

# **Fertility analysis with SILC: A quantification of measurement bias**

Aurélien Dasre (Université Paris Ouest Nanterre La Défense, GTM CRESPPA)

Angela Greulich (Université Paris 1 Panthéon Sorbonne)

November 2015

## **Abstract**

The aim of this paper is to quantify the potential bias of fertility measures in the European Survey of Income and Living Conditions (EU-SILC). This survey is more and more used in socioeconomic analysis as it provides harmonized measures of socioeconomic variables available for over 30 European countries. However, as an economic survey, it is not directly shaped to ensure a good measurement of demographic variables, in particular of fertility. This results from the fact that in the questionnaire, there is no question on the number of children ever born. By comparing the fertility measures of EU SILC to unbiased measures from the Human Fertility Data Base (HFDB), this paper quantifies the extent to which the missing question on children ever born leads to a systematic underestimation of different fertility measures in SILC. We hereby decompose the bias by age and child rank as well as by socioeconomic categories by applying a longitudinal as well as a cross-sectional perspective.

## **1. Motivation/context:**

EU-SILC provides internationally comparable measures of living and income conditions. Children are only observed when living in the households of respondents, which leads to a systematic underestimation of fertility. This represents a particular problem for demographic analysis trying to model fertility as a function of socioeconomic variables. However, socioeconomic analyses are also concerned as economic outcomes of individuals and households are most often controlled for family size and structure. More and more studies use the EU SILC because of numerous advantages discussed below, but only few of them are concerned about the potential bias of results that can emerge due to the problematic fertility measures. We therefore propose a quantification of the survey's measurement bias in fertility by answering the following questions:

- What countries are concerned most by the bias of fertility measures and why (register countries, countries with long cohabitation of children with parents, countries with high percentage of childlessness, countries with important birth postponement etc.)
- What age of parents and which rank of children is most concerned by the bias? Are the biases different for different socioeconomic categories?
- Do the biases also occur when measuring fertility in a retro perspective way or only when measuring fertility in a cross-sectional way?
- How to circumvent these biases at best?

For answering these questions, we proceed the following way:

- Based on the number of children observed in the household, we calculate fertility by age, which allows us computing SILC measures of completed fertility rates as well as of total fertility rates (as well as their decomposition by child rank). We compare these SILC measures with unbiased measures reported by the Human Fertility Data Base in order to quantify the measurement biases.
- We compile the probability of child birth (as well as of child departure) while comparing retro perspective measures of birth (and departure) obtained with the cross-sectional data base with measures obtained with the longitudinal database of SILC (4 year rotational panel). This procedure allows quantifying the measurement bias caused by attrition.
- In order to obtain information about measurement bias differentiated by socioeconomic categories, we compare the French SILC data that is integrate in the international data base to the original French survey on income and living conditions, which asks the respondents about their children ever born.

## 2. Data Description

The European Union – Survey on Income and Living Conditions (EU-SILC) is a European survey provided by Eurostat. This survey was created in 2003 as a replacement for the European Community Household Panel (ECHP) and now includes thirty-one European countries. It gathers harmonized and comparable data at the individual and at the household level on income and living conditions as well as on many individuals' characteristics (sex, age, education, labour market position, income etc.).

EU-SILC is composed of two datasets – one cross-sectional and one longitudinal. The annual cross-sectional data are produced from the longitudinal panel (integrated design). The longitudinal dataset of EU-SILC is a rotational panel of four years, which means that for the majority of countries, individuals are observed for a maximum period of four years. The integrated design allows for a high number of observations for the cross-sectional data base. In the cross-sectional data base,  $\frac{1}{4}$  of individuals are observed for the first time,  $\frac{1}{4}$  for the second time,  $\frac{1}{4}$  for the third time,  $\frac{1}{4}$  for the fourth time (which reduces measurement bias due to cumulated respondent burden and sample attrition).

The survey contains information on both individuals and households. It is possible to identify adult women, their partner – if they have any – and the children who live in the same household.

The EU-SILC database contains a large number of economic and social variables that capture the situation of individuals and households and that may be considered as determinants of the decision to have children. It displays basic information on age and education level but also variables on individuals' labour market status (reported on a monthly basis) and income. This information is rarely available in other, more 'demographic' surveys. One exception is the Gender and Generations Surveys, but this survey has very limited country coverage. Other surveys as the European Labour Force survey contain information on work, but not on income. Some surveys exist that contain both demographic and economic variables with individuals followed up for more than only four years, but the limit of these data sets is their national focus

since these long-run surveys are generally run in only one given country (the German Socioeconomic Panel or the American Panel Study of Income Dynamics for example).

The advantage of EU-SILC is to propose a comparative perspective of European countries using a harmonized database for a large set of economic variables. Measurement biases in term of fertility emerge, however, due to the fact that the EU SILC does not report information on the number of children directly. However, children are observed with a proper identification number when living in their parents' households, and households are followed when moving. Individuals who split from their original household are, however, hard to follow up, which causes attrition issues.

**3- Preliminary Results:**

Figure 1 illustrates the downward bias in the number of children for women aged 45 by country while figure 2 decomposes the fertility bias by age and rank, exemplarily for Belgium.

Figure 1: Comparison of fertility at age 45 (HFDB vs. EU-SILC CS 2009)

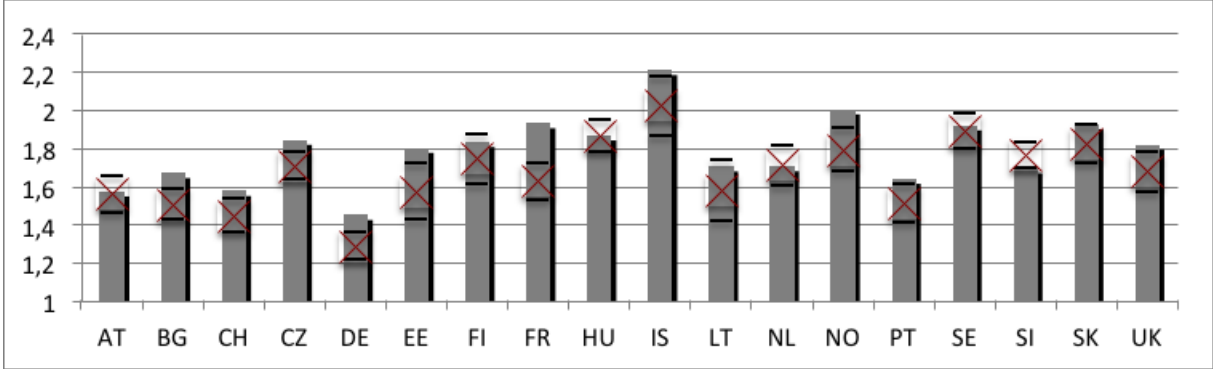
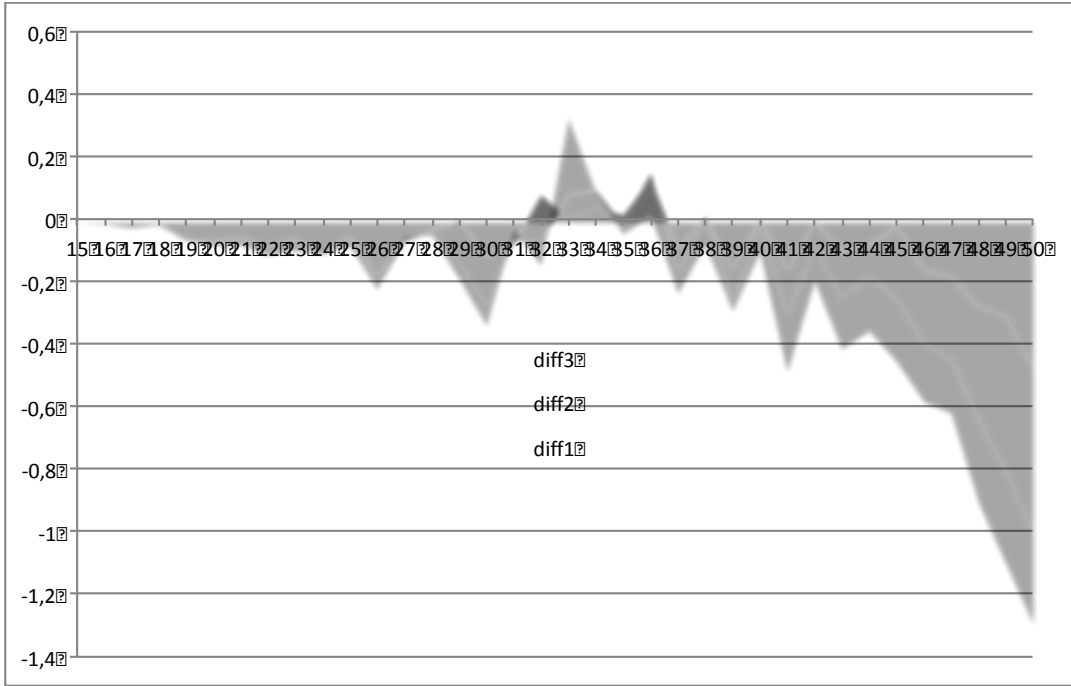


Figure 2: Fertility bias by age and rank (Belgium, EU-SILC CS 2009)



The strong downward bias in fertility for ages 40+ is mainly due to children moving out of their parents' households. Not only is the fertility measure downward biased, but the ranks of children are no longer reported correctly. Once the oldest child has left the household, the younger sibling will be falsely considered as being of rank  $n-1$ . The older the women and the younger her age at first childbirth, the higher is the potential risk of underestimating her number of children. Total fertility rates are therefore less biased than completed fertility rates, as for period measures of fertility, only new child births are taken into account.

However, total fertility rates also risk to be biased, mainly due to selection and attrition: Selection issues arise as single households are underrepresented in SILC and are difficult to follow up, especially at younger ages. Observing their first years of family formation is therefore also challenging, which leads to an underestimation of first child births. The weighting process applied in SILC is not sufficient to control for this sort of problem. Attrition issues arise mainly when individuals are not followed up when splitting from their original household. When entire households move, they are normally followed up in SILC. However, a child leaving the parents' household is most probably not followed up in SILC. If a child leaves the parents' household in order to found an own family, first child births of this child are thus likely to be underreported. In this case, attrition emerges as individuals risk dropping out of the survey due to child arrival.

Finally, the different measurement biases are likely to differ between socioeconomic categories. For example, the younger the women at first childbirth, the higher the risk of attrition and the risk of underestimating her completed fertility. This profile of women is also likely to be relatively low educated, to have a relatively unstable position on the labour market etc.