Using a Labour Force Survey to Derive Information on Fertility Histories

VALERIA BORDONE
FRANCESCO C. BILLARI
GIANPIERO DALLA ZUANNA

INDEX

1. THE SURPRISING LACK OF STATISTICAL INFORMATION ON BIRTHS IN ITALY

Fertility is an individual phenomenon, we can even define it as a private one. However it also has deep consequences at social level since couple's choices about procreation determine the replacement between generations and lead to economic and social development [Billari, 2004].

The new millennium, started with very low fertility levels in the ECE area and a lowest low fertility in some of these countries. Only few countries still have a TFR at or above the replacement level[1].

Differences between very low fertility and lowest low fertility are small. However, once fertility is under the replacement level (2.1 children per woman in fertile age in low-mortality countries), decimals assume a very important role for population changes.

Italy can be considered as a very particular Country because of its fertility behaviour. Over the last years, Italian fertility level has been one of the lowest in the world. Italy has been characterized by a collapse of the inclination to have more than 2 children per couple and by an increase in the postponement of childbearing [IRP, 2002]:

Earlier notions that fertility levels may naturally stabilize close to replacement level have been dispelled: in the early 1990s Italy and Spain were the first countries to attain and sustain lowest-low fertility levels[2]. A Total Fertility Rate (TFR) at lowest-low fertility level is clearly far from the demographic equilibrium. A sustained lowest-low fertility implies demographic, economic and social consequences, such as an annual decline of the population size by at least 1.5% in a stable population with an overall mean age of women at childbirth of 30 years.

Contrary to popular belief, Italy[3] has never had very high levels of fertility. Data prove that the Italian fertility decline has been firmly established by the end of the Word War I, when fertility control had already been practiced for half a century in many

[1] In 2000, countries with a TFR at or above 2 in the group of ECE area were Israel, Turkey and USA, as reported by Council of Europe [2002], Frejka and Sardon [2003], UNECE / PAU Demographic Database, UNECE Gender Statistics Database.

[2] According to Kohler, Billari and Ortega [2002] we consider a country with lowest-low fertility if its TFR (TFR) is at or below 1.3.

[3] And Spain, as described in Delgado Pérez and Livi-Bacci [1992].

regions. A constant feature of the fertility transition in Italy has been the large differences between geographic regions: in some of them the fertility decline started during the last decades of the 19th century, while in less developed areas it begun only in the 1940s and 1950s.

The two main surveys about Italian fertility (run in 1979 and 1996) show that a continuous and slightly decrease in the total fertility rate led to a negative TFR peak in 1995 (with 1.18 children per woman). Then the TFR seemed to level off at around 2 for a few years before declining again. In 2003 the TFR was 1.26 children per woman [Saraceno, 2004].

Even if fertility behaviour is a main problem of Italian demography, there is a lack of surveys that provide complete, recent and updated data about births and population in this country [Castiglioni, 2004].

The availability of data on births until 1996 could rely on the *Marital Status Source*: it showed characteristics of parents, children and also birth order. The period 1999 – 2001 was covered by surveys based on aggregate data without birth order. In 2002 the Ministry of Health started the *CEDAP FLUX survey* that provided characteristics of parents and children. This can be used for statistical analysis or public health purposes and gets back birth order at a regional level.

To recover lost information about births alive over the period 1999 – 2001, ISTAT provides *sample surveys*. The first survey of this type was conducted in 2002 about births between June 2000 and July 2001.

The *Register of births, marriages and deaths* is one of the sources where population is classified by sex, age and marital status at 31st of December of every year at municipality, province and region level[4].

The *Census* contains structural characteristics of the population, the educational level and the working activity (every 10 years) of everyone[5].

Therefore there is a lack of sources that provide at the same time data on population (with socio-economic characteristics) and number of children.

---

[4] This will be soon available also for stranger population, classified for sex and age, starting by a municipal level [Castigioni, 2004].

[5] Today Census provides just sex, age and marital status, family components, for Italian and stranger residents [Castigioni, 2004].

In this paper we discuss an alternative way to gather fertility information by using the *own-children method* (*OCM*) applied to a large dataset, that is the *Italian Labour Force Survey* (*ILFS*).

Section 2 describes the *ILFS* and the *OCM*. Section 3 shows how we apply this method to the *ILFS* data. This process allows to bridge the gap of fertility data by deriving information about children and the population at risk from parents' characteristics[6]. We therefore obtain a large-scale dataset for fertility analysis in Italy. Subsequently, in section 4, a quality assessment of the reconstructed fertility dataset is provided by comparing the estimated TFR[7] with the official existing TFRs that ISTAT and Eurostat calculated at regional and national level.

[6] Maria Castiglioni [2004] explains that to use this method we have to know household's composition and the relationship.
[7] The estimated TFR is the one calculated by using data provided in the *ILFS*.

2. THE USE OF LABOUR FORCE SURVEY AND THE OWN-CHILDREN METHOD

2.1 The Labour Force Survey

The *ILFS* is a quarterly survey that has been continuously led by ISTAT[8] since 1959. It is the principal statistical source about the Italian job market: the 4 waves carried each year[9] provide information on over 300,000 households (about 75,000 households selected by the sampled municipalities are interviewed each quarter). This figure corresponds to 800,000 individuals (1.4% of the total national population), which are distributed in 1,351 Italian municipality.

The sample is rotating, thus the same household is interviewed in the 3rd, 12th and 15th month after the first interview. This method provides data for micro-analysis of individuals' behaviour and of working condition changing.

Since 1959 the survey has met many changes in order to satisfy three main purposes. Firstly, to take the transformations of the Italian job market into account. Furthermore, to satisfy the demand of a better understanding of some Italian socio-economic characteristics. Finally, to become consistent with the European statistics on the job market[10].

The reference universe of the survey includes all the components of the households in Italy, even if temporary emigrated abroad.

The unit of the survey is the *de facto* family, as in the broad meaning of people linked by marriage, relationship, affinity, adoption or family tie who live together.

We consider the 2003 Labour Force Survey. Table 1 shows data provided by this survey, in terms of umber of interviewed households and individuals in the four parts of the survey.

---

8 ISTAT is the Italian National Institute of Statistics.
9 In 2004 the Labour Force Survey became a continuous survey instead of a quarterly one.
10 EUROSTAT attempts to harmonize different national surveys / statistics and to create a unique model.

Table 1. *Number of interviewed households and individuals in the four parts of the Labour Force Survey of 2003.*

	Interviewed Households	**Interviewed Individuals**
January	75,292	194,076
April	75,333	192,359
July	75,281	191,324
October	75,394	192,043

Notes: data in column "Interviewed Individuals" may contain duplications because these data are those of the dataset without any correction and besides an individual can be interviewed for two following survey.

In the Labour Force Survey there is neither a question about the number of children per woman nor birth dates of these children. However it is possible to match mothers and children over the 15-20 – year period foregoing the interview by using the *OCM*.

2.2 The Own-Children Method

As a large-scale dataset for fertility analysis in Italy is not available, we reconstruct birth histories by using the so-called *OCM*. This is a census or survey-based reverse-survival technique for estimating age-specific fertility over the years before the enumeration [Retherford and Cho, 1978].

The principal obstacle to the study of the growth component of the world population has been the absence or unreliability of vital statistics in the majority of nations. However, many of these nations with poor or nonexistent vital statistics have recently been conducting population censuses. Therefore this census information can be used to measure the growth components of the population in the absence of adequate vital statistics [Grabill and Cho, 1965]. The advised method to amend the fertility data deficit is the use of information on own-children as provided by national censuses [Grabill and Cho, 1965; Rindfuss, 1976] or other surveys. Since many censuses have the appropriate data, the own children technique has potential applicability in a large number of settings.

The *OCM* was formulated during the Sixties and it has been used in a lot of Countries, both in the developed and in the developing world. The diffusion in the second type of countries is due to its advantages given to its simplicity [Maffeini and

Rossi, 1984]. It is just required a particular data processing that matches each child with his mother.

The first step of the procedure is to match children and mothers within households (this is possible, of course, only when mothers are present) [Retherford and Bennett, 1977]. To do this, it is necessary to use the answers to questions about the relation with the head of the family, age, sex, marital status and number of children surviving or even born. These matched children, classified by own age and mother's age, are reverse-survived to estimate births by age of the mother in previous years.

Reverse survival is similarly used to estimate the number of women by age in previous years. After some adjustments for incorrect enumeration and unmatched ("other than own") children, age-specific birth rates and birth probabilities are calculated by dividing the first figure by the second [Retherford and Cho, 1978].

Information on own-children present in the home do not involve any new question on the census schedule. Still it can be obtained by performing a recording operation on the existing census schedules [Grabill and Cho, 1965]. Once the number of own-children present has been coded for a given woman on the census schedule, the data can be tabulated by other characteristics of the woman and her family that may be available from the census schedule or the survey one.

Own-children are actually defined as all children who can be identified as living with the mother. This definition includes some adopted children or stepchildren and excludes any offspring who may have died or moved away. Whether a child cannot be matched to a "mother" living in the same household, it is defined as a *non-own child* and, thus, excluded [Rindfuss, 1976].

In countries where early-age mortality rates are low and the nuclear family predominates, most of the children born in the years immediately preceding the census are living with their respective mother. Therefore they can be enumerated with her.

Thus, whether infant mortality rates are low, whether young people prefer to stay at home with parents rather than to live alone or with friends before marriage and whether children generally stay with their mothers in cases of separation or divorce, this method doesn't create great problems [Desplanques, 1994]. Italy is a country that satisfies everyone of these three peculiarities.

The essential innovation in the own-children method is the acknowledgement that both censuses and household surveys contain substantial implicit data on fertility. Moreover, certain provided information enable us to match children to their biological mothers. This is due to the enumeration of most children in the same household as their mothers. It is conventional to refer to these children as *own-children*.

In this way, it is possible the association between the survivors of the births occurred in previous years and their mothers by matching the respective features. Education and other characteristics included in the census or survey [Luther and Cho, 1988] allow to derive age-specific fertility rates.

The *OCM*'s outcomes are sometimes less satisfactory than those of the classic method of measurement since some children live with the father or with other relatives, such as grandparents, instead of the mother. The age specific fertility rates, and the resulting TFRs are therefore underestimated, being calculated only from children living with the mother.

A problems encountered in our analysis, is actually the *Grandmother Effect*: whether a young single mother lives with her parents, it is likely that her child is wrongly recorded as being her mother's child [Desplanques, 1994].

An obstacle not addressed in earlier expositions or applications of the *OCM* is the *sampling variability*. This is important, however, because usually the method is applied to a systematic sample of households (usually from 1 to 5%) taken from a total census count. Sampling variability of age-specific birth rates becomes a serious problem whenever small numbers of births and women are encountered. As described by Retherford and Bennett [1977], this occurs in three types of situations: whether the sample is small; whether detailed breakdown into categories of rates is desired; and whether detailed cross-classification of rates by characteristics is desired[11].

Although there is another drawback when the *OCM* is applied, it still remains a good method because it has some important advantages over other methods used to

[11] In the last two situations, problems of sampling variability may be severe even when the total sample is quite large, and may overshadow other sources of error [Retherford and Bennett, 1977].

obtain age-specific birth rates for previous years. In the *OCM* age-specific birth rates for previous years can normally be tabulated only according to characteristics at the time of census or survey, and not at the time that the births occurred [Retherford and Cho, 1978]. However, when the *OCM* is applied to an existing census or survey, it does not require any additional data [Retherford and Cho, 1978]. Moreover, the *OCM* is not very sensitive to life table estimation errors under mortality levels currently prevailing in most parts of the world[12].

The method has the further advantage that birth rates can be tabulated by characteristics, such as education and occupation. This kind of information is collected by the census or survey questionnaire[13].

Birth rates based on own-children therefore provide an excellent base for studying differential fertility and, when more than one census or survey is available, for studying the factors that influence fertility trends.

There will always be a certain share of children not-linked with a mother because of the absence or the inadequacy of necessary data, the death of the mother before the survey or different addresses of the mother and her children [Maffeini and Rossi, 1984].

However, even whether the classic method can be used, the *OCM* may prove both pertinence and complementarity. The own-children birth histories[14] contain substantial information on birth intervals and parity progression. An own-children birth history differs from a complete birth history only because it excludes births of a woman's deceased children and of surviving children who no longer live in the mother's household. Since the numbers of these omitted births are often relatively small for 15-year period that precedes the census or the survey, the own-children birth histories may include most of the complete birth histories.

---

[12] As the method is so self-contained, it is especially useful in countries with a lack in adequate vital registration statistics. [Cho, 1971].

[13] This is not possible for rates derived from vital statistics, since only a few basic characteristics of the mother are recorded on the birth certificate. [Retherford and Cho, 1978].

[14] That is the sequence of births in previous years corresponding to a woman's own children [Luther and Cho, 1988].

3. THE OWN-CHILDREN METHOD IN OUR ANALYSIS

This analysis considers four datasets (the first survey in 2003 – *January* –, the second one – *April* –, the one of *July* and the one based on the last interview of the year – *October* –). For every dataset, we look for likely mothers, likely children and then we try to link every child to the respective mother.

The initial hypothesis is that children live in the house with the respective mother[15].

In the first dataset, we can match 35,019 children to their likely mothers (34,498 are children of women declared "head of the family or partner of the head of the family" and 521 are children of women who are "daughters of the head of the family"). Dataset *April* lets us match 34,410 children with their respective mothers (33,876 + 533 + 1 son of a woman that is probably the mother of the head of the family or the mother of his partner). In the dataset *July* we can consider 34,103 children linked to their mothers (33,688 + 413 + 2) and in the dataset *October* 34,895 (34,467 + 426 + 2). In all these cases we consider just women born between 1946 and 1988 and we introduce a constrain that requires children to be born later than 1983[16].

These figures correspond to the percentages contained in the following table 2.

Table 2. *Percentages of children linked with respective mothers in the four parts of Labour Force Survey 2003 and in the three group of children, divided on the basis of their relationship with the head of the family.*

	Children of the head of the family	**% linked**	**Children of a daughter of the head of the family**	**% linked**	**Children of an ancestor of the head of the family**	**% linked**
January	35,809	96.3	1,226	42.5	95	0.0
April	35,223	96.2	1,187	45.0	109	1.0
July	34,898	96.5	1,225	43.6	153	1.3
October	35,683	96.6	1,323	41.5	159	1.3

Notes: "Children of a daughter of the head of the family" are those children classified as *other relatives* and "children of an ancestor of the head of the family" are those people classified as *head of the family or partner of the head of the family*.

[15] See the description of the OCM in paragraph 3.2.

[16] These requirements are a good solution for the "*grandmother effect*" described above: by considering mothers born between 1946 and 1988 and children born later than 1983, we basically exclude every situation in which we could match the "children" with the grandmother.

The analysis of this paper can be considered as the sum of the three shares of table 2: a first share where we consider the children of head of the family, a second share where we consider the children of a daughter of the head of the family and a third one in which we consider the children of an ancestor f the head of the family.

Some possible errors are excluded in all the four datasets taken into account:

- for all the three parts we exclude observations where mother's year of birth bigger than her son's one;
- for the second and the third parts we exclude cases in which the difference between mother's year of birth and her son's one is not included between 15 and 40 years;
- for the second part we exclude observations in which a woman seems to have at least a son, but there are not data on him and observations in which a woman seems to have at least two children, but there are data just on her first son; and
- in the second part we also have to cut observations that lead us to link a child to two women (generally, one is the actual mother and the other one is her sisters).

As we require that mother's year of birth is between 1946 and 1988 and that children's year of birth is later than 1983, it is possible to explain why we have so little matches between mothers and children in our third part of the analysis: we have to delete almost all the possible children because it is really difficult to find an household in which the head of the family is a boy or a girl living with his/her mother and who is at most 20 years old. For this reason, we consider as a normal situation to have a decrease in the number of children matched with respective mothers between the first and the second groups and even a cut when we change over the third group of our analysis.

We add the constrains described above to ensure that we are minimizing the possibility of mistakes and, as seen, we increase our requirements in second and third parts (where we consider just observations where the difference between mother's year of birth and her son's one is included between 15 and 40 years.

As described in Dalla Zuanna [1989], there are some sure matches and some uncertain matches. The sure ones are those between a mother who is *the head of the family or the partner of the head of the family* and a child who is *the son of the head of the family*. The uncertain matches are those between a mother who is a *son of the head* *of the family* and a child who is *other relative*, and those between a mother who is *ancestor of the head of the family or of his partner* and a son who is *the head of the family or the partner of the head of the family*. In these last two cases, we require that the difference between the birth year of the mother and that of her son is included between 15 and 45 years.

The use of the *OCM* outputs a table like the following table 3[17].

Table 3. *An example of the table created using the own-children method.*

KEY - VARIABLES	BIRTH DATE OF THE WOMAN	BIRTH DATE OF THE FIRST SON	BIRTH DATE OF THE SECOND SON	BIRTH DATE OF THE THIRD SON	BIRTH DATE OF THE Nth SON	NUMBER OF CHILDREN
…						

Key-variables are all the variables that are necessary to identify an individual, in our case they are household progressive general number, municipality, province, region, section, household number, list, interviewer code, number of household components, order number. These key-variables identify a woman who could be a mother. In column we have all the women (with and without children).

[17] In the Labour Force Survey there isn't any question about the number of children that women have had. Still, we can create this variable and add it in a new column of this table.

4. AN ASSESSMENT OF DATA QUALITY

To test if data provided by the *ILFS* are useful, we have to control them. We can do it by comparing the TFR calculated on the basis of these data and those official produced by ISTAT and Eurostat.

Total Fertility Rate (TFR) is the final intensity of fertility:

$$TFR = \sum_{x=\alpha}^{\beta} \sum_{o=1}^{\omega} {}^{o}f_{x}^{i} \qquad (1),$$

where:

f = children born

x = age

i = generation

o = birth order.

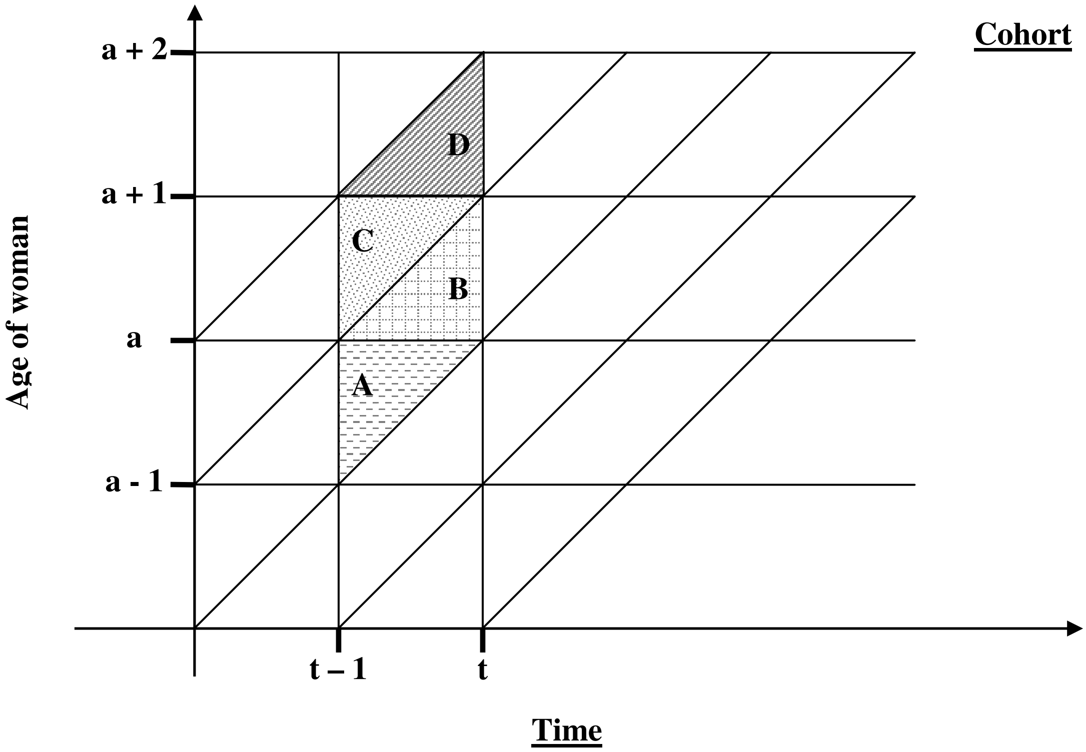
In our case, we don't take into account any distinction between different birth orders. Therefore expression (1) of TFR can be simplified as follows:

$$TFR = \sum_{x=\alpha}^{\beta} f_{x} \qquad (2).$$

The *Lexis diagram* describes graphically how to construct fertility life of women in the sample.

Figure 1 shows the Lexis diagram, which is a plot of a population's life experience in time vs. age. Births in the square *BC* are obtained by adding half the births in *AB* to half the births in *CD*. Mid-year population in *BC* is obtained by averaging the population at the beginning and at the end of the year, as represented by the left and right edges of the square *BC*.

Graph 1: *Lexis diagram.*



The current analysis reconstructs the fertility life of every woman in the sample. By using the Lexis diagram language, it is said that we follow the diagonal lines from the top to the bottom.

As described above, mothers are of three types: head of the family or partner of the head of the family, daughter of the head of the family or parents of the head of the family or of his/her partner. Despite of the absence of children with characteristics required by the third group, it is necessary to consider mothers of this type in the TFR's denominator as they are women in fertile age and therefore at risk of having a child.

4.1 The Italian Total Fertility Rate

Data from the 2003 *ILFS* lead to calculate the Italian TFR between 1990 and 2002. We consider women aged 14-58 years in 2003: using these data as base, we can outputs a table that contains the number of women in reproductive age (14 – 45 years old) over the 12-year period 1990-2002, under the hypothesis of absent geographical mobility.

Table 4. *The table of the women.*

AGE OF WOMAN	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
14														***
15														***
16														***
17														***
18														***
19														***
20														***
21														***
22														***
23														***
24														***
25														***
26														***
27														***
28														***
29														***
30														***
31														***
32														***
33														***
34														***
35														***
36														***
37														***
38														***
39														***
40														***
41														***
42														***
43														***
44														***
45														***
46														***
47														***
48														***
49														***
50														***
51														***
52														***
53														***
54														***
55														***
56														***
57														***
58														***

The starting point in the table 4 is the column shown with three asterisks (***). From these data it is possible to fill in all the squares on the left by following the diagonal black arrow that is drawn as an example.

The fertility life of these women is then reconstructed by matching every woman with the respective children and by attributing value 0 to the variable "number of children" for women without children.

The whole number of children is divided between "*linked children*" and "*not-linked children*" to distinguish those who can be matched with a mother and those who can't.

If the difference of age[18] between a mother and her linked child is not included in the interval 14-45, we consider this child in the group of *not-linked children*[19]

Whether we consider only the group of *linked children*, we are allowed to create a table similar to table 4. In its squares there is the number of children born by a mother with an age *x* in the year *y*, where 14<=*x*<=45 and 1990<=*y*<=2002.

Data in the table of children and those in the one of women allow the calculation of the Italian TFR between 1990 and 1992 by applying the formula number (2).

The number of children born in the year *y* is the sum of *linked children (LC)* and *not-linked children (NLC).* Therefore the following formulae hold:

$$BornChildren_{adjusted} = LC + NLC \qquad (3).$$

Given that $\%NLC = \frac{NLC}{LC + NLC}$ (4),

then $NLC = LC * \frac{\%NLC}{1 - \%NLC}$ (5).

The (3) can be written in the next way:

$$BornChildren_{adjusted} = LC + \left[LC * \frac{\%NLC}{(1 - \%NLC)} \right]$$

$$= LC + \left[1 + \frac{\%NLC}{(1 - \%NLC)} \right] \qquad (3b).$$

---

[18] Age is calculated in century months.
[19] We give them a different code to distinguish between the real not-linked children and those not-linked because of a mother too young or too old.

Similarly, we can calculate an adjusted TFR by adding the *not-linked children* to the linked ones in the way described by formula (6):

$$TFR_{adjusted} = TFR * \left[1 + \left(\frac{\%NLC}{1 - \%NLC} \right) \right] \qquad (6),$$

where $\%NLC$ represents[20] the percentage of children characterized by a known date of birth and a non-matching with any woman in the group of the mothers..

The existence of four datasets in a single year (*January*, *April*, *July* and *October*) makes the use of an average value necessary to obtain a single figure for the TFR. The $TFR_{adjusted}$ gives a TFR more similar to the official values provided by ISTAT and Eurostat.

[20] As calculated in formula (4).

Table 5. *Linked Children, Not-Linked Children and their percentages in Labour Force Survey dataset. Italian Total Fertility Rate estimated using Labour Force Survey and Italian Total Fertility Rate adjusted.*

	1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Linked Children	421559	411708	427288	413880	404090	408271	384666	394195	415350	417196	413150	401746	374515
Not Linked Children	20855	22452	22951	19686	17920	18175	20624	22330	18966	20318	24118	27894	25201
Born Children	442414	434159	450239	433567	422010	426445	405289	416524	434316	437514	437268	429640	399716
% Not Linked	5,05%	5,07%	5,11%	4,38%	4,34%	4,29%	5,06%	5,37%	4,47%	4,90%	5,49%	6,32%	6,40%
1 - % Not Linked	94,95%	94,93%	94,89%	95,62%	95,66%	95,71%	94,94%	94,63%	95,53%	95,10%	94,51%	93,68%	93,60%
TFR	1,32	1,29	1,31	1,26	1,22	1,23	1,15	1,18	1,23	1,24	1,25	1,22	1,14
$TFR_{adjusted}$	1,39	1,35	1,39	1,31	1,28	1,28	1,22	1,25	1,29	1,30	1,32	1,30	1,21

Notes: TFR = Estimated TFR and $TFR_{adjusted}$ = Estimated $TFR_{adjusted}$.

Table 6. *Official Total Fertility Rate in Italy.*

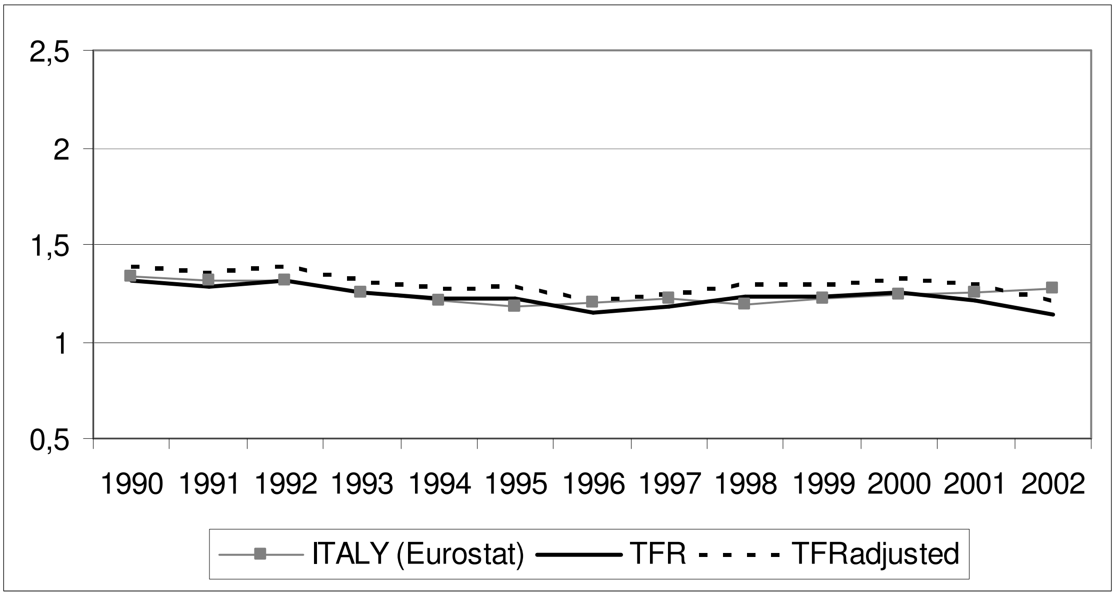
	1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
$TFR_{Eurostat}$	1,33	1,31	1,31	1,25	1,21	1,18	1,2	1,22	$1,19^{p}$	$1,22^{p}$	$1,24^{e}$	$1,25^{e}$	$1,27^{e}$
TFR_{ISTAT}	1,36	1,33	1,33	1,26	1,21	1,18	1,20	1,21	·	·	·	·	$1,27^{*}$

Notes: p = Provisional Value, e = Estimated Value and * = figure by www.demo.istat.it.
Sources: http://epp.eurostat.cec.eu.int for TFR by Eurostat, http://www.osservatorionazionalefamiglie.it/demografici.htm?8 for TFR by ISTAT and www.demo.istat.it for data with an asterisk.

The high number of linked children in this part is due to the fact that we are considering only the first and the second group of mothers and children. This means an exclusion of children who are *head of the family* because there are no children of this type who were born after 1990.

However in the TFR's denominator the *mothers of the head of the family* are included because they are women in fertile age during period of the analysis.

Graph 2. *Total Fertility Rate calculated using Labour Force Survey (TFR), Total Fertility Rate adjusted as described by formula (6) and Total Fertility Rate by Eurostat.*



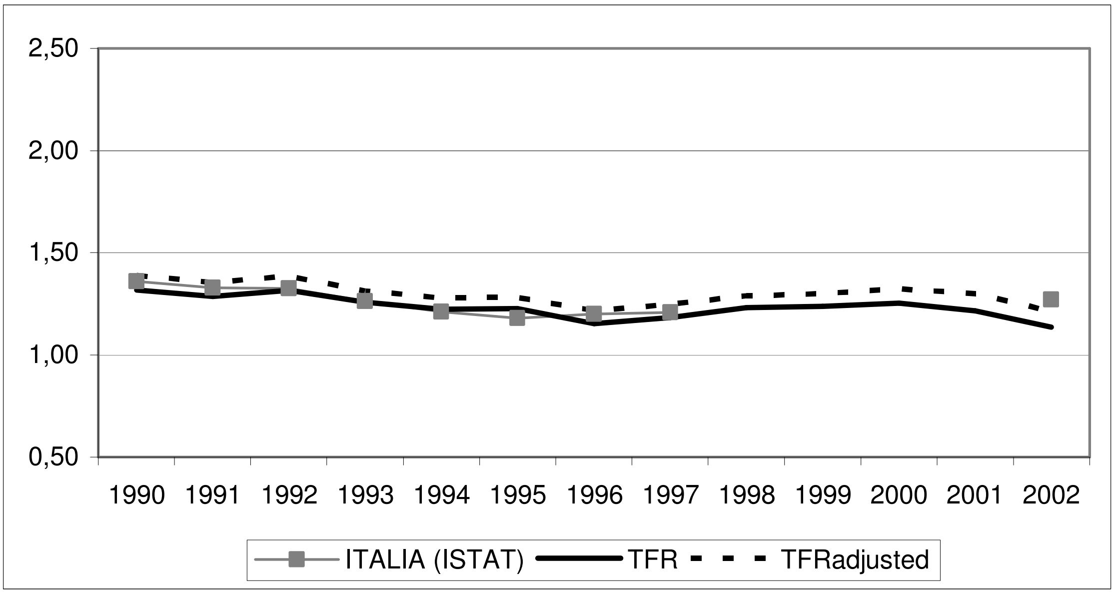
Notes: the source of "ITALY (Eurostat)" data is http://epp.eurostat.cec.eu.int.

Graph 2 suggests that estimated data from *ILFS* (the black line) are sometimes overestimating and sometimes underestimating the Official TFR (the grey line) by Eurostat. Data calculated in the current analysis are therefore quite similar to the real ones and their fluctuation around grey line can be read as a casual trouble with a zero average.

However, the trend of the $TFR_{adjusted}$ (the dotted line) appears as a constant overestimation of official data, except for 2002.

By using figures provided by ISTAT as official data instead of those by Eurostat, we obtain a similar trend (graph 3). The black line (the estimated TFR) is sometimes under and sometimes above the grey line (official data by ISTAT), while the dotted line (the estimated $TFR_{adjusted}$) is a constant overestimation of official data, except for 2002.

Graph 3. *Total Fertility Rate calculated using Labour Force Survey, Total Fertility Rate adjusted as described by formula (6) and Total Fertility Rate by Osservatorio Nazionale delle Famiglie.*



Notes: the source of "ITALIA (ISTAT.)" is the web site of "Osservatorio Nazionale delle Famiglie" http://www.osservatorionazionalefamiglie.it/demografici.htm?8 for data between 1990 and 1997 and www.demo-istat.it for data of 2002.

4.2 Total Fertility Rate in Italian Regions

The same method followed to calculate the Italian TFR by using the four datasets can be applied to the 20 Italian regions. The analysis consider again the 12-year period between 1990 and 2002.

Table 7 shows the data obtained as the average values of *January*, *April*, *July* and *October*'s TFRs.

As in the analysis at national level, it is necessary to create for every of the four datasets a table with the number of women in fertile age x ($14<=x<=45$) and a table with the number of children born in the year y ($1990<=y<=2002$). the only difference is that the same process is here applied to every single region instead of the whole country.

Table 7. *Total Fertility Rate in Italian Regions during the period 1990-2002 (Values estimated using data by Labour Force Survey).*

	1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Piemonte	0,99	1,00	0,92	1,09	1,03	1,00	1,03	1,13	1,03	1,11	1,22	1,20	1,16
Valle d'Aosta	1,10	0,98	0,85	1,18	1,38	1,10	0,84	0,94	1,30	1,30	1,27	1,43	1,25
Lombardia	1,10	1,06	1,11	1,11	1,10	1,10	1,03	1,10	1,13	1,15	1,14	1,22	1,20
Trentino-Alto Adige	1,35	1,33	1,43	1,20	1,30	1,42	1,36	1,30	1,33	1,41	1,62	1,44	1,42
Veneto	1,12	1,15	1,20	1,08	1,21	1,11	1,06	1,04	1,14	1,25	1,28	1,15	1,10
Friuli-Venezia Giulia	1,10	0,99	0,98	0,94	0,85	1,04	1,03	1,13	0,96	1,09	0,99	1,02	1,27
Liguria	0,83	0,95	1,02	0,95	0,93	0,78	1,03	0,97	1,11	0,86	0,96	1,10	0,96
Emilia Romagna	0,89	1,02	1,04	0,93	0,89	1,01	0,89	0,91	1,11	1,20	1,09	1,17	1,10
Toscana	1,14	1,06	0,93	0,95	0,97	0,94	0,91	0,89	1,17	1,11	1,07	1,17	1,04
Umbria	1,31	1,12	1,32	1,08	0,95	1,20	0,99	1,14	1,32	1,44	1,09	0,88	1,06
Marche	1,11	1,15	1,10	1,19	1,32	1,20	1,09	1,40	1,23	1,38	1,14	0,94	1,00
Lazio	1,28	1,18	1,21	1,19	1,17	1,11	0,99	0,94	1,04	0,94	1,15	1,05	0,93
Abruzzi	1,40	1,36	1,43	1,22	1,10	1,15	1,36	1,30	1,22	1,42	1,33	1,09	0,72
Molise	1,45	1,47	1,53	1,61	1,32	1,49	1,30	1,55	1,37	1,20	1,25	1,12	1,06
Campania	1,97	1,90	1,98	1,63	1,64	1,62	1,55	1,52	1,55	1,44	1,56	1,41	1,26
Puglia	1,66	1,59	1,57	1,74	1,47	1,59	1,49	1,36	1,62	1,34	1,29	1,37	1,14
Basilicata	1,90	1,65	1,41	1,67	1,51	1,15	1,28	1,22	1,36	1,55	1,26	1,27	1,18
Calabria	1,60	1,58	1,49	1,55	1,38	1,61	1,63	1,28	1,42	1,46	1,41	1,34	1,34
Sicilia	1,76	1,75	1,93	1,69	1,58	1,64	1,40	1,63	1,43	1,50	1,53	1,49	1,25
Sardegna	1,50	1,29	1,35	1,30	1,28	1,11	1,08	1,16	0,96	1,27	1,15	0,91	0,98

Formula (6) is still valid and leads to the calculation of an adjusted Total Fertility Rate ($TFR_{adjusted}$) for every single Italian Regions:

$$TFR_{adjusted} = TFR * \left[1 + \left(\frac{\%NLC}{1 - \%NLC} \right) \right] \quad (6).$$

The outcomes of the adjustment are grouped in table 8. The application of the formula (6) to the estimated regional TFRs (shown in table 7) provides these adjusted TFRs of table 8.

Table 8. *$TFR_{adjusted}$ in Italian Regions during the period 1990-2002 (Values estimated using data by Labour Force Survey).*

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Piemonte	1,01	1,01	0,94	1,14	1,09	0,99	1,11	1,15	1,02	1,14	1,28	1,21	1,30
Valle d'Aosta	1,13	1,01	0,86	1,20	1,41	1,13	0,86	0,90	1,39	1,41	1,29	1,33	1,28
Lombardia	1,17	1,11	1,16	1,16	1,10	1,13	1,08	1,11	1,18	1,21	1,21	1,30	1,29
Trentino-Alto Adige	1,38	1,35	1,50	1,24	1,33	1,44	1,37	1,36	1,37	1,43	1,64	1,56	1,54
Veneto	1,15	1,26	1,26	1,15	1,33	1,18	1,10	1,16	1,18	1,32	1,40	1,24	1,19
Friuli-Venezia Giulia	1,14	1,06	1,03	0,98	0,87	1,07	1,22	1,11	1,03	1,11	0,94	1,03	1,39
Liguria	0,90	1,05	1,01	0,97	0,98	0,83	1,12	1,01	1,21	0,88	0,94	1,16	1,07
Emilia-Romagna	0,95	1,09	1,10	0,99	0,94	1,06	0,93	0,94	1,13	1,27	1,13	1,29	1,18
Toscana	1,23	1,16	1,02	0,99	1,01	1,07	1,01	0,96	1,30	1,17	1,24	1,31	1,17
Umbria	1,35	1,24	1,43	1,25	1,04	1,31	1,10	1,30	1,34	1,59	1,24	0,97	1,18
Marche	1,17	1,30	1,10	1,30	1,39	1,39	1,06	1,53	1,31	1,41	1,22	1,11	1,05
Lazio	1,36	1,24	1,32	1,25	1,21	1,19	1,04	1,03	1,10	1,03	1,23	1,16	1,07
Abruzzo	1,51	1,35	1,49	1,23	1,15	1,20	1,39	1,40	1,45	1,47	1,36	1,18	0,80
Molise	1,53	1,51	1,56	1,60	1,30	1,63	1,32	1,59	1,52	1,20	1,31	1,21	1,13
Campania	2,10	2,01	2,08	1,70	1,70	1,66	1,62	1,68	1,66	1,48	1,65	1,48	1,48
Puglia	1,73	1,67	1,66	1,83	1,50	1,63	1,51	1,39	1,70	1,42	1,32	1,44	1,24
Basilicata	1,93	1,73	1,43	1,75	1,51	1,13	1,25	1,24	1,50	1,52	1,22	1,36	1,27
Calabria	1,73	1,66	1,53	1,58	1,45	1,73	1,66	1,35	1,54	1,51	1,47	1,43	1,42
Sicilia	1,90	1,81	2,02	1,76	1,59	1,66	1,48	1,72	1,51	1,51	1,61	1,61	1,35
Sardegna	1,52	1,26	1,42	1,38	1,35	1,15	1,09	1,17	1,01	1,38	1,13	0,94	1,09

Table 9. *Official Total Fertility Rate by ISTAT in Italian Regions during the period 1990-2002.*

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Piemonte	1,10	1,12	1,08	1,05	1,03	1,03	1,05	1,09	.	.	.	.	1,19
Valle d'Aosta	1,18	1,16	1,12	1,01	1,08	1,10	1,20	1,04	.	.	.	.	1,25
Lombardia	1,15	1,12	1,13	1,10	1,07	1,07	1,10	1,14	.	.	.	.	1,25
Trentino-Alto Adige	1,40	1,41	1,42	1,35	1,34	1,34	1,38	1,37	.	.	.	.	1,46
Veneto	1,16	1,12	1,14	1,09	1,06	1,07	1,10	1,13	.	.	.	.	1,24
Friuli-Venezia Giulia	1,03	1,02	1,04	0,95	0,94	0,94	0,98	1,11	.	.	.	.	1,12
Liguria	1,01	1,01	1,03	0,96	0,93	0,92	0,94	0,99	.	.	.	.	1,08
Emilia-Romagna	1,01	1,01	0,99	0,97	0,96	0,97	1,01	1,05	.	.	.	.	1,23
Toscana	1,08	1,05	1,05	1,02	0,98	0,98	1,00	1,04	.	.	.	.	1,17
Umbria	1,18	1,16	1,18	1,11	1,07	1,06	1,06	1,12	.	.	.	.	1,22
Marche	1,23	1,21	1,19	1,13	1,09	1,11	1,09	1,11	.	.	.	.	1,21
Lazio	1,28	1,23	1,26	1,21	1,17	1,11	1,12	1,14	.	.	.	.	1,22
Abruzzo	1,39	1,35	1,35	1,30	1,25	1,17	1,19	1,06	.	.	.	.	1,15
Molise	1,43	1,41	1,42	1,33	1,28	1,21	1,17	1,22	.	.	.	.	1,15
Campania	1,81	1,81	1,79	1,66	1,60	1,50	1,57	1,49	.	.	.	.	1,47
Puglia	1,65	1,60	1,58	1,49	1,44	1,37	1,37	1,36	.	.	.	.	1,30
Basilicata	1,66	1,56	1,57	1,44	1,36	1,33	1,27	1,04	.	.	.	.	1,24
Calabria	1,74	1,67	1,65	1,56	1,43	1,40	1,35	1,24	.	.	.	.	1,23
Sicilia	1,85	1,78	1,79	1,67	1,55	1,46	1,47	1,45	.	.	.	.	1,41
Sardegna	1,37	1,29	1,22	1,16	1,09	1,06	1,03	1,03	.	.	.	.	1,03

Source: Regional Total Fertility Rates since 1990 until 1997 are provided by http://www.osservatorionazionalefamiglie.it/demografici.htm?8 and figures for 2002 can be found on the web site www.demo.istat.it.

Table 9 shows what we have described in the second paragraph of this paper, about the lack of demographical data: official regional TFR can't be found for every year. In particular, over the period 1990 – 2002 it is available an official regional TFR since 1990 until 1997 and then for 2002, with an hole between 1998 and 2001.

However, the available data allow us to draw a graphical comparison between official data and our estimated values (both adjusted and not adjusted). The results of this confrontation are in Appendix 1 (graphs A1 – A20).

4.3 An evaluation of regional and national data quality

A further test of data quality is the Wilcoxon Signed-Rank Test. Since the one-sample sign test utilizes only the signs of the differences between each observation and the hypothesized median M_0, the magnitudes of these observations relative to M_0 are ignored. Assuming that such information is available, a test statistic which takes into account these individual relative magnitudes might be expected to give better performance. If we are willing to make the assumption of a symmetric population, the Wilcoxon signed-rank test statistic provides a test of location which is affected by both the magnitudes and signs of these differences.

The random sample has N observations $X_1, X_2, ..., X_N$ from a continuous cdf F with median M and we assume that F is symmetric about M.

Under the null hypothesis $H_0 : M = M_0$, the differences $D_i = X_i - M_0$ are symmetrically distributed about zero, so that positive and negative differences of equal absolute magnitude have the same probability of occurrence.

Therefore, the required assumptions are the independence of observations and a population which is continuous everywhere and symmetric.

The null hypothesis of the present analysis is that two observations come from the same distribution. In particular, we use the Wilcoxon test to compare first the estimated TFR and then the estimated $\text{TFR}_{\text{adjusted}}$ with the official regional TFR. Finally, we also apply the Wilcoxon test on data at national level.

For every region we compare first Estimated TFR with Official TFR and then $TFR_{adjusted}$ with Official TFR. Both the first and the second confrontation have 9 elements because we are considering figures over the period 1990 – 1997 and of 2002.

The last but one column in table 10 shows the outcomes of the Wilcoxon Test that compares the Estimated TFR with Official data, while the last column shows the outcomes of the Wilcoxon Test that compares $TFR_{adjusted}$ with Official data. In both the tests, the official data considered are those provided by ISTAT. This means that there is a lack between 1998 and 2001.

Table 10.

Region Code	Region	Population	MSE	MSE_{adj}	Wilcoxon Test	$Wilcoxon\ Test_{adj}$
1	Piemonte	4231334	0,007	0,008	0,227	0,977
2	Valle D'Aosta	120909	0,040	0,041	0,469	0,551
3	Lombardia	9108645	0,002	0,002	0,090	0,117
4	Trentino Alto Adige	950495	0,005	0,004	0,105	0,879
5	Veneto	4577408	0,007	0,014	0,910	0,055
6	Friuli Venezia Giulia	1191588	0,006	0,019	0,496	0,078
7	Liguria	1572197	0,009	0,007	0,070	0,805
8	Emilia Romagna	4030220	0,008	0,006	0,086	1,000
9	Toscana	3516296	0,008	0,006	0,043	0,383
10	Umbria	834210	0,012	0,025	0,969	0,023
11	Marche	1484601	0,024	0,047	0,813	0,219
12	Lazio	5145805	0,016	0,007	0,031	0,977
13	Abruzzo	1273284	0,034	0,036	1,000	0,547
14	Molise	321047	0,034	0,050	0,027	0,012
15	Campania	5725098	0,014	0,031	0,262	0,004
16	Puglia	4023957	0,017	0,026	0,313	0,016
17	Basilicata	596821	0,027	0,036	0,313	0,172
18	Calabria	2007392	0,022	0,029	0,910	0,117
19	Sicilia	4972124	0,014	0,021	0,680	0,039
20	Sardegna	1637639	0,012	0,024	0,031	0,008
ITALIA		57321070	0,003	0,003	0,148	0,078

To found a general rule about Wilcoxon Test results, we have to acknowledge that every possible rule doesn't work for southern regions. In fact, we would like to say that using a TFR adjusted improves our estimate. Still data in the last two columns of table 10 suggest that this is not true for southern regions and Veneto. Clearly, this fact influences the general situation of Italy, where we still have a good estimate.

The third column of table 10 shows the number of residents in every single regions. The two columns named MSE and MSE_{adj} contain the Mean Square Error, an old and proven measure of control and quality. MSE equals the mean of the squares of the deviations from target:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(x_i - T)^2 \qquad (7),$$

where:

x_i = ith value of a group of *m* values;

T = target or intended (i.e. desired) value for the product variable of interest.

In this analysis the MSE column is calculated as follows:

$$MSE = \frac{\sum(OfficialTFR - EstimatedTFR)^2}{9} \qquad (8)$$

and the next column is calculated as suggested by formula (9):

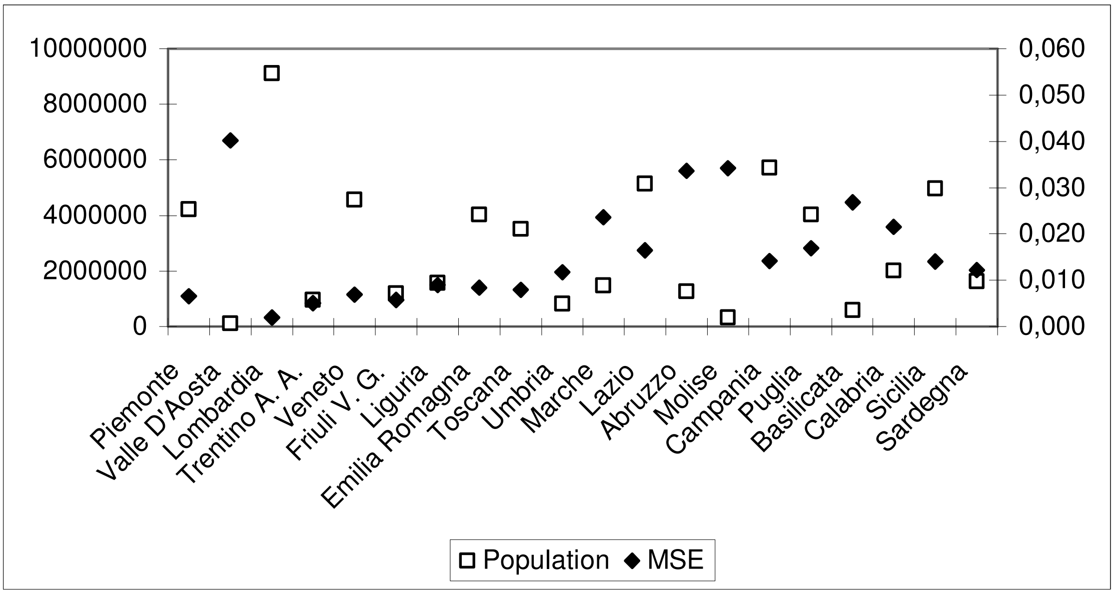
$$MSE_{adjusted} = \frac{\sum(OfficialTFR - TFR_{adjusted})^2}{9} \qquad (9).$$

Clearly, formulas 8 and 9 are an adaptation of formula 7 to the particular case that we are analysing.

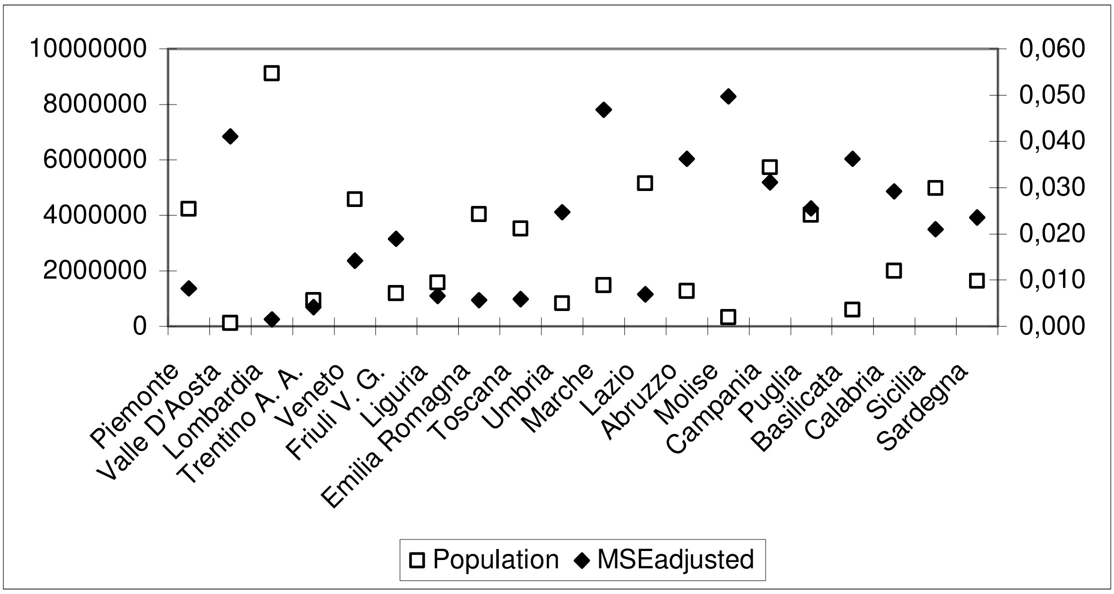
The expectation was an opposite relation between regional dimension (in terms of population) and mean square error. This relation can be seen graphically. However, as for the Wilcoxon Test, we have to acknowledge an imperfection: for the regions of Southern Italy it doesn't work in a good way whether we consider the MSE or wether we take the $MSE_{adjusted}$ into account.

Graphs 4 and 5 illustrate this relation.

Graph 4. *A graphical comparison between the size of Italian regions in terms of population and respective Mean Square Error, calculated with Official TFR and Estimated TFR.*



Graph 5. *A graphical comparison between the size of Italian regions in terms of population and respective Mean Square Error Adjusted, calculated with Official TFR and $TFR_{adjusted}$.*



5. CONCLUSIONS

As part 1 described, over the last years Italian fertility level has been one in the lowest in the world. Italian fertility model has two main characteristics [IRP, 2002]: the collapse of the inclination to have more than 2 children and the increase of the postponement of procreation.

The huge statistical lack of Italian data is given to an absence of sources that provide contemporary data on population (with socio-economic characteristics) and number of children.

Our work is based on the *ILFS* and the use of the *OCM*, that allows to derive children on the basis of parents' characteristics and population on the basis of the same characteristics[21]. In parts 2 and 3, we described the *OCM* and its application in the current analysis.

To test if *ILFS* gives us useful data, we provide a quality assessment for the reconstructed fertility dataset (part 4). We first calculated the TFR by using these data and then we compared the output with the existing data, provided by ISTAT and Eurostat.

The fertility life of every woman in the sample has been reconstructed by following the diagonal lines of the Lexis diagram from the top to the bottom.

Mothers are divided into three groups: head of the family or partner of the head of the family, daughter of the head of the family and parents of the head of the family or of his/her partner. Despite of the absence of children with the characteristics required by the third group, the mothers of this type are included in the denominator of TFR because they are women in fertile age.

The whole number of children is divided between "*linked children*" and "*not-linked children*", to distinguish those who can be matched with a mother and those who can't. If the difference of age[22] between a mother and her linked son isn't included in the interval 14-45, we consider this child in the group of *not-linked children*.

[21] Maria Castiglioni [2004] explains that the use of this method postulates household's composition and the kind of existent relationships.
[22] Age is calculated in century months.

The same process has been applied to calculate Italian TFR and the TFR of every single Italian region.

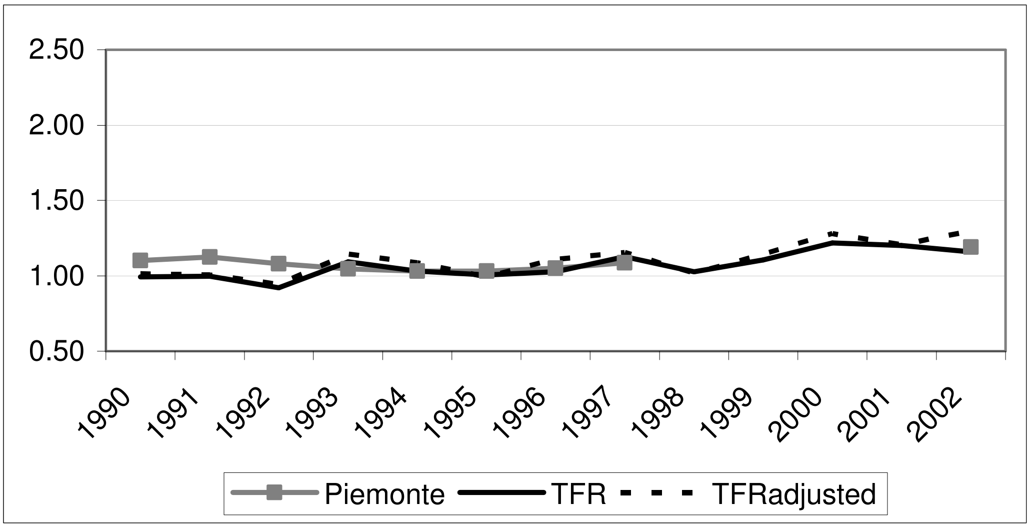
The evaluation of data quality is given by a comparison between the Estimated TFRs (both adjusted and not-adjusted) with Official data produced by ISTAT at regional level and by ISTAT and Eurostat at national level. We also provided a graphical confrontation. The introduction of the Wilcoxon Test suggests how good is the estimate of the TFR calculated by *ILFS* data.

The use of an alternative source, the *ILFS*, to bridge the gap of demographic Italian data and the use of the *OCM* to match children with their respective interviewed mothers, appears to give us usable results. We still acknowledge that our method's goodness is better at a national level compared with its applicability to a single region. There are also better results by looking at a period of time longer than a single year. Therefore, a punctual estimate in a single region and/or a single year is more risky than an observation of Italy over the period 1990 – 2002.

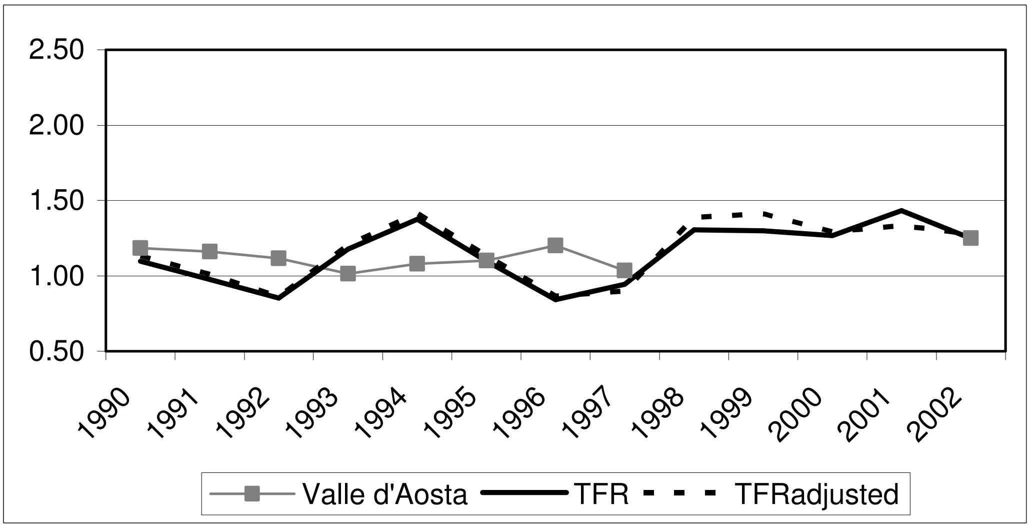
APPENDIX 1

This Appendix offers a graphical comparison, region by region, between Official TFR (by ISTAT[23]), estimated TFR (Total Fertility Rate estimated using data by Labour Force survey) and estimated $TFR_{adjusted}$.

Graph A1.

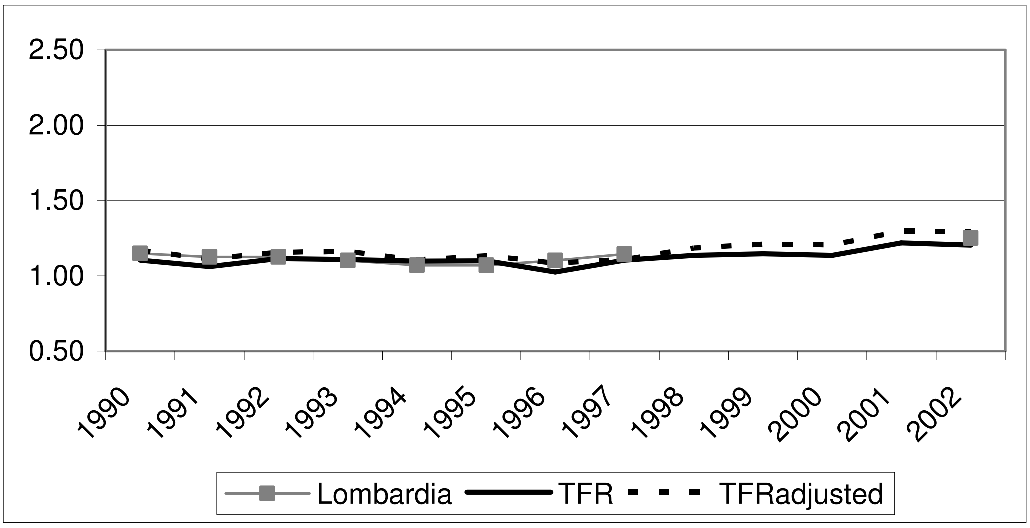


Graph A2.

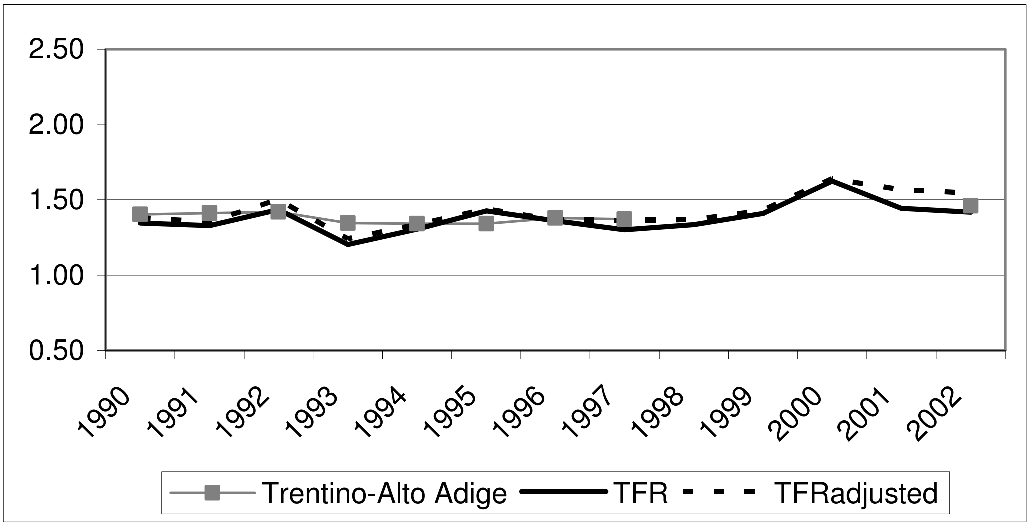


[23] *Source*: http://www.osservatorionazionalefamiglie.it/demografici.htm?8, web site of Osservatorio Nazionale delle Famiglie.

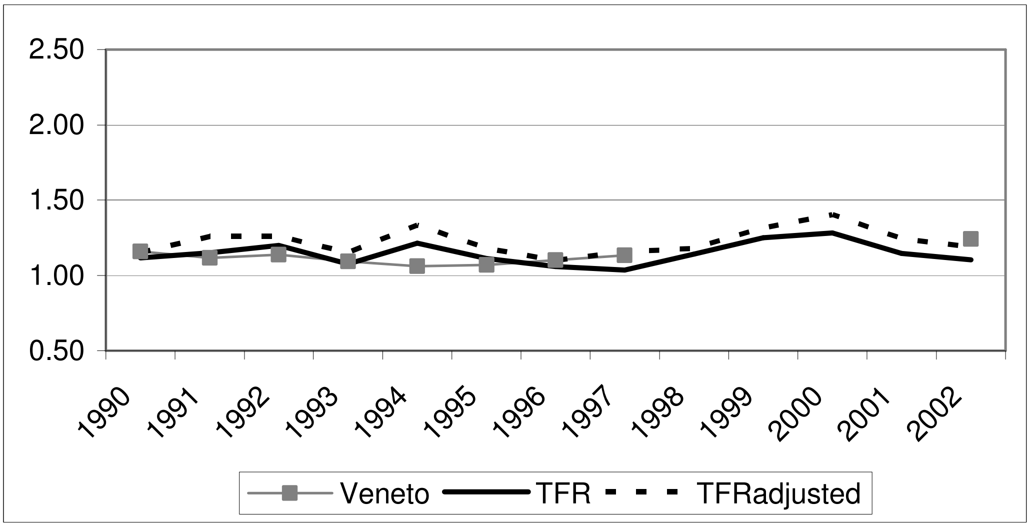
Graph A3.



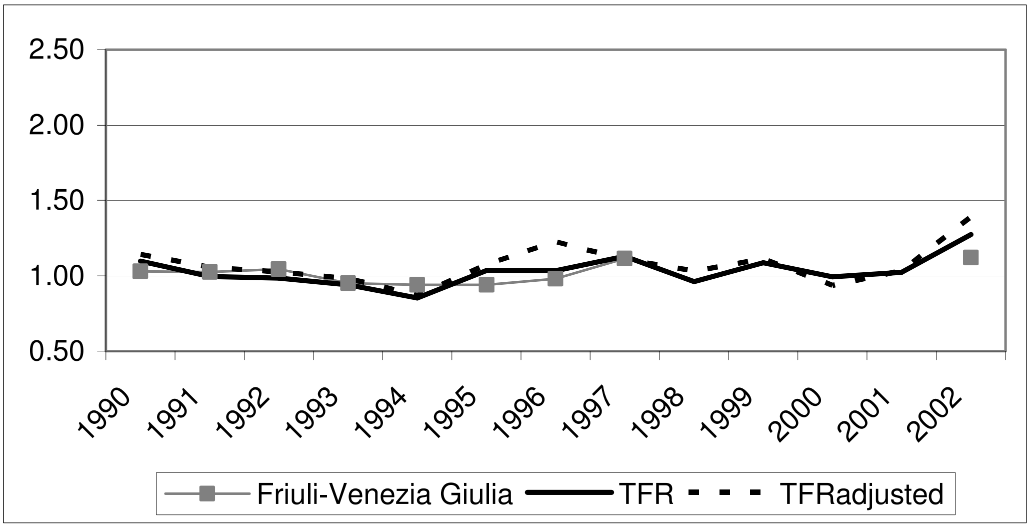
Graph A4.



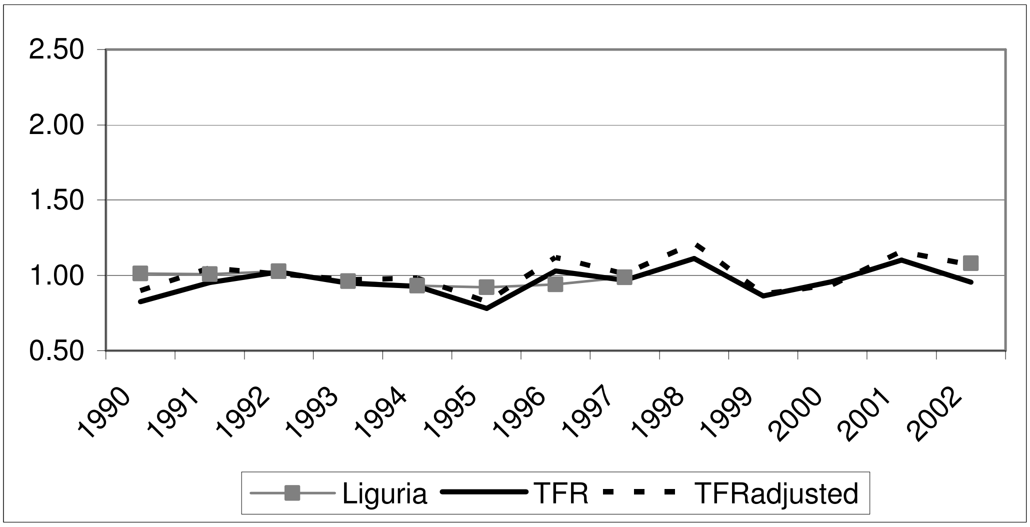
Graph A5.



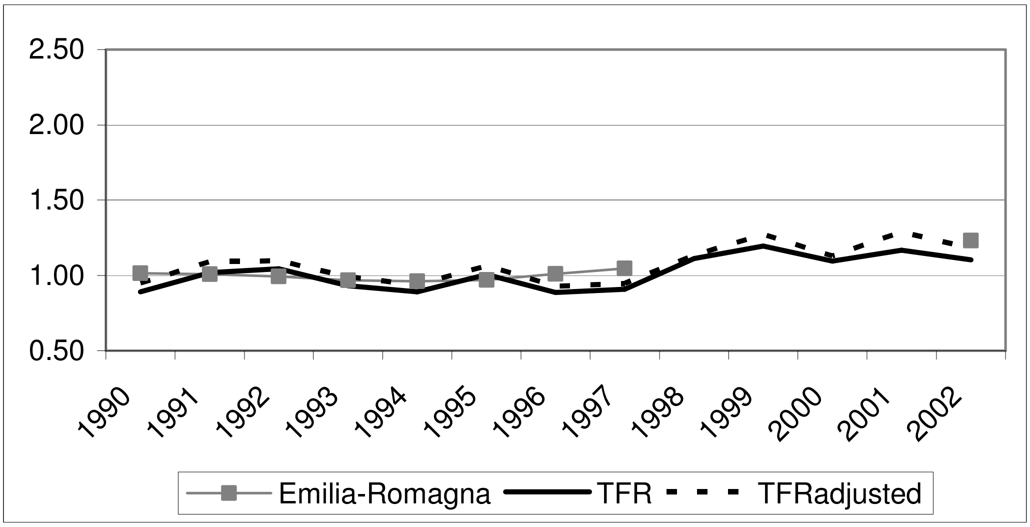
Graph A6.



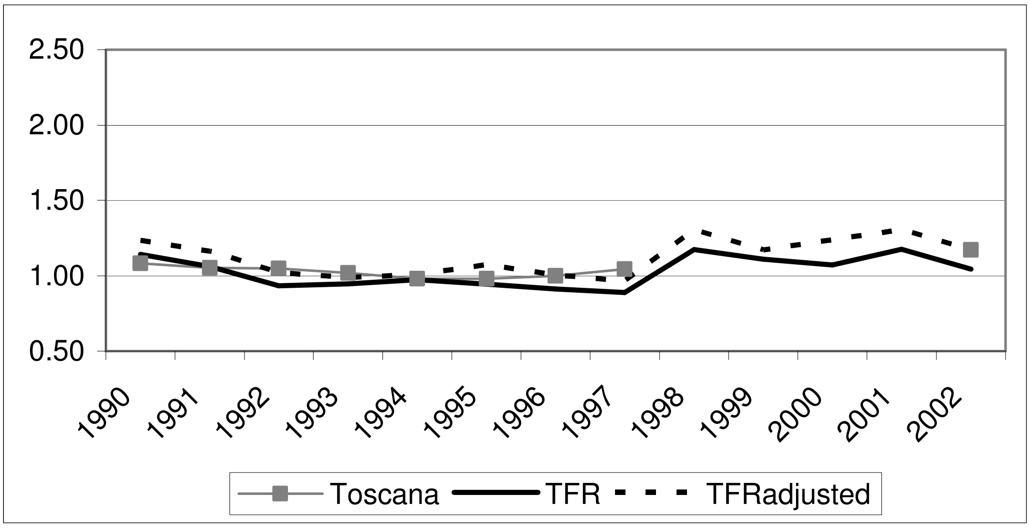
Graph A7.



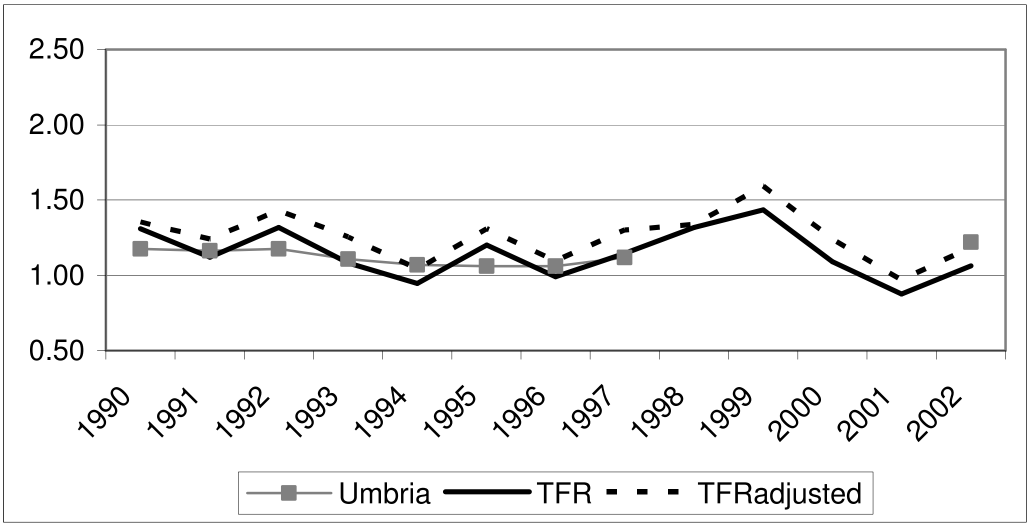
Graph A8.



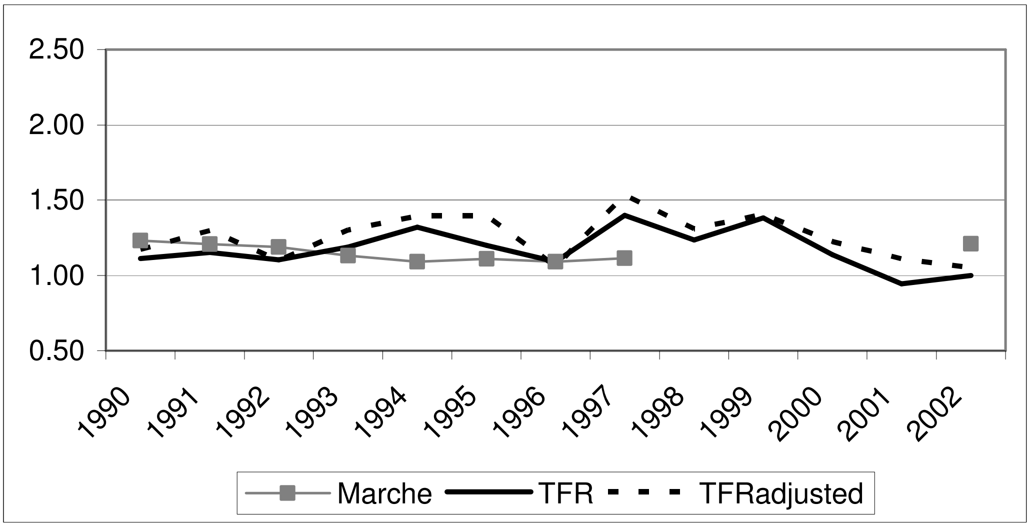
Graph A9.



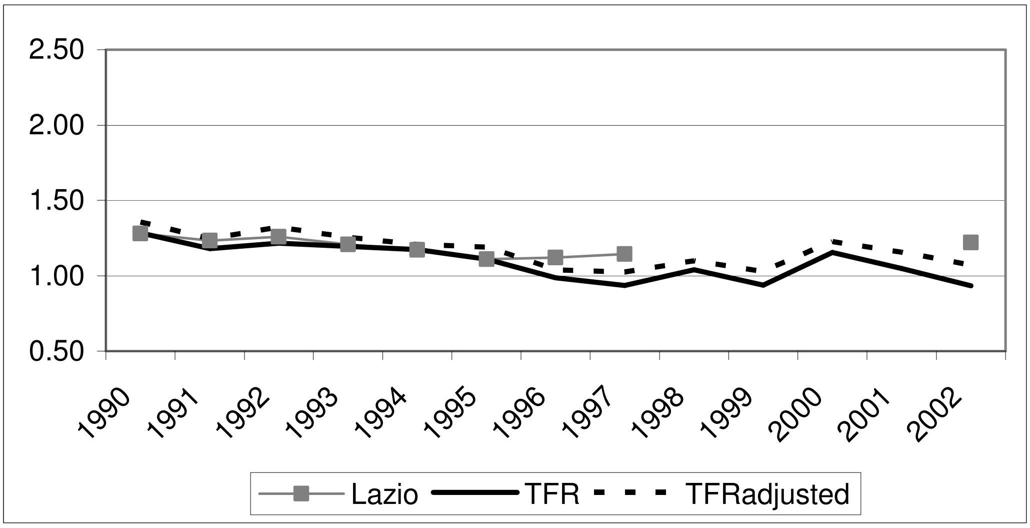
Graph A10.



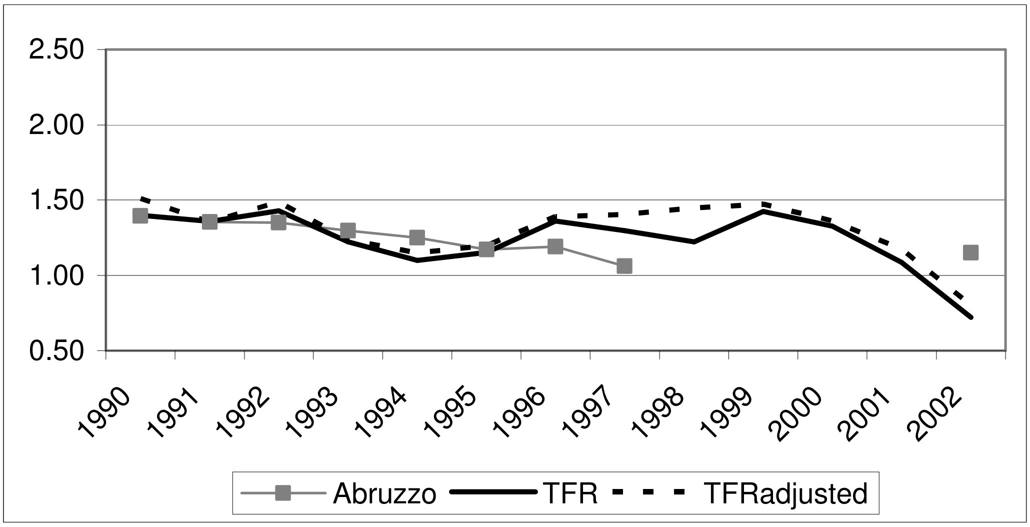
Graph A11.



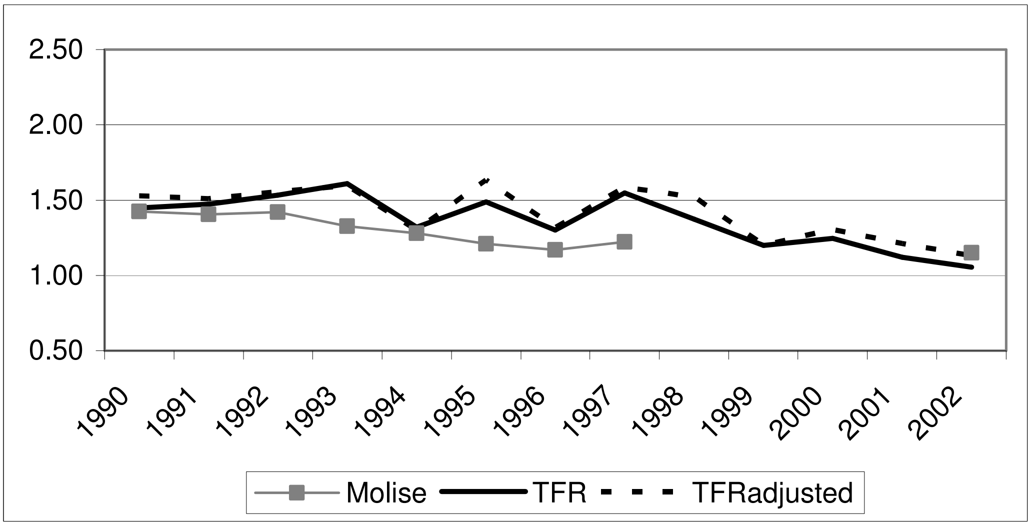
Graph A12.



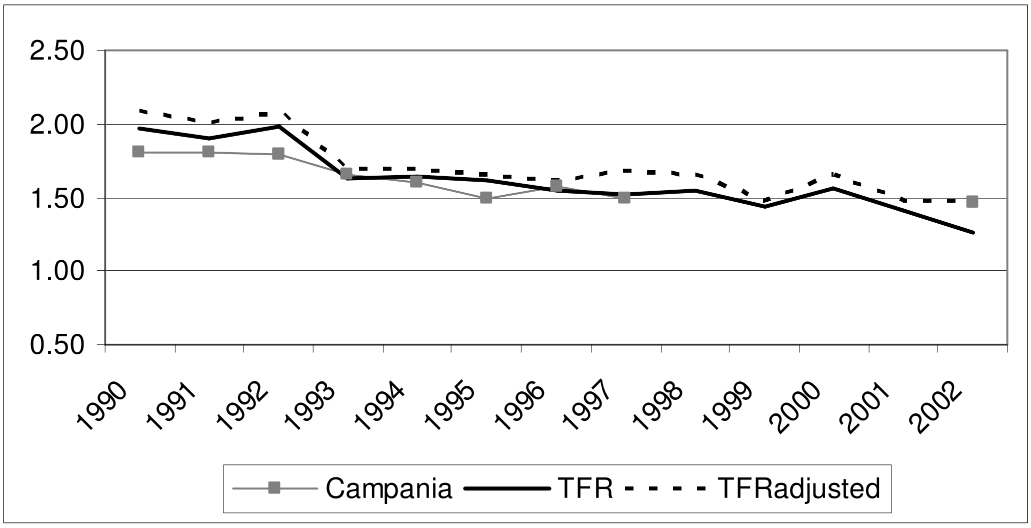
Graph A13.



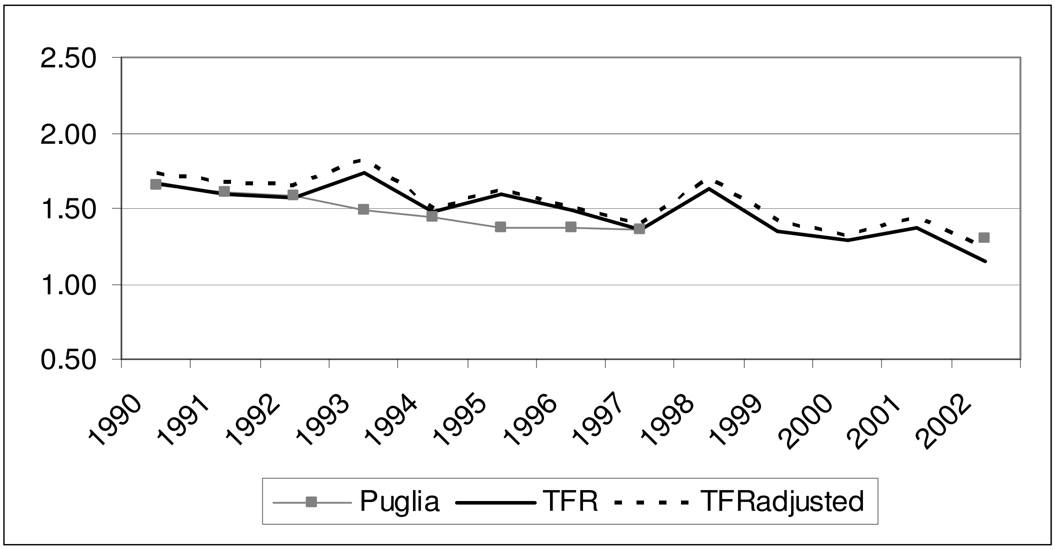
Graph A14.



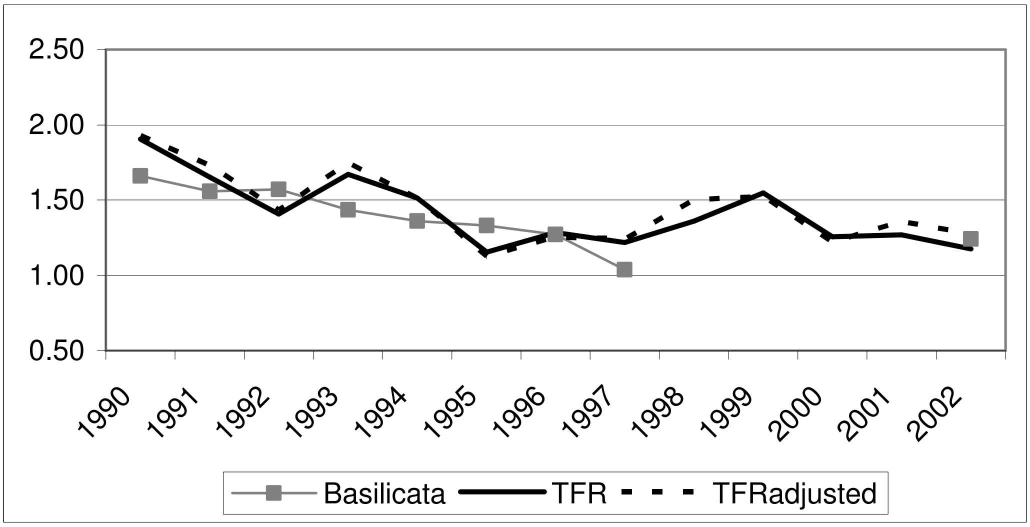
Graph A15.



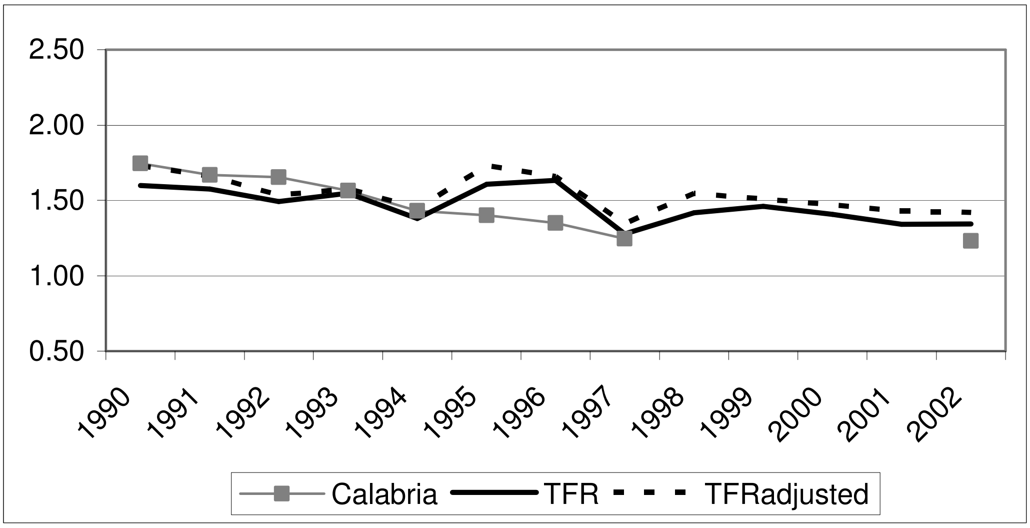
Graph A16.



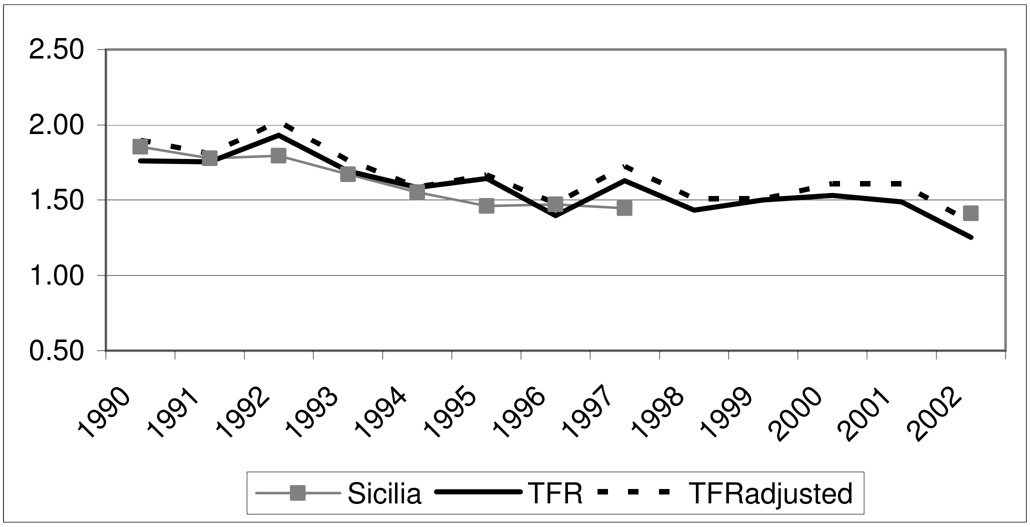
Graph A17.



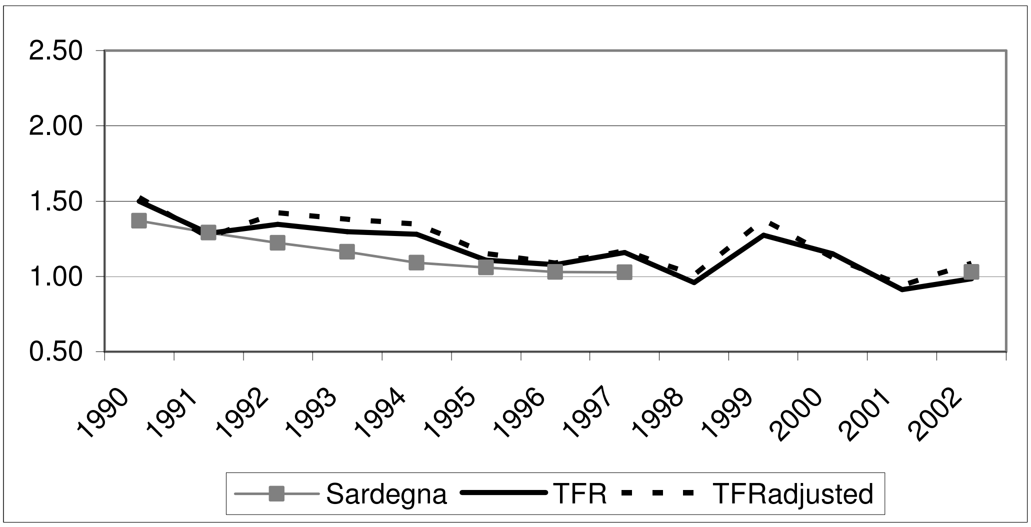
Graph A18.



Graph A19.



Graph A20.



REFERENCES

Billari F. C. 2004. Choices, Opportunities and Constraints of Partnership, Childbearing and Parenting: the Patterns in the 1990s. *Background Paper for the session "International migration: promoting management and integration", European Population Forum 2004*.

Castiglioni M. 2004. Dati necessari per indirizzare, monitorare e valutare l'impatto delle politiche per le nascite e le famiglie: la situazione dell'Italia di oggi e del prossimo futuro.

Cho L. J. 1968. Income and Differentials in Current Fertility. *Demography* 5: 198 – 211.

Dalla Zuanna G. 1989. Stima di dati su fecondità ed abortività in piccole aree. *Statistica* 1: 89 – 108.

Delgado Pérez M., and M. Livi-Bacci. 1992. Fertility in Italy and Spain: The Lowest in the World. *Family Planning Perspectives* 24: 162 – 167+171.

Desplanques G. 1994. Measuring Fertility Differentials from Census Information Alone. *Population: an English Selection* 6: 23 – 33.

Grabill, W. H., and L. J. Cho. 1965. Methodology for the Measurement of Current Fertility from Population Data in Young Children. *Demography* 2: 50 – 73.

Istituto di Ricerche sulla Popolazione. 2002. Rapporto sulla situazione demografica italiana.

Kohler H-P., F. C. Billari, and J. A. Ortega. 2002. The Emergence of Lowest-Low Fertility in Europe during the 1990s. *Population and Development Review* 28: 641 – 680.

Luther N. Y., and L. J. Cho. 1988. Reconstruction of Birth Histories from Census and Household Survey Data. *Population Studies* 42: 451 – 472.

Maffeini W., and F. Rossi. 1984. Stima di fecondità per età con dati italiani di censimento: problemi e prospettive. *Statistica* 3: 423 – 441.

Retherford R. D., and L. J. Cho. 1978. Age-Parity-Specific Birth Rates and Birth Probabilities from Census or Survey Data on Own Children. *Population Studies* 32: 567 – 581.

Retherford R.D., and N. G. Bennett. 1977. Sampling Variability of Own-Children Fertility Estimates. *Demography* 14: 571 – 580.

Rindfuss R. R. 1976. Annual fertility Rates from Census Data on Own Children: Comparisons with Vital Statistics Data for the United States. *Demography* 13: 235 – 249.

Saraceno C. 2004. Tra Vulnerabilità e Solidarietà Lunghe: Tendenze Attuali nelle Relazioni Familiari. *Versione Preliminare*, Milano.

Sobczak M., G. Swicegood, and C. Schmertmann. 2005. Constructing Fertility Measures with the Own-Children Technique: Old Methods for New Data. Paper presented at the *Annual Meeting of the Population Association of America*, Philadelphia, PA.

Wood C. H., and F. D. Bean. 1977. Offspring Gender and Family Size: Limitations of Own Children Data. *Journal of Marriage and the Family* 39: 445.

ONLINE SOURCES

- www.demo.istat.it
- http://epp.eurostat.cec.eu.int
- http://www.osservatorionazionalefamiglie.it/demografici.htm?8.