COINr: An R package for developing composite indicators

William Becker [1][¶], **Giulio Caperna**[2], **Maria Del Sorbo** [3], **Hedvig Norlén**[1], **Eleni Papadimitriou**[2], **and Michaela Saisana** [2]

1 Freelance consultant, Ispra, Italy **2** European Commission, Joint Research Centre, Italy **3** European Innovation Council and SMEs Executive Agency, Belgium ¶ Corresponding author

Summary

Composite indicators (CIs) are aggregations of indicators that aim to measure complex, multi-dimensional and typically socio-economic concepts such as sustainable development (Conceição, 2020), innovation (Dutta et al., 2020), globalisation (William Becker et al., 2021), gender equality (Equal Measures 2030, 2019), and many more. CIs are very widely used in policy-making, by international organisations, and they are equally well-covered in academic literature (El Gibari et al., 2019; David Lindén et al., 2021; Stefana et al., 2021). They are often used to rank and benchmark countries or regions to help direct policy making, but are also frequently used for advocacy (Cobham et al., 2015).

The *COINr* package, introduced in this article, aims to provide a harmonised development environment for composite indicators that includes all common operations from indicator selection, data treatment and imputation up to aggregation, presentation of results and sensitivity analysis. COINr enables development, visualisation and exploration of methodological variations, and encourages transparency and reproducibility.

Statement of need

Existing tools

Some dedicated tools for composite indicators exist: in Microsoft Excel, the *COIN Tool* is a spreadsheet-based system which allows users to build and analyse a composite indicator (W. Becker et al., 2019). In MATLAB, there are some packages addressing specific parts of index development: the *CIAO* package uses a nonlinear regression and optimisation approach to tune weights to agree with expert opinions (D. Lindén et al., 2021). In R (R Core Team, 2022) there is an existing package for composite indicator development, called *compind* (Fusco et al., 2018), focusing on weighting and aggregation, although this gives no special consideration to hierarchical structures, uncertainty and sensitivity analysis, and so on.

The Python library *CIF* gives a number of tools for building composite indicators, from loading data to aggregation and visualisation (Vraná, 2022). This is focused in particular on Business Cycle Analysis. Finally, there is a recently launched Web-based tool called the *MCDA Index Tool* (Cinelli et al., 2021). This is mostly focused on multi-criteria decision analysis, and doesn't include different levels of aggregation.

Why COINr

COINr is a significant step beyond existing composite indicator tools in many respects. COINr wraps all composite indicator data, analysis and methodological choices into a single S3 class

object called a "coin". A coin is a structured list including:

- Indicator data sets for each processing step (e.g. imputation and normalisation).
- Metadata pertaining to indicators and units (e.g. names and weights, but also the hierarchical structure of the index).
- A record of the COINr functions applied in constructing the coin.

This enables a neat and structured environment, simplifies the syntax of functions, and also allows comparisons between different versions of the same index, as well as global sensitivity analysis along the lines of Saisana et al. (2005) (for the distinction between "local" and "global" sensitivity analysis, see e.g. Saltelli et al. (2019)). COINr also supports time-indexed (panel) data, represented by the "purse" class (a data frame containing a time-indexed collection of coins). For more information on coins and purses, see the coins vignette.

All major COINr functions have methods for coins, and many have methods for purses, data frames, and numerical vectors. This means that COINr can be used either as an integrated development environment via coins and purses, but equally as a toolbox of functions for other related purposes.

COINr also offers a far wider range of functions and methodological options than any existing package. It not only includes a range of options for treating, imputing, normalising and aggregating indicator data (among others), but also has a suite of analysis tools to check data availability and perform correlation/multivariate analysis. Moreover, it has many options for plotting and visualising data using wrapper functions for the *ggplot2* package (Wickham, 2016). Many core COINr functions are written with hooks to link with other packages, for example allowing other imputation or aggregation packages to be used with coins.

Features

Primarily, COINr is used for *building* composite indicators: In practice this involves assembling a set of indicators (usually from different sources) and accompanying metadata, and assembling them into data frames that can be read by COINr to build a "coin" (see vignette). After that, the composite scores are calculated by applying COINr functions to the coin, which specify the methodological steps to apply, and how to apply them.

To give a flavour of COINr, we present a very short example using the built-in "ASEM" data set which comprises two data frames (one of indicator data, and the other of metadata). To build a coin, the new_coin() function is called:

```
# load COINr
library(COINr)
```

```
# build a coin with example data set
coin <- new_coin(iData = ASEM_iData, iMeta = ASEM_iMeta)
```

To see how these data frames are formatted, one can use e.g. str(ASEM_iData) or View(ASEM_iData) and refer to the coins vignette.

In the most simple case, we could build a composite indicator by normalising the indicators (bringing them onto a common scale), and aggregating them (using weighted averages to calculate index scores). This can be done in COINr using the Normalise() and Aggregate() functions respectively:

```
# normalise (scale) each indicator onto [0, 100] interval
coin <- qNormalise(coin, dset = "Raw", f_n = "n_minmax",
                   f_n_para = list(l_u = c(0, 100)))
```

```
# aggregate using weighted arithmetic mean
```

Becker et al. (2022). COINr: An R package for developing composite indicators. *Journal of Open Source Software*, *7*(78), 4567. https: //doi.org/10.21105/joss.04567.

```
# Note: weights are input in data frames when calling new_coin()
coin <- Aggregate(coin, dset = "Normalised", f_ag = "a_amean")
```

Both of these functions allow any other function to be passed to them, allowing more complex types of normalisation and aggregation. Here, the code simply uses the "min-max" normalisation method (scaling indicators onto the $[0, 100]$ interval), and aggregates using the weighted arithmetic mean, following the hierarchical structure and weights specified in the iMeta argument of new_coin().

To see the results in a table form, one can call the get_results() function:

```
# generate data frame with results at highest aggregation level (index)
get_results(coin, dset = "Aggregated") |> head()
```

```
  uCode Index Rank
1   DEU 75.23    1
2   GBR 68.94    2
3   FRA 65.92    3
4   CHE 62.61    4
5   NLD 61.24    5
6   SWE 60.59    6
```

We may also visualise the same results using a bar chart - here we see how countries rank on the "connectivity" sub-index (see Figure 1).
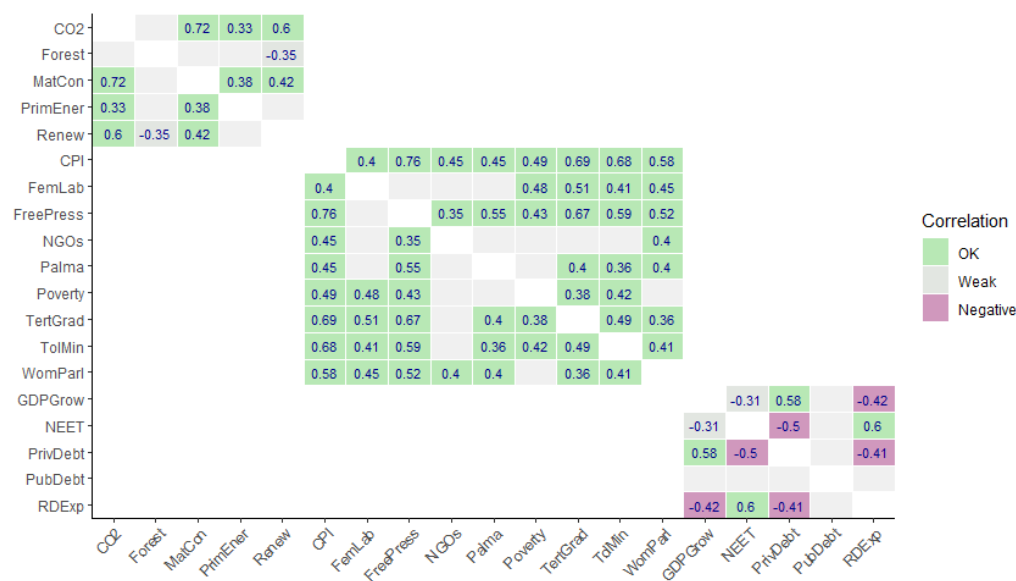
```
plot_bar(coin, dset = "Aggregated", iCode = "Conn", stack_children = TRUE)
```
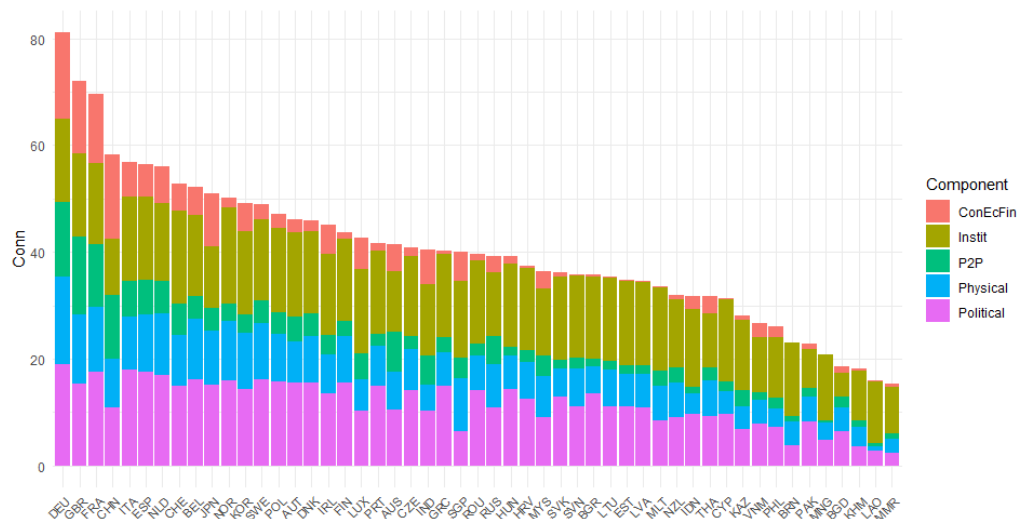

Figure 1: National connectivity scores broken down into component scores and sorted from highest to lowest.

As a final example, we show one of the analysis features of COINr: the possibility to plot and analyse correlations.

```
plot_corr(coin, dset = "Normalised", iCodes = list("Sust"),
          grouplev = 2, flagcolours = T, text_colour = "darkblue")
```

Becker et al. (2022). COINr: An R package for developing composite indicators. *Journal of Open Source Software*, *7*(78), 4567. https://doi.org/10.21105/joss.04567.

Figure 2: Correlations between sustainability indicators, with colouring thresholds. Only correlations within aggregation groups are shown.

The correlation plot in Figure 2 illustrates where e.g. negative correlations exist within aggregation groups, which may lead to poor representation of indicators in the aggregated scores.

COINr includes far more features than those shown here. Remaining features (with vignette links) include:

Building:

- Denomination by other indicators
- Screening units by data requirements
- Imputation of missing data
- Outlier treatment using Winsorisation and nonlinear transformations
- Weighting using either manual weighting, PCA weights or correlation-optimised weights.

Analysis:

- Analysis via indicator statistics, data availability, correlation analysis and multivariate analysis (e.g. PCA).
- Adjustments and Comparisons: checking the effects of methodological variations.
- Global uncertainty and sensitivity analysis of the impacts of uncertainties in weighting and many methodological choices

Others:

- A range of visualisation options, including statistical plots, bar charts and correlation plots
- Automatic import from the COIN Tool and fast export to Microsoft Excel.

For the full range of COINr features, see COINr documentation which is accessible at COINr's website.

Acknowledgements

References

Becker, W., Benavente, D., Dominguez Torreiro, M., Tacao Moura, C., Neves, A. R., Saisana, M., & Vertesy, D. (2019). *COIN Tool User Guide*. European Commission, Joint Research Centre. https://doi.org/10.2760/523877

Becker, William, Domínguez-Torreiro, M., Neves, A. R., Moura, C. T., & Saisana, M. (2021). Exploring the link between Asia and Europe connectivity and sustainable development. *Research in Globalization*, *3*, 100045. https://doi.org/10.1016/j.resglo.2021.100045

Cinelli, M., Spada, M., Kim, W., Zhang, Y., & Burgherr, P. (2021). MCDA index tool: An interactive software to develop indices and rankings. *Environment Systems and Decisions*, *41*(1), 82–109. https://doi.org/10.1007/s10669-020-09784-x

Cobham, A., Janskỳ, P., & Meinzer, M. (2015). The Financial Secrecy Index: Shedding new light on the geography of secrecy. *Economic Geography*, *91*(3), 281–303. https://doi.org/10.1111/ecge.12094

Conceição, P. (2020). *The 2020 Human Development Report*. United Nations Development Programme. ISBN: 978-92-1-126442-5

Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (2020). *The Global Innovation Index 2020: Who will finance innovation?* World Intellectual Property Organisation. https://www.globalinnovationindex.org/gii-2020-report

El Gibari, S., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: A review. *Journal of Business Economics*, *89*(1), 1–24. https://doi.org/10.1007/s11573-018-0902-z

Equal Measures 2030. (2019). *Harnessing the power of data for gender equality: Introducing the 2019 EM2030 SDG gender index*. Equal Measures 2030. https://www.data.em2030.org/2019-global-report

Fusco, E., Vidoli, F., & Sahoo, B. K. (2018). Spatial heterogeneity in composite indicator: A methodological proposal. *Omega*, *77*, 1–14. https://doi.org/10.1016/j.omega.2017.04.007

Lindén, D., Cinelli, M., Spada, M., Becker, W., & Burgherr, P. (2021). *Composite Indicator Analysis and Optimization (CIAO) tool, v.2*. https://doi.org/10.13140/RG.2.2.14408.75520

Lindén, David, Cinelli, M., Spada, M., Becker, W., Gasser, P., & Burgherr, P. (2021). A framework based on statistical analysis and stakeholders' preferences to inform weighting in composite indicators. *Environmental Modelling & Software*, 105208. https://doi.org/10.1016/j.envsoft.2021.105208

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 307–323. https://doi.org/10.1111/j.1467-985x.2005.00350.x

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic

review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

Stefana, E., Marciano, F., Rossi, D., Cocca, P., & Tomasoni, G. (2021). Composite indicators to measure quality of working life in Europe: A systematic review. *Social Indicators Research*, *157*(3), 1047–1078. https://doi.org/10.1007/s11205-021-02688-6

Vraná, L. (2022). *Composite Indicators Framework (CIF) for business cycle analysis*. Python Software Foundation. https://pypi.org/project/cif/

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4