Identifying Poverty Groups in Nairobi's Slum Settlements: A Latent Class Analysis Approach

Leontine Alkema, Ousmane Faye, Michael Mutua, Eliya Zulu*

Abstract

This paper aims to contribute to knowledge on conceptualizing and measuring urban poverty by categorizing households according to their socio-economic status.
We identify groups with similar profiles of socio-economic status using Latent Class Analysis (LCA). In LCA an unobserved, latent variable (poverty) explains the association between observed variables (indicators of socio-economic status). Compared to other methods for measuring poverty (such as Principal Component Analysis), in LCA the number and size of the poverty groups is determined by the data.
This study uses data from the longitudinal Nairobi Urban Health Demographic Surveillance System to identify poverty groups in two slums in Nairobi; Korogocho and Viwandani. In Korogocho we identify three groups, the poorest group accounting for 19% of all households. In Viwandani we identify four groups, with 27% of the households in the poorest group.

Leontine Alkema is Graduate Student, Department of Statistics, and Shanahan fellow at the Center for Studies in Demography and Ecology, University of Washington, Seattle, US (Email: alkema@u.washington.edu), Ousmane Faye is Postdoctoral Fellow, African Population and Health Research Center, Nairobi, Kenya (Email: ofaye@aphrc.org), Michael Mutua is Data Analyst, African Population and Health Research Center, Nairobi, Kenya (Email: mmutua@aphrc.org), Eliya Zulu is Deputy Director, African Population and Health Research Center, Nairobi, Kenya (Email: ezulu@aphrc.org).
This research is partly funded by the Wellcome Trust Foundation 078530/Z/05/Z, Hewlett Packard Foundation 2006-8376 and Rockefeller Foundation AR037.
The authors are grateful to the study communities in Korogocho and Viwandani and the participants of the workshop on poverty measurement held in Nairobi, February 2007 for sharing experiences and insightful discussions.

1. INTRODUCTION

The first Millennium Development goal is to eradicate extreme poverty and hunger (http://www.un.org/millenniumgoals). To work towards achieving the first MDG, those who suffer from extreme poverty and hunger need to be identified. The profiles of the poor need to be studied to get more insight into the factors that drive poverty outcomes, in order to design and carry out intervention programs. Identifying the poor, studying their profiles and monitoring progress with respect to poverty eradication require data on socio-economic status and appropriate methodologies for poverty measurement. Commonly collected indicators of socio-economic status (SES) are derived from data on asset ownership, amenities, income and expenditure, and household and individual food security. The major challenge in poverty analysis relates to how to derive poverty indicators that not only group individuals or households that are poor and not poor at any given point in time, but also how they move in and out of poverty over time.

Principal component analysis (PCA, Manly 1994) is a commonly used approach in poverty measurement in developing countries. It was first proposed by Filmer and Pritchett (1998) and commonly used by the World Bank and the Demographic Health Survey program (Rutstein and Johnson, 2004). In the PCA approach the dimension of an initial set of correlated variables (the SES indicators) is reduced by creating uncorrelated (perpendicular) components; each component is a weighted combination of the initial variables. The first component explains the highest proportion of the total variance for all indicators combined and is often referred to as the wealth index.
In the PCA approach households or individuals are grouped based on the ranking of the wealth indices, e.g. by using the quintiles of the wealth index set to define the 20% poorest.

The drawback of PCA is that information is lost when summarizing the indicators into one number and the arbitrary cut-off between the poorest and the rest. The cut-off between the poorest and the rest is based on the relative ranking and involves the decision which percentage of the poorest households to examine. This makes the cut-off arbitrary, are the groups really different? Households with similar characteristics might be forced into two different groups. Specifically, when examining poverty dynamics, changes in SES of households over time is of interest to get a better insight into how people fall into or move out of poverty. What is of interest is to identify households for which the socio-economic profile has changed significantly from one time point to the other, compared to the overall changes within the population. If the poorest group contains 30% of all households, moving from the lowest to the second lowest quintile is not a significant improvement. Whether or not households move out of the poorest group is of interest.

In this paper we use latent class analysis (LCA) for identifying groups of households with similar socio-economic profiles within the multidimensional poverty space. This approach overcomes the drawbacks of PCA of only taking into account one component (mapping the multidimensional space onto one line) and having to decide on the number of groups to be considered. In the LCA approach the grouping of households is datadriven: The number of classes and the size of each class is determined by the outcomes of the SES indicators.

The approach will be illustrated using data from the longitudinal Nairobi Urban Health Demographic Surveillance System (NUHDSS), set up in two slum settlements in Nairobi City. The DSS is set up by the African Population and Health Research Center (APHRC) to serve as the primary research tool for monitoring and evaluating health and poverty alleviation programs. Data on about 60,000 residents are updated every four months on a range of issues including: fertility, mortality, cause of death, vaccination for children, migration, marriage, schooling, housing conditions, household possessions and amenities (once every year).

In Section 2 we explain the latent class model that is used to identify poverty groups. In Section 3 we describe the data from the two study communities in Nairobi. In Section 4 we present the results of the LCA approach for identifying poverty groups in both slums. We end with a summary and discussion of the results, and ideas for future research in Section 5.

2. LATENT CLASS ANALYSIS FOR POVERTY MEASUREMENT

We use Latent Class Analysis (McCutcheon 1987, Goodman 2002) to identify groups of households with similar profiles of socio-economic indicators. The main assumption in latent class analysis (LCA) is that there is some unobserved variable / phenomenon (which is called the latent variable) which explains associations between a set of observed variables (also referred to as the manifest variables). The relationship between the latent variable and its indicators is probabilistic to allow for measurement error. The goal in LCA is to identify homogeneous classes (groups); the classes represent the different outcome categories of the latent variable, in each class the association between the indicators disappears. This is called local independence. The association between the observed variables is explained by the classes of the latent variable (McCutcheon 1987, Hagenaars 1990, 1993). In the latent class model (LC-model) each household gets assigned a probability that it belongs to a certain class. Based on the highest probability of class membership, the household is assigned to a certain class.

In poverty measurement the goal is to determine the “poverty status” of households (or individuals) based on differences in SES indicators. Poverty itself can be considered to be a latent variable, its manifest variables are the SES indicators. The LCA approach can be applied to poverty measurement in order to examine the different categories of poverty based on the SES indicators. Poverty groups are defined by pulling together combinations of indicators that are similar, e.g. a group with low asset ownership and low food security.

LCA has been used for poverty measurement in a number of European countries. Moisio (2004) used LCA to identify the poor in Finland, the Netherlands and the UK based on data on housing quality, income and a subjective measure of poverty. DeWilde (2004) examined the percentage of poor in Belgium and the UK over time using longitudinal panel data on housing quality, financial stress, and limited financial means.

In order to explain the grouping of households in the LC-model, define:

- M = Number of manifest variables (SES indicators)
- V_m = Manifest variable (indicator) m for $m = 1, \ldots, M$.
- r_s = Response pattern s (defined for $s = 1, \ldots, S = 2^M$) given by a combination of manifest variables: $r_s = \{V_1 = v_{1s}, \ldots, V_M = v_{Ms}\}$
- C = Number of latent classes
- π_c = Size of class c (proportion of households in class c)
- $P_c(V_m = v_m)$ = Probability of observing outcome v_m for manifest variable V_m in class c

Based on the assumption of local independence, the probability of observing response pattern r_s in class c is given by:

$$P_c(R = r_s) = \Pi_m P_c(V_m = v_m).$$

$P_c(R = r_s)$ is called the recruitment probability of class c for response pattern r_s. It follows that for a given number of classes C, class proportions p_c and conditional indicator probabilities $P_c(V_m = v_m)$ we can calculate the probability that household h belongs to class c:

$$P(\text{Household h Є Class c}) = p_c * P_c(R = r^{(h)}) \;/\; P(r^{(h)}),$$

with $r^{(h)}$ the response pattern for household h and $P(r^{(h)}) = \sum_c p_c * P_c(R = r^{(h)})$.
For a fixed number of classes the class proportions and conditional indicator probabilities determine the grouping of the households, therefore the model parameters of the LC-model are given by the class proportions and the conditional indicator probabilities.

The model parameters are estimated such that the model best fits the data, the outcomes of the SES indicators. These outcomes can be summarized with a frequency table in which each entry is given by $F_H(r_s)$: the observed frequency of the response pattern r_s for a sample of H households. Denote $E_H(r_s)$ as the expected frequency of response pattern r_s under the model, given by:

$$E_H(r_s) = H * \sum_c p_c * P_c(R = r_s).$$

In order to find the best model fit, the difference between the observed and expected frequencies is minimized using an iterative expectation-maximization (EM) algorithm (Dempster et al., 1977). The number of classes C is based on the Bayesian information criterion (Schwartz, 1978) which compares model fit while taking into account the number of parameters of each model.

The results of the model are the groups and for each household the probability that it belongs to a certain group. Model fit is assessed using the likelihood ratio Chi-square test statistic (Goodman 1970) and testing for local independence within the groups. The LC-model is fit using the R package poLCA (Linzer and Lewis, 2007).

3. DATA

This study uses data on household amenities and possessions, type of tenure, expenditure and food consumption from the NUHDSS to identify poverty groups in two slum settlements in Nairobi, Korogocho and Viwandani. The study uses data on household possessions, amenities, food security, and broad categories of expenditure collected from the NUHDSS study areas between September and December 2006. For household assets, the study asked respondents whether households had bought the assets in the previous year, whether they had disposed of assets that they had during the year, and whether they owned the assets in any other household apart from their household in the slum location. This distinction was necessarily to determine the extent to which slum residents invest in their rural and other homes outside the slums.

Table 1 shows descriptive statistics for both slums. The number of households is 1077 for Korogocho and 2971 for Viwandani. Viwandani is a slum settlement located in an industrial area, a large proportion of the population consists of migrant workers who come to Nairobi to work and do not settle in the slum for long The mean duration of stay is 5.4 years, while it is much higher in Korogocho at 12.5 years. The difference between the two slums also becomes clear when examining the proportion of households that own their dwelling (19% in Korogocho compared to only 7% in Viwandani) and whether they own assets in another place ("ownership at another place"). This indicator refers to the proportion of households that own a table, bed or mattress in another place outside of their slum dwelling. This indicator reflects linkages of the slum household with other households, e.g. part of their family left in their rural home. The proportion of households with ownership in another place is 24% in Korogocho compared to 75% in Viwandani.

For examining socio-economic status, the indicators ownership of dwelling and ownership at another place are taken into account. Additional indicators on asset ownership are ownership of radio, DVD-player and phone. Radio is a relatively common asset in the slums, around 80% of the households owns a radio. A DVD-player is a luxury good; only 9% of the households owns it. Ownership of a phone is 36% in Korogocho and 51% in Viwandani. Including ownership of phone into the model is of interest as phone is an asset and refers to accumulated wealth. Indeed, a phone is a useful tool to keep in touch with relatives outside Nairobi and find work (as many slum dwellers depend on casual labor). The outcome of the source of lighting (electricity or by another source) is also included into the model; 39% of the households in Korogocho have electricity, compared to 25% in Viwandani.

We include two indicators of food consumption/security. The first one is the outcome on the question: "Which of these statements best describes the food eaten by your household during the last 30 days?" The answers were 1. (Often/sometimes) your household did not have enough food to eat, 2. Your household had enough food, but not always the kinds of food it wanted, 3. Your household had enough of the kinds of food it wanted to eat.

In Korogocho the majority of outcomes are in the second category: 74% of the households responded that they had enough food to eat, but not always the kinds they wanted. 16% of the households did not have enough food to eat. For Viwandani the outcomes are more spread over the 3 outcomes: 41% of the households had enough food to eat and the kinds it wanted. However, of the remaining 59%, 26% did not have enough food to eat.

The second indicator on food consumption is expenditure on food. Households were asked to give the total expenditure on food in their household in the last week. Total expenditure was adjusted for household composition based on the OECD equivalence scale to get adult equivalent per capita household expenditure (Barrientos, 2003). Mean daily expenditure of an individual in Korogocho is just above 1 US$ a day, compared to around 1 US$ a day in Viwandani.

A subjective indicator of SES is given by self-rated wellbeing. Households were asked to rank themselves compared to other households in their community in terms of general wellbeing on a scale from one to ten, ten meaning richest. Mean ranking is around 4 in both slums; on average household rank themselves just below midpoint 5.

The indicators as given in Table 1 will be included in the LCA to derive homogeneous poverty groups. Because the two slums are different in terms of population and their socio-economic status, we will start by examining each slum separately.

Table 1: Descriptive statistics for Korogocho and Viwandani

	Korogocho	**Viwandani**
Sample size (No of households)	1077	2971
Mean duration of stay (years)	12.5	5.4
SES indicators:		
Ownership of dwelling	0.19	0.07
Ownership in other place	0.24	0.75
Ownership of radio	0.73	0.8
Ownership of DVD-player	0.09	0.09
Ownership of phone	0.36	0.51
Electricity	0.39	0.25
Food 1: Not enough food	0.17	0.26
Food 2: Enough, but not as wanted	0.74	0.32
Food 3: Enough, and as wanted	0.09	0.41
Expenditure on food (US$/week/adult)	8 US$	7 US$
Self-rated wellbeing (Scale 1-10)	3.9	4.1

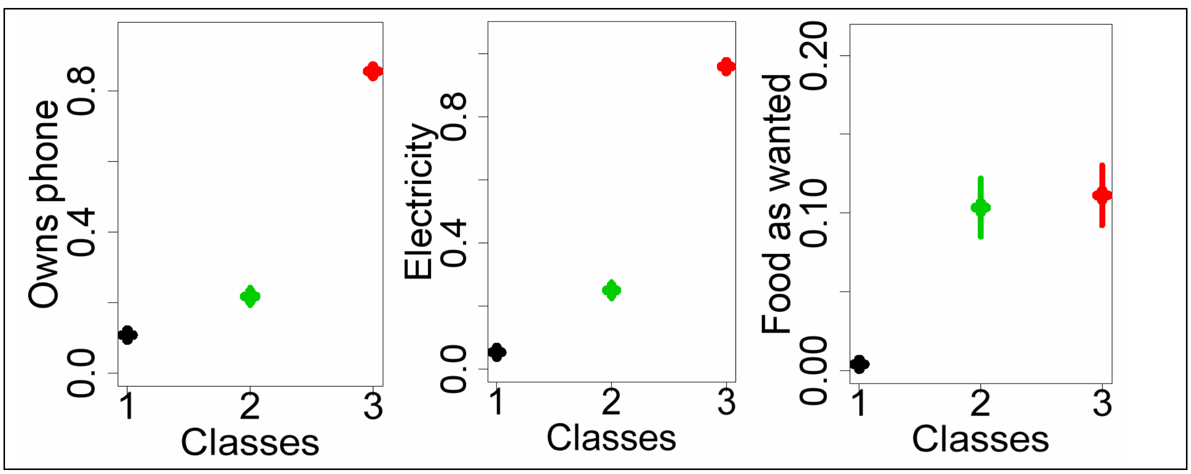
4. RESULTS

4.1 Poverty groups in Korogocho

Latent class analysis for Korogocho resulted in the selection of three groups of different size. The first group contains 19% of the households and can be identified as the poorest group. The second group is the largest group containing 56% of the households and has medium outcomes on the socio-economic indicators compared to the other two groups. The last group accounts for 25% of the population and is the richest group.

Outcomes for ownership of phone, electricity for lighting and having enough food and as wanted are shown in Figure 1 for each of the three groups, with 95% confidence intervals for the outcome in the total population of households in Korogocho. The plots show that ownership of a phone increase from 11% to 20% to 89% from the poorest to the richest group. The difference in electricity between the poorest and the richest group is 93%. The proportion of households in Korogocho that had enough food to eat and the kinds it wanted is 1% in the poorest group, compared to around 10% in the medium and richest group.

Figure 1: Indicator proportions with 95% confidence interval within each group. The groups are given by: 1. Poor (19%), 2. Median (56%), 3. Rich (25%)



The indicator proportions in each group are given in Table 2. As for ownership of phone and electricity, ownership of radio and DVD-player increase going from the poorest to the richest group. In the poorest group, 75% of the households report to not have enough food. Around 85% of the households in the medium and richest group have enough food to eat, and respectively 4 and 6% of the households do not have enough food. Self-rated wellbeing increases going from the poorest to the richest, it is 2.2 for the poorest group and 5.1 for the richest group.

Two variables that do not show an increase or decrease going from poor to rich are ownership of dwelling and ownership in another place. Ownership of dwelling is lowest in the medium group, 16% compared to 21% in the poorest and 27% in the richest group. Ownership in another place is highest in the medium group, 36% compared to 10% in the richest and 4% in the poorest group. A possible explanation is that the poorest as well as the richest group are the households that are more settled in the slum and have fewer ties with another place. This is supported by the mean duration of stay in each of the groups. The mean duration of stay is shortest in the medium group, 11.7 years, compared to 13.3 years in the richest and 14.7 years in the poorest group.

Table 2: Indicator proportions by group for the 3-class model for Korogocho

	Poor	Medium	Rich
Group proportion	0.19	0.56	0.25
Ownership of radio	0.18	0.79	0.98
Ownership of DVD-player	0	0	0.35
Ownership of phone	0.11	0.2	0.89
Electricity	0.06	0.23	0.98
Food 1: Not enough food	0.75	0.06	0.04
Food 2: Enough, but not as wanted	0.25	0.84	0.85
Food 3: Enough, and as wanted	0.01	0.10	0.11
Expenditure on food (US$/week/adult)	7.1	7.9	8.6
Self-rated wellbeing	2.2	3.9	5.1
Ownership of dwelling	0.21	0.16	0.27
Ownership in other place	0.04	0.36	0.1
Duration of stay (in years, not included in model)	14.7	11.7	13.3

4.3 Poverty groups in Viwandani

When fitting a LC-model with three classes to the data for Viwandani, we find a poor, medium and rich group, which contain respectively 28%, 50% and 22% of the households. The advantage of the LCA approach is that the number of classes is determined by the data. The model fits for models with varying number of classes can be compared using the Bayesian information criterion. This criterion combines the goodness of fit with a penalty term for the number of parameters in the model. E.g., a model with a larger number of classes is more likely to fit the data better than a model with a smaller number of classes, but the improvements in model fit might be very small compared to the extra number of parameters that are needed in the model.
For Korogocho a model with 3 classes fit the data best. For Viwandani the model with 4 classes fits the data slightly better than the 3-class model (or a model with a different number of classes). The 4-class model is very similar to the 3-class model with respect to the poorest and medium group: In the 4-class model we identify a poor group with 27% of all households and a medium group with 52% of the households, both groups are very similar in size and characteristics of the groups as determined by the 3-class model.

The richest group in the 3-class model is separated into 2 groups in the 4-class model, identifying the richest group with 13% of the households and an "extra" class with 9% of the households. The extra class is very similar to the richest group with respect to accumulated wealth. However, this class has lower self-rated wellbeing, lower outcomes on food security but higher expenditure on food than the richest group. In comparison with the other groups, the extra class has the lowest ownership in other places and the longest duration of stay in the slum. The indicator proportions for each class in the 4-class model are given in Table 3. This result illustrates the possibility of identifying different types of households with the LCA approach. The results are preliminary; the characteristics of the extra class will be examined further.

Table 3: Indicator proportions for the 4-class model for Viwandani

	Poor	Medium	Rich	Extra class
Group proportion	0.27	0.52	0.13	0.09
Ownership of radio	0.09	0.01	0.41	0.41
Ownership of DVD-player	0.53	0.87	1	0.96
Ownership of phone	0.05	0.55	0.98	0.92
Electricity	0.03	0.08	0.97	0.93
Ownership of dwelling	0.06	0.05	0.13	0.17
Ownership in other place	0.58	0.83	0.96	0.43
Self-rated wellbeing	2.8	4.5	5.3	3.9
Food 1: Not enough food	0.68	0.11	0	0.35
Food 2: Enough, but not as wanted	0.25	0.42	0.07	0.33
Food 3: Enough, and as wanted	0.07	0.47	0.93	0.32
Expenditure on food (US$/week/adult)	6.7	7.1	6.6	7.4
Duration of stay (in years, not included in model)	5.4	5.3	5.1	6.3

4. DISCUSSION AND FUTURE WORK

In this paper we use latent class analysis (LCA) for identifying poverty groups in two slums in Nairobi based on various indicators of socio-economic status. The advantage of the LCA approach is that the number of classes and the size of each class are determined by the data. In Korogocho we identified the poorest group with 19% of the households, the medium with 56% of households and the richest group with 25% of households. For Viwandani we identified four groups: the poorest households (27% of all households), the medium group (52%), the rich group (13%) and an extra group (9%) of households for which the outcomes of the SES indicators are similar to the richest group with respect to accumulated wealth and facilities, but different for the remaining SES indicators.

The results as presented in our paper are preliminary and based on a subset of the population in Korogocho and Viwandani. The final analysis will be based on a larger sample of households. We will examine the robustness/sensitivity of the results with respect to the choice of the indicators and compare with the results based on the PCA approach. We will also do the grouping based on the data of the two slums combined and interpret the results.
The determinants of poverty will be examined (e.g. education level of household members, gender of head of household) using latent class regression (Dayton and Macready, 1988; Hagenaars and McCutcheon, 2002). Using this analysis, we will be able to determine what explains the difference between two groups if, for example, the poorest as well as the richest group are both the long-term dwellers.

The data as presented in this paper were collected in 2006. In future work we will carry out a latent class analysis based on data from 2003 to get the grouping of households that were present at that time and determine how these groups changed by 2006. At household level, we will examine the groupings of the households that are present during the period 2003-2006 to answer questions related to the movement of households between poverty groups and the determinants; e.g. which households move out of or into the poorest group and why? At population level, we will be able to ascertain questions related to the composition of the poverty groups over time, the number of groups and their profiles, e.g. does the proportion of households that are being identified as poor or rich increase or decrease? Does the difference in SES outcomes between the groups increase or decrease? This will give more insight into poverty dynamics.

REFERENCES

Barrientos (2003). "What is the impact of non-contributory pensions on poverty? Estimates from Brazil and South Africa". Chronic Poverty Research Centre Working Paper No 33

Dayton, C. Mitchell and George B. Macready (1988). "Concomitant-Variable Latent-Class Models." Journal of the American Statistical Association 83(401): 173-178.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm (with discussion)." Journal of the Royal Statistical Society B 39: 1-38.

DeWilde, Caroline (2004). "The Multidimensional Measurement of Poverty in Belgium and Britain: A Categorical Approach". Social Indicators Research 68: 331-369.

Filmer D. and Pritchett L. (1998). "Estimating wealth effects without expenditure data -- or tears: An application to educational enrolments in States of India". World Bank Policy Research Working Paper No. 1994, Washington.

Goodman, Leo (1970). "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications." Journal of the American Statistical Association 65: 226-256.

Goodman, L.A. (2002). "Latent Class Analysis: The Emperical Study of Latent Types, Latent Variables, and Latent Structures". J.A. Hagenaars and A.L. McCutcheon (eds.), Applied Latent Class Analysis (Cambridge University Press, Cambridge): pp 3-55.

Hagenaars J.A. (1990). "Categorical Longitudinal Data. Log-linear Panel, Trend and Cohort Analysis". Stage Publications, Newbury park.

Hagenaars J.A. (1993). "Loglinear Models with Latent Variables". Sage University Paper Series on Quantitative Applications in the Social Sciences.

Hagenaars, Jacques A. and Allan L. McCutcheon, eds. (2002). "Applied Latent Class Analysis". Cambridge: Cambridge University Press.

Linzer, Drew A. and Jeffrey Lewis (2007). "poLCA: Polytomous Variable Latent Class Analysis." R package version 1.0. http://dlinzer.bol.ucla.edu/poLCA.

Manly BFJ (1994). "Multivariate statistical methods. A primer". 2nd Edition. London: Chapman and Hall.

McCutcheon, Allan L. (1987). "Latent class analysis". Sage Publications, Newbury Park.

Moisio, Pasi (2004). “A Latent Class Application to the Multidimensional Measurement of Poverty”. Quality & Quantity 38: 703-717.

Rutstein, Shea Oscar and Kiersten Johnson (2004). “DHS Comparative Reports No. 6: The DHS Wealth Index”.

Schwartz, G. (1978). “Estimating the dimension of a model." The Annals of Statistics. 6: 461-464.