

# A Measure of Poverty Based on the Rasch Model

Maria Cristiana Martini and Cristiano Vanin

**Abstract** Asking questions about income is often a hard task: non-response rates are high, and the reliability of the surveyed answers is sometimes poor. In this contribution we propose a measure of the economic status based on a set of wealth-related items. This measure is based on the Rasch model, which allows to estimate different relevance weights for each item: difficulty parameters indicate the severity of each of the situations described by the items, while ability parameters indicate poverty level of each unit. This poverty measure is computed on survey data collected on a sample of 2465 households living in Veneto. Some analyses conducted on the estimated poverty measure indicate a good consistency with the expected relationships, and confirm the possibility to get this kind of estimates from indirect measures.

**Keywords** Measuring poverty and inequality; Rasch model; Income; Composite indicators.

## 1 A Rasch Model to Measure Poverty

Income questions are among the most threatening questions in general population surveys, and often produce inadequate information. Non-response rates are frequently quite high, and family income tends to be underestimated because respondents are likely to report earning after taxes and deductions, or to forget about less common sources of income. Reluctance to report income data and difficulty to collect and summarise information from different family income sources has led to a number of strategies to collect data and to correct estimates. In this paper we propose a method to estimate wealth levels (or poverty levels) which does not require to ask threatening questions about income; this method is based on the Rasch model.

---

<sup>1</sup> M. Cristiana Martini, Università di Modena e Reggio Emilia; email: cmartini@unimo.it

Cristiano Vanin, Regione Veneto; email: cristiano.vanin@regione.veneto.it

The Rasch model was developed by Georg Rasch in 1960, aimed at objectively assessing the school achievement of Danish soldiers. Then, the most natural applications of the Rasch model are in psychometrics and education, but its flexibility and simplicity makes it suitable for other applications as well, such as quality measurement and social sciences in general. Broadly speaking, this model is a good instrument to measure latent variables.

The Rasch model can also be used for the study of poverty, but in literature this use is not frequent. Among the most recent works we can remind Fusco and Dickes (2008), while one of the first examples comes from the French Gailly and Hausman (1984).

To apply this psychometric model, we consider poverty as a latent variable. The Rasch model properties make it suitable to study poverty: the specific objectivity allows to obtain an objective measure of poverty, seen as a unidimensional trait in the model hypotheses; then, the possibility to rank the items and grade them in a hierarchical way, grants a definition of economic disease as the sum of different factors.

In the following application, the original and simpler version of the Rasch model is used, with items coded by dichotomous variables. The starting matrix is given by a set of units (households) observed in the values of a set of dichotomous items, regarding the ownership of some goods, or the manifestation of some situations, considered as deprivation indicators. The positive response  $x_{ij}=1$  indicates a state of deprivation for the  $i$ -th family with regard to the good or characteristic expressed by the  $j$ -th item, while the null response  $x_{ij}=0$  indicates absence of deprivation. The aim is to transform the answers in a synthetic measure that allows to place the latent variable on a continuous scale.

The peculiarity of this model consists in considering two factors to measure the latent variable: the individual ability and the item difficulty. So, the latent variable is evaluated for each individual not only on the basis of how many questions he answered positively, but also based on the degree of difficulty of each question: the answer  $x_{ij}$  given by the  $i$ -th respondent depends on his relative deprivation, i.e. his intrinsic poverty combined with the level of deprivation indicated by the  $j$ -th item. This method produces two separate rankings: one ranks individuals from the most to the least needy (controlling for the items difficulty parameters), the second orders questions according to the level of poverty they imply, taking into account how many individuals lack that good or are in that situation. The Rasch model represents these two factors by means of two families of parameters:  $k$  parameters of difficulty for the items ( $\beta_j, j=1, \dots, k$ ), and  $n$  ability parameters for the individuals ( $\theta_i, i=1, \dots, n$ ).

In this case, the parameter of difficulty  $\beta_j$  is renamed “parameter of severity”, since it corresponds to the idea of inequality: the corresponding ranking shows which are the goods whose lack mainly contribute to generate poverty, and describes the inequalities in the goods distribution among families. The higher the severity parameter, the more widespread is the ownership of the good, and the higher is the inequality degree corresponding to its lack. The underlying hypothesis is that lacking a good is more weighty if its ownership is common in the reference population (Micheli and Laffi, 1995). In this sense, we refer to a relative poverty approach, since the inequality scales are not built in absolute terms, but in comparison with other households.

The ability parameter  $\theta_i$  directly refers to poverty, and is renamed “position parameter”. Its meaning is clear, and it allows to rank the families from the most to the least poor. In fact, the higher the position parameter, the harsher its deprivation and poverty situation. Moreover, once we know the position parameter of a family and the

severity parameter for a certain good, we can calculate the probability for that family to suffer deprivation with respect to the item.

To formalise, the relationship between the latent variable and the observed items corresponds to the probability  $P(X_{ij}=1|\theta_i, \beta_j)$  of a positive answer to the  $j$ -th item for the  $i$ -th individual. The formula resembles that of a logistic regression model, with parameters belonging to two families:

$$P(X_i = 1 | \theta_i, \beta_j) = \frac{e^{(\beta_j - \theta_i)}}{1 + e^{(\beta_j - \theta_i)}}, \quad i = 1, \dots, n; j = 1, \dots, k$$

Compared with the psychometric applications, the use of Rasch models to study poverty has to square with non-experimental survey designs. In fact, while psychometrics and pedagogy studies are based on questionnaires especially designed to present batteries of valid and reliable tested items, all measuring the same construct, in the studies on economic disease the data are often collected according to a different rationale, and data must be recoded to satisfy the basic assumptions.

## 2 Application to Survey Data

The Rasch model has been applied to the data coming from a survey conducted on a sample of 2465 households living in Veneto. The survey, carried out from December 2004 to April 2005, has been conducted via computer assisted telephone interviews by the Statistics Department of the University of Padua, on behalf of the Local Centre for Household Documentation and Analysis. The Rasch model analysis responds to the demand for a variable expressing the situation of material poverty, meant as lack of goods or monetary shortage. In fact, the questionnaire includes a question about household income, but the wide response classes proposed<sup>1</sup> and the problem of tailoring these classes to the family size, do not allow to estimate income in an easy way. Then, in order to build a measure of material poverty, we consider a set of items that refer, to some extent, to the economic sphere.

Unlike Gailly and Hausman (1984), who consider 32 objective and subjective indicators of disease, we try to disentangle the material and non material components of poverty, therefore the Rasch model is fitted on the basis of a set of items all concentrated on different aspects of material deprivation, in order to obtain a unique principal latent trait. The selected items, dichotomised to indicate forms of material deprivation, are those reported in Table 1. All the variables take value 1 for households in a situation of deprivation with regard to the item.

In order to assess the internal consistency of the selected items, we calculate the Cronbach  $\alpha$  (Cronbach, 1951) and we perform an exploratory factor analysis to test the existence of a unique common latent factor. The Cronbach  $\alpha$  measures the internal

---

<sup>1</sup> The four wide classes are: “Less than 1500 € net per month”, “Between 1.500 € and 3.000 € net per month”, “Between 3.000 € and 6.000 € net per month”, “More than 6.000 € net per month”.

consistency of the items by measuring their correlations. In our case the value is 0.67, acceptable even if not high<sup>1</sup>.

The factor analysis is performed on the described items with principal components analysis and varimax rotation (Kaiser, 1958), and the eigenvalues scree-plot shows a unique latent factor, interpretable as a deprivation factor. The first factor, corresponding to the first eigenvalue ( $\lambda_1=3.5$ ), explains only 15% of the total variance. However, the reliability of the factorial procedure, expressed by Kaiser measure (Kaiser, 1974), equals 0.73, which indicates an “average” goodness according to the indications found in literature and given by the same Kaiser.

The algorithm used to select the final set of items is a backward procedure. First we apply the Rasch model to the whole matrix  $X$ , then we estimate the severity parameters  $\beta_j$  referred to the items, and the position parameters  $\theta_i$  for the families through the conditional maximum likelihood method (Andrich, 1988; Rasch, 1960/1980)<sup>2</sup>. Then, we check the statistical significance of the estimated severity parameters: the item corresponding to the less significant parameter is removed, and the Rasch model reestimated. The procedure is repeated until we obtain a set of items for which the model assumptions hold. Then, the final solution contains a set of items all referred to the same latent variable of economic deprivation, since the ones which less concur to the latent variable are excluded from the model.

After the initial estimate, only the parameter corresponding to the variable “no computers in the household” is statistically non significant, and is removed from the model. The second (and final) estimate of the Rasch model produces the parameters of severity shown in Table 1, all statistically significant at least at a 99% level.

We can note that the scale range is quite wide (ranging from  $-3.15$  to  $5.77$ ), making the set of items quite informative; in fact, when the items have similar severity parameters, the different response profiles tend to have the same probability, and the solution lacks continuity. Table 1 reports the items ordered according to the severity parameter: the ones with higher estimates of  $\beta_j$  mainly concur to generate inequalities, and can increase the probability of material poverty.

The first remark regards the ownership of material goods: these characteristics usually do not generate poverty, and are placed at the bottom of the ranking, with the

<sup>1</sup> Nunnally (1978) defines 0.70 an acceptable value for Cronbach  $\alpha$ ; 0.67 is quite close to that threshold.

<sup>2</sup> Among different possible estimation methods for the Rasch model, conditional maximum likelihood is the most frequent, but also other methods can be used, e.g. joint maximum likelihood or marginal maximum likelihood.

The method here used maximises the log-likelihood function conditional to the total score of each individual  $r_i = \sum_{j=1}^J x_{ij}$  :

$$L = -\sum_{j=1}^J s_j \beta_j - \sum_{r=1}^{I-1} f_r \log \gamma(r, \beta) \quad \text{with } s_j = \sum_{i=1}^I x_{ij}$$

where  $f_r$  indicates the number of individuals achieving a score equal to  $r$ ,  $J$  is the number of considered items and  $I$  the number of individuals, and  $\gamma(r, \beta)$  is the following symmetric function:

$$\sum_{x_{ij}=1}^r \exp(-\sum_{j=1}^J x_{ij} \beta_j) \cdot$$

Due to the property of sufficiency of the  $r_i$  scores, the maximum only depends on the severity parameters, which are the only elements in the estimate equations, resolvable through a Newton-Raphson procedure. Once the severity parameters have been estimated, their values can be used to estimate the position parameters, through a separate procedure.

only exceptions of the possess of car and TV. This confirms a predictable result, e.g. the ownership of rare goods such as satellite TV, motorcycle, DVD players or internet connection does not discriminate between poor and non poor households, while the lack of common goods as a car or a TV indicates economic deprivation. In fact, not having the first group of goods depends more on cultural factors and life style, since they are more frequent among young, highly educated households, while the elderly rarely hanker these devices.

**Table 1:** Severity parameters estimates and standard errors<sup>1</sup>, for the Rasch model fitted on the selected items. Ranking of the deprivation items according to their relevance.

| <i>Item</i>                                       | <i>Estimate of <math>\beta_j</math>:<br/>severity parameter</i> | <i>Standard-<br/>error</i> |
|---------------------------------------------------|-----------------------------------------------------------------|----------------------------|
| Eviction                                          | 5.77                                                            | 0.30                       |
| Use of savings or loans in the last year          | 3.83                                                            | 0.12                       |
| No TV in the household                            | 3.41                                                            | 0.10                       |
| Insufficient housing for the household needs      | 3.24                                                            | 0.10                       |
| Need for economic support in the last y           | 3.12                                                            | 0.09                       |
| No cars in the household                          | 2.45                                                            | 0.07                       |
| Lives in a building with 10 flats or more         | 2.34                                                            | 0.07                       |
| Rental house/apartment                            | 2.33                                                            | 0.07                       |
| No holidays in the last year for economic reasons | 1.79                                                            | 0.06                       |
| Savings consumption in the last year              | 1.48                                                            | 0.06                       |
| Lives in a building with less than 10 flats       | 1.14                                                            | 0.05                       |
| Household income less than 1.500 € per month      | 1.08                                                            | 0.05                       |
| No videotape recorder in the household            | 1.03                                                            | 0.05                       |
| Pension as the main household income              | 0.70                                                            | 0.05                       |
| Unable to save money in the last year             | 0.27                                                            | 0.05                       |
| No DVD player in the household                    | -0.26                                                           | 0.05                       |
| No internet connection in the household           | -0.39                                                           | 0.05                       |
| No motorcycle in the household                    | -1.30                                                           | 0.05                       |
| No satellite TV in the household                  | -1.55                                                           | 0.06                       |
| No ADSL connection in the household               | -1.83                                                           | 0.06                       |
| No laptops in the household                       | -2.16                                                           | 0.07                       |
| No paid domestic help in the household            | -2.36                                                           | 0.07                       |
| No flat screen TV in the household                | -3.15                                                           | 0.09                       |

<sup>1</sup> The standard errors are computed in the following way:

$$SE(\beta_j) = \left[ (1/\sum_{r=1}^I n_r) p_{rj} (1 - p_{rj}) \right]^{1/2}$$

where  $p_{rj}$  is the probability estimate to correctly answer the  $j$ -th item, for an individual with score  $r$ , and  $n_r$  is the number of individuals scoring  $r$ .

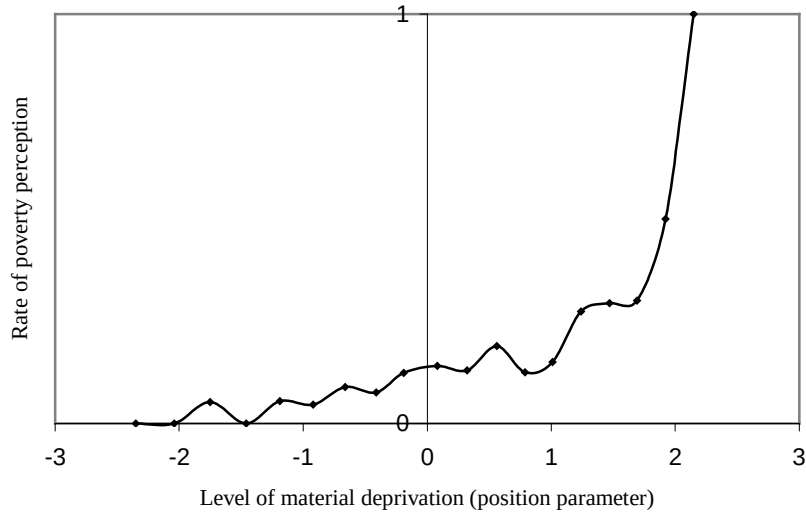
Among the most relevant items we can notice eviction and recourse to savings and loans; in fact, they are quite direct indicators of economic poverty and monetary trouble. Items concerning housing conditions are also quite relevant in the ranking, especially living in a rental house/apartment and in an apartment: this confirms the importance of the housing expenses to determine the economic situation. Moreover, Italian families traditionally aspire to buy their house, even at the cost of great economic sacrifice.

Finally, the parameter corresponding to the monthly income is placed in a medium ranking position, due to the extremely poor quality of the variable as collected in this survey.

### 3 Validation of the Poverty Scale

The variable resulting from the Rasch model fit should be a measure of poverty. In order to assess its validity, we must check if it behaves as expected. First of all, we find the poverty measure consistent with a variable on poverty self-perception asked in the same survey: families who perceive themselves in economic trouble have significantly higher levels of the new poverty measure (see Figure 1).

**Figure 1:** Percentage of households who perceive a poverty situation corresponding to each estimated value of the position parameters  $\theta_i$ .



Moreover, we model a regression tree where the (quantitative and continuous) criterion variable is the estimated position parameter resulting from the Rasch model.

The regression tree is estimated by maximising a criterion function  $\Phi(s,t)$ , giving a measure of diversity between the child nodes generated by partitioning the corresponding parent node  $t$ . In our application, the criterion function is Fisher's  $\eta^2$  index, derived from the heterogeneity function  $i(t)$  measuring the lack of homogeneity of a group with respect to the dependent variable. The  $i(t)$  function is the deviance of node  $t$ :

$$i(t) = \sum_i |y_i - \bar{y}_t|^2$$

where  $\bar{y}_t$  is the mean value  $y$  in the parent node  $t$ . Thus, we can estimate the mean value of  $y$  in each node, which turns to be a useful information to characterise the group.

The tree building process is very similar to a multiple regression model where the variables are selected through a forward procedure. Additionally, the regression tree allows to detect combinations of the predictive variables with interaction effects, and to insulate and describe groups of families particularly subjected to economic deprivation.

The regression tree is fitted by means of the library *tree* of software R (Venables e Ripley, 1999), devoted to Classification And Regression Trees (CART) analysis. The halt rules are set as follows: the minimum dimension of groups is set at 100 units to avoid the making of tiny groups; the maximum number of final nodes is set at 14 to avoid too complex branching; the minimum total deviance of a parent node is set at 1% of the starting deviance of  $y$  to avoid further partitioning of homogeneous nodes.

The analysis of the regression tree (not reported here) shows results which are consistent with what we expect from a poverty variable. The traditional family with an employed head of household is the wealthier type of family. These are couples, couples with children and extended families, whose principal income derives from employment or self-employment. Among these, the most protected from poverty are households with an average or high level of education. Also for households with lower education levels, steady families are less vulnerable to poverty, both when they have been formed from longer periods and when they have a stable structure: in this sense, the traditional couple with children is the less economically frail. A traditional and stable familiar structure seems to prevent from poverty independently on the working situation of the head of the household. In any case, for each type of family structure, children are a factor that protects from poverty; probably their presence is a stimulus to face toward critical situations and assure them a certain future.

Families whose main income comes from pensions or other sources have higher poverty levels. The traditional household structure (couple with children or couple with children and other members) is again less vulnerable to poverty. Moreover, the existence of more than one child further protects from material deprivation; since these are mostly old families with a retired head of the household, more children probably imply more incomes. The main difference, compared with the traditional families in the left side of the tree, concerns the couples without children: when the head of the household is not employed (hence, in most cases, old), this type of family is more frail to poverty. Finally, chronic invalids in the household correspond to higher levels of poverty, probably because they require medical attention and care expenses, in addition to the constant assistance of family members, thus subtracted to other activities.

## 4 Final Remarks and Future Perspectives

The Rasch model allows to estimate a poverty measure, also in default of a reliable and accurate measure of income. The necessary variables for the estimate pertain the economic area and the ownership of durable goods, and are therefore much easier to survey than income, and less prone to definition complexity, non responses and reluctant respondents.

A comparison between the new poverty measure and the perceived poverty, as well as the analysis of the relationship with some household variables likely to influence the poverty level, confirm the adequacy of this indicator. The lack of a reliable income variable in the survey data used for the application prevents from performing a complete validation of the obtained measure. The application could be replicated on a different dataset which includes an income variable, in order to compute an alternative measure of poverty independent from the income. The comparison of the two (direct and indirect) measures of poverty will also shed light on the possible reasons of discrepancy, and suggest future improvements.

## References

1. Andrich, D.: Rasch models for measurement. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-068. Sage Publications, Beverly Hills (1988)
2. Cronbach L.J.: Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297--334 (1951)
3. Fusco A., Dickes P.: The Rasch Model and Multidimensional Poverty Measurement. In: Kakwani N. and Silber J. (eds) Quantitative approaches to multidimensional poverty measurement, pp. 49-62. Palgrave Macmillan, New York (2008)
4. Gailly B., Hausman P. : Des Désavantages Relatifs à une Mesure Objective de la Pauvreté. In: Sarpellon G. (eds.) Understanding poverty, pp. 192-216. FrancoAngeli, Milano (1984)
5. Kaiser H.F.: The Varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**, 141--151 (1958)
6. Kaiser H.F.: An index of factorial simplicity. *Psychometrika* **39**, 31--36 (1974)
7. Micheli G.A., Laffi S. (a cura di): *Derive. Stati e percorsi di povertà non estreme*. FrancoAngeli, Milano (1995)
8. Nunnally J.: *Psychometric theory*. McGraw-Hill, New York (1978)
9. Rasch G.: Probabilistic models for some intelligence and attainment tests; Danish Institute for Educational Research, Copenhagen 1960. New edition, The University of Chicago Press, Chicago 1980 (1960/1980)
10. Venables W.N., Ripley B.D.: *Modern applied statistics with S-PLUS*. Springer, New York (1999)