SUMMARIZING MULTIPLE DEPRIVATION INDICATORS

Lorenzo Cappellari
(Università Cattolica di Milano and ISER, University of Essex)

and

Stephen P. Jenkins
(ISER, University of Essex)

ISER Working Paper
2006-40

Acknowledgement:

The research was supported by ISER's funding from the ESRC and the University of Essex. Earlier versions of the paper were presented at the IARIW 2004 and ECINEQ 2005 conferences and to seminars at the Work and Pensions Economists Group and DIW Berlin. We are grateful for comments and suggestions from Richard Berthoud, Bruce Bradbury, Mark Bryan, Steve McKay, John Micklewright, Brian Nolan, Lucinda Platt, Sophia Rabe-Hesketh, and Brendan Whelan, but responsibility for the views expressed lies entirely with the authors.

A version of this paper will appear as a chapter in *Inequality and Poverty Re-Examined*, edited by Stephen P. Jenkins and John Micklewright, Oxford University Press, Oxford, 2007.

Readers wishing to cite this document are asked to use the following form of words:

Cappellari, Lorenzo, and Jenkins, Stephen P. (August 2006). 'Summarizing multiple deprivation indicators', ISER Working Paper 2006-40. Colchester: University of Essex.

The on-line version of this working paper can be found at http://www.iser.essex.ac.uk/pubs/workpaps/

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and
- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Institute for Social and Economic Research, University of Essex, Wivenhoe Park,
Colchester. Essex CO4 3SQ UK
Telephone: +44 (0) 1206 872957 Fax: +44 (0) 1206 873151 E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

Deprivation scales derived from multiple, typically dichotomous, indicators, are widely used to monitor households' standards of living, and to complement measures of living standards based on income. We use an item response modelling (IRM) framework to address several issues concerning the derivation of deprivation scales in general and the use of sum-score deprivation indices in particular. Although we favour the IRM approach over the sum-score one in principle, we find in an illustrative analysis of basic lifestyle deprivation in Britain in the mid-1990s that both approaches provide very similar pictures of households' circumstances. We conclude with further discussion of the relative merits of the two approaches and highlight some topics for future research.

NON-TECHNICAL SUMMARY

It is widely agreed nowadays that being poor does not simply mean not having enough money. It means, more generally, a lack of access to resources enabling a minimum style of living and participation in the society within which one belongs – as in the definition of poverty adopted by the European Union, for example. In short, poverty is not only about low income, but also about deprivation.

These are not simply academic concerns. Assessments of deprivation are fundamental parts of national anti-poverty strategies in several countries. Summary indices of deprivation are used in combination with measures of low income to produce pictures of ‘consistent poverty’ in the National Action Plan Against Poverty and Social Exclusion in Ireland. In the UK, progress towards the eradication of child poverty is to be monitored not only using income poverty measures but also with measures of ‘material deprivation’. Deprivation indicators are included in the main EU surveys for social monitoring, i.e. the European Community Household Panel and the EU-SILC surveys, and are part of a wider portfolio of social indicators being developed at a European level.

This paper examines some methodological issues concerning the construction of a deprivation scale from multiple deprivation indicators, issues that have received little attention in the deprivation literature. We draw on the literature on item response modelling from psychometrics and educational testing as it has a long history of addressing similar measurement issues. Deprivation indicators are like test scores (i.e. whether an answer to a particular test question is right or wrong), and summarising deprivation indicators with a deprivation scale is like summarising test scores with a scale of academic ability. Our particular interest is in assessing the ubiquitous practice of constructing a deprivation scale as a raw (or weighted) sum of a relatively small set of dichotomous indicators.

We argue that the theoretical foundations of these ‘sum-score’ scales are relatively weak and that the item response modelling approach provides a more promising way to summarize multiple deprivation indicators. An application based on British Household Panel Survey data is used to illustrate the arguments.

We focus on ‘basic life style’ deprivation, summarized using seven binary indicator variables. The first six variables summarize responses to questions put to the household reference person asking whether he or she would like to be able to *PHRASE* but must do without *PHRASE* because they cannot afford it (an ‘enforced lack’), where *PHRASE* refers to:

- Keep your home adequately warm
- Eat meat, chicken, fish every second day
- Buy new, rather than second hand, clothes
- Have friends or family for a drink or meal at least once a month
- Replace worn out furniture
- Pay for a week’s annual holiday away from home

Each variable was scored one if there was an enforced lack of the relevant item or activity and zero otherwise; the percentage in parentheses is the fraction of the sample with an enforced lack. The seventh binary indicator variable summarized difficulties in meeting housing costs: i.e. whether the responding household

- had any difficulties paying for their accommodation in the last twelve months

Those reporting payment problems scored one on this variable; otherwise it was zero.

As it happens, both the item response modelling and the sum-score approaches provide very similar pictures of the patterns of basic lifestyle deprivation and their determinants, and so our results might be construed as providing an empirical rationale for the sum-score approach. We address this issue in the final sections of the paper, where we combine further discussion of the relative merits of sum-score and item response modelling approaches with suggestions of ways in which the latter approach could be developed further.

Introduction

It is widely agreed nowadays that being poor does not simply mean not having enough money. It means, more generally, a lack of access to resources enabling a minimum style of living and participation in the society within which one belongs – as in the definition of poverty adopted by the European Union, for example.[1] In short, poverty is not only about low income, but also about deprivation. The emphasis on deprivation reflects, in part, theoretical concerns that low income provides an 'indirect' measure rather than a 'direct' measure of poverty, as emphasized by Ringen (1988). In addition, there are more purely empirical concerns about an exclusive focus on low income. The snapshot picture provided by income measures from cross-section surveys may be misleading because, with income smoothing, current living standards may not reflect current income, and, in any case, there may be substantial measurement errors particularly at the bottom end of the income distribution. A large body of research has pointed out that the people who have a low income are not the same as the population who are most materially deprived: see inter alia Berthoud *et al.* (2004), Bradshaw and Finch (2003), Callan *et al.* (1993), and Perry (2002).

These are not simply academic concerns. Assessments of deprivation are fundamental parts of national anti-poverty strategies in several countries. Summary indices of deprivation are used in combination with measures of low income to produce pictures of 'consistent poverty' in the National Action Plan Against Poverty and Social Exclusion in Ireland (http://www.socialinclusion.ie/poverty.html). In the UK, progress towards the eradication of child poverty is to be monitored not only using income poverty measures but also with measures of 'material deprivation' (Department for Work and Pensions, 2003). Deprivation indicators are included in the main EU surveys for social monitoring, i.e. the European Community Household Panel and the EU-SILC surveys, and are part of a wider portfolio of social indicators being developed at a European level. See Atkinson *et al.* (2002) and Eurostat (2005).

This paper examines some methodological issues concerning the construction of a deprivation scale from multiple deprivation indicators, issues that have received little attention in the deprivation literature. We draw on the literature on item response modelling from psychometrics and educational testing as it has a long history of addressing similar

[1] 'Persons whose resources (material, cultural and social) are so limited as to exclude them from the minimum acceptable way of life in the Member State to which they belong' (EEC 1985).

measurement issues.[2] Deprivation indicators are like test scores (i.e. whether an answer to a particular test question is right or wrong), and summarizing deprivation indicators with a deprivation scale is like summarizing test scores with a scale of academic ability. Our particular interest is in assessing the ubiquitous practice of constructing a deprivation scale as a raw (or weighted) sum of a relatively small set of dichotomous indicators.

We argue that the theoretical foundations of these 'sum-score' scales are relatively weak and that the item response modelling approach provides a more promising way to summarize multiple deprivation indicators (Section 1). An application based on British Household Panel Survey data is used to illustrate these points (Section 2). As it happens, both approaches provide very similar pictures of the patterns of deprivation and their determinants, and so our results might be construed as providing an empirical rationale for the sum-score approach. We address this issue in the final sections of the paper, where we combine further discussion of the relative merits of sum-score and item response modelling approaches with suggestions of ways in which the latter approach could be developed further (Sections 3 and 4).

We are concerned with what Atkinson (2003) referred to as the 'counting' approach to deprivation. His cogent analysis discusses it from the perspective of social welfare measurement, considering the configurations of deprivation indicators that would allow one to say that deprivation is higher in one case than another for complete classes of summary indices – a dominance approach. By contrast, we consider the derivation of particular summary indices of deprivation and use statistical measurement models to provide the framework for assessing them.[3] Both Atkinson's (2003) and our approach serve to highlight the strong assumptions underpinning the sum-score approach to deprivation scale construction.

We focus on only one set of measurement issues concerning deprivation indicators. Issues such as which deprivation indicators should be included in a sample survey (McKay and Collard 2004), survey methods topics such as question wording, or whether different sets of indicators should be used for different population subgroups (McKay, 2004; Berthoud, Blekesaune and Hancock, 2006) are not considered here. Nor do we consider whether there is

[2] Similar methods were used by Kuklys (2004) to analyze housing and health 'functionings'. Moisio (2004) also related multiple indicators to a latent variable, as we do. The key difference is that our deprivation variable is a continuous one whereas his is discrete: he considers two latent classes – 'poor' and 'non-poor'.
[3] All our statistical models are parametric ones. For a non-parametric approach to related issues, see Spady (2006).

a critical level of deprivation above which households are judged to be in hardship, an issue analogous to the derivation of a poverty line when assessing income poverty.

1. Multiple deprivation indicators and a unidimensional deprivation scale

1.1 From deprivation indicators to a deprivation scale

There are many ways to define and measure 'deprivation', whether overall deprivation or specific dimensions of deprivation, but there are features common to them all:

- *multiple indicators* – the picture of household circumstances is based on multiple indicators of lack or possession of necessities (by contrast, income poverty is summarized using only one indicator);
- *combined into a single scale* – lack or possession of each item or activity (usually recorded as a zero or one in the indicator variables) is aggregated into a numerical scale (a simple or weighted sum).

Most derivations of scales of *overall deprivation* are inspired by and derive from Townsend's (1979) approach to poverty measurement. This was later refined in the Breadline Britain studies (Mack and Lansley, 1985; Gordon and Pantazis, 1997) and by Gordon *et al.* (2000).[4] In these studies, the multiple binary indicators refer to whether households lack various items and activities that are perceived as necessities and their lack is because they cannot afford them rather than because they do not want them, i.e. an 'enforced lack'. Examples of the indicators include 'having heating to warm living areas of the home', to 'able to visit friends and family', and 'having meat, fish or vegetarian equivalent every other day'. The Policy Studies Institute index of overall 'hardship' is similar in structure, except that it uses a prevalence-weighted sum of indicators rather than a simple unweighted sum (Vegeris and McKay, 2002; Vegeris and Perry, 2003).

Other studies have developed separate measures to summarize each of a number of *separate dimensions of deprivation*. For example, the ESRI Dublin research team have developed scales of basic life-style deprivation, secondary lifestyle deprivation, housing deprivation, and so on: see, for example, Nolan and Whelan (1986a, 1986b), Layte *et al.*

[4] Deprivation scales of the type considered in this paper are primarily a European phenomenon. We know of no similar US studies, for instance. There are US studies of material hardship and income: see e.g. Mayer and Jencks (1989).

(2001*a*, 2001*b*), and Whelan *et al.* (2001). (They have studied deprivation in Ireland and compared deprivation across EU countries.) A UK application using their methods is Calandrino (2003). The indices of material well-being and of accommodation and housing conditions developed by the Policy Studies Institute have a close familial resemblance (Vegeris and McKay, 2002; Vegeris and Perry, 2003). Although the measures cited each focus on different dimensions of deprivation, they are constructed in the same way as the measures of overall deprivation: multiple indicators are combined into a single numerical scale.

To simplify the arguments, we shall begin by assuming that one is interested in a single dimension of deprivation, call it 'basic lifestyle' deprivation. We do not observe basic lifestyle deprivation – it is a latent variable – but wish to make inferences about its distribution from a set of K dichotomous deprivation indicators observed for each of N households.[5] In practice, K is relatively small (often less than 10) and N is relatively large (several thousand).

The most commonly used deprivation scale is the sum of the dichotomous indicators. This 'sum-score' index D_i is

$$D_i = \sum_k I_{ik} \tag{1}$$

for each household $i = 1, \ldots, N$, and for each deprivation indicator I_{ik}, $k = 1, \ldots, K$. Alternatively, the sum-score index may be created as a weighted sum, $\sum_k w_k I_{ik}$. With prevalence weighting, for example, a higher weight (w_k) is given to an indicator for which the prevalence in the population is lower. (If few people in the population do not have an item, then arguably its lack should contribute less to overall deprivation.) We focus on (1).

The rationale for using the sum-score D_i as a deprivation scale is rarely considered. The view that is implicit in most studies is, we suspect, that the sum-score index is consistent with the classical measurement model:[6]

$$I_{ik} = D_i^* + \varepsilon_{ik}, \tag{2}$$

where D_i^* is the underlying 'true' but latent measure of deprivation and ε_{ik} is a measurement error term with zero mean, assumed to be independent of D_i^*, and mutually independent. The model implies that the average of the observed indicators for each household is equal to

[5] We refer to households as the unit of analysis as the deprivation indicators are typically collected in surveys using questions directed at one person who responds on behalf of the household as a whole. We assume that the choice of the indicators has already been resolved. The number of indicators for each household may in fact vary because of survey item response. We return to this issue in Section 4.

[6] For an authoritative discussion of measurement models in the psychometric literature, see Nunnally and Bernstein (1994).

$D_i^* + (1/K)\sum_k(\varepsilon_{ik})$. With sufficiently large K, the sample mean of the equation errors would tend to zero, so that the arithmetic average of the observed indicators for each household would equal the household-specific latent deprivation level. The sum-score which is what is typically used in practice – the total score rather than the average – preserves the ranking of households by D^*.

The problem with rationalizing the sum-score in this way is that the classical measurement model cannot hold in the current context because the observed deprivation indicators are dichotomous variables, not continuous ones. One needs an approach that incorporates this fundamental characteristic of the data. Item response models (IRMs) provide such a framework. How large K is will be an issue that we return to repeatedly.

1.2 One parameter item response models

The simplest IRM is the one parameter model, characterised by the following equations:

$$\begin{gathered} I_{ik}^* = \gamma_k + D_i^* + \varepsilon_{ik}, \\ I_{ik} = 1 \;\text{ if } I_{ik}^* > 0 \text{ and } I_{ik} = 0 \text{ otherwise.} \end{gathered} \tag{3}$$

The error terms, ε_{ik}, are independently distributed with mean zero, and have a fixed and common variance. The data structure corresponds to what economists would recognize as a balanced panel except that the repeated observations per household come from the different indicators rather than from different points in time. Model specification is completed by assumptions about the functional form for the distribution of the error terms (for example whether logistic or normal) and whether the household-specific measures of latent deprivation should be treated as a set of fixed parameters or as random effects.

The larger that γ_k is, the more likely that the value of the corresponding indicator I_{ik} is one given any level of deprivation D_i^*. Therefore each γ_k can be straightforwardly interpreted as representing the intrinsic cheapness of the indicator, expressed in latent deprivation terms.[7] Households are less likely report the lack of items that have smaller γ_k, other things being equal. The model also implies that the larger that a household's deprivation is, the greater the probability that each of the observed binary indicators equals one and, moreover, the effect of increasing D_i^* is the same for every item.

[7] In the item response modelling literature, $-\gamma_k$ summarizes the 'item difficulty' of a binary test score item in which a correct answer scores one and an incorrect answer scores zero.

The Rasch model is the one-parameter IRM arising when the error term has a logistic distribution and the D_i^* are treated as fixed effects. In this model, the observed sum-score D_i is a sufficient statistic. That is, given D_i, the *pattern* of responses on the K indicators provides no further information about D_i^*. All units with the same D_i have the same D_i^*.

But can one actually estimate D_i^* given information on K observed indicators? It is well known that conditional maximum likelihood methods are able to provide estimates of each γ_k as $N \rightarrow \infty$, given K fixed, but the D_i^* parameters cannot be estimated. In addition, standard maximum likelihood estimates of the D_i^* parameters are inconsistent as $N \rightarrow \infty$, given K fixed. Consistency requires $N \rightarrow \infty$, $K \rightarrow \infty$, and $N/K \rightarrow \infty$ (Mollenaar 1995), and yet the number of indicators is typically small. Intuitively, the problem is that, as far as the estimation of each D_i^* is concerned, the relevant sample size is the number of indicators, K. This number is usually small.

The standard way forward is to assume, instead, that the D_i^* are random individual effects. In this case, standard maximum likelihood methods may be used to estimate each of the intrinsic cheapness parameters γ_k. The main advance is that, in addition, one can derive predicted values for each D_i^* using 'empirical Bayes' (EB) methods.[8] The intuition is that one gets a good fix on each household's D_i^* by updating the information about the assumed shape of the latent variable distribution (the 'prior') using the information about household's observed responses and the item response parameters. The predicted deprivation score for each household is the expected value of this updated ('posterior') distribution. Put another way, to predict the latent variable for the given household, one combines the observed responses for a given household with the assumptions of the model relating observed indicators to the latent variable for every household. The 'empirical' tag arises because one does the predictions using sample estimates of the parameters (γ_k), rather than their true values, which are not observed. The EB predictor also has a nice interpretation of minimizing the mean square error of prediction over the sampling distribution of the responses taking the model parameters as known.

With EB prediction of latent deprivation, one has a more secure methodological foundation, with deprivation scales consistently founded on a measurement model. This is a substantial advantage. There are several points to note, however. First, the small-sample properties of EB predictors from IRMs are not well-known (Hoijtink and Boosma, 1995).

[8] See Skrondal and Rabe-Hesketh (2004, Chapter 7) or Hoijtink and Boosma (1995) for discussions of EB methods.

The relevant sample size is the number of deprivation indicators. Intuitively speaking, the larger that K is, the more information one has, and hence the better the prediction. Second, the orderings of households in terms of EB predictions and sum-scores are likely to be closely related, since the probability that a household is counted as deprived according to each and every observed indicator is an increasing function of D_i*. However, the association is not perfect (and also likely to vary with K). For any given sum-score value, there will be a distribution of EB predictions of D_i* because the same sum-score may be achieved from different combinations of indicator scores. We illustrate this later.

The third point is that the one parameter IRM incorporates strong assumptions that are likely to be unrealistic. For example, in the one parameter random effects probit IRM, the correlation between any pair of item deprivations is the same, regardless of which pair is considered: $\text{corr}(I_{ik}^*, I_{im}^*) = \rho$, for all $k \neq m$, where $\rho = \text{var}(D_i^*)/[1 + \text{var}(D_i^*)]$. This strong assumption may be tested using a multivariate probit model in which no restrictions are placed on the cross-equation correlations: $\text{corr}(I_{ik}^*, I_{im}^*) = \rho_{km}$. See Section 2. A more common way of avoiding the equi-correlation assumption is to incorporate additional parameters into the IRM. We consider this and other generalizations to the IRM specification next.

1.3 Two parameter item response models and other specification issues

The two parameter IRM weakens the assumption that a given change in D_i^* has the same impact on each deprivation indicator probability. This is done by introducing indicator-specific 'discrimination' parameters, otherwise known as 'factor loadings' into the one parameter random effects IRM:[9]

$$\begin{gathered} I_{ik}^* = \gamma_k + \lambda_k D_i^* + \varepsilon_{ik}, \\ I_{ik} = 1 \text{ if } I_{ik}^* > 0 \text{ and } I_{ik} = 0 \text{ otherwise.} \end{gathered} \tag{4}$$

For model identification, it is usually assumed that $\lambda_1 = 1$. The equi-correlation assumption no longer holds, since $\text{corr}(I_{ik}^*, I_{im}^*)$ is a function of λ_k and λ_m. The parameter estimates can be estimated by maximum likelihood, and one can derive estimates of D_i* by EB methods, subject to the caveats mentioned earlier.

[9] In the item response modelling literature, λ_k summarizes the extent to which the item (question) differentiates between individuals with different levels of academic ability (D_i*).

One of the emerging themes of the paper is the importance of having a relatively large number of deprivation indicators. One catch to this is that the more indicators that you use, the less likely it is that they refer to a single latent deprivation trait. This issue may be illustrated with reference to the research of inter alia Whelan *et al.* (2001). Using 24 deprivation indicators, they applied confirmatory factor analysis to identify three dimensions of deprivation: basic life-style deprivation, secondary lifestyle deprivation, and housing deprivation. Then they used a separate sum-score index to summarize deprivation within each dimension. IRMs such as (4) can be straightforwardly extended from being one factor models, as in (3) and (4), to having two or more factors (Goldstein, 1980; Skrondal & Rabe-Hesketh, 2004). The advantage of following the IRM approach is that both the specification of the number of factors, and the relationship between deprivation indicators and factors, are encompassed within a single model-based framework, and not split into two separate and potentially inconsistent steps.

IRMs also provide a consistent way in which to incorporate heterogeneity in household characteristics into the analysis, both in terms of modelling observed responses, and for exploring the determinants of latent deprivation itself. We consider these two aspects in turn.

First, we observe that binary deprivation indicators are typically derived from a two-part question. The first part asks whether the household has an item or participates in some activity and, if the response is negative, the second part asks whether the lack was because it could not be afforded. If the answer to this second part is yes, then the deprivation indicator scores one, and is zero otherwise. (Specific examples are provided in Section 2.) It is conceivable that there are systematic differences in observed responses because, even among households people with the same latent deprivation D_i*, there are heterogeneous views about what they 'want', about what they understand by affordability, or about the interpretation of specific questions (for example relating to what 'adequate' means). For example, some people may give greater priority to a warm home than to having friends around, and this may be reflected in their responses to whether they cannot afford something that they do not have.

In principle, it is straightforward to introduce covariates into the IRM to address this issue. For example, one may rewrite (4) as follows:

$$I_{ik}^* = \gamma_k + \lambda_k D_i^* + \beta_k X_{ik} + \varepsilon_{ik}, \qquad (5)$$
$$I_{ik} = 1 \text{ if } I_{ik}^* > 0 \text{ and } I_{ik} = 0 \text{ otherwise.}$$

Non-zero values of β_k indicate differential reporting propensities, or what is known as 'item bias' or 'differential item functioning' in the IRM literature. From this perspective, one may interpret the deprivation indicator regressions of Desai and Shah (1988) as being estimates of a one parameter IRM allowing for item bias but also assuming all cross-equation error correlations were equal to zero.

The IRMs discussed earlier can also be extended to model the determinants of the latent deprivation trait jointly with the estimation of the IRM parameters. The measurement component of the model is supplemented with a 'structural' equation of the form:

$$D_i^* = \alpha' Z_i + \xi_i, \tag{6}$$

where ξ_i is a normally distributed i.i.d. error term with mean zero and fixed variance. This is an example of a multiple-indicator multiple-cause (MIMIC) model. One of the issues that we consider in the empirical illustration to follow is whether the conclusions that one would draw about the impact of covariates on deprivation differs depending on whether they are derived from an IRM model supplemented with eqn. (6), or the conventional approach of regressing sum-scores on covariates.

Some people commenting on our research have objected to the incorporation of item bias parameters as in eqn. (5), stating that this conflates two distinct activities: the measurement of deprivation, on the one hand, and analysis of the determinants of deprivation, on the other hand. Their argument is that the level of deprivation should be assessed entirely in terms of deprivation indicator response patterns, and so characteristics should not play a role in the measurement model.

Our view is that there is an important distinction between analysis of the determinants of observed deprivation indicators (D_i), analysis of the determinants of the latent deprivation variable (D_i^*), and estimation of D_i^*.[10] Item bias refers to the first of these issues, i.e. how different people with the same latent deprivation may report different indicator prevalence, and the structural equation (6) is the framework for addressing the second issue. In principle, estimation of D_i^* – the third issue – may be achieved using EB methods applied to models incorporating item bias and an equation for the determinants of latent deprivation. The problem is that, in practice, it is difficult to estimate models that incorporate both item bias and a structural equation. Often the same characteristic appears in both parts (X_{ik} and Z_i have

[10] Analysis of how the observed responses on the indicator variables vary with characteristics is of interest in its own right, of course, quite separately from interest in underlying deprivation (the focus here).

common elements), and it is difficult to identify to identify its separate contributions from a statistical point of view. We ignore item bias from now on for this reason.[11]

To sum up so far, we have argued that an IRM approach provides a coherent approach to the derivation of a deprivation scale, and that its methodological foundations are more secure than those of the commonly-used sum-score approach. In the next section, we contrast the two approaches in an empirical illustration.

2. Empirical illustration: basic life-style deprivation in Britain

2.1 Data

We used data from wave 6 (survey year 1996) of the British Household Panel Survey (BHPS). The advantages of the BHPS data are that they are based on a large national population sample and (from wave 6 onwards) have contained a battery of questions about deprivation in addition to more conventional indicators of household living standards such as income. We used wave 6 data rather than some later year to minimize any potential impact of panel attrition on sample selection.

We focus on 'basic life style' deprivation, summarized using seven binary indicator variables. The first six variables summarize responses to questions put to the household reference person asking whether he or she would like to be able to *PHRASE* but must do without *PHRASE* because they cannot afford it (an 'enforced lack'), where *PHRASE* refers to:

- Keep your home adequately warm (1.9 per cent)
- Eat meat, chicken, fish every second day (3.1 per cent)
- Buy new, rather than second hand, clothes (5.3 per cent)
- Have friends or family for a drink or meal at least once a month (6.5 per cent)
- Replace worn out furniture (13.4 per cent)
- Pay for a week's annual holiday away from home (20.1 per cent)

Each variable was scored one if there was an enforced lack of the relevant item or activity and zero otherwise; the percentage in parentheses is the fraction of the sample with an

[11] An alternative way to address the heterogeneity in response issue would be to estimate different models for different population subgroups, for example elderly people versus younger people, or separately for different minority ethnic groups. This would also provide scope for using different sets of indicators for the different groups. The indicators of material deprivation recently introduced in the UK Family Resources Survey differ for adults and for children (Department for Work and Pensions 2003).

enforced lack. The seventh binary indicator variable summarized difficulties in meeting housing costs: i.e. whether the responding household[12]

- had any difficulties paying for their accommodation in the last twelve months (6.9 per cent)

Those reporting payment problems scored one on this variable; otherwise it was zero.

These seven indicators are representative of those used in the literature. They are a subset of those used by Townsend (1979) and the later Breadline Britain studies. They were introduced to the BHPS when that survey was used to contribute data to the UK component of the European Community Household Panel (ECHP) – the same variables were available on a harmonized basis for all countries in the survey. The list corresponds closely to those used to summarize basic life-style deprivation in the many ECHP-based studies of deprivation by the research team from ESRI Dublin: see inter alia Layte *et al.* (2001) and Whelan *et al.* (2001). See also Eurostat (2005). The indicators overlap with the ten indicators proposed for measurement of adult material deprivation by the UK Department for Work and Pensions (2003).

2.2 Summary statistics

There were 4,859 households with non-missing information on all seven indicators from an overall sample of 5,064 households. Sixty-nine per cent experienced no enforced lack according to any of the seven indicators; put another way, 31 per cent of the sample experienced an enforced lack of at least one item. Fifteen per cent were deprived of two items, and 8 per cent of three items, 4.4 per cent of four items, and 2.8 per cent were deprived of 4–7 items. Only one household was deprived of all seven items. The number of unique response patterns was 88, which is 66 per cent of the total number possible ($128 = 2^7$).

The 'reliability' of a sum-score deprivation scale is often assessed with reference to estimates of the Cronbach alpha statistic (α), even though the theory underlying it refers to a classical measurement model with continuous indicators. (See Nunnally and Bernstein 1994 for further discussion.) The α summarizes the extent to which the indicators in a summative scale are correlate well with each other. If each indicator were statistically independent of each of the other indicators, then $\alpha = 0$. At the other extreme, if all are perfectly correlated with each other, then $\alpha = 1$. Our estimate of α for the sample as a whole is 0.653, which lies

[12] Only individuals renting their accommodation (other than those receiving a 100 per cent rent rebate) or

within the bounds of what is usually considered to be acceptable, though is lower than the 0.81 reported by Whelan *et al.* (2001). At the same time, we should not put too much emphasis on the estimated α because we have discrete rather than continuous indicators.

2.3 Item response model estimates

Estimates of item response models are shown in Table 1.[13] Model 1 is the basic one parameter model. Reassuringly, the ordering of the indicators by the estimates of the 'intrinsic cheapness' parameters (γ_k) corresponds to the ordering by prevalence of enforced lack reported in the previous subsection. For any given level of D_i*, the probability of reporting an enforced lack is lowest for keeping the home adequately warm, and highest for having a week's holiday away. For example, if D_i* = 0, the probability of lacking an adequately warm home is 0.001, and of lacking a week's holiday away is 0.017. If D_i* = 1, the probabilities are 0.100 and 0.390.[14]

<Table 1 near here>

The equi-correlation assumption for the errors incorporated by Model 1 is relaxed in Model 2, and a likelihood ratio test rejects the former in favour of the latter (χ^2(303.8, d.f. = 20), p < 0.001). The ordering of the intrinsic cheapness parameters in terms of relative magnitude is the same, however. The same is true when we move to Model 3, the two parameter model that relaxes the assumption that a given change in D_i^* has the same impact on each deprivation indicator probability. Note also the substantial improvement in log-likelihood relative to the other models. There is substantial variation in the estimated factor loadings (λ_k), with relatively low values for difficulties in paying for accommodation keeping the home adequately warm, and the highest value for replacement of worn out furniture.

From these estimates, we can derive EB predictions of each household's latent deprivation score D_i^*, and see how these compare with their rankings by the sum-score. A comparison based on the two parameter IRM estimates (model 3) is shown in Figure 1. It is clear that the two scales order households in a very similar way. The scatterplot is close to a straight line, and the correlation between the two scales in 0.97. (The corresponding

buying with a mortgage were at risk of an enforced lack.

[13] All IRM parameter estimates and EB predictions were derived using the program modules `gllamm` and `gllapred` in Stata (http://www.gllamm.org). The exception was the multivariate probit version of the one parameter model: see Cappellari and Jenkins (2003).

[14] The precise values of the predicted probabilities depend on the type of binary model (logit, probit, cloglog) that one uses. This issue is discussed by Goldstein (1980).

scatterplot using EB predictions from the one parameter IRM is even more like a straight line.) There is only limited variation in predicted D_i^* scores at each sum-score value. And it is only at sum-scores of 5 and 6 that there are overlaps in predicted D_i^* scores.

<Figure 1 near here>

The close association between the two scales is underlined further when we examine the household type breakdown of the worst-off 30 per cent of the sample. Table 2, columns (a) and (b), show that the composition of this group is the same according to the two scales. The largest group, comprising almost one fourth of this worst-off group, is non-elderly working couples with children.

The determinants of deprivation are examined next. The impact of a set of covariates on latent deprivation D_i* is considered using the two parameter IRM supplemented with the specification shown in eqn. (6). This is compared with estimates from a regression of the sum-score on the same set of covariates, using ordered probit methods. (There are eight sum-score categories.) The regressors used are similar to those used in earlier deprivation studies: the numbers of adults and children in the household, the sex of the household head, the age of the household head and age-squared, whether the household contained at least one full-time worker, and the log of household annual income. See Table 3.

The two modelling approaches yield similar results in the sense that corresponding coefficient estimates have the same sign and are precisely estimated. Deprivation is higher the more adults or the more children there are in the household, or if the household head is a woman. Deprivation is lower for households with at least one full-time worker, and the higher the household income. The magnitudes of the corresponding coefficients are not entirely comparable because the scale of the dependent variables differs: for example, the sum-score ranges from 0 to 7, whereas the range of D_i^* is much smaller (see Figure 1). This explains why the magnitude of each coefficient in the sum-score regression is greater than that for its counterpart in the latent deprivation regression. But one can say that there are close similarities nonetheless. Ratios of coefficient estimates from one model are very similar to corresponding ratios from the other model. For example, the ratio of the estimated coefficient on the number of adults to the estimated coefficient on the number of children is 1.11 in the latent deprivation regression and 1.14 in the sum-score regression. Deprivation has an inverse U-shaped relationship with age in both regressions, with a maximum at age 35 according to the latent deprivation regression, compared to age 32 according to the sum-score regression.

3. Discussion

We have argued the case for an IRM approach to the derivation of deprivation scales from multiple deprivation indicators. Our empirical illustration has shown, however, that in practice, the IRM and conventional sum-score approaches yield very similar pictures of the distribution of deprivation in terms of the association between the distributions of scores, who is found to be worst-off, and also the determinants of deprivation. At one level, then, we have provided an entirely practical argument for the continued use of the sum-score approach. It is very simple to implement and to understand, and appears to provide the same conclusions.

This case for the sum-score approach is not decisive. There are some strong arguments in favour of exploring the IRM approach further in the deprivation context. The approach can handle missing indicator information in a straightforward manner, using what economists would call unbalanced panel methods. (For simplicity we did not use them in this paper.)

In addition, there are intrinsic advantages of using a consistent model-based framework for thinking about measurement. The framework can incorporate models of the relationship between the latent deprivation and explanatory variable, and can also be extended to have more than one latent deprivation variable. This approach contrasts with the two-step one which first uses confirmatory factor analytic methods to identify deprivation variables (even though, strictly speaking, these methods were developed for continuous variables), and then constructs sum-scores for each dimension identified at the first step. More generally, the specification of the IRMs has highlighted the nature of the assumptions underlying the construction of a deprivation scale. In the conventional sum-score approach, these assumptions are left implicit and typically ignored.

We have highlighted the important role played by the number of indicators available for the properties of the measures and estimation. Underlying this point is the common sense idea that there is little information that a small number of dichotomous indicators can communicate about a particular household's circumstances and or help us discriminate between different households. The maximum number of distinct response combinations is only 2^K.

This suggests that the more indicators there are, the better (subject to their being relevant to deprivation, of course). More information about the different circumstances of households might also be gained by using different types of indicators. For example, one could use polytomous variables with ordered categories, or indeed continuous variables. IRMs can be generalised to use combinations of dichotomous, ordered polytomous and continuous indicators, albeit at the cost of additional complexity. See Skrondal and Rabe-Hesketh (2004) for discussion of the principles and Ribar (2005) for an application.

Another way to get additional repeated observations on households is to use panel survey data in which there are responses on the same deprivation items at multiple points in time. The most extensive study to date of deprivation indicators using panel data is that by Berthoud *et al.* (2004), who considered the longitudinal evolution of a sum-score scale calculated at each annual interview. By contrast, we have in mind an extension to the IRM approach that takes explicit account of the repeated observations per household or individuals. In the same way that researchers have argued in favour of using repeated observations on income at each interview to calculate a measure of (unobserved) 'permanent' income, one could use the repeated observations on deprivation indicators over time to get a better measure of latent deprivation.

Precisely what the specification of a 'panel' IRM would look like is unclear, and an interesting topic for future research. (Ribar (2005) is the only related study that we are aware of.) For example, an empirical regularity identified by Berthoud *et al.* (2004) is that there is a decline in average deprivation sum-scores over time as living standards improve – in the same way that income poverty rates decline if the poverty line is fixed in real terms. This led them to standardize their sum-scores: the year-specific average score was deducted from each household's score and the result divided by the year-specific standard deviation (Berthoud *et al.* 2004, chapter 4). From an IRM perspective, one might ask what precisely it is that the passage of time is affecting – is it the intrinsic cheapness parameters or latent deprivation itself that changes over time, or both? If it is the former, then one might think of an IRM estimated from panel data in which there are interview-specific intrinsic cheapness parameters (γ_k varying with calendar time). If it is the latter, then one would incorporate interview-specific factor loading parameters (λ_k varying with calendar time).

4. Concluding remarks

There has been remarkable little discussion of fundamental measurement issues in the deprivation literature of the type that we have considered here, and especially little that takes account of the dichotomous nature of the indicators that are commonly used. In part, this may be because deprivation analysts have considered other measurement issues to have a greater priority for attention, for example the choice of the set of indicators itself, and the precise wording of questions about them in surveys.

We acknowledge that these are important issues. Nonetheless, we would argue that the issues we have raised also deserve some further consideration, especially as deprivation scores are being used increasingly to monitor social progress in national and cross-national contexts. Although we found in our illustrative application that IRM and sum-score approaches provided very similar descriptions of patterns of deprivation and their determinants, this need not be the case outside this setting. And it may partly reflect the small number of indicators in the first place.

There is an interesting contrast with this growing deprivation literature and the extensive literature on international comparisons of educational test scores based on harmonised surveys such as Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TMSS), and Progress in International Reading Literacy Study (PIRLS). In addition to survey issues, pure measurement issues have been given substantial attention, and IRM approaches are much used. One key difference is that these surveys provide a large number of indicators.

References

Atkinson, A. B. ‘Multidimensional deprivation: contrasting social welfare and counting approaches’, *Journal of Economic Inequality*, 1, 51–65.

Atkinson, A. B., Cantillon, B., Marlier, E., and Nolan, B. (2002), *Social Indicators. The EU and Social Exclusion*, Oxford University Press, Oxford.

Berthoud, R., Bryan, M. and Bardasi, E. (2004), *The Relationship between Income and Material Deprivation over Time*, Department for Work and Pensions Research Report 219, Corporate Document Services, Leeds.

Berthoud, R., Blekesaune, M. and Hancock, R. (2006), *Are ‘poor’ pensioners deprived?*, Department for Work and Pensions Research Report 364, Corporate Document Services, Leeds.

Bradshaw, J. and Finch, N. (2003) ‘Overlaps in dimensions of poverty, *Journal of Social Policy*, 32, 513–525.

Calandrino, M. (2003), 'Low-income and deprivation in British families', Working Paper Number 10, Department for Work and Pensions, London.

Callan, T. Nolan, B., and Whelan, C. T. (1993), 'Resources, deprivation, and the measurement of poverty', *Journal of Social Policy*, 22, 141–172.

Cappellari, L. and Jenkins, S. P. (2003), 'Multivariate probit regression using simulated maximum likelihood', *The Stata Journal*, 3, 278–294.

Department for Work and Pensions (2003), *Measuring Child Poverty*, Department for Work and Pensions, London

Desai, M. and Shah, A. (1988), 'An econometric approach to the measurement of poverty', *Oxford Economic Papers* 40: 505–522.

EEC (1985) *On Specific Community Action to Combat Poverty* (Council Decision of 19 December 1984) 85/8/EEC, Official Journal of the EEC, 2/24.

Eurostat (2005), 'Material deprivation in the EU', *Statistics in Focus – Population and Social Conditions*, 21/2005, Eurostat, Luxembourg.

Goldstein, H. (1980), 'Dimensionality, bias, independence and measurement scale problems in latent trait test score models', *British Journal of Mathematical and Statistical Psychology* 33: 234–246.

Gordon, D. and Pantazis, C. (1997), *Breadline Britain in the 1990s*, Ashgate, Aldershot.

Gordon, D., Adeleman, L., Ashworth, K., Bradshaw, J., Levitas, R., Middleton, S., Pantazis, C., Patsios, D., Payne, S., Townsend, P., and Williams, J. (2000), *Poverty and Social Exclusion in Britain*, Joseph Rowntree Foundation, York.

Halleröd, B. (1994), 'A new approach to the direct consensual measure of poverty', Social Policy Research Centre Discussion Paper No 50 (October), University of New South Wales, Sydney.

Hoijtink, H. and Boomsma, A. (1995), 'On person parameter estimation in the dichotomous Rasch model', pp. 53–67, Chapter 4, in: G.H. Fischer and I.W. Mollenaar (eds), *Rasch Models: Foundations, Recent Developments, and Applications*, Springer-Verlag, New York.

Kuklys, W. (2004), 'Measuring standards of living in the UK – an application of Sen's functioning approach using structural equation models', Working Paper on Strategic Interaction 11-2004, Max Planck Institute, Jena.

Layte, R., Nolan, B., and Whelan, C. (2000), 'Targeting poverty: lessons from monitoring Ireland's national Anti-Poverty Strategy', *Journal of Social Policy* 29: 553–575.

Layte, R., Maître, B., Nolan, B. and Whelan, C. T. (2001), 'Explaining levels of deprivation in the European Union', *Acta Sociologica*, 44, 105–122.

Lord, F. M. and Novick, M. R. (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading MA.

Mack, J. and Lansley, S. (1985), *Poor Britain*, George Allen & Unwin, London.

Mayer, S. E. and Jencks, C. (1989), 'Poverty and the distribution of material hardship', *Journal of Human Resources*, 33, 88–114.

McKay, S. (2004), 'Poverty or preference? What do 'consensual deprivation indicators' really measure?', *Fiscal Studies* 25, 201–223.

McKay, S. and Collard, S. (2004), 'Developing deprivation questions for the Family Resources Survey', IAD Research Division Working Paper No. 13, Department for Work and Pensions, London.

Moisio, P. (2004), 'A latent class application to the multidimensional measurement of poverty', *Quality and Quantity*, 38, 703–717.

Mollenaar, I. (1995), 'Estimation of item parameters', pp. 39–51, Chapter 3, in: G. H. Fischer and I. W. Mollenaar (eds), *Rasch Models: Foundations, Recent Developments, and Applications*, Springer-Verlag, New York.

Nolan, B. and Whelan, C. (1996*a*), 'Measuring poverty using income and deprivation indicators: alternative approaches', *Journal of European Social Policy*, 6: 225–240.

Nolan, B. and Whelan, C. (1996*b*), *Resources, Deprivation and Poverty*, Clarendon Press, Oxford.

Nunnally, J. C. and Bernstein, I. H. (1994). *Psychometric Theory*, third edition, McGraw-Hill, New York.

Perry, B. (2002), 'The mismatch between income measures and direct outcome measures of poverty', *Social Policy Journal of New Zealand*, 19, 101–127.

Ribar, D. (2005), 'The persistence of financial strains among low-income families: an analysis of multiple indicators', unpublished paper, Department of Economics, The George Washington University, Washington DC.

Ringen, S. (1988), 'Direct and indirect measures of poverty', *Journal of Social Policy*, 17, 351–365.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, CRC Press, Boca Raton FL.

Spady, R.H. (2006), ‘Identification and estimation of latent attitudes and their behavioral implications’, Working Paper CWP12/06, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, London.

Townsend, P. (1979), *Poverty in the United Kingdom*, Harmondsworth, Penguin.

Vegeris, S. and McKay, S. (2002), *Low/Moderate-income Families in Britain: Changes in Living Standards*, DWP Research Report No. 164, Corporate Document Services, Leeds.

Vegeris, S. and Perry, J. (2003), *Families and Children Study 2001: Report on Living Standards and the Children*, DWP Research Report No. 190, Corporate Document Services, Leeds.

Whelan, C., Layte, R., Maître, B., and Nolan, B. (2001), ‘Income, deprivation and economic strain: an analysis of the European Community Household Panel’, *European Sociological Review* 17: 357–472.

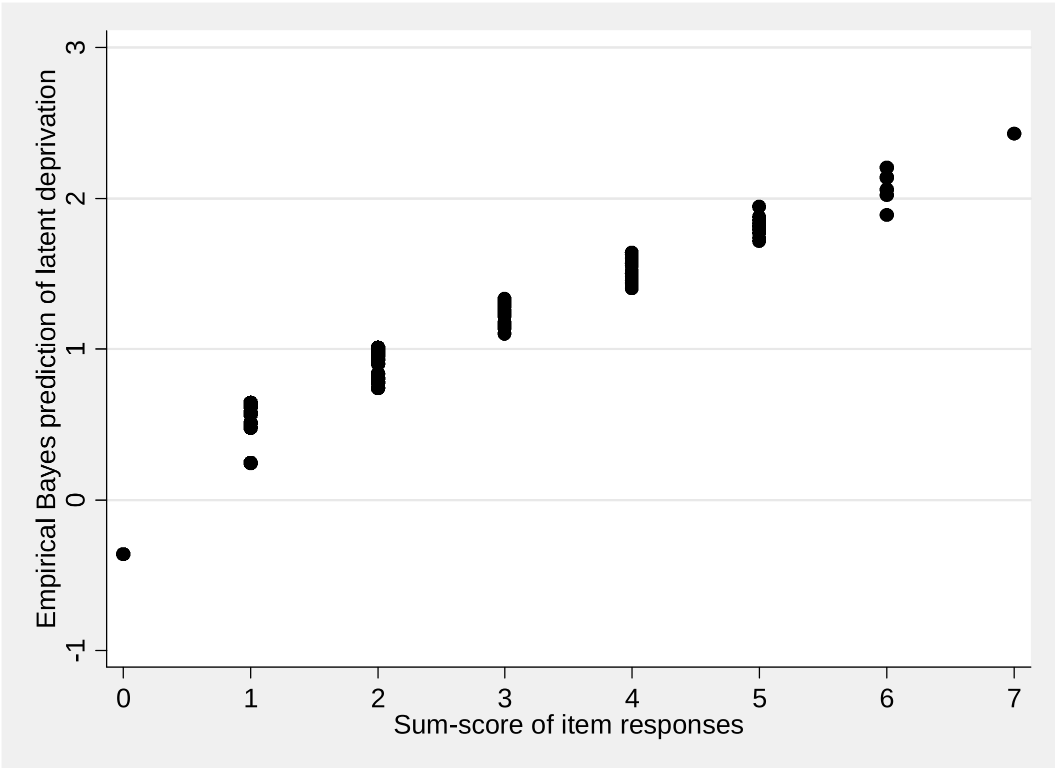
Table 1

Estimates of probit random effects item response models

Indicator	One parameter IRM (1)		One parameter IRM (multivariate probit) (2)		Two parameter IRM (3)			
	γ_k	(SE)	γ_k	(SE)	γ_k	(SE)	λ_k	(SE)
Home adequately warm	–3.12	(0.07)	–2.09	(0.04)	–1.83	(0.06)	1	
Meat etc. every second day	–2.80	(0.06)	–1.86	(0.04)	–1.59	(0.06)	1.34	(0.18)
New rather than second hand clothes	–2.44	(0.05)	–1.62	(0.03)	–1.26	(0.06)	1.54	(0.20)
Friends or family visit at least once a month	–2.29	(0.05)	–1.52	(0.03)	–1.08	(0.05)	1.44	(0.18)
Difficulties paying for accommodation	–2.22	(0.05)	–1.48	(0.03)	–1.07	(0.04)	0.80	(0.10)
Replace worn out furniture	–1.69	(0.04)	–1.11	(0.02)	–0.43	(0.05)	1.62	(0.20)
Week's annual holiday away	–1.28	(0.04)	–0.85	(0.02)	0		1.57	(0.19)
ρ	0.57	(0.01)	[a]		[b]			
	–7517.8		–7669.7		–7473.7			

Notes. [a]: Unrestricted cross-equation error correlations. Likelihood ratio test of model 1 versus model 2: χ^2(303.8, d.f. = 20), $p < 0.001$. [b]: estimates of cross-equation error correlations not shown. Estimate of var(D_i^*) from model 3 is 0.702 (SE = 0.15).

Figure 1
Empirical Bayes and sum-score deprivation scales are highly correlated



Note. EB estimates derived from two-parameter IRM (model 3 in Table 1). The correlation between the two series is 0.97.

Table 2
Composition of the worst-off 30 per cent, by deprivation measure

Column percentages

Household type	Sum-score	Two parameter IRM (a)	Two parameter IRM (b)
Elderly (household head of pension age)			
Single man	2.3	2.3	2.2
Single woman	9.9	9.9	9.8
Couple	6.3	6.3	6.2
Non-elderly			
Single, kids, full-time worker	5.5	5.5	5.3
Single, kids, no full-time worker	11.5	11.5	12.6
Single, no kids, full-time worker	5.8	5.8	5.6
Single, no kids, no full-time worker	9.1	9.1	9.9
Couple, kids, at least one full-time worker	24.4	24.4	23.9
Couple, kids, no full-time worker	7.1	7.1	7.3
Couple, no kids, at least one full-time worker	10.1	10.1	9.5
Couple, no kids, no full-time worker	3.8	3.8	3.9
Other	4.2	4.2	4.1
Total	100.0	100.0	100.0
N (households)	1430	1430	1425

Notes: (a): two parameter random effects IRM without covariates (Model 3 in Table 1). (b) As (a), except model also includes determinants of deprivation (see Table 3) .

Table 3
The determinants of deprivation: two approaches compared

	Dependent variable			
	Latent deprivation (a)		Sum-score (b)	
Regressors	Coeff.	(SE)	Coeff.	(SE)
Number of adults	0.136	(0.03)	0.165	(0.02)
Number of children	0.122	(0.02)	0.145	(0.02)
Female household head	0.170	(0.08)	0.199	(0.04)
Age of household head (years)	0.028	(0.01)	0.032	(0.01)
Age squared	–0.0004	(0.0001)	–0.0005	(0.0001)
One or more full-time workers	–0.219	(0.05)	–0.260	(0.05)
Log(income)	–0.491	(0.06)	–0.579	(0.03)
Constant	3.321	(0.42)		
log*L*	–7029.0		–4292.1	
N (households)	4671		4680	

Notes. (a): Specification based on eqn. (6) embedded in two parameter random effects IRM (other parameter estimates not shown). (b): Ordered probit regression of sum-score on covariates.