

# Technical note on generating market heat maps<sup>1</sup>

Pablo Acosta, Namsuk Kim and Illana Melzer  
Working Paper, Office of Development Studies, UNDP

July 2008

## Contents

1. Introduction.....	2
2. Survey Data.....	2
3. Income.....	2
4. Poverty .....	4
5. Locality .....	6
6. Employment.....	6
7. Housing ownership .....	9
8. Construction materials for the house .....	9
9. Water.....	9
10. Sanitation .....	9
11. Electricity.....	10
12. Telephone.....	10
13. Information and Communication Technology.....	10
14. Financial Access .....	10
15. Remittance .....	12
Reference .....	16

Annex 1: Analysis of Potential Nonresponse Bias in FinScope Data -Using Zambia 2005 Survey

---

<sup>1</sup> This paper was prepared as a background document for *Value for All: Strategies for Doing Business with the Poor*, UNDP, 2008 (<http://www.growinginclusivemarkets.org/>). Questions should be addressed to [namsuk.kim@undp.org](mailto:namsuk.kim@undp.org).

## 1. Introduction

The following note presents technical issues regarding generation of a market heat map, an analytical tool proposed in UNDP (2008). The definition and use of the heat map is illustrated in detail in Acosta et al (2008). The market heat map could convey information on the nature and extent of market inclusiveness of the poor. The specific focus is on the poor, defined as people living on less than \$2 a day; but for comparison, data on the non-poor can also be reported in the market heat map. The heat map can also be presented using disaggregated data by geographical region to serve the purpose of analysis at a sub-national level.

## 2. Survey Data

Data for constructing market heat maps come from national representative household survey data on living conditions. For African countries, i.e. Botswana, Namibia, South Africa and Zambia, the data are individual level surveys because not all household members are necessarily selected in the sample. The specific household surveys employed are:

- Bolivia (2002): Encuesta de Hogares, MECOVI
- Botswana (2004): FinScope, FinMark Trust
- Dominican Republic (2004): Encuesta de Condiciones de Vida, ENCOVI
- Guatemala (2000): Encuesta Nacional sobre Condiciones de Vida , ENCOVI
- Haiti (2001): Les Conditions de Vie en Haiti, IHSI/Fafo/UNDP
- Mexico (2002): Encuesta Nacional de Ingresos y Gastos de los Hogares, INEGI
- Namibia (2004): FinScope, FinMark Trust
- Pakistan (2004): Social and Living Standards Measurement Survey, Federal Bureau of Statistics
- Philippines (2003): Family Income and Expenditure Survey, National Statistical Office
- South Africa (2006): FinScope, FinMark Trust
- Zambia (2005): FinScope, FinMark Trust

### *Weights*

All the survey executing agencies provide the individual and household sample weights in the data sets. We use these official weights to compute all the statistics in the database.

## 3. Income

### *Income or Consumption*

It is well known that household consumption would be a better proxy for well-being than household income.<sup>2</sup> The main reasons listed in the literature can be summarized as follows: (i) if people can lend and borrow current consumption is closer to permanent income (or consumption) than current income, (ii) differential under-reporting by strata is usually a more severe problem for income than for consumption, and (iii) incomes are frequently reported before taxes, while consumption is an after-tax concept.

However, many of the heat maps can only use income data simply because there are not many countries collecting information on expenditure at the household level regularly. We use income data to identify the poor in general for convenience, and for international comparison. Incomes/expenditures are transformed into monthly basis if values are not reported on a monthly basis.

*Bolivia, Dominican Republic, Guatemala, Haiti, and Mexico*

Each map starts by classifying households as belonging to the poor or not, and for these countries, we use household income to identify the poor. First, it is necessary to compute per capita income measures per household. Household survey data typically include several income categories, those considered in the countries listed above are the following:

- Wages, bonuses
- Self-employment and entrepreneurial income
- Capital gains
- Rents
- Interests
- Transfers (domestic or international)

Our income measures do not typically include:

- Non-market agricultural production
- Imputed rent
- Imputed home production
- Asset sales
- Extraordinary incomes

*Botswana, Namibia, South Africa, and Zambia*

Per capita income has been calculated using data on household income as reported by FinScope surveys and dividing this by the number of people in the household. Household income data in the surveys is presented in terms of income bands. The mid-point of the income band has been used as a point estimate for household income. No adjustment has been made for the number of children in the household.

---

<sup>2</sup> Deaton and Zaidi (2002)

The number of households with missing information on the income is worrisome in some countries in Africa. If the missing observations are not randomly distributed in the income profile, the estimates from the survey might be biased. The Annex 1 shows that there might be a bias in the estimates in heat maps, but the sign of bias is constant which allows us to interpret the estimates as the upper bound of indicators.

#### *Pakistan and Philippines*

Per capita expenditure is used to identify the poor. The per capita expenditure aggregate in the survey is provided by the national statistical offices.

## **4. Poverty**

### *Methodological issues*

The measurement of the poor has been argued politically and economically in academic and practical reports, because there are neither normative nor objective clear arguments to set a single line that separates the poor and the non-poor (Deaton, 1997). Poverty reduction is one of the main policy goals set by international communities and the governments, although there is a huge conceptual ambiguity.

Because the definition of poverty is basically arbitrary, different definition has been used by different policy makers and researchers. Among others, three definitions of poverty have been widely used: (i) international poverty lines (USD 1 a day and USD 2 a day at PPP), (ii) national poverty lines (extreme and moderate), and (iii) relative poverty lines (50% of median income). Each definition has pros and cons, and should be chosen to best support the economic/social analysis and policy design.

The heat maps in UNDP (2008) adopt the first definition, especially USD 2 a day at PPP to identify the poor, mainly for computational preference and regional/international comparison purpose. Since the main focus of the heat maps is to look at the market penetration of the poor, authors can use different definitions to generate heat maps using other definitions if they suite their purpose better.

### *International PPP poverty lines*

The USD 1 a day at PPP prices is an international poverty line meant to define an international standard to measure the inability to pay for basic needs. The USD 2 a day at PPP price is also widely used, and usually refers to "moderate poverty." Ravallion and others (1991) proposed the USD1-a-day line and it has been used in various reports by international entities, such as World Bank and United Nations. In the original concept, it is a value measured in 1985 international prices and adjusted to local currency using purchasing power parities (PPP) to take into account local prices. The USD 1 standard

was chosen as being representative of the national poverty lines found among low-income countries.

Chen and Ravallion (2001) revised the poverty line in 1993 PPP terms at \$1.0763 a day. The value is multiplied by 30.42 to get a monthly poverty line. The USD-1-a-day and USD-2-a-day line are extensively used in cross country comparisons. Since it is simple and easy to calculate, it has been widely used for international poverty comparison. For instance, the United Nations' Millennium Development Goal 1 – eradicate extreme poverty and hunger – is stated in terms of USD-1-a-day poverty – halving between 1990 and 2015 the proportion of people whose income is less than USD 1 a day.

It is important to notice that our poverty figures may differ from other existing studies, such as World Development Indicator (WDI) and Socio-economic Database for Latin America and the Caribbean (SEDLAC), even if we used the same USD-2-a-day poverty line and the same survey data. The main reason is that our income/expenditure categories employed typically differ. For instance, SEDLAC usually includes estimates from in-kind labor income, agriculture production, home production and imputed rent. WDI excludes households reporting zero income, while we include them in our estimates. The share of households/individuals in each country who live on USD1/USD2 per day is as follows:

*Bolivia, Dominican Republic, Guatemala, Haiti, and Mexico*

Percentage of households under international poverty lines				
Country	Living on \$1 per day or less		Living on \$2 per day or less	
	Rural	Urban	Rural	Urban
Mexico	9.31	0.74	29.01	4.05
Bolivia	40.24	4.95	64.50	17.56
Dominican Rep	7.94	4.75	19.09	10.87
Guatemala	35.74	7.68	56.93	18.77
Haiti	67.36	28.24	84.74	46.28

*Botswana, Namibia, South Africa, and Zambia*

Percentage of Adults under international poverty lines		
Country	Living on \$1 per day or less	Living on \$2 per day or less
Zambia	56.86	76.47
Namibia	25.00	38.23
Botswana	28.57	42.85
South Africa	5.40	13.51

*Pakistan*

Percentage of population under international poverty lines				
Country	Living on \$1 per day or less		Living on \$2 per day or less	
	Rural	Urban	Rural	Urban
Pakistan	2.46	0.86	40.50	23.26

## *Philippines*

### Percentage of individuals with international poverty lines

	Living on \$1 per day or less	Living on \$2 per day or less
Total	13.09	49.74

## **5. Locality**

The Heat Map can be presented at any locality if the relevant information is provided in the data. The typical heat maps are displayed by urban and rural. The only exception is Philippines heat maps which are presented by 17 different regions, because the urban/rural information is not yet publicly available in Philippines 2003 household survey.

## **6. Employment**

Population in working age is divided to three groups: Employed, unemployed, and non-labor-force. The employment status can follow different definitions. For the heat map, we use the definition provided by the national statistical offices, and present the data by *working* (employed) and *not working* (unemployed and non-labor-force). As described in UNDP (2008), we consider the entire population in working age instead of focusing on labor force only, because heat maps based on labor force only would not be able to incorporate the non-labor-force who are discouraged by extended joblessness, social norms, and other factors.

## *Philippines*

Employment status concepts are as defined as follows in the technical notes by Philippines National Statistical Office.

(<http://www.census.gov.ph/data/technotes/notelfs.html#concepts>)

Labor Force or Economically Active Population. This refers to population 15 years old and over who are either employed or unemployed in accordance with the definitions described below.

Employed. Employed persons include all those who, during the reference period are 15 years and over as of their last birthday and are reported either:

- a. At work. Those who do any work even for one hour during the reference period for pay or profit, or work without pay on the farm or business enterprise operated by a member of the same household related by blood, marriage or adoption; or

- b. With a job but not at work. Those who have a job or business but are not at work because of temporary illness/injury, vacation or other reasons. Likewise, persons who expect to report for work or to start operation of a farm or business enterprise within two weeks from the date of the enumerator's visit, are considered employed.

Underemployed. Underemployed persons include all employed persons who express the desire to have additional hours of work in their present job or an additional job, or to have a new job with longer working hours. Visibly underemployed persons are those who work for less than 40 hours during the reference period and want additional hours of work.

Unemployed. Unemployed persons include all those who, during the reference period are 15 years old and over as of their last birthday who have no job/business and actively looking for work. Also considered as unemployed are persons without a job or business who are reported not looking for work because of their belief that no work was available or because of temporary illness/disability, bad weather, pending job application or waiting for job interview.

Persons Not in the Labor Force. Persons 15 years old and over who are neither employed nor unemployed according to the definitions mentioned. Those not in the labor force are those persons who are not looking for work because of reasons such as housekeeping, schooling, etc. Examples are housewives, students, disabled or retired persons.

Determination of Employment Status. The employment status of persons 15 years and over is determined on the basis of answers to a series of inter-related questions which are described below:

- a. "Did \_\_\_\_ do any work at all even for only one hour during the past week?" This question is asked to identify the employed persons. "Work at all" for purposes of this survey means that a person reported to his place of work and performed his duties/activities for at least one hour during the reference week. If a person reported that he or she did some work, not counting chores around the house, he is still considered in the employed category although most of his time was devoted to household chores. All persons not identified by the above question as employed are asked the following questions.
- b. "Although \_\_\_\_ did not work, did ---- have a job or business during the past week?" Some persons may not have work at all during the past week but may actually have jobs or businesses which they are temporarily not reporting to, as in the following cases: an employee on strike; a person temporarily laid off due to non-economic reasons like machine breakdown; a person with a new job to begin within two weeks from the date of interview; regular and temporary teachers, excluding substitutes, during summer vacation who still receive pay and who expect to go back to their jobs in the next school year. These persons are considered employed even though they are not actually at work.

- c. “Did \_\_\_\_\_ look for work at any time during the past week?” This question is asked to determine who among those who had no job/business had really done something to look for work. If a person looked for work, he or she is classified as unemployed, otherwise, the next question asked is to determine whether a person should be classified as unemployed or not in the labor force.
- d. “Why did \_\_\_\_\_ not look for work?” This question seeks to determine if the main reason for not looking for work is valid (see definition of unemployed) in which case the person is considered unemployed.

If the answer to this question is schooling, housekeeping, too young/old or retired/permanent disability or other reasons not considered valid, then the person is excluded from the labor force.

Philippine Concept on Unemployment. The Philippine Concept considered a person unemployed if he or she has no job/business during the reference week and is actively looking for work. Also considered as unemployed are persons without a job/business who are reported not looking for work because of the belief that no work available or because of temporary illness/disability, bad weather, pending job application or waiting for job interview.

International Labor Organization (ILO) Concept on Unemployment. The ILO Concept on employment states that a person is unemployed if he or she has no job or business during the reference week and is reported available and actively looking for work. Also considered as unemployed are persons who do not have job/business and are available for work but did not look for work because of the belief that no work is available, because of temporary illness/disability, bad weather, awaiting results of job application or awaiting for rehire/job recall.

We use the Philippine concept on unemployment in generating heat maps. But authors can easily switch to use ILO concept, if they wish, because the identification variable is already included in the survey by the national statistical office.

Occupation and Industry. The data on occupation and industry relate to the job held by employed persons during the past week. Occupation refers to the specific kind of work a person does while industry refers to the nature or character of the business or enterprise or the place wherein a person works. Persons employed with two or more jobs are reported in the job at which they worked the greatest number of hours during the past week. The new Philippine Standard Occupation Classification (1992 PSOC) and Philippine Standard Industry Classification (1994 PSIC) codes were used starting January 2001.

In the heat map, the category “professionals” includes armed forces, officials of government, managers, teachers, etc. “Manufacturing/service” workers includes

employees in manufacturing and service sectors. The category “agricultural” workers includes agricultural, forestry, fishery and related fields workers.

## **7. Housing ownership**

Usually housing is the main asset that most people own. Some household surveys report whether the house is owned by the family who lives in, and some also report the rental value of the dwelling.

## **8. Construction materials for the house**

Many household surveys report the materials used for the walls, roof and floor. But there is a huge difference among countries in the materials used for houses and in the coverage of these questions in the surveys. Materials that are a clear indicator of poverty in one locality may not be related to poverty in other country. Comparisons based on these variables should be made with care and preferably only within some locality, because these indicators are truly country-specific.

## **9. Water**

The access to water is relatively easy to identify in the survey. However, the quality of water is not easy to measure. We list the types of access to water generally by three categories: piped; trucks or buckets; hole or river. The original questions for access to water in the surveys widely differ across countries.

*Bolivia, Dominican Republic, Guatemala, Haiti, and Mexico*

Piped includes piped in house or outside, and piped public. Trucks or Buckets includes truck, bottled water and water by bucket. Hole or River includes hole, river or lake, rainfall and other.

*Pakistan*

Piped includes piped in house or outside tab. Trucks or Buckets includes tankers and vendors. Hole or River includes hand pump, motor pump, protected well, unprotected well, river and others.

## **10. Sanitation**

Similar to water, the access to sanitation is categorized for heat maps. The cross country comparison should be taken with care, because the quality of sanitation system varies across region/country.

*Bolivia, Dominican Republic, Guatemala, Haiti, and Mexico*

The estimates are presented in three categories: Drainage network or hole; latrine or other; don't have.

*Pakistan*

There are three categories: Flush, Latrine and Don't have. Flush includes flush to sewerage, flush to septic tanks and flush to drain. Latrine includes raised latrine, pit latrine and other.

## **11. Electricity**

The access to electricity is reported in many surveys. We use the information on access to electricity as the source of lighting. It must be noted that we measure only whether or not households have any access to electricity, not measuring to which extent they have reliable access to the power supply. The information on the reliable and continuous supply of electricity is rare in household survey data.

## **12. Telephone**

The access to a telephone is included in many surveys and the access to cell phone is included separately in many recent surveys as well.

## **13. Information and Communication Technology**

It is important to note that Information and Communication Technology (ICT) holds inherent public goods characteristics and therefore have an impact that goes beyond the individual user's welfare as consumption goods. ICTs are part of Goal 8 of the Millennium Development Goals which, amongst others, aims at making the benefits of new technologies — especially ICTs— more widely available.

*Botswana, Namibia, South Africa, and Zambia*

The access to ICT in the heat maps includes access to telephone, cell phone, public phone, computer, internet, email, and fax at home or elsewhere.

## **14. Financial Access**

*Bolivia, Dominican Republic, Guatemala, Haiti, and Mexico*

Access to credit includes any type of credit, from formal or informal institutions (i.e., shopkeepers, relatives, friends). This is important to consider since many studies, such as Navajas and Tejerina (2006), only include formal credit.

For credit source and use, since many households report more than one source/use of credit, figures reported are at the total credit-level.

The credit use is presented in three categories: (i) Agriculture, business and other business; (ii) Housing, education and durable goods; (iii) Non-durable consumption. The category labeled as “non-durable consumption” includes food, personal care and services, healthcare, transportation, recreation, clothing, funerals and gifts.

On figures showing the difference in access to credit by gender of the borrowing person, one concern is that the decision on whether, how much, and who to apply for a loan is made at the household level, not at the individual level (adult members collectively deciding which one of them to apply for loan). We could alternatively show differences in credit access and use for male or female headed households to distinguish opportunities for single female headed households. However, this would not necessarily reflect gender gap or differences in access/use of credit, but the effect of few collaterals (i.e., income generated by partner or spouse), which would naturally affect access to credit. Thus, with the limitations mentioned before, we still prefer to address gender differences by looking at the gender of the borrowing person, a methodology also followed by Bebczuk and Haimovich (2007).

#### *Botswana, Namibia, South Africa, and Zambia*

While the surveys contain incomplete data on household income, there appears to be sufficient data to explore product usage and access levels within the sample where data is available. Such an analysis would necessarily imply that access for the segment of the poor who have revealed their household income is a good proxy for access for the poor as a whole – which seems to be a reasonable assumption. Further, survey respondents are aged 16 or more. Penetration data for various products is therefore applicable to this segment of the market only.

Data on usage of funeral/burial insurance products is based on household heads as this is considered to be a household and not an individual product. Funeral insurance is provided by a range of sources, hence it was decided to divide them into 3 categories *viz.* formal (insurance co.’s, brokers, registered bodies, employers, etc.), semi-formal (funeral parlors), and informal (church contributions, burial and family societies).

Formal credit is defined as credit supplied from a formal source (banks, microfinance institutions, credit cards, store cards, etc.). Loans or credit from an employer is excluded. It should be noted that credit usage is often significantly understated in surveys.

The access strand divides the population into different categories of financial product usage *viz.* banks (have a bank account); other formal (do not have a bank account but

have other formal services e.g. microfinance products); informal (do not make use of banking or other formal products but do make use of informal services such as informal money lending and saving); and none (do not use any formal or informal services at all).

Formal products include bank accounts, home/personal/vehicle loans from registered formal organizations, subsidies and mortgages, store cards (purchase on account and pay later), retail/hire purchase, funeral policies from formally registered organizations or companies, asset insurance, life insurance, loss of earnings insurance, retirement annuities, pensions, investments such as saving policies, endowments, unit trusts, offshore investments and shares on the stock exchange, savings accounts with other registered formal organizations or companies such as the post office.

Informal products include belong to burial society, informal savings clubs such as Stokvels, safe guarding of money by someone else, loans from informal money lenders, clubs, or local shops.

### *Pakistan*

The access to credit is presented in the heat map in three different categories. Banks includes credits from banks. Lenders include money lenders, input suppliers, government, ROSCA (committee), usher and others. Shopkeepers include friends, shopkeeper and landlord.

For the use of credit, the category “agricultural business, other business” includes assets, investments and saving. “Housing, education, durable goods” includes housing and education and “non-durable consumption” includes food and beverages, clothing, fuel and lighting and other utilities, transportation/communications (including travel), health care and medicines, social functions and personal care, hygiene, upkeep, miscellaneous (including tobacco, daily use items)

## **15. Remittance**

Many of the surveys report whether the households receive remittance income from family member/friends inside/outside country. We use external remittances if the domestic and international remittances are distinguishable.

### *Bolivia*

Monthly remittance income from abroad is reported in the survey.

### *Dominican Republic*

We use the external income question to identify remittance recipients.

## Section External Income

Q31.D. Remittances from family member or friend.

*Guatemala*

There are questions regarding amount and origination of remittances in the survey.

Section E. Other Incomes and monetary subsidy: Remittances

Q7. Do you receive income from family members or family?

Q8. Where do they live?

1. Guatemala
2. United States
3. Mexico
4. Central America
5. Other countries

Q9. How much do you receive in last 12 months from the persons?

*Haiti*

The survey includes information on remittances broken down by the sender and also by monetary/ in kind transfer.

HE7. During the past 12 months, did any member of your household earn or receive income from any of the following activities... If yes, ask: What was the net total earning made from this activity during the past 12 months from each person involved in the activities?

5. Money transfers from relatives abroad
6. Transfers in kind from relatives abroad
7. Money transfers from relatives inside the country
8. Transfers in kind from relatives inside the country
9. Money transfers from other households/individuals abroad
10. Transfers in kind from other households/individuals abroad
11. Money transfers from other households/individuals inside the country
12. Transfers in kind from other households/individuals inside the country

*Mexico*

Monthly transfer income from abroad is reported in the survey.

*Botswana and Namibia*

Households with remittances are identified using following information. Notice that Botswana survey does not distinguish domestic and international remittances.

Q31.A. Different people get money in different ways. Please tell me what are your regular sources of income.

Q31.B. What are the regular sources of income earned by other members of your household?

Q32. How do you personally receive your money?

1. Cash
2. Cheque
3. Into Bank Account
4. Other

Item 13. Private transfers i.e. money from family/friends

### *South Africa*

The remittance recipients are identified by the following questions:

TR1-2. Receive money from a friend or family member living in South Africa

TR1-4. Receive money from a friend or family member living outside South Africa

In each question, households are asked of the means of transaction.

1. Bank
2. Post Office
3. Telegraph
4. With a friend/family member
5. Cheque
6. Using services such as Western Union, Thomas Cook, Rennies, MoneyGram.
7. With a paid/unpaid runner, e.g. taxi driver, bus driver
8. Store to store, that is, using services offered by shops such as Pick 'n Pay or Checkers
9. Electronic bank transfer (via ATM, Internet, telephone)
10. Mzansi account
11. Have not done this
12. Other means

Households with positive answer to the question TR1-4 with any means of transaction, except Item 11, are regarded as remittance recipients.

## *Zambia*

Households with access to remittance can be identified by several questions.

Q27. Different People receive money in different ways. Please tell me, what are your regular sources and receipts of money (income and other receipts).

Q28. What is your main source of income?

Q29. How do you personally receive your...?

- Cash in person
- Cash third party
- Swift transfer
- Western Union
- Cheque
- Into Bank Account
- Other

Q30. How often do you personally receive...?

- Daily
- Weekly
- Twice a month
- Monthly
- Less often

For each question, the survey asks various types of income source. Among them, the remittance related items are as follows:

11. Money from family in the country
12. Money from family outside the country
13. Money from friends in the country
14. Money from friends outside the country

The access to remittances in heat map is calculated from Question27, Item 12 and Item 14. It is possible to have a detailed heat map by mean of receiving remittances by using Q29, but there are not enough observations of remittance recipients for a reliable estimate.

## *Pakistan*

In Schedule H(1), the survey includes detailed information on the different sources of household income in the past year. It collects the information on the domestic and foreign remittances, separately.

## Reference

Acosta, Pablo, Illana Melzer, Ronald U. Mendoza, Namsuk Kim and Nina Thelen. 2008. "Exploring Challenges and Opportunities with Market Heat Maps." Working Paper. Office of Development Studies, United Nations Development Programme, New York.

Bebczuk, Ricardo, and Francisco Haimovich. 2007. "MDGs and Microcredit: An Empirical Evaluation for Latin American Countries." CEDLAS Working Paper No. 48. Universidad Nacional de La Plata, La Plata, Argentina [www.depeco.econo.unlp.edu.ar/cedlas/pdfs/doc\_cedlas48.pdf].

Brick, J. Michael, and Jonaki Bose. 2001. "Analysis of Potential Nonresponse Bias." Proceedings of the Annual Meeting of the American Statistical Association, 5-9 August. [www.fcsm.gov/committees/ihsng/brick\_bose\_jsm2001.pdf].

CEDLAS (Centro de Estudios Distributivos, Laborales y Sociales) and World Bank. 2006. *A Guide to the Socio-economic Database for Latin America and the Caribbean*. [www.depeco.econo.unlp.edu.ar/cedlas/sedlac/guide.htm].

Chen, Shaohua and Martin Ravallion. 2001. "How Did the World's Poorest Fare in the 1990s?" *Review of Income and Wealth* 47(3):283–300.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconomic Approach for Development Policy*. Washington D.C.: World Bank.

Deaton, Angus and Salman Zaidi. 2002. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis. *LSMS Working Paper 135*. World Bank, Washington D.C.

Fowler, Floyd, Jr. 2002. *Survey Research Methods, 3<sup>rd</sup> Edition*. Thousand Oaks, CA: Sage.

Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.

Kalton, Graham and Leslie Kish. 1981. "Two Efficient Random Imputation Procedures." American Statistical Association, ed., *Proceedings of the Survey Research Methods Section*. Alexandria, VA. [www.amstat.org/Sections/Srms/Proceedings/papers/1981\_031.pdf].

Korineka, Anton, Johan A. Mistiaena and Martin Ravallion. 2007. "An Econometric Method of Correcting for Unit Nonresponse Bias in Surveys." *Journal of Econometrics* 136(1): 213-35.

Lepkowski, James. 2005. "Non-observation Error in Household Surveys in Developing Countries." In United Nations Department of Economic and Social Affairs, eds.,

*Household Sample Surveys in Developing and Transition Countries*. New York: UN-DESA. [[http://unstats.un.org/unsd/hhsurveys/pdf/Household\\_surveys.pdf](http://unstats.un.org/unsd/hhsurveys/pdf/Household_surveys.pdf)].

Little, Roderick and Donald Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

McKnight, Patrick, Katherine McKnight, Souraya Sidani and Aurelio Figueredo. 2007. *Missing Data: A Gentle Introduction*. New York: Guilford Press.

Navajas, Sergio and Luis Tejerina 2006. "Microfinance in Latin America and the Caribbean: How Large is the Market?" Sustainable Development Department Best Practices Series: MSM-135. Washington D.C.: Inter-American Development Bank.

Ravallion, Martin, Gaurav Datt and Dominique van de Walle. 1991. "Quantifying Absolute Poverty in the Developing World." *Review of Income and Wealth* 37:345-61.

Särndal, Carl-Erik and Sixten Lundström. 2005. *Estimation in Surveys with Nonresponse*. New York: Wiley.

United Nations Development Programme (UNDP). 2008. Growing Inclusive Market. Mimeo, United Nations Development Programme, New York, NY.

World Bank. 2007. *World Development Indicators*. Washington D.C.

# Annex 1: Analysis of Potential Nonresponse Bias in FinScope Data -Using Zambia 2005 Survey

Namsuk Kim, UNDP

The potential bias from the high rate of item nonresponse for the individual/household income information in FinScope data sets should be evaluated in order to achieve reliable estimates. This paper suggests a few methods to assess the problem and summarizes preliminary findings of the effect of the nonresponse adjustments using Zambia 2005 Survey data. The result shows that there might be nonresponse bias, but the sign of bias is constant for the indicators of living conditions, that is, the estimates of access to basic needs from the response sample is always higher than those from the missing-observation-adjusted sample.

## 1. Introduction

FinScope, a FinMark Trust initiative, is a nationally representative study of consumers' perceptions on financial services and issues, which creates insight to how consumers source their income and manage their financial lives. The sample covers the entire adult population, rich and poor, urban and rural, in order to create a segmentation, or continuum, of the entire market and to lend perspective to the various market segments.<sup>3</sup> The statistical information from the Finscope data is highly demanded, and the survey is actively used to study not only the financial market but also general issues of economic growth and private sector.

Although the information on financial access is collected in the carefully designed survey, high nonresponse rates for a couple of survey variables pose questions on the reliability of the statistics that are calculated from the survey.<sup>4</sup> Especially, missing observations in income variables should be carefully dealt with.<sup>5</sup> The potential bias from the item non-responses can be evaluated using various statistical models.

Many studies have piled in the measurement and treatment of nonresponse bias. There is no doubt that minimizing nonresponse in the first place is a surer way to minimize nonresponse error in survey-based estimates (Fowler 2002, p.52). Nice summary would be Kalton and Kish (1981), Little and Rubin (1987) and Groves and Cooper (1998), regarding statistical adjustment to reduce the error in the estimates caused by nonresponse. Following these leads, this paper suggests a few exercises to assess and reduce the potential nonresponse bias in statistics calculated from the survey.

---

<sup>3</sup> <http://www.finscope.co.za/index.asp>

<sup>4</sup> For the impact of missing data, see McKnight et al (2007, p 6)

<sup>5</sup> Correction of nonresponse in income variable is not easy to solve. See Korineka et al (2007).

Among the countries in which the FinScope survey was conducted, Zambia survey is one of the most troublesome because its nonresponse rate is among the highest.<sup>6</sup> This note shows that the indicators of access to basic need are very conservative when the estimates are derived only from the response sample in Zambia. This type of exercise can be applied to other countries.

## 2. Measuring item nonresponse bias

While the level of nonresponse does not necessarily translate to bias, large differences in the response rates of subgroups may indicate that potential biases may exist. The bias is commonly expressed as:

$$bias = (1 - r)(\overline{y_r} - \overline{y_{nr}})$$

Where the subscripts *r* and *nr* denote respondents and nonrespondents, and (1-r) is the nonresponse rate. Therefore, if the response rates for high- and low-income individuals were very different, any difference between the means of the respondents and nonrespondents would result in a large bias.<sup>7</sup>

Measurement of nonresponse bias requires measurement of differences between respondents and nonrespondents on survey variables. The first analysis of nonresponse bias would be a simple comparison of response rates by selected variables in Table 1. Overall, the analysis shows that response rates are highly correlated with variables related to living condition (electricity, water, telephone). Consequently, the effect of adjustments should reduce nonresponse bias in estimates from those variables.

Table 1. Response rate by selected variables.

Variable	Percentage of individuals who responded to income question
Total	57.2
Region	
Rural	56.6
Urban	58.7
Access to Electricity	
Always, Often or Sometimes	61.2
Never or Rarely	55.6
Access to Water	
Always, Often or Sometimes	58.7
Never or Rarely	50.7
Access to Telephone at Home	
Yes	61.7

<sup>6</sup> The nonresponse rates are 43% for Zambia, 23% for Namibia, 15% for Botswana and 10% for South Africa, for example.

<sup>7</sup> Brick and Bose (2001) and Lepkowski (2005)

No	57.1
----	------

The second approach is to compare estimates from the respondent individuals (responding to income question) to the estimates computed from the entire sampling frame. Variables available both for the response sample and the entire sample can be used in such comparison. The difference between the response sample estimate and population value from the total sample is calculated and tested by t-statistics in Table 2.

Table 2 Difference of respondent sample and total sample mean

% of people who have access to...	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total sample mean	Response sample mean	Diff	t	H <sub>1</sub> : Diff<0 Pr(T<t)	H <sub>1</sub> : Diff=0 Pr( T >t)	H <sub>1</sub> : Diff>0 Pr(T>t)
Secondary education	.23	.26	-.03***	-2.65	.01	.01	.99
Urban residence	.29	.28	.00	.22	.58	.82	.41
Telephone	.40	.43	-.00	-.60	.27	.55	.72
Cell phone	.26	.29	-.03***	-3.19	.00	.00	.99
Bank	.14	.18	-.04***	-4.46	.00	.00	1
Cash income	.62	.67	-.05***	-3.86	.00	.00	.99
Electricity	.29	.31	-.02**	-1.64	.05	.09	.95
Medicine	.80	.84	-.04***	-4.05	.00	.00	1
Water	.81	.83	-.02**	-2.12	.02	.03	.98
Food	.80	.82	-.01*	-1.30	.09	.19	.90
Correct kinds of food	.44	.47	-.03***	-2.13	.01	.03	.98

Note: \*, \*\* and \*\*\* denote the estimate is statistically smaller than zero at 90%, 95%, and 99% confidence level, respectively.

Column (1) represents the total sample mean and Column (2) shows the response sample mean. If the nonresponse is randomly distributed throughout the sample, there would be no reason that those two sample means should statistically differ from each other. In that case, Column (3), which is Column (1) subtracted by Column (2), should be statistically equal to zero. Column (4) provides t-statistics to test if Column (3) is statistically equal, smaller or bigger than zero. If the value in Column (5) is very small, it means the probability of the difference being negative is very high. If the value in Column (6) is very small, it means the difference is very likely different from zero. Likewise, if the value in Column (7) is very small, it will indicate that the difference is very likely to be positive.

Table 2 shows that, in 11 comparisons, 10 of the differences in Column (3) are negative. Among these 10 negative differences, 8 of them are statistically smaller than zero, while none of them are statistically greater than zero at 90% confidence level. This result shows that the indicators for access to basic need are almost always higher when they are calculated from the response sample than when calculated from the total sample. There might be a systematic bias due to nonresponse, but the indicators are generally higher when they are calculated only from individuals responding to income question. If

nonresponse adjustment measures had been used, it is likely that the percentage of people with access to basic goods and services (food, water, electricity, education, banking service, etc.) would have been lower.

### 3. Reducing item nonresponse bias

The analysis involving imputation of data would be another way to examine the nonresponse bias and measure the effect of adjustment. Zambia 2005 Survey data is unique in the sense that there are a large number of individuals who answered to other questions but the income question. For those who provided enough information except per capita income, we can impute their income level and find which income group they might belong to.

Imputation is the procedure whereby missing values on one or more variables are filled in with substitutes. Imputed values can be classified into three major categories<sup>8</sup>:

1. values constructed with the aid of a statistical prediction rule
2. values observed not for the nonresponding elements themselves, but for similar responding elements
3. values constructed by expert opinion or best possible judgment

Here a commonly used regression imputation method, one of imputation by statistical rules, which is called "hot deck."<sup>9</sup> The missing information of nonresponse, per capital income, can be imputed from other information from the nonresponse through an estimation process as follows:

Step 1. Regress per capita income on other variables for respondents:

$$Y_r = X_r \beta_r + \Lambda$$

Step 2. Use the coefficient estimates to forecast income of nonrespondents:

$$\hat{Y}_{nr} = X_{nr} \hat{\beta}_r$$

Step3. Compare Response Sample (response-only sample) and Imputed Sample (response-plus-imputed sample).

Since the dependent variable, per capita income, is collected by range, we could use low/mid/high point of the range. A bivariate Logit or Probit model using income groups (poor vs non-poor, for example) as dependent variable is not recommended because the underlying income variable is semi-continuously observable.

A preliminary exercise in this approach suggests that the adjustment would change sample means in a single direction, that is, lower access to basic goods and services. The mid-point of per capita income is used as the dependent variable, and is regressed

---

<sup>8</sup> Sarndal and Lundstrom (2005, p.153)

<sup>9</sup> Lepkowski (2005, p.167) and Sarndal and Lundstrom (2005, p.160)

separately for each of 9 provinces on a number of explanatory variables to control for detailed household and individual characteristics.

Table 3 Difference of respondent sample and imputed sample mean for people living on less than \$2 per day, using the whole sample in the imputation

% of people who have access to...	(1) Imputed sample mean	(2) Response sample mean	(3) Diff	(4) t	(5) H <sub>1</sub> : Diff<0 Pr(T<t)	(6) H <sub>1</sub> : Diff~=0 Pr( T >t)	(7) H <sub>1</sub> : Diff>0 Pr(T>t)
Secondary education	.11	.14	-.03***	-3.19	.00	.00	.99
Urban residence	.14	.18	-.04***	-3.82	.00	.00	.99
Telephone	.01	.12	-.01*	-1.37	.08	.17	.91
Cell phone	.14	.18	-.04***	-4.36	.00	.00	1.00
Bank	.07	.08	-.01***	-2.20	.01	.02	.98
Cash income	.50	.59	-.09***	-5.51	.00	.00	1.00
Electricity	.13	.19	-.06***	-5.60	.00	.00	1.00
Medicine	.75	.82	-.07***	-5.26	.00	.00	1.00
Water	.79	.82	-.03***	-2.28	.01	.02	.98
Food	.74	.77	-.03***	-2.26	.01	.02	.98
Correct kinds of food	.35	.40	-.05***	-3.11	.00	.00	.99

Note: \*, \*\* and \*\*\* denote the estimate is statistically smaller than zero at 90%, 95%, and 99% confidence level, respectively.

Table 3 displays the difference of respondent sample and imputed sample mean for people living on less than \$2 per day. Column (3) represents the difference between Imputed sample mean and Response sample mean. They are all negative, suggesting the indicators for access to basic need are higher for response group. And Column (5) confirms that the differences are all statistically smaller than zero. Therefore, indicators for access to basic need for the poor calculated only from the response sample overstate the wellbeing of the total population of the poor.<sup>10</sup>

#### 4. Summary

The evaluation and adjustment for nonresponse bias presented in this memo can be summarized into two major findings. First, there is a systematic bias of nonresponse which is highly correlated with living condition of the households. Second, the bias always has the same sign and it suggests that indicators calculated from response sample overestimate the level of access to basic goods and services. Therefore, any living-

<sup>10</sup> We have not applied any a priori knowledge about the sample, getting the most conservative result. If the nonresponses are believed to share many things with the poor, we may apply this information to reduce the nonresponse bias further (following Sarndal and Lundstrom, 2005, p.162). For instance, instead of using the whole sample, the missing income variable is regressed on the observed income information of a subgroup of the sample, that is, the group with under \$2 per day. But we did not find the regression using sub group performs better than using the whole sample in terms of R square, mainly because of the sample size reduction.

condition-related indicator calculated from response sample for people living on less than \$2 per day could be, at most, very optimistic and arguably regarded as the upper-bound estimate.