBW) aus der Sicht der Teilnehmer
Arbeitsbericht Nr. 24

Semenenko, K., Becker, T. (2015): Entwicklung der Zufriedenheit der Landmaschinenhändler mit den Herstellern
Arbeitsbericht Nr. 25

Gebhardt, B. (2016): Beschreibung von 24 Nachhaltigkeitspreisen in Deutschland mit Relevanz für Unternehmen der Ernährungsbranche
Arbeitsbericht Nr. 26

Gebhardt, B., Ding, J.L., Feisthauer, P. (2018): Obsoleszenz - auch ein Thema bei Lebensmitteln: Ergebnisse einer Expertenbefragung
Arbeitsbericht Nr. 27

Gebhardt, B. (2020): Nachhaltigkeitswettbewerbe in Deutschland 2020. Übersicht und Methodik der Bestandsaufnahme
Arbeitsbericht Nr. 28

Gebhardt, B. (2020): Plant-based foods for future. Results of consumer and professional expert interviews in five European countries - EIT-Food Project „The V-Place“
Arbeitsbericht Nr. 29

Bozorov, A., Feuerbacher, A., Wieck, C. (2021): A 2014 Social Accounting Matrix (SAM) for Uzbekistan with a Focus on the Agricultural Sector
Arbeitsbericht Nr. 30

Gebhardt, B. (2021): Quo vadis? Ansätze der Qualitätssicherung von Nachhaltigkeitswettbewerben für Unternehmen. Ergebnisse eines Experten-Workshops
Arbeitsbericht Nr. 31

Kareem, O.I., Wieck, C. (2021): Mapping agricultural trade within the ECOWAS: structure and flow of agricultural products, barriers to trade, financing gaps and policy options. A research project in cooperation with GIZ on behalf of BMZ
Arbeitsbericht Nr. 32

Gebhardt, B. (2022): Status Quo und Potentiale des ökologischen, Heil-, Kosmetik- und Gewürzpflanzenanbaus in Baden-Württemberg. Studienbericht & Supplement
Arbeitsbericht Nr. 33

Gebhardt, B., Bermejo, G., Imort-Just, A., Kiefer, L., Zikeli, S., Hess, S. (2023): Zweinutzungshuhn – was ist das? Umfrage unter Landwirt*innen und Geflügelhalter*innen in Deutschland 2022
Arbeitsbericht Nr. 34

Bermejo, G., Imort-Just, A, Gebhardt, B., Hess, S., Kiefer, L., Zikeli, S., (2023): Status-Quo und Perspektiven von Zweinutzungshühnern in Baden-Württemberg: Ergebnisse eines World-Cafés im Rahmen des 1. Dialogforums des Projektes „ZweiWert“ am 2.3.2023.
Arbeitsbericht Nr. 35

Gebhardt, B., Maute, J., Kiefer, L. (2023): Zweinutzungshuhn – wie schmeckt das? Sensorische Beurteilung von Hühnerfleisch und Eiern von vier Zweinut-zungshuhn-Genetiken
Arbeitsbericht Nr. 36

Gebhardt, B., Hellstern, L. (2023): Die Kraft von Awards. Umfrage unter Unternehmen und Vergabeinstitutionen in Deutschland 2023
Arbeitsbericht Nr. 37

Gebhardt, B., Hellstern, L. (2024): SIEGER! Business-Awards als Instrument zur Steuerung der Nachhaltigkeitstransformation – Ansätze für Qualitätssicherung und Schärfung der strategischen Weiterentwicklung
Arbeitsbericht Nr. 38

Gebhardt, B. (2024): Sensorische Beurteilung von Zweinutzungshühnern in der Gemeinschaftsverpflegung. Ergebnisse einer Verkostung des Gerichts „Halbes Brathähnchen“ in drei Kantinen in Baden-Württemberg
Arbeitsbericht Nr. 39

Gebhardt, B. (2024): Nachhaltigkeitsexzellenz in der Landwirtschaft: Mehr Sichtbarkeit für die versteckten Leuchttürme der Alltagspraxis
Arbeitsbericht Nr. 40