Boom, echo, pulse, flow

(Open version)

Tim Riffe*[1], Kieron Barclay[1], Christina Bohk-Ewald[1], and Sebastian Klüsener[2]

[1]Max Planck Institute for Demographic Research
[2]Bundesinstitut für Bevölkerungsforschung

November 8, 2018

Abstract

Human population renewal starts with births. Since births can happen at any time in the year and over a wide range of ages, demographers typically imagine the birth series as a continuous flow. Taking this construct literally, we visualize the birth series as a flow. A long birth series allows us to juxtapose the children born in a particular year with the children that they in turn had over the course of their lives, yielding a crude notion of cohort replacement. Macro patterns in generational growth define the meandering path of the flow, while temporal booms and busts echo through the flow with the regularity of a pulse.

Keywords: Fertility, Population structure, Population momentum, Population renewal, Data visualization

1 Introduction

Usually demographers think of fertility as an age-regulated process. In any case it is bounded by menarche and menopause, both of which are anchored to age. These anchors may move, but not far or fast. And between these bounds, at least within acceptably homogeneous subpopulations, fertility patterns appear to conform to some regular schema. Since births can happen at any time throughout the year, and since demography usually deals in large numbers, it is common to imagine the birth flow as a continuous stream. This is so not only as a pragmatic assumption to allow for calculus, but it also gives us a heuristic understanding of fertility as a smoother of population structure (Arthur 1982). In this treatment, we retreat from rates, the material of projections, to the absolute number of babies born, the raw material of population renewal.

We aim to represent a historical view of Sweden's historical birth series in a single multilayered visualization. The birth series is rendered as a flow, in such a way as to be suggestive of novel analytic perspectives, and to invite newcomers and curious minds deeper into the discipline of demography. This image entails investment from the viewer, and this manuscript serves as a protracted legend and caption. Intellectual payoffs include a simultaneous sense of long term patterns of generational mixing and generational replacement, medium term baby booms and echos, and the short term shocks of population momentum. We challenge more experienced demographers to relate this image to the Lexis diagram, to imagine how the picture would change if fertility were indexed to fathers' age, or to reimagine this image of aggregates as immense set of lineages.

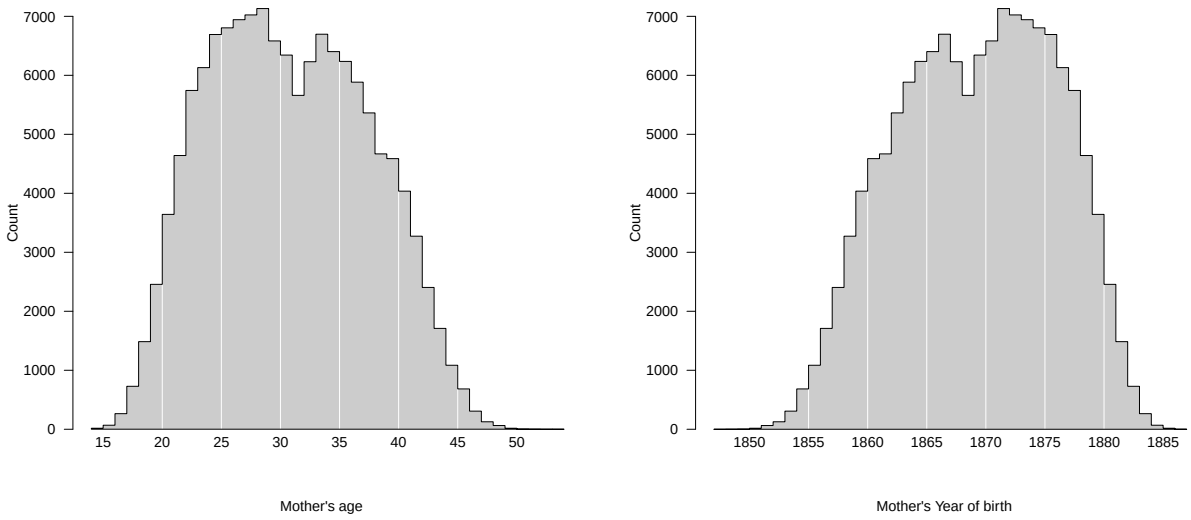
We use birth count data from Sweden, covering a total of 242 occurrence years from 1775 to 2016. Data for the years 1891 to 2016 is taken directly from the Human Fertility Database. (2018) without further adjustment. We augment the HFD series in both directions, including newly digitized data for

*riffe@demogr.mpg.de

the period 1775 to 1890 (Statistique Générale de la France 1907), which we have graduated and adjusted. We describe these adjustments in Appendix A. To complete our picture, we project the fertility of cohorts whose fertility careers are still incomplete (1972-2016) through age 45. We describe the details of this projection in Appendix A.4. The temporal spread from the earliest mother cohort in our final data set (1721) to the latest offspring cohort (2061) is 341 years.

2 Age and cohort-structured birth count distributions

A picture of the births in a year is for demographers most instinctively broken down by the age of mothers who gave birth in that year, Fig. 1a, or by the year of birth of mothers Fig. 1b. These two distributions are essentially identical, but appear as mirror images if chronological time is enforced in x.



(a) Births in 1900 by age of mother

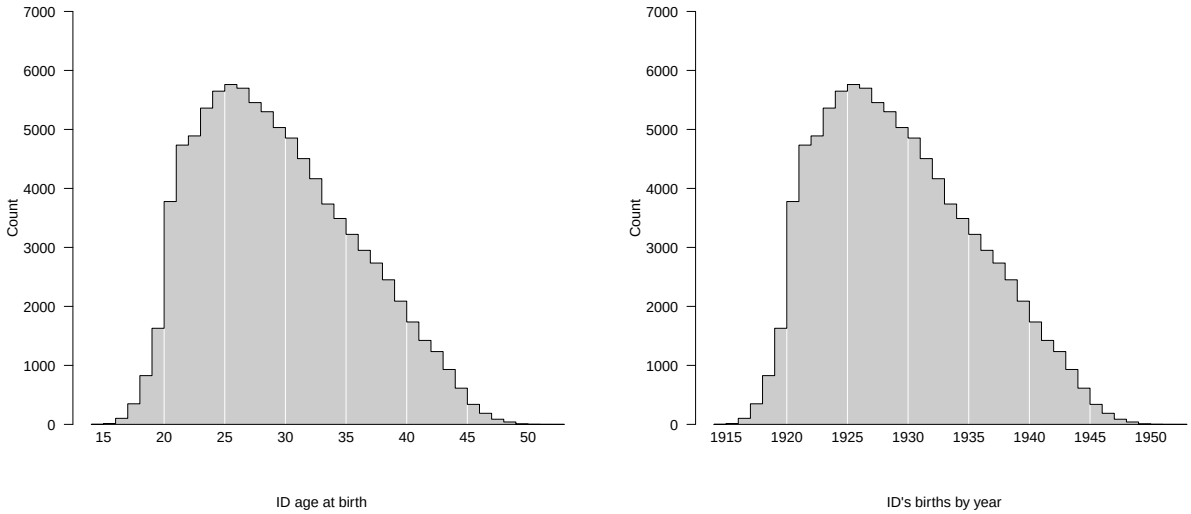
(b) Births in 1900 by year of birth of mother

Figure 1: Births in a year structured by mothers' age versus mothers' year of birth are a reflection over y and shift over x. Count distributions such as this may be jagged, even if the underlying rate distributions are smooth, due to population structure. The deficit around age 31 in 1a is due to a smaller number of potential mothers: the 1871 birth cohort was smaller than the surrounding cohorts.

If one disposes of a long-enough time series of births classified by mothers' year of birth, then one may further examine and break down the full reproductive career of the cohort of individuals born in a particular year, assuming no effects of migration. Since the childbearing of a cohort is spread over a synchronous span of ages and years, the classification by age (Fig. 2a) or year (Fig. 2b) yields identical and redundant distributions.

The births in a year are classified by mothers' cohort, i.e. cohort *origins* in Fig. 1b, whereas the births *from* a cohort are classified *to* time in Fig. 2b. The two distributions are different in kind, but relatable and both on a common scale. A fuller representation of their relationship would place them as two disjoint distributions on the same timeline, as in Fig 3.

The two distributions in Fig. 3 are related, and of comparable scale, but different in kind. The x coordinate of the left distribution is indexed to mothers' birth cohort, whereas the x coordinate of the right distribution is indexed to child cohort, occurrence year. In this way the respective x coordinates are two generations apart, relating to each other as grandmothers and grandchildren, where the *ego* generation is 1900. These are two quantities that we may wish to compare in various ways to get a better



(a) Births from mothers born in 1900 by age of mother

(b) Births from mothers born in 1900 by year

Figure 2: Births of a cohort structured by mothers' age versus mothers' year of birth are a shift over x. The births over the life of a cohort more often resemble the smoothness of fertility rate schedules,

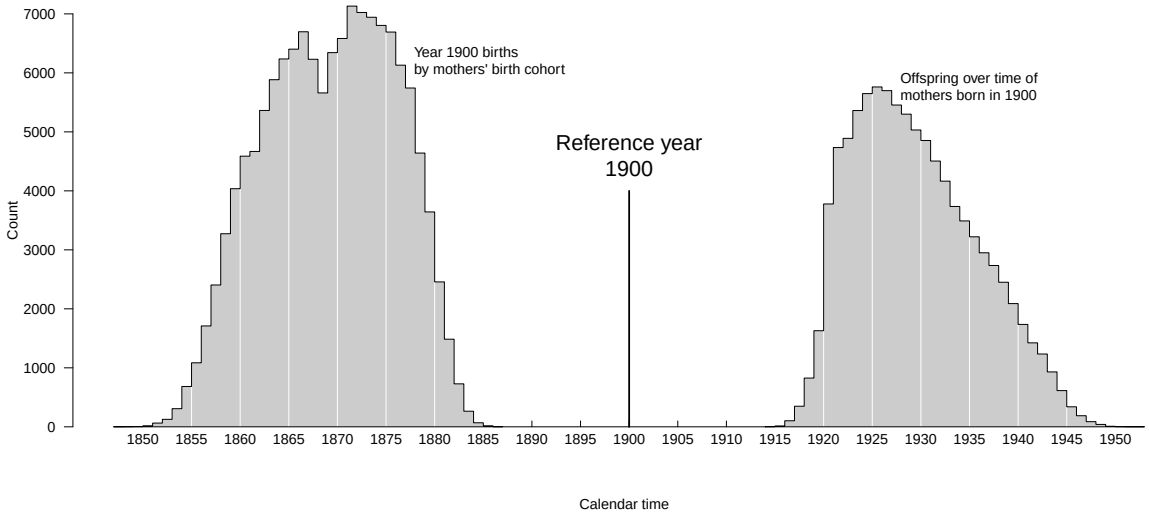


Figure 3: The cohort distribution of mothers who gave birth in 1900 and the births from mothers born in 1900 by year. These two distributions link three generations.

feel and understanding of the Swedish birth series.

For the case of these Swedish data, we have 241 such distribution pairs, making single-axis rendering impractical. An honest attempt might look like Fig. 4, where we reflect the Fig. 3 left distribution over y (**A**), keeping the Fig. 3 right-side distribution on top (**B**). These two distributions are linked by the year 1900, which of course overlaps with neither of them. In this representation, **A** and **B** are re-drawn for each possible ego year (1775-2016), and therefore imply a large sequential set of overlapping distributions. Each 20th distribution is highlighted, but despite attempts to make this graph legible, i) the high degree of overlapping and ii) the spatial dissociation of each **A** — **B** pair makes the intended comparison difficult over the series.

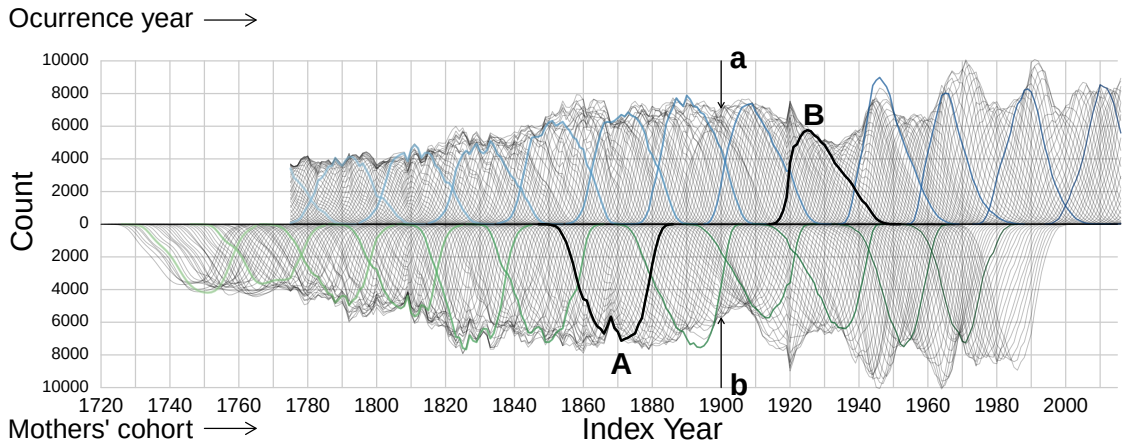


Figure 4: Two time series of birth count distributions. The top series is composed of offspring distributions of mother cohorts over time, indexed to occurrence years. The bottom series is composed of the offspring of a year indexed to mothers' birth cohorts. **B** is the offspring of mothers from the 1900 cohort indexed in x to occurrence year, and **A** are the births occurred in 1900 indexed in x to mothers' birth cohorts. The cross-section **a** gives **A** and the cross-section **b** gives **B**.

Fig. 4 produces at least two noteworthy artifacts that we may wish to preserve and clarify. 1) First order differences in the top series appear to cascade into the lower series— This derives from a specific kind of population momentum (Keyfitz 1971): larger cohorts have more offspring than smaller neighboring cohorts and vice versa, sudden fertility rate changes notwithstanding. 2) The composition of **A** in the bottom series is implied by the cross-section **a** of the top series, and the composition of **B** is implied by the cross-section **b**. This observation deserves further elaboration: The curve **A** is composed of all the births in 1900 indexed *back* to mothers' cohorts. Each point on the curve **A** comes from a different top-axis distribution as it crosses the year 1900 (and vice versa for the bottom). The cross-section of curves **a** is therefore a redundant encoding of the single highlighted curve **A**. The cross-section **b** is a redundant encoding of **B** in the same way. While **A** and **B** are disjoint, and difficult to relate, **a** and **b** share a single x coordinate, and so may lend themselves to comparison. The "problem" with the cross-sections **a** and **b** is that points from the corresponding distributions **A** and **B** are overlapped due to collapsation on a single x coordinate. It is basically impossible to work out what **A** (**B**) might look like if presented only with **a** (**b**) and its surroundings.

In this way the two distributions that we might wish to compare for a given ego year are already available at a like coordinate, but comparison is stifled by overplotting. If instead we stack the slices that are indecipherably overlapped in **a** (and likewise for **b**) we get something like that shown in Fig. 5, cumulative birth distributions.[1] Here the total bar length is proportional to the total cohort (offspring) size, and stacked bins reflect 5-year mother cohorts (occurrence years). From this representation it is clear that mothers born in the 20 years between 1860 and 1880 produced the bulk of the 1900 cohort (86%), which itself produced the majority of its offspring in the 20 years between 1920 and 1940 (90%). It is also quite visible that the 1900 cohort did not replace itself in a crude sense: 138,139 babies formed a cohort whose mothers gave birth to 95,379 babies over their life-course, a crude replacement of 69%. Other perspectives on reproduction that account for survival and attrition of the mother cohort through migration would give a more optimistic assessment (Henry 1965). The key feature of Fig. 5 is that the two distributions that were disjoint in Fig. 3 and hard to pick out in Fig. 4 can now be associated at a common x coordinate. This virtue allows us to view the time series of Fig. 4 with greater clarity and hopefully reveal some macro properties of the history of Swedish natality.

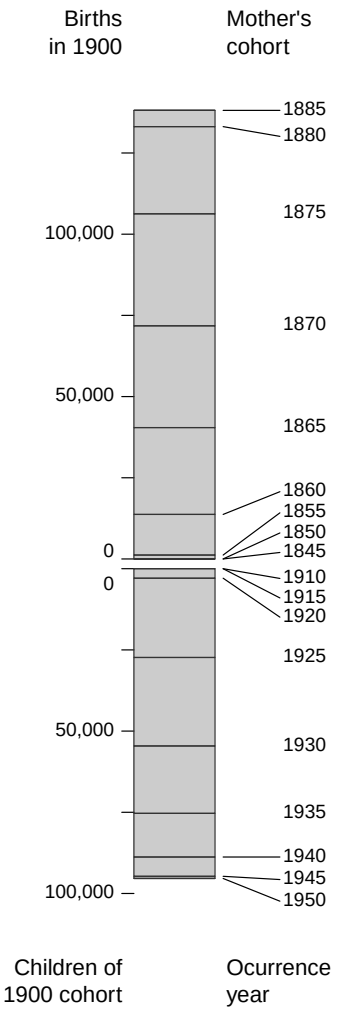


Figure 5: The 1900 cohort as a composite bar with its offspring reflected over y. The size of each bar stacked in the top composition is proportional to the area of its corresponding polygon in the left distribution of Fig. 3. The size of each bar stacked in the lower composition is proportional to the area of its corresponding polygon in the right distribution of Fig. 3.

Fig. 6 is a depiction of the exercise of Fig. 5, cohort bars on top reflected with offspring bars on the bottom. Equal bounded bins from Fig. 5 are joined into continuous regions. For the top region, filled polygons represent the births of mothers from quinquennial cohorts, spread over time. For the bottom region, filled polygons represent the mother-cohort origins of the births in quinquennial periods. Color darkness is proportional to the standard deviation of the birth distribution in each stacked polygon, with darker areas indicating more compact and lighter/brighter areas indicate wider distributions.

The meandering baseline of Fig. 6 is proportional to a smoothed time series of the crude cohort replacement rate. We overlay a horizontal line to indicate periods of approximate growth, replacement, and contraction. Periods where the meandering x-baseline is above this line (ca 1780 to 1860) indicate crude growth, and periods below the horizontal reference line (ca 1870 to 1930) indicate crude generation contraction. The nineteenth century was characterized by steady replacement by this measure.

[1] Young mothers are on top and older mothers on bottom for both distributions. It would also make sense to plot increasing (or decreasing) ages emanating out from the centerline in both directions.

[fold-out figure 4×a4 paper size at 100% in separate pdf, about here.]

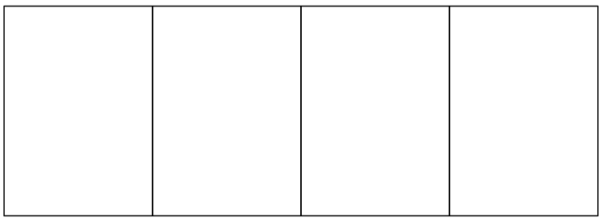


Figure 6: A time series of the same graphical construct as presented in Fig. 5. The x axis now meanders proportional to a smoothed time series of the crude cohort replacement ratio. Fill color darkness is proportional to the standard deviation of each birth distribution. The birth series now appears as a flow, but reveals echoes in cohort and offspring size, an odd periodicity in recent decades, and a long term dampening of the crude replacement rate. A 5-generation female lineage is annotated atop to serve as a guide.

To aid the viewer with interpretation, we overlay a known lineage of five female generations,[2] where x position is exact to the year, y position in the top region is matched to the mothers cohort, and y position in the bottom is matched to daughters' year of birth. Wider horizontal spacing between generations over time indicates increasing ages at maternity within this lineage (increasing from 23 to 39).

3 Discussion

[TODO: To be continued...This section and the following analytic perspectives section are only temporary. We probably won't want to introduce new] Several macro features come to the fore in this visualization. These are either known features of the Swedish birth series, or else merit further study. Echoes, booms, why is boom periodicity a recent phenomenon? Is there a dose-response to vertical reverberation in first derivative (can this be referred to as first or second order?) features.

3.1 Sebastian comments

One is that period effects seem to act differently in the three generations. If 1900 would have been a crisis year, especially younger grandmothers might have postponed births to a later year. Whether or not 1900 was a crisis year should be less relevant for when and in which quantity daughters of the 1900 cohort are born. Thus, my hunch would be that period effects most strongly affect the middle generation as it can only be born in one year. In addition, the effects of crisis conditions when the mother generation is born are likely to be more systematic for the grandmother compared to the daughter generation.

The issue of selective migration and conditional survival is briefly mentioned, but would require more discussion in case we stay in the perspective of generations. What is nice is that conditionality acts similar in both directions (though the effects of migration and conditional survival are likely to differ in quantity between generations). In order for mothers to be born in 1900, the grandmothers have to survive and to live in Sweden in 1900. In order to become an female offspring of the 1900 cohort, the mothers have to survive and to live in Sweden when giving birth. We should make more clear that this conditionality applies in all what we do. In that sense I also find it a bit confusing that we talk about crude reproduction. Perhaps we should rather call this conditional reproduction (conditional on survival and place of residence in Sweden at the time of giving birth).

The interpretation of generation arguments becomes particularly tricky as soon as a country faces substantial in-migration. For the immigrant mothers we lack information on conditional survival up to

[2]This lineage can be located in the public domain on `https://www.geni.com/people/Karin-Ottolina-Landsten/6000000022470480183`.

reproductive ages and the birth year distribution of the grandmothers. This makes statements about grandmaternal lags as in Figure 7 quite hypothetical as the immigrant mothers just contribute the generational length between mothers and daughters. Perhaps it would be safer to stay in the above described conditional thinking which would certainly support the use of the meandering x-baseline in the main plot. If we want to stick to thinking in generations, we might call them conditional generations.

4 Analytic perspectives

[TODO: this sort of thing ought to inspire discussion, but the stuff presently in this section may not belong in the paper. If so, then in online supplementary material perhaps, because the viability of the paper ought not depend on it.] A few macro patterns can be extracted from the data structure implied by the matrix $\mathbf{B}(c,t)$.

4.1 Analytic vignette 1

An example of macro patterns that can be extracted from this data structure include two-distribution location distance statistics, as demonstrated in Fig. 7. For each reference year in the range 1775 to [TODO: 1970] we have the mothers' cohort distribution and the next-generation childbirth year of occurrence distribution. Since both distributions are derived from the same table of counts by year of occurrence and mothers' age, it will help to use a simple notation, where the reference year is denoted with r, and following the index position convention $B(c,t)$ where mothers' cohort c takes the first and year of occurrence t the second position, respectively. In this way $B(c,r)$ are the births in year r to mothers from cohort c and $B(r,t)$ are the births in year t to mothers from cohort c (offspring). In this way, the weighted mean intergenerational lag $\overline{m2}$ is defined as

$$\overline{m2}(r) = \frac{\sum\sum(B(c,r) * B(r,t)) * (c-t)}{\sum\sum B(c,r) * B(r,t)} \quad , \tag{1}$$

and like-weighted distance quantiles may also be derived, as displayed in Fig. 7.

From Fig. 7 we learn that the mean two-generation maternal lag decreased in the mean and all quantiles in the 100 years from 1850 to 1950 by around 7 to 8 years. The interquartile, 95% and 99% spreads all decreased by about 1 year over the same period, and by more than another year in the following 20 years.

4.2 Analytic vignette 2

One of the most immediately visible features of Fig. 6 is the propagation of first differences in $B(t)$ to $B(c)$. The 1920 cohort is a particularly visible example: There were 23560 more births in 1920 than in 1919, an increase of 20.4%, and mothers from the 1920 cohort also gave birth to 20.7% more babies than the 1919 cohort. Fig. 8 displays the relationship in proportional first differences between matched birth cohort and offspring size. For the most part, the size of such structural echoes is maintained 1:1 in cohort offspring.

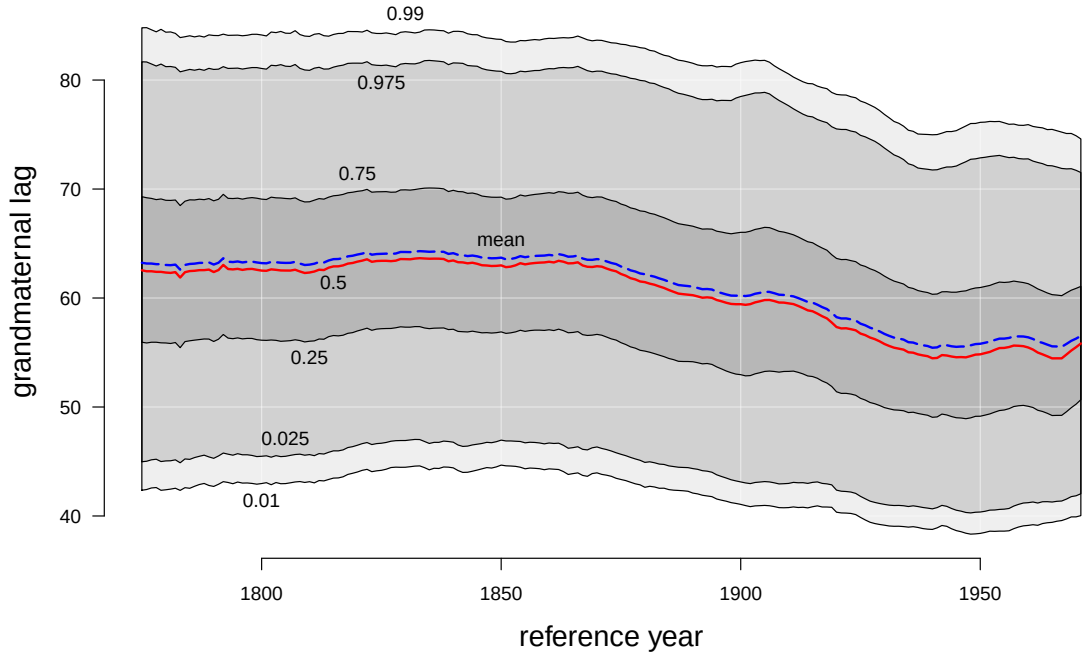


Figure 7: Time lag from mothers' cohort to next generation offspring year of birth, the *grandmaternal lag*, referenced to central (ego) cohorts. Decimals indicate birth distribution quantiles, where the red line indicates the median, and the blue dashed line the mean. The lag decreased broadly by all measures by ca 7-8 years in the 100 ego years from 1850 to 1950. The interquartile, 95% and 99% ranges also compressed by about 1 year in the same period.

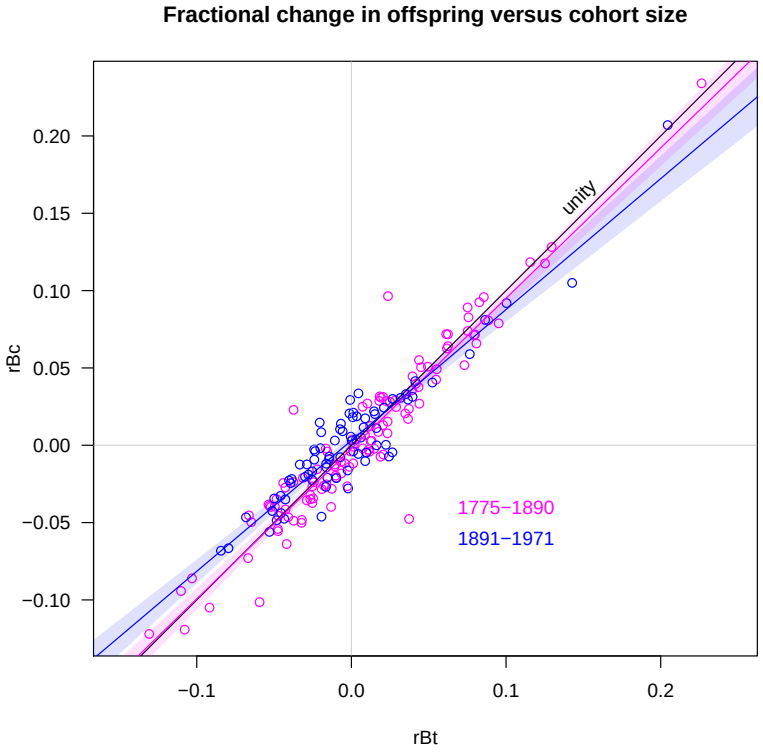


Figure 8: A roughly 1:1 dose-response relationship in relative size of structural echo.

References

W Brian Arthur. The ergodic theorems of demography: a simple proof. *Demography*, 19(4):439–445, 1982. doi: 10.2307/2061011.

Christina Bohk-Ewald, Peng Li, and Mikko Myrskylä. Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences*, 115 (37):9187–9192, 2018.

Norman H Carrier and AM Farrag. The reduction of errors in census populations for statistically underdeveloped countries. *Population Studies*, 12(3):240–285, 1959. doi: 10.1080/00324728.1959.10405023.

Joop de Beer. A time series model for cohort data. *Journal of the American Statistical Association*, 80 (391):525–530, 1985. doi: 10.1080/01621459.1985.10478149.

O Grigorieva, A Jasilioniene, DA Jdanov, P Grigoriev, T Sobotka, K Zeman, and VM Shkolnikov. Methods protocol for the human fertility collection, 2015. URL `https://www.fertilitydata.org/docs/methods.pdf`.

Louis Henry. Réflexions sur les taux de reproduction. *Population (french edition)*, 20(1):53–76, 1965. doi: 10.2307/1526986.

Human Fertility Collection. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). online, 2018. Available at www.fertilitydata.org (data downloaded on [Oct., 2018]).

Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). online, 2018. Available at www.humanfertility.org (data downloaded on [Oct., 2018]).

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany), 2018. Available at www.mortality.org or www.humanmortality.de (data downloaded on Oct., 2018).

Nathan Keyfitz. On the momentum of population growth. *Demography*, 8(1):71–80, 1971. doi: 10.2307/2060339.

Tim Riffe, Sean Fennell, and Jose Manuel Aburto. *DemoTools: Standardize, Evaluate, and Adjust Demographic Data*, 2018. URL `https://github.com/timriffe/DemoTools`. R package version 0.10.000.

Silvia Rizzi, Jutta Gampe, and Paul HC Eilers. Efficient estimation of smooth distributions from coarsely grouped data. *American journal of epidemiology*, 182(2):138–147, 2015. doi: 10.1093/aje/kwv020.

Henry S Shryock, Jacob S Siegel, and Elizabeth A Larmon. *The methods and materials of demography*. US Bureau of the Census, 1973.

Thomas Bond Sprague. Explanation of a new formula for interpolation. *Journal of the Institute of Actuaries*, 22(4):270–285, 1880. doi: 10.1017/S2046167400048242.

Statistika Centralbyrån. Historisk statistik för sverige. del 1, befolkning 1720-1967, 1969.

Statistique Générale de la France. *Statistique internationale du mouvement de la population daprès les registres détat civil: Résumé rétrospectif depuis lorigine des statistiques de létat civil jusquen 1905*. Imprimerie national, 1907.

A Data sources and adjustments

Data presented here are from several different data sources, covering different time periods. These data originate in different APC bins, and some are derived using indirect methods or projection methods. These full list of sources is outlined in Tab. 1. This appendix describes all steps taken to bring data into a standard format suitable for this study. The data format required to build these visualizations consists in birth counts tabulated by single year and cohort. Input data cover the years 1736 until 2016, with an oldest mother cohort of 1687. We complete the fertility of incomplete cohorts through 2016, bringing the latest year of occurence to [TODO: 2066].

From	To	Data	Bins	Use	Source
1736	1750	birth counts	totals only	constraint	SCB
1751	1755	life tables	single ages 0-110	reverse survival	HMD
1751	each	population census	abridged ages	base for retrojection	HMD
1751	1774	ASFR	single-age, 5-year	infer births 1751-1774 & derive retrojection standard	HFC
1751	1774	population exposures	single-age and year	infer births	HMD
1751	1774	birth counts	totals only	constraint	HMD
1775	1890	birth counts	abridged ages	constraint	SGF
1891	2016	birth counts	single ages	as-is	HFD
X	2016	ASFR	single ages	rate projection	HFD
2017	2060	Population projections	single ages	infer completed fertility	[TODO: SCB / UNPD]

Table 1: Data sources

A.1 Years 1736 - 1750

Estimating birth counts by single year of age for the 15 year period from 1736 to 1750 requires several steps of data processing and some strong assumptions. First, we reverse-survive females observed in the 1751 mid-year population census of Sweden, as extracted from the Human Mortality Database (2018) input database. Since this is a July 1 census, we take it as an acceptable proxy for exposure. The census originates in abridged ages $[0, 1, 3, 5, 10, 15 \ldots 90+]$. Examination of five year age groups suggests a underlying pattern age heaping, and for this reason we first smooth them using the so-called United Nations method (see Carrier and Farrag 1959) as implemented in the `DemoTools R` package (Riffe et al. 2018). We then graduate to single ages using the Sprague method (Sprague 1880, Shryock et al. 1973) as implemented in the `DemoTools R` package. This population is now the basis population to be reverse-survived through each single year until 1736, where we only make use of the fertile ages.

To reverse-survive, we use a standard survival curve defined as the age-specific arithmetic mean of the five single-age life table survival functions for the years 1751-1755 (Human Mortality Database 2018). The mid-year population count at age x, t years before the 1751 census $P(x, 1751 - n)$ is estimated as:

$$P(x, 1751 - n) = P(x + n, 1751) * \frac{\ell(x)}{\ell(x + n)} \quad , \tag{2}$$

where $\ell(x)$ is the standard survival function described.

The next step is to derive a standard ASFR curve, $F(x)$. ASFR for the years 1751-1775 is given by the Human Fertility Collection. (2018) in single ages and 5-year bins. If we rescale $F(x)$ in each 5-year period to sum to 1, one sees that there was very little shifting or shape changes in the period 1751: each unity-scaled $F(x)$ curve is for practical purposes equivalent. We therefore take the standard fertility rate curve, $F^\star(x)$ to be their age-specific arithmetic mean, and we assume that it is valid for the year-range 1736 until 1750.

A first pass of unscaled birth counts at age x, t years before 1751 is taken as the product of estimated exposure and $F^\star(x)$.

$$\widehat{B}^\star(x, 1751 - n) = P(x, 1751 - n) \cdot F^\star(x) \tag{3}$$

The first pass of birth estimates implies a TFR of 1 in each year. Total births for these years $B(1751 - n)$ is known (Statistika Centralbyrån 1969, Tab. 27 & Tab. 28), so we derive our final estimate of age specific births, $\widehat{B}(x, 1751 - n)$ as:

$$\widehat{B}(x, 1751-n) = \widehat{B}^{\star}(x, 1751-n) \cdot \frac{B(1751-n)}{\sum_{x=12}^{50} \widehat{B}^{\star}(x, 1751-n)} \tag{4}$$

At this stage of processing, birth count estimates for the years 1736-1750 are given in single ages (AP Lexis squares). Further adjustments are carried out in common with later periods and described in the following sections.

A.2 Years 1751 - 1774

Estimating birth counts in single years and by single year of age for the 24 year period from 1751 to 1774 follows a similar logic, but it requires no retrojection. The Human Fertility Collection. (2018) provides ASFR in single ages[3] in 5-year bins. The HMD provides exposure estimates $P(x,t)$ in single ages and years over this same period. A first-pass estimate of birth counts, $\widehat{B}(x,t)$ is given by:

$$\widehat{B}(x,t) = P(x,t) \cdot F(x,t') \qquad , \tag{5}$$

where t' denotes the 5-year bin in which t happens to fall. Year bins in the data are 5-years wide, and shifted up by 1, ergo 1751-1755, 1756-1760, and so forth. Following the convention of indexing to the lower bound, t' is defined as:

$$t' = 5\lfloor t/5 \rfloor + 1 \tag{6}$$

Births by single year of mothers' age are then rescaled to sum to the annual totals reported in the HMD:

$$B(x,t) = \widehat{B}(x,t) * \frac{B(t)}{\sum_{x=12}^{50} \widehat{B}(x,t)} \tag{7}$$

At this stage of processing, birth count estimates for the years 1751-1774 are given in single ages (AP Lexis squares). Further adjustments are carried out in common with later periods and described in the following sections.

A.3 Years 1775 - 1890

Birth counts for the 116 year period from 1775 to 1890 are available from Statistique Générale de la France (1907). These data are age-period classified, and given in a mixture of age classes, with a predominance 5-year age classes (especially for ages 20-50), but also sometimes single ages (especially for ages 15-19), and time-varying top and bottom open ages. We standardize these data in a few simple steps.

First, births of unknown maternal age were redistributed proportionally to the distribution of births of known maternal age. Second, counts are graduated to single ages using the graduation method proposed by Rizzi et al. (2015) and implemented in `R` in the package `ungroup`.

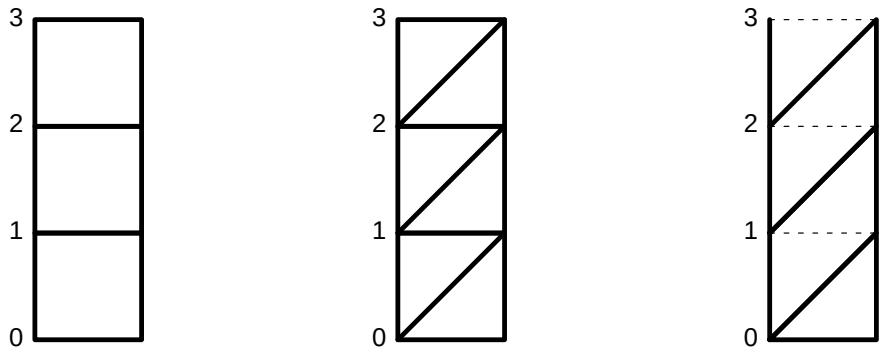
Third, counts were shifted into period-cohort Lexis bins assuming that half of the births in each single age x bin go to the lower triangle of age $x+1$ and half to the upper triangle of the age-reached-during-the-year (PC) parallelogram at age x, as diagrammed in Fig. 9.

At this stage data are binned and Lexis-conformable with HFD data for years 1891 and forward. With data processed as of step three, one could produce two time series represented in Fig. 6, with a subtle artifact visible in Fig. 10. In area **A** of this figure, birth counts in age bins have been graduated using the previously mentioned pclm method, which has the usually-desired artifact of smoothness. For the affected range of years, mother cohorts are identified via the identity $C = P - A - 1$.[4] Since age patterns of counts are smooth, these sum in Lexis diagonals to a smooth time series of cohort total offspring, as seen in the profile of area **B** of the same figure. Area **C** of this figure delimits years 1876 until 1971, where both cohort and matched offspring sizes are directly observed, and where fluctuations would appear to co-vary quite strongly. In the first instance for reasons of aesthetic continuity, and in the second instance for the sake of a more sensible count graduation, we have opted to adjust the counts in area **B** to carry the pattern of fluctuation observed over cohort size from 1775 to 1890.

This adjustment works by extracting the fluctuation pattern from **A** and transferring it to **B**. We do this by first smoothing the annual time series of total cohort size $B(t)$ according to some smoothness

[3] This data was graduated by the HFC from 5×5 Lexis cells according to the HFC methods protocol (Grigorieva et al. 2015).

[4] One subtracts 1 because data are in period-cohort bins.



(a) AP square bins (b) Split evenly to triangles (c) Regroup to PC bins

Figure 9: The count regrouping procedure for years 1776 to 1890, step three of data adjustment. Data are graduated to single ages (Fig. 9a), then split in half (Fig. 9b) and regrouped to period cohort (PC) bins (Fig. 9c).

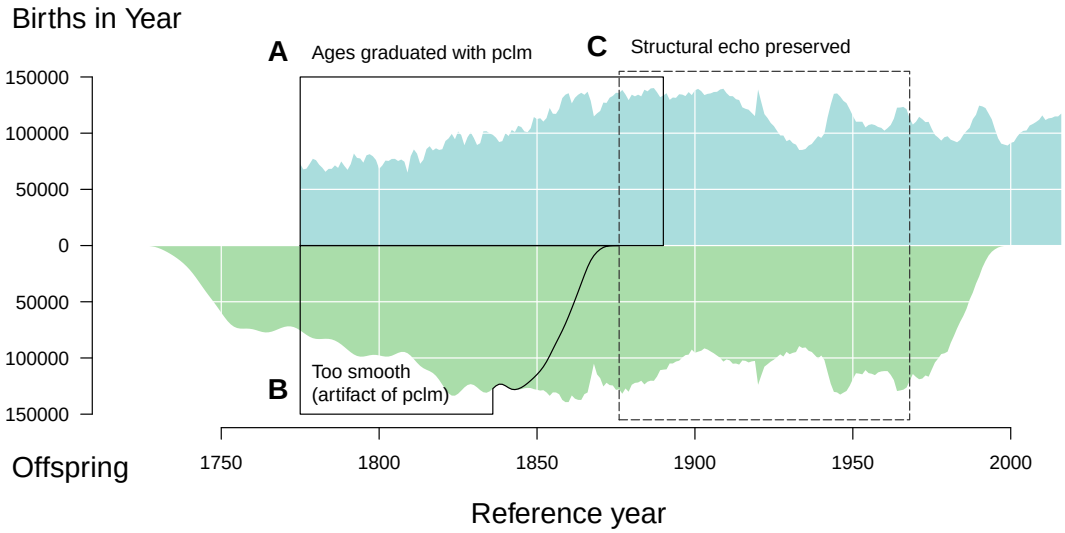


Figure 10: In reference years $\geq$ 1891 both births by year and cohort offspring are directly observed in single year bins, which means that the structural echo between total birth cohort and offspring size is preserved for reference years $\geq$ 1876 (**C**). Total per annum births in years $\leq$ 1890 (**A**) are presumed accurate, and so first differences of these are observed. Offspring from cohorts born in years $\leq$ 1876 (**B**) were partially (1836–1876) or entirely ($<$ 1836) born in years $\leq$ 1890, implying a smooth redistribution over single years of mother cohorts. We wish to adjust the births in **B** to recuperate the kind of structural echo in **C**.

parameter, λ.[5] The ratio of $B(t)$ to the smoothed birth series $B(t)^s$ defines the multiplicative adjustment factor, $adj(t) = B(t)/B(t)^s$. Total offspring size $B(c)$ is then adjusted as $B(c)' = adj(t) * B(c)$, for $c = t$. Counts in single ages are then rescaled to sum to the original totals in 5-year age groups, and counts for years $>$ 1890 are unaffected. The smoothing parameter is selected such that the linear relationship in fractional first differences $rd(B(t)) = \frac{B(t+1)-B(t)}{B(t)}$ between the annual birth series and adjusted offspring series $rd(B(c)')$ for years 1775-1890 matches that for the reference years 1877-1971 as closely as possible. Specifically, we select λ so as to minimize the sum of the difference in the slope and residual standard deviation for the periods before and after 1891. Further clarifications about this adjustment, and code for diagnostic plots can be found in the annotated code repository. The end effect is to adjust the series to look like Fig. 11.

[5] For the present case we've used a loess smoother, using the `R` function `loess()` with smoothing parameter $\lambda =$ `span`. It would be straightforward to swap this smoothing method out with a different one.

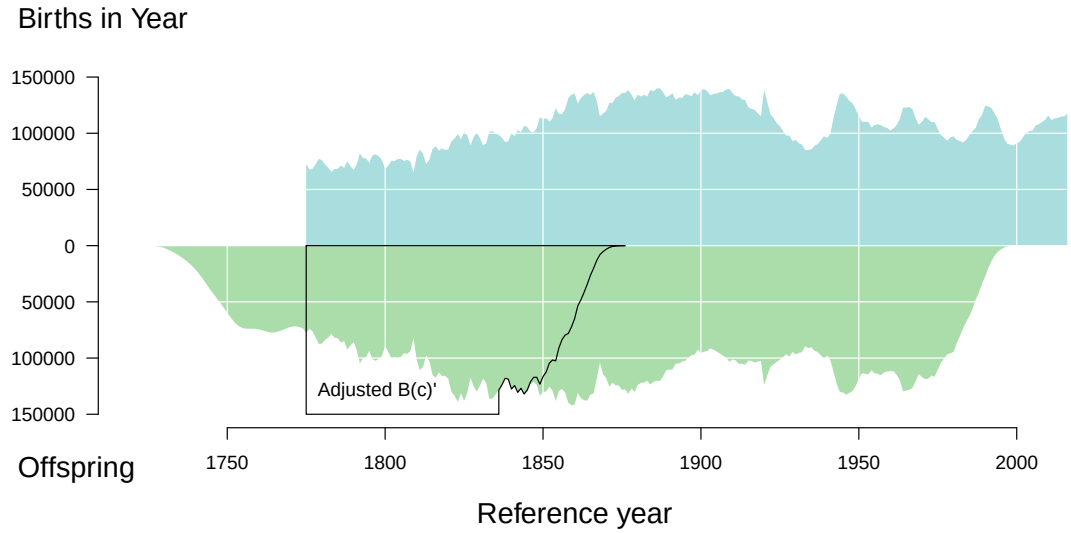


Figure 11: The adjusted birth series. Annual total births $B(t)$ on top axis and annual total offspring $B(c)$ on bottom axis, with adjusted offspring counts $B(c)'$ outlined.

We adjusted in this way for the sake of a more nuanced time series of total offspring, but this approach may be used to good effect in graduating age-structured counts (births, deaths, populations) whenever time series are long enough to permit information on birth cohort size to propagate through the Lexis surface. These aspects are visible to some degree in the shaded polygons of Fig. 6 in years < 1891.

A.4 Projected birth counts

[TODO: complete when exercise done] Offspring counts by year of occurrence, $B(c, t)$ are only fully observed for years ≤ 1971. To complete the reflection, we have opted to project birth counts for cohorts whose fertility careers are incomplete. This is done by combining a projection of cohort fertility rates using the method proposed by de Beer (1985) with a standard projection of population denominators (mortality projection too? Could also just take the pop projection from Statistics Sweden.). The method used is parsimonious, and it performed very well in a comprehensive assessment of fertility forecast methods (Bohk-Ewald et al. 2018). Light documentation to follow here, as well as an update of Fig. 11.

A.5 Meandering baseline

A peculiar feature of Fig. 6 is the meandering baseline, which replaces the standard straight-line x-axis. The baseline is derived from the crude cohort replacement rate $\mathbb{R}(c)$, defined as $\mathbb{R}(c) = B(c = r)/B(t = r)$. This measure is not a replacement for the classic measure of net reproduction R_0, which differs in a few key ways: i) crude replacement is not sex-specific (our birth series is composed of boy and girl births combined), whereas R_0 is typically defined for females only. ii) while births arise from fertility rates over the life course, the number of potential mothers over the life course is not a mere function of mortality, but of migration as well, and the Swedish birth series will have been affected by heavy out-migration from 1850 until the Second World War (cite SCB), and some in-migration in more recent decades. Cohort R_0 is purged of population structure such as this (except to the extent that subgroups have differential vital rates), whereas $\mathbb{R}(c)$ is not, and for this reason we call it *crude*.

The series of $\mathbb{R}(c)$ is rather smooth without further treatment, save for 11 periodic breaks between 1970 and 1840, a period of rupture between 1865 and 1880, and another set of at least four breaks since the great depression in the 1930s. Rather than preserve these ruptures, we opt to smooth them out and instead capture long term trends in $\mathbb{R}(c)$ in the baseline. Fig. 12. Keeping the baseline meander smooth minimizes the visual penalty in assessing the variation in $B(c)$ or $B(t)$ separately, and it enhances our ability to see the long term pattern.

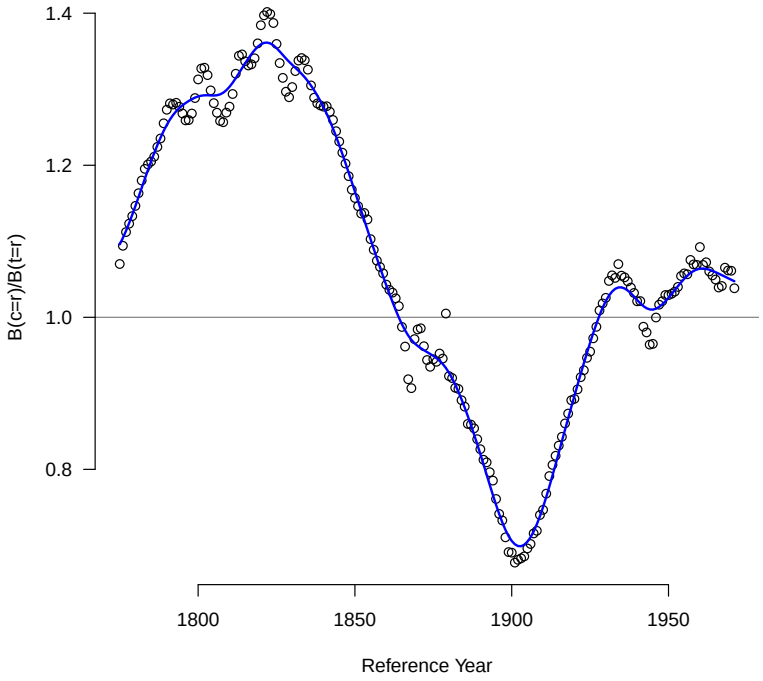


Figure 12: The time series of crude cohort replacement, $\mathbb{R}(c)$, and its smooth pattern (blue line) on which the Fig. 6 meandering baseline is based.