

Reconstructing long term fertility trends with pooled birth histories

Bruno SCHOUMAKER, Centre de recherche en démographie et sociétés,
Université catholique de Louvain (Belgium).

1. Introduction

The accurate measurement of past fertility trends has important practical, theoretical and policy implications. Assumptions about future fertility in population projections are largely influenced by past fertility trends. The evaluation of population policies is also (notably) based on the reliable measurement of fertility trends. The measurement of the timing and of the speed of fertility decline is also central to theories of fertility changes.

In developing countries - where civil registration systems are deficient - fertility trends are often inferred with a few estimates from censuses and surveys (United Nations, 2011). Such an approach tends to mask changes that occurred between two data points, especially when there are few estimates. Fertility trends computed in this way may also be largely influenced by varying data quality between surveys (Schoumaker, 2010).

In this paper, we will present a method for reconstructing and smoothing fertility long term fertility trends by combining birth histories from multiple surveys, using Poisson regression and restricted cubic splines. Data used in the paper come from World fertility surveys and Demographic and Health Surveys¹. In the first part, we present the method and illustrate its application by combining several fertility surveys in Colombia. Next, we will use simulated birth histories to validate the method in a variety of situations (only one presented in this abstract). Finally, we apply the method to several countries, from various parts of the world, with varying numbers of surveys and with different data quality issues. Finally, these trends will be compared with published trends (United Nations Population Division and DHS) (not presented in this abstract).

¹ Data from any other survey with birth histories (such as some MICS) can also be used.

2. Reconstructing fertility trends from birth histories – a review of approaches

DHS reports typically publish age-specific fertility rates by 5-year periods before the survey that allow reconstructing partial TFRs (sometimes computed in reports). Garenne and Joseph (2002) used DHS birth histories to compute births and exposure by age groups and calendar years on pooled data from several surveys, and estimated total fertility rates over long periods in several countries. They smooth TFRs by fitting polynomials on the estimated TFRs. More recently, Machiyama (2010) computed partial TFRs from birth histories in several sub-Saharan African countries, and used Loess techniques to smooth fertility trends. In addition, she corrected displacements of recent births by transferring some births from one year to the preceding year before computing rates².

The approach presented in this paper is a generalization of an approach used by the author in previous research (Schoumaker, 2004; Schoumaker 2010; Schoumaker, 2012). It uses Poisson regression to reconstruct fertility trends from several surveys pooled together. The major difference with Garenne and Joseph (2002), and Machiyama's work (2010) is to use a regression framework to reconstruct fertility trends. This allows estimating TFRs between 15 and 49 with truncated data in a straightforward way (with a few assumptions), whereas other authors estimate TFRs between 15-34. Our method also allows smoothing fertility trends in a straightforward way by including appropriate variables in the regression (e.g. with splines), and allows testing for changes in speed of fertility decline. Data quality problems can also to some extent be corrected including appropriate variables in the regression model. Finally, our approach is validated by using micro-simulation.

3. Method

The method consists in reconstructing trends in total fertility rates (15-49) with Poisson regression, and smoothing trends using restricted cubic splines. It is first described with data from a single survey; next, it is presented with several surveys that are pooled together.

3.1. Fertility data from a birth history

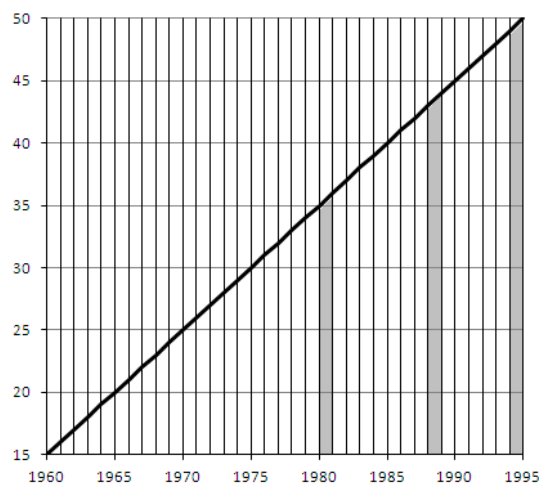
The basic principle can be best illustrated with a Lexis diagram showing the data that are obtained from a birth history (figure 1). A typical fertility survey collects

² Other recent research on the reconstruction of fertility trends includes Alkiema et al. (2012).

birth histories among women aged 15-49 at the time of the survey. In this example, the survey is conducted on the 31st of December 1994. The corresponding data needed to compute fertility rates is shown in Table 1. Data come from the 1995 DHS in Colombia.

Using data from the full birth history, it is possible to compute the number of births and the total exposure (number of women-years) for each year and each age-group below the black diagonal³. Computing fertility rates thus necessitates producing a table containing for each year and each age group, the number of births and the total exposure. Each cell in the following table corresponds to a rectangle in the Lexis diagram. As is clear from this table and the Lexis diagram, some data are missing because of the truncation.

Figure 1: Lexis diagram illustrating birth history data.



For the year preceding the survey (1st year, or year 15), births and exposure are available for all age groups. In year 9 (7th year before the survey), no data is available for the age group 45-49, and the age group 40-44 is only partially covered. In the 15th year before the survey (1980), births and exposure are complete for the first four age groups (15-19, 20-24, 25-29, 30-34), and a very small part of the 35-39 age group is observed (9 births, 78 years of exposure). No data is available for age groups 40-44 and 45-49.

Using classical demographic methods, the total fertility rates are obtained by computing age-specific fertility rates in each of the seven 5-year age groups,

³ A stata command (tfr2) was produced by the author

summing these age-specific fertility rates⁴ and multiplying them multiplied by 5. For the last year, the TFR can be computed between 15 and 49. As one goes back in time, the age range becomes more limited. For instance, the TFR for 1988 can only be obtained for women 15-44, and the rate for this last age group will be slightly biased, since no information is available for women aged 44 in 1988. In 1979, the TFR can only be computed among women aged 15-34.

Table 1. Table of births and exposure from a retrospective birth history, Data from 1995 Colombia DHS

Period	Year before survey	Age-group	Births	Exposure (years)
15	1 st	15-19	181	2099
15	1 st	20-24	326	1937
15	1 st	25-29	286	1822
15	1 st	30-34	151	1578
15	1 st	35-39	61	1385
15	1 st	40-44	26	1192
15	1 st	45-49	2	848
...				
9	7 th	15-19	189	1937
9	7 th	20-24	325	1822
9	7 th	25-29	222	1578
9	7 th	30-34	165	1385
9	7 th	35-39	75	1192
9	7 th	40-44	21	848
...				
0	15 th	15	162	1652
0	15 th	20	261	1417
0	15 th	25	285	1247
0	15 th	30	150	985
0	15 th	35	9	78

When reconstructing TFRs from birth histories, the usual approach consists in computing truncated TFRs. For instance, the TFR for the 10 years preceding the survey would be computed over the age range 15-39.

3.2 Reconstruction of TFRs from a single survey with Poisson regression

The method consists in estimating TFRs by using Poisson regression with age groups and years as independent variables (Schoumaker, 2004; Schoumaker, 2012). Poisson regression can be used with individual data or grouped data, leading to strictly identical results (Powers and Xie, 2000). Because it is computationally less

⁴ The computation of age-specific fertility rates consist in dividing the number of births in a rectangle of the Lexis diagram by the exposure in that rectangle.

intensive, using grouped data allows fitting the models more quickly. For this reason, we use grouped data like in Table 1⁵.

The model we use can be summarized using this equation.

$$\log(\mu_{it}) = \log(e_{it}) + f(\text{age}) + g(\text{time}) \quad [\text{Eq. 1}]$$

μ_{it} is the expected number of births (births column) at each age (i) and each period (t), e_{it} is the total exposure at each age and each period (exposure column), $f(\text{age})$ is a function of age, and $g(\text{time})$ is a function of the calendar time. The term $\log(e_{it})$ is the *offset*, and has a fixed coefficient equal to one.

This equation can be reorganized by dividing the number of births by the exposure.

$$\log\left(\frac{\mu_{it}}{e_{it}}\right) = \log(\lambda_{it}) = f(\text{age}) + g(\text{time}) \quad [\text{Eq. 2}]$$

$$\lambda_{it} = \exp[f(\text{age})] * \exp[g(\text{time})] \quad [\text{Eq. 3}]$$

λ_{it} is the fertility rate at age i and period t, and is equal to the product of an age effect and a time effect. This log-rate model is estimated with Poisson regression (Powers and Xie, 2000).

An assumption made here is that the shape of the age-specific fertility rates is constant over time, i.e. that there is no interaction between age effects and time effects. In other words, time has a multiplicative effect on fertility rates that is similar at all ages. Comparisons with trends in partial TFRs between 15-34 and simulations indicate that this assumption does not have a strong influence on fertility trends over a relatively short period of time (10-15 years). As discussed later, the assumption of a constant age fertility schedule can be relaxed by using changing age schedules in the offset. This is what is done when several surveys are pooled together.

The following model illustrates this method for reconstructing fertility over the 15 years preceding the 1995 DHS in Colombia⁶. In this example, age is included as a set of dummy variables for five-year age groups (Table 2). Calendar time is

⁵ However, using individual data leads to strictly identical results, and may be more flexible in some applications. See Schoumaker (2004) for the presentation of the method with individual data.

⁶ Although the method can be used to reconstruct fertility trends from a single survey over a longer period than 15 years, possible omissions of births in the past and smaller sample sizes, as well as the departure from the assumption of constant age fertility schedule lead us to limit the period to the last 15 years.

measured by dummy variables to model annual variations in fertility. The model is written in the following way.

$$\log(\mu_{it}) = \log(e_{it}) + \alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{kit} + \sum_{h=2}^{15} \delta_h \cdot T_{hit} \quad [\text{Eq. 4}]$$

Alternatively, the rate can be expressed with this equation.

$$\lambda_{it} = \exp \left[\alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{kit} \right] * \exp \left[\sum_{h=2}^{15} \delta_h \cdot T_{hit} \right] \quad [\text{Eq. 5}]$$

α is the constant term, A_{kit} are dummy variables for the 6 age groups from 20-24 to 45-49, and T_{hit} are 14 dummy variables for the years after the reference year.

Predicting the fertility rate for a single age group (e.g. age group 25-29) for a specific year (e.g. year 5) is straightforward. The dummy variables are equal to 1 for the specific age group and year (and 0 for the other age groups and years), and the rate is just a function of the constant, the regression coefficient for the 25-29 age group, and the regression coefficient for the 5th year dummy variable.

$$\lambda_{it} = \exp[\alpha + \beta_{25-29}] * \exp[\delta_5] \quad [\text{Eq. 6}]$$

The total fertility rate (15-49) for year t is equal to 5 times the sum of age-specific fertility rates, multiplied by the exponential of the regression coefficient of the dummy variable for year h.

$$TFR_h = 5 * \left(\exp[\alpha] + \sum_{k=20-24}^{45-49} \exp[\alpha + \beta_k] \right) * \exp[\delta_h] \quad [\text{Eq. 7}]$$

Regression coefficients of the Poisson regression are reported in the second column of Table 1. From these regression coefficients, age-specific fertility rates for the reference year (1) are computed. The fertility rate for a specific age group (column 4) is equal to the exponential of the sum of the constant of the model and of the coefficient of the age group (the sum of the coefficients are log(rates), in column 3). The sum of the age-specific fertility rates for all the age groups multiplied by 5 is equal to the TFR for the reference year (1st year of the 15 year period, equal to 3.90 in Table 1). The TFRs for the following years are obtained by multiplying the TFR of the reference year by the exponentials of regression coefficients of the following years. Figure 2 shows the annual variations of the TFRs (same values as in Table 1), with 95% confidence intervals. Standard errors for the TFR are computed using

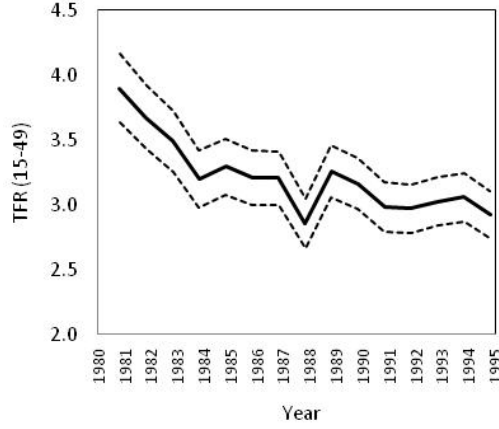
the delta method⁷. As will be shown later, the trend can be smoothed with restricted cubic splines.

Table 2: Age specific fertility rates and reconstruction of fertility trends over the fifteen calendar years preceding the 1995 DHS survey in Colombia with Poisson regression.

Independent variables (1)	Regression coefficients (2)	Log(rates) ($\alpha+\beta$) (3)	Rates $\exp[\alpha + \beta]$ (4)	TFR (5)
Constant	<i>A</i> -2.212			
	<i>B</i>			
15-19	(REF)	-2.212	0.110	
20-24	0.669	-1.543	0.214	
25-29	0.559	-1.652	0.192	
30-34	0.248	-1.963	0.140	
35-39	-0.234	-2.451	0.086	
40-44	-1.183	-3.395	0.034	
45-49	-3.116	-5.327	0.005	
	δ			
Year 1	(REF)			3.90
Year 2	-0.060			3.67
Year 3	-0.111			3.49
Year 4	-0.197			3.20
Year 5	-0.169			3.29
Year 6	-0.194			3.21
Year 7	-0.196			3.21
Year 8	-0.311			2.86
Year 9	-0.180			3.26
Year 10	-0.209			3.16
Year 11	-0.268			2.98
Year 12	-0.272			2.97
Year 13	-0.253			3.03
Year 14	-0.243			3.06
Year 15	-0.287			2.93

⁷ A Stata command (called tfr2) was prepared by the author for computing TFRs and reconstructing fertility trends from a single survey (see Schoumaker, 2012).

Figure 2: Reconstruction of TFRs (15-49) and 95% confidence intervals over the 15 years preceding the survey, Colombia 1995 DHS.



Another equivalent way of estimating the model is to control the age pattern of fertility in the offset (controlling both for exposure and age pattern of fertility). The model is written in the following way, where a_{it} measures the age pattern of fertility.

$$\log(\mu_{it}) = \log(e_{it}) + \log(a_{it}) + \gamma + \sum_{h=2}^{15} \delta_h \cdot I_{hit} \quad [\text{Eq. 8}]$$

Comparing equation [8] with equation [4], we see that

$$\log(a_{it}) = \alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{kit} - \gamma \quad [\text{Eq. 9}]$$

and

$$a_{it} = \frac{\exp\left[\alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{kit}\right]}{\exp[\gamma]} \quad [\text{Eq. 10}]$$

The numerator of this expression is the fertility rate for the reference year. The denominator is the constant of the second model. By constraining the sum of a_{it} over the 7 age groups to be equal to 1, a_{it} are proportionate fertility rates. Then

$$\exp[\gamma] = \sum_{k=20-24}^{45-49} \exp[\alpha + \beta_k] \quad [\text{Eq. 11}]$$

$$a_{it} = \frac{\exp\left[\alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{kit}\right]}{\sum_{k=20-24}^{45-49} \exp[\alpha + \beta_k]} \quad [\text{Eq. 12}]$$

The total fertility rate is computed as:

$$TFR_h = 5 * \exp[\gamma] * \exp[\delta_h] \quad [\text{Eq. 13}]$$

In other words, if the age pattern of fertility is controlled in the offset, the TFR for the reference year is the exponential of the constant multiplied by five; the TFR for the other years are obtained by multiplying the TFR of the reference year by the exponential of dummy variables for the year h.

In summary, this alternative approach requires first estimating the age-specific fertility rates for the reference year and dividing these rates by their sum to obtain proportionate age-specific rates. These proportionate rates are then included in the offset⁸, and a new model is estimated. Although controlling the age pattern of fertility in the offset has no interest when working with a single survey – and involves running two models leading to identical fertility trends – it proves useful when several surveys are pooled together⁹. The age pattern can be estimated for each survey separately, and included in the offset when surveys are pooled together in order to relax the assumption of constant age-schedule.

In the next section, the approach is extended to be used with several birth histories, and trends are smoothed with restricted cubic splines.

3.3 Reconstruction of TFRs from several surveys with Poisson regression

The method we just described can be applied to a pooled data set of birth histories from several surveys. We first illustrate it in the case of Colombia with 7 surveys (1976 WFS, 1986 DHS, 1990 DHS, 1995 DHS, 2000 DHS, 2005 DHS, 2010 DHS)¹⁰.

Before pooling the surveys, fertility was reconstructed over the last 15 years for each of the seven surveys separately (Figure 3), in the same way as in section

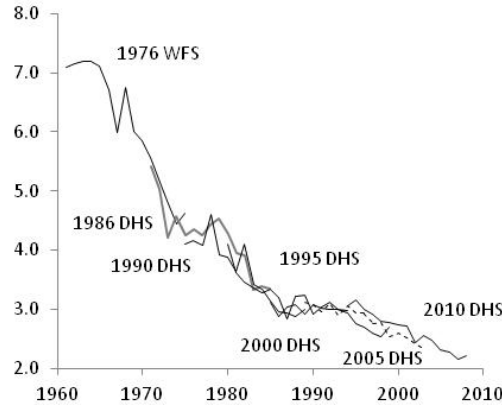
⁸ Multiplying exposure by the proportionate rate and taking its logarithm.

⁹ It is also possible to include an age pattern from another source – this is not discussed in this presentation.

¹⁰ The situation of Colombia is to some extent exceptional, because of the large number of surveys, their relatively good quality, and their large sample sizes.

3.2. The seven surveys match quite well, and show a downward trend, punctuated by slowdowns in the 1970s and 1990s.

Figure 3: Reconstruction of TFRs (15-49) over the 15 years preceding seven surveys, Colombia.



The next step consists in pooling the seven data sets, and reconstructing the fertility trend using Poisson regression. Pooling the data sets consists in preparing a data set similar as in Table 1 for each survey, and appending the seven data sets (seven surveys) to for a single data file. In order to compute fertility rates by calendar year, the datasets for each survey are prepared for the fifteen calendar years preceding each survey.

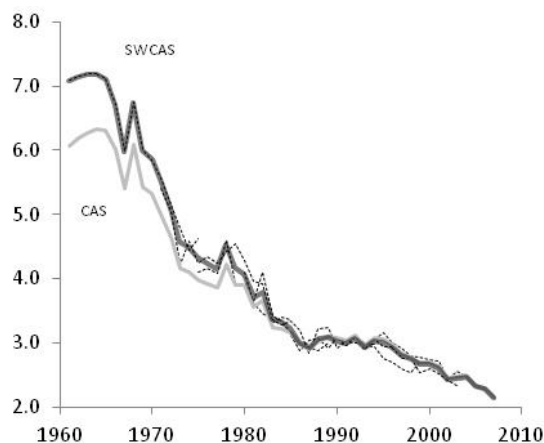
Assuming a constant age pattern of fertility over a 50 year period may lead to biased fertility trends. If fertility decreases more quickly among older women, this would lead to underestimating the TFR in the past¹¹. The assumption of constant age schedule can be relaxed by controlling for the age pattern in the offset, and allowing the age pattern to vary over time. In this paper, this is done by considering a survey-specific age schedule. For each survey, the proportionate age-specific fertility rates are computed (see previous section) and included in the offset. The age pattern is considered constant for each survey, but variable across surveys. Although this assumption does not strictly hold, simulations show it performs well¹².

¹¹ This is because rates above 35 are reconstructed from the model, and are assumed to have declined at the same pace as the fertility rates at young ages (the opposite would be true if fertility declined more rapidly at younger ages).

¹² Since birth histories from several surveys overlap, the assumption of constant age schedule for each survey does not mean the age schedule is constant over time between for each period covered by the surveys.

Figure 4 below illustrates the reconstruction of the TFR in Colombia between 1961 and 2008 using Poisson regression with (a) a constant age schedule (CAS), and (2) survey-wise constant age schedules (SWCAS). Dotted lines represent estimates from the 7 separate surveys (as shown on Figure 3). This figure shows that the SWCAS approach fits the separate estimates very well. Before the mid 1980s, the CAS approach leads to an underestimation of the TFR, due to the more rapid fertility decrease at advanced ages. In the following part of this illustration, we use the SWCAS approach.

Figure 4: Reconstruction of TFRs (15-49) between 1961 and 2008 by pooling 7 surveys, Colombia. Comparison of Constant age schedule (CAS) and Survey-wise constant Age schedule (SWCAS)



3.3. Restricted cubic splines

Smoothing fertility trends is performed with restricted cubic splines, i.e. piecewise polynomial functions constrained to join at predefined years (knots) (Andersen, 2009). Cubic splines are flexible and allow fitting a large variety of shapes with relatively few parameters. (Harrell, 2001). To fit restricted cubic splines with K knots, K-1 variables (functions of time periods) are created. The construction of these variables depends on the number and the location of knots¹³. The new variables are introduced as explanatory variables in the Poisson regression model, in place of the dummy variables¹⁴. The predicted total fertility rate is also obtained for each year using the coefficients of regression. The model is of the same form as

¹³ The *mkspline* command in Stata is used to create these variables after defining the number and the location of the knots (StataCorp, 2007).

¹⁴ The rest of the model is similar as in the previous section : the age pattern is controlled in the offset.

in [Eq. 1]. The only difference is that $g(time)$ is not modeled as a series of dummy variables, but as a linear function of the $K-1$ variables (RCS) created to fit the restricted cubic splines.

$$\log(\mu_{it}) = \log(e_{it}) + \log(a_i) + \gamma + \sum_{h=1}^{K-1} \delta_h \cdot RCS_{hit} \quad [\text{Eq. 14}]$$

For instance, when there are five knots, the model is

$$\log(\mu_{it}) = \log(e_{it}) + \log(a_i) + \gamma + \delta_1 \cdot RCS_{1it} + \delta_2 \cdot RCS_{2it} + \delta_3 \cdot RCS_{3it} + \delta_4 \cdot RCS_{4it} \quad [\text{Eq. 15}]$$

The TFR is estimated as:

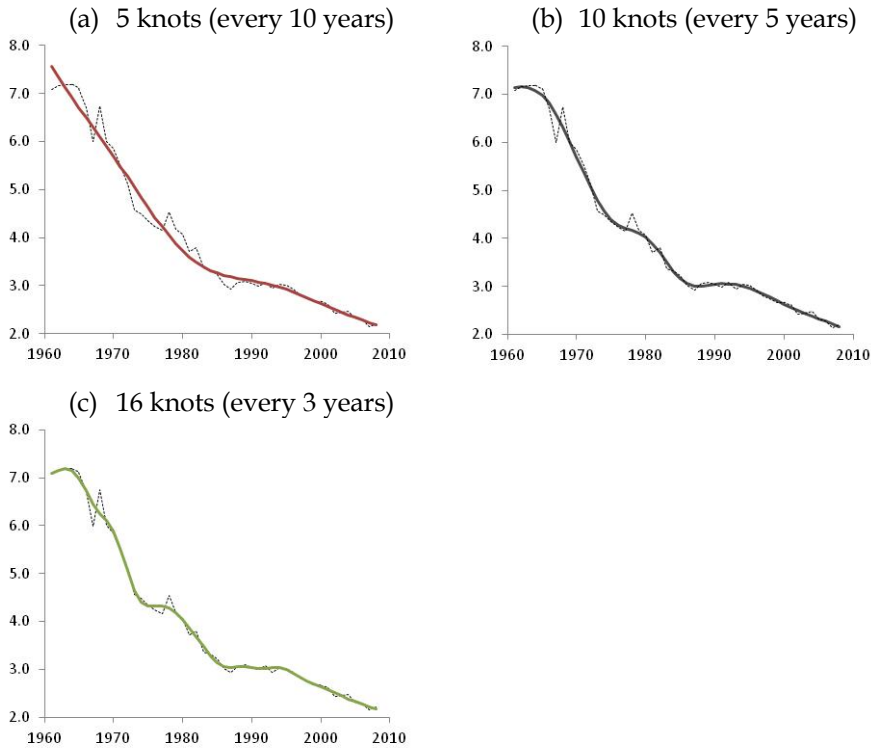
$$TFR_t = 5 * \exp[\gamma] * \exp[\delta_1 \cdot RCS_{1it} + \delta_2 \cdot RCS_{2it} + \delta_3 \cdot RCS_{3it} + \delta_4 \cdot RCS_{4it}] \quad [\text{Eq. 16}]$$

The number and location of knots have to be defined before adjusting the restricted cubic splines. The shape of the smoothing function is not very sensitive to the location of the knots (Harrell, 2001; Andersen, 2009; Dupont, 2009), but is more sensitive to the number of knots. Figure 5a to 5c compare reconstructed TFRs using restricted cubic splines with knots located every 10 years, 5 years and 3 years¹⁵. Spacing knots by 10 years leads to overlooking some changes in the pace in the fertility transition. Locating knots every five years gives a much better fit, and indicates two slowdowns in the fertility transition. Finally, locating knots every three years fits the TFR only marginally better. In the rest of this paper, knots are located every five years, as a compromise between flexibility and parsimony¹⁶.

¹⁵ The location of knots is done backward, starting from the last knot. The last knot is located $(I/2)$ years before the last year, where I is the width of the interval. For instance, with a 5-year interval, and the last year in 2008 (mid-point 2008.5), the last knot is located on 2006. The next to last knot is located on 2001, and so on. Different locations of knots were tested, and have a very small impact on the shape of the smoothing function.

¹⁶ Secondly, locating knots every five years allows identifying stalls in fertility transitions. Several authors suggest that a stall should be considered as such if it lasts for at least five years (Machiyama and Sloggett, 2009; Moultrie et al., 2008)? Knots need not be spaced evenly.

Figure 5: Adjustment of restricted cubic splines to reconstructed TFRs (15-49) between 1961 and 2008, Colombia.



3.4. Testing for stalls

An additional interest of the RCS approach is to allow testing changes in the rhythm of fertility transitions, and notably to identify stalls. With our method, this can be done by testing if the slope of the RCS is significantly negative. This is done by computing marginal effects (the first derivative) of the RCS function, as well as their standard errors (Buis, 2009)¹⁷. A marginal effect that is not significantly negative is interpreted as a stall¹⁸.

¹⁷ This is done in stata with the user-written command *mfxrcspline* (Buis, 2009).

¹⁸ A one-tailed test is performed by constructing 90% confidence intervals around the marginal effects. If 0 is not included in the 90% confidence interval, the slope is significantly negative with a 95% confidence.

Figure 6: Marginal effect of RCS and 90% confidence interval, TFRs (15-49) between 1961 and 2008, Colombia.

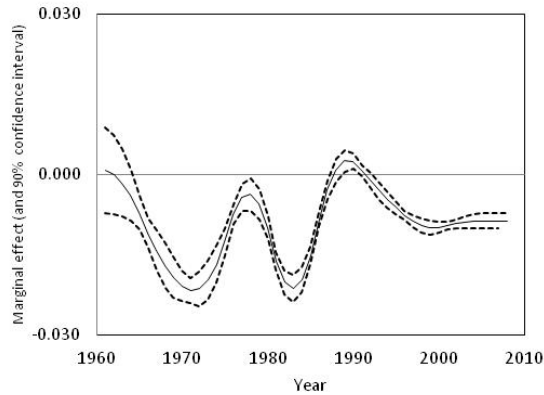
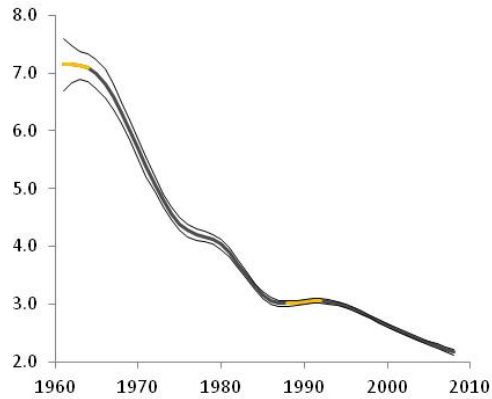


Figure 6 shows the marginal effects and their 90% confidence interval for the RCS with knots located every five years (corresponding to figure 5b). In the first few years, the slope is not significantly negative. The rate of fertility decline reaches -1.5% per year around 1970, and the fertility decline slows down in the late 1970s. A new acceleration of fertility decline occurs in the early 1980s, followed by a deceleration and a stall between 1988 and 1992. After 1992, the decline resumes, and fertility currently declines at a rate of 0.7% per year.

Figure 7 summarizes this information by representing years when fertility decreases significantly in black, and years when fertility is either stable or increasing (stall) in orange (light grey in black and white). The 90% confidence interval for the TFRs is also shown on these figures (black smooth lines above and below the reconstructed trends)¹⁹.

¹⁹ The sample design is taken into account using the jackknife method (correcting for clustering), and standard errors of TFRs are computed from standard errors of the coefficients with the delta method. In the case of Colombia, the sample sizes of the various surveys are large, resulting in small standard errors. In most situations, confidence intervals are larger

Figure 7: Reconstructed TFR (15-49) in Colombia (1961-2008), 90% confidence intervals, and identification of periods of no significant decline (orange).



4. Validation with micro-simulation

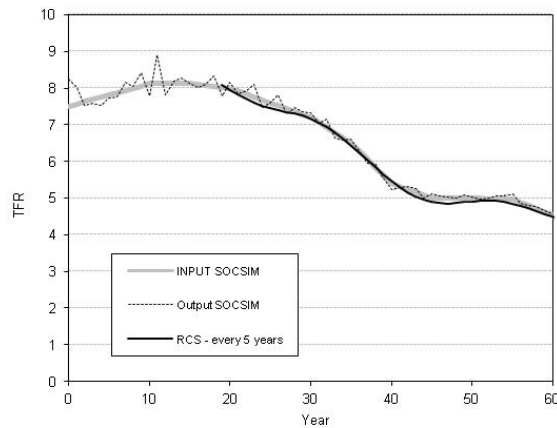
Several scenarios of fertility transitions are used to generate birth histories through micro-simulation (using SOCSIM)²⁰. These scenarios differ by the speed of fertility decline, as well as the changing age pattern of fertility. Currently two scenarios have been generated. Other scenarios will be used to generate different types of fertility transitions and to reproduce typical data quality problems (four to six others scenarios will be used). The first scenario is characterized by a more rapid fertility decline at low ages (like in, say, Morocco). Mean age at childbearing increases, at least in the first phase of the fertility decline. The second scenario represents a situation where fertility decreases more quickly at high ages (like in Colombia for instance). Mean age at childbearing decreases.

DHS-like samples are selected from these simulated birth histories at several points in time. The method described above is used to reconstruct fertility trends from these samples of women (birth histories), and these trends can be compared to the trend of the TFR used as the input of SOCSIM (as well as the TFR computed from the SOCSIM output, that is birth histories).

Preliminary results indicate the method performs well. As shown on Figure 8 (scenario of decreasing mean age at maternity), the black line (reconstructed fertility trends) is very close to the SOCSIM input and output, indicating the method based on pooling DHS allows replicating the original data.

²⁰ We thank Carl Mason from the University of California for generating simulated birth histories.

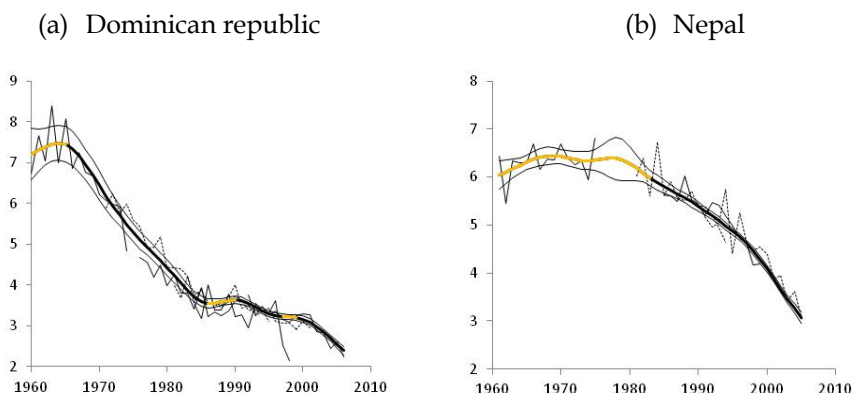
Figure 8: Comparison of SOCSIM input (TFRs) and outputs (TFRs from simulated birth histories), and reconstructed fertility trends from simulated DHS-like surveys at 5-year intervals (second scenario – decreasing mean age at maternity)

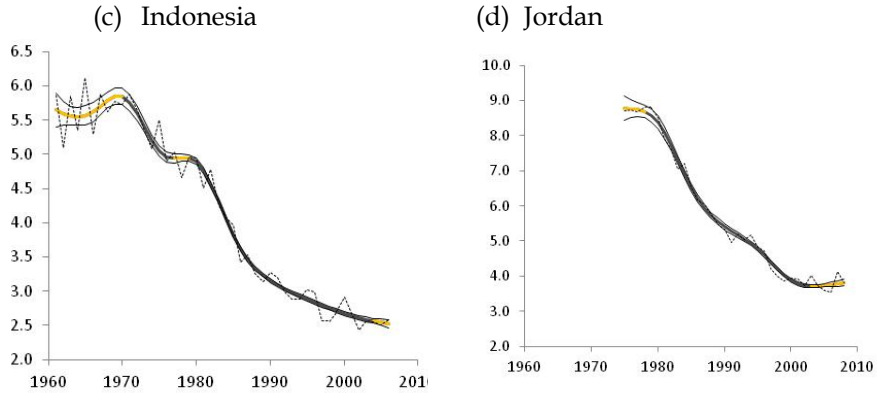


5. Preliminary results in a few countries

The method will be applied to a wide variety of countries from various parts of the World. Preliminary results include the Dominican Republic, Indonesia, Nepal and Jordan (Figure 9a to 9d). These four examples show that fertility trends are not linear, and are characterized by accelerations and slow downs, which are usually not apparent in published data. Stalls are visible in the Dominican Republic, Indonesia and Jordan.

Figure 9: Reconstructed TFR (15-49) in Dominican Republic, Nepal, Indonesia and Jordan, with 90% confidence intervals, and identification of periods of no significant decline (orange).





Further analyses will include countries with surveys affected by discrepancies across surveys (e.g. Morocco, Fertility in the first survey (WFS) seems to be underestimated, Figure 10a) and/or data quality problems within surveys (e.g. displacements and omissions of births Ghana, Figure 10b). Random effect models and models correcting for displacements and omissions of births will be used to produce corrected estimates of fertility trends in countries affected by data quality problems. Preliminary analyses indicate that the random effect approach leads to much wider confidence intervals (Figure 11, Morocco), and also leads to revising upward (by about 1 child) the fertility estimates from the World Fertility Survey in the 1970s.

Figure 10: Reconstructed TFR (15-49) for the last 15 years from several surveys in Morocco and Ghana.

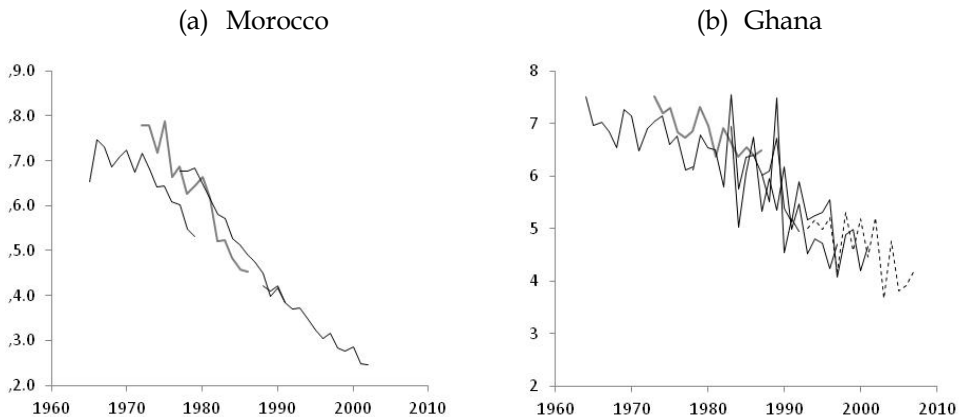
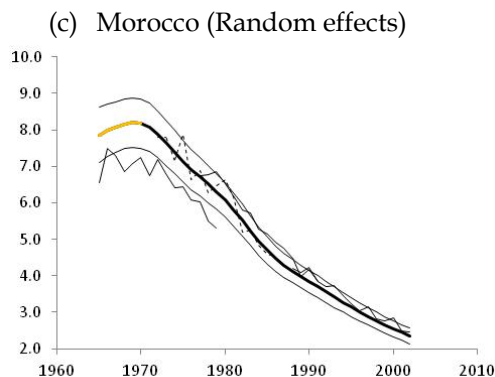


Figure 9: Reconstructed TFR (15-49) in Morocco, using random effects models, , with 90% confidence intervals, and identification of periods of no significant decline (orange)



References

- Alkema L., Raftery A., Gerland P., Clark S., Pelletier F. (2012), “Estimating Trends in the Total Fertility Rate with Uncertainty Using Imperfect Data. Examples from West Africa”, *Demographic Research*, 26(15), 331-362.
- Andersen, R., 2009, “Nonparametric Methods for Modeling Nonlinearity in Regression Analysis”, *Annual Review of Sociology*, vol. 35, pp. 67-85.
- Buis, Maarten L. 2009. "POSTRCSPLINE: Stata module containing post-estimation commands for models using a restricted cubic spline" <http://ideas.repec.org/c/boc/bocode/s456928.html>
- Garenne M. and V. Joseph (2002), “The Timing of the Fertility Transition in Sub-Saharan Africa.” *World Development*, 30(10), 1835-1843.
- Harell, F., 2001, *Regression Modeling Strategies*, Springer Verlag, Secausus, 568 p.
- Machiyama K. (2010), A Re-examination of Recent Fertility Declines in Sub-Saharan Africa, *DHS Working Paper*, 68, ICF Macro, Calverton.
- Machiyama K. and A. Sloggett, 2009, “Is Fertility Decline Stalling in Sub-Saharan Africa? Re-Examination of Fertility Trends”, *Meeting of the Population Association of America*, Detroit, May 2009.
- Schoumaker B. (2012, under revision). “tfr2: A Stata module for computing fertility rates and TFRs from birth histories”, *Demographic Research*.

Schoumaker B. (2010), “Reconstructing Fertility Trends in Sub-Saharan Africa by Combining Multiple Surveys Affected by Data Quality Problems”, *Meeting of the Population Association of America*, Dallas.

United Nations, Department of Economic and Social Affairs, Population Division (2011). *World Fertility Report 2009*, United Nations, New York.