Reconstructing Brazilian birth histories based on census data: a comparative analysis of two methodologies

Adriana Miranda-Ribeiro ♣

Eduardo L.G. Rios-Neto ♦

I- Introduction

In Brazil, vital statistics are not reliable due to coverage problems both in birth and death registration. In the absence of good vital statistics, the most reliable sources for the calculation of fertility measurements are household surveys and demographic censuses. Most of the censuses and surveys (except for the DHS surveys, that do not have large sample size) are well fit to the calculation of period total fertility rate but they usually lack birth histories for the calculation of more refined fertility measurements (such as parity age total fertility rate, tempo adjusted rates, etc).

Birth histories can be reconstructed using demographic censuses with large sample size, information on parity, age, the number of children by age and living in a household in which they can be allocated to their mothers. Birth histories also enable the calculation of birth intensities, tempo, and quantum period measurements of fertility. The standard measurement of fertility based on household surveys and censuses that lack birth histories are based on incidence rates, while birth histories enable the calculation of birth intensities.

This paper will review three methodologies that allow the measurement of fertility for a 15 years period before the survey and will compare two birth reconstruction methodologies to be applied to demographic censuses.

The first methodology, developed by Grabill and Cho (1965) is the Own Children Method (OCM), a reverse-survival (or reverse projection) technique designed to calculate Total Fertility Rates (TFR) for a 15 year period before the survey. This methodology was very popular for the calculation of TFRs based on printed cross tabulations of data. It is revised here as a foundation for the development of micro level data techniques to complete incomplete birth histories, and it will be also used to compare the results in the aggregate level (TFR series). On one hand, the OCM is not capable of generating birth

♣ PhD Candidate in Demography, Cedeplar, Federal University of Minas Gerais, Brazil.

♦ Professor, Demography Department and Cedeplar, Federal University of Minas Gerais, Brazil.

intervals and birth histories, because it was developed prior to the widespread use of micro data. On the other hand, the OCM provides an accurate estimation of aggregated TFR histories, the aggregated TFR generated by the other two methodologies will be compared with the OCM ones.

The second methodology (henceforth BHP standing for birth history probabilistic), developed by Luther and Cho (1996), is an extension of the OCM to micro level data. It reconstructs the birth histories of interviewed women based on the allocation of mother and child. The reconstruction of the birth history is made by a probabilistic process, which attributes to each woman the probability of having an omitted child, based on her cohort fertility schedule.

The third methodology (henceforth BHM standing for birth history matching) also consists on the reconstruction of the birth history of the interviewed women, but it uses a matching process of women with complete and incomplete birth histories to input the age of omitted children in the women with of incomplete birth histories. Computational algorithms for probabilistic matching will allow the matching of each woman with incomplete birth history with the most similar women with complete birth history in the data set.

The main purpose of this paper is to apply the two methodologies of birth history reconstruction – BHP and BHM – to the Brazilian data, in order to compare the results and to test the efficiency of the matching algorithm. Other purposes of the paper are:

(i) to apply the OCM to the Brazilian data;

(ii) to apply the Kohler & Ortega method to disentangle tempo, parity, and quantum effects of fertility during the Brazilian fertility decline.

The two methodologies will be applied to the last two editions of the Brazilian Demographic Censuses: 1991 and 2000. Then, a fertility time series from 1977 to 2000 will be built.

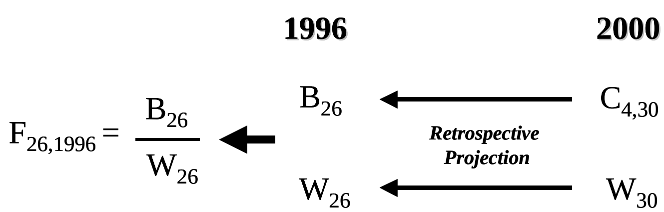
II- METHODOLOGY

II.1 The Own Children Method (OCM)

The OCM is a methodology that allows the estimation of the total fertility rates and age specific fertility rates for a 15 year period before the survey. [1] Basically, the method consists of backward projection of children's cohorts – allocated according to the mother's and children's ages in the date of the survey. The backward projection generates the number of births in every year, previous to the survey, according to the mother's age at birth. The number of women by age in the previous years is obtained by backward projection of the enumerated women. The relationship between the number of births and the number of women allows the calculation of both total and age specific fertility rates every year.

Figure 1 depicts the OCM reasoning. Suppose a survey in 2000, in which: $C_{4,30}$ denotes the number of 4 years old children, matched to women aged 30 at the time of survey; W_{30} denotes the number of women aged 30 enumerated in the survey. Four years ago, $C_{4,30}$ were births to women aged 26. Then, B_{26} is the number of births to women aged 26 in 1996, W_{26}; $F_{26,1996}$ is the age specific fertility rate for women aged 26 in 1996.

Figure 1: Reverse Survival diagram for own-children analysis



Source: Adapted from CHO, RETHERFORD & CHOE, 1986 (p.2)

According to Cho, Retherford & Choe (1986), there are advantages and limitations in the application of OCM, in relation to the use of other available data which allow similar measurements.

The advantages are:

(i) Its usefulness in places where the vital registration is deficient, although in these places censuses may also be deficient, they are, in general, more accurate. In addition, Besides that, if the omission in the censuses is due to coverage of households, then the relationship child/woman should be little

[1] "The estimates are not usually computed further back than fifteen years because births must then be based on children aged 15 or more at the enumeration, a large proportion of whom do not reside in the same household as their mother and hence cannot be matched." (CHO, RETHERFORD & CHOE, 1986, p.2)

influenced by the undercount and the OCM rates maintain the same trend of the real rates;

(ii) The surveys generally used in the application of OCM allow the tabulation of the fertility information by a series of attributes, which makes it possible a wide range of studies on fertility differentials;

(iii) The sample size is much larger than those covered by surveys with birth histories, this fact warrants sample representativness even at the more disaggregated levels;

(iv) The absence of age truncation at the initial matching of children to mothers, which is done for women up to 64 years old at the time of enumeration, allows the estimation of period fertility rates for a 15 years period.

The limitations are:

(i) Both age pattern of fertility and fertility trends may be distorted by age misreporting, what can be diminished by the aggregation of estimates over several calendar years;

(ii) The estimates may be biased due to migration, if they are not a small proportion of the population and compose a select group, in terms of fertility behavior;

(iii) The OCM does not yield birth interval statistics, once it is not based on the reconstruction of individual level birth histories[2];

(iv) The results may be biased by the matching errors and misallocation of non-own children.

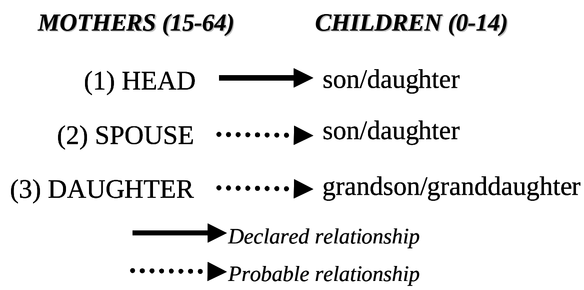
The first step of the OCM is the allocation of children (0-14) to their mothers (15-64). In the 2000 Brazilian Census, the allocation is based on the information of children's and women's relationship to the family head. In the 1991 Brazilian Census, the allocation is made easier by the availability of a variable that identifies the child's mother when both live in the same household.

[2] This is the main difference between the OCM and the other two methods revised here.

Allocation of children to their mothers based on the relationship to family head

The allocation procedure based on the relationship to the family head depends on the information available, and it may differ, from one survey to another. In the 2000 census, the variable 'relationship to family head' allows the identification of: (i) head; (ii) spouse; (iii) son/daughter/stepson/stepdaughter; (iv) grandson/ granddaughter; (v) mother/mother in-law/father/father in-law; (vi) brother/sister; (vii) other relatives; (viii) others. Using those categories, it is possible to identify the relationship mother/child between: (1) woman who is the head of family and those declared as son/daughter [3] of the head of family; (2) woman who is spouse of the head and those declared as son/daughter [4] of the head of family; (3) woman who is daughter of head and those declared as grandson/granddaughter of the head of family. These relationships are shown on Figure 2.

Figure 2: Possible relationships to identify the relationship mother/child within a family



Own elaboration

Contrary to what happens in the first case, in the second and third cases the relationship mother/child is presumed and some consistency checks must be made. The checks must guarantee that: (i) ages of mother and children are compatible (to the age at the beginning of the reproductive period and to the age of the last born child); (ii) the number of allocated children is compatible to the number of children ever born and the number of surviving children. Inconsistent cases are removed and the children are then denominated 'non-own-children'. The same denomination is given to the children that, for any reason, were not allocated to any woman.

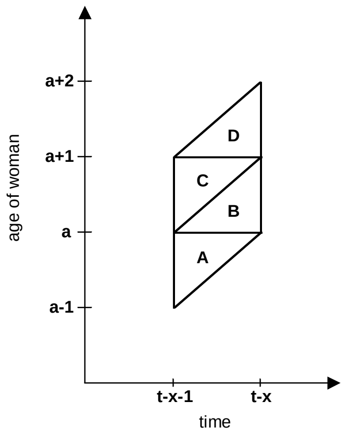
After the allocation of children to their mothers, a cross tabulation indicates the number of own-children, by age, allocated to mothers, by age. The non-own children classified

[3] This category does not differentiate legitimate, adopted or step children, what may result in biases.
[4] Idem.

by age are, then, proportionally distributed, taking into account the age of the women's own-children.

The calculation of central age-specific birth rates (corresponding to the square BC of Figure 3), for a census enumeration at time t, are done by Equations 1 to 3:

Figure 3: Lexix Diagram for visualizing the calculation of F_a(t-x-1)



Source: CHO, RETHERFORD & CHOE, 1986 (p.11)

$$B_a(t-x-1) = C_{x,a+x+1} * U^{c}_{x,a+x+1} * V_x * r_{0\leftarrow x} \quad \textbf{(1)}$$

$$W_a(t-x-1) = W_{a+x+1} * U^{w}_{x,a+x+1} * R^{f}_{a\leftarrow a+x+1} \quad \textbf{(2)}$$

$$F_a(t-x-1) = \frac{[B_{a-t}(t-x-1) + B_a(t-x-1)]}{[W_a(t-x-1) + W_a(t-x)]} \quad \textbf{(3)}$$

In equations *1 to 3:* $B_a(t\text{-}x\text{-}1)$ denontes to births on parallelogram CD; $B_{a\text{-}t}(t\text{-}x\text{-}1)$ corresponds to births on parallelogram AB; $W_a(t\text{-}x\text{-}1)$ denotes the women on left edge of square BC; $W_a(t\text{-}x)$ denotes tth women on right edge of square BC; $r_{0\leftarrow x}$ denotes the children's reverse-survival factor from age group *x,x+1* to exact age 0; $R^f_{a\leftarrow a+x+1}$ denotes the women's reverse-survival factor from age *a+x+1* to exact age *a*;[5] $U^c_{x,a+x+1}$ and $U^w_{x,a+x+1}$ denote the adjustment factors for underenumeration and age misreporting of children and women;[6] *Vx* denotes the adjustment factor for non-own children aged *x, x+1*.

Births on square BC are obtained by adding half of the births in AB to half of births in CD. The midyear population in BC is likewise obtained by averaging the population at the beginning and end of the year, as represented by the left and right edges of the square BC. The symbol *t* denotes the time of census and *a* denotes women's age and *x* denotes the child's age at time *t*.

II.2- Reconstruction of birth histories based on census data

The OCM just described was helpful to show the relevance of a dataset that has information on the mothers and their own children. This methodology was developed in a time that micro data use was not so widespread as today. Furthermore, the OCM dealt with the problem of omitted children at the aggregate level, making it no well suited to build birth histories with birth intervals. The two methodologies discussed below will allocate the children to their mothers at the micro level and following the logic developed in the OCM. Once children are allocated to their mothers, births histories are constructed, by transforming the variable 'age of children' in 'year of birth'.

Although the definition of completeness differs between the two methodologies, both intend to turn into complete the partial birth histories of the enumerated women, by assigning an age to omitted children. [7] Migration and mortality are assumed to be not selective by fertility and eventual misreport problems are not selective by the important characteristics. These are assumptions that should hold in any birth history.

---

[5] Considering constant mortality over the estimation period, $R_{a\leftarrow b} = L_a / L_b$ and $r_{0\leftarrow x} = l_0 / {}_1L_x$. Mortality functions were taken form Carvalho (1974, 1978), Carvalho & Pinheiro (1986) e Cedeplar (1999).

[6] In the case of Brazilian censuses, no post enumeration survey is available. Thus $U^c_{x,a+x+1}$ and $U^w_{x,a+x+1}$ were simply set to one.

[7] There are two omission conditions: (i) child is alive at time of census, but does not live in the same household of mother (non-own child); (ii) child is dead by the time of census. For the second condition, the assigned age is the age that the child would have if was alive.

II.2.1- Reconstruction of birth histories based on a probabilistic process (BHP)

The methodology developed by Luther & Cho (1988) is considered as an extension of the OCM, because it is based on the allocation of the children to their respective mothers – as described on the OCM. It is an advance with respect to OCM regarding the treatment of the omitted children and the correction at the micro level.

If a woman has all her children alive and living with her at the time of census, the birth history is complete. Otherwise, the birth history is partial. The procedure to complete partial birth histories is based on a probabilistic process, by which a "partial birth history is used as input, together with supplementary information on number of children ever born and to her and number of her children still surviving. From these data the number of her children ever born who are deceased and the number who are surviving but no longer living in her household can be calculated. The total number of such children is the number of births that must be assigned probabilistically to estimate her complete birth history" (Luther & Cho, 1988, p.452). Given a woman at a particular age, a model fertility curve to the cohort schedule is estimated, which may be derived from period schedules of age-specific fertility rates. For fitting each cohort schedule, the authors suggest to use Luther's fertility model. [8]

$$\tilde{f}(x) = (x-s)^2 * \exp\left[\frac{-2(x-s)}{m}\right], s \le x \le u + 0.5 \quad \textbf{(4)}$$

The parameters *s* and *m* are given by:

$$s = \frac{\{[12{,}9210 + 2{,}04642 * \ln(r - 0{,}535)] * \sqrt{r} + 7{,}20 - 0{,}5\bar{a}\}}{\sqrt{r} - 0{,}5} \quad \textbf{(5)}$$

and $$m = \frac{2(\bar{a} - s)}{3} \quad \textbf{(6)}$$

In Equations 4, 5 and 6, *x* denotes the age of woman and *u* is the lowest age possible for that woman between age at time of census and 46; $\bar{a}$ denotes the mean age of childbearing for the cohort schedule of fertility rates; *r* denotes the ratio of cohort age specific fertility rates at ages 20-24 relative to ages 15-19; *s* denotes the cohort's age at fertility initiation. The determination of *m* and *s* for each cohort leads to the determination of a particular form of the fitted fertility model curve.

The next step is the derivation of the probabilities to be used in the assignment of ages at time of census to the deceased and 'non-own' children, for each woman. The

[8] Which employs a particular form of the Pearson Type III.

assignment probabilities are derived by using the fertility model curve that fits the cohort schedule of age-specific rates corresponding to the woman's age at time of census. The curve predicts the relative probability of the event *B(j)*, that a child was born to this woman *j* completed years before the census. With *g* denoting the age of this woman at the time of census, this probability is found for all ages *j* of a child for which *g-j*, the woman's average age at birth, ranges from *s* up to the smallest value between *g* and 46.

The probability of having a birth *j* years from the census is given by the equation:

$$P[B(j)] = H(g - j + 1/2) - H(g - j - 1/2) \quad \textbf{(7)},$$

where $$H(x) = 0{,}5m\left\{0{,}5m^2 - \exp\left[\frac{-2(x-s)}{m}\right]\left[(x-s)^2 + m(x-s) + 0{,}5m^2\right]\right\} \quad \textbf{(8)}$$

Once finding *P[B(j)]*, the probabilities to be used in the assignment process can be derived. The probabilities are different according to the omission condition. For a deceased child, suppose that *D* is the event that a child of any age dies before the time of the census or survey. Then, the relative probability of being born *j* completed years and dying before the census is:

$$P[B(j) \cap D] = P[B(j)]P[D \mid B(j)] \quad \textbf{(9)}$$

where *P[D|B(j)]* denotes the conditional probability that a child born *j* completed years before the census has died before the time of census, which can be taken from life tables.[9]

For a non-own child, suppose that *N* is the event that a child of any age is non-own (surviving and unmatched) at the time of census. Equation 7 gives the relative probability of being born j completed years before the census and being a non-own child at the time of census.

$$P[B(j) \cap N] = P[B(j) \cap \tilde{D} \cap N] \quad \text{or}$$

$$P[B(j) \cap N] = P[B(j)](1 - P[D \mid B(j)])P[N \mid \tilde{D} \cap B(j)] \quad \textbf{(10)}$$

where $\tilde{D}$ is the complement of *D*, i.e., the event that a child is surviving at the time of census. Thus $P[N \mid \tilde{D} \cap B(j)]$ is the conditional probability that a surviving child aged j completed years at the time of census is a non-own child and is obtained from the

[9] For details, see Luther & Cho (1988), p.459.

census by dividing the total number of non-own children aged *j* by the total number of children (no-own plus own) aged *j*. [10]

Equations 9 and 10 give the general form of the probability functions for each cohort. For each woman, the probabilities must be recalculated, taking into account the own-children's age. For an own-child aged *j*, the probability of having another child aged [g-j-5/4, g-j+5/4] is zero. Once the probabilities are recalculated the assignment procedure starts. For each woman, a random process using a computer-generated random numbers is applied to each of her omitted children, using the relative probabilities given by Equation 9 if the child is deceased and by Equation 10 if the child is 'non-own'. This will determine an age *j* at the time of census or survey. This is done randomly once at a time for each child. After the first omitted child has an assigned age, the probabilities are recalculated and, then, the procedure is repeated, until all woman's omitted children have an assigned age. [11]

II.2.2- Reconstruction of birth histories based on a matching process (BHM)

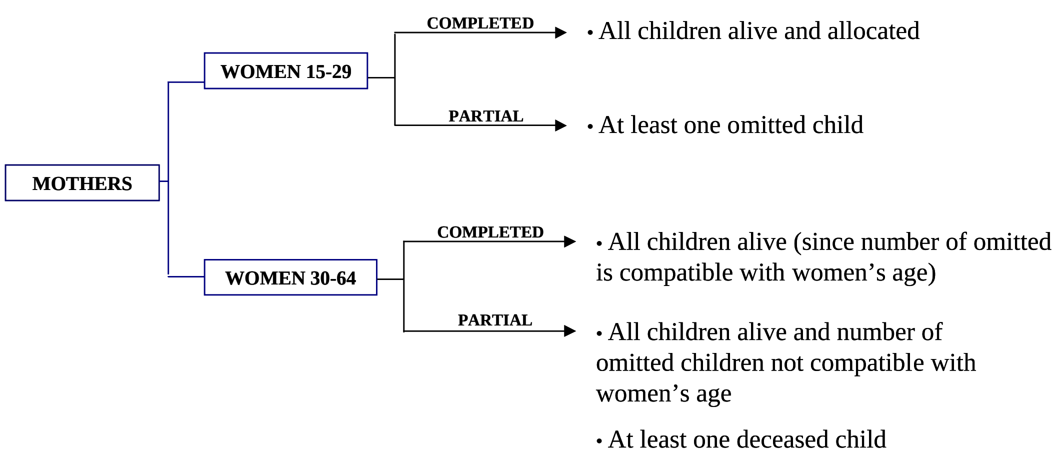
The BMP methodology is being developed at Cedeplar/UFMG [12] and also aims to complete partial birth histories of women aged 15-64. Generally, the method consists in comparing each partial birth history for women with incomplete birth histories in the data set with the women with completed birth histories, in order to find the most similar completed birth history, which then will be used to input the lack information on the partial birth history. Using computational algorithms, a series of reproductive variables are compared, in order to find the 'best match'. This is a non-parametric method that seems to be easier to be implemented than the BHP method.

As in the previous method reviewed, the allocation procedure developed by the OCM is applied. The age and sex of each own-child determines the child's year of birth and sex of the born child. The reconstruction period is limited to 15 years, since it is assumed that most children aged up to 14 live with their mother (or are enumerated in the same household). This assumption is also important in the definition of completed birth histories, for women aged 30 and more.

[10] For details, see Luther & Cho (1988), p.460.
[11] The authors point out that the reconstruction procedure as described does not allow for deceased or 'non-own' children of multiple births, since the chance of assigning an age identical to a previous one is zero.
[12] In cooperation with José Antonio Ortega (Universidad of Salamanca) and Vania Candida da Silva (Cedeplar/UFMG).

In order to define the birth histories that will compose the completed and partial data sets, the women are divided in two different groups: the first group is comprised by the women aged 15 to 29 and the second group by the women aged 30 to 64. Assuming that the reproductive period begins at age 15, the women of the first group can only have children aged 0 to 14. Thus, their birth histories will be completed if all their children are own-children (i.e. surviving at the time of census and allocated). The definition of completed birth history is different for the second group of women. This group of women can have children aged 15 or more and, following the assumption, they can have surviving children that do not live in their mother's household. Thus, in the case of women in the second group with all children surviving, the birth history is considered as completed if the number of non-own children is compatible with women's age. [13] The assumption there is that the children who were not allocated are aged 15 or more. Partial birth histories are the complement of the completed ones. For women aged 15 to 29, the birth history will be partial if they have, at least, one omitted child (i.e., number of allocated children less than the number of children ever born). For women aged 30 to 64, the birth history will be partial if they have, at least, one deceased child. In both groups of women, those who had children but none were allocated must be included in the partial birth history data set. Figure 4 illustrates the path to build the birth histories.

Figure 4



Source: Own elaboration

[13] This condition is to avoid, for example, that a woman aged 30, mother of three (surviving) children - two of them omitted – is incorporated in the completed birth histories database. Unless those 2 omitted children are twins aged 15, there is no other possibility to consider her birth history as completed. For this, the number of omitted children is considered incompatible to the woman's age and her birth history is considered as a partial one.

In the birth histories data sets – partial and completed – each line is a woman with a vector of selected and constructed variables indicating the age and sex of the allocated children. At this point, the data sets are ready for comparison. The comparison is done by a *linkage software*, in this case 'Reclink', developed by Camargo Jr. and Coeli (2000). [14] Basically, the software compares the two data sets – completed and partial – and classifies, for each partial birth history, a group of completed birth histories by a score, defined by the number of correct matches. The highest score defines the best match (or best pair), which will be the completed birth history to be used to input the missing variables of the partial birth history.

In order to start the comparison procedure, it is necessary to define two types of variables: the *block* variables[15] and the *parameters*. The *block* variables divide the data sets in blocks of logical and homogeneous characteristics. This is important to optimize the comparison process because each partial birth history will be compared to all completed birth histories within the block. It also guarantees that the comparison will occur among homogeneous birth histories, what increases the probability to find the best set of pairs. The *parameters* are the variables used in the comparison among the partial birth history and the completed ones. Suppose a comparison between a pair of birth histories. For each matched *parameter*, an individual score is computed; the sum of the individual scores gives the total score of that pair, which is recognized as a link between the two registers.

Two unexpected situations occurred during the comparison process: (i) in some cases, the highest score appeared more than once; (ii) and some partial birth histories did not 'find' a pair. In the first case a random process was used to select a pair. In the second case, the comparison process was repeated with a different set of *block* variables.[16]

The availability of variables in each Brazilian demographic census determined the *block* variables and *parameters* used in the comparison procedure (Table 1).

[14] CAMARGO Jr., K. R. de e COELI, C. M. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. **Caderno de Saúde Pública**, Jun 2000, vol.16, no.2, p.439-447. Available at: www.scielo.br/scielo.php (Feb 23th, 2006)

[15] In our case, it was important not to use socioeconomic variables in the block, so that we could generate a pure demographic match. Parity, age, and year of last birth were the most important dimensions in the block. Other matching techniques such as propensity scores stress the use of socioeconomic variables.

[16] The variable 'age of mother' was removed from the set of *block* variables.

Table 1

1991 and 2000 Brazilian Demographic Censuses: definition of *block* variables and *parameters* used in *Reclink* subroutine

	***Block* Variables**	***Parameters***
1991 Census	Age of woman Parity Year of last birth	Birth in specific year (15 variables) Sex of child born specific year (15 variables) Still-birth (by sex) Number of children living elsewere (by sex)
2000 Census	Age of woman Parity Year of last birth	Birth in specific year (15 variables) Sex of child born specific year (15 variables) Still-birth

Own elaboration

II.2.3- Comparing BHP and BHM for the Best Reconstruction of Birth Histories

As it was stated previously, the main advance of BHP and BHM with respect to the OCM is the use of micro level birth histories and the construction of birth intervals from census data. The aggregated time series of TFR derived from the OCM is assumed to be quite robust by the literature, and there would be no need of other methods if one did not want to know about the *quantum* and *tempo* of fertility.

A comparison of methods can be performed using several statistical tools and simulations, the comparison between BHP and BHM performed in this paper is rather simple. We take the OCM method as the gold standard for the aggregated generation of a TFR time series. Then, we generate the same series using BHP and BHM. The methodology that best match the series generated by the OCM will be considered the best fitted for future use.

After deciding on the best procedure to reconstruct micro level birth histories, we perform the application of those birth histories to the Brazilian case using indicators generated by the Kohler and Ortega method to be described below.

II.3- Methodologies Applied to birth histories: Kohler and Ortega Method (K-O)

The birth histories data sets allow the application of Kohler & Ortega method, which leads to estimates of tempo, parity, and quantum effects of fertility. Based on birth histories, we obtain the parity and age distribution of women. For every year, we have the number of births by parity and age, and compute the birth intensities. Finally, the computed birth intensities are used to analyze fertility trends. The formulas developed below are based on Ortega & Kohler (2002).

Birth intensities are defined as the number of births of order "i" divided by the number of women of parity "ï-1", as defined in (11):

$$m_c(a) = \frac{B_c(a)}{E_c(a)} \quad \textbf{(11)}$$

(1)

$B_c(a)$ are births of women aged a and class c.

$E_c(a)$ are women aged a and class c (parity "ï-1").

Incidence rates generate the age specific fertility rates:

$$f_c(a) = \frac{B_c(a)}{E\ (a)} \quad \textbf{(12)}$$

$B_c(a)$ are births of women aged *a* and class *c* and $E\ (a)$ are women aged *a*.

The tempo effect ($r_j(a)$) of parity "*j*" at age a is:

$$r_j(a) = \gamma_j + \delta_j(a - \bar{a}_j) \quad \textbf{(13)}$$

γ_j and δ_j are the parameters gamma and delta, calculated interactively following Kohler e Philipov (2001). Gamma is the change in the mean age at childbearing, while delta is the proportional change in the standard deviation and $\bar{a}_j$ é the mean age at the adjusted fertility function. The adjusted functions follow the two formulas below.

For intensities:

$$m_j'(a) = \frac{m_j(a)}{1 - r_j(a)} \qquad \textbf{(14)}$$

For incidences:

$$f_j'(a) = \frac{f_j(a)}{1 - r_j(a)} \qquad \textbf{(15)}$$

TFR e TFR_C (class c) are calculated by the formulas below:

$$TFR_C = \sum_a f_c(a) \quad \textbf{(16)}$$

$$TFR\ = \sum_c TFRc = \sum_a f(a) \qquad \textbf{(17)}$$

A fertility table is built following the formulas below. The age and parity specific probability of birth:

$$q_j(a) = 1 - \exp[-m_j(a)] \qquad \textbf{(18)}$$

The birth probabilities are the basis for the other measures of the table.

$$b_j(a) = D_j(a).q_j(a) \qquad \textbf{(19)}$$

where $D_j(a)$ is the number of women of parity *j* and exact age *a* and $b_j(a)$ is the number of births of parity *j* to the women at age *a*.

An interactive procedure using formulas (19) and (20) allows the calculation of the number of women and births.

$$D_j(a+1) = D_j(a) - b_j(a) + b_{j-1}(a) \qquad \textbf{(20)}$$

For parity *J*, the last parity, including parities *J* and above, the two interactive formulas are:

$$b_J(a) = D_j(a) f_J(a) \qquad \textbf{(21)}$$

$$D_J(a+1) = D_J(a) + b_{J-1}(a) \qquad \textbf{(22)}$$

The number of women at the initial condition of the table (radix), that is to say age "α" and parity "j=0", is N. At the exact age "α" and parity "j>0" the number of women is null.

The mean number of births for women in the synthetic cohort is defined by (23).

$$b_{j1,j2}(a_0, a_1) = \sum_{a=a0}^{a1} \sum_{j=j1}^{j2} b_j(a) \qquad \textbf{(23)}$$

The completed fertility of the fertility table corresponds to (23) and is also called Parity and Age Total Fertility Rate (PATFR) as in Rally and Toulemon (1993). When PATFR is built based on tempo adjusted intensities, it is free from both tempo and parity compositional effects.

$$b_{0,J}(\alpha,\omega)/N = b_{0,J}(\alpha)/N \qquad \textbf{(24)}$$

III- EMPIRICAL RESULTS

The empirical analysis is divided in three parts. First, we compare the BHP and BHM methods with the OCM, in order to decide for the best method. Then, we go on to complete the analysis of TFR trends based on the estimation of the preferred method. Finally, we apply the KO methodology to the birth history constructed by the preferred method in order to understand the trends observed in Brazil.

III.1- Some results of the application of BHP and BHM: comparisons to OCM and traditional TFR

The TFR series resulted from the application of the OCM and the two methodologies of birth history reconstruction – BHP and BHM – to the 2000 Census. The traditional TFR[17] (taken from 1991 and 2000 Censuses) are shown on Figures 5 to 9. The results serve two simultaneous purposes: first, to compare the different conditions leading to fertility decline in Brazil; and second, to compare the data quality of the two methods. Five Brazilian states (one in each of the five Brazilian Regions) were chosen for this exercise. In general the BHM series fits better the OCM series than the BHP one. The latter fits well only in the states that presumably of good quality data, these are also states with lower fertility levels in the beginning of the series.

Figures 5, 6 and 7 show the TFT series for Rio Grande do Sul (South Region), Distrito Federal (Center-West Region) and Espirito Santo (Southwest Region), which are states with good data quality and low fertility levels during all the period. Figures 8 and 9 show the results for Pará (North Region), and Pernambuco (Northwest Region), these are states with worse data quality and higher fertility levels.

For Rio Grande do Sul (Figure 5), BHM and BHP curves are well fitted to the OCM curve and to the traditional TFR during all period. Fertility is around 2,5 in 1986 and around the replacement level during the 90's.

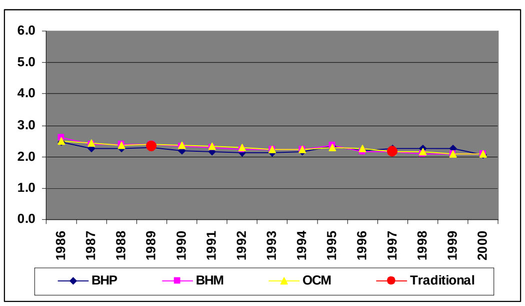
In Figure 6, the results also show low fertility levels during all period. Fertility reaches the replacement level during the 1990's and at the end of the series fertility is below replacement level. The BHM curve is well fitted both to the OCM curve (during all period) and to the traditional TFR, while BHP is well fitted only at the end of the series. Notice that until 1995 the BHP series is only at a slightly lower level.

[17] The traditional TFR were calculated from the 1991 and 2000 censuses, by the P/F Brass technique (BRASS, 1975) and refers to a period of approximately three years previous to the Censuses.

The results for Espirito Santo (Figure 7) show that the BHM curve fits the OCM curve and the traditional TFR during all the period. The BHP curve shows the same fertility decline pattern, but at a slightly lower level during all period.

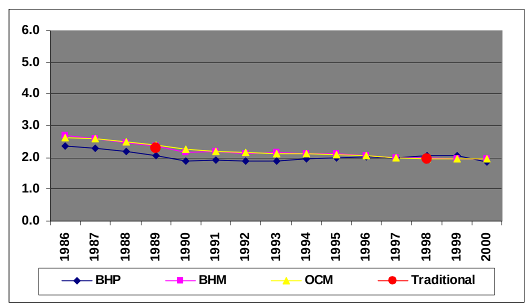
Figures 8 and 9 show that for Pará and Pernambuco the results for the application of the BHP are not good, compared to the results of OCM, BHM and traditional TFR. While the OCM and BHM curves are well fitted (excepted at the very beginning of the series), the BHP curve does not fit. BHP also presents a completely different pattern of fertility decline. At the end of the series, the BHP fertility levels are closer to the BHM and OCM levels, especially in Pernambuco (Figure 9).

Figure 4
Rio Grande do Sul, Brazil: TFR series from the application of OCM, BHP and BHM and traditional TFR



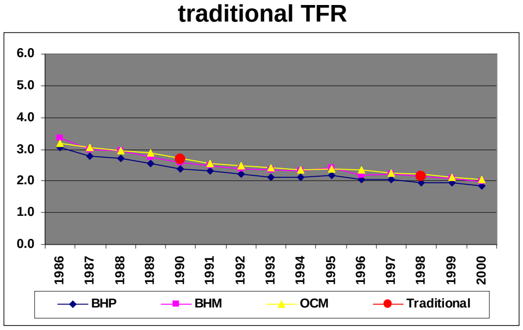
Sources: 1991 and 2000 Brazilian Demographic Censuses

Figure 5
Distrito Federal, Brazil: TFR series from the application of OCM, BHP and BHM and traditional TFR



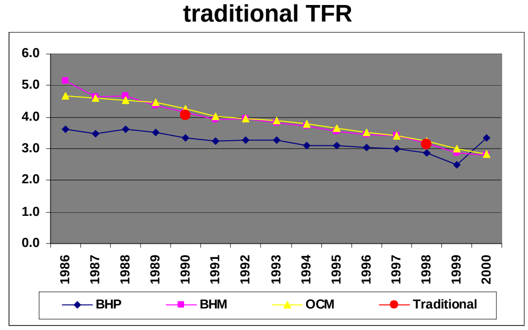
Sources: 1991 and 2000 Brazilian Demographic Censuses

Figure 6
Espírito Santo, Brazil: TFR series from the application of OCM, BHP and BHM and traditional TFR



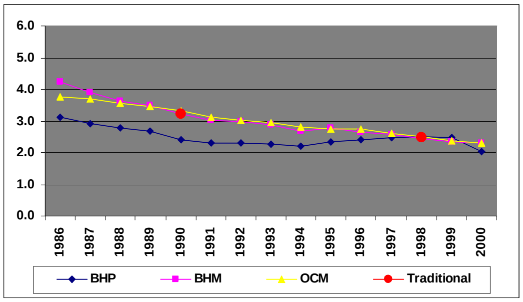
Sources: 1991 and 2000 Brazilian Demographic Censuses

Figure 7
Pará, Brazil: TFR series from the application of OCM, BHP and BHM and traditional TFR



Sources: 1991 and 2000 Brazilian Demographic Censuses

Figure 8
Pernambuco, Brazil: TFR series from the application of OCM, BHP and BHM and traditional TFR



Sources: 1991 and 2000 Brazilian Demographic Censuses

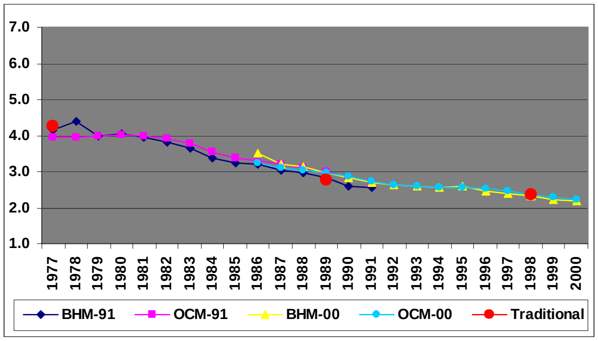
The results just discussed indicated that the BHM is superior to the BHM in this exercise. If we take into consideration that the estimation of birth histories is easier to be done in the BHM than in the BHP, then there is another justification for adopting the former method.

III.2- Results of BHM methodology: the 1977-2000 series

Figures 5 to 9 showed that BHM series were well fitted to the OCM series and traditional TFR, independently of fertility decline pattern and data quality condition. Thus, Figures 10 to 15 present the completed TFR series (from 1977 to 2000) obtained by the application of BHM and OCM to the 1991 and 2000 Censuses data. The traditional TFR for the period (1980, 1991 and 2000 Censuses) are also calculated. The 1991 BHM and OCM gives the 1977-1991 series and the 2000 BHM and OCM gives the 1986-2000 series. Figure 10 presents the series for Brazil as a whole and Figures 11 to 15 present the five Brazilian regions series. The results are satisfactory and the curves are well superposed, showing the efficiency of BHM method.

At the end of 1970's Brazilian TFR was around 4.0 (Figure 10), after 1980, Brazil experienced a steady fertility decline (stronger in the eighties), reaching 2.8 in 1990, going to around 2.2 in 2000. The 2000 series suggests that the fertility decline will continue, but perhaps in a low pace.

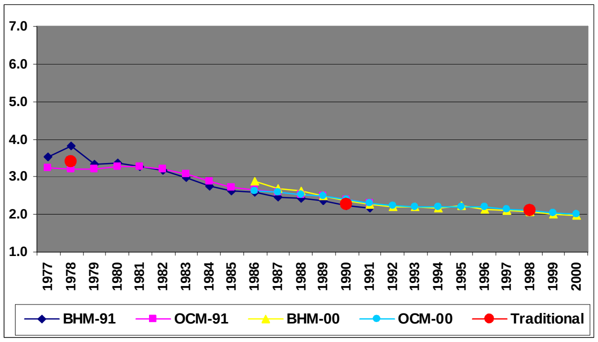
Figure 9
Brazil, 1977-2000: Traditional TFR and TFR series from the application of OCM and BHM



Sources: 1980, 1991 and 2000 Brazilian Demographic Censuses

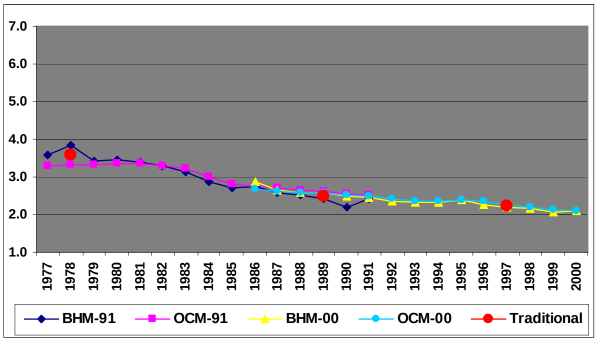
The Southeast Region, as shown in Figure 11, presents a different pattern of fertility decline. At the end of seventies, fertility was stable, around 3.2. It declined slowly since 1982, reaching the replacement level in the mid-nineties. Although almost stable, fertility will possibly fall a little bit after 2000.

Figure 10
Southeast Region, Brazil, 1977-2000: Traditional TFR and TFR series from the application of OCM and BHM



Sources: 1980, 1991 and 2000 Brazilian Demographic Censuses

Figure 11
South Region, Brazil, 1977-2000: Traditional TFR and TFR series from the application of OCM and BHM



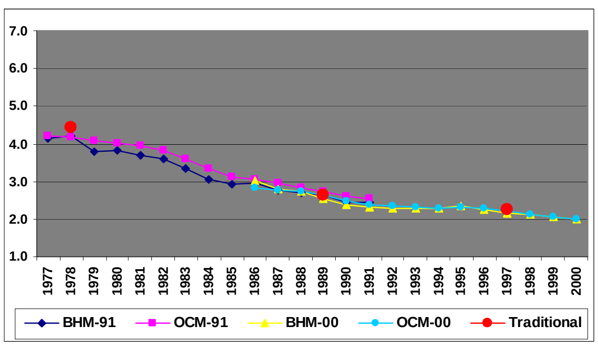
Sources: 1980, 1991 and 2000 Brazilian Demographic Censuses

The South Region, as shown in Figure 12, presents a pattern of fertility decline similar to the one observed in the Southeast Region. Nevertheless, TFR reached a more stable and flat level during the nineties, at a TFR level that was higher (around 2.4) than in the Southeast region. TFR is below replacement level in 2000, and it will probably continue declining.

The TFR trends in the Center-West region is observed In Figure 13. TFR declined during all period, with a stronger pace in the eighties. During the nineties, fertility was almost stable around 2.3. It reached the replacement level in the end of nineties. In 2000, TFR was already below replacement level.

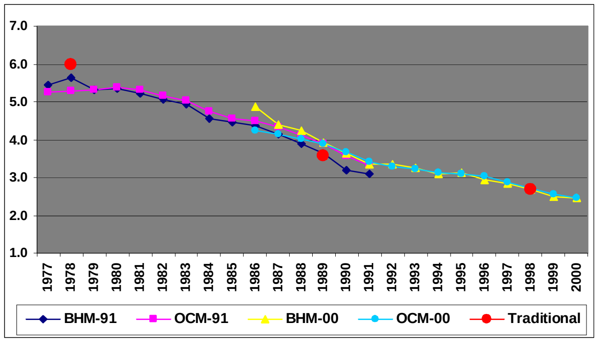
Northeast and North Regions (Figures 14 and 15) show different fertility levels in Brazil. In the Norhteast, TFR was over 5.0 in 1982, reaching 4.0 around 1988, 3.0 around 1996, and 2.5 in 2000. The pattern of decline in the second half of the nineties suggests that fertility will continue falling. In the North Region, TFR presents the highest fertility levels during the period. Its level was around 6.0 at the end of seventies, reaching 5.0 in the middle eighties, 4.0 at the end of eighties and 3.0 at the end of nineties. TFR in 2000 is still near 3.0, the pattern suggests that TFR will continue falling.

Figure 12
Center-West Region, Brazil, 1977-2000: Traditional TFR and TFR series from the application of OCM and BHM



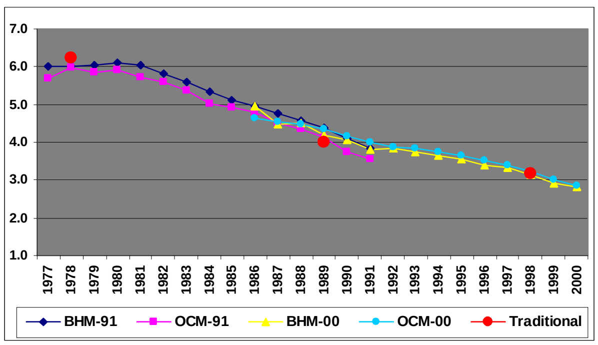
Sources: 1980, 1991 and 2000 Brazilian Demographic Censuses

Figure 13
Northeast Region, Brazil, 1977-2000: Traditional TFR and TFR series from the application of OCM and BHM



Sources: 1980, 1991 and 2000 Brazilian Demographic Censuses

Figure 14
North Region, Brazil, 1977-2000: Traditional TFR and TFR series from the application of OCM and BHM



Sources: 1980, 1991 and 2000 Brazilian Demographic Censuses

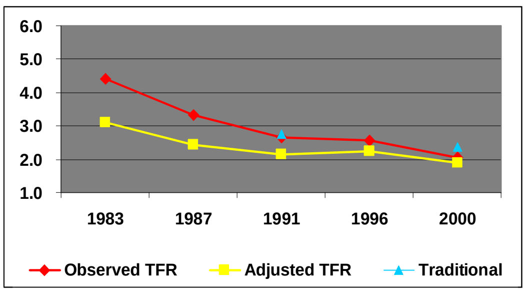
III.3- Kohler & Ortega Methods: quantum, tempo and parity effects

The birth history micro data allow the calculation of some fertility measures that are not directly available in the censuses, for example, the intensities (defined in Equation 11), which allow the appliance of the K-O method, to disentangle tempo, quantum and parity effects of fertility.

Figures 16 to 27 display the result of the application of K-O method to the Brazilian birth histories database. Figures 16 to 19 display results for Brazil, Figures 20 to 23 display results for low educated women (0-3 years of study), and Figures 24 to 27 display results for mid educated women (4-7 years of study). The results give a picture of tempo, parity, and quantum effects of Brazilian fertility in 1983, 1987, 1991 (1991 Census), 1996, and 2000 (2000 Census).

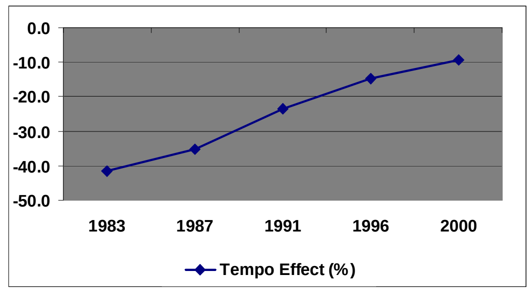
The Brazilian case is interesting to the international literature because the fertility transition towards replacement level took place without the presence of a positive and significant tempo effect – on the contrary, there was a strong negative tempo effect that declined from -40% in 1983 to -10% in 2000 (Figure 17). Because of the negative and declining tempo effect, observed TFR was always higher than adjusted TFR and the difference between observed and adjusted TFR diminished from almost 1.5 children in 1983; to almost zero in 2000 (Figure 16).

Figure 15
Brazil: Observed and Adjusted Total Fertility Rate, 1983-2000



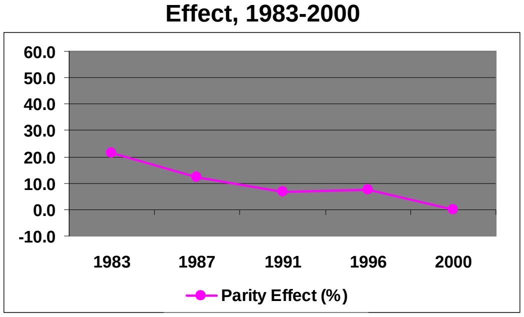
Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 16
Brazil: Fertility Tempo Effect, 1983-2000



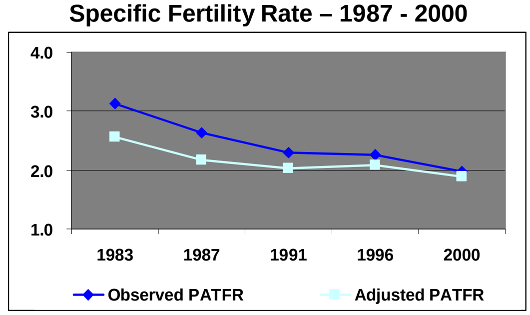
Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 17
Brazil: Parity Composition Effect Net of Tempo Effect, 1983-2000



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 18
Brazil: Observed and Adjusted Parity and Age Specific Fertility Rate – 1987 - 2000



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

The adjusted PATFR is the pure index of fertility (Ortega and Kohler, 2002), that is to say, the fertility free from tempo and compositional distortions. This index of fertility is the pure quantum effect, free from all non-behavioral effects. Results in Figure 19 suggest that the "pure quantum component" was already near the replacement level in 1987, and the below replacement level was reached between 1991 to 2000. The trend depicted is not strong enough for letting someone to predict that the pure quantum will reach the lowest low fertility level (around 1.4) in the short run. What makes adjusted TFR different from the pure quantum index (adjusted PATFR) is the parity and age distribution of fertility.

The parity composition effect is displayed in Figure 18. It declined from 20% in 1983 to 10% in 1987, remaining constant under 10% in 1991 and 1996. It declined to zero in 2000. The strong parity composition effect until 1996 favored a TFR higher than the one observed in the pure quantum index (Figure 19). It is important to remind that the parity composition effect is derived from the comparison between adjusted TFR and adjusted PATFR. The former measure is affected by the parity composition derived from past behavior, while the latter is a summary index derived solely from the current set of birth probabilities at parity and age. Thus, if fertility behavior does not change for a long time adjusted PATFR is the quantum to be observed.

Figures 20 to 23 and 24 to 27 compare fertility results between two educational strata of women: low educated (0 to 3 years of study) and mid educated (4-7 years of study). Although tempo effects had opposite trends from 1983 to 1991 and from 1996 to 2000, the magnitude of the negative tempo effect do not differ much from 1991 to 2000 for the two educational strata (Figures 21 and 25).

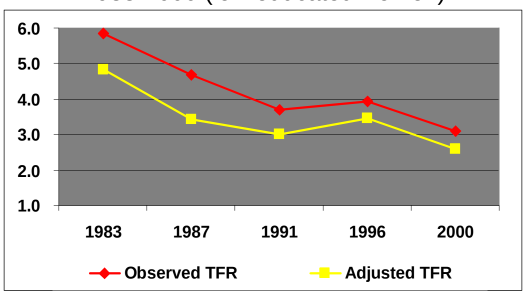
Negative tempo effect implies in observed TFR higher than adjusted TFR. Comparing observed and adjusted TFR (Figures 20 and 24), it is interesting to notice that the difference is always higher for the low educated women. In 2000, the mid educated women's adjusted TFR is at the replacement level, while the low educated women's adjusted TFR is 2.6.

The parity composition effect (Figures 22 and 26) is quite strong among low educated women, presenting a decline during the period (the effect goes from 60% in 1983 to 30% in 1987, 20% in 1991, 25% in 1996, and 11% in 2000). This parity composition effect is much lower among mid educated women (12%, 7%, 0%, 3% and 0%, respectively). The small parity composition effect among mid-educated women can be suggesting that this group has experienced a transition towards low fertility well before the low-educated group.

Low educated women had a fertility quantum (pure fertility index) of 3.0 in 1983, declining to 2.3 in 2000. The fertility quantum of mid-educated women was 2.6 in 1983 and at the replacement level in 2000 (2.12). This suggests that the recent decline in the Brazilian fertility index (1983 to 2000) is more strongly determined by the decline observed among low educated women.

Figure 19

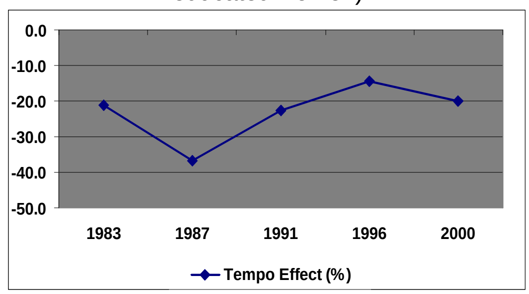
Brazil: Observed and Adjusted Total Fertility Rate, 1983-2000 (low educated women)



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 20

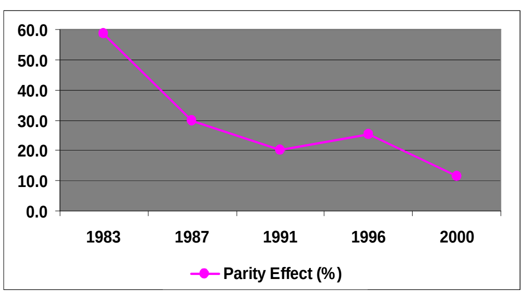
Brazil: Fertility Tempo Effect, 1983-2000 (low educated women)



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 21

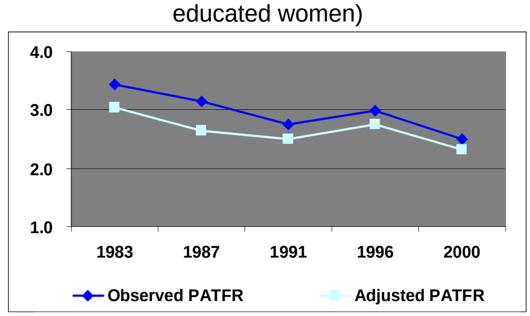
Brazil: Parity Composition Effect Net of Tempo Effect, 1983-2000 (low educated women)



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 22

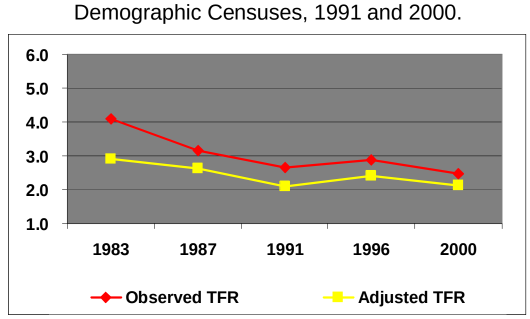
Brazil: Observed and Adjusted Parity and Age Specific Fertility Rate – 1987 – 2000 (low educated women)



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

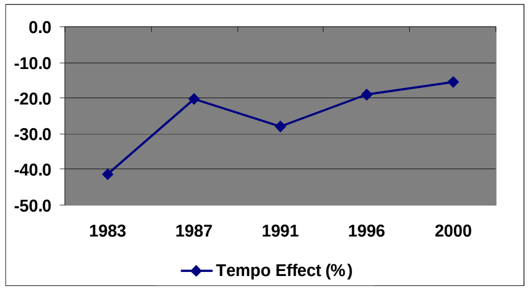
Figure 23

Brazil: Observed and Adjusted Total Fertility Rate, 1983-2000 (mid educated women)



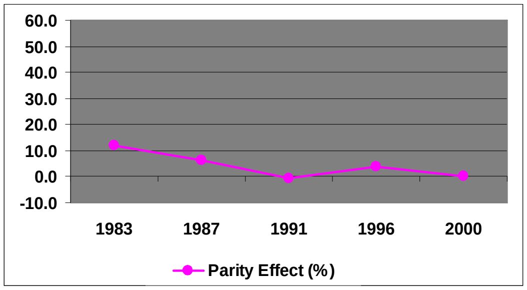
Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 24
Brazil: Fertility Tempo Effect, 1983-2000 (mid educated women)



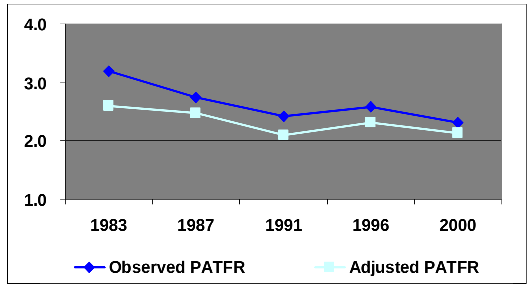
Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 25
Brazil: Parity Composition Effect Net of Tempo Effect, 1983-2000 (mid educated women)



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Figure 26
Brazil: Observed and Adjusted Parity and Age Specific Fertility Rate – 1987 – 2000 (mid educated women)



Source: Birth History Reconstruction from the Brazilian Demographic Censuses, 1991 and 2000.

Final Remarks

This paper compared two methodologies of birth history reconstruction – BHM and BHP – that were applied to the Brazilian case using census data. The results were compared contrasting the TFR series obtained from these two methodologies to the TFR series obtained by the application of the OCM. The idea was to test and to compare the results of the methodologies, so that the methodology with more satisfactory results would be used to generate the database of birth histories.

Results indicated that no significant differences were observed among the TFR series in places with good quality data and low mortality and fertility levels. However, in places with bad data quality and relatively high mortality and fertility levels, the TFR series derived from the application of BHP presented significant differences with respect to the OCM's TFR series. This did not happen in the case of the BHM's TFR series. Thus, BHM was adopted for the reconstruction of birth histories of all women enumerated in the 1991 and 2000 Brazilian Censuses.

The reconstruction of birth histories in large surveys and censuses can be considered a great advance in the field of fertility studies in developing countries. Due to the low quality of birth registration and to the absence of good birth histories in countries like Brazil, this procedure can shed light to the process of post-transitional fertility in developing countries. It stresses points overlooked in the traditional analysis such as quantum and tempo fertility, as well as the analysis of birth spacing.

Taking the Brazilian example, the country has experienced a steady fertility decline during the last thirty years, and the only empirical evidence used by the experts to analyze it has been the measurement of period TFRs (incidence rates). The possibility of measuring other fertility indexes, as intensities, enabled us to understand a little more about the Brazilian fertility transition. We now know that the tempo effect in the Brazilian transition operated in the opposite direction from the one found in the developed countries. Also know that there was a strong parity composition effect, also forcing TFR to higher levels. Finally, we know that the Brazilian pure fertility index (quantum) is already at the replacement level, so that if Brazil enters a postponement stage it will inevitably reach TFRs at below replacement levels.

References

BRASS, W. **Methods for estimating fertility and mortality from limited and defective data**. Chapel Hill, North Carolina. The North Carolina Center, 1975.

CARVALHO, J.A.M.; PINHEIRO, S.M.G; UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Fecundidade e mortalidade no Brasil - 1970/80.** Belo Horizonte: CEDEPLAR, 1986. 151p.

CARVALHO, J.A.M; UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Fecundidade e mortalidade no Brasil - 1960/1970.** Belo Horizonte: Centro de Desenvolvimento e Planejamento Regional, 1978. 102p.

CARVALHO, J.A.M; UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Tendências regionais de fecundidade e mortalidade no Brasil.** Belo Horizonte: CEDEPLAR, 1974. 95p.

CEDEPLAR, Dados preliminares para a Projeção Populacional por sexo e grupos de idades qüinqüenais, das Unidades da Federação, Brasil, 1990/2020, 1999. Relatório de atividades do Pronex (Uso restrito).

CHO, L-J., RETHERFORD, R.D., CHOE, M.K. **The own-children method of fertility estimation**. Honolulu, Hawaii: University of Hawaii Press, 1986. 188p.

FEENEY, G. A simpler matrix approach to polynomial interpolation. Honolulu: East-West Population Institute, 1974.Cited by CHO, L-J., RETHERFORD, R.D., CHOE, M.K. The own-children method of fertility estimation. Honolulu, Hawaii: University of Hawaii Press, 1986. 188p.

GOLDANI, A.M. (2002) What will happen to Brazilian fertility? **Completing the fertility transition***.* ESA/P/WP.172/Rev.1. UN Population Division, pp. 358-375.

KOHLER, H.-P. and J. A. ORTEGA (2002): "Tempo-Adjusted Period Parity Progression Measures, Fertility Postponement and Completed Cohort Fertility", **Demographic Research**, vol. 6(6), pp. 91-144.

KOHLER, H-P and D. PHILIPOV (2001). Variance effects in the Bongaarts-Feeney formula. **Demography** 38(1), 1-16.

LUTHER, N.Y, CHO, L-J. Reconstruction of birth histories from census and household survey data. **Population Studies**, 42: 451-472, 1988.

ORTEGA, J.A. and H-P. KOHLER (2002) Measuring low fertility: rethinking demographic methods. **Demographic Research**, Working Paper 2002-001, Max Planck Institute for Demographic Research, Rostock, Germany (disponível em http://www.demogr.mpg.de).

SILVA, V.C, MIRANDA-RIBEIRO, A., RIOS-NETO, E.L.G. A period decomposition of fertility decline in Brazil: pure fertility index, tempo and parity composition effect. In the CD-ROM of XXV General Population Conference of International Union for the Scientific Study of Population (IUSSP), Tours, France, 18-24 July, 2005.