Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques
International Networks for Studies in Technology, Environment, Alternatives and Development

IRISS Working Papers

IRISS-C/I

An Integrated Research Infrastructure in the Socio-economic Sciences

Item response theory and the measurement of deprivation: Evidence from PSELL-3

by

Monica Raileanu Szeles
Alessio Fusco

IRISS Working Paper Series 2009-05 May 2009

Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques
International Networks for Studies in Technology, Environment, Alternatives and Development

Item response theory and the measurement of deprivation: Evidence from PSELL-3

Monica Raileanu Szeles
Transylvania University of Brasov, Romania and CEPS/ INSTEAD, Luxembourg
Alessio Fusco
CEPS/ INSTEAD, Luxembourg and CEMAFI University of Nice Sophia Antipolis

Abstract Item Response Theory (IRT) has recently been proposed as a framework to measure deprivation. It allows deriving a latent measure of deprivation from a set of dichotomous observed items of deprivation and analyzing determinants of deprivation. We investigate further the use of IRT models in the field of deprivation measurement. Firstly, the paper emphasizes the importance of item selection and the Mokken Scale Procedure is applied in order to select the items to be included in the scale of deprivation. Secondly, we apply the one and the two-parameter probit IRT models for dichotomous items on two different sets of items, in order to highlight different empirical results. Finally, we introduce a graphical tool, the Item Characteristic Curve (ICC) and analyse the determinants of deprivation in Luxembourg. The empirical illustration is based on the fourth wave of the Luxembourg socioeconomic panel Liewen zu Lëtzebuerg (PSELL-3).

Reference IRISS Working Paper 2009-05, CEPS/INSTEAD, Differdange, Luxembourg

URL http://ideas.repec.org/p/irs/iriswp/2009-05.html

Monica Szeles is sponsored by a post-doc grant ("Bourse Formation Recherche" BFR05/098-LB) financed by the Luxembourg *Ministère de la culture, de l'enseignement supérieur et de la recherche* and administered at CEPS/INSTEAD.

(CEPS/INSTEAD internal doc. #07-09-0470-E)

ITEM RESPONSE THEORY AND THE MEASUREMENT OF DEPRIVATION: EVIDENCE FROM PSELL-3[1]

Monica Raileanu Szeles[2]
Transylvania University of Brasov, Romania
CEPS/ INSTEAD, Luxembourg
Alessio Fusco
CEPS/INSTEAD, Luxembourg
CEMAFI – University of Nice Sophia Antipolis

Abstract:

Item Response Theory (IRT) has recently been proposed as a framework to measure deprivation. It allows deriving a latent measure of deprivation from a set of dichotomous observed items of deprivation and analyzing determinants of deprivation. We investigate further the use of IRT models in the field of deprivation measurement. Firstly, the paper emphasizes the importance of item selection and the Mokken Scale Procedure is applied in order to select the items to be included in the scale of deprivation. Secondly, we apply the one and the two-parameter probit IRT models for dichotomous items on two different sets of items, in order to highlight different empirical results. Finally, we introduce a graphical tool, the Item Characteristic Curve (ICC) and analyse the determinants of deprivation in Luxembourg. The empirical illustration is based on the fourth wave of the Luxembourg socioeconomic panel "Liewen zu Lëtzebuerg" (PSELL-3).

Key words: Item Response Theory, deprivation, latent trait, Mokken scale, PSELL3.
JEL: I32

---

[1] This paper was written when Monica Szeles was sponsored by a post doctoral grant financed by the Luxembourg Ministère de la Culture, de l'Enseignement Supérieur et de la Recherche, whose financial support is gratefully acknowledged. The authors would like to thank Philippe Van Kerm, Paul Dickes, Jean-Benoît Hardouin, Tudorel Postolache and the participants to the research conference in Maastricht "The pursuit of certainty: Applications in measuring poverty" (2007) for their helpful comments and suggestions. We are also grateful to Lorenzo Cappellari and Stephen P. Jenkins who made available part of their programme. The usual disclaimer applies.

[2] Corresponding author. Address: Transilvania University of Brasov, Romania, e-mail: monica.szeles@unitbv.ro;

1. Introduction

For Kakwani and Silber (2008a:xv), “the most important development of poverty research in recent years is certainly the shift of emphasis from a uni- to a multidimensional approach to poverty”. This theoretical advance, recognised by many researchers, has recently given rise to an abundant literature (*inter alia*, Jenkins and Mickelwright, 2007 or Stewart *et alii*, 2007). The concept of deprivation is part of this growing literature on the multidimensional approaches of poverty, together with other approaches such as Sen’s capability approach.[3]

Following the definition proposed by Townsend (1987:125), “deprivation may be defined as a state of observable and demonstrable disadvantage relative to the local community or the wider society or nations to which an individual, family or group belongs”. It can therefore be considered as the inability to possess the goods and services and engage in the activities that are ordinary in the society or that are socially perceived as necessities. These definitions shed light on the fact that the situation of disadvantage should not be assessed solely on the basis of a lack of financial resources, that is in an indirect way, but that there are some advantages to focus directly on the achievements of the individuals (Sen, 1979; Ringen, 1988).

The theoretical improvement provided by the concept of deprivation is illustrated, at the institutional level, by the recent adoption of a commonly agreed indicator of material deprivation to be included in the list of social inclusion indicators at the European Union level. The rationale behind inclusion of indicators of material deprivation in this list is that they provide a better understanding of the living conditions of the poor and give information about domains that are not dealt with in the relative income approach of poverty (Guio, 2005, 2009).

The growing importance of this concept in the field of poverty research has led to the proposal of a wide variety of techniques to measure deprivation (see e.g. Kakwani and Silber 2008b). The main task usually consists in summarising the information available from a set of categorical items of deprivation. In this paper, deprivation is conceptualised as a latent construct, which needs to be inferred from a number of manifest deprivation indicators. In

[3] See Kakwani and Silber (2008a) or Fusco (2007) for a review of the different multidimensional approaches of poverty.

this context, we argue that the Item Response Theory (IRT), an extension of Classical Test Theory (CTT) with historical roots in mathematics and psychology, provides a suitable methodological framework to analyse deprivation.

IRT is a traditional technique of educational and psychological measurement. Its application to the analysis of deprivation goes back to *inter alia* Gailly and Hausman (1984) or Dickes (1989). See Cappellari and Jenkins (2007) and Fusco and Dickes (2008) for recent applications.[4] These papers make use of IRT mainly as measurement models in order to derive deprivation scales; in addition, Cappellari and Jenkins (2007) also use IRT as explanatory models when analysing the determinants of deprivation.

The focus of our paper is similar to the one of Cappellari and Jenkins (2007). It consists in investigating the advantages of the IRT framework when measuring deprivation. In addition, we extend they work by also examining comparatively the one and two-parameter IRT models, emphasizing the importance of item selection through the application of the Mokken Scale Procedure and introducing other IRT tools such as the item characteristic curve.

The paper is structured as follows. The concepts and methods are described in Section 2. Section 3 describes the data from the Luxembourgish socioeconomic panel 'Liewen zu Lëtzebuerg' (PSELL-3) while results and analysis are reported in Section 4. Finally, Section 5 summarises the main findings and provides a final discussion.

2. On the use of IRT to measure deprivation

Deprivation can be considered to be a latent concept that needs to be inferred through its manifest indicators. The responses of individuals to deprivation items questionnaire represent the manifest or observed indicators. Deprivation is usually assessed by collecting data on the extent to which households possess certain commodities, engage in certain activities or are subject to financial pressures (Whelan, 1993).

The steps leading to the computation of a measure of deprivation are the choice of the relevant dimensions/domains and the set of elementary indicators representing them, the evaluation of deprivation on each of these items and dimensions, the aggregation of the elementary indicators into a composite index for each dimension and, if considered relevant,

[4] See also Lancaster and Green (2002a, 2002b) for recent applications aiming at analyzing the link between deprivation and ill-health.

the aggregation of the different dimensions into an overall index of deprivation (Chiappero Martinetti, 2000; Nolan and Whelan, 1996). In this paper, after having introduced the link between items testing and items selection (2.1), we focus on the step of aggregation of different items of deprivation into a synthetic index of multiple deprivation through the use of IRT models (2.2) and explain how the analysis of the determinants can be done (2.3).

2.1. Items testing and selection

Theoretical as well as empirical requirements are needed to build a reliable scale of deprivation. Selecting the relevant domains and items to measure deprivation should first rely on theoretical criteria in order to operationalize the concept of deprivation. This important issue is out of the scope of this paper (see e.g. Alkire, 2001). From an empirical point of view, potential items of deprivation must fulfil a set of assumptions and requirements in order to ensure that they are referring to the same latent construct. Some IRT models requires very restrictive set of assumptions (e.g. the Rasch model) whereas others are less restrictive (e.g. the two-parameter IRT). In all cases, IRT models rely on a set of fundamental hypotheses (Hardouin, 2005):

- Unidimensionality of latent trait: The first central assumption in the IRT is that the items measure just one latent trait. This hypothesis implies that a single dominant trait gives the probability of item endorsement.
- Local independence: The second central assumption is the local independence. According to this assumption, after controlling for dominant factors, item pairs should not be associated. The local independence relates to the unidimensionality in the sense that no other characteristic of the individual influences the response probabilities.
- Monotonicity: The third assumption states that the probability of presenting a disadvantage is a non-decreasing function of the latent trait; the higher is the position of an individual on the latent trait, the higher is his probability of answering correctly to a given item.

Testing the fulfilment of these fundamental hypotheses allows indirectly identifying a set of items respecting the IRT assumptions. This can be considered as a selection procedure. Some authors give a high importance to this question (e.g. Hardouin, 2005 or Fusco and Dickes, 2008) when others do not put much emphasis on it (e.g. Cappellari and Jenkins, 2007).

2.2 Construction of a summarizing index

IRT models have been used extensively in the fields of educational and psychological measurement, where the objects of analysis, such as ability, personality or intelligence, are often of a latent nature. This method is derived from the item-based test theory, that belonged itself, originally, to the wide field of psychological measurement (Lawley, 1943; Minton, 1988).[5] Over time, the IRT has been continuously developed, which has led to its acceptance as a specific technique of psychometrics (Lord, 1952; Rasch, 1966; Baker, 1992).

Let y_{ij} be the answer of individual i=1..n to item j=1..m. Items used to measure deprivation are usually dichotomous with y_{ij}=1 if individual i presents a disadvantage on item j and 0 otherwise.[6] y_{ij} are repeated observations pertaining to each individual and provide the data from which measurements can be inferred. An unweighted score of deprivation can be computed $S_i = \sum_{j=1}^{m} y_{ij}$. The method consisting in using S_i to analyze deprivation is called the sum-score approach, which is a simple counting method.[7] The underlying measurement model to the unweighted raw score is the so called classical test theory (CTT). This theory presupposes that effects between answers of individuals are only due to variation in the ability; all potential sources of variation are assumed to be constant or to have a nonsystematic (random) effect (van der Linden and Hambleton, 1997). Hence, the observed score is partly due to an underlying true score and partly to measurement error:

$$y_{ij} = D_i^* + \varepsilon_{ij} \qquad [1]$$

Where D_i^* is the latent true score for individual i and ε_{ij} are measurement errors. ε_{ij} have an expected value of zero, are independent from the true score and are mutually

[5] In psychometrics and educational testing, IRT is also known as Latent Trait Analysis (LTA). LTA studies the relation between latent continuous variables and observed categorical variables (binary or ordered).

[6] In the case of categorical or continuous items of deprivation, y_{ij} can take more values. Fuzzy sets approach provide functions allowing to assign gradual degree of deprivation to the different modalities of the item (see Chiappero Martinetti, 2000) Moreover, IRT models can also handle non dichotomous items (see De Boeck and Wilson, 2004, Zheng and Rabe-Hesketh, 2007 or Lancaster and Green, 2002a).

[7] The computation of S_i above mentioned considers that all the items have the same importance. To relax this strong assumption it is possible to include weight in the computation of the deprivation index $S_i = \sum_{j=1}^{m} w_j y_{ij}$, where w_j is the weight of the item j. If the items have the same importance then w_j=1 for all item j=1..m and S_i is an unweighted raw score. We do not consider the weighting issue in this paper.

independent. The classical measurement model implies that, in the long run, the average score is expected to be equal to the true score if the number of deprivation indicators is large enough (De Boeck and Wilson, 2004; Cappellari and Jenkins, 2007).

However, this model suffers from several weaknesses. From a psychometric point of view, the legitimacy of summing the different items in the same scale is not verified within the classical measurement model (Fusco and Dickes, 2008). From a more technical point of view, the sum score approach cannot satisfactorily be incorporated into the framework of the classical measurement model when the observed deprivation indicators are dichotomous and not continuous ones (Lancaster and Green, 2002; Cappellari and Jenkins, 2007).

IRT models have been built on the basis of the critique formulated toward the CTT. They consist in making an explicit parametrization of both the latent ability and the properties of the items, by modeling the relationship between the observed items and the latent variable. Indeed, as stated by Molenaar (1995:4), "IRT is built around the central idea that the probability of a certain answer when a person is confronted with an item, ideally can be described as a simple function of the person's position on the latent trait plus one or more parameters characterizing the particular item."[8] IRT is generally written in terms of the generalised linear model formulation, where the conditional probability of a particular answer given the latent trait, called the item characteristic curve, is specified by a link function (Skrondal and Rabe-Hesketh, 2004:71).

Two types of link function are generally specified. The normal IRT models are based on the cumulative normal probability distribution function while the logistic IRT models are based on the logistic function.

A number of advantages arise when moving from the CTT to IRT. When applied to the analysis of deprivation, in the CTT the unit of analysis is the scale of deprivation while in the IRT the unit of analysis is the item itself. Therefore in the CTT, data are modelled at the level of deprivation score rather than at the item level. The IRT relates both the characteristics of items (item parameters) and the characteristics of individuals (deprivation score) to the probability of providing a particular response (Lord, 1980). This leads, in the

[8] Ostini and Nering (2006:2) put it like this: "The mathematical foundation of IRT is a function that relates the probability of a person to responding to an item in a specific manner to the standing of that person on the trait that the item is measuring. In other words, the function describes, in probabilistic terms, how a person with a higher standing on a trait (i.e. more of a trait) is likely to provide a response in a different response category to a person with a low standing on the latent trait. This mathematical function has a pre-specified form (usually a logistic ogive) and is now generally referred to as an item response function (IRF)."

case of the IRT, to an optimal scale design, as items and individuals are represented together on the same scale. The general form of the one-parameter IRT equation is:

$$y_{ij}^{*} = \beta_j + D_i^{*} + \varepsilon_{ij}$$

$$y_{ij} = 1 \quad \text{if} \quad y_{ij}^{*} > 0 \quad \text{and} \quad y_{ij} = 0 \quad \text{otherwise} \qquad [2]$$

Where y_{ij}^{*} is the latent score of deprivation for item j and individual i, β_j represents the item parameter and D_i* is the person parameter. ε_{ij} are independent with mean zero. The larger β_j the higher the probability that $y_{ij} = 1$ given D_i^{*}. Hence, households are less likely to present a disadvantage linked to an item with smaller β_j which Cappellari and Jenkins term intrinsic cheapness parameter. In psychometrics, it is a common practice to analyse $-\beta_j$ which is interpreted as an item difficulty parameter. In the framework of deprivation, $-\beta_j$ can be called the parameter of severity of the item and D_i^{*} the parameter of deprivation of the individual (Fusco and Dickes, 2008). The difficulty parameter is the point on the latent deprivation scale where an individual has a 0.5 probability of being deprived of an item. In the one-parameter model, for a given level of deprivation D_i^{*}, the probability of presenting a disadvantage decreases with the difficulty parameter of the item, and for a given level of difficulty $-\beta_j$, the probability of presenting a disadvantage increases with the level of deprivation (Skrondal and Rabe-Hesketh, 2004). This property of double monotonicity implies that items and individuals are strictly ordered.

As mentioned by Cappellari and Jenkins, the completion of the specification of the model requires making assumptions about the functional form of the distribution of error terms and about the treatment of the latent deprivation variable D_i^{*}.

The Rasch model is a well known one-parameter item response model where D_i^{*} are treated as fixed parameters and where the error term has a logistic distribution (Molenaar, 1995). Under these assumptions, the unweighted sum score is a sufficient statistic of the individual ability parameter given the item parameters. This means that the simple aggregation of the indicators respecting the Rasch model assumptions contains all the statistical information on the value of the unknown ability parameter. Conditional maximum likelihood can be used to estimate the item parameters (Skrondal and Rabe-Hesketh, 2004).[9] The property of specific objectivity stipulates that "comparison of the ability of two subjects should only depend on the ability of these subjects (and not the ability of others) and that the comparison should yield the same results whatever item the comparison is based on" (Skrondal and Rabe-Hesketh, 2004:73). The hard task with this model consists in finding indicators fulfilling its restrictive properties. It is not only the Rasch model, but it is the one-parameter IRT model in general that imposes strong assumptions. From a theoretical point of view, Fusco and Dickes (2008) consider that the Rasch model, is well suited to operationalise a definition of poverty in terms of an accumulation of disadvantages.[10]

Another strategy consists in treating D_i^* as individual random effects. In this case, the standard maximum likelihood provides estimates of the parameter β_j and predicted value of D_i^* can be estimated by empirical Bayes methods that make use of both the assumed latent variable distribution and the pattern of observed responses of the posterior distribution and the item parameters. As mentioned by Cappellari and Jenkins, this method presents some caveats due, *inter alia*, to the unknown small-sample properties and to the strong assumption underlying the model such as the equi-correlation between any pair of item.

The use of the two-parameter IRT model allows relaxing the assumption of equi-correlation between several items through the introduction in the model of a second item parameter. λ_j are factor loadings and are called discrimination parameters as they reflect the discriminating power of the items between individuals whose latent score of deprivation are below and above the item difficulty. The larger λ_j is, the better the discriminating power of the item.

$$y_{ij}^* = \beta_j + D_i^* \lambda_j + \varepsilon_{ij}$$

$$y_{ij} = 1 \quad if \quad y_{ij}^* > 0 \quad and \quad y_{ij} = 0 \quad otherwise \qquad [3]$$

[9] Cappellari and Jenkins (2007:171) underline a potential problem: if conditional maximum likelihood can estimate β_j when n tends to infinite and given m fixed, the D_i^* parameters cannot be estimated. Standard maximum likelihood estimates of D_i* are inconsistent as n tends to infinite given m fixed. Hence the usual small number of items available in the study of deprivation is a problem for the estimation of the model.

[10] To estimate the Rasch model with Stata, see the command –raschtest- by Jean-Benoît Hardouin (2007).

The item parameters β_j and λ_j can be estimated by maximum likelihood and D_i^* by empirical Bayes methods. For model identification, the scale of D_i^* is fixed by anchoring, i.e. assuming that $\lambda_1 = 1$. In the two-parameter IRT model a change in the latent score of deprivation does not equally affect the items of deprivation.

The natural question arising from the existence of the variants of the models are which one we should choose (Wilson, 2003). The two-parameter IRT model is less constringent on the data so that it often has a better fit than the one-parameter IRT model. From a statistical point of view, this can be seen as an advantage (Cappellari and Jenkins, 2007); in terms of operationalisation of a concept of poverty as an accumulation of disadvantage, it is not necessarily better (Fusco and Dickes, 2008).

2.3. Analysis of determinants

Equations [2] and [3] refer to measurement model. To move from a measurement model to an explanatory model, in order to analyse the determinants of deprivation, a structural equation can be incorporated into the model. The structural equation introduces the covariates which explain the latent score of deprivation as follows:

$$D_j^* = \alpha Z_j + \xi_j \qquad [4]$$

where Z_j is a vector of observed covariates, α is the vector of regression coefficients and ξ_i is a normally distributed error term with mean zero and fixed variance. The addition of the structural model (equation [4]) with a measurement model (equation [2] or [3]) is similar to a Multiple Indicator Multiple Cause (MIMIC) model in the framework of structural equations modelling.[11]

In conclusion, IRT is a complex but powerful methodological tool for measuring deprivation as it allows deriving a deprivation scale, computing a deprivation score and also analysing the determinants of deprivation, as a MIMIC model.

3. An empirical illustration based on the PSELL-3 data

[11] See Kuklys (2005) for an application of MIMIC model to the operationalisation of the capability approach.

In this section, we apply IRT models to measure deprivation and analyse its determinants in Luxembourg. The analysis uses data from the Socio-Economic Panel "Liewen zu Lëtzebuerg" (PSELL-3) which is the luxembourgish component of the EU-*Community Statistics on Income and Living Conditions* (EU-SILC). PSELL-3 was launched in 2003, with an initial sample of 3500 households that were representative of the population living in private households in Luxembourg. In this paper, we focus on the fourth wave of PSELL-3, conducted in 2006 and the unit of analysis is the household.

Following the logic of the previous section, the aim of this empirical section is threefold: to illustrate the use of testing the IRT hypothesis in order to select the items of deprivation that will be included in the scale (3.1), to estimate and choose between the one- and two-parameter IRT models (3.2) and finally to analyse the determinants of deprivation in Luxembourg (3.3). These three points are studied subsequently on two sets of items. The first one refers to items of financial stress and the second contains items from different domains of life in order to identify a 'global' scale of deprivation. The application of the same models on two different sets of items aims at illustrating different types of results and different choice of models.

3.1. Items selection and testing

The item testing and selection play an important role in our analysis. In order to test the three fundamental hypotheses of IRT, i.e. unidimensionality, local independence and monotonicity, we use the Mokken Scale Procedure (MSP) (Hemker *et alii*, 1995).[12] In this case, MSP can be seen as an automatic item selection procedure to identify a set of items pertaining to a unique scale and respecting the three hypotheses of IRT (Hardouin, 2005). MSP is a hierarchical scaling method having a probabilistic nature, where the reproducibility is measured by the Loevinger's H coefficient. The closer the Loevinger's H coefficient is to 1, the better is the scale. In addition, we compute the Cronbach alpha coefficient to assess the reliability of the selected sets of items. The value of the Cronbach alpha coefficient increases when the internal consistency (correlation) between the items increases.

[12] The Mokken Scale Procedure has been applied using the Stata modules -msp-.and -loevH- . Others tests can be applied to check whether a set of items respect the parametric IRT model properties. In particular, Hardouin (2005, 2007) presents a set of local and global tests for the Rasch model (see the –raschtest- package in Stata).

PSELL-3 contains the usual items used in the framework of deprivation measurement (see e.g. Layte *et alii*, 2001). Through an application of MSP on this initial set of items, we identify two scales of deprivation that we call scale of global deprivation and scale of financial stress. The Loevinger H coefficient of 0.72 for the scale of financial stress and 0.59 for the global scale suggest that both sets of items show good scale properties (Mokken, 1971). Additionally, the Cronbach alpha coefficient for the financial stress scale is 0.85 and for the global scale of deprivation 0.56. These values indicate an acceptable internal consistency for both scales.

The global scale of deprivation includes nine dichotomous items pertaining to the enforced lack of durable goods, housing facilities and the capacity to afford basic requirements. The items are listed below according to their increasing deprivation rates in Luxembourg computed on 3001 observations:

1. Cannot afford to have a washing machine (if wanted to) (0.4%)
2. The dwelling has no bath (0.6%)
3. Cannot afford keeping home adequately warm (0.6%)
4. Cannot afford to have a car (if wanted to) (1.6%)
5. Cannot afford eating meat or equivalent every second day (2.2%)
6. Cannot afford one week annual holiday away from home (11.9%)
7. Cannot afford facing unexpected expenses (20.5%)

The variables of the global scale of deprivation do not only form a reliable scale, but they are also often used in the literature to operationalise deprivation.

The scale of financial stress is constituted of three dichotomous items:

1. Arrears on hire purchase instalments or other loan payment (1.3%)
2. Arrears on mortgage or rent payment (1.4%)
3. Arrears on utility bills (1.8%)

The area of the financial stress indicators is broader, but we limit at three items that are suggestive to describe this domain of deprivation.

3.2. Estimation of IRT models

We estimate the one and two-parameter IRT models on the two sets of items defined earlier. This allows ranking the items of deprivation according to their parameter of severity, computing the latent score of deprivation and comparing the one and two-parameter models for each scale. Table 1 displays the estimates of the IRT models for the scale of global deprivation.[13] According to the item difficulty parameter, the items “washing machine” and “bath” are the most severe in the one-parameter IRT. This implies that the probability that an individual who has no bath (or no washing machine) to be deprived of the other items of this scale is higher than 0.5. For the two-parameter IRT models, the item “washing machine” is again the most difficult item whereas the rank of the “bath” is lower compared to the one-parameter model. “Ability to face unexpected expenses” is the easiest item in both models. It is interesting to note that the models yield different rankings of deprivation items in terms of their difficulty, which is different from what Cappellari and Jenkins (2007) report. This may be due partly to the fact that, contrary to Cappellari and Jenkins, some of the items composing our global scale of deprivation show very close proportions of deprivation and partly to the fact that the range of discrimination parameter is larger in our case (0.74 to 2.70) compared to Cappellari and Jenkins (0.80 to 1.62).

[13] The Stata commands –gllamm- and –gllapred- were used for all the computations (see Rabe-Hesketh *et alii*, 2004).

Table 1: The estimates from the probit IRT models: global scale

	One parameter IRT		Two parameter IRT	
	Est	(SE)	Est	(SE)
intercept *(β_i)*				
Washing machine	4.52	0.17	3.46	0.33
Bath	4.51	0.16	3.08	0.22
Keep home warm	3.95	0.13	3.17	0.24
Car	3.56	0.11	2.77	0.16
Meat or equivalent every second day	3.49	0.11	3.26	0.26
Holiday away from home	1.89	0.07	2.72	0.38
Ability to face unexpected expenses	1.14	0.06	1.13	0.08
discrimination parameter (λ_j)				
Washing machine	1		1.00	Fixed
Bath	1		0.71	0.22
Keep home warm	1		1.13	0.29
Car	1		1.07	0.25
Meat or equivalent every second day	1		1.50	0.36
Holiday away from home	1		2.70	0.77
Ability to face unexpected expenses	1		1.69	0.37
Variance D_i^*	2.28	0.19	0.81	0.34
Log likelihood	-3528.32		-3503.93	

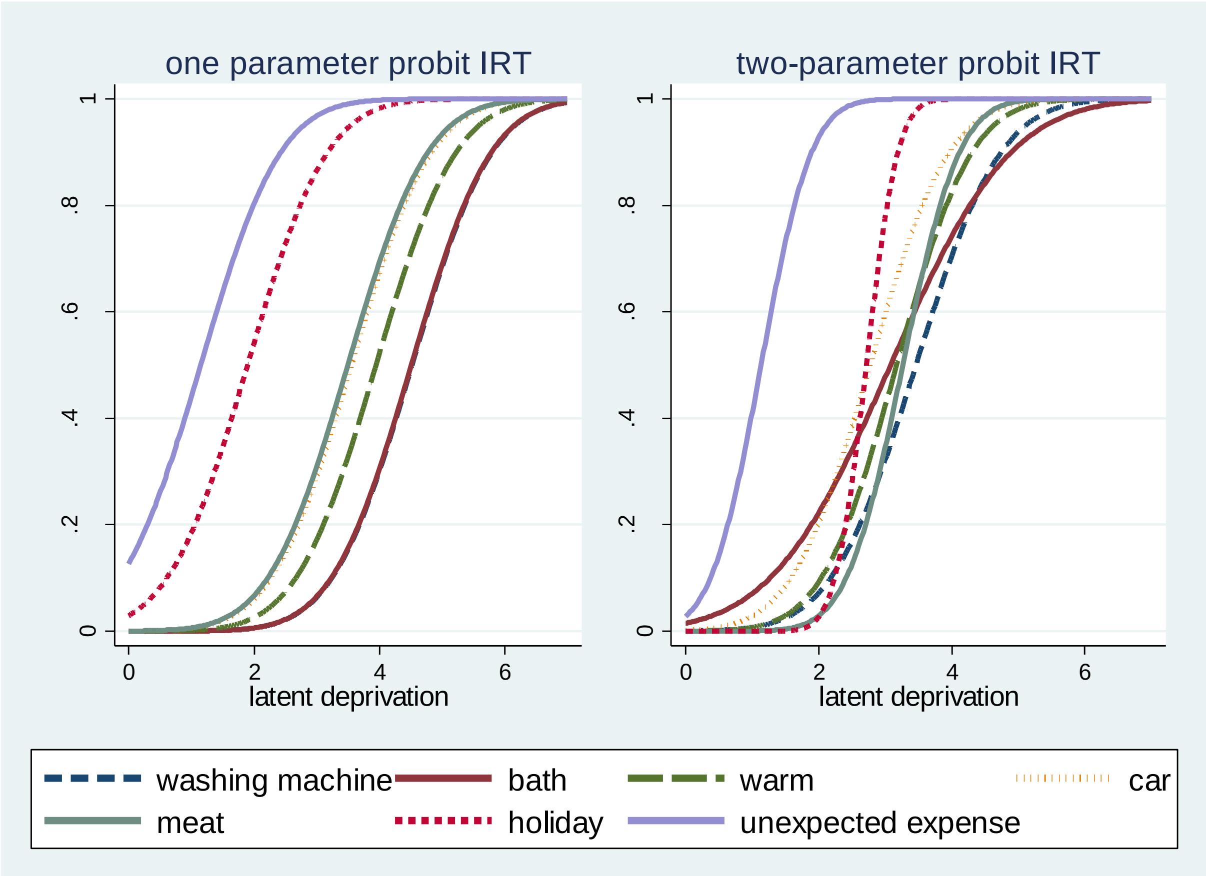
Source: PSELL-3, CEPS/INSTEAD, 2006, authors' computation

Note: all coefficient are significant to the 1% threshold; Likelihood ratio test of the two-parameter IRT model over the one-parameter IRT model LR chi2 (6) =48.77, Prob > chi2 = 0.00.

In addition to the item difficulty parameter, the discrimination parameter reflects the impact that a change in the latent measure of deprivation has on each deprivation indicator probability. The most discriminatory item is about "affordability of one week annual holiday away from home", while "not having a bath" is the least discriminatory items explaining the lower rank in terms of difficulty in the two-parameter model. The discrimination parameters have values ranging from 0.74 to 2.70, suggesting that in this case, the two-parameter IRT brings additional valuable information in comparison with the one-parameter IRT where implicitly, all the discrimination parameters are equal to one. The variance of the household latent deprivation is 2.28 for the one-parameter model and 0.81 for the two-parameter model. Also, the likelihood ratio test indicates that the two-parameter IRT fits better our data than the one-parameter IRT.

A useful tool to complement the previous table is the Item Characteristic Curve (ICC - also known as item response function or trace line). ICC is a graphical representation describing the relationship between the latent variable and the probability of giving a positive answer to an item. It plots the conditional probability of a particular response, given the latent trait $P(y_{ij} = 1 | D_i^*)$. In our case, the ICC is a plot of the household latent score of deprivation over the probability of being deprived upon an item. The higher the latent score of deprivation, the higher the probability of being deprived for a given item. The ICC gives insights on the location of the item parameters. The difficulty parameter is a location index describing where the item stems along the deprivation scale. The higher the difficulty parameter is, the less likely a given individual is to be deprived of that item. Figure 1 displays the ICC for the global scale of deprivation.

Figure 1. Item characteristic curves for the scale of global deprivation



Source: PSELL-3, CEPS/INSTEAD, 2006, authors' computation.

In the one-parameter IRT model, all items have a discrimination parameter equal to one. This implies that all the ICCs have the same slope and that they do not intersect. The ICC of the easiest item ("unexpected expense") is the one the most on the left whereas the ICC of the most difficult items ("bath" and "washing machine") are almost overlapping on the right. In the two-parameter IRT model, all items have different discrimination parameters describing how well the item differentiates between individuals having deprivation scores below or above the item location. The steepness of the ICC in its middle section reflects the discrimination power of items. The steeper the curve is, the higher is the discriminating power of the item as the probability of being deprived of the item is very different between individuals whose latent score of deprivation are right below and right above the item difficulty. The flatter the curve is, the less the item is discriminatory. The most discriminatory items are "unexpected expenses" and "holiday"; the less discriminatory items are "washing machine" and "bath". The ICC of "bath" crosses the ICC of other items.

The ICC represents a graphical tool of analysis in the IRT modelling. Even though the ICC curves do not bring any additional evidence about the items but they provide a friendly graphical interpretation to the underlying IRT models.

Estimates of the probit IRT models and the ICCs related to the financial stress scale are reported in the annex (Table A1 and Figure A1). In terms of item difficulty parameter, the ability of paying loans (one-parameter IRT) and the ability of paying bills (two-parameter IRT) are the most difficult, while the ability of paying rents (one-parameter IRT) and the ability of paying rent (two-parameter IRT) are the easiest. The discrimination parameter has lowest value for the ability to pay loans and the highest value for the ability to pay bills. The likelihood ratio test suggests that all the three items have close discrimination powers as the one-parameter model fits better the data than the two-parameter model.

This result indicates that the two-parameter IRT model is not always a better choice than the one-parameter IRT. However, it should be noted that this result is less common. We have applied the two IRT models to several scales of deprivation and the cases where the one-parameter IRT model is preferred to the two-parameter IRT model are very rare. Moreover, the high household level variance as well as the presence of some missing values in the data asks for cautiousness when interpreting the results brought by this scale.

3.3. The determinants of deprivation in Luxembourg

In this section, the determinants of deprivation in Luxembourg are examined using the one and two-parameter IRT models. As explained in a previous section, the analysis of determinants is done by incorporating a supplementary structural equation into the general IRT model (equation 4). By so doing, the derivation of the individual deprivation scores and the analysis of determinants are integrated into one single model, which functions as a MIMIC model (Skrondal and Rabe-Hesketh, 2004). We analyse the impact on deprivation of a set of covariates including socio-economic characteristics of the head of the household (gender, age and citizenship), the status on the labour market of the head of the household, the household type and the logarithm of the household equivalised disposable income.[14] The results are presented in Table 2.

[14] For a more in depth analysis of social exclusion in Luxembourg, see Raileanu Szeles (2008).

Table 2. The determinants of global deprivation

	one-parameter IRT			two-parameter IRT		
	Coefficient		SE	coefficient		SE
female	0.21	**	0.08	0.13	*	0.05
age	0.01		0.01	0.01		0.01
age squared	0.00		0.00	0.00		0.00
Portuguese	0.57	***	0.09	0.36	***	0.09
other non-native	0.32	***	0.08	0.20	**	0.06
lone-parent family	0.43	***	0.13	0.29	**	0.10
large family	0.25	*	0.11	0.16	*	0.08
unemployed	0.85	***	0.14	0.56	***	0.14
inactive	0.20		0.10	0.12		0.07
log income	-1.61	***	0.08	-1.01	***	0.21
constant	14.77	***	0.87	9.23	***	1.97
log likelihood	-3015.1			-2987.3		

Source: PSELL-3, CEPS/INSTEAD, 2006, authors' computation

Notes. 1) One-parameter IRT: var(D_i^*): 1.15 (SE: 0.11); Two-parameter IRT: var(D_i^*): 0.48 (SE: 0.20)

2) * p<0.05, ** p<0.01, ** p<0.001

The table above shows that the effects of different covariates of the latent deprivation score are close, in relative terms, although there is a difference between them, in absolute terms. In the one-parameter IRT model, deprivation is higher for households with low income and for lone family. There is also, on average, a higher risk of deprivation when the head of the household is unemployed, non Luxembourgish and for female-headed household. The age of the head of the household doesn't appear to have an impact on deprivation levels. The same conclusions can be reached with the two-parameter model.

In table A2 of the annex, the impact of the same list of covariates on the scale of financial stress is presented. Very few variables have a significant impact on latent deprivation at the 1% level. In the one-parameter IRT model, unemployed-headed households are the most exposed to financial deprivation, as well as low income households. In the two-parameter IRT model, the same effects exist but to a lower level of significance.

4. Conclusions

In this paper, we have investigated the use of IRT models for summarising a number of dichotomous items into a synthetic index of multiple deprivation. IRT provides a methodological framework to derive deprivation scales, to summarize the items into a score of deprivation and to analyze the determinants of deprivations. In addition to this objective, that is similar to the one of Cappellari and Jenkins (2007), we also examine comparatively the one and two-parameter IRT models, emphasize the importance of item selection through the application of the Mokken Scale Procedure, analyse the determinants of deprivation in Luxembourg and introduce other IRT tools such as the item characteristic curve.

The items testing and selection represent an important step in the IRT analysis. There is a broad range of tests checking whether the items designed to form a scale fulfil the particular requirements of the Rasch model or the IRT general assumptions. In our paper, the Mokken Scale Procedure has been used as to determine whether the set of items used respect the IRT fundamental assumptions.

Even though in the deprivation literature both the one and the two-parameter IRT models carry the general advantages of latent models over the traditional models, there are some differences between them. By adding an item discrimination parameter, the two-parameter IRT model relaxes the strong assumption of equi-correlation imposed by the one-parameter IRT model and often yields a better fit. There are rare cases where this is not the case. In this paper, we applied the one and two-parameter IRT models on two different sets of deprivation items. In the case of the global scale of deprivation, the likelihood ratio test shows a better fit for the two-parameter IRT whereas in the case of the financial stress scale, the one-parameter IRT model fits better the data. These empirical findings suggest that the two-parameter IRT does not always bring an improvement in comparison with the one-parameter IRT, even though this is rarely the case. However, the choice between the one- or the two-parameter IRT model should also be based on theoretical consideration. For example, when studying the cumulative nature of economic and social disadvantages in order to describe deprivation, the one-parameter IRT such as the Rasch model, can be considered to be a better choice (Fusco and Dickes, 2008).[15] Finally, when studying the determinants of deprivation, it appeared that the two IRT models provide similar conclusions.

[15] van der Linden and Hambleton (1997 :12) underline this point : « no general recommendation can be made with respect to this choice between a more stringent model with excellent statistical tractability and a more

The measurement of deprivation through the IRT can be extended in many directions, at both cross-sectional and longitudinal levels. At a cross-sectional level, multidimensional IRT allows deriving several scales of deprivation (such as monetary and non-monetary) from a set of items and a number of models (for example, polytomous IRT models) could be used to describe and summarize other types of data than the dichotomous ones. IRT can also be used to evaluate differential item functioning or the non-equivalence of measurement items across groups of people, by examining the probabilities of item endorsement across these groups. The extension of the IRT at a longitudinal level could also be done in several directions and it may give insights into the process of change over time.

---

flexible model likely to fit a larger collection of items. Additional factors such as (1) the nature of the misfit, (2) the availability of substitute items, (3) the amount of time available for rewriting items, (4) the availability of a sufficiently large sample to properly estimate item parameters for more general models, and –probably most important- (5) the goal of the testing procedure play a significant role in the handling of items that are not fit by the Rasch model."

References

1. Alkire S.: Valuing freedoms. Sen's Capability Approach and Poverty Reduction. Oxford University Press, Oxford (2001)
2. Baker, F.B.: Item Response Theory: Parameter estimation techniques. Marcel Dekker, New York (1992)
3. Cappellari, L., S.P. Jenkins: Summarizing Multiple Deprivation Indicators. In: Jenkins, S.P, Micklewright, J. (eds) Inequality and Poverty: Re-examined, pp. 166-184. Oxford University Press, Oxford (2007)
4. Chiappero Martinetti, E.: A multidimensional assessment of well-being based on Sen’s functioning approach. Rivista Internazionale di Scienze Sociali **108**, 207-31 (2000)
5. De Boeck, P., Wilson, M.: Explanatory Item Response Models: A generalized linear and nonlinear approach. Springer, New-York (2004)
6. Dickes, P.: Pauvreté et conditions d'existence.Théories, modèles et mesures, Document PSELL **8**, Walferdange CEPS/INSTEAD, Luxemburg (1989)
7. Fusco, A.: La pauvreté, un concept multidimensionnel, L’Harmattan, Paris (2007)
8. Fusco, A., Dickes, P.: The Rasch model and multidimensional poverty measurement. In: Kakwani, N., Silber, J. (eds) Quantitative approaches to multidimensional poverty measurement, pp. 49-62. Houndmills Palgrave Macmillan (2008)
9. Gailly, B., Hausman, P.: Désavantages relatifs a une mesure objective de la pauvreté. In: Sarpellon, G. (ed.) Understanding Poverty, pp. 192-216. Franco Angeli, Milan (1984)
10. Guio, A.C. : La privation matérielle dans l’UE. Statistiques en bref : Populations et conditions sociales. **21**, Eurostat (2005)
11. Guio, A.C.: What can be learned from deprivation indicators in Europe?. Paper presented at the SPC Indicators Sub-Group, 10 February 2009
12. Hambleton, R. K.: Item Response Theory: A Broad Psychometric Framework for Measurement Advances. Psicothema. **6(3)**, 535-556 (1994)
13. Hardouin, J.B.: Rasch analysis: Estimation and tests with raschtest. Stata J. **7(1)**, 1-23 (2007)
14. Hardouin, J.B.: Manual for the SAS macro-programs LoevH and MSP and the Stata modules LoevH and MSP, http://anaqol.org/biblio/msp.pdf, (2005)

15. Hemker, B. T., Sijtsma, K., Molenaar, I. W.: Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. Appl. Psych. Meas. **19(4)**, 337-352 (1995).
16. Jenkins, S.P, Micklewright, J. (eds): Inequality and Poverty: Re-examined. Oxford University Press, Oxford (2007)
17. Kakwani, N., Silber, J. (eds): Many Dimensions of Poverty. Houndmills: Palgrave Macmillan, London (2008)
18. Kakwani, N., Silber, J. (eds): Quantitative Approaches To Multidimensional Poverty Measurement. Houndmills: Palgrave Macmillan, London (2008)
19. Kuklys, W.: Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications. Springer Verlag, Berlin (2005)
20. Lancaster, G., Green, M.: Deprivation, ill-health and the ecological fallacy. J. Royal Stat. Soc. A. **165**, Part2, 263-78 (2002)
21. Lancaster, G., Green, M.: Latent variable techniques for categorical data. Statistics and computing **12**, 153-161 (2002)
22. Lawley, D.N.: On problems connected with item selection and test construction. In Proc. Roy. Soc., Edinburgh, **61**A, 273-287 (1943)
23. Layte, R., Nolan, B., Maître, B., Whelan, C.T.: Explaining deprivation in the European Union. Acta Sociologica **44(2)**, 105-122 (2001)
24. Lord, F.M.: A theory of test scores. Psychometric Monogr **7**, New York (1952)
25. Lord, F.M.: Applications of Item Response Theory to practical testing problems. NJ: Lawrence Erlbaum, Hillsdale (1980)
26. Marlier, E., Atkinson, A.B., Cantillon, B., Nolan, B.: The EU and social inclusion: Facing the challenges. Policy Press, Bristol (2007)
27. Minton, H., Lewis, L., Terman, M.: Pioneer in psychological testing. New York University Press, New York (1988)
28. Molenaar, I.W.: Some Background for Item response Theory and the Rasch Model. In: Fischer, G.H., Molenaar, I. W. (eds) Rasch Models, Recent Developments and Applications, pp.7-15. Springer-Verlag, New -York (1995)
29. Mokken, R.J.: A theory and procedure of scale analysis. Mouton, The Hague (1971)
30. Nolan, B., Whelan, C.T.: Resources, deprivation and poverty. Clarendon Press, Oxford (1996)
31. Ostini, R., Nering, M.: Polytomous item response theory models. Quantitative Applications in the Social Sciences **07-144**, SAGE Publications series (2006)

32. Rabe-Hesketh, S., Skrondal, A., Pickles, A.: GLLAMM Manual. University of California, Berkeley, Division of Biostatistics, Working Paper **160** (2004)
33. Raileanu Szeles, M.: The Forms and Determinants of Social Exclusion in the European Union: The Case of Luxemburg. Int. Adv. Econ. Res. **14(4)**, 369-380 (2008)
34. Rasch, G.: An individualistic approach to item analysis. In: Lazarsfeld, P., Henry, N. (eds.) Reading in mathematical social science. MIT Press, Cambridge (1966)
35. Ringen, S.: Direct and indirect measures of poverty. J. Soc. Pol. **17**, 351-366 (1988)
36. Sen, A.K.: Issues in the Measurement of Poverty. Scand. J. Econ. **81**, 285-307 (1979)
37. Skrondal, A., Rabe-Hesketh, S.: Generalized Latent Variable and Modeling: Multilevel, Longitudinal, and Structural Equation Models. Chapman & Hall/CRC, Boca Raton (2004)
38. Stewart, F., Ruggeri, C., Laderchi, C., Saith, R.: Introduction: Four Approaches to Defining and Measuring Poverty. In: Stewart, F., Saith, R., Harriss-White, B. (eds) Defining poverty in the developing world. Palgrave MacMillan, Houndmills (2007)
39. Townsend, P.: Deprivation. J. Soc. Pol. **16 (1)**, 125-146 (1987)
40. van der Linden, W.J., Hambleton, R.K.: Item response theory: Brief history, common models, and extensions. In: van der Linden, W.J., Hambleton, R.K. (eds) Handbook of Modern Item Response Theory, pp. 1-28. Springer, New York (1997)
41. Whelan, B.: Non-monetary indicators of poverty. In: Berghman, J., Cantillon, B. (eds) The European Face of Social Security: Essays in honor of Herman Deleeck, 24-42. Avebury (1993)
42. Whelan, C., Layte, R., Maître, B., Nolan, B.: Income, deprivation and economic strain: an analysis of the European Community Household Panel. Eur. Sociol. Rev. **17(4)**, 357-472 (2001)
43. Wilson, M.: On choosing a model for measuring. Meth. Psych. Res. **8(3)**, 1-22 (2003)
44. Zheng, X., Rabe-Hesketh, S.: Estimating parameters of dichotomous and ordinal item response models with gllamm. The Stata Journal **7(3)**, 313-333 (2007)

ANNEX

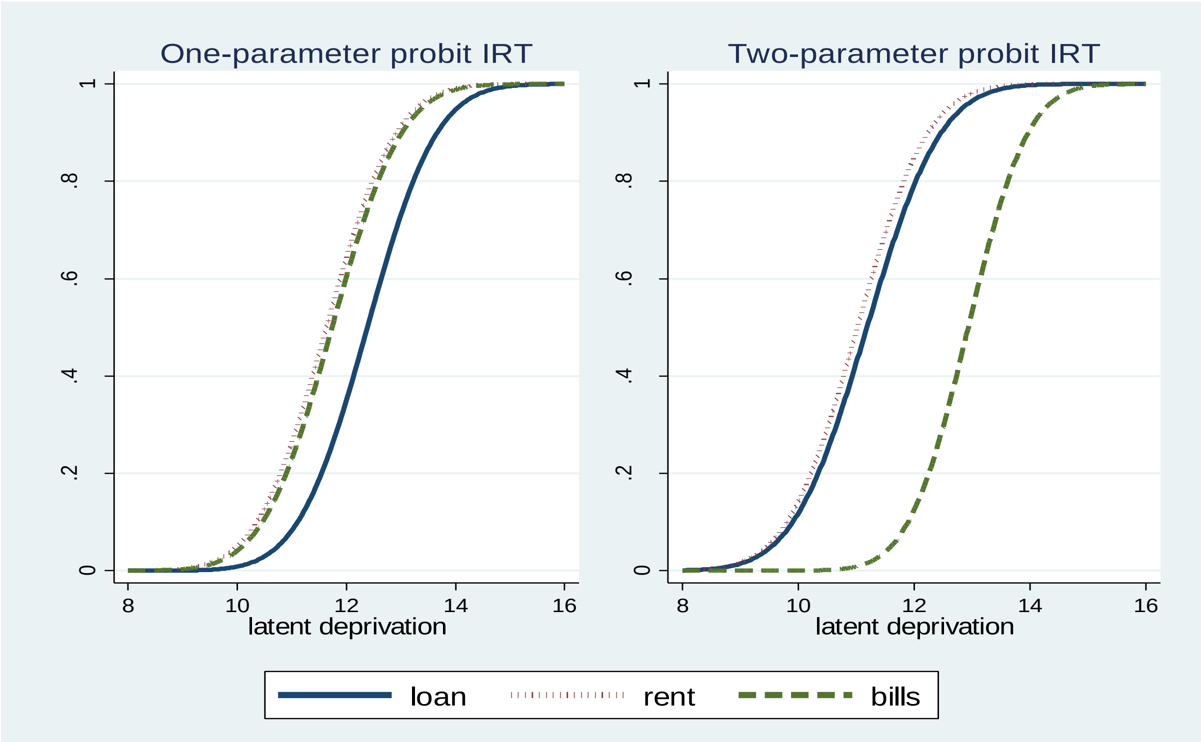
Table A1. The estimates of the probit IRT models: Scale of financial stress

	One-parameter IRT		**Two-parameter IRT**	
	Coef.	(SE)	Coef.	(SE)
Difficulty parameter (β_i)				
loan	12.38	1.09	11.18	1.98
rent	11.63	1.04	11.02	2.13
bills	11.74	1.04	12.93	2.21
Discrimination parameter (λ_j)				
loan	1		1.00	fixed
rent	1		1.05	0.28
bills	1		1.24	0.33
Variance D_i^*	49.95	9.89	40.24	15.52
Log likelihood	-735.78		-735.42	

Source: PSELL-3, CEPS/INSTEAD, 2006, authors' computation

Notes. All coefficients are significant at a level of 1%; Likelihood ratio test of the two-parameter IRT model over the one-parameter IRT model: LR chi2 (2) = 0.73, Prob > chi2 = 0.69.

Figure A1. Item characteristic curves for the scale of financial stress



Source: PSELL-3, CEPS/INSTEAD, 2006, authors' computation

Table A2. The determinants of financial stress

	one-parameter IRT			two-parameter IRT		
	coefficient		SE	coefficient		SE
female	0.08		0.33	0.10		0.44
age	0.11		0.08	0.15		0.12
age square	0.00		0.00	0.00		0.00
Portuguese	0.74	*	0.38	1.04		0.66
other nationality	0.44		0.37	0.60		0.54
lone family	-0.05		0.55	-0.04		0.74
large family	0.52		0.42	0.68		0.62
unemployed	2.52	***	0.52	3.34	*	1.30
inactive	0.81		0.42	1.10		0.70
log income	-2.22	***	0.34	-3.01	*	-1.17
constant	13.58	***	3.35	18.51	*	8.08
log likelihood	-669.10			-668.40		

Source: PSELL-3, CEPS/INSTEAD, 2006, authors' computation

Notes. 1) One-parameter IRT: var(D_i^*): 9.93 (SE: 2.38); Two-parameter IRT: var(D_i^*): 17.92 (SE: 13.16)

2) * p<0.05, ** p<0.01, *** p<0.001

IRISS Working Papers

The IRISS Working Paper Series has been created in 1999 to ensure a timely dissemination of the research outcome from the IRISS-C/I programme. They are meant to stimulate discussion and feedback. The working papers are contributed by CEPS/INSTEAD resident staff, research associates and visiting researchers.

The fifteen most recent papers

Raileanu Szeles M. & Fusco A., 'Item response theory and the measurement of deprivation: Evidence from PSELL-3', IRISS WP 2009-05, May 2009.

Bazyl M., 'Factors influencing tenure choice in European countries', IRISS WP 2009-04, April 2009.

Valentova M., 'Employment breaks due to childcare in the Czech Republic. Before and after 1989. ', IRISS WP 2009-03, March 2003.

Sohn C., Reitel B. & Walther O., 'Cross-border metropolitan integration in Europe (Luxembourg, Basel and Geneva)', IRISS WP 2009-02, February 2009.

Zhelyazkova N. & Valentova M., 'Womens Perceptions of Consequences of Career Interruption due to Childcare in Central and Eastern Europe', IRISS WP 2009-01, January 2009.

Pi Alperin M., 'A comparison of multidimensional deprivation characteristics between natives and immigrants in Luxembourg', IRISS WP 2008-14, December 2008.

Lincaru C., Predosanu G. & Brinza R., 'Process of transition from school-to-work: generator for the initial stage of path dependence in career development', IRISS WP 2008-13, December 2008.

Pavlopoulos D. & Fouarge D., 'Escaping low pay: do male labour market entrants stand a chance?', IRISS WP 2008-12, December 2008.

Abatemarco A., 'Measuring Inequality of Opportunity through Between-Group Inequality Components', IRISS WP 2008-11, November 2008.

Nichols A., 'Trends in Income Inequality, Volatility, and Mobility Risk', IRISS WP 2008-10, November 2008.

IRISS-C/I 10th Anniversary Conference on 'The measurement of discrimination, inequality and deprivation: Recent developments and applications: Book of abstracts', IRISS WP 2008-09, November 2008.

Ayllón S., 'Modelling poverty transitions in Spain: Do attrition and initial conditions really matter? ', IRISS WP 2008-08, October 2008.

Neocleous T. & Portnoy S., 'A Partially Linear Censored Quantile Regression Model for Unemployment Duration', IRISS WP 2008-07, September 2008.

Kankara M. & Moors G., 'Measurement Equivalence and Extreme Response Bias in the Comparison of Attitudes across Europe', IRISS WP 2008-06, May 2008.

Prejmerean M. & Vasilache S., 'What's a university worth? Changes in the lifestyle and status of post-2000 European Graduates.', IRISS WP 2008-05, February 2008.

Electronic versions

Electronic versions of all IRISS Working Papers are available for download at
http://www.ceps.lu/iriss/wps.cfm

IRISS-C/I is a visiting researchers programme at CEPS/INSTEAD, a socio-economic policy and research centre based in Luxembourg. It finances and organises short visits of researchers willing to undertake empirical research in economics and other social sciences using the archive of micro-data available at the Centre.

What is offered?

In 1998, CEPS/INSTEAD has been identified by the European Commission as one of the few *Large Scale Facilities* in the social sciences, and, since then, offers researchers (both junior and senior) the opportunity to spend time carrying out their own research using the local research facilities. This programme is currently sponsored by the European Community's 6th Framework Programme. Grants cover travel expenses and on-site accommodation. The expected duration of visits is in the range of 2 to 12 weeks.

Topics

The major resource offered to visitors is access to a series of internationally comparable longitudinal surveys on living conditions at the household and individual level. The anonymised micro-data provide information on wages and income, health, education, employment and professional activities, accommodation, social relations,... Comparable micro-data are available for EU countries, Central European countries, as well as the USA. These data offer opportunities to carry out research in fields such as *survey and panel data methodology*, *income distribution and welfare*, *income and poverty dynamics*, *multi-dimensional indicators of poverty and deprivation*, *gender, ethnic and social inequality*, *unemployment and labour supply behaviour*, *education and training*, *social protection and redistributive policies*, *fertility and family structures*, *new information technologies in households and firms*, ...

Who may apply?

All individuals (doctoral students as well as experienced academics) conducting research in an institution within the EU-25 or an FP6 Associated State. IRISS-C/I can be meeting place for groups of researchers working on a joint project. We therefore encourage joint proposals by two or more researchers.

For more detailed information and application form, please consult our website: http://www.ceps.lu/iriss or contact us at

IRISS-C/I, CEPS/INSTEAD
BP 48, L-4501 Differdange, G.-D. Luxembourg
Tel: +352 585855 610; Fax: +352 585588
E-mail: iriss@ceps.lu