The University of Manchester
Global Development Institute

Global Development Institute

Working Paper Series

2016-003

July 2016

ISBN: 978-1-909336-37-7

A harmonised proxy means test for Kenya's National Safety Net programme

Juan M. Villa[1]

[1] Honorary Research Fellow, Global Development Institute, The University of Manchester, United Kingdom.

Email: juan.villa@manchester.ac.uk

Cite this paper as:

Villa, Juan M. [2016] A harmonised proxy means test for Kenya's National Safety Net programme. GDI Working Paper 2016-003. Manchester: The University of Manchester.

Abstract

Four cash transfer programmes are part of what is known as the National Safety Net Programme in Kenya. Current targeting practice of each intervention entails the use different proxy means tests based on the estimation of consumption expenditure from household surveys. This paper presents the new Living Conditions Score (LCS), a proxy means test, which harmonises the identification of households in poverty based on an alternative categorical principal component analysis. Richer household information from the latest national census is employed in this analysis. The new LCS is supported by lower inclusion and exclusion errors as well as better internal validity in identifying households with the lowest living conditions.

Keywords

Anti-poverty programme, Targeting, Proxy means test, Kenya, National Safety Net Programme.

1. Introduction

Targeting has become an essential component for determining eligibility in the implementation of anti-poverty programmes in developing countries. The success of a targeting component of anti-poverty programmes, in particular what Walle (1998) refers to narrow targeting, depends on how accurate, and less distortive, it is to make income transfers reach the poor (Sen, 1995). In developed countries the traditional practice in targeting beneficiaries of social transfers is based on means tests by looking into easy-to-verify individual earnings or income records (Kathy, 2002). Contrarily, given the lack of accurate information on individual earnings and the dominance of an informal economy, several targeting methods have evolved in the anti-poverty practice of developing countries. Community targeting, categorical targeting, self-targeting, proxy means tests or the combination of these have been in the heart of the implementation of anti-poverty programmes all over the developing world (Coady et al., 2004). This paper is focused on proxy means tests and their current application to Kenya's National Safety Net Programme (NSNP).

The absence of reliable income or consumption information poses a real challenge to the identification of individuals or households in poverty in developing countries (Barrientos, 2013, Ch. 6). Proxy means tests (PMTs) are the traditional response when other targeting methods are discarded. What we know about the traditional design of PMTs is mainly based on the prediction of household's welfare by using observational data. In particular, Grosh and Baker (1995) have suggested the construction of PMTs by linear regression, in which the left-hand variable has typically been per-capita income or consumption and, as right-hand variables, a group of household hard-to-hide characteristics. Muller and Bibi (2010) suggest the use of a quantile regression approach which focuses on the quantile around the poverty line to reduce exclusion errors. These methods are highly dependent on the availability of household income-consumption data that require precise calculations. To the extent that one needs to collect or impute information on quantities and prices of purchased goods and services that often are difficult to observe or remember (Deaton, 2016). There is an increasing concern that consumption expenditure and income data from household surveys are not trustworthy due to underreporting or do not make use of appropriate measurements methods. Even in high income countries underreporting of self-employed workers may reach 25 percent (Hurst et al., 2013). Similarly, the collection of consumption data requires time-demanding questionnaires that must follow rigorous expenditure tracking (Deaton and Muellbauer, 1980). For instance, in Kenya consumption data is collected by a household survey administered by an enumerator within a limited time window, instead of following the standard practice, especially in rural areas where auto-consumption is dominant over purchased goods.[1] PMTs inherit these issues. However, as estimands of household per capital consumption expenditure, PMTs are not perfect

[1] For instance, food auto-consumption in rural Kenya averages 43 percent of total food consumption.

but they offer a second-best solution through which a significant amount of participants in anti-poverty programmes receive demand-side transfers around the world.

Yet, the debate continues about the best method to measure a proxied welfare in developing countries. As this paper explores, other methods have advanced in the calculation of PMTs through linear combinations of observable characteristics. These methods bypass the use of consumption and income data and employ the multivariate analysis technique of principal component analysis (PCA). These PMTs are widely used by anti-poverty programmes Latin America, such as the Suiben in Dominican Republic (Lavigne and Vargas, 2013), the former Ficha CAS in Chile (Carneiro et al., 2015), and the Sisben in Colombia (Bottia et al., 2012). In spite of the fact that PCA methods are based on the correlation among observable characteristics without an explicit dependent variable, Filmer and Pritchett (2001) have shown that the results of a welfare measure from a PCA are similar to those of a linear regression with consumption data.

To date, four cash transfer programmes employing consumption expenditure-based PMTs are part of the Kenya's National Safety Net Programme (NSNP) (World Bank, 2013). They are the Cash Transfer for Orphan and Vulnerable Children (CT-OVC), the Older Persons Cash Transfer (OPCT), the Cash Transfer for People with Severe Disability (CT-PWSD) and the Hunger Safety Net Programme (HSNP). The four of them employ several levels of identification of beneficiaries in their targeting process. Proxy means tests, based on household consumption data, play a pivotal role in their implementation, combined with community-based selection and validation. Despite belonging to the same anti-poverty policy, to date each programme acts independently in the delivery of the transfers. This implies that each programme employs a proxy means test and participatory selection with different concepts of poverty. This paper seeks to remedy this practice by proposing a new and harmonised PMT based on census data and constructed by PCA. The new PMT is equipped with the feature that it can be adopted simultaneously by the four cash transfer programmes of the NSNP.

The major objectives of this paper are to present (i) a review of current PMTs employed by the programmes belonging to the NSNP in Kenya, (ii) the development of a new PMT that corrects existing anomalies and harmonises this targeting component of the NSNP by using the available data from household surveys and national census. Instead of relying on consumption expenditure data and linear regression, the new PMT is based on a PCA. Additionally, this paper presents a new targeting questionnaire and the generation of a new PMT formula that integrates the targeting needs and objectives of each cash transfer of the NSNP. By this token, this paper makes a major contribution to the implementation of the cash transfer programmes in Kenya with an analysis that can be replicated in other contexts.

The paper has been organised in the following way. After this introduction the second section relates the background and context that motivates the objectives of this paper. The third section describes and analyses current PMT practices in current cash transfer programmes in Kenya. The fourth section presents the development of a new harmonised PMT for the NSNP. Finally, the fifth section shows the conclusions of the paper.

2. Background

After several years of implementation, poor-targeted cash transfer programmes in Kenya have been recognized as an effective anti-poverty initiative. Several evaluations of the Cash Transfer for Orphan and Vulnerable Children (CT-OVC) have evidenced the effects of this programme on children's human capital and household consumption (Jagero, 2011; Njuga, 2013; Taylor, 2012; The Kenya CT-OVC Evaluation Team, 2012). Cash transfers are relevant for households in poverty in the sense that an additional income lifts the constraints that keep them in poverty, as they can be used for better education prospects for children and a higher adult investment in productive assets and inputs, such as livestock, seeds and fertilisers (Barrientos, 2012). To date, four relevant cash transfer programmes are being implemented by the Kenyan government with international cooperation from multilaterals and bi-laterals organisations. They are the Cash Transfer for Orphan and Vulnerable Children (CT-OVC), the Older Persons Cash Transfer (OPCT), the Cash Transfer for People with Severe Disability (CT-PWSD) and the Hunger Safety Net Programme (HSNP). Taken together, these programmes shape the National Safety Net Programme (NSNP). Even though the four programmes are part of the same anti-poverty policy, they operate with independent targeting methods that attempt to deliver income transfers to households with consumption levels below the poverty line.

A broad aspect of current targeting criteria used by the four cash transfer programmes is detailed in Table 1 below:

Table 1: Targeting methods in the cash transfer programmes of the NSNP

Programme	Programme objective	Targeting
CT-OVC	To provide regular cash transfers to families living with OVC to encourage fostering and retention of children at school and to promote their human capital development.	The identification of priority locations is followed by a two-stage survey to identify beneficiaries. During the first survey, Form 1, local community members identify households that are poor. The second survey, Form 2, is used to gain information from households so that potential beneficiaries can be subjected to a PMT. o

OPCT	To strengthen the capacities of older persons and improve their livelihood while alleviating integrated poverty through sustainable social protection mechanisms. This is to be achieved by providing regular and predictable cash transfer to vulnerable older persons in identified households and building the capacities of beneficiaries in order to improve their livelihoods	Extreme poor households with members 65 years of age and above, not enrolled in other cash transfer programme, a non-recipient of pension, has resided in a particular location for more than a year and must be a Kenyan citizen. Similar to the CT-OVC, the OPCT combines community selection with proxy means test.
CT-PWSD	To support persons with Severe Disabilities who require permanent care, they continue to depend on parents, care givers and well-wishers.	Categorical targeting based on the definition of a person with severe disability at every constituency. The community prepares a list of potential beneficiaries that can apply for the programme.
HSNP-II	To deliver unconditional cash transfers aimed at reducing poverty, food insecurity and malnutrition, and promote asset retention and accumulation.	In order to determine the allocation of total HSNP resources for each of the 4 selected northern and arid counties, a modified version of the Commission for Revenue Allocation (CRA) formula was applied. During the initial registration process for HSNP-II, all households were also required to take part in a wealth ranking exercise. PMT data were used to rank all households registered in wealth order. Household scores were then combined with community validation. The combination of these two rankings provide an overall ranking of eligible households.

Source: with information from GIZ (2013) and implementation manuals.

The Government of Kenya through the Ministry of Labour, Social Security and Services has expressed interest in integrating and harmonising the targeting process of the four cash transfer programmes with the aim of consolidating the NSNP. The main objective of this integration is the generation of synergies that would combat poverty with the same criterion for every beneficiary household. As it can be noted from the Table 1 above, current targeting criteria employed by these programmes are clearly shaped by their objectives in the categorical selection of individual recipients (e.g. children or the elderly). While other levels of selection are determined by categorical criteria and community participation, the basic targeting approach of all four programmes orbits around the identification of households in moderate or extreme poverty. Current poverty identification in the four programmes is being made with different methods and different understanding of proxied consumption expenditure.

3. Existing proxy means tests

The four programmes of the NSNP rely on the collection of a household questionnaire whose information is used to construct a PMT score. Once a community-based screening is complete registry of pre-selected households, enumerators collect the information following programme-specific questionnaires that are then digitalised and converted into a welfare indicator with a defined cut-off point. The prediction of the household consumption is achieved by linear regression employing household data from the Kenya Integrated Household Budget Survey (KIHBS) carried out in 2005/2006.

Five elements were taken into account to look into current PMT practice. First, estimated coefficients can be used to predict household consumption expenditure only if questions are asked preserving their integrity from the primary source. Any alteration to the questions taken from the KIHBS would yield incorrect predictions. Second, questions are sensible, in the sense that the questionnaires are easy to understand and implement by the enumerators and respondents. The administration of the questionnaire is straightforward and time-limited. Third, the resulting PMT formula makes correct use of econometric techniques, to the extent that linear regression assumptions are accomplished and selected variables as statistically significant. Four, the PMT formula makes use of relevant information collected by the PMT questionnaire. Some questions are included in que questionnaire but are not finally used in the PMT formula, this may add noise to the administration of the questionnaire and make it unnecessarily longer. Finally, and more importantly, the PMT formula and the resulting household welfare prediction should be consistent with community participation in the targeting process. If the PMT formula yields a household ordering that does not reflect what the community observes, then the whole targeting process may be at risk of increasing the social costs of the process with undesired consequences for the implementation of the anti-poverty policy.[2]

3.1. PMT in the CT-OVC programme

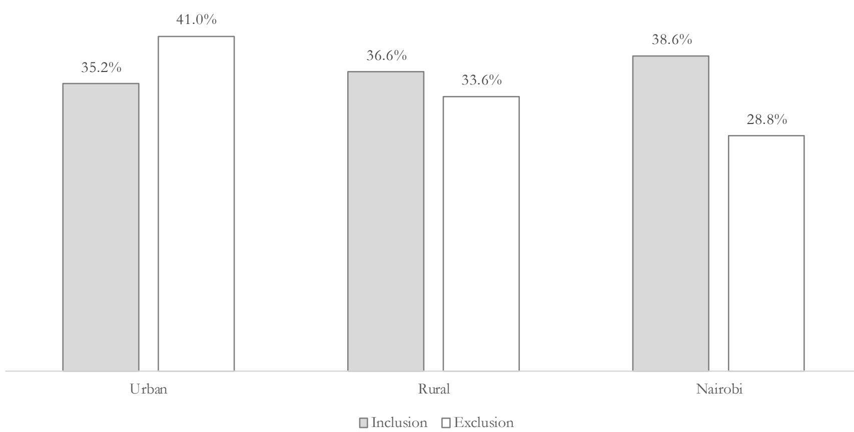
The CT-OVC generates a PMT from 27 questions taken from the KIHBS 2005/2006, 19 of which are shared with the PMT from other programmes. Several points are worth noting from this PMT. First, it does not preserve the definition of household head as specified by the KIHBS and, instead, it changes the household structure in relation to

[2] With respect to the last point, evidence shows that several conflicts could arise from the entire process as people classified as poor with the community may result classified as non-poor by the PMT. For instance, in an evaluation of the community participation in the targeting process of the HNSP, some opinions were collected to illustrate this issue, stating that "*We don't want a computer to pick our poorest we know better than anyone else who is needy here and we should be able to identify them*" (Fitzgibbon, 2014, p. 34).

the main "caregiver." This leads to the fact that some weights from the PMT formula that were estimated for the household head are applied to the "caregiver," mixing two different definitions. Second, the labour force questions in the PMT and the KIHBS questionnaires refer to two different time periods in terms of the economic activity of the members of the households. And third, the options for the dwelling materials are altered from the KIHBS to the PMT questionnaire, a fact that can make the enumerator chose a different answer from a different set of options.

The inclusion an exclusion errors of the CT-OVC's PMT has been obtained by replicating the regression algorithm which predicts household consumption. The predicted household consumption is then compared with the actual consumption and poverty levels. Inclusion errors are defined as the proportion of predicted poor households that are not actually poor. In this case the inclusion errors are between 35.2% and 38.6% in the three areas. This implies that more than one third of those households that would be selected by the PMT would not be actually poor.

Figure 1: inclusion and exclusion errors of the CT-OVC's PMT



Source: Author calculations with KIHBS 2005/2006 data

Similarly, Figure 1 above shows the exclusion errors, that is, the proportion of actual poor households that would not be selected by the PMT. In this case, exclusion errors can be 41 percent in urban areas, 33.6 percent in rural areas and 28.8 percent in Nairobi.

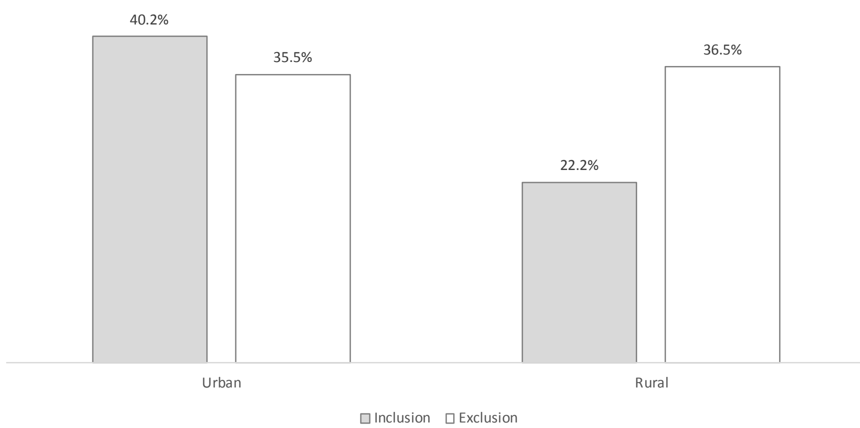
3.2. PMT in the HSNP

The PMT questionnaire for HSNP administers 34 questions and is also based on the KIHBS 2005/2006. In this case the questionnaire also alters the definition and identification of the head of the household by asking about the "Main Provider." However, some questions refer to the head of the household, a concept that is mixed with the "Main Provider." The questionnaire also includes questions that may seem unnecessary, like the ownership of towels or frying pans that are actually included in the PMT formula.

In econometric terms, the PMT formula includes 78 and 62 parameters in urban and rural areas, respectively. However, only 34 and 19 of those parameters are statistically significant in urban and rural areas, respectively. This means that the PMT formula contains parameters that are not relevant to the household consumption but are forced to play an important role in its prediction with an ambiguous consequence, e.g. ownership of livestock is non-significant in Nairobi but forced to be in the PMT formula. For instance, the number of adults working is a non-significant parameter in the PMT formula, which has positive sign in urban areas but a negative sign in rural areas. The negative sign of this parameter in rural areas apparently is inconsistent with the prediction of household consumption, which is an embedded error in the PMT formula.

A similar exercise was done to examine the inclusion errors. In this case, Figure 2 below shows that this PMT leads to inclusion errors of 40.2 percent and 22.2 percent in urban and rural areas, respectively. Exclusion errors are similar in both areas, implying that more than one third of poor households would be excluded by the PMT.[3]

Figure 2: inclusion and exclusion errors of the HSNP's PMT



Source: author calculations with KIHBS 2005/2006 data

3.3. PMT in the OPCT and CT-PWSD programme

The PMT questionnaires and formulas of the OPCT/CT-PWSD are divided in two. The first one consists of the PMT tool that is being actually employed. The questionnaire and PMT formula were derived from an analytical exercise with no specific reference to the prediction of household consumption and the employment of any household survey. The questionnaire is composed of 101 questions that make it the longest questionnaire among the programmes. The weights of the PMT formula were assigned

[3] See Sabates-Wheeler et al. (2014) for an extended evaluation of the HSNP targeting process.

without a clear criterion and apparently without distinction between urban and rural areas.

4. Harmonised targeting approach for the NSNP

A new PMT to be used by the programmes belonging to the NSNP should consider the significant community participation involve in the whole targeting process. Recall that the community plays a pivotal role in the selection of households as leaders first select the households to which the PMT questionnaire is administered. Thereafter, the community validates the lists and household rankings that result from the application of the PMT formula. While existing PMTs are based on the prediction of household consumption expenditure calculated with data of the KIHBS, the communities focus on asset possession, household demographic composition and work category of adult members. For instance, in an exercise with 20 western Kenyan villages, Krishna et al. (2004) collected qualitative information from focal groups on their views of poverty. These views are related with deprivation in several dimensions and apart from a quantitative focus on income or consumption. Only food can be considered as a dimension that a food poverty line can capture with a PMT. Other priorities, such as clothing, housing and poultry are still considered part of a poverty measure. Having cattle, furniture or being able to send children to secondary education are deemed to be part of a higher living standard. In sum, poverty measures defined by the community differ from the one internalised by the programme design, particularly because a technical measure and definition of poverty is not always similar to the one held by the community members. If a PMT reflects a different realty from the one observed from the community, the operation of the cash transfer programme could be compromised.

Several considerations are taken into account for the design of a new PMT for the NSNP. First, Communities do not rank or select households in poverty according to their observed consumption. Second, the definition and construction of the household consumption from a household survey obey to scientific criteria that are not observed by the community, such as imputations (value of the rent of own or provided dwellings and self-provided consumption expenditure) and adult equivalences from statistical agencies. Similarly, household consumption contains items that are acquired on a yearly or weekly basis but the final measure is fetch on a monthly basis. As these facts are not observed by the community, while household consumption is consistent with the measure of national poverty headcount, its combination with the community participation in a targeting process can lead to higher social costs, like grievances and rejection of the selection of beneficiary households. Third, the estimations of household consumption based on linear regressions are endowed with a constant term that assigns a floor of consumption that prevent the programmes from reaching the very poor households that are identified by the community. Therefore, the new PMT proposed here is not based on the direct prediction of household consumption, but based on the generation of a selection score that denotes the living conditions of all households.

4.1. Methodology

The new PMT score has a significant number of attractive features. It provides a linear combination of coefficients that summarises the variance of included variables (Johnson and Wichern, 2002) that, in this case, denote the living conditions of Kenyan households. It is also equipped with similar discriminating power between poor and non-poor households according to the observed consumption expenditure. Instead of relying on linear regression it is calculated by a PCA. Alike linear regression, this method does not directly predict household consumption, does not set a constant term that implies the calculation of a minimum consumption and does not depend on a parametric algorithm that yields significant or not significant parameters of variables. While linear regression is based on the correlation of household characteristics and household consumption, PCA is based on the relation between each household characteristic and the rest of characteristics.

In fact, while linear regression is based on the covariance between consumption expenditure and each explanatory variable, the resulting index or score from the PCA is based on the correlation or covariance matrix of selected variables. For example, a household endowed with electricity can show more consumption expenditure than other without it, but electricity is also related with the ownership of appliances and children's ability to study at night that the PCA is able to account for (Filmer and Pritchett, 2001). On other hand, linear regression yields the predicted household consumption based on a linear formula with a constant term from with household consumption deviates depending on the household characteristics, while The PCA produces an index based on the correlation among household characteristics, no constant term is involved. Finally, the predicted consumption from a linear regression and a score obtained from a PCA are both correlated with actual household consumption: low predicted consumption and low PCA scores are related to low actual consumption. In a comparison of a PCA score with household consumption expenditure carried out by Filmer and Pritchett (2001), it was demonstrated that a PCA score calculated in several contexts was highly correlated with consumption expenditure in Indonesia, Pakistan and Nepal. It was also observed that the PCA is endowed with internal coherence, to the extent that poorest households with consumption data were related with poorest living conditions predicted by the PCA. Similarly, the PCA showed consistency and stability to the inclusion and exclusion of certain variables, demonstrating its robustness. Moreover, predictions by a linear regression were not better than those of a PCA score in terms of internal validity and inclusion errors (Castaño, 2002).

An important issue arising from the estimation of consumption expenditure or PCA is the role of categorical variables interacting with continuous ones. The conventional approach in this matter has been the employment of binary variables for each category, creating as many binary variables as categories. This practice has demonstrated to perform poorly compared with other techniques that quantify those categories and treat them as continuous. On one hand, a group of multivariate methodologies focus on the quantification of ordinal categories through the implementation of optimal scaling methods with PCA, like the one implemented in some Latin American countries (Castaño, 2002). On the other, categorical variables are quantified with the method known as polychoric correlations. Kolenikov and Angeles (2009) offer details and demonstrate that polychoric correlations perform better than other methods used with a PCA. Given the latter, this paper focuses on the use of polychoric correlations for the quantification of ordinal variables in the generation of a new PMT for the NSNP.

4.2. Data

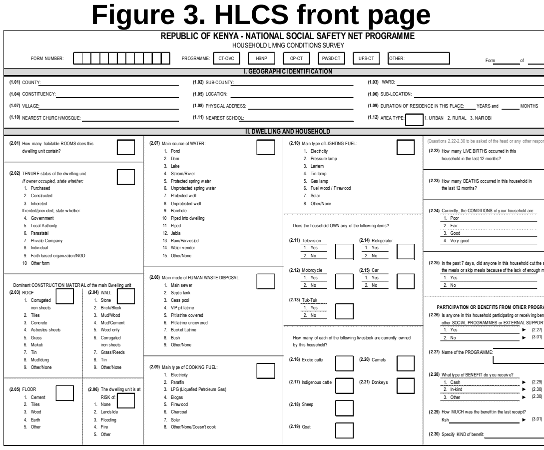
Two micro-data sources are suitable for the development of a new PMT with a PCA in Kenya. First, the KIHBS 2005/2006 contains the information do this analysis with several modules that capture information on households' living conditions, human capital endowment and labour market participation. However, several downsides dominate the generation of a new PMT from the KIHBS. Indeed, as Kenya is divided between counties, sub-counties, divisions, constituencies, locations, sub-locations and villages, the KIHBS is representative only at an aggregate level which does not allow the construction of variables at geographic levels below the county level. Similarly, the KIHBS is based on a sampling method with sampling errors and the fact that it was carried out over a decade ago makes it unreliable for some practitioners.

Second, the 2009 Census is also a suitable micro-data source for the development of a new PMT. The 2009 Census details all variables at sub-location level, with which a richer geographic discrimination is possible for the generation of a PMT. While the Census does not contain the same number of question modules as the KIHBS does, it is endowed with the relevant and sufficient characteristics to develop a PMT tool. The Census is equipped with the modules of household demographics (including recent births and deaths), livestock and assets, living conditions, dwelling construction materials, source of water and the provision of other utilities. As a consequence, since the new PMT is not interested in the prediction of the household consumption, the 2009 Census (with more than 44 million observations) outweighs the KIHBS 2005/2006 (with nearly 66,000 observations) in the sense that no sampling errors are implicit in the analysis, and any geographic disaggregation is possible at any level.

4.3. A new PMT questionnaire

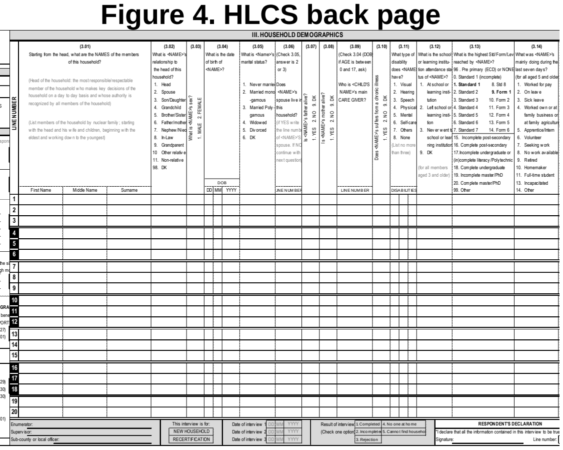
To generate a new PMT for the NSNP, the first step is the definition of a new questionnaire to which we refer to the Household Living Conditions Survey (HLCS). The HLCS is aimed at capturing information on (i) household physical living conditions, such as dwelling materials and the provision of water, sewage, electricity and the cooking fuel; (ii) household endowment of assets like TV, refrigerator (essential for food preservation), car or motorcycle (as a means of transportation and work) and tuk-tuk as a source of income. This section has dismissed the inclusion of assets that are considered unnecessary in the generation of a PMT formula (towels, frying pans, animal carts, etc.); (iii) relevant to rural areas and some regions, the ownership of livestock is captured by the survey. Although in some regions the ownership of them is not highly relevant, the questionnaire should include all of them regardless the region it is administered; (iv) the questionnaire includes questions that may be irrelevant to the PMT formula but help the flow of the questions, while the respondent may feel that these are the ones defining programme participation; and (v) household composition and human capital. The HLCS includes questions that will help us establish the structure of each household and the identification of several nuclear families within the household. Starting from the head of the household, the questionnaire determines the education attainment of each member, as well as his/her health status. These questions contribute to understanding the health and education endowments of the household in terms of human capital, while the household composition helps us characterise the basic social capital endowment (number adults, old members and children, and economic activity).

Figure 3. HLCS front page



Source: author based on Kenya's 2009 Census.

Figure 4. HLCS back page



Source: author based on Kenya's 2009 Census.

The suggested new questionnaire is composed of five parts ordered from the easiest to the most difficult questions (see structure in Figures 3 and 4). It identifies the programme that is collecting the information, the identification of the household and the observable living conditions following literally the questions of the 2009 Census. The dwelling and household module (module II) collects information on the dwelling construction materials, provision of water and the ownership of some assets and livestock. This module also records information on the number of births and deaths in the last 12 months as an indication of health and demographics within the household. It also asks for the respondent's impression of the household status and food security. Finally, this module collects information on the participation of the household in other social transfer programmes. The household demographics module (module III) records information on every member of the household in relation to the household head (as mandated by the 2009 Census). For every child, this module identifies who is the main caregiver and whether his/her parents are still alive. Here persons with disabilities or chronically ill are identified and the education and work status are recorded. The last questions ask for each member's earned income that is not used in the PMT but works as a distraction for those households interested in cheating during the questionnaire administration.

4.4. Results: new PMT formula

To develop a new PMT formula, the analysis here is based on the available information that can be captured by the HLCS and the capacity of the resulting questions to identify households with the worst living conditions. In this sense, the information from the HLCS allows us to identify 177 variables or parameters to be included in a Living Conditions Score (LCS). The LCS is a PCA index rescaled in the range 0 - 100, with 0 denoting the poorest household and 100 the wealthiest. Instead of predicting household consumption, the PCA provides a score, the LCS, based on the correlation of each variable or parameter with the rest of variables included in the analysis. The correlations among the 177 variables or parameters are summarized in one single number on a scale 0 - 100 with different scales in Nairobi, urban and rural areas.

Table 2 below shows the variables taken into account in the PCA that can be drawn from the HLCS or external sources. The fact that this analysis is based on the 2009 Census allows the inclusion of variables at the sub-location level which strengthen the discriminating capacity of the analysis across communities within the same region or county (previous PMTs had been able to include aggregate variables but at the county level). These variables are grouped by:

Region: defined by the county of residence of the household. These are excluded from the Nairobi LCS;

Geographic characteristics at the sub-location level: this includes population, deaths and birth rates. Also the precipitation and elevation data were included as indicators of the conditions of the terrain. Precipitation rates can determine food production, while elevation accounts for weather (it is not the same living in a house with poor construction materials in the coast or in the mountains);

Dwelling conditions and services: this comprises the tenure of the dwelling, main construction materials, human waste disposal and cooking and lighting fuels;

Household assets: ownership of durable assets and those associated with transportation and income generation (tuk tuk);

Livestock: number of animals owned by the household (especially in rural areas). These are excluded from the Nairobi LCS;

Household characteristics: household age and education composition, presence of orphans or chronically ill members;

Labour force: labour conditions of the head and spouse of the household, proportion of workers among adults and presence of child workers; and

Labour force (sub-location): characteristics of the labour force at the sub-location level.

Table 2: Variables in the LCS.

Variable(s)	Description
Regions	
Central	Binary variable for Central region
Mombasa	Binary variable for Mombasa region
Coastal	Binary variable for Coastal region
Upper Eastern	Binary variable for Upper Eastern region
Mid-Eastern	Binary variable for Mid-Eastern region
Lower Eastern	Binary variable for Lower Eastern region
North Eastern	Binary variable for North Eastern region
Nyanza	Binary variable for Nyanza region
North Rift	Binary variable for North Rift region
Central Rift	Binary variable for Central Rift region
South Rift	Binary variable for South Rift region
Western	Binary variable for Western region
Geographic characteristics	
Population (sub-location)	Number of inhabitants at the sub-location level
Death rate (sub-location)	Number of deaths/1000 inhabitants at sub-location level
Birth rate (sub-location)	Number of births/1000 inhabitants at sub-location level
Mean precipitation (sub-location)	Mean annual precipitation rate (ml) at sub-location level
Mean elevation (sub-location)	Mean elevation (mts) at sub-location level
Dwelling conditions and services	
Dwelling tenure	Tenure status of the dwelling unit
Household size (members)	Number of household members
Rooms per persons	Number of habitable rooms per persons
Wall construction material	Dominant construction material of the walls
Roof construction material	Dominant construction material of the roof
Floor construction material	Dominant construction material of the floor
Main source of water	Main source of water (for all purposes)
Main mode of human waste disposal	Main source of human waste disposal
Main type of cooking fuel	Main type of cooking fuel
Main type of lighting fuel	Main type of lighting fuel
Household assets	
TV	Household or any household member owns a TV
Motorcycle	Household or any household member owns a motorcycle
Car	Household or any household member owns a car
Refrigerator	Household or any household member owns a refrigerator
Tuk tuk	Household or any household member owns a tuk tuk
Livestock	
Number of exotic cattle	Number of exotic cattle
Number of Indigenous cattle	Number of Indigenous cattle
Number of sheep	Number of sheep
Number of goat	Number of goat
Number of camel	Number of camel
Number of donkeys	Number of donkeys
Household characteristics	
Male head	Head of the household is a male
Spouse in the household	Spouse of the head of the household lives in the household
Monogamous head marriage	Head of the household is monogamously married
Proportion of male members	Number of males / household size
Age of head of the household	Age of the head of the household
Mean age of the household	Sum of all members' age / household size
Age of spouse	Age of head's spouse
Dependency ratio	Number of members under 15 and over 65 / members between 15 and 65 years of age
Proportion of children under 6	Number of children under 6 / household size.
Orphan in the household	Whether there is an orphan child in the household (mother or father not alive)

Proportion of children under 12 attending school	Number of children under 12 attending school / total number of children under 12
Head's education	Education level of the head of the household
Spouse's education	Education level of the head's spouse (if more than one spouse, pick the highest)
Maximum education of any member	Maximum education of any member, including head and spouse
Any member 15-65 with disability	Any member in working age with disability
Proportion of household members dead in the last 12 months	Number of deaths in the last 12 months / household size
Proportion of household members born in the last 12 months	Number of live births in the last 12 months / household size
Labour force	
Head works	Head of the household worked in the last seven days
Spouse works	
Proportion of working members 15-65	Number of members between 15-65 years of age who worked in the last seven days/Number of members 15-65 years of age
Proportion of working children 6-15	Number of members between 15-65 years of age who worked in the last seven days/Number of members 6-15 years of age
Labour force (sub-location)	
Proportion of wage workers (sub-location)	
Proportion of agricultural workers (sub-location)	
Proportion of self-employed (sub-location)	

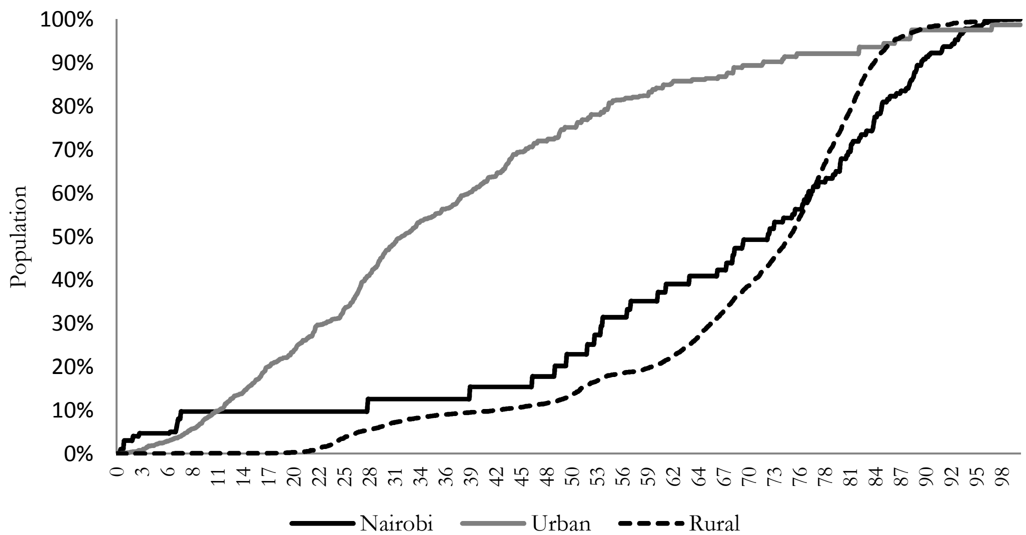
Source: Author based on the HLCS.

The variables in the PCA, that lead to the LCS (the new PMT score), are not directly associated with the prediction of household consumption expenditure, but rather, they are correlated with the rest of parameters that lead to the definition of the LCS.

Due to confidentiality restrictions, the weights of the LCS are not presented here. However, some findings are worth noting. The weights or scoring coefficients are small or large according to their contribution to the final score. As it can be seen, unlike the regression analysis previously used for the calculation of the PMT score, the presence of variables in the PCA is not determined by their significance in the prediction of household consumption but by the extent to which they contribute to the LCS. For instance, the PMT of the CT-OVC only considers the wall materials of stone and wood, while the others are excluded because they have low prediction power. Instead, in the PCA that leads to the LCS all construction materials are in the formula ordered according to their relevance. Therefore, the new LCS is a PMT considered more consistent with community participation as (i) none of the household characteristics are omitted from the formula and (ii) household consumption, which is not directly observed by the community, is not predicted.

Figure 5 below shows the poverty estimates of the LCS for each geographical area. As it can be seen, the proportion of the population below the score in Nairobi and Rural LCS increases more slowly than the one in urban areas. However, make no mistake: these three scores are not comparable. One advantage of the LCS is that there is no fixed cut-off point. In fact, the cut-off point for the selection of potential beneficiaries is determined by the defined coverage of each programme. To cover of 50 per cent of the population the cut-off score should be 72.1 in Nairobi, 31.6 in Urban and 74.5 in Rural areas.

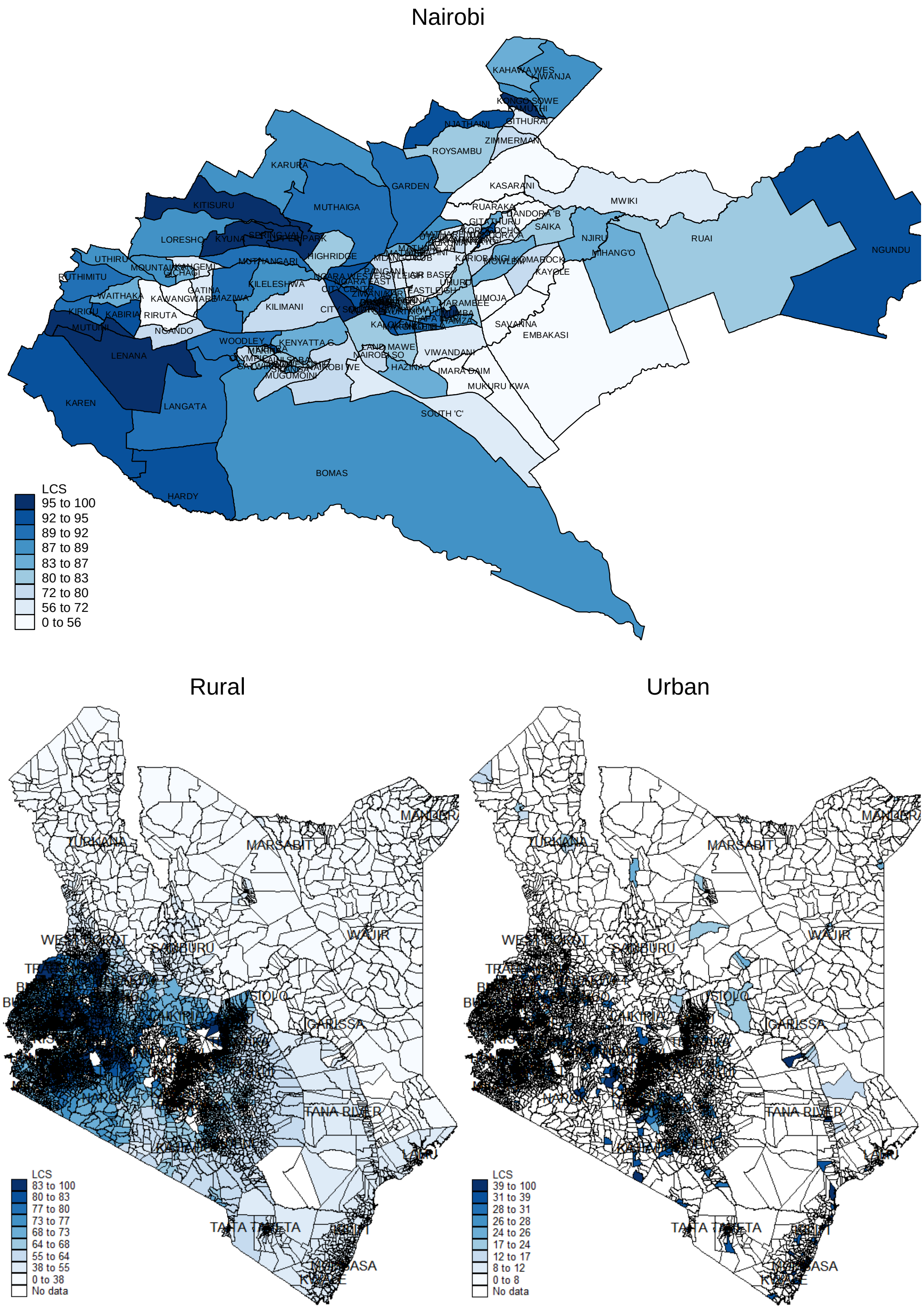
Figure 5. Population below the LCS.



Source: author with data from 2009 Census.

Figure 6 below shows a set of maps of the average of the LCS by sub-location for Nairobi, rural, and urban areas. Colours vary in a range of deciles of the LCS:

Figure 6. Mapping of the LCS in rural, urban and Nairobi areas.



Source: author with data from 2009 Census.

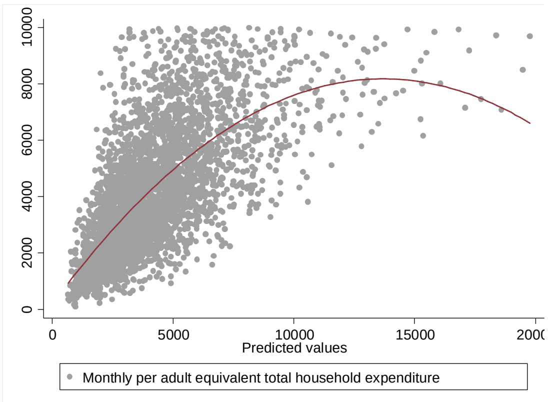
It is apparent from the maps above that the areas with the best living conditions in Nairobi are located around the sub-location of Central Square and the west part of the city. By the same token, the rural map shows that living conditions are better in the south-west of the country and around Nairobi and Mombasa, while the northern counties of the country are highly deprived, such as Turkana, Marsabit and Wajir and Mandera. Finally, the urban map reveals many sub-locations with no data as most of the country is predominantly rural. There are 947 sub-locations with urban households only, very few of them are notably deprived.

4.5. Comparison with KIHBS 2005/2006

An additional exercise is to compare the PMT obtained from a linear regression (existing method) and the LCS. Only one source of data allows this comparison, that is, the KIHBS 2005/2006. This exercise uses the variables used in the construction of the LCS with census data but taken from the KIHBS and introduced in a PCA and a linear regression for comparison purposes. However, sub-location level data cannot be used and, as a consequence, the capacity of the LCS is compromised, thus the resulting LCS is a reduced version of the one obtained with the 2009 Census. The same variables were used for both methods. However, the linear regression output results from a two-step procedure in which all variables are included in the analysis and then only significant variables at 5 percent are left in a second approach. As a first approach, Figures 7 and 8 show respectively the actual household consumption on the vertical axis against the predicted consumption from a conventional linear regression (like the one currently used by the programmes) and the new LCS adapted in the KIHBS 2005/2006. It is apparent from this comparison that the LCS has a better fit to the observed consumption data than the prediction obtained from a linear regression, allowing for a better dispersion and more targeting efficiency: the concentration of households around the origin (close to zero) of the predicted income leads to higher exclusion errors, this is not the case with the LCS.

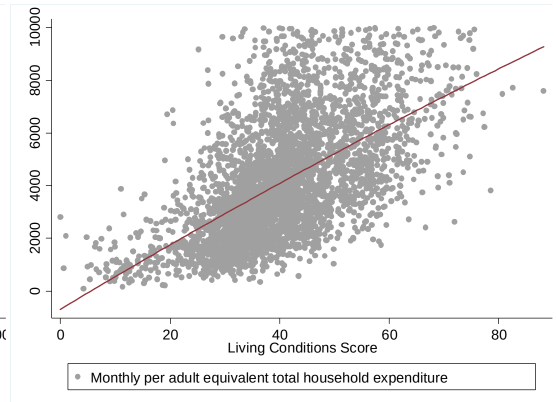
There is not a defined cut-off point for the LCS. In a hypothetical case that cash transfer programmes were to select 20 percent of the poorest population of the country, the exclusion and inclusion errors of the LCS and the predicted consumption would be 27.6 and 27.5 percent, and 35 and 26.2 percent, respectively.

Figure 4: Actual household consumption vs. predicted consumption.



Source: Author with data from KIHBS 2005/2006.

Figure 5: Actual consumption vs. an adapted version of the LCS in the KIHBS.



Source: Author with data from KIHBS 2005/2006.

An additional exercise of internal validity of the LCS and the predicted consumption consists of the comparison of the variables included in both PMTs. Table 3 shows that the LCS is equipped with a better discrimination of the bottom 40 percent of households with lowest scores than the predicted consumption in rural areas. For instance, while the 55.9 percent of poorest households according to the LCS live in dwellings with roofs of grass/mud/dung, 42 percent of households according to the predicted consumption live with the same condition. The LCS also has better discrimination of the top 20 percent of the households, it discriminates those with the best living conditions. For instance, while 19.3 percent of the to 20 percent of households according to the LCS live in dwellings with stone-made walls, 16.3 percent of households in the top 20 percent according to the predicted consumption live in the same condition. The LCS also has a better identification of households with high dependency ratio and with orphan children. Finally, the bottom of the table shows the poverty indicators according to consumption data. According to the LCS, the bottom 40 percent of households have consumption expenditure slightly higher than the one indicated by the predicted consumption method. This is an expected fact, as the LCS is not based on the prediction of household consumption expenditure. However, this translates into a lower poverty headcount incidence in the bottom 40 percent in term of absolute and extreme poverty.

Table 3: Comparison of the LCS from a PCA with predicted consumption from a linear regression.

Rural areas	LCS			Predicted consumption		
	Bottom 40%	Middle 40%	Top 20%	Bottom 40%	Middle 40%	Top 20%
Wall construction material						
Grass/Reeds	13.4	0.09	0.00	12.4	1.03	0.12
Stone	0.37	3.15	19.3	1.15	3.83	16.3
Roof construction material						
Grass/Mud/Dung	55.9	4.43	0.12	42.9	15.4	4.30
Corrugated iron	24.4	93.8	97.3	40.3	80.6	91.8
Floor construction material						
Earth	95.8	78.2	35.8	95.0	78.0	38.0
Cement	3.12	20.1	61.7	3.83	20.3	59.8
Water provision						
River/pond/stream/Dam	35.2	28.7	18.1	33.0	30.9	18.0
Piped	1.59	5.70	19.6	1.28	6.08	19.51
Ownership of assets						
TV	0.72	6.98	34.5	0.59	7.39	34.0
Motorcycle	0.09	0.44	4.11	0.06	0.22	4.61
Refrigerator	0.03	0.12	2.43	0.03	0.16	2.37
Ownership of livestock						
Cattle	7.85	13.1	17.3	11.7	11.7	12.4
Sheep/goats	11.2	15.8	16.2	16.1	13.4	10.9
Donkey	3.93	3.05	2.12	5.02	2.49	1.06
Household						
Male head	68.4	67.1	76.6	68.0	68.1	75.5
Dependency ratio	43.7	36.0	22.2	37.4	40.0	26.9
Orphan in household	20.9	16.0	9.73	17.9	18.8	10.2
Head illiterate	87.7	66.2	27.6	90.6	62.3	29.7
Early mother (< 15 yo)	0.16	0.19	0.19	0.06	0.19	0.37
Poverty						
Food expenditure (Kshs)	1,087	1,336	1,859	937	1,376	2,075
Total expenditure (Kshs)	1,525	2,069	3,536	1,290	2,120	3,902
Poverty headcount	69.4	40.3	12.3	73.0	37.3	9.16
Extreme poverty	38.6	16.5	3.18	41.0	10.9	1.50

Source: author based on KIHBS 2005/2006.

An additional comparison here focuses on the ranking performance of the linear regression approach and the LCS. As the community participation of the NSNP plays a pivotal role in the selection of beneficiary households, it is worth examining how the PMTs rank the population from the poorest to the wealthiest household. It is important that the PMT prevent the community from raising grievances or complaints against the objective selection of households. To complete this comparison, it is necessary to note the difference between the observed ordering of households according to their income or consumption (ord^o) and the predicted ordering of the PMT (ord^p). Thus, a new ranking performance (*RP*) formula is introduced here:

$$RP = \frac{\sum_i^N |ord_i^o - ord_i^p|}{(N)^2} * 2 \qquad (1)$$

It denotes the extent to which the predicted or ranking or order of households deviates from the observed order in relation to the potential ordering deviation of the population (N^2). The lower the *RP*, the lower the difference between predicted and observed orders. In other words, an *RP* of 0 denotes perfect ranking performance and an *RP* of 1 denotes complete imperfect ranking performance. As an illustrative example, Table 4 below shows three cases of RP based on hypothetical predicted orderings.

Table 4: Ranking performance of linear regression and LCS predictions.

Unit	Observed welfare ranking	RP = 0 (1)	RP = 100 (2)	RP = 50 (3)
A	1	1	6	2
B	2	2	5	1
C	3	3	4	5
D	4	4	3	3
E	5	5	2	6
F	6	6	1	4

Source: author.

The application of the *RP* formula is shown in Table 4 below. As it can be observed, the ranking of households according to total equivalent household consumption performs better for the linear regression prediction, with a particular relevance in rural areas. This is not surprising, as the linear regression is based on the estimation of this variable. On the other hand, if equivalent adult scales are omitted, the ranking of households according to total consumption expenditure and food expenditure performs better with the LCS, except for purchased food expenditure in rural areas. In urban areas, the ranking of households with the LCS outperforms that of linear regression.

Table 5: Ranking performance of linear regression and LCS predictions.

RP according to…	Rural		Urban	
	Linear regression	LCS	Linear regression	LCS
Total equivalent adult consumption	38.93	47.18	30.01	35.34
Total consumption expenditure	55.12	53.70	51.65	44.96
Food expenditure	59.81	58.25	60.02	55.03
Purchased food expenditure	58.68	59.96	60.25	55.51
Auto consumption food expenditure	65.93	62.86	60.31	57.99

Source: author based on KIHBS 2005/2006.

Taken together, this results show that an alternative PMT can be used to harmonise the NSNP. The LCS uses richer data from the national Census and provides similar predictive capacity of poverty as the predicted consumption-based PMT does. The algorithm based on the PCA is based on the correlation of all the variables included in the analysis, instead on focusing on household consumption expenditure. A priori, it can be considered that the LCS is more friendly to the community participation component of the whole targeting process, as it does no focus on other parameters that the community members do not see, such as the imputed factors of household consumption expenditure.

5. Conclusions

This paper has been focused on a new harmonised proxy means test (PMT) for the Kenyan National Safety Net Programme (NSNP). Currently, four cash transfer programmes are part of the NSNP but are still being implemented with different targeting processes, which involve several versions of a PMT. The main objectives of this paper were to describe current PMT practice in Kenyan cash transfer programmes and to propose a new formula and algorithm based on principal component analysis and the employment of the national census. Two main PMT methods were tested. The principal component analysis facilitated the calculation of a living condition score (LCS) which was compared to the conventional approach based on the estimation of consumption expenditure. The paper showed that the LCS performs similar to the estimation of the consumption expenditure, with the ability of employing census data with a focus on variables with geographical representation at sub-location levels.

The analysis of this paper provides an important input for the implementation of the NSNP with a harmonised PMT. A new questionnaire was presented with a structure that can combine the calculation of the LCS and the collection of categorical data that help the interventions to identify eligible households. Thus, the new LCS allows several practical applications. Firstly, the LCS facilitates the standardisation of the information of beneficiaries or eligible households for the administration of the four programmes. Second, the standardisation of beneficiary information also facilitates the creation of synergies among programmes, in the sense that the HLCS collected by one programme can be used by the others. Third, the standardisation of household information and the creation of synergies can be reflected on potential savings in resources and costs that can be instead allocated to other aspects of the implementation of each programme.

This paper also extends our knowledge on the general PMT approach of social transfers in developing countries. Several methods have been tested and compared in the literature in terms of their efficiency in reducing leakage and under coverage of the poor. The PCA approach employed here has shown here that it can achieve efficiency rates as good as other regression-based methods. The employment of the PCA here has been motivated by the fact that the Kenya's NSNP operates with a relevant community participation component, where the calculation methods of household per capita consumption expenditure is barely observed by community members. In particular, in rural communities where auto-consumption accounts for almost a half of total consumption expenditure the estimation by linear regression has shown little acceptance by community members. In this sense, the LCS derived from a PCA showed better ranking performance than the PMT obtained from a linear regression.

Therefore, the PCA presented in this paper is, to some extent, designed to addressed these issues and can inspire its replicability in similar contexts.

An issue that was not addressed in this paper was the comparison of the full version of the LCS with the estimated consumption expenditure approach. The use of the national Census allowed the employment of rich household data and the generation of geographical variables at low levels of sub-national divisions. A reduced version of the LCS was used in the comparison of the LCS with the estimated consumption expenditure approach based on the KIHBS, which did not facilitate the disaggregation of variables at sub-national levels. Notwithstanding this limitation, the paper found that the LCS has an outstanding performance in terms of inclusion and exclusion errors and association with observed living conditions.

6. References

Barrientos, A., 2013. Social Assistance in Developing Countries. Cambridge University Press, Cambridge, UK.

Barrientos, A., 2012. Social Transfers and Growth: What Do We Know? What Do We Need to Find Out? World Development 40, 11–20. doi:10.1016/j.worlddev.2011.05.012

Bottia, M., Sosa, L.C., Medina, C., 2012. El SISBEN como mecanismo de focalización individual del régimen subsidiado en salud en Colombia: ventajas y limitaciones. Revista de Economía del Rosario 15.

Carneiro, P.M., Galasso, E., Ginja, R., 2015. Tackling social exclusion: evidence from Chile. World Bank Policy Research Working Paper.

Castaño, E., 2002. Proxy means test index for targeting social programs : two methodologies and empirical evidence. Lecturas de Economía 133–144.

Coady, D., Grosh, M.E., Hoddinott, J., 2004. Targeting of Transfers in Developing Countries: Review of Lessons and Experience. World Bank Publications.

Deaton, A., 2016. Measuring and Understanding Behavior, Welfare, and Poverty. American Economic Review 106, 1221–1243. doi:10.1257/aer.106.6.1221

Deaton, A., Muellbauer, J., 1980. Economics and Consumer Behavior. Cambridge University Press.

Filmer, D., Pritchett, L.H., 2001. Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. Demography 38, 115–132. doi:10.2307/3088292

Fitzgibbon, C., 2014. HSNP Phase II Registration and Targeting Lessons Learned and Recommendations (Final Report). UK Department of International Development (DFID).

GIZ, 2013. Identification of the Poor: An Overview of Current Targeting Approaches and their Applicability in Developing Countries with a Focus on Kenya. GIZ Health Sector Program in Kenya.

Grosh, M., Baker, J.L., 1995. Proxy means tests for targeting social programs. Living standards measurement study working paper 118, 1–49.

Hurst, E., Li, G., Pugsley, B., 2013. Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed. Review of Economics and Statistics 96, 19–33. doi:10.1162/REST_a_00363

Jagero, N., 2011. An analysis of social costs of secondary education in Kenya after the introduction of subsidised secondary education in 2008. International Journal of Education Economics and Development 2, 213–224. doi:10.1504/IJEED.2011.042402

Johnson, R.A., Wichern, D.W., 2002. Applied Multivariate Statistical Analysis. Prentice Hall.

Kathy, L., 2002. Survey of Social Assistance in OECD Countries (Volume I: Cross-Country Paper). Abt Associates Inc.

Kolenikov, S., Angeles, G., 2009. Socioeconomic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer? Review of Income and Wealth 55, 128–165. doi:10.1111/j.1475-4991.2008.00309.x

Krishna, A., Kristjanson, P., Radeny, M., Nindo, W., 2004. Escaping Poverty and Becoming Poor in 20 Kenyan Villages. Journal of Human Development and Capabilities 5, 211–226.

Lavigne, M., Vargas, L.H., 2013. Social protection systems in Latin America and the Caribbean: Dominican Republic.

Muller, C., Bibi, S., 2010. Refining Targeting against Poverty Evidence from Tunisia*. Oxford Bulletin of Economics and Statistics 72, 381–410. doi:10.1111/j.1468-0084.2010.00583.x

Njuga, E.M., 2013. Unicef cash transfer programme for orphans' and vulnerable children: a case study of its effectiveness in kilifi county, kenya. (Thesis). University of Nairobi.

Sabates-Wheeler, R., Hurrell, A., Devereux, S., 2014. Targeting social transfer programmes: Comparing design and implementation errors across alternative mechanisms (Working Paper Series No. UNU-WIDER Research Paper WP2014/040). World Institute for Development Economic Research (UNU-WIDER).

Sen, A., 1995. The Political Economy of Targeting, in: Public Spending and the Poor. Johns Hopkins University Press, Baltimore and London.

Taylor, J.E., 2012. A Methodology for Local Economy-Wide Impact Evaluation (LEWIE) of Cash Transfers (Working Paper No. 99). International Policy Centre for Inclusive Growth.

The Kenya CT-OVC Evaluation Team, 2012. The impact of the Kenya Cash Transfer Program for Orphans and Vulnerable Children on household spending. Journal of Development Effectiveness 4, 9–37. doi:10.1080/19439342.2011.653980

Walle, D. van de, 1998. Targeting Revisited. World Bank Res Obs 13, 231–248. doi:10.1093/wbro/13.2.231

World Bank, 2013. Kenya National Safety Net Program for Results (Technical Assessment). World Bank.