CLC Estimator: a tool for latent construct estimation via congeneric approaches in survey research

Giacomo Marzi
IMT School for Advanced Studies Lucca
Piazza San Francesco 19, 55100, Lucca, Italy
giacomo.marzi@imtlucca.it
ORCID: https://orcid.org/0000-0002-8769-2462

Marco Balzano
Department of Management
Ca' Foscari University of Venice
873 Cannaregio, S. Giobbe, 30121 Venice, Italy
marco.balzano@unive.it
&
KTO Research Center
SKEMA Business School, Sophia Antipolis
ORCID: https://orcid.org/0000-0002-2452-631X

Leonardo Egidi
Department of Economics, Business, Mathematics, and Statistics "Bruno de Finetti"
University of Trieste, Trieste, Italy
legidi@units.it
ORCID: http://orcid.org/0000-0003-3211-905X

Alessandro Magrini
Department of Statistics, Computer Science, Applications "G. Parenti",
University of Florence, Florence, Italy
alessandro.magrini@unifi.it
ORCID: http://orcid.org/0000-0002-7278-5332

Published in Multivariate Behavioral Research

Published Version (DOI):

https://doi.org/10.1080/00273171.2023.2193718

To cite this article:
Marzi, G., Balzano, M., Egidi, L., Magrini, A. (2023). CLC Estimator: A Tool for Latent Construct Estimation via Congeneric Approaches in Survey Research. Multivariate Behavioral Research. DOI: 10.1080/00273171.2023.2193718

CLC Estimator: a tool for latent construct estimation via congeneric approaches in survey research

The present paper proposes the Shiny app 'CLC Estimator' –Congeneric Latent Construct Estimator– to address the problem of estimating latent unidimensional constructs via congeneric approaches. While congeneric approaches provide more rigorous results than suboptimal parallel-based scoring methods, most statistical packages do not provide easy access to congeneric approaches. To address this issue, the CLC Estimator allows social scientists to use congeneric approaches to estimate latent unidimensional constructs smoothly. The present app provides a novel solution to the challenge of limited access to congeneric estimation methods in survey research.

Keywords: latent construct estimation; congeneric approaches; Shiny app; scale scores; social sciences.

Introduction

This study addresses the limitations of congeneric approaches to estimating unidimensional latent constructs. Specifically, the study aims to provide users with an accessible statistical tool for estimating unidimensional latent constructs, even if they have limited familiarity with commonly available statistical software.

The rationale for this study is grounded on the importance of latent constructs in predicting behavioural patterns. Indeed, social scientists rely on latent constructs to measure abstract entities such as personality traits, leadership attitudes, and organisational characteristics.

Abstract entities are conceptualised in the form of latent constructs, which often require the adoption of multiple items to estimate them (Graham, 2006; McNeish and Wolf, 2020). Let ξ be the latent construct and $X_1, \ldots, X_p$ be p items. A factor model is usually defined as:

$$X_i = \mu_i + \lambda_i \xi + \varepsilon_i \qquad i = 1, \ldots, p$$

$$E(\varepsilon_i) = 0 \quad Var(\varepsilon_i) = \psi_i$$

$$E(\xi) = 0 \quad Var(\xi) = 1$$

where μ_i , λ_i and ε_i are, respectively, the mean, the (unstandardised) loading and the measurement error of the item X_i. Each item is expressed as a linear function of the latent construct plus a measurement error in this definition. Therefore, the latent construct is conceived as a common causal structure underlying the items. Consequently, changes in the construct are reflected by changes in the items, and the items are expected to be highly correlated and interchangeable: dropping an item should not alter the conceptual meaning of the construct.

The estimation of latent constructs based on multiple items follows two main approaches: parallel and congeneric (Jöreskog, 1971). In parallel approaches, items have the same loading and the same error variance, i.e., $\lambda_i = \lambda \wedge \psi_i = \psi \ \forall i$ (parallel assumption). Therefore each item contributes equally to the latent construct. The most popular parallel approaches include sum and average scores, which define the latent construct as the raw sum or average of item scores. In congeneric approaches, each item has a unique loading and a unique error variance (Millsap and Everson, 1991; Graham, 2006; McNeish and Wolf, 2020). Thus, the higher the correlation between an item and the latent construct, the higher the loading. For example, if I1 is more closely related to the latent variable than I2, then I1 will have a greater weight than I2 in the calculation of the latent construct. This implies that, unlike parallel approaches, the measurement of each item contributes differently to the latent construct.

McNeish and Wolf (2020) argue that parallel approaches often lack theoretical justification and supporting evidence, and scholars often rely on them because of their computational immediacy. In addition, researchers should maintain the same approach to estimating the latent constructs when adopting previously validated scales. Otherwise, the results may be inconsistent with the theoretical and methodological assumptions of the original validation process (McNeish and Wolf, 2020). As scales are usually validated using congeneric approaches, our app is an appropriate tool for estimating latent constructs in survey research when researchers adopt previously validated measures.

To address the above problem, we developed a user-friendly Shiny app that allows social scientists to estimate latent constructs based on congeneric approaches. The app can estimate a unidimensional latent construct based on several different methods, given a data file provided by the user, including item measurements. Our app is helpful when a statistical package requires programming expertise to perform the congeneric estimation. For example, congeneric-based estimates are typically not ready for use in software that performs regression analysis, QCA, ANOVA, MANOVA, and other types of analysis, making it difficult for users unfamiliar with statistical environments to obtain a proper estimate of congeneric latent constructs.

The remainder of the article is structured as follows: in the next section, we present our Shiny app. Next, we report a question-and-answer section to provide users with additional guidance; Section 4 includes concluding remarks.

Presenting the Shiny app

The CLC Estimator Shiny app is written in R for Statistical Computing (R Core Team, 2022) based on the packages 'shiny' (Chang et al., 2022), 'psych' (Revelle, 2022) and 'mirt' (Chalmers et al., 2022).

The app is available at: https://plsdeams.shinyapps.io/CLC_Estimator/.

When users run the app online, it is standalone and requires only an updated web browser. The app is stored on shinyapps.io, which ensures compliance with data security and regulatory standards for sensitive data in various jurisdictions.[1] The CLC Estimator Shiny app does not store any data during the calculation process and deletes all the data when the user closes the app. For archiving purposes, the source code (updated at the acceptance date of the present paper) is available on FigShare at: https://doi.org/10.6084/m9.figshare.c.6361907.v2. For computational use, users are encouraged to use the latest version of the app available at the aforementioned link or the latest version of the source code, available on GitHub at: https://github.com/LeoEgidi/clc

To run the app offline, please enter the following code in the R console[2]:

```
install.packages('devtools')  #install devtools package

library('devtools')  #load devtools package

install_github('leoegidi/clc')  #install clc package from GitHub

library('clc')  #load clc package

clc()  #load clc estimator
```

Moving to the presentation of the CLC Estimator Shiny app, Figure 1 shows its graphical interface.

[1] The hosted solution provides users with an extra layer of security when using the CLC estimator Shiny app online.
[2] "devtools" requires Rtools installed that can be downloaded from: https://cran.r-project.org/bin/windows/Rtools/

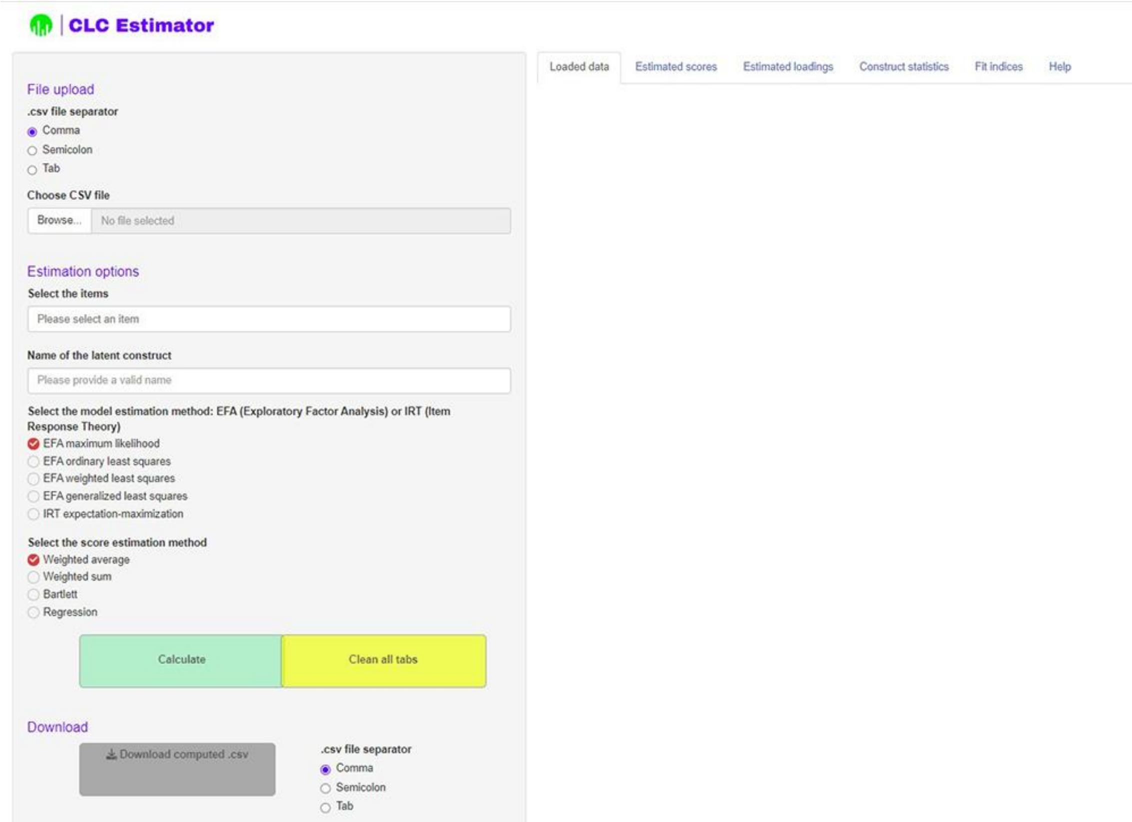

Figure 1 - Interface of the CLC Estimator Shiny app

The interface of the app has been designed according to the principles of simplicity and user-friendliness. As shown in Figure 1, the interface of the app consists of a single page that allows performing all the steps necessary for calculating the latent construct and presenting the results.

Users interact with the left panel of the app, starting at the top of the panel (loading the .csv file), gradually moving to the bottom of the panel while selecting the desired options, and ending with the computed .csv file ready for download.

Specifically, the user starts by loading the .csv file with raw data and selecting the correct delimiter for the values. Once the file is loaded into the app, the items to be used for estimating the latent construct are displayed in the 'Loaded Data' panel.

Next, users are asked to indicate which items will be used to estimate the latent construct. Users enter the name of the latent construct to be estimated, select the model estimation and scoring method, and complete the process by pressing the green 'Calculate' button.

The Shiny app allows for two alternative factor models: Exploratory Factor Analysis (EFA) (Grice, 2001; Thompson, 2004) and Item Response Theory (IRT) (Bock & Aitkin, 1981; Samejima, 1969). Model estimation methods are available for the EFA model: Maximum Likelihood (ML), Ordinary Least Squares (OLS), Weighted Least Squares (WLS) and Generalised Least Squares (GLS) as implemented in the 'psych' package (Revelle, 2022). For the graded IRT model, the estimation method is maximum likelihood using the EM algorithm implemented in the 'mirt' package (Chalmers et al., 2022).

The scoring methods available for both EFA and IRT models include weighted sum and weighted average, while for the EFA model, it is possible to select either regression or Bartlett's scoring (Grice, 2001), and for the graded IRT model, one of Maximum a Posteriori (MAP), Expectation a Posteriori (EAP) and ML (Embretson & Reise, 2000). Unweighted average and unweighted sum are also available in the app to explore the properties of parallel and congeneric models.

Missing data are handled based on the standard procedures of the R packages 'psych' and 'mirt': for the EFA model, the correlation matrix is computed based on pairwise complete observations, while for the graded IRT model, the full information maximum likelihood method is used. The CLC Estimator immediately warns the user if the selected items contain missing values.

Next, the results of the estimation are displayed in the 'Estimated scores' panel, while the 'Estimated loadings' panel shows the estimated loadings, and the 'Construct statistics' and 'Fit indices' panels provide summary statistics and fit indices of the estimated construct. Several different constructs can be computed consecutively from the same dataset, and each time the estimated scores are added to the previous ones in the 'Estimated scores' panel. Users can press the yellow 'Clear all tabs' button to clear the panels and move on to calculating a new construct.

Finally, the estimated scores and raw data can be downloaded in a .csv file by selecting the .csv delimiter and pressing the "Download computed .csv" button.

The 'Help' panel of the app refers to this article and other useful information for users on how to use the Shiny app, as well as key questions and answers to facilitate the use of the CLC Estimator.

Overall, the app is intended to be an accessible tool for estimating latent constructs via congeneric approaches, while any other type of assumption or data preparation procedure will need to be performed by users in complementary statistical software.

Additional guidance on using the CLC Estimator:

In the present section we present key questions and their answers to facilitate the correct use of the CLC Estimator.

- *What can the CLC Estimator do?* The app only estimates unidimensional latent constructs based on congeneric approaches. CLC Estimator is not intended for statistical procedures commonly available in existing statistical packages, such as exploratory factor analysis, principal component analysis or confirmatory factor analysis.
- *When should the CLC Estimator be used?* CLC Estimator performs latent construct estimation via congeneric approaches when the available statistical packages do not include this function.
- *At what point in the data analysis process the CLC Estimator should be used?* CLC Estimator should be integrated into the data analysis process after review (including reliability analysis) of the items to be retained in the latent construct estimation.

- *When the CLC Estimator should not be used?* CLC Estimator assumes a common factor model, i.e. a reflective latent construct. Therefore, it is not appropriate when the latent construct is a composite (Rhemtulla et al., 2020).
- *How to deal with missing data, outliers, and other data issues in the CLC Estimator?* CLC Estimator does not handle missing data, outliers or incomplete data. Users are encouraged to address these issues before using the CLC Estimator. Specifically for missing data, note that the CLC Estimator uses the standard missing data handling procedures embedded in the 'psych' and 'mirt' R packages. For the EFA model, the correlation matrix is computed based on pairwise complete observations. For the graded IRT model, full information maximum likelihood is used. If the data contain missing values, the CLC Estimator warns the user immediately.
- *How to create a usable .csv file for the CLC Estimator?* CLC Estimator supports standard .csv files generated by popular statistical packages. The first row of the .csv file should contain the labels of the variables (i.e. the header), while the other rows should contain the data. The .csv file loading interface allows the user to select the delimiter between values (comma, semicolon, tab). The .csv file should be formatted in UNICODE, with decimal symbols expressed in points according to international scientific conventions. A sample dataset in .csv format (comma delimiter) is available at: https://dx.doi.org/10.6084/m9.figshare.21786335
- *What type of model estimation method should be used to estimate the latent constructs?* When social scientists use previously validated scales, the same method of estimation should be used for those scales. If neither the estimation method is specified by the original developers of the scales nor such scales have

been previously validated, maximum likelihood estimation (MLE) is recommended (for a detailed discussion of this topic, see Thompson, 2004).

- *What type of output does the CLC Estimator produce?* CLC Estimator generates a standard .csv file with variable names in the first row and values in the remaining rows. The generated .csv file output can be easily imported into popular statistical packages.

Conclusions

In the present article, we propose a user-friendly Shiny app for estimating unidimensional latent constructs using congeneric approaches. The app allows advanced and basic users to easily estimate and visualise latent constructs. Although many studies rely on unidimensional latent constructs, we plan to extend our app to estimate multidimensional latent constructs in future releases. Missing data imputation will be another future addition to the application.

References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4): 443-459. https://doi.org/10.1007/BF02293801

Chalmers P., Pritikin J., Robitzsch A., Zoltak M., Kim K., Falk C.F., Meade A., Schneider L., King D., Liu C.W., Oguzhan O. (2022). *mirt: Multidimensional Item Response Theory*. R package version 1.37.1. https://CRAN.R-project.org/package=mirt

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2022). shiny: Web Application Framework for R. R package version 1.7.4. https://CRAN.R-project.org/package=shiny

Embretson, S. E. & Reise, S. P. (2000). Item Response Theory for Psychologists. Lawrence Erlbaum Associates Publishers.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, 66(6), 930–944. https://doi.org/10.1177/0013164406288165

Grice, J. W. (2001), Computing and evaluating factor scores. *Psychological Methods*, 6, 430-450. https://doi.org/10.1037/1082-989X.6.4.430

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109-133. https://doi.org/10.1007/BF02291393

McNeish, D.M., & Wolf, M.G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52(6), 2287-2305. https://doi.org/10.3758/s13428-020-01398-0

Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, *26*(3), 479-497. https://doi.org/10.1207/s15327906mbr2603_6

Pek J., Hoyle R. H. (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass*, 10, 150–163. https://doi.org/10.1111/spc3.12237

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

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 2.2.9, https://CRAN.R-project.org/package=psych

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, 25(1), 30–45. https://doi.org/10.1037/met0000220

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. American Psychological Association. https://doi.org/10.1037/10694-000