MULTIPOL-C

An R program for fitting unipolar log-logistic IRT multidimensional models to continuous item responses

USER'S GUIDE

Prepared by:

David Navarro-González

Pere J. Ferrando

Fabia Morales-Vives

José M. Casas

Please reference this document as:

Navarro-Gonzalez, D., Ferrando, P.J., Morales-Vives, F., & Casas, J.M. (2025). *MULTIPOL-C User's Guide. Technical Report*. Universitat Rovira i Virgili. Tarragona.

Available at: https://psico.fcep.urv.cat/utilitats/MULTIPOL-C/

Contents

-1. Theoretical Bases and features

In non-cognitive (e.g. personality and attitude) measurement, the multidimensional version of Samejima's (1974) continuous response model (CRM; see also Bejar, 1977, and Wang & Zeng, 1998), is a well-known Item Response Theory (IRT) model that is expected to have increasing relevance in the future due to the also growing interest in continuous item response formats (e.g. Huang, 2024). Explicitly or implicitly, the CRM is intended for traits that can be viewed as bipolar dimensions (and so, that are equally meaningful at both ends of the continuum; see e.g. Morales-Vives et al., 2023) and that are (at least approximately) normally distributed in the population of interest. For most non-cognitive traits, particularly normal-range personality traits, these assumptions seem quite reasonable (extraversion-introversion is a paradigmatic example). However, certain traits such as clinical or psychiatric traits, addictions, extreme beliefs or maladaptive personality traits can be more plausibly modeled as unipolar dimensions (Lucke, 2013, 2015; Reise & Waller, 2009). So, the low end of the dimension merely reflects the absence of trait manifestations while the high end reflects the different levels of trait severity or extremeness. Furthermore, it is reasonable to scale this type of dimensions as adopting only positive values (Lucke, 2013) and to assume that they are more informative and meaningful at its upper end. Finally, when traits of this type are measured in community populations, an expected result is that the items and test scores have extreme (rightly skewed) distributions, because most of the individuals are expected to have low trait levels and be piled-up at the lower end (e.g. Morales-Vives et al., 2023). If this is so, the latent trait distribution is expected to have, intrinsically, a low mean and high variance both leading to a pronounced right skewness (e.g. Magnus & Liu, 2018; Morales-Vives et al., 2023).

In the unidimensional case, a unipolar transformed version of the CRM can be derived by assuming the latent trait distribution to be lognormal (0,1). This distribution is rightly skewed and intended for a dimension that adopts only positive values. The model that results from this transformation is the Log-Logistic CRM (LL-CRM; Ferrando et al. 2024). Please, consult UNIPOL-GC-Guide for a description of unidimensional LL models.

A multidimensional extension of the LL-CRM (Md-LL-CRM has been recently proposed by Ferrando et al. (submitted). From a substantive and practical view, this extension has a clear interest: a review of psychometric developments in the domains in which log-logistic models can be appropriate, shows that many of them use a multidimensional framework (e.g. Reise & Waller, 2009). The present guide describes the implementation of this model in a program named MULTIPOL-C (for MULTIdimensional uniPOLar-Continuous).

-1.1 Description of the Md- LL-CRM: main features

In this section we shall only provide a conceptual summary of the model implemented in the program. More technical presentations will be available in Ferrando et al. (submitted) to which the interested reader is referred.

We shall start by first describing the estimation and fitting procedures used in MULTIPOL-C. They are based on a two-stage: Calibration and Scoring (e.g. Mislevy & Bock, 1990) approach. In the calibration stage, the parameters of the items are estimated and model-data fit at the structural level is assessed. In the scoring stage, provided that model-data fit is judged to be acceptable, the item parameter estimates are taken as fixed and now, and used to obtain individual score estimates and accompanying measures of score accuracy. This summary will be used for describing the main features of the Md-LL-CRM in terms of both, (a) item parameters and functioning (calibration stage), and (b) individual score estimates and measures of score accuracy (scoring stage).

Let's start by considering first the unidimensional LL-CRM. Suppose that the item scores can be treated as (approximately) continuous, and, that, for practical and interpretative purposes, they are scaled to have values between 0 and 1. With this scaling, for a fixed trait level θ_U (the *U* stands for unipolar), the expected score in item *j* is given by:

$$E(X_j \mid \theta_U) = \frac{\alpha_j \theta_U^{\beta_j}}{1 + \alpha_j \theta_U^{\beta_j}} . \qquad (1)$$

The trait θ_U is assumed to follow a lognormal distribution with parameters μ_U=0 and σ_U=1. So: (a) θ_U is anchored to zero and has no upper limit, and (b) $ln(\theta_U)$ is normally distributed with zero mean and unit variance. Now, as a function of θ_U, the conditional expectation (1) is the Item Response Function (IRF) of the LL-C model, and is governed by two item parameters: α_j and and β_j. Both are restricted to have positive values, and are "easiness" and form-curvature parameters, respectively. More in detail, the higher α_j is, the higher the expected score for item *j* becomes (so the interpretation as easiness is clear). As for the β_j's, at low values the expected IRF is flatter at almost all trait levels, and, the higher the β_j value becomes, the more the IRF increases at low trait levels.

A useful auxiliary item location parameter can be defined from (1) as:

$$\delta_j = \left(\frac{1}{\alpha_j} \right)^{\frac{1}{\beta_j}}. \qquad (2)$$

The δ_j location parameter, which again is always positive, can be interpreted as the θ_U trait level at which the expected item score is 0.5 (i.e. the midpoint response scale), which, conceptually, is the response scale value that marks the transition from a tendency to disagree with the item to a tendency to agree with it (see Ferrando, 2009).

The IRFs defined by equations (1) and (2) are graphically illustrated below by using two hypothetical items.

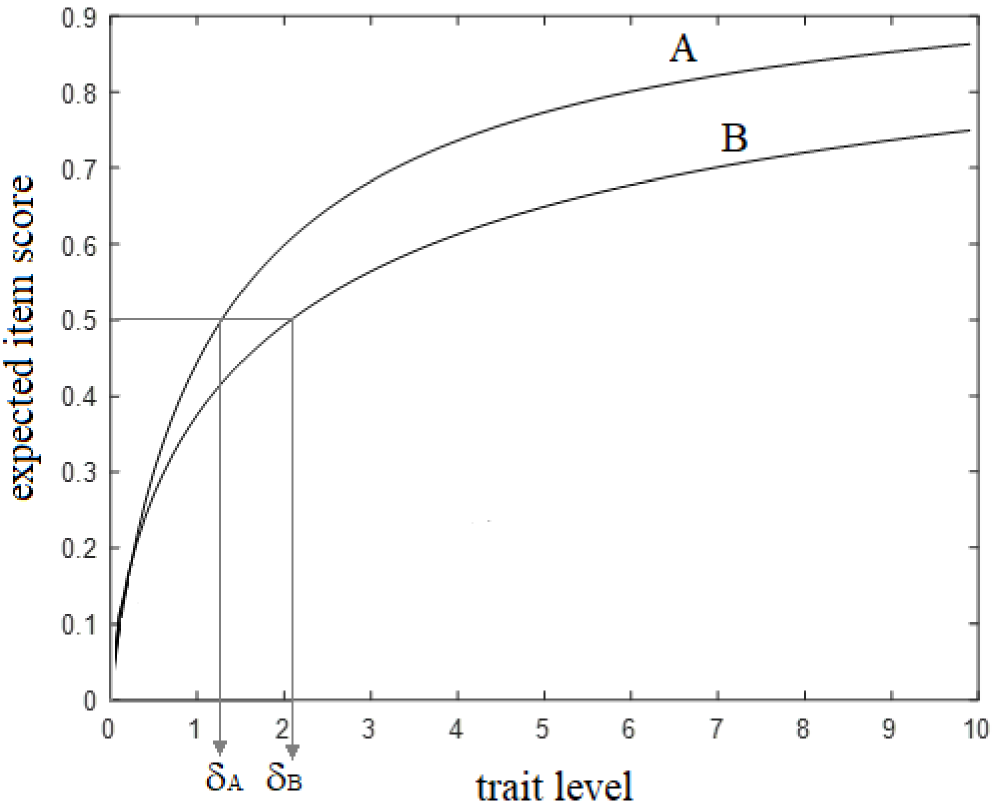


Figure 1. Item response functions (curves) in the LL-CRM

To start with, note that the IRFs in figure 1 clearly depart from the usual CRM (or GRM) ogives. Here, each curve is a power function (e.g. Stevens, 1975), and its general trend is that the curve is concave downward, and its slope tends to increase more strongly for trait values close to zero and flattens as θ_U increases. Conceptually, this trend implies that the item score becomes progressively less sensitive to the trait level as this level increases. Note also, that unlike of what occurs in the standard CRM (and standard IRT models in general) the slope of the IRF is not maximal when the expected score is 0.5 (i.e. when the $\theta_U = \delta_j$). Rather, the slope increases without bound as θ_U approaches zero. Specifically, the slope at a given θ_U level is:

$$Slope(\theta_U) = (\frac{\beta_j}{\theta_U})E(X_j|\theta_U)(1 - E(X_j|\theta_U)) \quad (3)$$

The approach we propose for appraising the discriminating power of an LL-C item is based on (3). We interpret the β_j parameter as an overall item discrimination index, because, other things constant, the higher the β_j value, the higher the slope (3) is. Next, as a summary of the discriminating power the item has at different levels, we also report "quartile" slopes at the trait levels at which the expected item scores are 0.25, 0.50 (i.e. $\theta_U = \delta_j$) and 0.75 (see Figure 1 to graphically understand this approach).

The LL-CRM is described in more detail in UNIPOL-GC-Guide. We shall now generalize the unidimensional results so far to the Md-LL-CRM. To do so, consider that the test under study measures now a set of r unipolar traits: $\boldsymbol{\theta}_\mathbf{U} = \theta_{U1}$,.. θ_{Uk} … θ_{Ur} each of them distributed as lognormal $(0,1)$. The expected item score for fixed $\boldsymbol{\theta}_\mathbf{U}$, is:

$$E\left(X_j\middle|\boldsymbol{\theta}_\mathbf{U}\right) = \frac{\alpha_j {\theta_{U1}}^{\beta_{j1}} \ldots {\theta_{Ur}}^{\beta_{jr}}}{1+\alpha_j {\theta_{U1}}^{\beta_{j1}} \ldots {\theta_{Ur}}^{\beta_{jr}}}. \quad (4)$$

And describes not an Item Response Function (IRF) but a (r-dimensional) Item Response Surface (IRS). The location and form of this IRS is governed by a single parameter α_j (as in the unidimensional case) and by a set of r β_{jk} parameters respectively. To better understand the shape of the IRS defined by (3) we shall consider an example obtained from a typical item in the simplest bidimensional case.

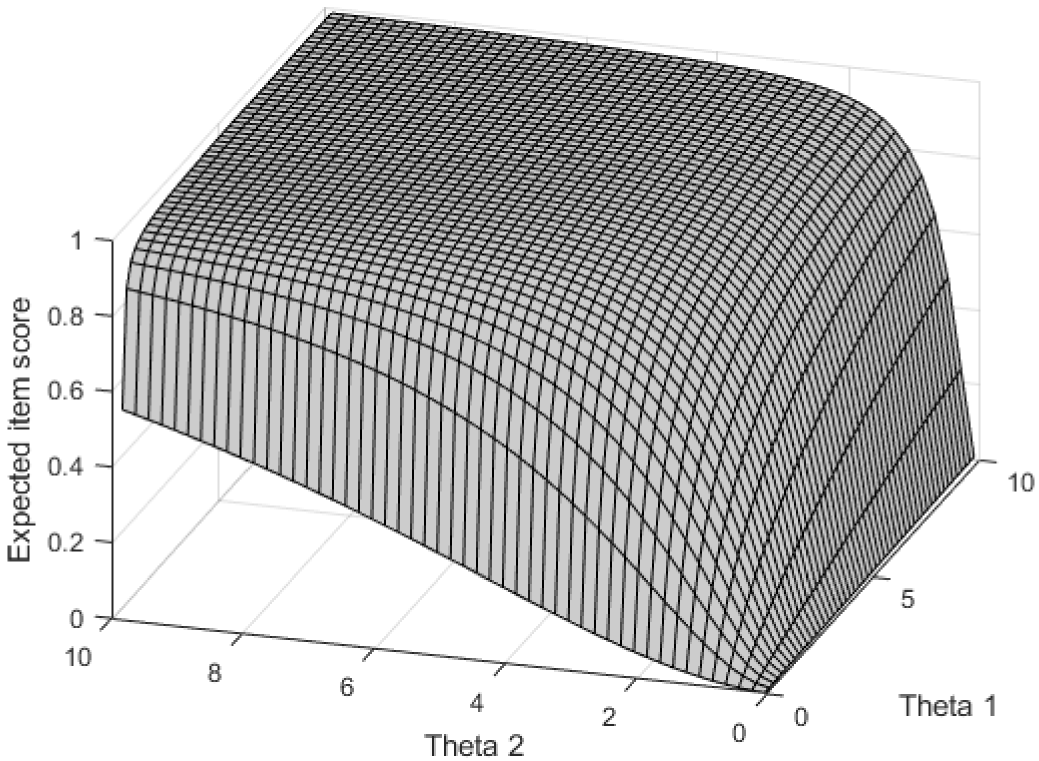


Figure 2. Item response surface of an Md- LL-CRM item

Clearly, the surface in figure 2 is a natural generalization of the unidimensional IRFs in Figure 1. At low trait levels, the slope of the surface strongly increases in any direction on the Theta1-Theta2 plane, and, as the levels on both traits increase it flattens.

If the items in Md- LL-CRM were "clean" in the factor analytic sense (see below), then, each item would only have a substantial β_{jk} value in one of the traits, and the remaining β_{jk} values would be zero. In this case, the item surfaces in Figure 2 would reduce to the item curves in Figure 1(as it should be) and the measurement properties of the items could be assessed in the same way as in the unidimensional case. For the more general case in which the items are "complex" (i.e. more than one β_{jk} parameter is substantially different from zero), and based on Reckase et al. (1985) we propose the following multidimensional extensions for a better understanding of the measurement properties of the items

Consider again the IRS in figure 1. The starting point is to determine the direction on the Theta1-Theta2 plane along which the slope of the IRS is maximal. Once determined, a multidimensional item difficulty index, denoted by Md-δ_j, is defined as the distance, along the direction of maximal slope, from the origin to the point at which the expected item score is 0.5. The mathematical expression for Md-δ_j, is complex, but the general structure and interpretation are similar to those based on expression (2).

A multidimensional discrimination index can next be defined by considering the slope of the IRS along the direction at which this slope is maximal. The result is simpler here, and is a direct extension of (3). If we denote by θ_{Uc} the trait composite along the direction of maximal slope, then, the slope at a given θ_{Uc} value is:

$$Slope(\theta_{Uc}) = \left(\frac{\sum_k \beta_{jk}}{\theta_{Uc}}\right) E\left(X_j|\theta_{Uc}\right)\left(1 - E\left(X_j|\theta_{Uc}\right)\right). \tag{5}$$

As a natural extension of the unidimensional proposal, we shall consider the sum of the β_{jk} weights as an overall measure of (multidimensional) item discrimination, and report quartile slopes along θ_{Uc} at which the expected item scores are 0.25, 0.50 (i.e. $\theta_U = \delta_j$) and 0.75. Overall, the item indices we propose for appraising the functioning of an item in the Md- LL-CRM are: (a) the easiness parameter α_j ; (b) the multidimensional item difficulty index Md-δ_j; (c) the r β_{jk} parameters; (d) the overall item discrimination parameter (the sum of the r β_{jk} parameters), (e) the slope (5) reported at the three quartiles above, and (f) the projections of the θ_{Uc} composite on the θ_{Uk} axes (see figure 2).

Guidelines for interpreting the reported indices above shall be now provided. The easiness α_j parameter is closely related to the expected item score in the 0-1 scale and so, it provides essentially the same information. Values of $\alpha_j > 1$ point out that the expected item score is above the threshold response scale point 0.5, and so, that the item is "easy". The items suitable for LL modeling are generally "difficult" (see above), and so, the expected α_j values when fitting the Md- LL-CRM will generally be smaller than 1. To provide some references, when $\alpha_j = 0.5$, the expected item score is about 0.35, and when $\alpha_j = 0.25$, the expected item score is about 0.20, which is rather "difficult" (severe or extreme if appropriate).

With regards to the multidimensional item difficulty Md-δ_j, we shall start by noting that the lognormal (0,1) distribution has a mean of 1.65, a median of 1, and a standard deviation of 2.16 (e.g. Aitchison & Brown, 1957), and also that the 1.65 mean is already the 69th percentile. The reference values are the same as those in the unidimensional case (see the UNIPOL-GC-Guide), values below 1 would be interpreted as that the item is "easy" in the sense that it only takes a low trait level to reach the threshold in the item response scale. Values at the 1.65 mean would already be interpreted as that the item is difficult, and values greater than, say, one standard deviation above the mean (3.5-4 or more) would be interpreted as that the item is extremely difficult (please, take these recommendations with a grain of salt). We shall further note that the easiness parameter (or the expected item score) and the multidimensional item difficulty are not redundant, but complementary indices of location. Thus, the easiness parameter/expected score is a general or marginal index of difficulty that does not depend on any particular direction on the Theta$_{jk}$ space, and, furthermore, this difficulty is expressed in the response scale metric (i.e. 0-1). In contrast, Md-δ_j, expresses item difficulty in a particular direction: the "optimal" direction of maximal slope. Furthermore, the index is expressed in the trait (or trait composite) metric: lognormal (0,1).

We turn now to the measures of item discriminating power. The sum of the β_{jk} parameters is interpreted as an overall item discrimination index along the direction in which this discrimination is maximal. As for reference values, they are again the same as those in the unidimensional case (see the UNIPOL-GC-Guide). Tentatively, we can consider values about between 0.3 and 0.7 to be normal range while values above 1 would already indicate a high level of overall discriminating power. However, these references are provisional, and more evidence is needed. As for the slope values reported across the "optimal" direction, they do not have so much a numerical interpretation but rather a trend interpretation: to appraise how the slope changes as the level in the trait composite increases, and at what point the slope becomes virtually flat.

The projections of the θ_{Uc} composite on the θ_{Uk} axes, finally, indicate to us which the optimal directions of functioning for each of the items are. This information is important if item comparisons in terms of difficulty or discriminating power are to be made. In both cases, for the items to be meaningfully compared, the items must have (at least approximately) the same direction (see Reckase et al., 1985).

The same as it occurs in the unidimensional case, by making appropriate (exponential) transformations of the person ($\boldsymbol{\theta}_U = \theta_{U1}$,.. θ_{Uk} …θ_{Ur}) and the item easiness parameters (α_j), The IRS's in (3) can be transformed to the corresponding surfaces in the standard Multidimensional CRM (Md-CRM; Bejar, 1977, Samejima, 1974). We note that the β_{jk} parameters do not need to be transformed as they are the same in both models. Technical details on this reparameterization are given in Ferrando et al. (submitted), and here we shall only discuss the practical and substantive implications of these results. First, the result that the Md-LL-CRM can be obtained as a reparameterization of the Md-CRM allows us to consider the first of them as a transformation of the second in which the scale for the traits ($\boldsymbol{\theta}_{\mathbf{U}} = \theta_{U1}$,.. θ_{Uk} …θ_{Ur}) is changed (from standard normal to lognormal). Second, at the calibration stage the Md-LL-CRM can be calibrated using the same procedures used in the Md-CRM. And, since the latter can be calibrated using well-established factor-analytic procedures (e.g. Bejar, 1977) so can the present model. However, the second result also implies that, at the calibration stage, both models (standard and LL) are expected to reach the same degree of model-data fit when calibrated in the same dataset. Thus, we have a case of alternative models that fit the data equally well but that are based on different principles and philosophies (see e.g. Samejima, 1996).

We shall now discuss the results concerned with the scoring stage. The outcomes of this stage are: (a) the individual trait point estimates (i.e. individual scores) in each of the *r* traits, (b) the conditional measures of accuracy corresponding to each individual estimate in each trait, and (c) the marginal reliability estimates. Because reliability is a unitless and bounded measure that applied researchers are familiar with, in our proposal we have chosen to use reliability as a measure of score accuracy. So, for each respondent, in addition to the individual score estimate per trait, we shall provide the corresponding conditional reliability estimate.

As discussed above, the score (point) estimates obtained from the Md-LL-CRM are, essentially, non-linear (exponential) transformations of the score estimates that would have been obtained from the standard Md-CRM. So, if the model is fitted with the sole purpose of ranking the individuals according to their trait levels, then it is not really worth using the Md-LL-CRM instead of the Md-CRM, because the rank order would be the same as in both cases.

In terms of conditional reliabilities, however, both models will generally function in diametrically opposite ways. Under the conventional Md-CRM, the items that are appropriate for the Md-LL-CRM would be modelled as "difficult" (recall that the proportions of item endorsements when these models are appropriate are very low). So, in the standard model, the conditional reliability of the scores will be considered to be maximal at high trait levels. Conceptually, these results mean that, for each trait, the test will be considered to be highly appropriate for differentiating between the (few) individuals who have very high levels but will not be sensitive enough to differentiate between individuals at lower levels. On the other hand, under the Md-LL-CRM, and, again for each trait, the conditional reliability curve as a function of the trait level will be decreasing, reaching its highest values at very low trait levels. This result implies that, according to this model, the test will accurately differentiate between those individuals who have no, or virtually no trait manifestations and those who clearly do have them. However, it will not be sensitive enough to make finer differentiations between those with high levels. More details on this result are provided in UNIPOL-GC-Guide.

For each trait, finally, the marginal reliability estimate is obtained as the average of the conditional reliabilities across the trait levels. However, because the conditional reliability curves typically will change considerably over the trait level, the marginal reliability cannot be taken as a good summary of the general accuracy behavior of the scores. So, it is a useful auxiliar index of overall score accuracy but has to be interpreted together with the corresponding conditional reliability functions.

-1.2 Calibration and scoring procedures used in MULTIPOL-C

As advanced above, by using an appropriate "linearizing" transformation, the logit transformation (see Ferrando et al. 2024, 2025), the Md-LL-CRM can be formulated and fitted as factor-analytic (FA) model. The program described approach uses this result and assumes that the user has transformed the 0-1 scaled scores to logit scores, and has fitted the appropriate FA solution (see section 1.3) to the inter-item **covariance** matrix. Provided that she/he is satisfied with the appropriateness of the fitted solution, the only information needed to perform the implemented procedures will be: (a) the data matrix with the item scores in the original 0-1 scaling, (b) the covariance-based factor pattern (the factor loadings or weights are directly the same as the β_{jk} parameters in equation (4)), and (c) the inter-factor correlation matrix. Once this initial FA estimates have been provided, the program will carry out the reparameterization, and provide all the Md-LL-CRM item parameter estimates described in section 1.1. This would complete the calibration stage.

In the scoring stage, the item parameter estimates above are taken as fixed and known (see Mislevy & Bock, 1990) and individual point estimated scores and accompanying reliability estimates (section 1.1) are obtained. The same as in the unidimensional proposals, but with more reason in this case, the chosen scoring procedure is Bayes expected a posteriori (EAP, Bock & Mislevy, 1982) by specifying the prior distribution for each θ_U to be lognormal (0,1).

At this stage, we wanted procedures that were robust and able to provide finite and plausible estimates for all the respondents in the sample under analysis, which led us to choose Bayes expected a posteriori (EAP, Bock & Mislevy, 1982) score estimation, by specifying the prior distribution for each θ_{Uk} to be lognormal (0,1).

Please note that, for the moment, the implementation so far described is aimed at users with a certain level of IRT knowledge and skilled enough to carry out data transformations and fit structural FA solutions to covariance matrices. In the future, we plan to extend the Unipolar programming so that the data could be fully calibrated and scored from the beginning by providing only the raw data matrix as input.

-1.3 Substantive and practical considerations for using the Md-LL-CRM

Most of the conditions for considering the Md-LL-CRM as an appropriate model are the same required in the unidimensional case, and are discussed in in detail in UNIPOL-GC-Guide. Here we shall only specifically discuss the added requirements imposed by the multidimensional generalization. The first important additional restriction of the model is that all the β_{jk} parameters in equation (4) must be positive. Now, if the structural solution that fits well the data was a "clean" solution as described above, which is technically denoted as an independent-cluster solution, this condition could always be achieved by appropriate orientation and (if needed) item reversals (McDonald, 2000). In complex solutions, the positive orientation of all the β_{jk}'s is not always possible, and a previous process of item selection may be required to attain the condition.

In practice, we admit that in most applications a certain degree of complexity is unavoidable (Lucke 2005) and it is for this reason that we have developed the multidimensional indices described above. However, the designer should always strive for attaining a solution as clean as possible if the Md-LL-CRM is the basis model for calibrating and scoring a measurement instrument. An we clearly state that the Md-LL-CRM is not meant to be directly used as an exploratory model in complex datasets. However, if, as mentioned above, a previous process of item selection procedure has to be used for obtaining such clean instrument, this process can be based on standard exploratory FA, because the Md-LL-CRM can be linearized to take the form of an FA model.

-2. Installation and setup

The proposal so far discussed has been implemented as R script called `MULTIPOL`, which has been developed in R Version 4.4.1 and runs with R versions more recent than 3.5.0.

The program is released in a compressed folder, which includes two main R scripts: `MULTIPOL` and `MULTIPOL_GUI`.

The first script, `MULTIPOL`, was designed for advanced R users, where the input values have to be provided through code and the output is also printed in the R console or returned silently. When using this function, the user must source the function before its utilization:

```
>source("MULTIPOL.R")
```

The second script, `MULTIPOL_GUI`, is a shiny app (Chang et al. 2024), which provides a Graphical User Interface version of the script, more suitable for users less familiar with R language. It requires shiny package to be installed

Finally, a data folder is also provided, containing an example dataset as well as all the required input data for using `MULTIPOL`.

The program can be downloaded from:

https://psico.fcep.urv.cat/utilitats/MULTIPOL-C/index.php

-3. Program Usage I: Entering and importing data

As mentioned, `MULTIPOL` can be used through code version or through the shiny app (`MULTIPOL_GUI`). In both cases, the inputs required are the same, but in the code version, the inputs must be R variables, and in the GUI version, the inputs are expected to be external data files, ideally, .csv files with no format.

The usage of `MULTIPOL` though the script (once sourced) is the following:

```
>MULTIPOL(X, B, PHI, method = "fast", output = TRUE)
```

While the usage of `MULTIPOL_GUI` is the following:

```
>library(shiny); runApp('MULTIPOL_GUI.R')
```

The inputs required for `MULTIPOL` are described as:

`X`: A Respondents ×items data matrix, containing the item scores, scaled between 0 and 1.

`B`: An Items ×dimensions pattern matrix, containing factor loadings based on the covariance FA (see section 1.2)

`PHI`: Inter-factor correlation matrix. (see section 1.2)

`method`: The precision of the EAP score estimation, where "fast" uses 25 quadrature nodes, and is only recommended when the computing time is a constrain, and "accurate", which uses 200 quadrature nodes and provides robust and accurate scores, but is high time consuming.

output: Determines if the output will be displayed in the console, TRUE by default. If it is TRUE, the output is printed in console and if it is FALSE, the output is returned silently to the output variable. When using the GUI version, this option is always TRUE, since it needs to print the results for the user to see them.

-3. Program Usage II: Output

The output can be presented in the console (if `output = TRUE`) and/or returned as an R list, including all the output variables from the analysis. This list includes:

`Alpha`: Item easiness parameter estimates for each item.

`Beta`: Item discrimination parameter estimates in each factor. They are the same as those in the **B** matrix given in the input

`Overall_item_discriminations`: (see section 1.1)

`Multidimensional_item locations`: (Mdif; see section 1.1)

`Multidimensional_slopes`: (Obtained at the quantiles .25, .5, .75; see section 1.1)

`Orthogonal_projections`: Orthogonal item projections on the factor axes (see section 1.1).

`Scores`: A matrix containing for each factor: the Point Estimate scores, the corresponding standard errors (PSDs) and the conditional reliability estimates, in a matrix format. (see section 1.1)

When using `MULTIPOL_GUI` version (shiny app), the results will be only printed in the console.

-4. Illustrative example

The following simulated dataset illustrates how to use `MULTIPOL`. The database used is located in the file named **X10_2F.csv**. The data consists of 1000x10 matrix, containing item scores scaled in the (0,1) interval and conforms two a bidimensional complex solution in which each item has a dominant loading on one factor and a secondary loading on the other. The provided input is the inter-factor correlation matrix (**PHI.csv**), and the covariance-based loadings which are the same as the Item beta parameter estimates (**Beta.csv**).

As previously described, the program can be runned through code, using `MULTIPOL.R`, or through GUI, using `MULTIPOL_GUI.R`, we are going to run both cases, using the same example data.

4.1 Using `MULTIPOL` through code:

Open `MULTIPOL.R`, and source the file, using the function source(), including within the parentheses the path of the `MULTIPOL.R` file.

Next, we need to import the three required inputs. This can be done in different ways, for example, using the function read.csv(), which is native to R. Once all the three files are imported, and `MULTIPOL` sourced, we can begin the analysis:

```
> output <- MULTIPOL(X = X10_2F, B = Beta, PHI = PHI, method =
"fast", output = TRUE)
```

As you can see, we decided to store the output in a new variable called output, which will be a list containing the results of the analysis, described in the previous section. Plus, since we decided to keep the output to be printed in the console, the results will be shown in console as well as saved in the output variable.

First, the calibration results will be shown:

```
CALIBRATION

Item easiness parameter estimates:
0.1965
0.2390
0.2418
0.2831
0.2852
0.2043
0.2554
0.2619
0.3077
0.3021

Item Beta parameter estimates:
1.0000    0.4000
1.0000    0.4000
1.1000    0.3000
1.9000    0.3000
2.0000    0.3000
0.4000    1.0000
0.4000    1.0000
0.3000    1.1000
0.3000    1.9000
0.3000    2.0000
```

```
Overall item discriminations:
1.4000
1.4000
1.4000
2.2000
2.3000
1.4000
1.4000
1.4000
2.2000
2.3000

Multidimensional locations (difficulty):
4.3123
3.7494
3.5745
2.1658
2.0939
4.1935
3.5756
3.3759
2.0854
2.0421

Multidimensional slopes:
    .25      .5     .75
 0.1334 0.08116 0.02777
 0.1535 0.09335 0.03194
 0.1610 0.09792 0.03351
 0.3138 0.25394 0.11559
 0.3321 0.27461 0.12774
 0.1372 0.08346 0.02856
```

0.1609 0.09789 0.03350

0.1704 0.10368 0.03548

0.3259 0.26374 0.12005

0.3405 0.28157 0.13098

Orthogonal projections on the factor axes:

3.6446	2.3050
3.1688	2.0041
3.1684	1.6547
2.0128	0.7998
1.9525	0.7562
2.2415	3.5441
1.9112	3.0219
1.5627	2.9924
0.7701	1.9380
0.7375	1.9043

The calibration results agree with what can be expected when the unipolar modeling is appropriate. Note that the easiness parameters are all below one (which means that the marginal expected item scores are quite low), and, correspondingly, the multidimensional locations are rather high. So, in all of the items, a high level in the "optimal" trait composite is required for attaining a midpoint score of 0.5. Please, recall that the median in the latent trait distribution is 1, and all the multidimensional difficulties are far above this value. As for item discriminating power, the multidimensional discriminations are very high (see section 1.1) and the slopes are non-negligible only at the 0.25 quantile and then level off.

Next, the scores matrix will be presented. Please, keep in mind that, when using large datasets, this matrix can be too large to fit the console limits. The scores for the first 10 simulated participants are as following:

SCORES

	Point estimate 1	PSD 1	Reliability 1	Point estimate 2	PSD 2	Reliability 2
1	2.777	1.617	0.855	0.617	0.844	0.918
2	5.337	1.061	1.519	0.099	0.506	0.998
3	1.051	1.051	0.001	0.000	1.000	1.000
4	2.713	1.054	0.760	0.056	0.876	0.999
5	2.297	1.102	0.607	0.229	0.921	0.989
6	1.051	1.051	0.000	0.000	1.000	1.000
7	1.051	1.051	0.001	0.000	1.000	1.000
8	1.665	1.882	0.630	0.687	0.915	0.899
9	19.304	3.646	3.712	1.132	0.010	0.725
10	15.339	4.949	4.035	1.568	0.010	0.473

The most relevant pieces of information above are the EAP point estimates in each factor and the corresponding conditional reliabilities. Note, that as discussed in section 1.1, the conditional reliabilities tend to decrease the higher the point EAP estimate is.

Finally, under the scores matrix, the marginal reliabilities for both factors are presented:

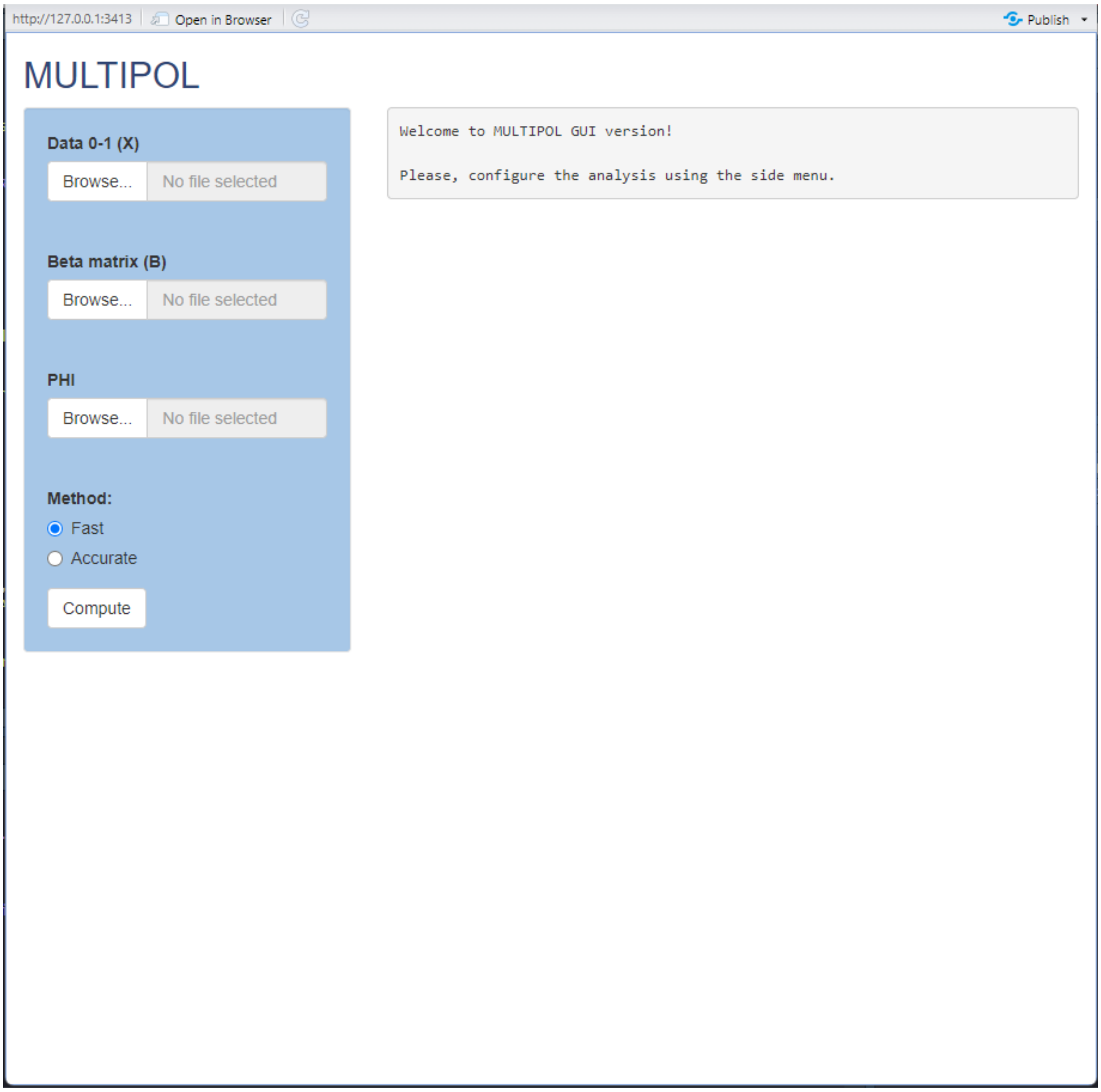
```
Marginal reliability 1: 0.9160214

Marginal reliability 2: 0.9094496
```

As can be seen, both marginal reliabilities are rather high, but see the notes in section 1.1 regarding the limitations of these indices.

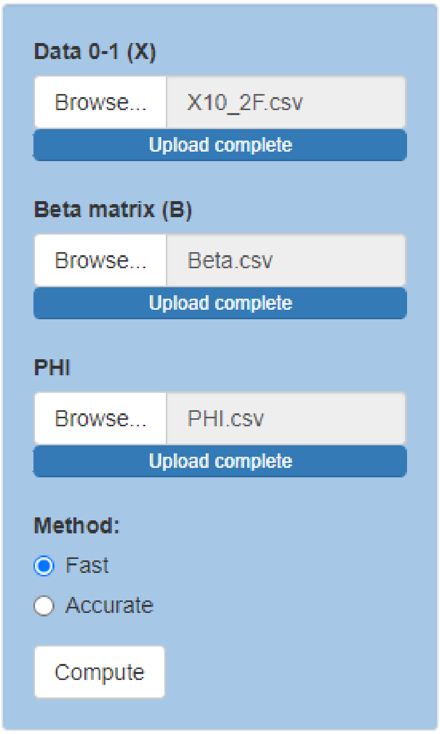
4.2 Using `MULTIPOL` though GUI:

If the user is interested in using the GUI version, just open the `MULTIPOL_GUI.R` file and run the app, which can be done through the function runApp(), or using "Run App" button on RStudio. Please, remember that `MULTIPOL_GUI` has some dependencies, mainly the shiny package. The first time that the app is launched, additional packages can be installed.

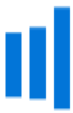
Once launched, the app should look like the following:



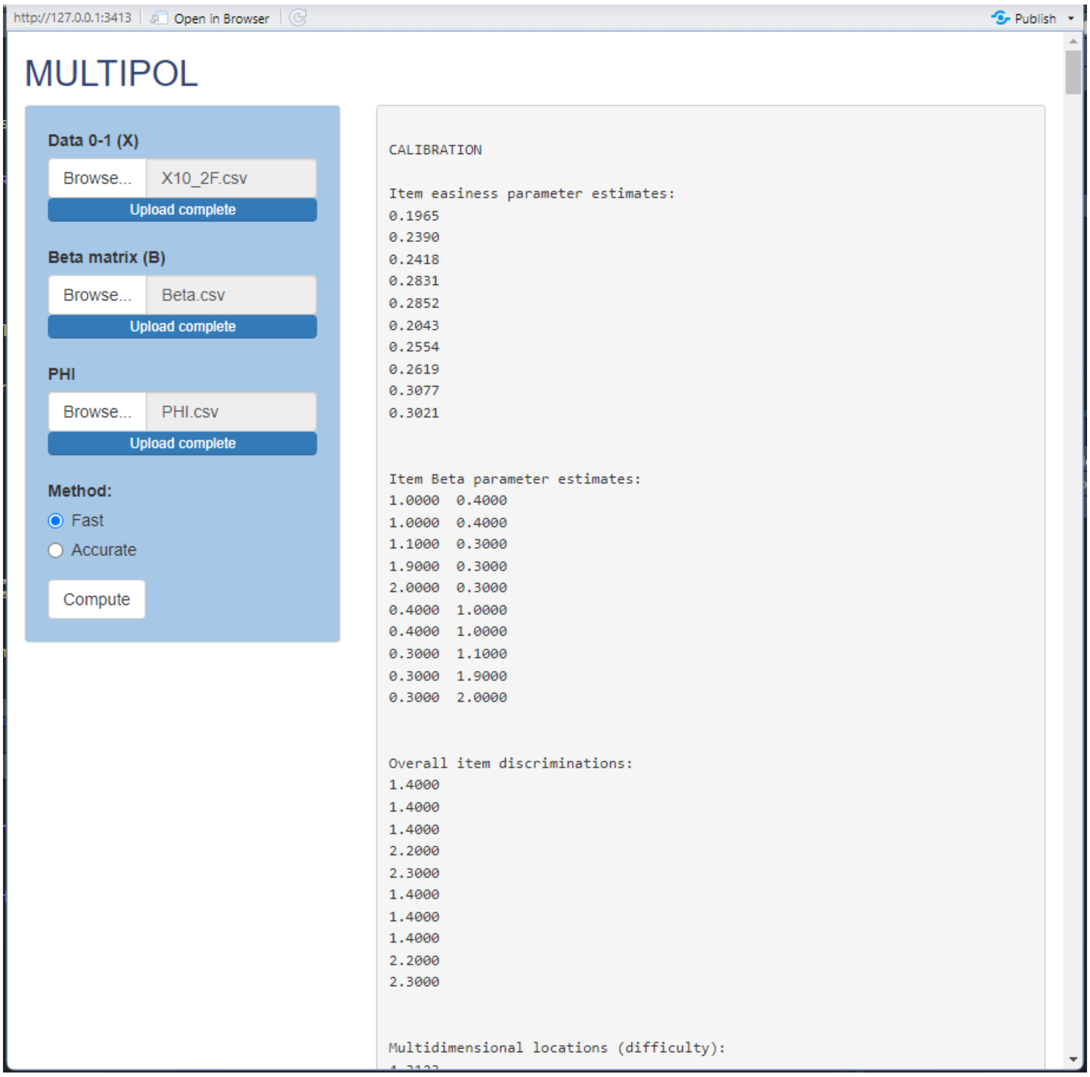
Next, the user has to provide the path to the three required inputs, using the "Browse…" buttons. The user can select if the fast or the accurate method will be used, in this example, we used the fast one. The input panel should look like this:



Finally, the user must click on Compute and wait until the analysis is finished. A moving symbol will appear on the right side of the application, indicating that the function is busy:



Once it is done, the output would be presented on the right side, containing the printed version of the output. Please, remember that, when using the GUI version, the output cannot be saved as a variable.



As shown, the results are exactly the same as the previously presented in the code version of MULTIPOL, so the interpretation remains identical.

References

Aitchison, J., & Brown, J.A.C. (1957). *The Lognormal Distribution.* Cambridge University Press.

Bejar, I. I. (1977). An application of the continuous response level model to personality measurement. *Applied Psychological Measurement*, *1*(4), 509-521. https://doi.org/10.1177/014662167700100407

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied psychological measurement*, *6*(4), 431-444. https://doi.org/10.1177/014662168200600405

Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2025). shiny: Web Application Framework for R. R package version 1.10.0.9000, https://github.com/rstudio/shiny, https://shiny.posit.co/

Ferrando, P.J. (2009). Difficulty, discrimination, and information indices in the linear factor analysis model for continuous item responses. *Applied Psychological Measurement*, *33*(1), 9-24. https://doi.org/10.1177/0146621608314608

Ferrando, P. J., Morales-Vives, F., & Hernández-Dorado, A. (2024). Measuring unipolar traits with continuous response items: Some methodological and substantive developments. *Educational and Psychological Measurement*, *84*(3), 425-449.

Huang, H. Y. (2025). Exploring the Influence of Response Styles on Continuous Scale Assessments: Insights From a Novel Modeling Approach. *Educational and Psychological Measurement*, *85*(1), 178-214.

Lucke, J.F. (2013). Positive trait item response models. In R. E. Millsap, L. A. van der Ark, D. M. Bolt, and C. M. Woods (Eds.), *New developments in quantitative psychology* (pp. 199–213). Springer.

Lucke, J.F. (2013). Positive trait item response models. In R. E. Millsap, L. A. van der Ark, D. M. Bolt, and C. M. Woods (Eds.), *New developments in quantitative psychology* (pp. 199–213). Springer.

Lucke, J.F. (2015). Unipolar item response models. In S. P. Reise and D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 272–284). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315736013

Magnus, B.E., & Liu, Y. (2018). A zero-inflated Box-Cox normal unipolar item response model for measuring constructs of psychopathology. *Applied Psychological Measurement*, *42*(7), 571-589. https://doi.org/10.1177/0146621618758291

Morales-Vives, F., Ferrando, P. J., & Dueñas, J. M. (2023). Should suicidal ideation be regarded as a dimension, a unipolar trait or a mixture? A model-based analysis at the score level. *Current psychology*, *42*(25), 21397-21411.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied psychological measurement*, *9*(4), 401-412.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual review of clinical psychology*, *5*(1), 27-48.

Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, 39, 111-121.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied psychological measurement*, *1*(2), 233-247. https://doi.org/10.1177/014662167700100209

Samejima, F. (1996). Evaluation of mathematical models for ordered polychotomous responses. *Behaviormetrika*, *23*, 17-35.

Stevens, S.S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. Transaction Publishers.

Wang, T., & Zeng, L. (1998). Item parameter estimation for a continuous response model using an EM algorithm. *Applied Psychological Measurement*, *22*(4), 333-344. https://doi.org/10.1177/014662169802200402

Wang, T., & Zeng, L. (1998). Item parameter estimation for a continuous response model using an EM algorithm. *Applied Psychological Measurement*, *22*(4), 333-344. https://doi.org/10.1177/014662169802200402