The Stata Journal (2001)**1**, Number 1, pp. 1–28

The own-children method of fertility estimation using Stata

Abstract. This article presents `ocm`, a program to calculate age-specific fertility rates using an advanced version of the Own-Children Method, originally proposed by Grabill and Cho (1965). The software provides a graphical representation of average fertility patterns by age over the last 15 years, generates weighted estimates for various population subgroups, enhances accuracy by focusing calculations on biological connections between children and their mothers, and delivers 15 reproductive measures. These measures include total and net fertility rates, mean age at childbearing, the percentage of teenage pregnancies, the proportion of childless women, the percentage of unmatched children, and the replacement level of fertility. I demonstrate the capabilities of `ocm` using 2010 Brazilian Census microdata sourced from IPUMS-International.

Keywords: st00!!, children, age, demography, fertility, reproduction, own-children, Brazil, survival

1 Introduction

The Own-Children Method (OCM) is an appealing reverse-survival technique for estimating the average number of children women and men have over their lifecycles (Cho et al. 1986; Dubuc 2009; Grabill and Cho 1965; Schoumaker 2017, 2019). The method generates fertility levels and trends over the past 15 years, across population subgroups, and in countries without comprehensive birth registration. In addition, despite imperfections in data and minor deviations from its assumptions, OCM consistently delivers reliable results (Abbasi-Shavazi et al. 2013; Avery et al. 2013; Retherford and Cho 1978; Rindfuss 1976; Timaeus 2021).

The application of OCM, however, has been limited by its extensive computational demands. These include algorithms for associating children with their potential mothers, determining maternal age at the time of each birth, special tabulations of births by mother's age at single-year intervals, identifying and reallocating children with unidentified mothers, reverse survival calculations for children and women prior to a census, and averaging age-specific fertility rates across different ages and time periods. To streamline these processes, this article introduces `ocm`, a Stata command that efficiently computes reproductive measures and fertility rates using OCM (United Nations 1983).[1]

Designed to work with census and survey data, `ocm` is particularly effective with the Integrated Public Use Microdata Series (IPUMS) sources, such as IPUMS-USA, IPUMS-

1. A user-friendly dialog box for implementing `ocm` is accessible through the Stata menu under User → Statistics → Fertility Estimation using the Own-Children Method (ocm), or by entering `db ocm` in the Command window.

International, and IPUMS-CPS (Ruggles et al. 2024b,a; Flood et al. 2023). I reccomend the use of IPUMS data because it provides standardized, high-quality demographic variables essential for applying the OCM, ensuring robust fertility estimates. As of April 2024, for example, these variables were available for approximately 100 countries and 511 periods (Ruggles et al. 2024a).[2] The data required by OCM is also, by far, the most available and covering the largest age range in Demographic and Health Surveys (Schoumaker 2017, 806).

The following section details the origins, assumptions, robustness, and prior applications of OCM across various contexts and populations. Subsequent sections describe the algorithms of `ocm` and illustrate its capabilities using 2010 Brazilian census microdata categorized by race (White, Brown, or Black).

2 Background

Wilson Grabill and Lee Jay Cho were the first to propose a strategy to estimate fertility rates by linking children in the household to their presumed mothers (Grabill and Cho 1965). This technique, which became known as the Own-Children Method, employs reverse-survival to calculate the number of women (denominator) and the number of births (numerator) over the preceding 15 years who were living in the same family and who had a mother-child tie. Age-specific fertility rates can then be estimated by dividing the number of births by the number of women at each year preceding the enumeration (Cho et al. 1986).

The Own-Children Method has several distinct advantages: (i) it generates detailed age-specific fertility by single years of age and across socioeconomic characteristic such as income, race, religion, marital status, geographical location, occupation, and education; (ii) it only requires data from a single census or cross-sectional survey to produce estimates for up to 15 years before the survey date; (iii) it remains effective in populations with low migration rates during the periods previous to enumeration; and (iv) it exhibits a high tolerance to errors in recall and temporal fluctuations in mortality rates (Feeney 1975). The accuracy of OCM, however, depends on the quality of the input data and the validity of its underlying assumptions (Grabill and Cho 1965; Retherford and Cho 1978). Specifically, the OCM may be susceptible to five sources of error or assumptions that need to be critically evaluated.

The first and most significant source of error in OCM is age misreporting (Myers 1940, 1954). Incorrect reporting of children's age can skew the observed age pattern and trends in fertility leading to apparent fluctuations from year to year (Cho et al. 1986). However, the impact of age misstatement can be mitigated by smoothing population counts by single years of age or aggregating estimates across five-year age groups over several years (Abbasi-Shavazi 1997; Abbasi-Shavazi et al. 2013).

The second source of error is enumeration underreporting. When the underenumer-

2. For example, the key variable MOMLOC, which establishes the mother-child link within the household, is consistently available across all these datasets Sobek and Kennedy (2009).

ation of children and women is not uniform, fertility estimates will be inherently biased. Nevertheless, the risk of differential underenumeration is reduced when both the numbers of births and women come from the same data source. Moreover, reverse-surviving older children to estimate the numerator of fertility rates minimizes the undercount of newborns in censuses and surveys (Hattersley and Creeser 1995). In most developed countries, the underenumeration of births is below 2%, but it can get as high as 59% in the least developed regions (UNICEF 2023). In Brazil, for instance, the birth registration completeness exceeds 95% (Hunter and Sugiyama 2018).

The third underlying assumption of OCM is low rates of migration, or what is alternatively termed the "constancy of group membership" (Rindfuss 1976, 1977). If there is substantive change in the composition of the group status (such as occupation, education, religion, race etc.) over the 15 years preceding the data collection, fertility estimates could be misleading. Stated differently, if there is moblity across categories over time, then the constancy of group-membership assumption is violated, which would bias historical fertility estimates. However, if migration rates are low or the fertility patterns of migrants resemble those of non-migrants during the estimation period, the impact of violating the constancy of group membership assumption may be negligible (Retherford and Cho 1978).

The fourth potential issue arises from mismatches between mothers and their own-children. Situations where the mother has died, does not reside in the same household, or the connection to her offspring is social rather than biological – as in cases of adoption or stepchildren – can result in numerous "non-own" children. To avoid underestimating fertility rates, these unmatched children must be redistributed among identified mothers, typically under the assumption that fertility and mortality patterns are consistent across both identified and unidentified mothers. Although mismatches can lead to bias, it is generally less significant than errors caused by age misreporting, and can be further reduced by ensuring that the mother-child linkages are confinedto biological relationships (Abbasi-Shavazi 1997; Krapf and Kreyenfeld 2015). Moreover, even when the proportion of non-own children is as high as 30%, the impact on fertility estimates remains minimal (Cho et al. 1986).

The fifth assumption concerns the temporal constancy of mortality rates. Theoretically, if there have been substantial changes in the survivorship of children and women over the past 15 years, it would be imprudent to apply a constant life table for reverse survival calculations. However, in practice, OCM exhibits low sensitivity to innaccuracies in life table estimates or variations in mortality. Spoorenberg (2014) demonstrated that even significant shifts in life expectancy – ranging between 10 and 20 years – have a negligible effect on fertility estimates.

Overall, minor deviations from these assumptions have slight impacts on OCM fertility estimates. The method's reliability stems from its ability to provide accurate fertility trends even with unadjusted data (Rindfuss 1976), and it continues to yield results comparable to those obtained from birth history analysis, even when faced with potential errors (Abbasi-Shavazi 1997; Avery et al. 2013; Dubuc 2009; Schoumaker 2017; Retherford and Alam 1985). Due to its robustness, OCM has been applied to a multitude of settings, particularly where data from vital registration systems are incomplete or unavailable for certain population subgroups (Abbasi-Shavazi et al. 2013; Dubuc 2009; Hacker 2003). In Brazil, the method has been employed to estimate fertility across different years, regions (Lima et al. 2018; Miranda-Ribeiro 2007; Wong 1983a,b), and income classes (Muniz 2012), highlighting its versatility and effectiveness in diverse contexts.

3 Required data

The implementation of the Own-Children Method relies on three specific datasets:

1. The probabilities of child survivorship from age 0 to 15 years;
2. The probabilities of survivorship for adult females from age 15 to 65 years;
3. A set of variables identifying persons and households, ideally extracted from IPUMS (Ruggles et al. 2024b,a; Flood et al. 2023).

The first two datasets, containing probabilities of survival, must be provided by external `.dta` files before evoking `ocm`.[3] Additionally, the third dataset must include the following variables:[4]

```
. use Brazil_2010_census.dta, clear
. ds, detail
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
country         byte    %8.0g      COUNTRY    country
year            int     %8.0g      YEAR       year
serial          double  %12.0g                household serial number
pernum          byte    %8.0g                 person number
perwt           double  %12.0g                person weight
momloc          byte    %8.0g                 mother´s location in household
stepmom         byte    %8.0g      STEPMOM    probable stepmother
age             byte    %8.0g      AGE        age
age2            byte    %8.0g      AGE2       age, grouped into intervals
sex             byte    %8.0g      SEX        sex
race            byte    %8.0g      RACE       race or color
```

These variables, available via IPUMS for over 100 countries, are crucial for implementing OCM. To illustrate the procedure I use microdata from the 2010 Brazilian census. Note that the name of these variables across datasets must be consistent as `ocm` references them internally. The variables “person weight” (`perwt`) and “race or color”

3. In section 5, I demonstrate how to derive survival probabilities for children and women by single years of age using an iterative method of mortality estimation (Merli 1998; Muniz 2023).
4. I utilized `texdoc` for integrating Stata outputs into LaTeX(Jann 2016).

(`race`) are optional but useful. The first expands the sample to represent total population counts, while the second facilitates the analysis of fertility rates across different racial subgroups.

The IPUMS's constructed variables `momloc` and `stepmom` are particularly important because they identify the link between children and their potential mothers within the same household. Variable `momloc` indicates whether a child's mother resides in the same household, distinguishing between social (e.g., stepmothers, adoptive mothers) and biological relationships. This linkage is essential to OCM's ability to accurately estimate fertility. More details on the matching criteria between children and their likely mothers are discussed in Sobek and Kennedy (2009).

4 On mathematical notation

I utilize matrix notation to present the methodologies employed in the own-children method for two primary reasons: firstly, because `ocm` extensively uses Mata for computation; and secondly, to ensure clarity and avoid visual clutter in the presentation. Matrices are denoted by bold uppercase letters ($\mathbf{C}, \mathbf{W}, \mathbf{F}$); vectors are represented by bold lowercase letters ($\mathbf{p}, \mathbf{c}, \mathbf{aw}, \mathbf{k}$); and individual elements or scalars are indicated by italicized lowercase letters.

To differentiate similar vectors or matrices referring to different periods, or those with varying lenghts and elements, I use superscripts (e.g., $\mathbf{p}^c$ and $\mathbf{p}^w$). I use subscripts to identify the row and column positions of elements within vectors or matrices ($\mathbf{k}_{16}, \mathbf{C}_{a,16}$). Multiplication by a vector of ones ($\mathbf{1}_{51}, \mathbf{1}_{18}$) represents the summation of elements in vectors and matrices. Other mathematical symbols and operations are defined locally as needed, including operations such as element-wise division ($\oslash$), multiplication ($\odot$), transpose ($^\intercal$), among others.

5 The OCM algorithms

The own-children method of fertility estimation involves a sequence of ten consecutive steps:

Step 1. Estimation of survivorship probabilities for children aged 0 to 15 years

These probabilities must be independently obtained and specified in `ocm`. They are employed to back-project the number of children reported in the date of the survey or census 15 years into the past (e.g., in step 7). Children's survivorship probability, denoted as $\mathbf{p}^c$, is equal to one minus the probability of dying between successive ages ($\mathbf{q}^c$):

$$\mathbf{p}^c = \mathbf{1} - \mathbf{q}^c \tag{1}$$

Here, $\mathbf{1}$ is an all-ones vector, and $\mathbf{q}^c$ is a column vector with entries representing the probability of death between ages consecutive ages x and $x+1$, ${}_{x+1}q_x^c$. The vetor spans from ${}_1q_0^c$ to ${}_{16}q_{15}^c$. There are several ways to obtain a life table containing these probabilities (Preston et al. 2001; United Nations 1983)¿ However, for precision and efficiency, I use an iterative method of mortality estimation developed by Merli (1998) and implemented in Stata by Muniz (2023).

This method requires two age-specific population distributions and the average number of deaths between them. For this study, I use the 2010 and 2022 population counts from the Brazilian censuses (IBGE 2024), along with and the mortality data from the Brazilian Ministry of Health (Ministério da Saúde 2024). Below is the Stata code snipet used to generate the survival probabilities for ages 0 to 15:

```
. use br2010_22.dta, clear
. qui ilt, interval(12) // from Muniz(2023)
. mata p=1:-qi //Probability of surviving
. getmata p
. forval x=1(5)80 {
  2.          qui insobs 4, after(`x´)
  3.                         }
. qui replace age=_n-1
. foreach var of varlist p {
  2. ipolate `var´ age, gen(`var´c)
  3.                         }
. drop pop_2010 pop_2022 deaths p
. qui drop in 17/l
. lab var age "Exact age"
. lab var pc "Prob. of surviving btw. ages x and x+1"
. qui save Lkids.dta, replace
. list, noobs clean

    age          pc
      0   .97902375
      1   .98287902
      2   .98673428
      3   .99058955
      4   .99444481
      5   .99830008
      6   .99818876
      7   .99807744
      8   .99796612
      9    .9978548
     10   .99774349
     11   .99650878
     12   .99527407
     13   .99403936
     14   .99280465
     15   .99156994
```

The output variable, `pc`, indicates survival probabilities. It shows, for instance, a 97.90% probabilitiy of surviving the first year of life, and a 99.16% probability from age

15 to 16. The saved dataset (Lkids.dta) is indispensable for ocm in step 7.

Step 2. Estimation of survivorship probabilities for adult females aged 15 to 65 years

This step mirrors the process described for children, adapting it to adult females between ages 15 and 65 years. The key difference is that the input data used in ilt – population counts and the average number of deaths over the period – refer to adult women instead of children. Namely:

```
. use br_women2010_22.dta, clear
. qui ilt, interval(12)
. mata p=1:-qi
. getmata p
. forval x=1(5)80 {
  2.          qui insobs 4, after(`x´)
  3.                         }
. qui replace age=_n-1
. foreach var of varlist p {
  2. ipolate `var´ age, gen(`var´w)
  3.                         }
. drop pop_2010 pop_2022 deaths p
. qui drop in 1/15
. qui drop in 52/l
. qui list, noobs
. lab var age "Exact age"
. lab var pw "Prob. of surviving btw. ages x and x+1"
. qui save Lfemale.dta, replace
```

The variable in the second column of Lfemale.dta is similar to (1), and can be expressed as

$$\mathbf{p}^w = \mathbf{1} - \mathbf{q}^w \tag{2}$$

where $\mathbf{p}^w$ is a column vector representing the probabilities of female survival, by single ages between 15 and 65 years. The resultant dataset, Lfemale.dta, is crucial to implement step 8 in OCM.

Step 3. Tabulation of children by age and their mothers' age

Matrix **C**, also referred to as "the own-children tabulation", is compiled by cross-tabulating two variables: 1) the mother's single year of age (spanning ages 15 and 65 years, represented by rows), and 2) the child's age (ranging from 0 to 15 years, represented by columns). The age of the mother is obtained from momloc, a variable included in most census data provided by **?**. This variable indicates whether or not the person's mother lived in the same household. It identifies social (stepmother and adopted mother) and biological relationships, and is available for analysis in over 100 countries.

A snapshot of this 51 by 16 tabulation matrix, using unweighted 2010 Brazilian census data, is

```
. use Brazil_2010_census.dta, clear
. qui ocm, p_children(Lkids.dta) p_female(Lfemale.dta) nodraw
. mata C
```

(*Output partially omitted*)

$$\mathbf{C} = \begin{pmatrix} 3515 & 814 & 216 & 110 & 75 & 57 & 43 & 41 & \cdots & 18 & 14 \\ 6698 & 3527 & 842 & 287 & 158 & 129 & 104 & 79 & \cdots & 34 & 15 \\ 9716 & 6654 & 3579 & 971 & 351 & 193 & 161 & 134 & \cdots & 43 & 27 \\ 12288 & 10077 & 6733 & 3662 & 1083 & 448 & 230 & 208 & \cdots & 55 & 65 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 11 & 12 & 13 & 28 & 19 & 20 & 45 & 48 & \cdots & 219 & 277 \end{pmatrix}$$

The first element of this matrix indicates that in 2020 there were 3515 newborn children whose mothers were 15 years old. COnversely, the last element shows that women aged 65 years had 277 children aged 15 years old in 2010. The sum of all elements in **C** represents the total number of children whose mothers were identified in the same household.

Step 4. Number of children with unidentified mothers, classified according to their own age

The count of children with unidentified mothers ($\mathbf{c}^u$) is obtained in a similar manner by tabulating instances where the mother's age is missing, against the children's own age (from 0 to 15 years). This is illustrated as follows:

$$\mathbf{c}^u = \begin{pmatrix} 7660 & 11339 & 14625 & 17883 & 20942 & \cdots & 49564 & 57093 \end{pmatrix}$$

This row vector reflects the age-specific counts of children whose mothers were not recorded in the dataset. For instance, there were 7,660 unidentified cases for newborns and 57,093 cases for 15-year-olds. These children will be proportionally reallocated among identified mothers in subsequent steps (specifically, steps 6 and 7) to better estimate the overall fertility rates.

Step 5. Tabulation of women (regardless of motherhood status) by single year of age

The number of women by single year of age, ranging from 15 and 65 years, encompasses those who have been exposed to the risk of childbearing. Notice that the vector **aw** includes counts of adult women beyond the reproductive period (i.e., older than 49 years) because these women will be backprojected to their earlier reproductive ages in step 8, thereby accounting for the entire fertility history of the cohort.

$$\mathbf{aw} = \begin{pmatrix} 197466 \\ 187722 \\ 183980 \\ \vdots \\ 66025 \\ 66186 \end{pmatrix}$$

Each entry in vector **aw** represents the number of women at a specific age who were considered potentially fertile at some point. For instance, the first element of **aw** shows that there were 197,466 women aged 15 and 66,186 women aged 65 according to the 2010 Brazilian census. These figures illustrate the structure of the female population within the studied age range and provide the foundation for further demographic and fertility analyses in subsequent steps.

Step 6. Adjustment factors for redistributing children with unidentified mothers

The reallocation of children with unidentified mothers involves computing an expansion factor, denoted as **k**, to account for the proportional redistribution of these children across various age groups from 0 to 15:

$$\mathbf{k} = \mathbf{1_{16}} + \mathbf{c}^u \oslash \mathbf{1_{51}}\mathbf{C} = \begin{pmatrix} 1.0263 & 1.0395 & 1.0508 & \cdots & 1.1442 & 1.1660 \end{pmatrix}$$

In this equation, $\mathbf{1_{16}}$ and $\mathbf{1_{51}}$ are row vectors of ones, sized 16 and 51, respectively; $\oslash$ represents the element-wise division; and $\mathbf{1_{51}C}$ calculates the total children with identified mothers by age through summing the elements in $\mathbf{C}$. The row-vector $\mathbf{k}$ is then used to adjust the counts of identified children by age, increasing, for example, the number of identified children at age 0 by 2.63% and at age 15 by 16.60%.

This procedure of redistributing the unmatched children assumes that the age pattern of (unobserved) non-coresident mothers is similar to that of mothers living with the their children. Timaeus (2021), however, points that this assumption would distort the age pattern of fertility because maternal orphans tend to be concentrated among older women. As a result, assuming that the distribution of unmatched children independs of the mother's age will likely overestimate the fertility of younger women and underestimate that of older ones. To overcome this challenge, Timaeus suggests that only children whose mother is alive should be included in the numerator of age-specific feritlity rates. This is an ingenious solution, but it depends on the availability of the question "Is this child's mother alive?" in the census or survey under investigation.[5] Alternatively, the proportion of children with living mothers could also be obtained

5. The variable (e.g., `mortmot`) indicating whether the person's biological mother was still living at the time of the census was available for 42 countries and 83 periods in IPUMS-I (Ruggles et al. 2024a).

through a predefined linear equation, but this would require a new set of underlying assumptions (Timaeus 2013).

Another alternative to include the unmatched children into the numerator of age-specific rates would be to create a rule to assign them to living mothers in the data set. Schoumaker (2017), for example, has used this strategy to randomly link unmatched children to fathers. In the case of female fertility, however, there is no justifiable reason to assume that the distribution of unmatched children among living mothers is random. Moreover, inputting the mother's age to unmatched children – using Stata's `mi` suite of commands, for example – would increase `ocm`'s computation time, making it inefficient and likely generating results and biases similar to those produced by the traditional OCM approach. The bias caused by the missalocation of unmatched children on recent age-specific and total fertility rates is in general negligible, especially when the number or orphans is small, but these adjustments are crucial for accurate demographic analysis (Timaeus 2021, 834-835)].

Step 7. Reverse survival of children

The number of births ($\mathbf{b}$) occurring in a past year $(t-15+x)$ to mothers between 15 and 49 years is given by the average of children surviving to these mothers at successive ages a and $a+1$. Thus

$$\mathbf{b}_a^{t-15+x} = \frac{\mathbf{k}_{17-x}\mathbf{C}_{a,17-x}}{\mathbf{p}_{17-x}^c} \tag{3}$$

$$\mathbf{b}_{a+1}^{t-15+x} = \frac{\mathbf{k}_{17-x}\mathbf{C}_{a+1,17-x}}{\mathbf{p}_{17-x}^c} \tag{4}$$

and

$$\mathbf{b}^{t-15+x} = \frac{\mathbf{b}_a^{t-15+x} + \mathbf{b}_{a+1}^{t-15+x}}{2} \tag{5}$$

In this formulation, subscript a denotes mothers' ages spanning from rows $(17-x)$ to $(51-x)$ in matrix $\mathbf{C}$, corresponding to ages 31 to 65. Subscript $a+1$ indexes the elements of $\mathbf{C}$ located between its $(18-x)$th and $(52-x)$th rows. The other subscript, $(17-x) = 16, 15, 14...2$, indexes the position of the column in vectors $\mathbf{k}$, $\mathbf{p}^c$, and in matrix $\mathbf{C}$. It also represents the child's age x years prior to the survey. For example, using data from the 2010 Brazilian census to compute the number of births in 1996, we find:

$$\mathbf{b}_a^{1996} = \frac{\mathbf{k}_{16}}{\mathbf{p}_{16}^c}\mathbf{C}_{a,16} = \frac{1.1660}{.99156994}\begin{pmatrix} 4426 \\ 7475 \\ 12080 \\ \vdots \\ 295 \\ 246 \end{pmatrix} = \begin{pmatrix} 5204 \\ 8790 \\ 14205 \\ \vdots \\ 347 \\ 289 \end{pmatrix} \tag{6}$$

$$\mathbf{b}_{a+1}^{1996} = \frac{\mathbf{k}_{16}}{\mathbf{p}_{16}^{c}}\mathbf{C}_{a+1,16} = \frac{1.1660}{.99156994}\begin{pmatrix} 7475 \\ 12080 \\ 151164 \\ \vdots \\ 246 \\ 277 \end{pmatrix} = \begin{pmatrix} 8790 \\ 14205 \\ 17831 \\ \vdots \\ 289 \\ 326 \end{pmatrix} \tag{7}$$

Thus, the estimated number of births in 1996 to women aged 15 to 49 is derived by averaging vectors from (6) and (7):

$$\mathbf{b}^{1996} = \frac{\mathbf{b}_{a}^{1996} + \mathbf{b}_{a+1}^{1996}}{2} = \begin{pmatrix} 6997 \\ 11497 \\ 16018 \\ \vdots \\ 318 \\ 307 \end{pmatrix} \tag{8}$$

For operational convenience, these vectors of birth counts spanning the last fifteen years are consolidated into a single matrix $\mathbf{B}$, containing 35 rows (representing mother's ages from 15 to 49 years) and 15 columns (for each of the past 15 years):

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}^{1996} & \mathbf{b}^{1997} & \cdots & \mathbf{b}^{2010} \end{pmatrix} = \begin{pmatrix} 6997 & 7567 & \cdots & 5384 \\ 11497 & 12194 & \cdots & 8847 \\ 16018 & 16590 & \cdots & 11758 \\ \vdots & \vdots & \ddots & \vdots \\ 307 & 294 & \cdots & 114 \end{pmatrix} \tag{9}$$

This matrix provides a comprehensive view of historical birth trends and is instrumental in further demographic analyses.

Step 8. Reverse survival of women

The number of women who were alive and exposed to the risk of having children in the past 15 years ($\mathbf{w}^{t-15+x}$) is obtained by back-projecting the female adult population using the probabilities of survival in $\mathbf{p}^{w}$. This process is delineated as follows:

$$\mathbf{w}_{a}^{t-15+x} = \mathbf{aw}_{a} \oslash \mathbf{p}_{a}^{w} \otimes \mathbf{p}_{r}^{w} \tag{10}$$

$$\mathbf{w}_{a+1}^{t-15+x} = \mathbf{aw}_{a+1} \oslash \mathbf{p}_{a+1}^{w} \otimes \mathbf{p}_{r}^{w} \tag{11}$$

and

$$\mathbf{w}^{t-15+x} = \frac{\mathbf{w}_{a}^{t-15+x} + \mathbf{w}_{a+1}^{t-15+x}}{2} \tag{12}$$

where a and $a+1$ are as in (3) and (4), r indexes the probabilities of female survival between rows 1 and 35 (or ages 15 and 49 years) of $\mathbf{p}^w$, and $\otimes$ is the element-wise multiplication of column vectors. The number of women in 1996, for instance, is

$$\mathbf{w}_a^{1996} = \begin{pmatrix} 181048 \\ 164707 \\ 165617 \\ \vdots \\ 66025 \end{pmatrix} \oslash \begin{pmatrix} .9928 \\ .9922 \\ .9915 \\ \vdots \\ .9039 \end{pmatrix} \otimes \begin{pmatrix} .9965 \\ .9963 \\ .9962 \\ \vdots \\ .9703 \end{pmatrix} = \begin{pmatrix} 181725 \\ 165399 \\ 166390 \\ \vdots \\ 70879 \end{pmatrix} \tag{13}$$

$$\mathbf{w}_{a+1}^{1996} = \begin{pmatrix} 164707 \\ 165617 \\ 157417 \\ \vdots \\ 66186 \end{pmatrix} \oslash \begin{pmatrix} .9922 \\ .9915 \\ .9909 \\ \vdots \\ .8972 \end{pmatrix} \otimes \begin{pmatrix} .9965 \\ .9963 \\ .9962 \\ \vdots \\ .9703 \end{pmatrix} = \begin{pmatrix} 165427 \\ 166418 \\ 158252 \\ \vdots \\ 71580 \end{pmatrix} \tag{14}$$

and thus

$$\mathbf{w}^{1996} = \frac{\mathbf{w}_a^{1996} + \mathbf{w}_{a+1}^{1996}}{2} = \begin{pmatrix} 173576 \\ 165909 \\ 162321 \\ \vdots \\ 72854 \\ 71230 \end{pmatrix} \tag{15}$$

Similar to (9), the vectors containing the number of women between ages 15 and 49 in the past 15 years can be arranged into a single matrix $\mathbf{W}$ of the same order as $\mathbf{B}$, with 35 rows and 15 columns:

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}^{1996} & \mathbf{w}^{1997} & \cdots & \mathbf{w}^{2010} \end{pmatrix} = \begin{pmatrix} 173576 & 165909 & \cdots & 71230 \\ 178210 & 173547 & \cdots & 72705 \\ 16018 & 16590 & \cdots & 11758 \\ \vdots & \vdots & \ddots & \vdots \\ 185898 & 182948 & \cdots & 116051 \end{pmatrix} \tag{16}$$

This matrix serves as a key tool for further demographic analyses and planning.

Step 9. Calculation of age-specific fertility rates

The number of births to women by age group is defined by the following matrix operation:

$$\mathbf{F} = \mathbf{B} \oslash \mathbf{W} \tag{17}$$

Here, $\mathbf{F}$ is a 35 x 15 matrix containining single age-specific fertility rates for women between 15 and 49 years (rows), over the past 15 years (columns). However, because of age-reporting errors – such as age-heaping and the under enumeration of children below age 5 –, single age estimates are likely erratic. To mitigate these issues and facilitate comparison with alternative fertility estimation methods which utilize five-year age group estimates, I adopt a smoothing technique recommended by the United Nations (1983), focusing on three-year moving averages for these rates, aggregated into five-year age groups ($\mathbf{F}^{3,5}$). These three-year moving averages for five-year age groups are calculated as follows:

$$\mathbf{F}^{5,3} = \frac{1}{3} \sum_{j=1,4\ldots13}^{j+2} \frac{1}{5} \sum_{i=1,6\ldots31}^{i+4} \mathbf{F}_{i,j} \tag{18}$$

In this formula, the subscripts i and j respectively index the rows and columns of $\mathbf{F}$. Matrix $\mathbf{F}^{3,5}$ represents a smoothed version of $\mathbf{F}$, obtained after averaging age-specific fertility rates for five-year age groups and for moving periods of three years observed in the past 15 years.

Step 10. Calculation of total fertility rates

The Total Fertility Rate (TFR) is a vital demographic indicator that quantifies the average number of children a woman would bear if she survived to the end of her reproductive life span while experiencing specific age-specific fertility rates at each age. As defined by (Preston et al. 2001, 95), the TFR can be mathematically expressed as follows:

$$\mathbf{tfr} = \mathbf{1}_7 \mathbf{F}^{5,3} 5 \tag{19}$$

where $\mathbf{1_7}\mathbf{F}^{5,3}$ is the column-sum of age-specific rates in $\mathbf{F}^{5,3}$, and the scalar 5 represents the length of the age interval. The elements of `tfr` correspond to average total fertility rates observed for moving periods of three years in the 15 years before the survey enumeration. This summary measure provides critical insight into the fertility levels and shifts within the population over time.

6 Summary measures of reproduction

The `summary` option in `ocm` calculates 15 demographic measures based on information available at the time of the survey. This section delineates the formulas for these measures as they are displayed in the output. Readers interested in a deeper understanding of these calculations are encouraged to review this section and to look further into its references.

The first measure is the mean age of the population (A_p), calculated as a weighted average of the proportions of people in each age group:

$$A_p = \frac{\mathbf{1}_{18}(\mathbf{a} \otimes \mathbf{n})}{\mathbf{1}_{18}\mathbf{n}} \tag{20}$$

Here, $\mathbf{a}$ is a column-vector of the midpoints of age intervals ranging from 2.5 to 87.5 years; $\mathbf{n}$ corresponds to the count of individuals within each age group; and $\mathbf{1}_{18}$ is a row vector of ones, with a length of 18, indicating the number of age groups considered.

The second imeasure, the mean age at childbearing (A_m), is defined as:

$$A_m = \frac{\mathbf{1}_{18}(\mathbf{a} \otimes \mathbf{m})}{\mathbf{1}_{18}\mathbf{m}} \tag{21}$$

where $\mathbf{m}$ is the maternity rate, i.e., the number of daughters born per woman, defined as

$$\mathbf{m} = \frac{\mathbf{F}_5^{5,3}}{1 + srb} \tag{22}$$

In this equation, $\mathbf{F}_5^{5,3}$ refers to the fifth column of $\mathbf{F}^{5,3}$, which contains the most recent age-specific fertility rates. The term srb stands for the sex ratio at birth, which is the empirical ratio of male to female births.

The crude birth rate (cbr) quantifies the average number of births in the population over a specific period, calculated as:

$$cbr = \frac{\mathbf{1}_{18}(\mathbf{n} \otimes \mathbf{m})}{\mathbf{1}_{18}\mathbf{n}} \tag{23}$$

The gross reproduction rate (grr) measures the number of daughters a woman would expect to have if she survived through her entire reproductive period under current fertility rates:

$$grr = \mathbf{1}_{18}\mathbf{m}5 \tag{24}$$

Here, grr accounts for the sum across the age-specific maternity rates, scaled by the length of each age interval (5 years).

If mortality is taken into account, the grr becomes the net reproduction rate (nrr), or the average number of daughters that a cohort of women would have if subjected to current materrnity and mortality rates throughout their lives. This rate is approximated by the probability of surviving to the mean age of the maternity function ($\mathbf{p}^w_{A_m}$):

$$nrr \approxeq \mathbf{p}^w_{A_m} grr \tag{25}$$

where $\mathbf{p}^w_{A_m}$ is the survival probability at the mean age of childbearing, approximated by the element of $\mathbf{p}^w$ closest to A_m.

The percentage change in total fertility rates observed for consecutive three-year periods in the last 15 years is calculated as

$$\mathbf{\Delta tfr}_j = \frac{\mathbf{tfr}_j - \mathbf{tfr}_{j+1}}{\mathbf{tfr}_j} 100 \tag{26}$$

where subscript $j = 1, 2...4$ indexes the column position of the elements in vectors $\mathbf{\Delta tfr}$ and $\mathbf{tfr}$. Thus the mean percentual change is

$$\overline{\Delta tfr} = \frac{\mathbf{\Delta tfr 1}_4^\intercal}{4} \tag{27}$$

where $\mathbf{1}^\intercal$ is a transposed column vector of ones, and the standard deviation of the temporal change in total fertility rates is

$$\sigma_{\mathbf{\Delta tfr}} = \sqrt{\frac{\sum(\mathbf{\Delta tfr}_j - \overline{\Delta tfr})^2}{4-1}} \tag{28}$$

The next six measures refer to the size of the female population and to the proportions of women who were "social" mothers, biological mothers, teenage mothers, mothers after age 45, and childless. Because these are simple percentages represented by ratios between women in specific conditions (e.g., biological mothers) and the total female population, there is no need to specify formulas for them. The last two summary measures, however, are the following:

$$c^u = \frac{\mathbf{c}^u \mathbf{1}_{16}^\intercal}{\mathbf{1}_{51} \mathbf{C} \mathbf{1}_{16}^\intercal} 100 \tag{29}$$

where c^u is the percentage of children with unidentified mothers within the population. And the approximate replacement level for the total fertility rate ($\widehat{tfr}$) is

$$\widehat{tfr} \approxeq \frac{1 + srb}{\mathbf{p}^w_{Am}} \tag{30}$$

These measures provide a comprehensive view of reproductive patterns and demographic shifts within a given population.

7 Extended examples

I use microdata from the 2010 Brazilian Census (Ruggles et al. 2024a) combined with survival probabilities estimated in Steps 1 and 2 of Section 5 (files `Lkids.dta` and `Lfemale.dta`) to illustrate the capabilities of `ocm`. In Stata, you can generate a table of three-year fertility rates by five-year age groups, a graph of age-specific fertility rates over time, and summary measures (using the `summary` option) by executing the following command:

```
. ocm [iw=perwt], p_children(Lkids.dta) p_female(Lfemale.dta) biological summary
```

Estimated three-year fertility rates, by five-year age group

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.0857473	.0806032	.069226	.0628264	.0583855
20-24	.1444407	.1320281	.1097815	.0987235	.0900545
25-29	.1241954	.1135419	.0966923	.0886688	.081921
30-34	.0833242	.076241	.0665445	.0640034	.0614972
35-39	.044946	.0417357	.0360891	.0341968	.0331808
40-44	.0187524	.015657	.0128764	.0117665	.0106866
45-49	.004971	.0042944	.0027362	.0021256	.0018413
TFR	2.531884	2.320507	1.96973	1.811555	1.687834

Note: These estimates consider 6.74e+07 women.
Summary measures

Measure	Value
Mean age of the population	32.1065
Mean age of childbearing	26.9143
Crude birth rate(CBR)	0.0143
Gross reproduction rate(GRR)	0.8304
Net reproduction rate(NRR)	0.8257
% change in TFR(mean)	9.5812
% change in TFR(sd)	3.7476
Female population	9.735e+07
% of social mothers *	38.7556
% of biological mothers	37.2750
% of mothers(10-19 years)	0.4941
% of mothers(after 45 years)	0.3149
% childless women	62.7250
% of unmatched children **	13.3579
Replacement level of TFR	2.0442

* Includes stepmother, biological, and adopted mothers.
** Children with unidentified mother in the current estimate.

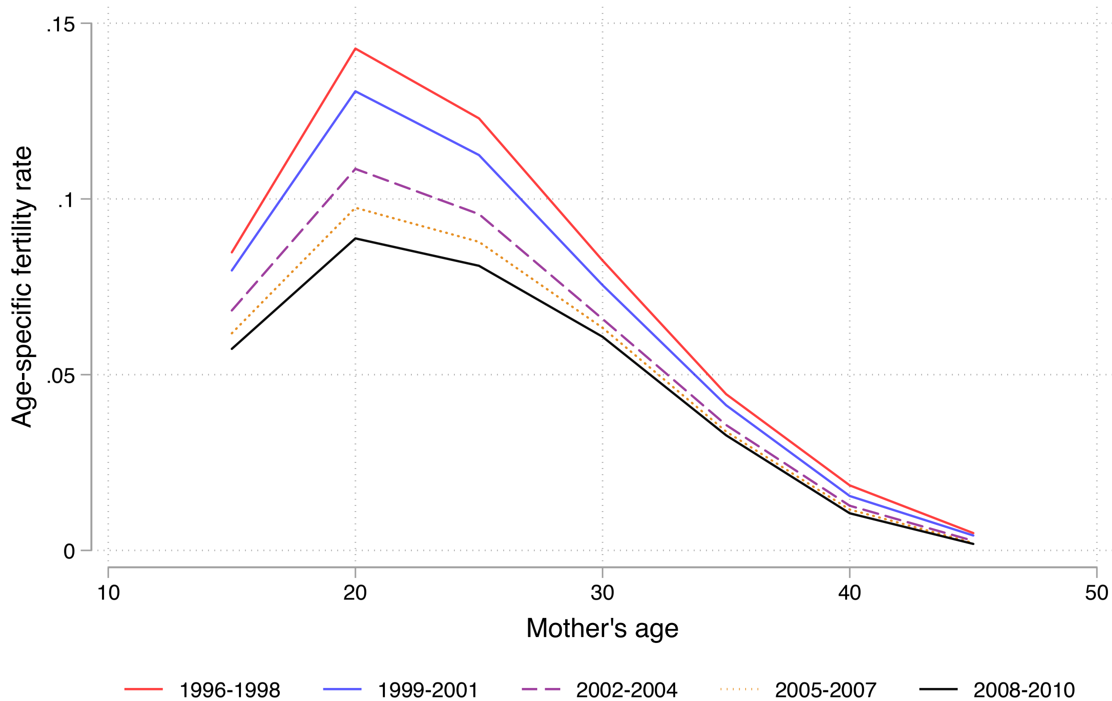

Figure 1: Age-specific fertility rates over time

The bottom of the first table shows a decrease in total fertility rate from 2.7172 to 1.7991 children per woman between 1996 and 2010. Figure 1 depicts the evolving age pattern and level of fertility over time. The summary measures table provides additional statistics characterizing the reproductive behavior of women in 2010. Notably, the total fertility rate declined at an average rate of 9.75% over the past 15 years, with a standard deviation of 3.02 percentage points between three-year periods. Moreover, the net reproduction rate was 0.8801 daughters per woman; 0.55% were teenage mothers; 0.34% of women became mothers after age 45; 62.31% of women had never had a child; and the estimated replacement level of fertility was 2.04, a threshold surpassed by the Brazilian population since the 2005-2007 period. The next section reveals the full potential of `ocm`, focusing on generating fertility estimates for different population subgroups.

7.1 Estimates for population subgroups: `by:` and `by()` options

The own-children method was originally developed to estimate fertility differentials among population subgroups (Grabill and Cho 1965). To illustrate the versatility of `ocm`, I use race-specific data from the 2010 Brazilian census and utilize the `by:` prefix and the `by()` option. First, evoking the `by:` prefix applies uniform survival probabilities across all subgroups:

```
. drop if race==30 | race==40 | race==99 // Drops Indigenous, Asian and unknown race
> s from the dataset.
(326,463 observations deleted)
. by race, sort: ocm [iw=perwt], p_children(Lkids.dta) p_female(Lfemale.dta) ///
>         biological nodraw
--------------------------------------------------------------------------------
-> race = white
Warning: bysort race: (or by race, sort:) assume that children and females in all
categories of race have the same survivorship probabilities, as presented in the
second column of files Lkids.dta.dta and Lfemale.dta.dta.
Estimated three-year fertility rates, by five-year age group
```

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.0656933	.0621505	.0539061	.0515144	.0514852
20-24	.1188535	.1101529	.0931235	.0872303	.0846161
25-29	.1112948	.1036539	.0914317	.0878638	.0851269
30-34	.0755213	.0714274	.0664034	.0677285	.0692097
35-39	.0369409	.0368001	.0342639	.0353326	.0368615
40-44	.0133596	.0117893	.0106207	.0107532	.0107266
45-49	.0031166	.0025901	.0018581	.0016675	.0015683
TFR	2.1239	1.99282	1.758038	1.710451	1.697972

```
Note: These estimates consider 3.28e+07 women.
--------------------------------------------------------------------------------
-> race = black
Estimated three-year fertility rates, by five-year age group
```

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.0775387	.0691464	.0546296	.0474448	.0400053
20-24	.125254	.1060494	.0841255	.0705858	.0563596
25-29	.1059857	.089866	.071571	.0589738	.0481831
30-34	.0745853	.0615404	.0478207	.0410652	.0352795
35-39	.0423464	.0362369	.0285884	.0229165	.0191704
40-44	.0190253	.0153288	.0114356	.0097056	.0068821
45-49	.0059639	.004514	.0030864	.0020416	.0015629
TFR	2.253497	1.913409	1.506286	1.263666	1.037215

```
Note: These estimates consider 5120910 women.
```

```
-> race = brown (b
Estimated three-year fertility rates, by five-year age group
```

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.1075563	.1011032	.0864082	.0751156	.066074
20-24	.1731568	.1578532	.1301474	.1136571	.098986
25-29	.1395587	.1264063	.1044814	.0927945	.0822612
30-34	.0926134	.0827312	.0681084	.0620067	.0558466
35-39	.055095	.0481477	.0386998	.0337883	.0303078
40-44	.025827	.0206747	.0158253	.0131427	.011044
45-49	.0074919	.0066255	.003864	.0027881	.0022835
TFR	3.006495	2.717709	2.237672	1.966466	1.734015

```
Note: These estimates consider 2.84e+07 women.
```

In this scenario, `ocm` assumes a single mortality schedule (from `Lkids.dta` and `Lfemale.dta`) for the three racial categories: White, Black, and Brown. This configuration models a hypothetical scenario where all racial groups share the same mortality as the general population, as stated in the output's headline. The results indicate notable fertility differentials; for instance, Blacks have historically exhibited the lowest fertility rates, ranging from 2.25 to 1.04 children per woman between 1996 and 2010, while Browns have shown the highest, ranging from 3.01 to 1.73 children per woman during the same period.

The second approach involves applying race-specific survival probabilities for children (`kids_race.dta`) and adult women (`female_race.dta`):

```
. ocm [iw=perwt], p_children(kids_race.dta) p_female(female_race.dta) ///
>         biological nodraw summary by(race)
Processing category: white
Estimated three-year fertility rates, by five-year age group
```

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.0655885	.0621276	.053904	.0514963	.0513741
20-24	.1187241	.1101568	.0931464	.0872151	.0844407

25-29	.1112381	.1037049	.0914858	.087868	.0849585
30-34	.0755313	.0714973	.0664654	.0677467	.0690797
35-39	.036971	.0368557	.0343087	.0353498	.0367955
40-44	.0133796	.0118137	.0106391	.0107614	.0107089
45-49	.0031234	.0025968	.0018621	.0016692	.0015659
TFR	2.122779	1.993764	1.759058	1.710533	1.694617

Processing category: black

Estimated three-year fertility rates, by five-year age group

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.0774475	.0690524	.0545661	.0473879	.039808
20-24	.1248893	.1057836	.0839741	.0704761	.0560728
25-29	.1054076	.0894659	.0713485	.0588414	.047926
30-34	.0739411	.0611196	.0475861	.0409256	.035074
35-39	.0418276	.0358892	.0283948	.022812	.0190478
40-44	.0187141	.0151344	.011333	.0096486	.0068345
45-49	.0058379	.0044388	.0030502	.0020261	.0015509
TFR	2.240326	1.904419	1.501264	1.260588	1.03157

Processing category: brown (brazil)

Estimated three-year fertility rates, by five-year age group

Age group	1996-1998	1999-2001	2002-2004	2005-2007	2008-2010
15-19	.1077551	.1011444	.086413	.0751586	.066316
20-24	.173361	.1578338	.1301055	.1136967	.0993401
25-29	.1396053	.126311	.1043984	.0927984	.0825458
30-34	.0925463	.0826057	.0680207	.0619905	.0560311
35-39	.0549847	.0480284	.0386235	.0337663	.0304036
40-44	.0257372	.0206	.0157814	.0131273	.0110758
45-49	.0074508	.0065929	.0038497	.0027831	.0022894
TFR	3.007201	2.715581	2.235961	1.966605	1.740009

Summary measures by race

Measure	race=10	race=20	race=51
Mean age of the population	33.6729	34.0101	30.0611
Mean age of childbearing	27.4096	26.6321	26.4548
Crude birth rate(CBR)	0.0137	0.0090	0.0153
Gross reproduction rate(GRR)	0.8363	0.4884	0.8560
Net reproduction rate(NRR)	0.8325	0.4844	0.8502
% change in TFR(mean)	5.3847	17.5906	12.7320
% change in TFR(sd)	4.7615	2.7274	3.4373
Female population	4.720e+07	6.911e+06	4.167e+07
% of social mothers *	37.3819	33.1082	40.6286
% of biological mothers	36.1442	31.5265	38.9482
% of mothers(10-19 years)	0.5170	0.4046	0.7300
% of mothers(after 45 years)	0.2505	0.3716	0.3829
% childless women	63.8558	68.4735	61.0518
% of unmatched children **	10.9426	18.4190	15.2782
Replacement level of TFR	2.0355	2.1296	2.0466

* Includes stepmother, biological, and adopted mothers.

```
** Children with unidentified mother in the current estimate.
```

This method highlights that the Own-Children Method is relatively insensitive to variations in mortality rates. Despite differences in life expectancy at birth – 72.57 years for Whites, 68.02 for Browns, and 66.96 for Blacks as noted by Muniz et al. (2024) – fertility estimates are similar to those generated under the uniform mortality asumption. This similarity suggests that, within the Brazilian context, mortality differences among racial groups do not significantly influence race-specific fertility estimates, reinforcing the robustness of OCM in demographic research.

Finally, the table titled `Summary measures by race` provides valuable insights and facilitates comparisons of reproductive dynamics among Whites (`race=10`), Blacks (`race=20`), and Browns (`race=51`). It reveals that Blacks experienced the steepest decline in total fertility over the last 15 years, with a percentual change of 17.59%. Additionaly, the table highlights that Blacks have the lowest net reproduction rate at 0.48 daughters per woman. This group also shows the highest percentage of childless women (68.47%) and largest proportion of unmatched children (18.42%). These comparative results are particularly important because they contrast those reported by United Nations Population Fund (2018), which suggests that Whites had the lowest fertility rates since the early 2000s. This discrepancy underscores the influence that differences in data sources and methodological approaches can have on fertility estimations.

8 The `ocm` command

8.1 Syntax

`ocm` [`if`] [`weight`], `p_children(`*string*`)` `p_female(`*string*`)` [`by(`*groupvar*`)` `biological` `summary` `nodraw`]

8.2 Description

The `ocm` command calculates age-specific fertility rates and reproductive measures, leveraging the methods originally introduced by Grabill and Cho (1965) and subsequently refined by Cho et al. (1986). The program applies reverse-survival techniques on cross-sectional census or survey data from a single year to compute three-year fertility rates across five-year age groups over the previous 15 years. It offers several functionalities: the `by(varname)` option facilitates the analysis of fertility rates within subgroups identified by a categorical variable; the `biological` option restricts analysis to biological mothers; and the `summary` option generates demographic indicators that describe the reproductive profile of the current population. To effectively implement the Own-Children Method for fertility estimation, three data inputs are necessary:

1. a data extract (preferably from IPUMS) containing, as a minimum, eight variables:
   - the year of data collection (`year`),
   - the household serial number (`serial`),
   - the identifier for individuals within the household (`pernum`),
   - position of the mother within the household (`momloc`),
   - a variable to restrict the analysis to biological mothers (`stepmom`),
   - age of the individual (`age`),
   - age categorized in five-year intervals (`age2`),
   - sex of the individual (`sex`).

   For analysis involving weights, include the person's weight (`perwt`). For subgroup analyses, also include a relevant categorical variable in the extracted dataset, such as `race, religion`, occupation, education, or geographical location;

2. a file with survival probabilities for children at single year ages from 0 to 15; and

3. a file containing survival probabilities for women at single year ages from 15 to 65.

The datasets must conform to the structure and variable ordering outlined in section 3 to ensure the correct operation of the `ocm` functionalities. Proper preparation and configuration of these data inputs are crucial for the effective application of the Own-Children Method.

8.3 Options

`p_children(`*string*`)` specifies the file containing the survival probabilities for children. This file should include two columns: the first column indicates the child's age, and the second the probabilities of surviving from age x to $x + 1$. This option is mandatory.

`p_female(`*string*`)` specifies the file containing the survival probabilities of women. It includes two variables: the age of the women and the probabilities of surviving between ages x and $x + 1$. The `p_female( )` option is mandatory.

`by(`*groupvar*`)` specifies the categorical variable that defines the subgroups for analysis. This option allows the calculation of fertility rates across different demographic groups.

`biological` directs `ocm` to compute fertility rates exclusively for children linked to their biological mothers. Without this option, the program considers all maternal relationships, including social ones such as stepmothers and adoptive mothers.

`summary` generates a set of 15 measures reflecting the current population's reproductive behavior. These indicators include crude birth rates, gross and net reproduction rates, mean age at childbearing, the proportion of teenage mothers, the proportion of mothers over age 45, the percentage of childless women, and the rate of change in total fertility over the past 15 years, among others.

`nodraw` prevents `ocm` from automatically displaying a graph of age-specific fertility rates. By *default*, `ocm` displays a graph, and if the `by()` option is specified, it saves individual category-specific graphs as `gr`*i*`.gph` files in the working directory before combining them in a new graph named `gr_combine.gph`.

8.4 Output

As a minimum, `ocm` displays a table and a graph of age-specific and total fertility rates for women at reproductive ages (15 to 49 years old). The results show patterns, levels, and trends of fertility over the past 15 years.

Additionally, when enabled through the `summary` option, `ocm` generates key demographic measures for either the general population or specific subgroups defined by the `by(`*groupvar*`)` option. These measures provide deeper insights into the reproductive behavior and characteristics of the population or the targeted subgroups.

8.5 Stored results

After execution, the `ocm` command stores computed age-specific and total fertility rates, summary demographic measures, and other relevant variables in system matrices. These matrices are accessible under the following names:

Stata matrices

`r(asfr)`	Returns a matrix with age-specific fertility rates, averaged over three years and categorized by five-year age groups from the past 15 years
`r(tfr)`	Returns a row-vector of total fertility rates calculated over the past 15 years
`r(asfr_`*i*`)`	For subgroups specified by the `by(`*groupvar*`)` option, this matrix contains age-specific fertility rates, averaged over three years, sorted into five-year age groups for each category *i* of *groupvar* from the past 15 years
`r(tfr_`*i*`)`	Returns a row-vector with total fertility rates over the last 15 years for each subgroup *i*, as defined by the `by(`*groupvar)* option
`r(summary)`	Specifying the `summary` option returns a matrix with 15 demographic measures characterizing the reproductive pattern and level of the current population
`r(summary_by)`	Returns a matrix with 15 demographic measures for each subgroup *i*, specified by the `by(`*groupvar*`)` option

Mata matrices

`asfr`	Matrix of single age-specific fertility rates for mothers aged 15 to 49 years between times $t - 15$ and t
`asfr5`	Matrix of five-year age-specific fertility rates for mothers aged 15 to 49 years between times $t - 15$ and t
`m`	Maternity rates matrix, detailing female births by women aged 15 to 49 years
`p`	Matrix of population counts, structured by five-year age groups
`women`	Matrix of the number of reverse-survived women by single year age categories over the last 15 years
`children`	Matrix detailing the number of reverse-survived children, structured by the single year age of the mother over the past 15 years
`asfr`*i*	Matrix of age-specific fertility rates by single years for mothers aged 15 to 49, calculated for each subgroup *i*, as defined in the `by(`*groupvar)* option, from $t - 15$ to t
`asfr5`*i*	Matrix of age-specific fertility rates by five-year age groups for mothers aged 15 to 49, calculated for each subgroup *i* from $t - 15$ to t, as defined in the `by(`*groupvar)* option
`mat`*i*	Maternity rates matrix for each subgroup *i*, detailing female births by women aged 15 to 49 years, as categorized by the `by(`*groupvar)* option
`p`*i*	Matrix of population counts for each subgroup *i*, structured by five-year age groups, as defined in the `by(`*groupvar)* option
`women`*i*	Matrix of the number of reverse-survived women by single year age categories for each subgroup *i* over the last 15 years, as defined in the `by(`*groupvar)* option
`children`*i*	Matrix detailing the number of reverse-survived children for each subgroup *i*, structured by the single year age of the mother over the last 15 years, as defined in the `by(`*groupvar)* option

9 Conclusions

This article highlights the capabilities of `ocm`, a Stata program designed to estimate trends and levels of fertility using the Own-Children Method (Cho et al. 1986; Grabill and Cho 1965). The software not only generates graphs of age-specific fertility rates but also produces estimates tailored to different population subgroups by accounting for their category-specific mortality rates. Additionally, it computes 15 summary measures that characterize the reproductive patterns of the current population. Ideally, this program should be used with IPUMS data, which provides the standardized census variables necessary for implementing the procedure. However, microdata from sources such as the Demographic and Health Surveys, the World Fertility Survey, Multiple Indicator Cluster Surveys (MICS), and the Pan Arab Program Project for Family Health (PAPFAM) can also be used to construct equivalent variables required for fertility estimation using the Own-Children Method (Schoumaker 2019).

`ocm` mitigates potential fluctuations and biases caused by age misreporting by averaging single-year fertility rates into three-year periods for five-year age groups. Additionally, the program enhances the accuracy of fertility estimates by offering an option to limit the calculation of fertility rates to biological mothers, rather than social mothers. This adjustment prevents the erroneous allocation of adopted and stepchildren to women, thereby refining the precision of the fertility estimates.

The analysis of fertility patterns and trends among Brazil's racial groups (Whites, Browns, and Blacks) over the last 15 years demonstrates validity but is not completely immune to the violations of two critical assumptions of the Own-Children Method (OCM): mortality constancy and the stability of categorical membership. The application of a single life table across the period under study inherently assumes stable mortality rates, a presumption that is unlikely given Brazil's overall demographic changes and specific variations across racial groups. However, evidence indicates that OCM can still provide reliable results even when actual survival rates differ significantly from those observed 15 years ago (Spoorenberg 2014).

Concerns about the stability of racial categorization are particularly relevant in Brazil, where racial reclassification is a documented phenomenon (Miranda 2015; Muniz 2020; Muniz and Bailey 2022). Research by (Muniz et al. 2024, 463) indicates that, on average, a Brazilian had a 23% chance of reclassifying from Black to Brown between 2017 and 2019. Over extended periods, the likelihood of such transitions increases, potentially compromising the constancy of membership assumption. According to (Rindfuss 1977, 14), the most significant impacts on fertility estimates arise when the assumption of constant group membership is breached by substantial inflows or outflows of individuals who differ markedly from those staying or departing. However, his analysis showed minimal effects, a finding echoed by Muniz and Bailey (2022), who reported limited influence of racial shifting on racial inequality outcomes in Brazil. The extent and nature of demographic changes necessary to substantially alter fertility estimates remain questions for further research, indicating a need to investigate how these assumption departures might affect fertility patterns, levels, and trends in different settings.

Despite its inherent limitations and assumptions, the Own-Children Method continues to be a trustworthy, valid, and informative approach for estimating the fertility of population subgroups and in regions where vital registration systems are lacking or deficient. With the introduction of `ocm`, this method has now also become more accurate, appealing and efficient, cementing its relevance in demographic research and beyond.

10 References

Abbasi-Shavazi, M. J. 1997. An assessment of the own-children method of estimating fertility by birthplace in Australia. *Journal of the Australian Population Association* 14(2): 167–185.

Abbasi-Shavazi, M. J., M. Hosseini-Chavoshi, F. Banihashemi, and A. Khosrvi. 2013. Assessment of the own-children estimates of fertility applied to the 2011 Iran Census and the 2010 Iran-MIDHS.

Avery, C., T. S. Clair, M. Levin, and K. Hill. 2013. The 'own children' fertility estimation procedure: a reappraisal. *Population Studies: a journal of demography* 67(2): 171–183.

Cho, L.-J., R. D. Redford, and M. K. Choe. 1986. *The own-children method of fertility estimation*. 1st ed. Honolulu, HI: Population Institute.

Dubuc, S. 2009. Application of the own-children method for estimating fertility by ethnic and religious groups in the UK. *Journal of Population Research* 26: 207–225.

Feeney, G. 1975. The Own-Children Method of Estimating Age-Specific Fertility Rates: Lecture Notes Prepared for a Demographic Analysis and Data Evaluation Workshop. http://demographer.com/gfeeney/teaching-notes/1975-own-children-notes/1975-own-children-notes.pdf.

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, D. Backman, A. Chen, G. Cooper, S. Richards, M. Schouweiler, and M. Westberry. 2023. Integrated Public Use Microdata Series CPS. https://doi.org/10.18128/D030.V11.0.

Grabill, W. R., and L.-J. Cho. 1965. Methodology for the measurement of current fertility from population data on young children. *Demography* 2(1): 50–73.

Hacker, J. D. 2003. Rethinking the "early" decline of marital fertility in the United States. *Demography* 40(4): 605–620. https://doi.org/10.2307/1515199.

Hattersley, L., and R. Creeser. 1995. Longitudinal study 1971-1991: history, organisation and quality of data. Technical report, Office of Population Censuses and Surveys.

Hunter, W., and N. B. Sugiyama. 2018. Making the Newest Citizens: Achieving Universal Birth Registration in Contemporary Brazil. *The Journal of Development Studies* 54(3): 397–412. https://doi.org/10.1080/00220388.2017.1316378.

IBGE. 2024. Sidra: Banco de Tabelas Estatísticas [eletronic resource]. https://sidra.ibge.gov.br/Tabela/9606.

Jann, B. 2016. Creating LaTeX documents from within Stata using texdoc. *Stata Journal* 16(2): 245–263. http://repec.sowi.unibe.ch/stata/texdoc.

Krapf, S., and M. Kreyenfeld. 2015. Fertility assessment with the own-children method: a validation with data from the German mikrozensus. Technical report, Max Planck Institute for Demographic Research, Rostock, Germany.

Lima, E. E. C., B. L. Queiroz, and K. Zeman. 2018. Completeness of birth registration in Brazil: an overview of methods and data sources. *Genus* 74(11). https://doi.org/10.1186/s41118-018-0035-9.

Merli, G. 1998. Mortality in Vietnam, 1979-1989. *Demography* 35(3): 345–360.

Ministério da Saúde. 2024. Departamento de Informática do SUS – DATASUS. Coordenação Geral de Informações de Saúde[eletronic resource]. http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt10uf.def.

Miranda, V. 2015. A ressurgence of black identity in Brazil? Evidence from an analysis of recent censuses. *Demographic Research* 32(59): 1603–1630. https://doi.org/10.4054/DemRes.2015.32.59.

Miranda-Ribeiro, A. 2007. Reconstrução de Histórias de Nascimentos a partir de Dados Censitários: aspectos teóricos e evidências empíricas.

Muniz, J. O. 2012. Demographic dynamics of poverty and income inequality: the case of Brazil. *Revista brasileira de estudos de população* 29(2): 323–348. https://doi.org/10.1590/S0102-30982012000200007.

———. 2020. Multistate life tables using Stata. *The Stata Journal* 20(3): 721–745.

———. 2023. Iterative intercensal single-decrement life tables using Stata. *The Stata Journal* 23(3): 813–834.

Muniz, J. O., and S. R. Bailey. 2022. Does race response shift impact racial inequality? *Demographic Research* 47(30): 935–966. https://doi.org/10.4054/DemRes.2022.47.30.

Muniz, J. O., A. Saperstein, and B. L. Queiroz. 2024. Racial classification as a multistate process. *Demographic Research* 50(17): 457–472. https://doi.org/10.4054/DemRes.2024.50.17.

Myers, R. J. 1940. Errors and bias in the reporting of ages in census data. *Transactions of the Actuarial Society of America* 41(104): 395–415.

———. 1954. Accuracy of age reporting in the 1950 United States census. *Journal of the American Statistical Association* 49(268): 826–831. https://doi.org/10.2307/2281542.

Preston, S. H., P. Heuveline, and M. Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers.

Retherford, R. D., and I. Alam. 1985. Comparison of fertility trends estimated alternatively from birth histories and own children. *Papers of the East-West Population Institute* 94.

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(3): 567–581.

Rindfuss, R. R. 1976. Annual fertility rates from census data on own children: comparisons with vital statistics data fro the United States. *Demography* 13(2): 235–249.

———. 1977. Methodological difficulties encountered in using own-children data: illustrations from the United States. *Papers of the East-West Population Institute* 42.

Ruggles, S., L. Cleveland, R. Lovaton, S. Sarkar, M. Sobek, D. Burk, D. Ehrlich, Q. Heimann, and J. Lee. 2024a. Integrated Public Use Microdata Series, International: version 7.5 [dataset]. https://doi.org/10.18128/D020.V7.5.

Ruggles, S., S. Flood, M. Sobek, D. Backman, A. Chen, G. Cooper, S. Richards, R. Rodgers, and M. Schouweiler. 2024b. Integrated Public Use Microdata Series, USA: version 15.0 [dataset]. https://doi.org/10.18128/D010.V15.0.

Schoumaker, B. 2017. Measuring male fertility rates in developing countries with demographic and health surveys: an assessment of three methods. *Demographic Research* 36(28): 803–850.

———. 2019. Male fertility around the world and over time: how different is it from female fertility? *Population and Development Review* 45(3): 459–487.

Sobek, M., and S. Kennedy. 2009. The Development of Family Interrelationship Variables for International Census Data.

Spoorenberg, T. 2014. Reverse survival method of fertility estimation: an evaluation. *Demographic Research* 31(9): 217–246. https://doi.org/10.4054/DemRes.2014.31.9.

Timaeus, I. 2013. Indirect estimation of adult mortality from orphanhood. In *Tools for demographic estimation*, ed. T. Moultrie, R. Dorrington, A. Hill, K. Hill, I. Timaeus, and B. Zaba, 222–243. Paris: International Union for the Scientific Study of Population. https://demographicestimation.iussp.org/content/indirect-estimation-adult-mortality-orphanhood.

Timaeus, I. M. 2021. The own-children method of fertility estimation: the devil is in the detail. *Demographic Research* 45(25): 825–840. https://doi.org/10.4054/DemRes.2021.45.25.

UNICEF. 2023. UNICEF global databases: based on DHS, MICS, other national surveys, censuses and vital registration systems. https://data.unicef.org.

United Nations. 1983. Manual X: Indirect techniques for demographic estimation. Technical report, Department of International Economic and Social Affairs.

United Nations Population Fund. 2018. Fecundidade e dinâmica da população brasileira. Technical report, Fundo de População das Nações Unidas.

Wong, L. L. R. 1983a. Fecundidade no Brasil (urbano e rural): aplicação do Método dos Filhos Próprios ao Censo de 1970. *Informe Demográfico* 1(9): 53–98.

———. 1983b. Níveis e tendências da fecundidade nas diversas regiões do Brasil: aplicação do Método dos Filhos Próprios ao Censo de 1970. *Informe Demográfico* 1(9): 99–140.