What is poverty mapping

Nikos Tzavidis [1]

Methods at Manchester
Manchester, May 20, 2010

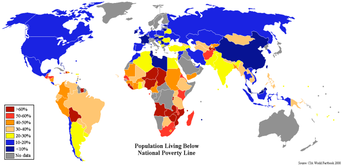
[1]Social Statistics & Centre for Census and Survey Research, University of Manchester, nikos.tzavidis@manchester.ac.uk

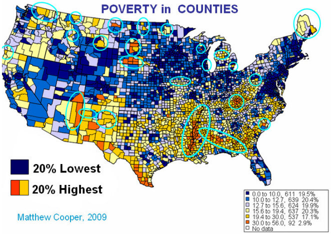
What is Poverty Mapping?

Wikipedia Definition

Methodology for providing a detailed description of the spatial distribution of poverty and inequality within a country. It combines individual and household (micro) survey data and population (macro) census data with the objective of estimating welfare indicators for specific geographic area as small as village or hamlet.

The output


>60%
50-60%
40-50%
30-40%
20-30%
10-20%
<10%
No data
Population Living Below
National Poverty Line
Source: CIA World Factbook 2008


POVERTY in COUNTIES
20% Lowest
20% Highest
Matthew Cooper, 2009
0.0 to 10.0, 611 19.5%
10.0 to 12.7, 639 20.4%
12.7 to 15.6, 624 19.9%
15.6 to 19.4, 637 20.3%
19.4 to 30.0, 537 17.1%
30.0 to 56.0, 92 2.9%
No data

Uses of Poverty Maps

- Guiding intervention mechanisms
  - formulating social and economic policies
  - allocation of government funds
  - regional planning
  - business decision making
- Examples of applications of poverty mapping
  - South Africa: Maps of poverty, safe water and disease mapping
  - Nicaragua: Poverty maps to guide expansion of health services
  - Guatemala: Poverty maps guide need for roads (network data)
  - Cambodia: Poverty maps to guide food aid
  - Poverty estimation in European countries and in the U.S.

Historical Origins

World Bank (2000)

Poverty is the pronounced deprivation in well-being.

- Since the 1880s alternative concepts of well-being
    1. Command over commodities - Income/Consumption based
    2. Include non-monetary measures - Health, Nutrition, Education
    3. Capability to function in a society: Well-being as multidimensional phenomenon
- Poverty mapping is a relatively new concept
    - Recent advances in GIS make poverty mapping possible
    - Powerful statistical software and computing power
    - Development of new statistical methodologies - Small Area Estimation

Measuring Poverty - Targets of Poverty Mapping

- Focus of poverty mapping: Mainly measures under definition 1
- A class of poverty measures are defined by Foster, Greer & Thorbecke (FGT) (1984)
- Denote by t the poverty line

$$F_{\alpha j} = [\frac{t - y_{ij}}{t}]^{\alpha} \mathrm{I}(y_{ij} \leqslant t)$$

  - $a = 0$ - Head Count Ratio / Incidence of Poverty
  - $a = 1$ - Poverty Gap
  - $a = 2$ - Poverty Severity

Poverty Mapping & Small Area Estimation

Target geographies of poverty mapping: Villages, Lower/Middle Super Output Areas, Municipalities

- Surveys are used to provide estimates for large populations
- Estimates for smaller populations are of high importance
- <u>Direct Estimates</u>: Use only the domain-specific sample data
- Direct estimates may suffer from low precision

Small area estimation is concerned with the development of statistical procedures for producing efficient (more precise) estimates for domains <u>planned</u> or <u>unplanned</u> with <u>small</u> or <u>zero</u> sample sizes

Data Requirements & Procedure

- Survey data: Available for y and for x related to y
- Census/Admin Data: Available for x but not for y

Poverty Mapping Implementation - 4 Steps

1. Use sample data to estimate statistical models that link y to x
2. Combine the estimated model parameters with Census x to create a synthetic population of y values
3. Use these predictions to estimate the target parameters and associated precision estimates
4. Map the results

Data Requirements (Cont'd)

- Access to good auxiliary information is crucial
- Ideal Scenario: Access to census/admin microdata

Successful implementation of poverty mapping requires

- A good working model linking y with x
- Detailed auxiliary information for out of sample units
- Appropriate estimation procedures

Poverty Mapping - Step 1: Small Area Models

Models with Random Area Effects - Multilevel Models (Pfeffermann, 2002 ; Rao, 2003)

Include random area effects to account for the between area variation beyond what is explained by covariates

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_j + \epsilon_{ij},\ i = 1, ..., n_j,\ j = 1, ...d$$

M-quantile Small Area Models (Chambers and Tzavidis, 2006)

A less parametric, outlier robust approach to small area estimation

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}(\theta_j) + \epsilon_{ij},\ i = 1, ..., n_j,\ j = 1, ...d$$

Estimation Framework - Steps 2 & 3

The target of inference - Focus on Incidence of Poverty

$$F_j = N_j^{-1}\Big[\sum_{i \in s_j} \mathrm{I}(y_i < z) + \sum_{i \in r_j} \mathrm{I}(y_i < z)\Big]$$

How do we estimate F_j?

Poverty Mapping Methodologies

- The World Bank Method (ELL)
  (Elbers et al. 2003)
- The Empirical Best Predictor (EBP)
  (Molina & Rao 2009)
- The M-quantile approach
  (Chambers and Tzavidis 2006, Tzavidis et al. 2009)

The ELL Approach

Key Features

Assumes a linear model with random area effects

- Fit the model with cluster random effects and transformed y
- Estimate β, σ_u, σ_ϵ
- Generate bootstrap cluster effects u_j^* & random errors ϵ_{ij}^*
- Construct a bootstrap population using

$$y_{ij}^* = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_j^* + \epsilon_{ij}^*$$

- Repeat the process L times each time estimating the target

$$\hat{F}_{0j}^{ELL} = N_j^{-1} \Big[\sum_{i \in U_j} \mathrm{I}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_j^* + \epsilon_{ij}^* < z) \Big]$$

- Average the results over L

The EBP Approach

Target

$$\hat{F}_{0j}^{BP} = N_j^{-1}\Big[\sum_{i\in s_j} F_{0j} + \sum_{i\in r_j} \hat{F}_{0j}^{BP}\Big]$$

$\hat{F}_{0j}^{BP} = E_{y_r}(F_{0j}|y_s)$, $y_r|y_s \sim N(\mu_{r|s}, V_{r|s})$

- Fit the model with area random effects and transformed y
- Generate bootstrap cluster effects u_j^* & random errors ϵ_{ij}^*
- Construct a bootstrap population using

$$y_{ij}^* = \mathbf{x}_{ij}^T\boldsymbol{\beta} + v_j + u_j^* + \epsilon_{ij}^*$$

- Repeat the process L times each time estimating the target

$$\hat{F}_{0j}^{BP} = N_j^{-1}\Big[\sum_{i\in s_j} \mathrm{I}(y_i < z) + \sum_{i\in r_j} \mathrm{I}(\mathbf{x}_{ij}^T\boldsymbol{\beta} + v_j + u_j^* + \epsilon_{ij}^* < z)\Big]$$

The M-quantile Approach

Key Features

$$\hat{F}_{0j}^{MQ} = N_j^{-1}\Big[\sum_{i\in s_j} F_{0j} + \sum_{i\in r_j} \hat{F}_{0j}^{MQ}\Big]$$

- Fit the MQ model with raw y to the sample data and estimate β, q_j
- Generate bootstrap errors e_{ij}^* from the edf of the errors
- Construct a bootstrap population using

$$y_{ij}^* = \mathbf{x}_{ij}^T \boldsymbol{\beta}(\hat{q}_j) + \epsilon_{ij}^*$$

- Repeat the process L times each time estimating the target

$$\hat{F}_{0j}^{MQ} = N_j^{-1}\Big[\sum_{i\in s_j} \mathrm{I}(y_i < z) + \sum_{i\in r_j} \mathrm{I}(\mathbf{x}_{ij}^T \boldsymbol{\beta}(\hat{q}_j) + \epsilon_{ij}^* < z)\Big]$$

Estimating the uncertainty - Mean Squared Error Estimation

Every point estimate must be accompanied by a measure of variability.

For all three approaches precision measures are computed by bootstrap

- ELL and EBP - Parametric bootstrap
- MQ - Non-parametric bootstrap

A Case Study: Poverty Mapping in Italy

- Use EU-SILC data from Italy
- Census micro-data:
  Available from ISTAT for Tuscany, Lombardia and Campania
- Estimation targets & geography:
  - The incidence of poverty
  - The income distribution
  - Geography: Provinces (NUTS 3) and municipalities (NUTS 4)

A closer look at the data

Survey Data:

- Household level data for Lombardia, Tuscany, & Campania
- Household level variables:
  - Equivilsed income
  - Household size
  - Facilities, number of rooms, size of house
- Head of household variables:
  - Age, sex, marital status, ethnicity
  - Education, employment status

Census Data

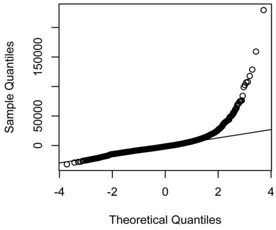
- Availability of Census microdata for most key variables
- Only the facilities variables are available at area level

Preliminary Modelling of the Italian EU-SILC

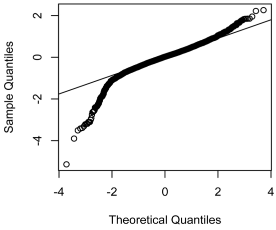
- Working model: 2-level random effects (Province)
- Outcome variable: Equivalised income
- Covariates: Household and head of household variables
- Percentage of variability explained - 18%
- Intracluster correlation coefficients - 4.5%
- Normal QQ plots show departures from normality

Model Diagnostics

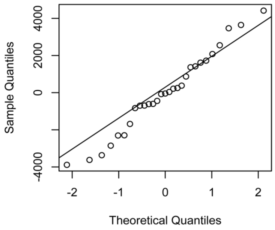
First level residuals, raw scale



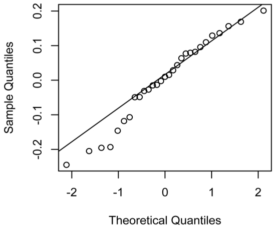
Second level residuals, log scale



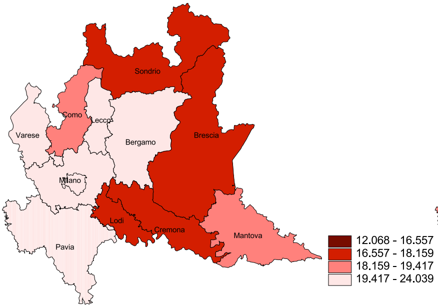
First level residuals, raw scale

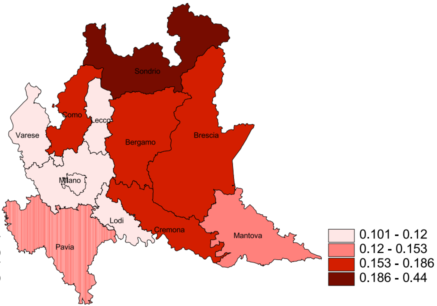


Second level residuals, log scale

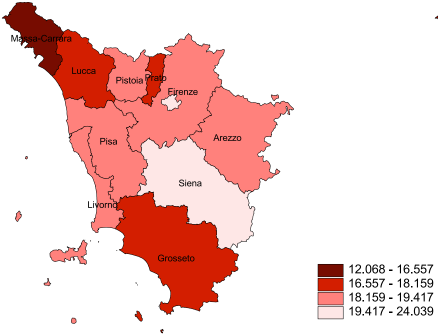


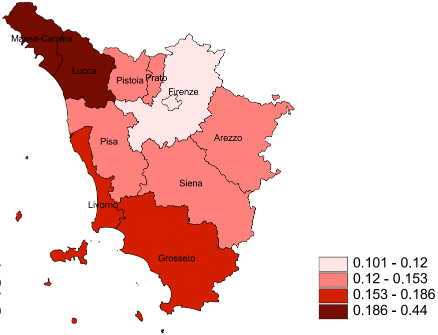
A picture of the welfare in Lombardia



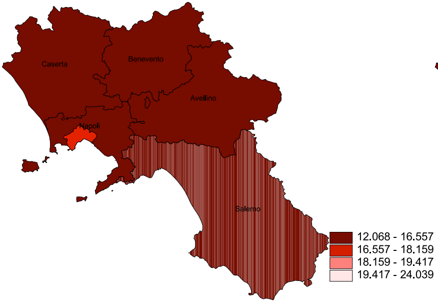


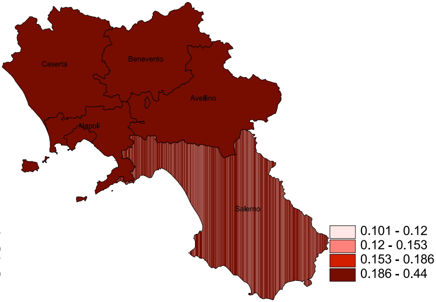
A picture of the welfare in Tuscany





A picture of the welfare in Calabria





Current developments

- Two funded projects - EU 7th Framework
  - SAMPLE - http://www.sample-project.eu/en/the-project/
  - AMELI - http://www.uni-trier.de/index.php?id=24474
- Aim 1: Develop inequality and poverty indicators
- Aim 2: Develop models for estimating these indicators and corresponding accuracy measures at small area level (NUTS 3 and NUTS 4 level)

SAMPLE Project: Structure

- WP1 New indicators and models for inequality and poverty with attention to social exclusion, vulnerability and deprivation (CRIDIRE / WSE / GUS / PP / UNIPI-DSMAE / SR)
- WP2 Small area estimation of poverty and inequality indicators (UNIPI-DSMAE / CCSR / UC3M / UMH)
- WP3 Integration of EU-SILC data with administrative data (PP /SR / UNIPI-DSMAE)
- WP4 Standardization and application development - Software for living conditions estimates (SR)
- WP5 Management (UNIPI-DSMAE / ALL)
- WP6 Information and dissemination of results (SR / ALL)

SAMPLE Project: Work Package 2 in Detail

- Task 2.1 Estimate the cumulative distribution function of income at small area level - (UNIPI-DSMAE and CCSR)
- Task 2.2 Small area estimates of poverty with spatial models - (UC3M, UNIPI-DSMAE, CCSR)
- Task 2.3 Small area estimates of poverty with temporal models - (UMH)
- Task 2.4 Small area estimates of poverty with spatial-temporal models - (UMH, UC3M, UNIPI-DSMAE)

Current Research Directions

- Robust methods in poverty estimation
- Spatial methods in poverty estimation
- Models that account for geographical positioning (longitude, latitude, altitude)
- Software development - R functions (open source)
- Poverty mapping with fuzzy set measures
- Multidimensional measures of poverty - Multivariate models

References

- Breckling, J. and Chambers, R. (1988). M -quantiles. *Biometrika*, **75**, 761–771.
- Chambers, R. and Dunstan, R. (1986). Estimating distribution function from survey data, *Biometrika*. **73**, 597-604.
- Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93**, 255-68.
- Chambers, R., Chandra, H. and Tzavidis, N. (2008). On robust mean squared error estimation for linear predictors for domains. [*Paper submitted for publication. A copy is available upon request*].
- Elbers, C., Lanjouw J. and Lanjouw P., (2003). Micro-level estimation of poverty and inequality, *Econometrica*, **71**, 355-64.
- Molina, I. and Rao, J.N.K. (2009). Small area estimation of poverty indicators. Under review by *The Canadian Journal of Statistics*.
- Pfeffermann, D. (2002). Small Area Estimation- New Developments and Directions. *International Statistical Review*, **70**, 125-43.
- Prasad, N. G. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators,*Journal of the American Statistical Association*, **85**, 163-71
- Rao, J.N.K. (2003). Small area estimation. Wiley
- Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R. (2008). Small area estimation via M-quantile geographically weighted regression. [*Paper submitted for publication. A copy is available upon request]*.
- Sinha, S. and Rao, J.N.K.(2008). Robust small area estimation. [*Paper submitted for publication.]*.
- Tzavidis, N., Marchetti, S. and Chambers, R. (2009). Robust prediction of small area means and quantiles. To appear in *The Australian and New Zealand Journal of Statistics*.
- Tzavidis, N., Salvati, N., Pratesi, M. and Chambers, R. (2008). M-quantile models for poverty mapping. *Stat.Meth.and Applications*, **17**, 393-411.