How to measure material deprivation?
A Latent Markov Model based approach

Francesco Dotto[1]

Joint work with:
Alessio Farcomeni[2], Maria Grazia Pittau[3] and Roberto Zelli[3]

Trieste, 22/11/2019

[1] Dipartimento di Economia, Università degli studi di Roma Tre
[2] Dipartimento di Economia e Finanza, Università di Roma "Tor Vergata"
[3] Dipartimento di Scienze Statistiche, Università di Roma La Sapienza

Outline

Material Deprivation Measurement

The status of material deprivation is not directly observable.

European Union Commission (2004) definition refers to an enforced lack of commodities and/or dimensions

1. *Social welfare approach* - based on a *suitable* welfare function
2. *Counting approach* - based on counting the number of *dimensions* in which people suffer deprivation.

Furthermore it is intrinsically a relative concept

Material Deprivation *philosophically* speaking

The status of material deprivation is not directly observable.

Furthermore is intrinsically a relative concept

"*By necessaries I understand not only the commodities which are indispensably necessary for the support of life, but whatever the custom of the country renders it indecent for creditable people, even of the lowest order, to be without. A linen shirt, for example [....] a creditable day-laborer would be ashamed to appear in public without a linen shirt ....* ".

Adam Smith, The Wealth of Nations, 1776, vol.II, V.2.148

How does EUROSTAT measure material deprivation?

- $R = 9$ items/attributes households can or cannot afford

  1. to keep home adequately warm;
  2. one week annual holiday away from home;
  3. a meal with meat, chicken and fish or a protein equivalent every other day;
  4. to face unexpected expenses;
  5. a telephone;
  6. a color TV;
  7. a washing machine;
  8. a car;
  9. to pay rent or utility bills (whether the household has arrears).

- **household deprived**: at least 3 out of 9 lacking items
- **household severe deprived** at least 4 out of 9 lacking items

Our proposal

Our proposal consists in implementing a Latent Markov Model [4] for classifying individuals based on their deprivation status

This approach has, in our opinion, two main advantages:

1. Arbitrary thresholds are not needed
2. Allows to classify individuals by their intertemporal deprivation status.

Furthermore we also provide an optimal weighting scheme aimed at reducing the dimensionality of the outcome.

[4] more details in Bartolucci et al. (2012)

Latent Class analysis....why and how

A brief (non exaustive) recap

Latent Class analysis is the cornerstone of many different statistical models.

The common assumption standing these models is the existence of latent *characteristic* which is used to explain unobserved heterogeneity possibly affecting response variables and covariates.

Observed / Latent	Continous	Discrete
Continous	Factor Analysis	Mixture Modelling
Discrete	Item Response Theory	Latent Class Models

A sketch of the model

Introduction

Response vector

Let $Y_{it} = (Y_{it1}, Y_{it2}, \ldots, Y_{itR}) \in [0,1]^R$ with $i = 1, 2 \ldots, n$ and $t = 1, 2 \ldots, T$. $Y_{itr} = 1$ indicates that the i-th individual is deprived in the item r at the time t.

Latent Variable

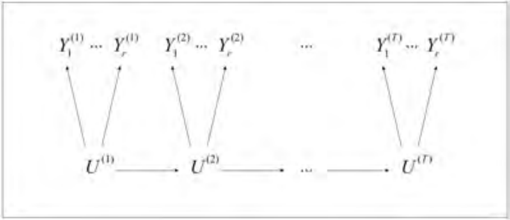
Furthermore, let U_{it} be the latent state of the i-th individual at time t. We assume that $U_{it} = \{1, 2\}$ corresponding to the non deprived/deprived latent status, respectively.

Model's assumptions

Let $Y_{i1}, \ldots, Y_{iR}$ be the vector of the values of the categorical response variables[5] for the i-th individual and U be a latent variable having k support points.

1. *Local independence:* The latent process *fully* explains the observable behavior of a subject

2. *Markovianity:* The latent process follows a first order *inhomogeneous Markov chain*



[5] The R items

The key quantities

Our model belongs to latent Markov models for longitudinal data (Bartolucci et al. (2012))). The quantities involved in likelihood the function (1) are:

1. The *manifest distribution* $\mathbb{P}(Y_{itr} = 1 | U_{it} = j) = p_{jr}$ with $j = 1, 2$
2. The initial distribution $\mathbb{P}(U_{i1} = j) = \pi_j$ with $j = 1, 2$
3. The *inhomogeneous transition probabilities*:
   $\mathbb{P}(U_{it} = j | U_{i,t-1} = h) = \pi_{jth}$ with $t = 2, \ldots, T$.

$$L(\theta) = \prod_{i=1}^{n} \left[\sum_{U_{i1}=1}^{2} \sum_{U_{i2}=1}^{2} \cdots \sum_{U_{iT}=1}^{2} \Pr(U_{i1}) \prod_{t=2}^{T} \Pr(U_{it} | U_{i,t-1}) \times \right. \\ \left. \times \prod_{t=1}^{T} \prod_{r=1}^{R} \Pr(Y_{itr} | U_{it}) \right]^{s_i}, \tag{1}$$

1. We applied the proposed model to the component of EU-SILC released in August 2016.
    - 4 time occasion involved: 2010, 2011, 2012, 2013.
    - 3 different countries involved: Greece, Italy and UK.
2. The 9 deprivation items explained in the introduction have been considered.

Model's output

We focus on the following key quantities (more details in Dotto et al. (2019))

1. Material Deprivation can be evaluated in terms of *Posterior Probability of being deprived* $\tilde{w}(y) = \mathbb{P}[\mathbf{Y_{it}} | U_{it} = 2]$
2. *Sensitivity* ($\hat{p}_{2r} = \mathbb{P}[Y_{ijtr} = 1 | U_t = 2]$) and *Specificity* ($1 - \hat{p}_{1r} = \mathbb{P}[Y_{ijtr} = 0 | U_t = 1]$) of the items.
3. Optimal weights

❶ Deprivation Probability

Deprivation rate according to a continuum of thresholds

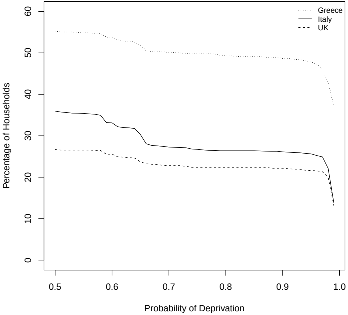


Figure 1: Year 2010

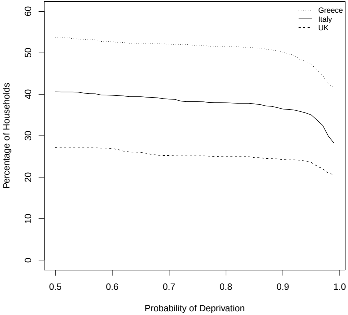


Figure 2: Year 2011

❶ Deprivation Probability

Deprivation rate according to a continuum of thresholds

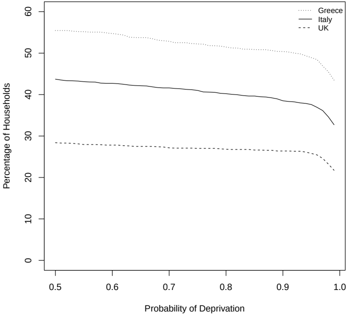


Figure 3: Year 2012

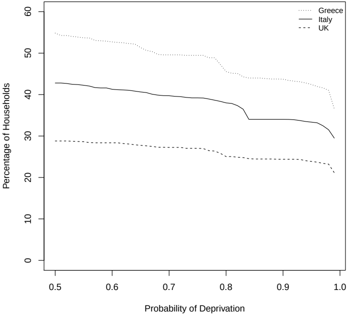


Figure 4: Year 2013

2 Sensitivity and Specificity

Some comments

1. Sensitivity Estimated probability of being deprived ($j = 2$) in a specific item given that the latent variable assumes the status of deprivation
2. Specificity Estimated probability of not lacking item r given that the household is not materially deprived ($j = 1$).

Some more specific comments:

- Generally durable goods (telephone, TV, washing machine) are specific, but not very sensitive, attributes.
- Incapacity of having one week annual holiday away from home and of facing unexpected expenses are sensitive, but not very specific, items.

❷ Specificity and sensitivity

In each country

- $\hat{p}_{2r}$: Sensitivity
- $1 - \hat{p}_{1r}$: Specificity

Table 1: sensitivity for Greece, Italy, and UK separately and for the three countries as a whole, wave 2010–2013.

		Greece		Italy		UK	
Item	description	$\hat{p}_{2r}$	$1-\hat{p}_{1r}$	$\hat{p}_{2r}$	$1-\hat{p}_{1r}$	$\hat{p}_{2r}$	$1-\hat{p}_{1r}$
1	keep the house warm	49.6	92.9	43.4	98.0	21.8	98.1
2	one week holiday	88.9	76.0	92.4	82.4	81.0	95.7
3	afford a meal	31.7	99.0	30.8	98.9	20.9	99.8
4	unexpected expenses	87.3	88.8	83.4	90.3	85.3	91.5
5	telephone	1.2	100.0	0.8	100.0	0.2	100.0
6	color TV	0.1	100.0	0.8	100.0	0.3	100.0
7	washing machine	2.5	99.7	0.9	100.0	1.6	100.0
8	car	15.5	97.6	7.9	99.8	17.9	99.2
9	arrears	58.5	82.9	26.8	98.3	28.7	99.5

❷ Specificity and Sensitivity

In the *Pooled* model

- $\hat{p}_{2r}$: Sensitivity
- $1 - \hat{p}_{1r}$: Specificity

		Pooled	
Item	description	$\hat{p}_{2r}$	$1 - \hat{p}_{1r}$
1	keep the house warm	34.5	98.0
2	one week holiday	87.4	87.5
3	afford a meal	25.8	99.5
4	unexpected expenses	83.5	90.9
5	telephone	0.7	100.0
6	color TV	0.5	100.0
7	washing machine	1.3	100.0
8	car	12.3	99.5
9	arrears	29.8	98.5

❸ Optimal weighting

Why?

Recap:

Each of the 2^9 configurations are mapped in a posterior probability

$$\tilde{w}(y) \ : \ \{0,1\}^R \rightarrow [0,1],$$

BUT

It is impractical to work with 9-dimensional vectors

THUS WE NEED

weights associated to each item $\tau_1, \ldots, \tau_R$ and a one-dimensional score $S(Y) = \sum_{r=1}^{R} \tau_r Y_r$:

❸ Optimal weighting

How?

Let:

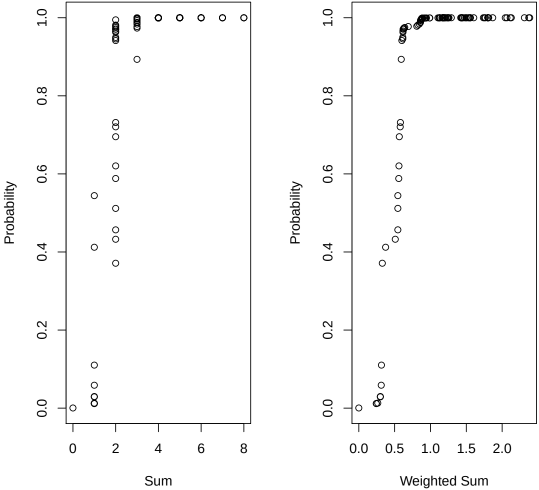
- $\tilde{w}_{(1)}, \tilde{w}_{(k)} \ldots, \tilde{w}_{(2^R)}$ are the (ordered) posterior probabilities of being deprived given the configuration Y
- Let also define as $S_{(k)}(\tau)$ the k-th ordered score given weighting $\tau_1, \ldots, \tau_R$.

We need to minimize:

$$\inf_{\tau} \sum_{k=1}^{2^R} (S_{(k)}(\tau) - \tilde{w}_{(k)})^2. \tag{2}$$

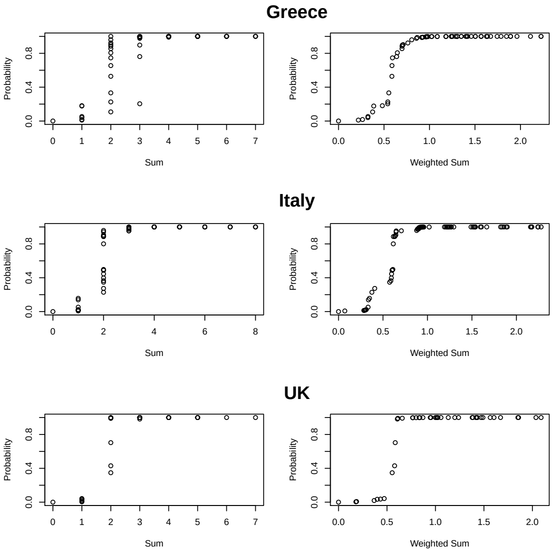
Genetic algorithm (Simon 2013; Scrucca et al. 2013; Scrucca 2017) to solve (2) is needed

❸ Optimal weighting



❸ Optimal weighting

Results in the pooled model



❸ Optimal weighting

Different country...different weights

item	description	Greece	Italy	UK	Pooled
1	keep the house warm	0.134	0.106	0.041	0.074
2	one week holiday	0.180	0.122	0.159	0.123
3	afford a meal	0.192	0.102	0.262	0.086
4	unexpected expenses	0.133	0.116	0.188	0.110
5	telephone	0.143	0.153	0.046	0.132
6	color TV	0.005	0.006	0.042	0.074
7	washing machine	0.061	0.143	0.004	0.172
8	car	0.090	0.112	0.038	0.110
9	arrears	0.061	0.139	0.221	0.120

- The null hypothesis that weights are equal is rejected
- The null hypothesis that weights are equal across countries is rejected too

3 Optimal weighting

Final considerations

- Our score is arguably better at predicting poverty status there are specific combinations of two lacking items that lead to high probabilities to be poor.
- At the same time there are configurations of three lacking items that lead to low proability of being poor
- inverting the distribution of the optimally weighted sums, we can obtain a pooled threshold for deprivation

> With a threshold given by Optimal Weights we can cluster new observations without reestimating the whole model!

Conclusions

Done:

- We treated the status of deprivation as a latent state
- Provided a relative importance score for each item
- Assessed **transitions from and to** material deprivation status

Further direction of research

To do:

- Consider all EU countries
- Insert covariates in the latent distribution

Would it be fair to insert the country of residence as a covariate?
In this case to care about:

- Assessment of Measurement Invariance
  (work in progress with A. Farcomeni, R. Di Mari and A. Punzo)

In other words:

Given an item Y_r, and a covariate X_j, does equation (3) hold?

$$Y_r \perp X_j | U1 \tag{3}$$

References I

Bartolucci, F., A. Farcomeni, and F. Pennoni
2012. *Latent Markov models for longitudinal data*. CRC Press.

Commission, E.
2004. *Joint report on social inclusion*. Office for Official Publications of the European Communities.

Dotto, F., A. Farcomeni, M. G. Pittau, and R. Zelli
2019. A dynamic inhomogeneous latent state model for measuring material deprivation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):495–516.

Scrucca, L.
2017. On some extensions to ga package: Hybrid optimisation, parallelisation and islands evolutionon some extensions to ga package: hybrid optimisation, parallelisation and islands evolution. *The R Journal*, 9(1):187–206.

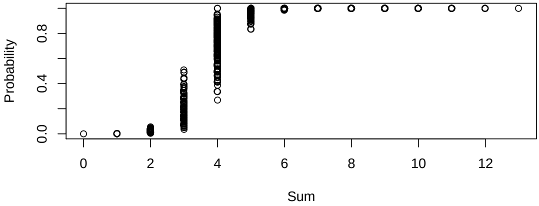
Scrucca, L. et al.
2013. Ga: a package for genetic algorithms in r. *Journal of Statistical Software*, 53(4):1–37.

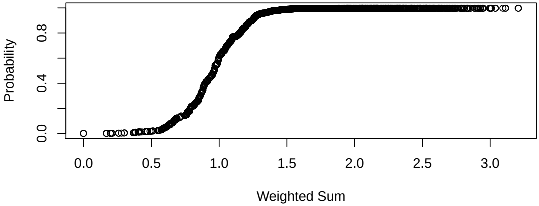
References II

Simon, D.
2013. *Evolutionary optimization algorithms*. John Wiley & Sons.

First Spoiler

Computation of optimal scores on extended deprivation item list





Second spoiler

Maybe a LASSO-type penalty on the likelihood?

$$
\begin{aligned}
l(\boldsymbol{\theta}) = &+ \lambda_1 \sum_{hj} \sqrt{\sum_{tk} \eta_{htkj}^2} + \lambda_2 \sum_{htj} \sqrt{\sum_{k} \sum_{l \geqslant k} (\eta_{htkj} - \eta_{htlj})^2} \\
&+ \lambda_3 \sum_{hkj} \sqrt{\sum_{t} \sum_{s \geqslant t} (\eta_{htkj} - \eta_{hskj})^2}
\end{aligned} \tag{4}
$$

where η_{htkj} denotes the coefficient associated with the j-th dummy variable X_{itj} with respect to item h at time t conditionally on $U_{it} = k$.