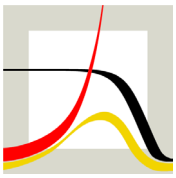

Max-Planck-Institut für demografische Forschung

Max Planck Institute for Demographic Research

Konrad-Zuse-Strasse 1 ▪ D-18057 Rostock ▪ Germany ▪ Tel +49 (0) 3 81 20 81 - 0 ▪ Fax +49 (0) 3 81 20 81 - 202 ▪ www.demogr.mpg.de

MPIDR Technical Report TR 2018-002 I January 2018

R programs for splitting abridged fertility data into a fine grid of ages using the quadratic optimization method

Anatoli I. Michalski I mpoctok@yandex.ru
Pavel Grigoriev I grigoriev@demogr.mpg.de
Vasily P. Gorlishchev I gpa15@yandex.ru

For additional material see www.demogr.mpg.de/tr/

This technical report has been approved for release by: Dmitri A. Jdanov (jdanov@demogr.mpg.de), Head of the Laboratory of Demographic Data.

R programs for splitting abridged fertility data into a fine grid of ages using the quadratic optimization method

Anatoli I. Michalski (mpoctok@yandex.ru)

Pavel Grigoriev (grigoriev@demogr.mpg.de)

Vasily P. Gorlishchev (gpa15@yandex.ru)

Abstract. Occasionally**,** it is necessary to split aggregated fertility data into a fine grid of ages. In this technical report, we present an application of the quadratic optimization (QO) method for splitting abridged fertility data. The QO approach allows us to split not only age-specific fertility rates, but also birth counts in cases in which population exposures are available. Unlike the existing methods, the QO algorithm simultaneously satisfies five criteria of the 'ideal' disaggregation method: 1) *shape* – the estimated fertility curves should be plausible and smooth*; 2) fit* **–** the predicted values should be as close as possible to the observed values; 3) *non-negativity* **–** only positive values should be returned for the birth-order data; *4) balance* – the estimated five-year age group totals should match the input data; and 5) *parity* – the balance by parity must be maintained. The examples provided in this report demonstrate the simplicity, flexibility, and high degree of effectiveness of the proposed technique.

1.Background

The challenge of having to split aggregated fertility data into single years of age is often encountered by demographers. To tackle this problem, several reasonably effective disaggregation methods have been developed and tested (McNeil et. 1975; Smith, Hyndman, Wood, 2004; Liu, et al. 2011; Schmertmann 2012; Jasilioniene et al. 2012, Grigoriev and Jdanov, 2015). The need for disaggregation tends to arise when demographers are using historical data or data from developing countries that do not have functioning systems of vital registration. Splitting is often required for the purposes of data harmonization; i.e., for ensuring that the data are comparable across time and countries. The issue of splitting is particularly relevant in discussions about the maintenance of large international databases such as the Human Fertility Database (http://www.humanfertility.org) and the Human Fertility Collection (http://www.fertilitydata.org). Currently, the HFD has its own splitting protocol, the HFD method (Jasilioniene et al. 2012); while the HFC uses the Calibrated Spline (CS) estimator (Schmertmann, 2012) to disaggregate age-specific fertility rates (Grigorieva et al. 2015). Both methods work reasonably well, but they also have certain disadvantages (see Grigoriev et al. 2018). These challenges motivated us to develop a new method that can be universally applied to both the high-quality HFD data and the heterogeneous and noisy HFC data. Our new method simultaneously satisfies the following criteria:

1) *Shape* - the estimated fertility curves should be plausible and smooth;

2) *Fit* – the predicted values should be as close as possible to the observed values;

3) *Non-negativity* – only positive values should be returned;

4) *Balance* – the estimated five-year age group totals should match the input data; and

5) *Parity* – the balance by parity must be maintained.

To our knowledge, none of the existing methods fully meets the first four criteria, and no attempt has been made to extend the restrictions to criterion (5).

2. Description of the algorithm

The problem of splitting aggregated fertility data into single years of age is formulated as the reconstruction of a smooth curve, which satisfies criteria (1-5). The smoothness of the disaggregated age-specific fertility rates (ASFR) is determined by the quadratic function, which equals the sum of the squares of the second order divided differences for the ASFR, which have to be as small as possible. In this way, the problem of splitting aggregated fertility data is reduced to the problem of quadratic optimization under a set of linear constraints derived from criteria (2-5).

Mathematically, the quadratic optimization problem for the age-specific fertility rate calculation can be formulated as

Minimize $\|F' fr\|^2$ criterion (1)

subject to $G' fr - b = 0$, criterion (4)

$fr_j \geq 0$, j=1,...,n. criterion (3)

Here, n is the number of years for which the age-specific fertility rate is calculated, fr is the vector of the one-year ASFR values, F is the $(n-2) \times n$ matrix for the calculation of the second-order divided differences, b is the vector of the observed number of births by m five-year age groups, and G is the $m \times n$ matrix of the number of women exposed to a birth event at a specific age in a specific five-year age group.

The quadratic optimization problem takes on a more complex form when both age- and parity-specific fertility rates have to be estimated. That is,

minimize	$\|F' fr\|^2 + \sum_{p=1}^{5} \|F' fr^p\|^2$		criterion (1)
subject to	$G' fr - b = 0$		criterion (4)
	$G^p{}' fr^p - b^p = 0$	p=1,...,5	criterion (4)
	$E_j fr_j - \sum_{p=1}^{5} E_j^p{}' fr_j^p = 0$	j=1,...,n	criterion (5)
	$fr_j \geq 0$	j=1,...,n,	criterion (3)
	$fr_j^p \geq 0$	j=1,...,n, p=1,...,5	criterion (3)

The new notations fr^p are - the vectors of one-year ASFRs by parities 1,2,3,4, and 5+; b^p - the vectors of the observed number of births by m five-year age groups and by parities 1,2,3,4, and 5+; G^p - $m \times n$ matrixes of the number of women exposed to births at a specific age in a specific five-year age group and parities 1,2,3,4, and 5+; E_j - the number of women in the j-th one-year age group exposed to births; and E_j^p - the number of women in the j-th one-year age group who had already given birth to p-1 children. In the case of unconditional estimates for the age-specific fertility rate, all of the G^p matrixes are equal to the G matrix, and the vectors E and E^p are the rows of the G matrix.

3. Requirements

The scripts were tested using R version 3.1. The application of the QO method requires the use of the '*quadprog'* package, which must be installed from a standard R repository (CRAN).

4.Usage

The .zip file included in the technical report contains *R scripts* and the data files used in the examples. R functions *QOSplit.R* and *QOSplitPar.R* have to be loaded into an R environment using the *source* command. A description of these functions is provided below.

4.1. R functions

4.1.1. QOSplit.R()

This function splits the aggregated age-specific fertility rates (ASFRs) into single years of age using the quadratic optimization procedure.

Usage: *QOSplit (Fx ,L, AgeInt)*

The arguments are as follows:

Fx – the vector of the aggregated ASFRs to be split into single ages;

L – the vector containing the lower limits of the age intervals (e.g., L=10 for age interval 10-14); and

AgeInt - the vector containing the length of the age intervals.

All of the arguments should have the same length.

Example: *QOSplit (Fx*= c(*0.00008, 0.02655, 0.10538, 0.09983, 0.05105, 0.01904, 0.00359, 0.00033, 0.00000), L=seq(10,50,5), AgeInt=rep (5,9))*

QOSplit.R returns the data frame in the following format:

Age	ASFR
10	0.00000000
11	0.00000000
12	0.00000000
13	0.00000000
14	0.00040106
15	0.00330651
16	0.01036136
17	0.02223489
18	0.03862065
19	0.05823645
20	0.07882435
21	0.09715068
22	0.11096405
23	0.11899532
24	0.12095762
25	0.11754635
26	0.11043916
27	0.10104228
28	0.09049050
29	0.07964717
30	0.06910420
31	0.05918206
32	0.05010353
33	0.04199370
34	0.03487994
35	0.02869193
36	0.02326165
37	0.01846799
38	0.01423672
39	0.01054049
40	0.00739889
41	0.00487835
42	0.00299531
43	0.00171613
44	0.00095713
45	0.00058459
46	0.00041477
47	0.00031853
48	0.00022143
49	0.00010364
50	0.00000000
51	0.00000000
52	0.00000000
53	0.00000000
54	0.00000000

By default, the minimum and the maximum ages (here, 10 and 54) in the output correspond to the minimum and the maximum ages of the first (10-14) and the last (50-54) age intervals in the input data.

4.1.2. QOSplitPar.R()

This function splits the aggregated age- and parity-specific fertility data (either rates or births counts) into single years of age using a quadratic optimization procedure. In cases in which the population exposures by single year of age are not available, the function can only return the ASFRs.

Usage: *QOSplitPar<-function(CY,Fx,Pop5,Pop1=NULL, L, AgeInt, Rates=TRUE)*

The arguments are as follows:

CY – the country-year string (e.g., “AUT1984”);

Fx – the data frame containing the aggregated birth counts to be split;

Pop5 – the vector of the population exposures (aggregated): length(Pop5)=length(Fx);

Pop1 – the vector of the population exposures by single year of age: length(Pop1)=length(AgeOutput) (Note: By default, the Pop1=NULL – data are not available.);

L – the vector containing the lower limits of the age intervals (e.g., L=10 for the age interval 10-14)

AgeInt - the vector containing the length of the age intervals*;* and

Rates – in response to the question of whether rates or birth counts should be returned, Rates=FALSE - birth counts should be returned, and otherwise rates should be returned (Note: By default, Rates=TRUE because Pop1=NULL).

Example:

QOSplitPar(CY=”FIN2004”,Fx=d5[,5:10],Pop5=d5$Exposure,Pop1=NULL, L=d5$Age,AgeInt=d5$AgeInt,Rates=TRUE)

QOSplitPar.R returns the data frame in the following format:

Age	ASFR0	ASFR1	ASFR2	ASFR3	ASFR4	ASFR5P	CountryYear	Diff
12	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	FIN2004	0.00000
13	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	FIN2004	0.00000
14	0.00012292	0.00012292	0.00000000	0.00000000	0.00000000	0.00000000	FIN2004	0.00000
15	0.00168088	0.00171299	0.00000000	0.00000000	0.00000000	0.00000000	FIN2004	-0.00003
16	0.00463514	0.00466725	0.00000000	0.00000000	0.00000000	0.00000000	FIN2004	-0.00003
17	0.00896009	0.00886673	0.00009336	0.00000000	0.00000000	0.00000000	FIN2004	0.00000
18	0.01497605	0.01384366	0.00109648	0.00000000	0.00000381	0.00000000	FIN2004	0.00003
19	0.02289256	0.01923820	0.00336921	0.00022474	0.00002830	0.00000000	FIN2004	0.00003
20	0.03274254	0.02486426	0.00688077	0.00092371	0.00008883	0.00000000	FIN2004	-0.00002
21	0.04431083	0.03078099	0.01128111	0.00200885	0.00027071	0.00000000	FIN2004	-0.00003
22	0.05722504	0.03690366	0.01625608	0.00341569	0.00062999	0.00002558	FIN2004	-0.00001
23	0.07097391	0.04298532	0.02150905	0.00508500	0.00117509	0.00019845	FIN2004	0.00002
24	0.08488892	0.04863515	0.02677927	0.00698116	0.00188521	0.00057731	FIN2004	0.00003
25	0.09814433	0.05331850	0.03184188	0.00909216	0.00271028	0.00115783	FIN2004	0.00002
26	0.10975555	0.05635845	0.03650950	0.01143118	0.00357262	0.00187426	FIN2004	0.00001
27	0.11893407	0.05734987	0.04057096	0.01394311	0.00441011	0.00266440	FIN2004	0.00000
28	0.12508750	0.05615939	0.04379126	0.01650449	0.00517618	0.00346962	FIN2004	-0.00001
29	0.12781954	0.05292544	0.04591159	0.01892354	0.00583987	0.00423481	FIN2004	-0.00002
30	0.12692999	0.04805820	0.04664930	0.02094015	0.00638577	0.00490845	FIN2004	-0.00001
31	0.12241476	0.04223964	0.04569795	0.02222590	0.00681403	0.00544255	FIN2004	-0.00001
32	0.11451635	0.03603456	0.04297412	0.02258389	0.00711714	0.00580521	FIN2004	0.00000
33	0.10372386	0.02989062	0.03861743	0.02194878	0.00727990	0.00598062	FIN2004	0.00001
34	0.09077299	0.02413831	0.03299052	0.02038678	0.00727942	0.00596907	FIN2004	0.00001
35	0.07664605	0.01899098	0.02667907	0.01809562	0.00708515	0.00578692	FIN2004	0.00001
36	0.06257194	0.01454481	0.02049179	0.01540460	0.00665884	0.00546662	FIN2004	0.00001
37	0.04947378	0.01084595	0.01500616	0.01258647	0.00599630	0.00503835	FIN2004	0.00000
38	0.03796894	0.00789047	0.01056844	0.00985745	0.00512737	0.00452997	FIN2004	0.00000
39	0.02836902	0.00562444	0.00729367	0.00737723	0.00411599	0.00396707	FIN2004	-0.00001
40	0.02067985	0.00394383	0.00506566	0.00524897	0.00306011	0.00337296	FIN2004	-0.00001
41	0.01460149	0.00269460	0.00353701	0.00351929	0.00209175	0.00276862	FIN2004	-0.00001
42	0.00987457	0.00175819	0.00244416	0.00220198	0.00129860	0.00217492	FIN2004	0.00000
43	0.00628030	0.00105151	0.00160745	0.00127797	0.00072401	0.00161257	FIN2004	0.00001
44	0.00364045	0.00052696	0.00093109	0.00069533	0.00036702	0.00110209	FIN2004	0.00002
45	0.00181735	0.00017243	0.00040313	0.00036932	0.00018232	0.00066390	FIN2004	0.00003
46	0.00071393	0.00001129	0.00009552	0.00018234	0.00008030	0.00031824	FIN2004	0.00003
47	0.00017192	0.00000000	0.00000000	0.00007620	0.00002417	0.00009388	FIN2004	-0.00002
48	0.00000000	0.00000000	0.00000000	0.00002406	0.00000190	0.00000000	FIN2004	-0.00003
49	0.00000000	0.00000000	0.00000000	0.00000420	0.00000000	0.00000000	FIN2004	0.00000
50	0.00001559	0.00000000	0.00000000	0.00000000	0.00000353	0.00001169	FIN2004	0.00000
51	0.00002992	0.00000000	0.00000000	0.00000000	0.00000739	0.00002244	FIN2004	0.00000
52	0.00003304	0.00000000	0.00000000	0.00000000	0.00000847	0.00002478	FIN2004	0.00000
53	0.00002182	0.00000000	0.00000000	0.00000000	0.00000570	0.00001637	FIN2004	0.00000
54	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	FIN2004	0.00000

Columns ASFR0, ASFR1, ASFR2, ASFR3, ASFR4, and ASFR5P refer to the total (0) age-specific fertility rate, and to the ASFRs by parities 1,2,3,4, and 5+, respectively. Variable *Diff* indicates a small difference between ASFR0 and the sum of ASFR1–ASFR5P.

The output containing birth counts has the same structure.

QOSplitPar(CY="FIN2004",Fx=d5[,5:10],Pop5=d5$Exposure,Pop1=d1$Exposure, L=d5$Age,AgeInt=d5$AgeInt,Rates=FALSE)

Age	B0	B1	B2	B3	B4	B5P	CountryYear	Diff
12	0.00	0.00	0.00	0.00	0.00	0.00	FIN2004	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	FIN2004	0.00
14	3.97	3.97	0.00	0.00	0.00	0.00	FIN2004	0.00
15	53.04	54.06	0.00	0.00	0.00	0.00	FIN2004	-1.02
16	143.29	144.28	0.00	0.00	0.00	0.00	FIN2004	-0.99
17	269.70	266.89	2.81	0.00	0.00	0.00	FIN2004	0.00
18	464.36	429.24	34.00	0.00	0.12	0.00	FIN2004	1.00
19	736.14	618.63	108.34	7.23	0.91	0.00	FIN2004	1.03
20	1081.99	821.65	227.38	30.52	2.94	0.00	FIN2004	-0.50
21	1472.33	1022.77	374.84	66.75	8.99	0.00	FIN2004	-1.02
22	1855.53	1196.60	527.11	110.75	20.43	0.83	FIN2004	-0.19
23	2249.15	1362.19	681.62	161.14	37.24	6.29	FIN2004	0.67
24	2702.97	1548.60	852.69	222.29	60.03	18.38	FIN2004	0.98
25	3155.60	1714.33	1023.80	292.34	87.14	37.23	FIN2004	0.76
26	3523.61	1809.34	1172.11	366.99	114.70	60.17	FIN2004	0.30
27	3888.03	1874.80	1326.29	455.81	144.17	87.10	FIN2004	-0.14
28	4070.70	1827.59	1425.09	537.10	168.45	112.91	FIN2004	-0.44
29	4026.63	1667.28	1446.33	596.14	183.97	133.41	FIN2004	-0.50
30	3743.00	1417.17	1375.63	617.50	188.31	144.74	FIN2004	-0.35
31	3445.39	1188.85	1286.18	625.55	191.78	153.18	FIN2004	-0.15
32	3408.00	1072.39	1278.91	672.10	211.81	172.76	FIN2004	0.03
33	3202.95	923.01	1192.49	677.77	224.80	184.68	FIN2004	0.20
34	2869.14	762.96	1042.76	644.38	230.09	188.67	FIN2004	0.28
35	2565.04	635.55	892.84	605.59	237.11	193.67	FIN2004	0.28
36	2232.32	518.90	731.07	549.58	237.56	195.03	FIN2004	0.18
37	1781.12	390.47	540.24	453.13	215.87	181.39	FIN2004	0.02
38	1375.56	285.86	382.88	357.12	185.76	164.11	FIN2004	-0.17
39	1047.48	207.67	269.31	272.39	151.98	146.48	FIN2004	-0.35
40	774.03	147.61	189.60	196.47	114.54	126.25	FIN2004	-0.44
41	548.34	101.19	132.83	132.16	78.55	103.97	FIN2004	-0.36
42	369.16	65.73	91.37	82.32	48.55	81.31	FIN2004	-0.12
43	233.77	39.14	59.83	47.57	26.95	60.03	FIN2004	0.25
44	136.06	19.69	34.80	25.99	13.72	41.19	FIN2004	0.67
45	66.62	6.32	14.78	13.54	6.68	24.34	FIN2004	0.96
46	26.60	0.42	3.56	6.79	2.99	11.86	FIN2004	0.98
47	6.59	0.00	0.00	2.92	0.93	3.60	FIN2004	-0.86
48	0.00	0.00	0.00	0.94	0.07	0.00	FIN2004	-1.01
49	0.00	0.00	0.00	0.16	0.00	0.00	FIN2004	-0.16
50	0.61	0.00	0.00	0.00	0.14	0.46	FIN2004	0.01
51	1.19	0.00	0.00	0.00	0.29	0.89	FIN2004	0.01
52	1.30	0.00	0.00	0.00	0.33	0.98	FIN2004	-0.01
53	0.87	0.00	0.00	0.00	0.23	0.65	FIN2004	-0.01
54	0.00	0.00	0.00	0.00	0.00	0.00	FIN2004	0.00

4.2. Format of the input data

The input data appear in the form of .csv (comma separated values) files, with the first line containing the variable names (header). Five data files are used in the examples:

- ‘HFD abridged data 7 groups.csv’,
- ‘HFD abridged data 9 groups.csv’,
- ‘HFD abridged data 7 groups BO.csv’,
- ‘HFD abridged data 9 groups BO.csv’, and
- ‘BirthsExposASFR1x1_BO.csv’.

Files (1) and (2) refer to the input data with seven and nine age groups, respectively. These files have to be used by *QOSplit.R.*

Fragment of 'HFD abridged data 7 groups.csv':

```
head(dta2,10)
 Age CountryYear Code Year Births  Exposure            ASFR AgeGrLabel AgeInt
  15     AUT1951  AUT 1951   7350 215582.54 0.0340936701089      15-19      5
  20     AUT1951  AUT 1951  27775 242049.79 0.1147491183529      20-24      5
  25     AUT1951  AUT 1951  32974 291065.68 0.1132871453618      25-29      5
  30     AUT1951  AUT 1951  17917 217312.90 0.0824479356725      30-34      5
  35     AUT1951  AUT 1951  11638 257812.81 0.0451412790544      35-39      5
  40     AUT1951  AUT 1951   4702 293646.64 0.0160124427101      40-44      5
  45     AUT1951  AUT 1951    393 554677.92 0.0007085192791      45-49      5
  15     AUT1952  AUT 1952   7184 208671.92 0.0344272482853      15-19      5
  20     AUT1952  AUT 1952  27770 234527.53 0.1184082738602      20-24      5
  25     AUT1952  AUT 1952  32427 280115.63 0.1157629083390      25-29      5
```

Fragment of 'HFD abridged data 9 groups.csv':

```
head(dta1,10)
 Age CountryYear Code Year   Births  Exposure            ASFR AgeGrLabel AgeInt
  10     AUT1951  AUT 1951    15.00 140071.63 0.00010708806630      10-14      5
  15     AUT1951  AUT 1951  7350.00 215582.54 0.03409367010891      15-19      5
  20     AUT1951  AUT 1951 27775.00 242049.79 0.11474911835288      20-24      5
  25     AUT1951  AUT 1951 32974.00 291065.68 0.11328714536183      25-29      5
  30     AUT1951  AUT 1951 17917.00 217312.90 0.08244793567248      30-34      5
  35     AUT1951  AUT 1951 11638.00 257812.81 0.04514127905437      35-39      5
  40     AUT1951  AUT 1951  4702.00 293646.64 0.01601244271005      40-44      5
  45     AUT1951  AUT 1951   379.59 287396.02 0.00132079073329      45-49      5
  50     AUT1951  AUT 1951    13.41 267281.90 0.00005017174751      50-54      5
  10     AUT1952  AUT 1952    18.00 163855.71 0.00010985274788      10-14      5
```

Note: For testing purposes, these data files include more variables than are actually needed. Only four variables are used by function *QOSplit.R*: *Age, CountryYear, ASFR, and AgeInt*.

Files (3) and (4) refer to the parity-specific fertility data aggregated into seven and nine age groups. These files have to be used by *QOSplitPar.R*

Fragment of 'HFD abridged data 7 groups BO.csv':

```
head(dat5,10)
 Age CountryYear Code Year    B0    B1    B2   B3   B4 B5p  Exposure           ASFR0
  15     AUT1984  AUT 1984  8332  7422   852   56    2   0 313799.64 0.0265519743745
  20     AUT1984  AUT 1984 34072 20574 11223 1966  270  39 323330.00 0.1053784059629
  25     AUT1984  AUT 1984 28383 10538 11728 4600 1165 352 284304.52 0.0998330944580
  30     AUT1984  AUT 1984 12735  2668  4353 3341 1524 849 249448.19 0.0510526855296
  35     AUT1984  AUT 1984  4640   635  1139 1234  826 806 243700.60 0.0190397561598
  40     AUT1984  AUT 1984   992    94   158  228  184 328 276387.81 0.0035891597390
  45     AUT1984  AUT 1984    67     0     8   14    7  38 412547.14 0.0001624056829
  15     AUT1985  AUT 1985  7486  6742   705   39    0   0 306332.55 0.0244374944811
  20     AUT1985  AUT 1985 32629 19738 10775 1819  247  50 327347.34 0.0996769975281
  25     AUT1985  AUT 1985 28774 10811 11963 4514 1149 337 293083.48 0.0981768061441
           ASFR1            ASFR2            ASFR3            ASFR4           ASFR5P
 0.0236520347824 0.00271510827737 0.00017845782105 0.000006373493609 0.00000000000000
 0.0636315838308 0.03471066712028 0.00608047505644 0.000835060155259 0.00012061980020
 0.0370658897720 0.04125154253615 0.01617983421438 0.004097718882556 0.00123810905293
 0.0106956077733 0.01745051748020 0.01339356280757 0.006109485099892 0.00340351236864
 0.0026056562848 0.00467376772975 0.00506359032354 0.003389404868105 0.00330733695362
 0.0003401018301 0.00057166052294 0.00082492784324 0.000665731241910 0.00118673830080
 0.0000000000000 0.00001939172333 0.00003393551583 0.000016967757915 0.00009211068582
 0.0220087613935 0.00230142046609 0.00012731262153 0.000000000000000 0.00000000000000
 0.0602968088881 0.03291610678737 0.00555678870035 0.000754550197353 0.00015274295493
 0.0368871012450 0.04081772196782 0.01540175515863 0.003920384731340 0.00114984304131
 AgeInt AgeGrLabel
      5      15-19
      5      20-24
      5      25-29
      5      30-34
      5      35-39
      5      40-44
      5      45-49
      5      15-19
      5      20-24
      5      25-29
```

Fragment of ‘HFD abridged data 9 groups BO.csv’:

```
head(dat5,10)
 Age CountryYear Code Year    B0    B1    B2   B3   B4 B5p  Exposure            ASFR0
  12     AUT1984  AUT 1984    13    13     0    0    0   0 162068.54 0.00008021297656
  15     AUT1984  AUT 1984  8332  7422   852   56    2   0 313799.64 0.02655197437448
  20     AUT1984  AUT 1984 34072 20574 11223 1966  270  39 323330.00 0.10537840596295
  25     AUT1984  AUT 1984 28383 10538 11728 4600 1165 352 284304.52 0.09983309445801
  30     AUT1984  AUT 1984 12735  2668  4353 3341 1524 849 249448.19 0.05105268552961
  35     AUT1984  AUT 1984  4640   635  1139 1234  826 806 243700.60 0.01903975615981
  40     AUT1984  AUT 1984   992    94   158  228  184 328 276387.81 0.00358915973899
  45     AUT1984  AUT 1984    67     0     8   14    7  38 203900.99 0.00032859085186
  50     AUT1984  AUT 1984     0     0     0    0    0   0 208646.15 0.00000000000000
  12     AUT1985  AUT 1985    21    21     0    0    0   0 153656.41 0.00013666855812
           ASFR1             ASFR2             ASFR3             ASFR4            ASFR5P
0.00008021297656 0.00000000000000 0.00000000000000 0.000000000000000 0.0000000000000
0.02365203478245 0.00271510827737 0.00017845782105 0.000006373493609 0.0000000000000
0.06363158383076 0.03471066712028 0.00608047505644 0.000835060155259 0.0001206198002
0.03706588977200 0.04125154253615 0.01617983421438 0.004097718882556 0.0012381090529
0.01069560777330 0.01745051748020 0.01339356280757 0.006109485099892 0.0034035123686
0.00260565628480 0.00467376772975 0.00506359032354 0.003389404868105 0.0033073369536
0.00034010183011 0.00057166052294 0.00082492784324 0.000665731241910 0.0011867383008
0.00000000000000 0.00003923472858 0.00006866077502 0.000034330387508 0.0001863649608
0.00000000000000 0.00000000000000 0.00000000000000 0.000000000000000 0.0000000000000
0.00013666855812 0.00000000000000 0.00000000000000 0.000000000000000 0.0000000000000
 AgeInt AgeGrLabel
      3      12-14
      5      15-19
      5      20-24
      5      25-29
      5      30-34
      5      35-39
      5      40-44
      5      45-49
      5      50-54
      3      12-14
```

Note: For testing purposes, these data files include more variables than are actually needed. The following variables are used by function *QOSplitPar.R*: *Age, CountryYear, B0, B1, B2, B3, B4, B5p, Exposure, and AgeInt*.

Both of the *BO* files were generated from the original birth counts and the population exposures by single year of age (‘BirthsExposASFR1x1_BO.csv’). This file provides the information on the population exposures (argument *Pop1* of the function *QOSplitPar.R)*. Additionally, this file contains the original fertility estimates we used to assess the model fit.

Fragment of ‘BirthsExposASFR1x1_BO.csv’:

```
head(dat1,10)
 CountryYear Code Year Age Agegroup    B0    B1    B2  B3 B4 B5p Exposure       ASFR0
     AUT1984  AUT 1984  12    10-14     0     0     0   0  0   0 51383.50 0.000000000
     AUT1984  AUT 1984  13    10-14     3     3     0   0  0   0 53807.85 0.000055800
     AUT1984  AUT 1984  14    10-14    10    10     0   0  0   0 56877.19 0.000175817
     AUT1984  AUT 1984  15    15-19    96    94     2   0  0   0 60561.21 0.001585173
     AUT1984  AUT 1984  16    15-19   405   389    15   1  0   0 61930.79 0.006539558
     AUT1984  AUT 1984  17    15-19  1243  1183    58   2  0   0 62726.76 0.019816104
     AUT1984  AUT 1984  18    15-19  2417  2217   190   9  1   0 63175.63 0.038258423
     AUT1984  AUT 1984  19    15-19  4171  3539   587  44  1   0 65405.25 0.063771639
     AUT1984  AUT 1984  20    20-24  5688  4422  1161  99  6   0 66104.74 0.086045267
     AUT1984  AUT 1984  21    20-24  6581  4415  1908 222 35   1 66588.15 0.098831399
       ASFR1       ASFR2       ASFR3       ASFR4    ASFR5P
 0.000000000 0.000000000 0.000000000 0.000000000 0.000000
 0.000055800 0.000000000 0.000000000 0.000000000 0.000000
 0.000175817 0.000000000 0.000000000 0.000000000 0.000000
 0.001552149 0.000033000 0.000000000 0.000000000 0.000000
 0.006281205 0.000242206 0.000016100 0.000000000 0.000000
 0.018859574 0.000924645 0.000031900 0.000000000 0.000000
 0.035092646 0.003007489 0.000142460 0.000015800 0.000000
 0.054108806 0.008974815 0.000672729 0.000015300 0.000000
 0.066893842 0.017563037 0.001497623 0.000090800 0.000000
 0.066303088 0.028653747 0.003333927 0.000525619 0.000015
```

5. Examples

The R scripts *ExamplesQOSplit.R* and *ExamplesQOSplitPar.R* included in the .zip file contain various examples in which the functions *QOSplit* and *QOSplitPar* are used. As the scripts are accompanied by detailed comments, even inexperienced R users should find the QO method fairly simple to use.

References

Grigoriev P, Jdanov DA (2015). Splitting abridged fertility data using different interpolation methods: is there the optimal solution? Presentation at the 80th PAA Meeting, San Diego, USA. http://www.humanfertility.org/Docs/paa/Grigoriev_Jdanov.pdf

Grigoriev P, Michalski AI, Gorlishchev VP, Jdanov DA, Shkolnikov VM (2018). New methods for estimating detailed fertility schedules from abridged data. MPIDR Working Paper WP-2018-001. Rostock. Available at http://www.demogr.mpg.de/papers/working/wp-2018-001.pdf

Grigorieva O, Jasilioniene A, Jdanov DA, Grigoriev P, Sobotka T, Zeman K, Shkolnikov VM (2015). Methods Protocol for the Human Fertility Collection. Available at: http://www.fertilitydata.org/docs/methods.pdf

Jasilioniene A, Jdanov DA, Sobotka T, Andreev EM, Zeman K, Nash EJ, and Shkolnikov VM. (with contributions of Goldstein J, Philipov D, and Rodriguez G) (2012). Methods Protocol for the Human Fertility Database. Available at: http://www.humanfertility.org

Liu Y, Gerland P, Spoorenberg T, Kantorova V, Andreev K (2011). Graduation methods to derive age-specific fertility rates from abridged data: a comparison of 10 methods using HFD data. Presentation at the First Human Fertility Database Symposium, Max Planck Institute for Demographic Research, Rostock, Nov 2011. http://www.humanfertility.org/Docs/Symposium/Liu-Gerland%20et%20al.pdf

McNeil DR, Trussell TJ, Turner JC (1977). Spline interpolation of demographic data. *Demography* 14(2): 245-252.

Schmertmann C (2012). Calibrated spline estimation of detailed fertility schedules from abridged data. MPIDR Working Paper WP-2012-022. Rostock.

Smith L, Hyndman R, Wood S (2004). Spline interpolation for demographic variables: the monotonicity problem. *Journal of Population Research* 21 (1), pp. 95-97.

Williams HP. (2013). Model Building in Mathematical Programming. Chichester: John Wiley & Sons.