Estimation of demographic statistics in the Cambodian Socio-Economic Survey (CSES) 2004-2011

Pen Socheat,

National Institute of statistic, Cambodia, Email: pen.socheat@hotmail.com

Since 2007 the CSES is the only source that every year has provided estimates of Cambodian demographic statistics such as population and household sizes by geographical region, sex, marital status and ethnicity. This "work" describes the methodology used to make the estimates consistent with Census projections taking into account change in household sizes. The methods are based on assumptions and they will adjust the weights of the survey. The assumptions need to be evaluated continuously and this presentation is the first documentation and evaluation. Using time series of some key statistics, the work also presents and discusses the demographic development in Cambodia. A future decision if the methodology in the CSES has to be completely changed or just updated must be made when the Inter-censual result from 2013 are available.

Keyword: Consistency, Census Projections, Weighting method.

1. Introduction

Since Cambodia's 2008 Census, the official demographic statistics are based on yearly projections of the Census results. The statistics and the underlying methodology of the projections are described in NIS (2011) and Arriaga (1994). The National Institute of Statistics in Cambodia has, with the assistance of Statistics Sweden and the financial support of the Swedish International Development cooperation Agency, conducted the Cambodian Socio-Economic Survey (CSES) every year since 2007. An important part of the CSES is its estimates of demography and migration. They are used as background information and for statistical comparisons of all the other subject matter content in the CSES. The CSES contain modules of country wide sample of households and household members about housing conditions, education, economic activities, household production and income, household level and structure of consumption, health, victimization, etc. There are also questions related to people in the labour force, e.g. labour force participation.

CSES was conducted intermittently in the period 1993 to 2004 but since 2007 the survey is annual. The 2004 and 2009 were large sample surveys (12,000 households), whereas the years between have small samples (about 3,600 households).

The data from the CSES provide important information about living conditions in Cambodia and have a wide range of use. Results from CSES are used for monitoring the National Strategic Development Plan (NSDP) and progress towards the Millennium Development Goals.

It is important that the demographic estimates of the CSES and other Cambodian surveys to a certain degree are consistent with the official Census statistics. Not primarily because the projections from the Census are particularly accurate in a given domain of study, but rather to keep Cambodian official statistics harmonized. This paper describes the method used up to now, and discusses its properties and future developments. The outline is the following. After initially defining necessary notation and parameters, I will describe the methods that are used to estimate the population

growth and the changes in household compositions. Thereafter I will present demographic estimates and comment.

2. Background of the Cambodian Socio-Economic Survey

The CSES is a multistage sample survey of Cambodian normal[1] households (and persons). Presently, the sampling design consists of three stages. The primary sampling units come from nationwide frame of villages. These are divided into a rural and an urban stratum and ordered by geographical location. From each stratum a systematic sample with probabilities proportional to size (number of households per village) is selected. The secondary sampling units are enumeration areas in the villages selected by simple random sampling and in the final stage ten households are selected by systematic sampling. Data are collected every month in face to face interviews. The two main parameter of interest considered in this paper are,

Y_t: The total popuation (POP) of Cambodia in year t and
M_t: The number of Households (HHs) in Cambodia in year t

Estimates of Y_t is given by, Y'_t: The Census estimate of the Cambodia population adjusted for undercount for normal households (HHs) and $\hat{Y}_t$: The estimated population of Cambodia from CSES adjusted with Calibrated weights. Furthermore, define $\bar{X}_t$: as the estimated average HH Size of year t, where, $\bar{X}_t = Y'_t / M_t$.

When we estimate demographic characteristics of Cambodia, we need to take into account two different trends over time. First the growth number of persons and second the change in HHs composition. The household sizes are decreasing. I will begin by describing method used for the population growth. The method has been used since CSES 2010 and it uses the census projection and the census estimate from 2008, Y'_{2008}, as a base.

3. Methods and assumptions to estimate demographic change in the CSES Population Growth

Let $D_{y',t}$ be the population change since 2008 according to Census projections. Then, the yearly estimate of change from March 2008 to March 2009 for the whole Cambodia is given by $D_{y' 2009} = \frac{Y'_{2009}}{Y'_{2008}} = 1.021$ according to Population Census of Cambodia 2008. Since the CSES is stratified by Urban/Rural, similar estimates are computed for Urban/Rural separately. This will give us a series of estimates of change from 2008- 2012.

2009	2010	2011	2012
2.1%	3.7%	5.3%	6.9%

Table 1: Estimate of demographic change since 2008 in CSES

From the census projections some adjustments must be made for the number of persons living in normal HHs or not and adjustment also needs to be made for administrative changes happening in the provinces since the census.

4. Change in HHs Size

The other trend that is necessary to estimate, is the change in HHs composition. Define $D_{\bar{X},t}$: as the change in average household size since 2008 according to census

---

[1] Normal households are all household except homeless, Institutional household, boat population, transient population

projection. For the change in household size the CSES estimate have been based on a crude model using the observed change between 1998 census and 2008 census. In 1998, the average household size was 5.18 persons per household and in 2008 the household size had decreased to 4.66, according to census data. This means an annual decrease in household size by approximately 1% . This is used for a projection model of the annual change in average household size based on the assumption of the same annual change every year in the period 1998 to 2008, that is

$$\delta = \left(\frac{\overline{X}_{2008}}{\overline{X}_{1998}}\right)^{0.1} = \left(\frac{4.66}{5.14}\right)^{0.1} = 0.9902$$

Hence, a decrease by approximately 1 percent every year is assumed, however household size cannot continue to decrease linearly, so the yearly projection formula applied to the change in the average household size is $D_{\overline{X},t} = \left(1 + \left(0.9 * (\delta - 1)\right)\right)^{k}$ where $k = 0$ for 2008, $k = 1$ for 2009, and so on.

Therefore, $D_{\overline{X},2009} = 0.99118$ gives the projection $\overline{X}_{2009} = 4.73$ for normal household size in 2009. The projected number of normal households are then given by the following computation. E.g for 2009 we get

$$M_{2009} = \left(D_{y',2009} * M_{2008}\right)/D_{\overline{X},2009} = 2,902,389$$

Taking into account the administrative changes and the two trends, (i) increasing population and (ii) decreasing household sizes we can compute a series of estimated total number of households per year using the above formula.

5.   *Estimation in the CSES and Calibration estimate of the population size.*

The CSES estimate of the population size is calibrated with weights to satisfy the equation $\hat{Y}_t = Y_t'$. Theory about calibration can be found in Särndal and Lundström (2005). The information used in the estimation is the projections Y_t' from the Census. Since the CSES has practically no nonresponse the estimator is of the form $\hat{Y}_t = \sum_s w_k y_k$ with y_k, being a general notation of a study variable which when counting the population here is equal to one for each unit. However it can also be a domain indicator counting the population in a e.g. a province. The weight $w_k = d_k v_k$ consists of two parts, d_k is the design weight from the sampling design, and v_k is the calibrating factor determined as to satisfy $\sum_s w_k = Y_t'$. In our case it gives us a simple expansion estimator with $v_k = \frac{Y_t'}{\sum_s d_k}$.

The calibration technique guarantees that the CSES estimates of the population size will be equal to the forecasts from the 2008 Census and this assures consistent results in the Cambodian population statistics. The weights that are computed will also influence other CSES estimates. A possible problem is that the accuracy of the census population forecasts are decreasing as time passes. Comparative studies of Cambodian population forecasts and results from later censuses and an intercensal survey show quite disturbing inconsistencies (See NIS 2004 and NIS 2008). It is important that the CSES is calibrated against information that is as reliable as possible, otherwise the calibrated weights may introduce an unnecessary source of error to the survey results. Continuing to use the above method based on the projections from the 2008 census is not a good alternative. The Cambodian 2013 intercensal survey will provide new information that can be used for the 2013 CSES results and onwards until the next planned Census in 2018.

6. Demographic estimates

In this section we illustrate the demographics estimated in recent CSES compared with General Population Census 2008.
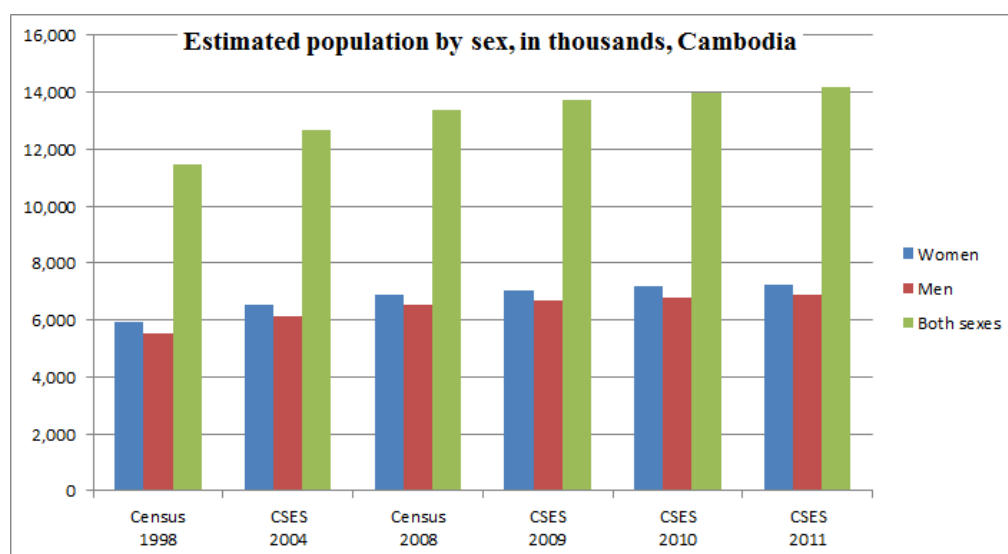

Figure 1: Population development in Cambodia 1998 to 2011

The figure shows the population increase in Cambodia. The numbers for Census 1998, 2008 for estimated population by sex in thousand presented here differ a little bit from those presented in the CSES 2004, 2009, 2010 and 2011 reports. The numbers presented in the CSES reports do not take into account changes in the definition of urban area between 1998 and 2008 and for 2004 and 2009 also did not adjust for areas uncovered in the 1998 census (see NIS (2010)).

A review of the estimation procedure for 2009 revealed that the procedure gave a slight upward bias. The procedure has consequently been adjusted and the 2009 estimates have been updated.

The population of Cambodia distributed by sex is shown in figure 1 above. The sex ratio (men in relation to women) has increased significantly between the two censuses but in recent years there seems to be no change, but it might be described as "a normal" at the national level that still denotes an excess of women than men varying over a rather narrow range from about 10,5000 to 14,000 in thousand for both sexes.

Household size demographics

One of the interesting parts of the CSES is the household size demographics. Figure 2 below shows projected average household size based on currently used model. In trend the project average household size is going to decrease every year in entire Cambodia, both urban and rural.
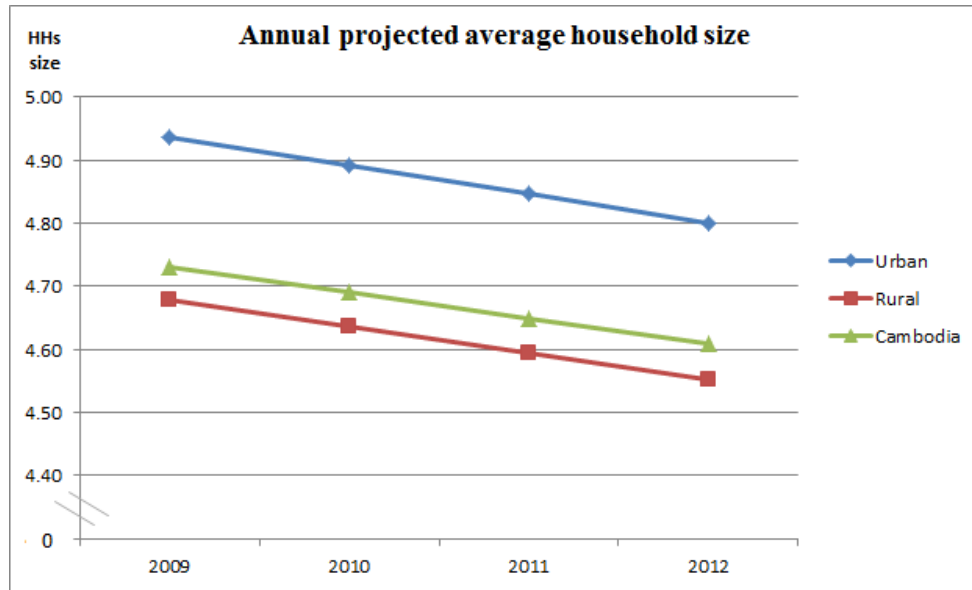
Figure 2: Projected average household size based on currently used model

The estimated household size for 2009 are greater than those projected for 2010, 2011 and 2012 including urban and rural (see figure 2 above).

However, figure 3 below shows a different picture, with estimates from the 2010 and 2011 CSES illustrating a steeper decrease in average household size than the modelled assumptions suggest. We also have indications from the migration statistics in CSES 2010 that the population in Cambodia was moving to Phnom Penh and other abroad. Most for job opportunity and in some cases (Marries, Repatriation or return after displacement, Insecurity, Transfer of work place. Etc.). In these cases the household size is decreasing.

Figure 3 shows the estimated mean household size in the CSES surveys and the mean household size in normal households according to census 2008. The census figure, is recorded after adjusting for the undercount.
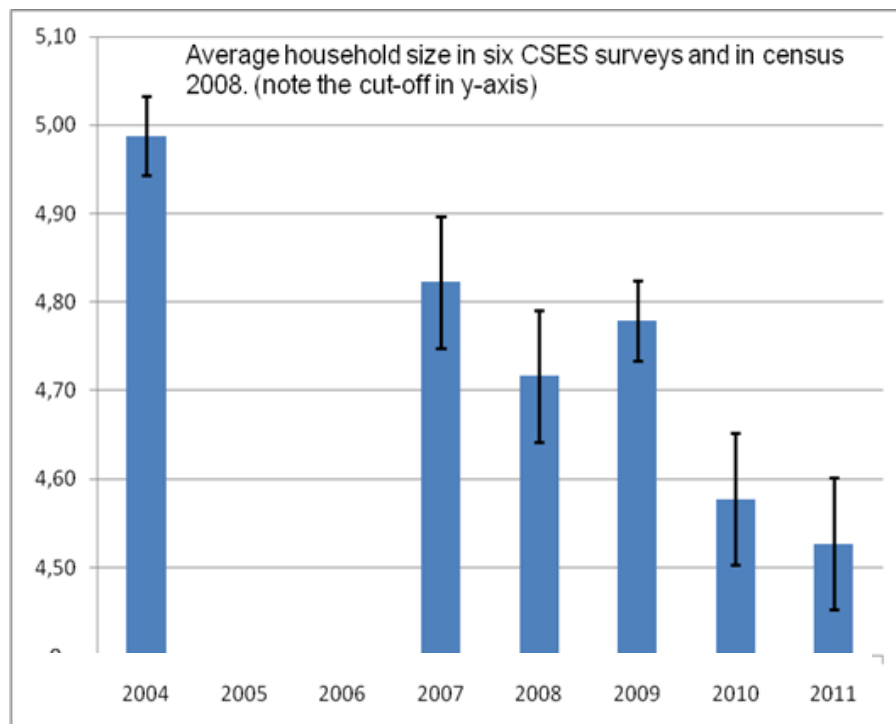

Figure3: Average household size in six CSES and Census 2008

Since we believe that systematic measurements bias is small in recording the CSES household sizes, we suspect from comparing figures 2-3 that the assumptions of change in household sizes needs to be revised. There are signs that the decrease is faster than previously assumed. The CSES 2012 and the Intercensal survey 2013 will spread new light on this.

7. Final Comments and conclusion

For a country, it is of outmost importance that the population statistics are reliable. Another requirement that sometimes appears is consistency in the numbers and estimates, hence that many sources provide if not the same but at least a similar picture of reality. The Cambodian 2008 census and its population projections have up till today been the baseline for improving the estimates of the CSES. This paper describes this estimation methodology and it also raises some questions and doubts, which are the key findings of this paper.

For example, the census results are getting older and there are signs that the accuracy is decreasing. Moreover, the methodology used to project the household sizes do also show signs of being off the mark. The future plans are to compare these findings with the results from the ongoing Cambodian Intercensal survey. It will be used to assess the quality of the census projections and to draw conclusions of how to improve future demographic estimates of the CSES.

8. References

Arriaga, E. (1994). Population Analysis with Microcomputers, US Bureau of the Census, New York

Pettersson, H. (2010). Average household sizes in the CSES-surveys, Internal report, National Institute of Statistics Cambodia.

Särndal, C-E. and Lundström, S. (2005) Estimation in Surveys with Nonresponse. New York, Wiley.

National Institute of Statistics Cambodia (2004) First revisions of Population Projections for Cambodia 1998-2020. (June 2004).

National Institute of Statistics Cambodia. (2010). Analysis of the Census Results Report 11, Family and Households. NIS June 2010.

National Institute of Statistics Cambodia. (2011) General Population Census of Cambodia 2008, Analysis of the Census Result Report 12 Population Projection of Cambodia. (January 2011).