PENNSTATE

1855

Center for Statistical Ecology and Environmental Statistics

Hotspot Detection and Mapping of Poverty

By Gianni Betti, Francesca Ballini, and Laura Neri

University of Siena, Italy

This material is based upon work partially supported by the National Science Foundation under Grant No. 0307010. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the agencies.

[Presented at the 7th Annual International Conference on Digital Government Research]

Technical Report Number 2006-0533
TECHNICAL REPORTS AND REPRINTS SERIES
May 2006

Department of Statistics
The Pennsylvania State University
University Park, PA 16802

G. P. Patil
Distinguished Professor and Director
Tel: (814)865-9442 Fax: (814)865-1278
Email: gpp@stat.psu.edu
http: //www.stat.psu.edu/~gpp
http://www.stat.psu.edu/hotspots
DGOnline News

Hotspot Detection and Mapping of Poverty

Gianni Betti
University of Siena
P.za S. Francesco, 7
53100 Siena, Italy
tel. 00390577235084
betti2@unisi.it

Francesca Ballini
University of Siena
P.za S. Francesco, 7
53100 Siena, Italy
tel. 00390577235084
ballini@unisi.it

Laura Neri
University of Siena
P.za S. Francesco, 7
53100 Siena, Italy
tel. 00390577235084
neri@unisi.it

ABSTRACT
In this paper we aim at detecting poverty hotspots and preparing the corresponding maps. Poverty and inequality maps — spatial descriptions of the distribution of poverty — are most useful to policymakers and researchers when they are finely disaggregated, i.e. when they represent small geographic units, such as cities, municipalities, regions or other administrative partitions of a country. Moreover, when poverty hotspots are detected, policymakers can use them to propose appropriate programmes and anti-poverty policies. We demonstrate the construction of detailed poverty maps, primarily using data from a Population Census in conjunction with an intensive small-scale national sample survey. The methodology adopted combines census and survey information to produce finely disaggregated maps. The basic idea is to estimate a linear regression model with local variance components using information from the smaller and richer sample data - in the case of Albania the Living Standard Measurement Study (LSMS) conducted in 2002 — in conjunction with the large-scale but limited information from the 2001 Population and Housing Census.

Categories and Subject Descriptors
H.2.33 [**Languages**]: *Database (persistent) programming languages*.

General Terms
Measurement, Design, Languages.

Keywords
Poverty and inequality mapping, hotspot detection.

1. INTRODUCTION
In this paper we aim at detecting poverty hotspots and performing poverty mapping primarily using data from a Population Census, in conjunction with an intensive small-scale national sample survey. The methodology adopted, described in [1], combines census and survey information to produce finely disaggregated maps which describe the spatial distribution of poverty in the country under investigation. We intend to adopt this methodology to the case of Albania. The basic idea is to estimate a linear regression model with local variance components using information from the smaller and richer sample data — in the case of Albania the Living Standard Measurement Study (LSMS) conducted in 2002 — in conjunction with aggregate information from the 2001 Population and Housing Census of Albania, supplemented by some other sources (e.g. the General Census of Agricultural Holdings).

The estimated distribution of the dependent variable in the regression model (monetary variable) can therefore be used to generate the distribution for any sub-population in the census, conditional to the sub-population's observed characteristics. From the estimated distribution of the monetary variable in the census data set or in any of its sub-populations, estimates can be made of various poverty measures, such as the Sen and the Foster-Green-Thorbecke indices. To assess the precision of the estimates, standard errors of the poverty measures need to be computed using an appropriate procedure such as bootstrapping. Four important aspects of this methodology should be noted at the outset. Firstly, information from the Census is required at micro (household and individual) level; however micro-level linkage between Census and survey data is not required. Secondly, the vector of covariates utilised in the regression model implies that those variables have to be present in both sources. Thirdly and most importantly, the common variables in the sources must be sufficiently comparable; comparability requires the use of common concepts, definitions and measurement procedures. Moreover, especially in Transition Countries such as Albania, with rapid changes in living conditions, it is important that reference periods for the data sets are as close as possible to each other. Poverty indices are calculated for the whole of Albania and disaggregated at three levels: the 12 Prefectures; the 36 Districts; and the 374 Communes/Municipalities. For a complete discussion of the poverty mapping results and policy implications see [2].

2. HOTSPOT DETECTION
In [3] Hotspot is defined as "…something unusual, anomaly, aberration, outbreak, elevated cluster, critical resource area, etc…"; in our context a Poverty Hotspot represents an area characterized by certain local characteristics which could expand and affect also other neighbouring areas.

We have adopted the "upper level sets (ULS) scan statistic" in order to identify poverty hotspots at three levels in Albania: Prefectures, Districts and Communes. The most interesting results are those obtained for Districts and Communes, and are reported respectively in Tables 1 and 2. From [2] it is known that in Albania there is wide heterogeneity of poor people among Prefectures, but Districts within the same Prefecture show similar level of diffusion (in terms of the so-called Head Count Ratio). In fact, Table 1 shows how at any stopping criteria (20% to 50%) the ULS scan statistic detects one large poverty hotspot comprising Districts concentrated in two or three Prefectures.

Table 1. Hotspot detection for Districts

Stopping criteria	# units	p-value	MLE	Risk ratio
20%				
Hotspot 1	11	0.001	41443.636	1.529
Hotspot 2	2	0.001	2377.16	1.193
Hotspot 3	1	0.001	418.41	1.166
30%				
Hotspot 1	13	0.001	42499.799	1.511
Hotspot 2	2	0.001	2377.15	1.193
Hotspot 3	1	0.001	418.41	1.166
40%				
Hotspot 1	13	0.001	42499.799	1.511
Hotspot 2	2	0.001	2377.15	1.193

When we disaggregate the analysis at Commune level (see Figure 1), we note a much higher heterogeneity among Communes in the same District, and observe three large poverty hotspots in the northern and central parts of Albania (at 10% stopping criteria). Table 2 shows that the poorest 10% of all poor people in Albania are concentrated in around 30% of Communes (140 out 374); in fact poverty is concentrated in smaller Communes and in isolated areas like the northern mountain region.

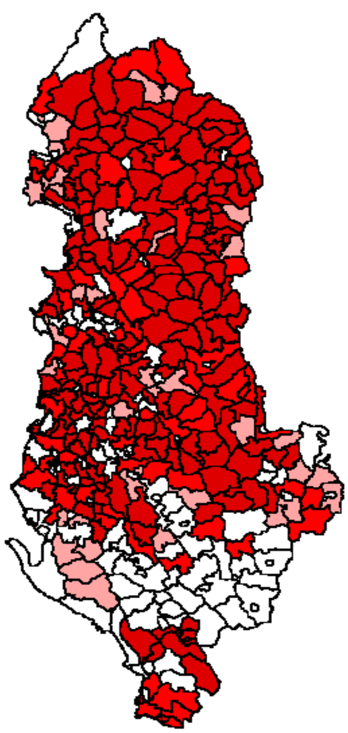

Figure 1. Poverty Hotspots at 20%, 30%, 40% and 50% stopping criteria for Communes.

Table 2. Hotspot detection for Communes

Stopping criteria	# units	p-value	MLE	Risk ratio
10%				
Hotspot 1	69	0.001	77646.343	1.963
Hotspot 2	44	0.001	18058.545	1.482
Hotspot 3	27	0.001	10555.89	1.525
20%				
Hotspot 1	88	0.001	81689.615	1.857
Hotspot 2	75	0.001	31111.446	1.530
Hotspot 3	3	0.001	2346.215	1.556
Hotspot 4	7	0.001	471.795	0.810
30%				
Hotspot 1	176	0.001	140721.17	1.974
Hotspot 2	7	0.001	471.795	0.810
Hotspot 3	6	0.001	339.848	0.768
Hotspot 4	1	0.001	266.794	1.990

3. CONCLUDING REMARKS

The findings of research like the present one are potentially very useful for policymakers. We find that in Albania there is considerable heterogeneity of poverty rates across administrative units. The particular spatial pattern of this heterogeneity and the characteristics underlying the hotspots have important policy implications for poverty alleviation programmes: at the highest level we observe a large spatial heterogeneity among Prefectures; this spatial heterogeneity is much less pronounced among Districts within the same Prefecture; however, it is pronounced again at the lowest level among Communes within the same District. What this means for the practitioner and the policymaker is that it is important to disaggregate down to the Commune level when analysing issues and planning interventions, as this will add substantially in terms of precision of the targeting of resources when compared to stopping the analysis at the District level.

4. ACKNOWLEDGMENTS

We would like to thank Prof. Ganapati Patil and Dr. Pushkar Patankar for having provided the MLE estimates for hotspot detection and Prof. Vijay Verma for helpful comments.

5. REFERENCES

[1] Elbers, C., Lanjouw, J. O., and Lanjouw, P. Micro-level Estimation of Poverty and Inequality. *Econometrica, 71* (2003), 355-364.

[2] Neri, L., Ballini, F., and Betti, G. Poverty and Inequality mapping in Transition Countries. *Statistics in Transition, 7* (2005), 135-157.

[3] Patil, G. P., and Taillie, C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics, 11* (2004), 183-197.