



2 Census, 10 years of household surveys in Bolivia. Time to re-weight and define a comparable sample design for Welfare indicators

Alvaro Chirino Gutierrez
ARU Foundation, La Paz-Bolivia, achirino@aru.org.bo

Abstract

This paper proposes a sample design for the period 2002-2012 of the series of household surveys conducted by the National Institute of Statistics of Bolivia; the design seeks to achieve optimum sample performance for the indicators of moderate poverty and extreme poverty to level departmental (first national sub-unit) and national. There are two main reasons for this new design; the first, the existing sample designs in surveys of the series differ in several features such as: stratification, clustering, sampling frame, the allocation of the sample, the sample size, et.al. These variations do not allow a consistent reading of the evolution of welfare indicators over the years. The second motivation is the relevance, in 2012 the last Census of Population and Housing was conducted, the results of this census show a different structure to projections. And following the recommendations of the United Nations 2010, on the opportunities of a new census, this is an ideal to suggest a new design that will bring the reweighting of the sample.

Keywords: Re-wight; Bolivia; Household surveys; Welfare indicators.

1. Introduction

The National Institute of Statistics of Bolivia (NIS), like most countries in the region, development in recent years, household surveys (HS) designed to measure welfare, this practice was the initial incentive program “MECOVI” from the World Bank. The 2001 Bolivia conducted the National Census of Population and Housing (NCPH) and as one of the natural results of the census, the information obtained by this became the basis for the development of surveys, in November of 2012 was carried the NCPH again. In the intercensal period a total of 10 rounds of surveys were conducted from 2002 to 2012, some peculiarities on this period are; 2003 and 2004 a survey of continuous type intended to build a new family budget for the consumer price index, in 2002 to 2005 the sampling frame was a master sample, in 2010 no survey was conducted, since 2011, the survey almost doubled the size of sample households.

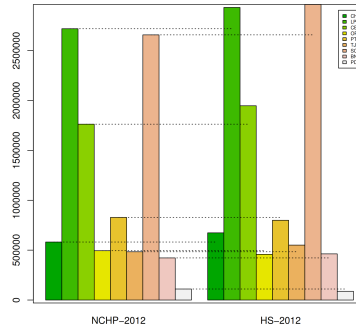
1 presents an overview of the sample design of the intercensal period verified that there is no relationship in the series on the design from one year to another, indeed, almost any year is similar in design to another, this motivates think that the estimates come from these surveys are not fully comparable. Also, the results of NCPH-2012 show that the structure of the country is not the same which expand the HS, eg HS-2012 expands to a population of about 10,874,551, however, NCPH-2012 recorded a total of 10,059,856 more than 800 thousand of difference.

Table 1: design, Survey of household in Bolivia 2002 to 2012

Year		2002	2003-2004	2005	2006	2007	2008	2009	2011	2012
Frame		Master Sample, NCPH-2001	Master Sample, NCPH-2001	Master Sample, NCPH-2001	NCPH-2001	NCPH-2001	NCPH-2001	NCPH-2001	NCPH-2001	NCPH-2001
Stage		2	2	2	2	2	2	2	2	2
Strata		Urban Rural	3 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3
Cluster		Group of census sector	Group of census sector	Group of census sector	Group of census sector	Group of census sector	Group of census sector	Group of census sector	Group of census sector	Group of census sector
Stratification, first stage	Stage 1	Population, 5 levels	Population, 5 levels	Population, 5 levels	Area and Basic needs, 8 levels	Area and Basic needs, 8 levels	Population and consumption	Area and Basic needs, 8 levels	Population and Basic needs, 20 levels	Population and Basic needs, 20 levels
Selection mechanism	Stage 2	rps, housing	rps, housing	rps, housing	rps, housing	Sistematic rps, housing	Sistematic rps, housing	Sistematic rps, housing	Sistematic rps, housing	Sistematic rps, housing
	Last stage	rps, housing	rps, housing	rps, housing	rps, housing	Unknown	Unknown	Unknown	Sistematic	Sistematic
	People	679	547	166	311	369	304	301	790	729
	Household	5746	5767	2855	4006	4699	3856	3932	8551	8360
	People	24933	38500	16511	16511	16804	15030	15665	33821	31935

A result, more important than national are referred to the department, these are the first national sub units of Bolivia and are the first local actors of public policy, therefore, adequate information to the department level is necessary, the structure reflected by the HS-2012 at the departmental level differs from that obtained

Figure 1: Estimated population structure of the HS-2012 vs the results of NCPH-2012



in NCPH-2012, this difference is show in 1, where it is clear the department with sub and over coverage.

This paper presents a expost sample design, with the aim of generally comparable and consistent estimates to the results of NCPH-2012, the final sample design is the result of a contrast in performance sampling of 5 alternative design proposals and an adjustment in the expansion factors about the growth rate of the period between 2001-2012, the analysis of sample performance is based on indicators of moderate and extreme poverty to departments levels

The document presented in Section 2, the methodology, data sources used are described, defining poverty in HS, treatment for sampling components such as stratification and clustering, construction of the theoretical probabilities, the factor expansion and projection used for the intercensal period, in section 3 the results are presented, the performance of sampling designs, the evolution of poverty and the final sample design. Finally, in Section 4 conclusions are given.

1.1. Objectives

The objective of this paper is:

“Developing a comparable sample design for welfare indicators from the series of household surveys 2002-2012, adjusted for population growth patterns of NCPH 2001 and 2012”

2. Methology. In order to obtain a comparable sample design for the series of surveys of 2002-2012, are produced five different sample designs and their performances are compared for the indicators of moderate and extreme poverty, for this, the data is defined, the definition of poverty of the document is established, clustering and stratification characteristics are defined, the projection for the intercensal period is done, the theoretical probabilities are constructed and finally the sample weights are defined adjusted to projections.

2.1 Data. Databases and methodological documentation household surveys were obtained from social data bank of portal NIS , databases of NCPH-2001 that is used as the basis of the sampling frame corresponds to which distributes the communication department of NIS. Finally, information NCPH-2012 was obtained from the website Redatam of the Census.

2.2 Poverty. All household surveys of the period of interest define poverty based on income and its location above or below the poverty line, following the methodology Foster, James, J. Greer, and Eric Thorbecke. 1984. Where moderate poverty for an individual is defined as:

$$p_i = I(y_i < z_k)$$

Where $I(\cdot)$ is an indicatrix function with respect to income (y_i) of the individual i and moderate poverty line (z_k) in the region k of the country. Similarly, extreme poverty is defined as:

$$pe_i = I(y_i < ze_k)$$

Where $I(\cdot)$ is an indicatrix function with respect to income (y_i) of the individual i and extreme poverty line (ze_k) in the region k of the country.

The study used as an aggregate, the incidence of extreme and moderate poverty, this indicator is defined within an area k , as:

$$P_0 = \frac{\sum_k p_i}{N_k}$$

$$P_{0e} = \frac{\sum_k pe_i}{N_k}$$

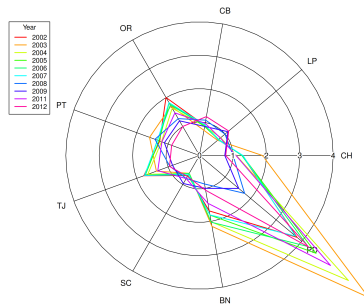
Where P_0 is the incidence of moderate poverty and P_{0e} the incidence of extreme poverty.

The study employed since 2006 the poverty variables reported in databases, for the lower years, the information comes from Hernani and Villarroel (2012), this a reason for the absence of the variables of poverty in the databases before 2006.

2.3. Cluster. During the period of the survey series there were two primary sampling units (PSU), the first was from 2002 to 2009 and was a grouping of census sectors that on average had 120 households in the HS 2011 and 2012 PSU were the census sectors that on average were 90 households, all these units based on CNPV-2001, as showed in Table 1. Some survey years, there were two stages in rural areas, however, in the methodological documents the criterion of definition and selection of these areas is not explained, therefore, for purposes of this paper defined the existence of two stages; cluster selection and selection of household. The PSU will be the grouping of census sectors, given their presence in most of the surveys and in the HS 2011 and 2012 they can be identified without difficulty.

2.4. Strata. Stratification is the main form of which depends on the structure of a complex sample design, along survey series have been employed various strata, all pre-2011 methodology documents define strata nationally and exclude treatment at the departmental level, however, to analyze databases, it is evident that there was a selection criterion based on departments, given the pattern of over and under representation of some departments respect to a proportional sampling frame structure. 2 shows the pattern, where the rings represent the ratio of the assigned sample and necessary, the case of Pando, which almost always received more than 3 times the necessary sample clearly shows, this shows that there was a selection based on departments.

Figure 2: Over and under sampling of HS 2002-2012



Five types of strata to PSUs level are defined, each combined with the 9 departments, the strata are:

- Population: 5 levels within departments; capital cities, intermediate cities, major population centers, minor population centers and dispersed area, the stratum corresponds to a definition of NIS
- Geographic: 3 levels; Altiplano, Valle and Llano, the stratum corresponds to a definition of NIS
- Consumption: 12 levels within departments, was built based on a regression model based on the HS-1999-2002 and was used for the HS-2008
- Principal Component: 5 levels within departments, was built based on the technique of principal components using information NCPH-2001
- Basic Needs: 4 levels within departments, was built based on the technique of unsatisfied basic needs index, the stratum corresponds to a definition of INE

2.5. Projections. For population projections in the intercensal period, exponential growth rates was used, this is defined as:

$$r = \frac{\ln(P_{2012}) - \ln(P_{2001})}{Y_{2012-2001}}$$

Where P_t corresponds to the population census the year t and $Y_{2012-2001}$ the elapsed years, from 2001 to 2012 period, the rate corresponds to November 21, 2012 (day of NCPH-2012). The formula used for the projection is:

$$Pr_i = P_{2001} * e^{r*a}$$

Where corresponds to the projection for the year and elapsed time in years from the base year 2001 to year projection. 2 shows the annual growth rates by department and area.

Table 2: Annual growth rate and population of NCPH 2001 and 2012 for department and area

Department	NCPH-2001			NCPH-2012			Annual growth rate (%)		
	Urbano	Rural	Total	Urbano	Rural	Total	Urbano	Rural	Department
Chuquisaca	218126	313396	531522	283123	298224	581347	2.325	-0.442	0.799
La Paz	1552146	797739	2349885	1814148	905196	2719344	1.390	1.126	1.302
Cochabamba	856409	599302	1455711	1200912	561849	1762761	3.013	-0.575	1.706
Oruro	236110	156341	392451	316757	177830	494587	2.619	1.148	2.062
Potosí	239083	469930	709013	336412	491681	828093	3.044	0.403	1.384
Tarija	247736	143490	391226	314510	169008	483518	2.127	1.459	1.888
Santa Cruz	1545648	483823	2029471	2160579	497183	2657762	2.985	0.243	2.404
Beni	249152	113369	362521	308690	113318	422008	1.910	-0.004	1.354
Pando	20820	31705	52525	53831	56605	110436	8.467	5.166	6.624
Total	5165230	3109095	8274325	6788962	3270894	10059856	2.436	0.452	1.742

Source: Author

2.6. Theoretical probability. Based on the clusters and the strata, the theoretical probabilities associated with the two stages of the design are defined. For the first stage, it is assumed probability proportional to size without replacement, using the total housing of the PSU recorded in the sampling frame of NCPH-2001, this takes the form of:

$$\pi_{Ii} = n_{Is} * \frac{t_{is}}{\sum_S t_{is}}$$

Where π_{Ii} is the probability of the first stage for the PSU i , n_{Is} the simple size of PSU at level s of the strata, and t_{is} the total of dwellings in the PSU i at the level s of the stratification.

The probability of the household k , is defined as the product of the probability of first stage and the last stage.

$$\pi_{ki} = \pi_{Ii} * \frac{n_{Iii}}{t_{is}}$$

Where π_{ki} is the probability for the individual k of the PSU i , n_{Iii} the simple size of the last stage.

For analysis were eliminated PSU with $n_{IIi} == 1$, to avoid an increase in the variances.

2.7. Weights. The construction of the expansion factors for the 5 different types of strata began with the theoretical factor, this is the inverse of the probability of selection:

$$W_{ki} = \left(\pi_{ki} * \frac{n_{IIi}}{t_{is}} \right)^{-1}$$

About this factor, an adjustment based on projections P_l was included, this adjustment did at the departmental level and by area.

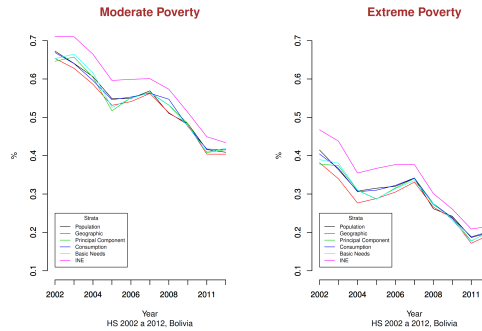
$$aW_{ki} = W_{ki} * \frac{P_l}{\sum_l W_{ki}}$$

High factors were adjusted based on the value of the 99th percentile of their distribution, that the recommendation of the Naciones Unidas. (2009), in order to reduce the variance of the estimates of poverty.

3. Results This section presents the results of the 5 types of design, showing the evolution of poverty, the performance of sampling and choosing the final design.

3.1. Poverty evolution. 3 shows the evolution of poverty for the series of interest and considering the 5 types of defined strata, the official estimate of INE is included. It is observed that all types of strata estimate a lower incidence official but keep the trend.

Figure 3: Poverty Evolution by Strata



3.2. Sampling performance. For sampling performance is considered the sample coefficient of variation and the design effect for the two types of poverty at national and regional level, 4 shows the evolution of poverty by design type includes the coefficient of variation and confidence intervals at 95%.

It is noted that one of the strata with lower performance is the geographic and that consumption and components are the best performers. 3 summarizes the performance at departmental and national levels for moderate and extreme poverty. It is observed that consumption is the strata that achieves the best performance, followed by the main components, then the basic needs of the population and lower performance is geographical. The design effect is greater than 7 in all cases, this is because the high homogenized in the PSU.

3.3. Final design. It is recommended to adopt the sampling design based on consumption strata, since achieves better performance.

4. Conclusion The conclusions are:

Figure 4: Sampling performance by strata

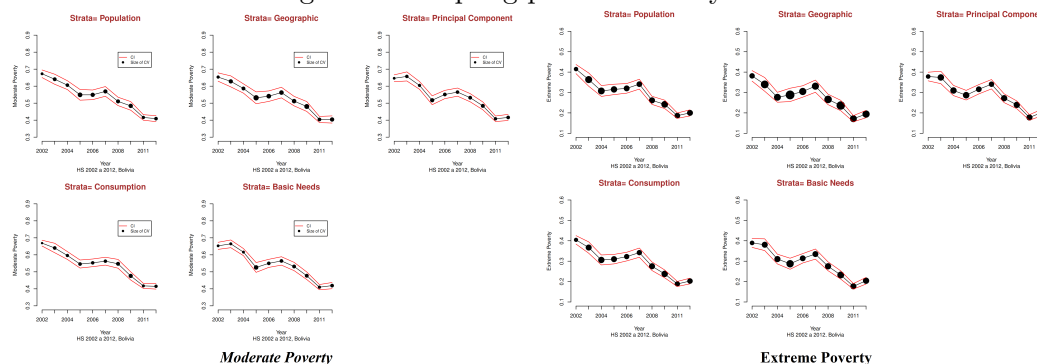


Table 3: Sampling performance by strata

Moderate poverty				
Strata	National		Department	
	CV	DEFF	CV	DEFF
Population	2.364	13.180	8.567	10.295
Geographic	2.679	16.077	9.427	12.735
Principal Component	2.155	10.749	8.007	8.521
Basic Needs	2.120	10.460	7.880	8.667
Consumption	2.070	10.028	7.719	7.918
Extreme poverty				
Strata	National		Department	
	CV	DEFF	CV	DEFF
Population	3.883	12.512	15.032	10.505
Geographic	4.767	17.117	18.081	13.918
Principal Component	3.837	11.824	15.268	9.675
Basic Needs	3.937	12.353	15.356	9.995
Consumption	3.595	10.593	14.336	8.626

- The sample design that improves the strata performance indicators of the sample is based on consumption strata design
- It is necessary to reduce the sample inside the PSU and include more PSU in the sample in order to improve the design effect
- The patterns of moderate and extreme poverty are over estimated with the official design that does not fit the growth rate that existed in Bolivia in the intercensal period.
- For the next intercensal period is necessary to preliminary planning of sample design for the entire series, in order to establish an integrated national survey system.

References

- Cochran, W. G. (1953). Tcnicas de Muestreo. Mexico: John Wiley and Sons
- Foster, James, J. Greer, and Eric Thorbecke. 1984. A Class of Decomposable Poverty Measures. *Econometrica* 52 (3): 76165
- Hernani W., Villarroel P. (2012) .Evolution of poverty in Bolivia: a multidementional approach INE, Bolivia. (2002-2012). Documento Metodolgico, Encuesta a Hogares 2002 - 2012. La Paz, Bolivia.
- Kish, L. (1955). Survey Sampling. New York: John Wiley and Sons.
- Lohr, S. L. (2000). Muestreo: Diseo y Anlisis.
- Naciones Unidas. (1986). Sampling frames and sample design for integrated household survey programmers.
- Naciones Unidas. (2009). Diseo de muestras para encuestas de hogares Directrices prticas (p. 252). Nueva York.
- Rosn, B. (1997). On sampling with probability proportional to size. *J. Statistics , Plann Inference.*
- Srndal, Swensson, y Wretman. (1992). Model Assisted Survey Sampling. Canada: Springer, Verlang.