(Submitted as Research Note)

Title   Boom, Echo, Pulse, Flow: 385 Years of Swedish Births

Short title   Boom, Echo, Pulse, Flow

Authors   Tim Riffe*, Kieron Barclay, Sebastian Klüsener, Christina Bohk-Ewald

(*)   Corresponding: Tim Riffe `riffe@demogr.mpg.de`

Author information

Name	Affiliation	ORCID	🐦
T. Riffe	Max-Planck-Institute for Demographic Research	0000-0002-2673-4622	`@timriffe`
K. Barclay	Max-Planck-Institute for Demographic Research		`@BarclayKieron`
S. Klüsener	Federal Institute for Population Research	0000-0003-0436-3565	`@demomapper`
C. Bohk-Ewald	Max-Planck-Institute for Demographic Research		

Disclosure statement   We have no conflicts of interest with any aspect of this work.

Supplemental online material   We propose to locate the appendices in supplementary online material, since they are 1/2 of the manuscript. Per correspondence with the editors prior to submission, the very large Fig. 7 will likely need to go at the end of the print article rather than where it is indicated in the text. In this case we would change the text to make sure it is referred to appropriately. We propose to offer a redundant copy of this figure in the supplementary online material as well.

Boom, echo, pulse, flow: 385 years of Swedish births

Author(s) redacted

January 25, 2019

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 Swedish 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 biological 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 throughout the fertile age range, and since demography usually deals in large numbers, it is common to imagine births at the population level as a continuous stream or flow. 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 the present exposition, we retreat from rates, the basis of projections and stable population theory, to the absolute number of babies born, the raw material of population renewal.

We aim to represent a long time series of birth counts as a multilayered view of population renewal, a perspective enabled by Sweden's long history of population statistics. The birth series is rendered as a flow, in such a way as to simultaneously suggest several analytic perspectives, and to invite newcomers and curious minds deeper into the discipline of demography. This image, a large fold-out insert, entails investment from the viewer, and this manuscript serves as a protracted legend and caption. It is not a concise analytic plot, but rather a composite of hundreds of distributions, rendered both in the period and cohort perspectives. 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.

Our time series of birth counts stretches from 1735 until 2016, and it is augmented by a projection of the completed fertility of cohorts through 2016, bringing the latest year of birth to 2071. The temporal spread from the earliest mother cohort in our final data set (1687) to the latest year of birth occurrence (2071) is 385 years. We describe our input data and adjustments to it briefly in Sec. 2, and in a detailed set of appendices. Sec. 3 relates different age-structured birth count distributions to a common calendar. These distributions become the basic elements of our visualization. Sec. 4 gives a plain language description of how to interpret the visualization and a guide through some of its major features. Sec. 5 discusses the strengths and limitations of this particular visualization, and Sec. 6 concludes with a summary of this work.

2 Data and methods

We would like for our visualization to be an exposition on human renewal in general, but this work will be most useful if based on a long data series of good quality, ergo a single population with its own history and peculiarities. As Perozzo (1880), we base our exposition on Sweden because the data are available

and of adequate dimension, but there are two ways in which this fact is not to be taken for granted: i) The data comes from multiple sources and formats, it varies in quality, and it must be brought to a common Lexis resolution (i.e., a uniform grid of period and cohort bins) to be useful, and ii) although the visualization shows a history of birth in Sweden in particular, it may serve as a stimulus for multiple demographic models of a general nature.

The visualization is based on birth count data from Sweden in single-year bins by year of occurrence (period) and mother cohort. Birth counts for years 1736 to 1750 are reconstructed from a variety of sources (Human Fertility Collection, 2018, (HFC); Human Mortality Database, 2018, (HMD); Statistika Centralbyrån, 1969) using indirect methods (see App. A.1). Data for years 1751-1774 are derived via adjustment from HFC estimates and HMD exposures and birth totals (see App. A.2). Data for the period 1775 to 1890 come from Statistique Générale de la France (1907), which we have graduated from mostly five-year age groups into single ages (see App. A.3). The merged birth count series for years 1736 to 1890 is subject to a global adjustment to retain information on cohort size in period birth distributions (see App. A.4). Data for the years 1891 to 2016 is taken directly from the Human Fertility Database (2018) (HFD) without further adjustment, although the HFD itself did split single age birth counts into Lexis triangles for years 1891 to 1969 (Persson et al. 2018). To complete our picture, we project the fertility of cohorts whose fertility careers were still incomplete as of 2016 (1962-2016) through age 55 (see App. A.5). These steps are fully reproducible, and further details can be found in an open data and code repository.

3   Birth count distributions in period and cohort perspectives

The main fold-out visualization we present (Fig. 7) is a composition of smaller elements, each a birth distribution. We therefore build up to its full description by making explicit some basic notions of birth distributions and the calendar time line, drawing a distinction between two kinds of birth distributions featured in the final plot. This is intended to aid in the interpretation of the main figure.

In the period perspective, a picture of the births 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. This is key: In the period perspective with births arranged on an age axis, young mothers are on the left and older mothers on the right, but the reverse is true on a cohort axis. Count distributions such as this may be jagged, even if the underlying rate distributions are smooth, due to population structure.[1]

Given a long-enough time series of births classified by age and/or mother cohort, the full reproductive career of the cohort of individuals born in a particular year can be represented as a distribution (assuming no effects of migration). Since the childbearing of a cohort is spread over a synchronous span of ages and years, the x-indexing by age (Fig. 2a) or year (Fig. 2b) yields identical and redundant distributions:
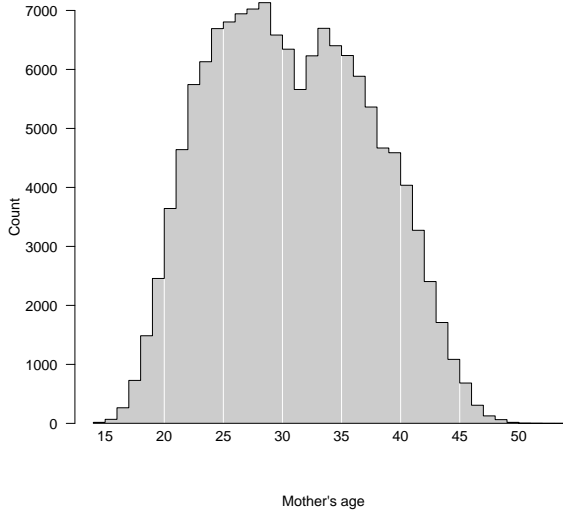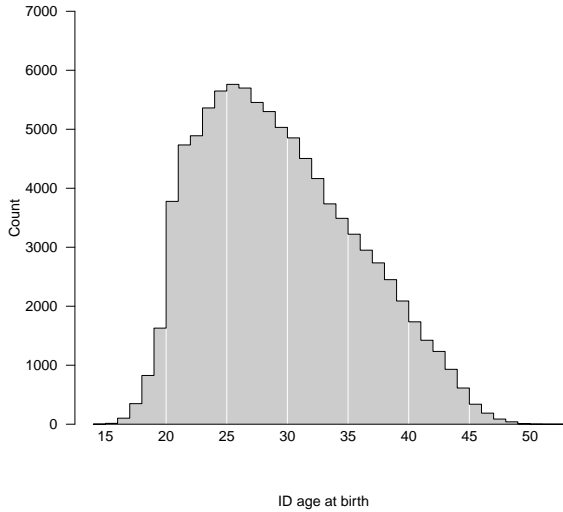
(a) Births in 1900 by age of mother

(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.

there is no reflection of young and old mothers when toggling between age and period classifications in the cohort perspective: In both cases young mothers are on the left and older mothers on the right. The distribution of births over the lifecourse of a cohort often resembles the smoothness of fertility rate schedules, but this is not necessarily the case.

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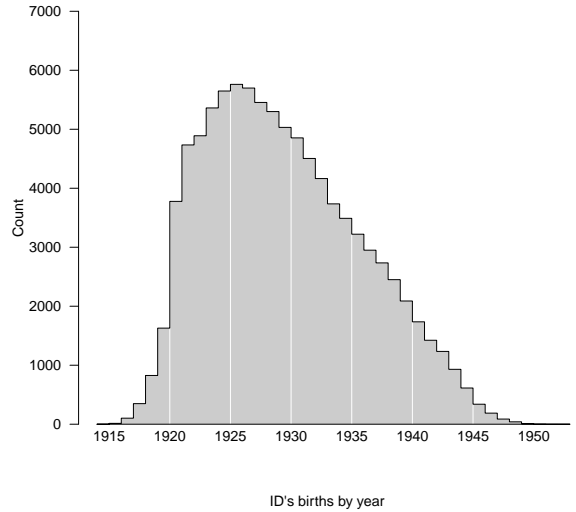
(b) Births from mothers born in 1900 by year

Figure 2: Births of a cohort structured by mothers' age versus mothers' year of birth are a shift over x.

The births in a year are classified by mothers' cohort, i.e. cohort *origins* in Fig. 1b, whereas the births *from* a cohort are classified *to* period in Fig. 2b. The two distributions are different in kind, but relatable

6

and both on a common scale. Importantly, these two distributions are indexed to the same calendar line. A fuller representation of their relationship would place them as two disjoint distributions on the same timeline, as in Fig 3.
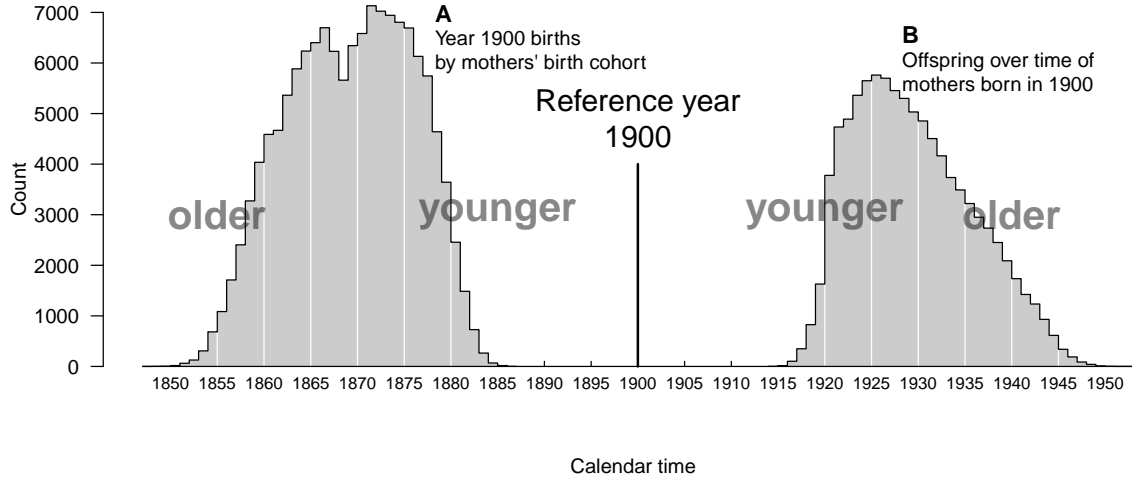

Figure 3: The cohort distribution of mothers who gave birth in 1900 (**A**) and the births from mothers born in 1900 by year (**B**). These two distributions imply *three* generations: the mothers of the 1900 cohort, the cohort of 1900 itself, and the offspring of the 1900 cohort.

The two distributions on the calendar axis in Fig. 3 are related, and of comparable scale, but different in kind. The x coordinate of the left distribution **A** is indexed to mothers' birth cohort (old to young), whereas the x coordinate of the right distribution **B** is indexed to child cohort, occurrence year (young to old). In this way the respective x coordinates are two generations apart, relating to each other as grandmothers and grandchildren. These are two distributions that we may wish to compare in various ways to better understand the birth series, but doing so for the entire Swedish dataset presents a practical challenge for a single visualization.

For the case of our extended Swedish birth series, we have 281 distribution pairs such as the year 1900 dual shown in Fig. 3, making simultaneous rendering impractical unless we transform the data in some way. An honest attempt might look like Fig. 4, where we reflect the left distribution **A** over y, keeping **B** on the top axis. These two distributions are linked by the year 1900, the reference year, which of course overlaps with neither of them. In this representation, distributions **A** and **B** are re-drawn for each possible reference year from 1736 to 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. These two drawbacks hide some macro patterns present in the data.

Still, the reflected axes in Fig. 4 produce at least two noteworthy artifacts that we may wish to preserve
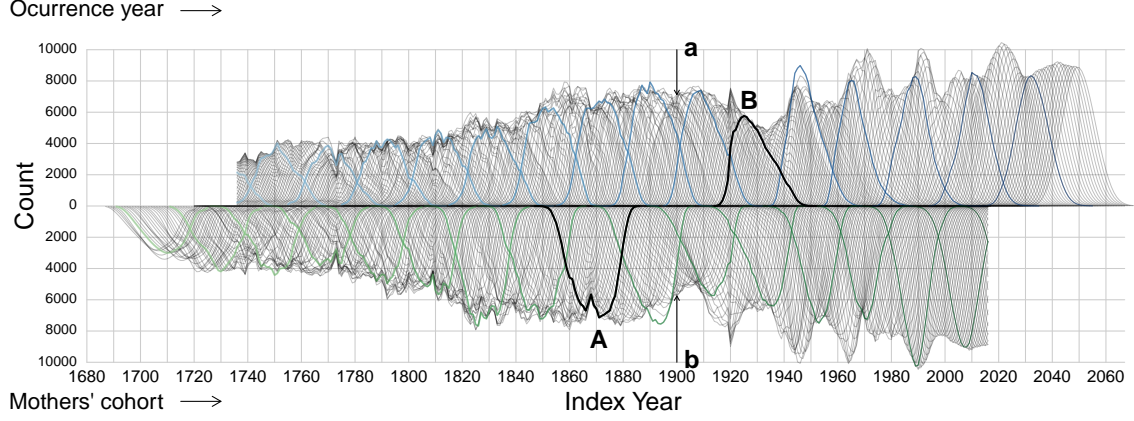
Figure 4: One time series of birth count distributions, under the period and cohort perspectives. The top series is composed of cohort offspring distributions indexed to period. The bottom series is composed of period birth distributions indexed to mothers' cohort. **A** and **B** are the same as in Fig. 3. The cross-section **a** gives **A** and the cross-section **b** gives **B**.

and clarify: i) First order differences in the top series appear to cascade into the lower series— This highlights a small constituent of population momentum (Keyfitz 1971): larger cohorts tend to have more offspring than smaller neighboring cohorts and vice versa, sudden fertility rate changes notwithstanding. ii) 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 second observation merits further elaboration: Since we have re-plotted the same series twice, each birth and each distribution is present in the plot twice. **A** and **B**, the two distributions we would like to juxtapose in bulk, are found indexed to the same reference year in the cross sections **a** and **b**. Alignment on a single reference year ought to facilitate comparison, but this visual task is stymied because the points constituting cross-sections **a** and **b** are heavily overlapped.

If instead we stack the slices that are indecipherably overlapped in **a** and **b** we get something like that shown in Fig. 5, cumulative birth distributions. Here the total bar length is proportional to the total cohort size of **a** and total offspring size of **b**. Stacked bins reflect 5-year mother cohorts in **a** and 5-year periods in **b**. 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 eventual mothers gave birth to 95,379 babies over the course of their lives, a crude replacement of 69%. Other perspectives on reproduction that account for survival and attrition or growth of the mother cohort through migration would give a more optimistic assessment of reproductivity (Henry 1965), a subject we return to in the discussion. The key feature of Fig. 5 is that the two distributions that are disjoint when drawn on a calendar line in Fig. 3, and that are hard to pick out in Fig. 4 can now be associated at a common x coordinate, and with distinguishable bins. This transformation allows us to view the time series of Fig. 4 with greater clarity, and it is the plot element from which our main Fig. 7 is composed.
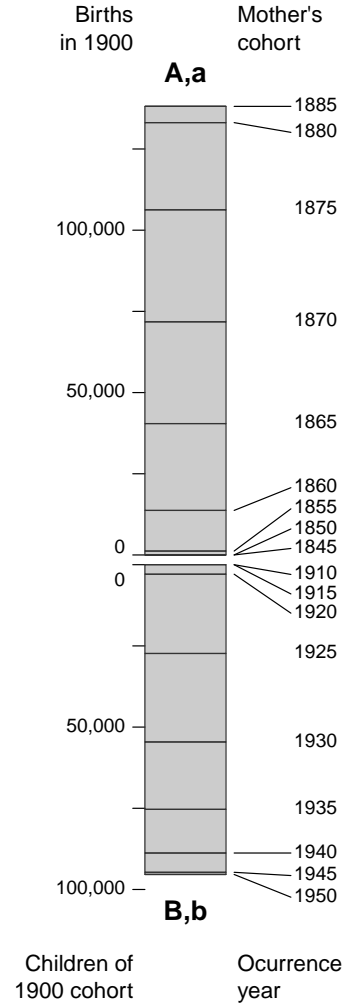

Figure 5: The 1900 cohort as a composite bar with its offspring reflected over y. The size of each bar stacked in the top composition is proportional to the area of its corresponding polygon in the left distribution **A** of Fig. 3. The size of each bar stacked in the lower composition is proportional to the area of its corresponding polygon in the right distribution **B** of Fig. 3. This is the same as stacking the slices of **a** and **b** from Fig. 4

If we repeat the exercise of Fig. 5 for each reference year in our data, and then merge neighboring like-bounded bins, we get something like Fig. 6. This is just the same as taking the distributions of Fig. 4, grouping in quinequennial bins, and then stacking them instead of overlapping them. That is, the filled polygons on the top axis represent the births of mothers from quinquennial cohorts, spread over occurrence years. On the bottom axis, filled polygons show the births in quinquennial periods indexed in x to mothers' cohort. This is the skeleton of our main visualization.
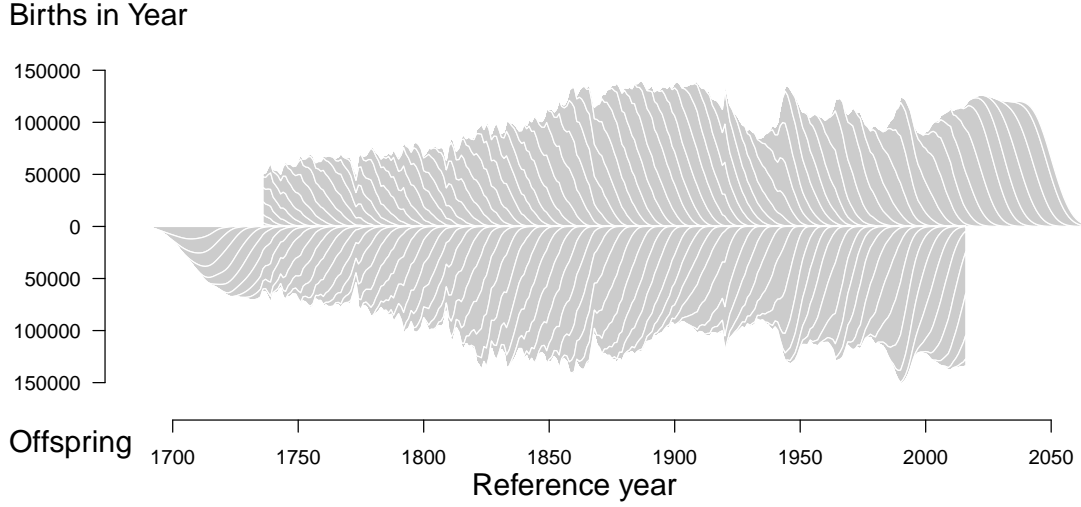
9

Figure 6: One time series of birth count distributions, under the period and cohort perspectives. The top series is composed of stacked polygons representing the quinquennial mother cohort birth distributions indexed to period. The bottom series is composed of stacked polygons representing quinquennial period birth distributions indexed to mothers' cohort. This is the underlying structure of fold-out Fig. 7.

4   A guide to the visualization

This section serves as a guide to interpreting our primary visualization (fold-out Fig. 7). We first describe the primary visual attributes so that the reader may seek meaning in patterns, and we then narrate some of the more salient features of the graph.

Meandering baseline   We start with the stacked polygon representation in Fig. 6, and subject it to a single y transformation before re-plotting in our primary figure. This transformation is a y shift proportional to a smoothed time series of the crude cohort replacement ratio (see App. A.6 for details), which in effect creates a meandering baseline rather than a flat x-axis. We call this crude replacement because the ratio takes no account of mortality or migration. A horizontal white line crosses the middle region of the plot showing where the baseline would have to be to indicate exact crude replacement. Regions where the baseline is above the horizontal cut point, such as the first half of the 19th century, are cohorts that produced a crude surplus of offspring, and vice versa. This transformation distorts the series somewhat, but we think in such a way as to make long term macro patterns more visible without losing the ability to compare proximate points on the calendar.

Profile   The height of each series vis-à-vis the baseline represents the total births in each year on the top series and the total cohort offspring on the bottom series (the children they had). Due to the meandering baseline, long term trends must be sought out with some effort by comparing the profile with the nearest meandering y grid line. This is not the kind of variation that the visualization focuses on. We consider

short and medium term fluctuations to be of greater interest in this visualization. The pattern of single-year matched fluctuations in cohort and offspring size is retained, as is the matched sinusoidal pattern of medium term booms and busts in the 20th century.

Stacked distributions   Shaded areas depict two kinds of birth distributions. On the top axis, each polygon is the offspring of a mother cohort, indexed to the calendar years in which the births occurred. Each distribution runs from left to right, stacked atop the previous cohort distribution. Therefore, births of young mothers are those closest to the outer profile, and births of older mothers approach the baseline. On the bottom axis, each polygon is the distribution of births in an occurrence year, back-indexed to the mother cohort x position on the calendar. Each distribution runs from older mothers on the left (outer profile) to younger mothers on the right (approaching baseline).

Shading color   Each polygon stacked in this visualization represents a birth distribution. We would like to highlight some characteristic of these distributions that may enhance our ability to detect patterns in this series. We opt to color based on the spread of each distribution, as defined by the birth-weighted standard deviation of maternal age at birth. The top and bottom shade colors are drawn from the same colorblind-friendly palette (Garnier 2018), where light colors mark wider distributions and dark colors mark relatively compact distributions. Underlying values range from 4.9 to 6.7 for period distributions (bottom) and from 5.1 to 6.7 for cohort distributions (top). There is a clear shift from wider to narrower distributions for cohorts born after around 1910 and occurrence years after about 1960.

Lineage   To aid the viewer with interpretation, we overlay a five-generation maternal lineage, which includes Alva Myrdal (born Reimer), who as much as anyone ought to remind us of the endogenous forces in the birth series we depict.[2] Since the top and bottom portions of the graph are alternative depictions of the same data, each member of this lineage appears twice: once below her mother at the same x position, and once again on the top axis x-indexed to the year of birth and y indexed to mother cohort.

Anna Lisa was born in 1829, destined to become Alva's great grandmother. Anna Lisa gave birth to Anna Sofia in 1856. Anna Sofia therefore belongs to the offspring of the 1829 cohort, appearing directly below Anna Lisa and inside the polygon for births occurring in years 1856. The arc connecting Anna Sofia in 1829 on the bottom axis (offspring) with herself in 1856 on the top axis (appearance in birth series) is a self-link, where displacement in x is equal to her mother Anna Lisa's age at birth (27), and top y position places her inside the polygon for births from mothers born 1829. Anna Sofia gave birth to Alva's mother Lovisa in 1877, who gave birth to Alva in 1902, who in turn gave birth to Sissela in 1934. This descendancy continues, but births beyond Sissela are not depicted. The lineage narrative may aid the viewer in understanding how the top and bottom portions of the graph relate as two perspectives on one and the same series.
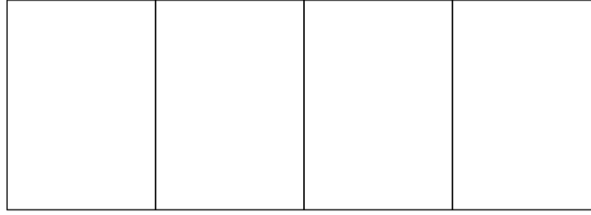
Figure 7: A period and cohort representation of the Swedish birth series. The top y axis indexes the births occurred in each year, broken down by mother cohort. The bottom y axis indexes the births that each mother cohort had over the course of their lives, broken down by year of occurrence. The x axis meanders proportional to a smoothed time series of the crude cohort replacement ratio. Fill color darkness is proportional to the standard deviation of the time spread of each birth count distribution: darker colors indicate more concentrated birth distributions and light colors indicate wider distributions. The birth series now appears as a flow, but reveals echoes in mother cohort size and respective total offspring size, a strong periodic series of booms and busts in recent decades, and a long term dampening of the crude replacement ratio. Five generations of a female lineage are annotated atop to serve as a guide.

Event timeline Several of the larger first differences in the period birth series (which cascade into cohort offspring size) have likely explanations, in most cases owing to mortality and health shocks. A selection of these are labelled, and Utterström (1954) provides complementary discussion. These labels are included primarily to satisfy inevitable curiosity, but the feature of the data that we wish to draw attention to is the very cascading of such deviations into the series of offspring size. For example, the depression in births in 1919 due to pregnancy loss during the Spanish influenza pandemic and surge in 1920 due to recovery from the same (Boberg-Fazlic et al. 2017) is naturally an event of scientific interest. This prominent deviation is for us a discussion point for its twofold echo— both in terms of offspring size (bottom axis) and imaginably as an amplifier of the well-known baby boom. Mothers from the 1920 cohort were the largest single contributor to cohorts born in the eight years from 1943 to 1950, the so-called first wave of the Swedish baby boom.[3] Only two other cohorts in this series, 1792 and 1811, might have been so dominant.[4]

5 Discussion

Several macro features of the Swedish birth series come to the fore in our visualization. These are either known features of the Swedish birth series, illustrations of some aspect of demographic thinking, or else merit further study. We briefly discuss how this visualization may inspire reflection on a set of stylized themes, including feedback, mixture, female dominance, and reproductivity. These are for the sake of provocation, and other themes may also come to the fore in the eye of the viewer. That these themes may be evoked does not mean that our visualization would be an appropriate analytic graphic to represent the given phenomena or concept. The value of this visualization may be in the simultaneous appeal to

such thinking.

Vibration, echoes, and cyclicity This visualization highlights the matched deviations in the size of cohorts and their offspring, which make the series appear to vibrate, especially in reference years before 1860. We have not found literature on the mirrored pattern of single-year deviations in cohort and offspring size. Since these are in the first place an artifact of structure, they may not be mysterious enough to study, but by extension we also know nothing of their consequences at any dimension or scale of observation. There is a literature that seeks to understand the medium term periodic fluctuations in birth cohort size (This literature largely derives from Lee 1974), and more often the interplay between demographic cycles and society (e.g., Easterlin 1987). Cycles of this kind are visible in the medium-term smooth waves whose peaks (troughs) have been roughly equally spaced since 1920 (1935). The fertility and population projections we use to complete the fertility of cohorts through 2016 reveal no continuation of this pattern, but the most recent wave is all but guaranteed to continue in the form of an offspring echo. We estimate this echo will be centered on the 1990 cohort, which will end up having more offspring than any before it in Sweden.

Temporal mixture Following the color pattern from left to right with blurred eyes, one gets a sense that the speed and rhythm of generational mixing has changed over time in Sweden, and this we think is a primary message of this visualization. To see how, consider a reference cohort as a kind of time transfer, relating a distribution of mothers to a distribution of offspring (see. Fig. 3). These two distributions may never overlap in human populations, due to the age of menarche[5], and their respective compactness may vary independently. At least two macro patterns in the distributions can be seen through the color pattern: i) The series undergoes a structural shift from light hues coloring wider and more overlapped distributions to dark hues coloring narrower and more steeply stacked distributions. This shift is centered around reference year 1940 — that is, on the mother cohort 1910 or occurrence year 1965. ii) Gentle undulations approximately 30 years (or a generation) apart prevail through both the early and late parts of the series, but their amplitude is dwarfed by the previous mentioned structural shift.

A female dominant view These observations and conjectures are predicated on the assumption that father age distributions either do not matter or else that they change in sync with mother age distributions. This begs the question as to whether the revealed patterns (vibration, replacement, mixture) are preserved, dampened, or magnified when indexed instead to father cohorts. Such sex comparisons have been made for migration-adjusted male and female net reproductivity (Hyrenius 1951) in Sweden, with males' net reproductivity often higher than females' between 1850 and 1950. However, we do not know how the degree of first difference reflection (vibration) would compare for males and females: if the correlation turned out to be stronger for males, what would that say about the common assumption

of female dominance in models of human renewal? Certainly the joint consideration of both mothers' and fathers' cohorts would lead to a more complete temporal mixture than what we may surmise from mothers' cohorts alone: Changes in age heterogamy over time would lead to a change in the pattern of temporal mixture, with more age-homogamous parentage leading to more directed time transfers in much the same way as compact birth distributions. Such questions could be investigated to some degree using large genealogical databases.

Reproductivity Birth cohort sizes more than doubled from the mid 18th Century until the late 19th Century in Sweden, after which time periodic fluctuations in size have ranged between 80000 and 140000, with no obvious long term trend of growth or decrease in the last century or over our projected horizon. Offspring size doubled from the mid 18th Century until the mid 19th Century, falling until around 1900, after which time it has grown in periodic spurts, with the 1990 cohort's offspring projected to be the largest in our series at over 150000. These two series relate in our baseline pattern of crude generation replacement, which shows several clear turning points. In the graph, the baseline relates to the transecting horizontal line, with points above showing crude growth and points below crude decrease. This measure of replacement reached a maximum in reference year 1822, a series minimum in 1901, with a projected local maximum around 1999. We reckon it impossible for a trained demographer to gaze at this visualization without either embracing or fighting against it as a literal representation of Sweden's reproductivity. As a picture of births alone, there is no account of population stocks in general, of attrition from mortality or emigration, or of increment through immigration. Accounting for any of these other phenomena would involve transformations away from the scale of absolute births and towards something closer to demographic behavior. There are other measures of reproductivity that take into account mortality (Kuczynski 1932), or both mortality and migration (Hyrenius 1951, Ortega and del Rey 2007, Preston and Wang 2007, Wilson et al. 2013, Ediev et al. 2014, inter alia), and which could therefore offer alternative values to arrange on the Lexis grid to recreate this work. Structure-driven shocks would in this case likely diminish or disappear, whereas behavior driven shocks might amplify, but the corresponding visualization would be one of forces or behavior rather than our tangible unit of births.

Questions These reflections can be translated into a set of tangible questions of varying scope and import. For example: What are the consequences of short term anomalies in cohort and offspring size, such as the 1920 cohort or those prior to the 20th Century? How do deviations in offspring size scale with the size of cohort size fluctuations? How do period and cohort cycles in the *spread* of the birth distribution relate, and what causes and consequences do they have? How would this picture (and its derived questions) change if indexed instead to father cohorts? What if the value depicted were rates rather than counts? What if the value depicted were stocks at some other age, but indexed in the same

way? We think that it may provoke many more questions than these.

6   Conclusions

We offer a visualization of the time series of Swedish birth counts, structured by two time measures: year of occurrence (period) and mothers' cohort. We index the same series twice, once to period (top) and once to cohort (bottom), which leads to a reflected set of axes, aligned to a single calendar. Each series consists in a set of sequentially stacked distributions. The period-indexed time series is a set of stacked distributions binned by mother cohorts. The cohort-indexed series is of stacked distributions binned by occurrence years. Each stacked distribution could be summarized in many ways, and we opted to color the distributions by the birth-weighted standard deviation of the maternal age at birth. The color pattern shows a transition to more compact birth distributions after the baby boom. Rather than *detrending* the series, we inject the series with a trend embedded in the calendar x-axis, which helps reveal the long-term pattern in crude generation replacement. A maternal lineage centered on Alva Myrdal in rendered atop the series to orient the viewer's interpretation of the visualization, as is a timeline of selected demographic shocks. In the discussion we highlight a few stylized themes and questions, omitting others for the sake of brevity.

Demographers are used to exploratory or explanatory graphs focused on a single diagnostic or pattern. Exploratory graphs are usually of a familiar form, and are quick to interpret, while explanatory graphs are further distilled to deliver a clear message. Our visualization is neither of these, but rather an excuse to give pause and reflect on the fundamentals of population renewal. The investment required to understand this visualization ought to draw the reader deeper into the joys and frustrations of demographic thought and practice.

Acknowledgements

(appendices probably for online-only supplementary material)

A   Data sources and adjustments

Data presented here are from several different data sources, covering different time periods. These data originate in different age-period-cohort (APC) bins, and some are derived using indirect methods or projection methods. The full list of sources is outlined in Tab. 1. Fig. 8 depicts the discrete Lexis bins of input data for the historical period before 1891. This appendix describes all steps taken to bring data into a standard format suitable for this study, and sections follow the order of the data processing steps undertaken. The data format required to build the figures in this manuscript consists in birth counts tabulated by single year of occurrence and mother cohort, the period-cohort (PC) Lexis shape (also denoted as 'VV' in HFD documents). Input data cover the years 1736 until 2016, with an oldest mother cohort of 1687. We complete the fertility of incomplete cohorts through 2016 up to age 55, bringing the latest year of occurrence to 2071.

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
1966	2016	ASFR	single ages	rate projection	HFD
2017	2071	Population projections	single ages	infer completed fertility	SCB

Table 1: Data sources
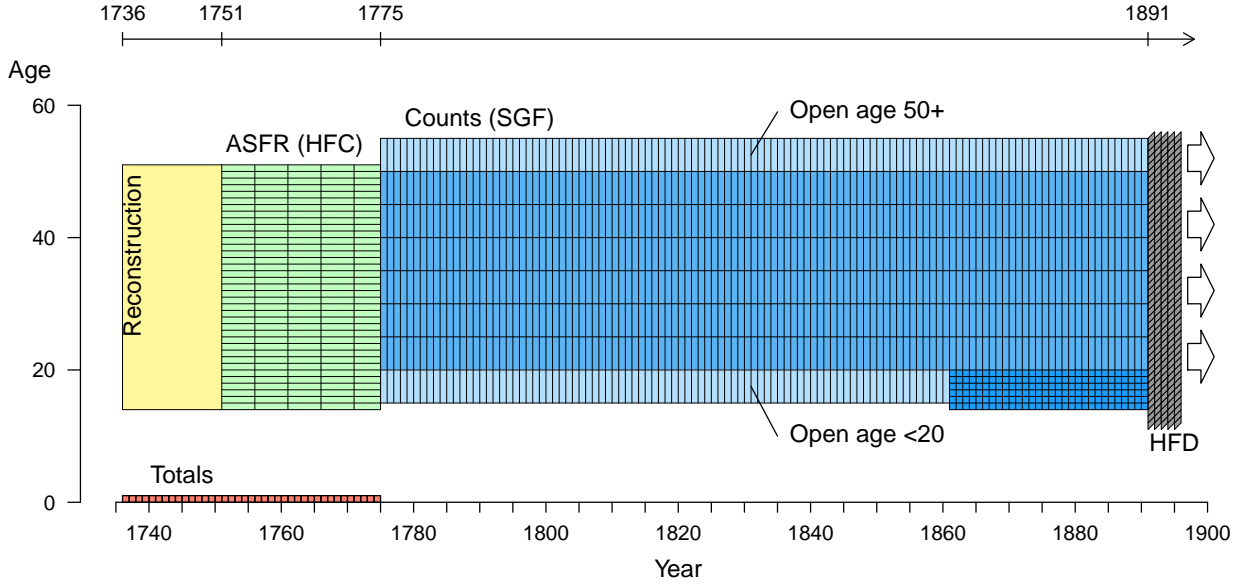
Figure 8: Lexis diagram depiction of fertility data sources and discrete Lexis bins for the period 1736-1890. Total counts are used as constraints for years 1736-1750, denoted with red boxes at age 0. The yellow box from 1736 to 1750 denotes the Lexis region over which birth counts are reconstructed using indirect techniques. The blue boxes from 1775 to 1890 indicate various age-period bins of birth counts. Light blue indicates lower and upper open age groups. Dark blue boxes in ages 14-19 from 1861-1890 indicate single year age-period bins. The HFD provides single-year period-cohort bins for birth counts starting in 1891 (gray).

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 $[0, 1, 3, 5, 10, 15 ... 90+]$. 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, n years before the 1751 census $P(x, 1751 - n)$ is estimated as:

$$P(x, 1751 - n) = P(x + n, 1751) * \frac{\ell(x)}{\ell(x + n)} \qquad , \qquad (1)$$

17

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

The next step is to derive a standard age-specific fertility rate (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 1751-1774: 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 $F^{\star}(x)$.

$$\widehat{B}^{\star}(x, 1751 - n) = P(x, 1751 - n) \cdot F^{\star}(x) \tag{2}$$

The first pass of birth estimates implies a total fertility rate of 1 in each year. Total births in each of 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:

$$\widehat{B}(x, 1751 - n) = \widehat{B}^{\star}(x, 1751 - n) \cdot \frac{B(1751 - n)}{\sum_{x=12}^{50} \widehat{B}^{\star}(x, 1751 - n)} \tag{3}$$

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[6] 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:

$$\widehat{B}(x, t) = P(x, t) \cdot F(x, t') \qquad , \tag{4}$$

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:

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

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

$$B(x, t) = \widehat{B}(x, t) * \frac{B(t)}{\sum_{x=12}^{50} \widehat{B}(x, t)} \tag{6}$$

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 (Pascariu et al. 2018).

A.4   Years 1736 - 1890

At this stage of processing all birth counts for years 1736 until 1890 are in single age-period (AP) bins, and datasets covering the three periods are merged into a common dataset. Two further adjustments are performed, the first to move AP into PC bins. The second adjustment compensates for the smoothness of graduation methods so as to preserve the expected relationship between a cohort's size its total offspring size.

A.4.1   Adjustment to PC bins

Counts were shifted from AP Lexis bins into PC 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. This simple assumption could be made more sensitive, but it would have no noticeable impact on our visualization. At this point data are in a common format with the HFD data for the years 1891-2016, and these are merged into a single dataset.

| (a) AP square bins | (b) Split evenly to triangles | (c) Regroup to PC bins |
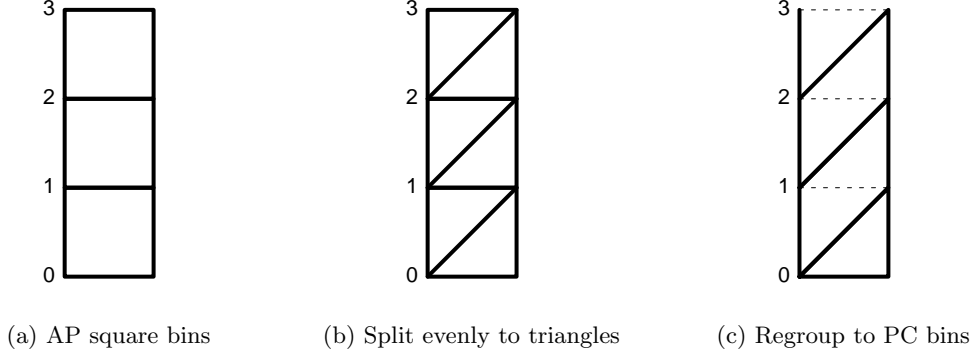
Figure 9: The count regrouping procedure for years 1736 to 1890. Data are graduated to single ages (Fig. 9a), then split in half (Fig. 9b) and regrouped to period-cohort (PC) bins (Fig. 9c).

A.4.2 Cohort size adjustment

At this stage of processing we have a harmonized dataset comprising a single series from 1736 until 2016 in consistent single-year PC bins. As such, one could produce the two time series represented in Fig. 7, albeit 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 $C = P - A - 1$.[7] 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 1875 until 1971, where both cohort and matched offspring sizes are directly observed,[8] and where fluctuations would appear to co-vary quite strongly. For the sake of a more sensible count graduation and for reasons of aesthetic continuity, we have opted to adjust the counts in area **B** to carry the pattern of fluctuation observed over cohort size from 1736 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 parameter, λ.[9] 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 $\mathbb{B}(c)$ is then adjusted as $\mathbb{B}(c)' = adj(t) * \mathbb{B}(c)$, 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 $rd(\mathbb{B}(c)')$ for years 1736-1876 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 slopes and residual standard deviations 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.

We adjusted in this way for the sake of a more nuanced time series of total offspring, but this approach
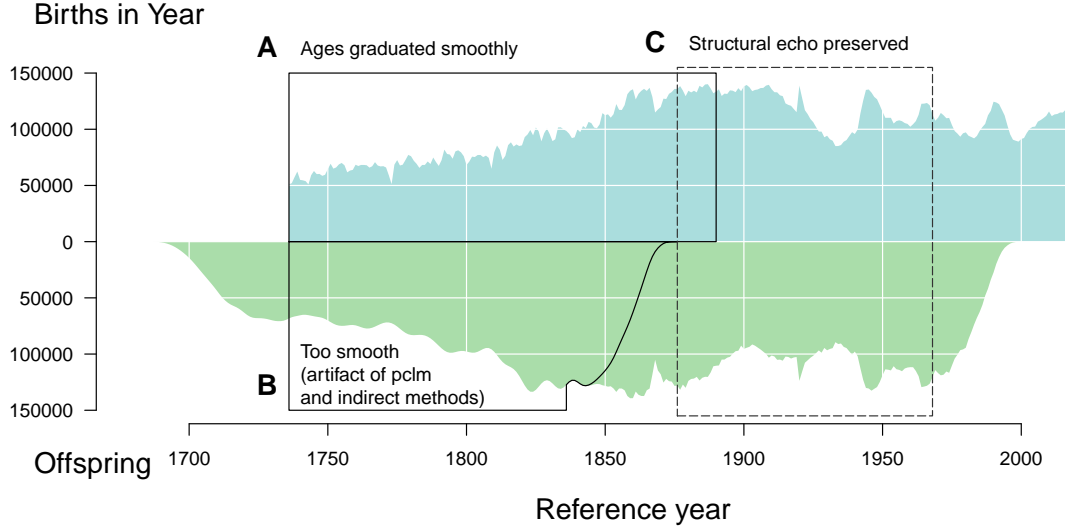
Figure 10: In reference years ≥ 1891 both births by year and cohort offspring are directly observed in single year bins, which means that the structural echo between total birth cohort and offspring size is preserved for reference years ≥ 1876 (**C**). Total per annum births in years ≤ 1890 (**A**) are presumed accurate, and so first differences of these are observed. Offspring from cohorts born in years ≤ 1876 (**B**) were partially (1836–1876) or entirely (< 1836) born in years ≤ 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**.
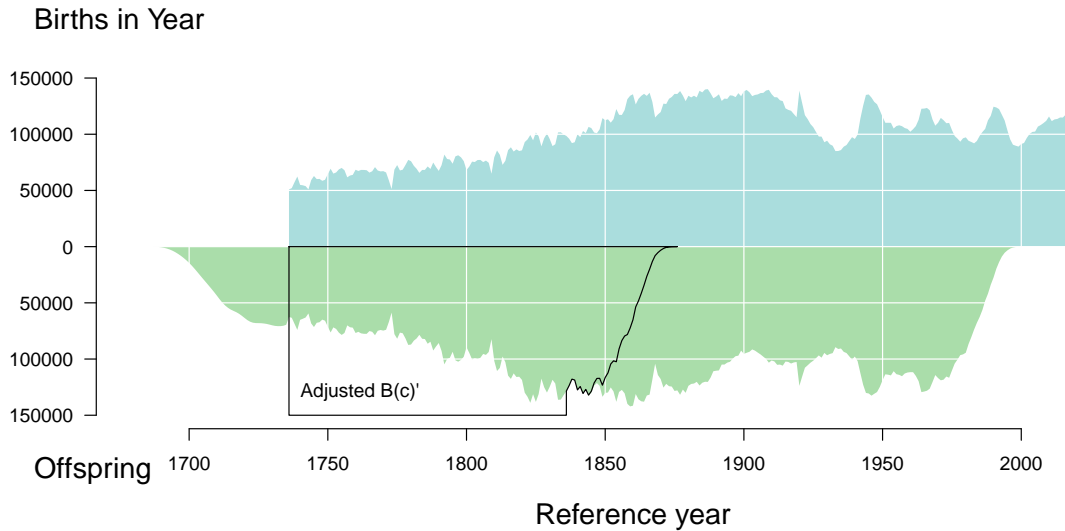
Figure 11: The adjusted birth series. Annual total births $B(t)$ on top axis and annual total offspring $\mathbb{B}(c)$ on bottom axis, with adjusted offspring counts $\mathbb{B}(c)'$ outlined.

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. 7 in years < 1891.

A.5   Projected birth counts

Offspring counts by year of occurrence, $B(c,t)$ are only fully observed for years ≤ 1961. To complete the offspring reflection through the final reference year 2016, we have opted to project birth counts for cohorts whose fertility careers are incomplete (1962-2016). This is done by combining a projection of cohort fertility rates using the method proposed by de Beer (1985) with Sweden's official projection of population denominators (Statistika Centralbyrån 2018a;b) to derive the implied birth counts by single year of age and time. The method used is parsimonious, and it performed very well in a comprehensive assessment of fertility forecast methods (Bohk-Ewald et al. 2018). This outstanding forecast performance can partly be explained with the model's ability to capture changing levels of fertility across ages and cohorts with two integrated ARIMA time series processes. In our application for Swedish birth counts the model of de Beer is fitted to observed fertility rates at ages 12 through 55 in calendar years 1967 to 2016 in order to complete the fertility for cohorts 1962 through 2004 and fully forecast the fertility of cohorts born 2005 through 2016.
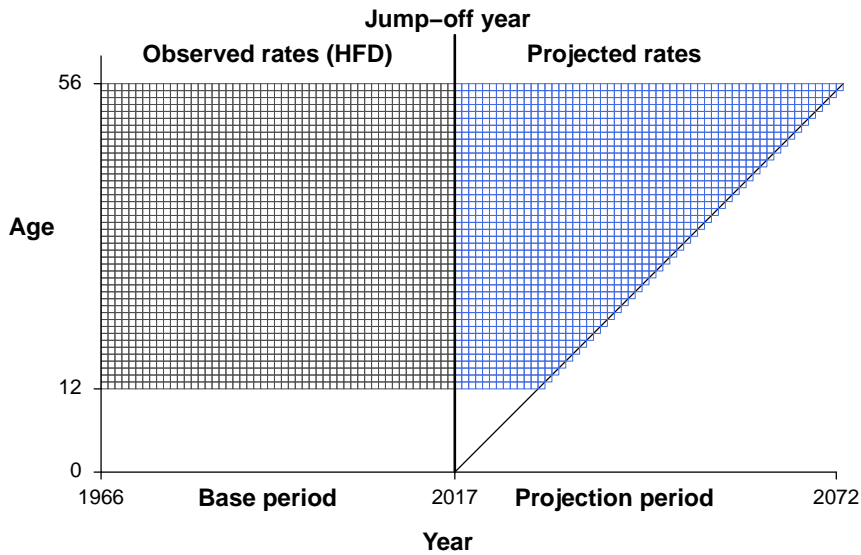
Figure 12: Diagram of the age and year-range of fertility rates used for model fitting and projection. We fit to 50 years of data (1966 until 2016), and project until the 2016 cohort (right bound of 2017) reaches age 55 in the year 2071 (right bound of 2072). The diagonal line delimits the data range after splitting to period-cohort bins.

Projected rates are then multiplied with SCB projected female population counts to derive implied births by age of mother. The projected birth counts are in single year age-period bins. These are then split to vertical parallelograms using the simple method described in Appendix A.4.1. The upward-slanting diagonal line in Fig. 12 marks the edge of the 2016 cohort fertility.

A.6 Meandering baseline

A peculiar feature of Fig. 7 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 $\mathbb{R}(c) = \mathbb{B}(c = r)/B(t = r)$. 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, 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, as seen in Fig. 13. Keeping the baseline meander smooth minimizes the visual penalty in assessing the variation in $\mathbb{B}(c)$ or $B(t)$ separately, and it enhances our ability to see the long term pattern. Specifically, we use use the `smooth.spline` from the `stats R` package, to smooth the $ln(\mathbb{R}(r))$ with smoothing parameter $\lambda = 0.00001$. The baseline that appears in Fig. 5 is the smooth prediction multiplied by 100,000.
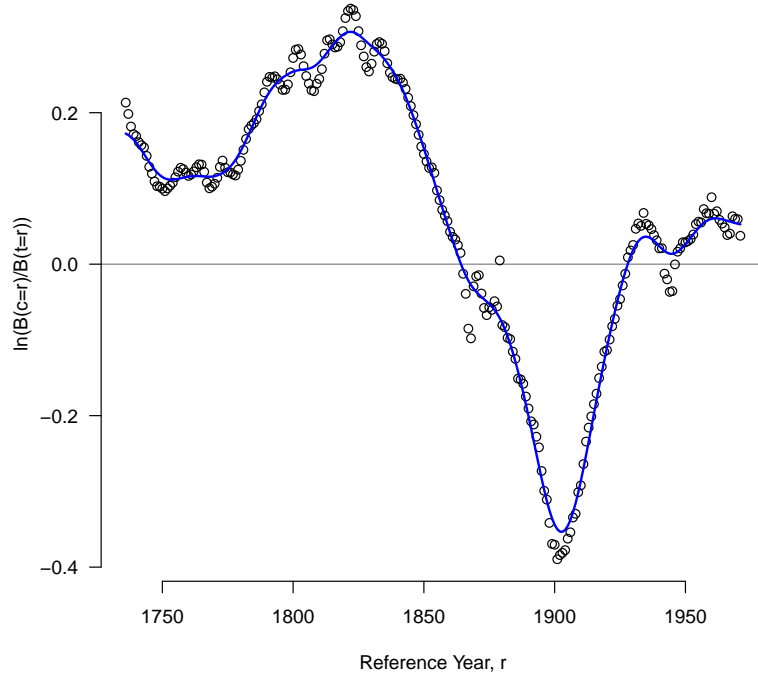
Figure 13: The time series of crude cohort replacement, $\mathbb{R}(c)$, and its smooth pattern (blue line) on which the Fig. 7 meandering baseline is based.

Notes

[1] The deficit around age 31 in 1a is due to a smaller number of potential mothers: Cohorts born between 1867 and 1869 were smaller than the surrounding cohorts due to a famine in those years.

[2] Alva Myrdal designed policies to make childbearing more compatible with women's work, to improve child well-being, and she was instrumental in other aspects of the Swedish welfare state. She also received a Nobel Peace Prize in 1982 for her work with the United Nations on disarmament, and for her influential writings on the topic of disarmament.

[3] By rough arithmetic, we estimate that the excess size of the 1920 cohort accounts for around 1% of first-wave (1940-1950) baby boomers in Sweden, or around 4-5% of the excess births making the first wave of the baby boom stand out in the first place. That is, there would have been a boom anyway, but we reason it was amplified by the 1920 birth anomaly.

[4] In our adjusted series, the 1792 mother cohort is the largest contributor to the nine cohorts born 1817 to 1825. The 1811 mother cohort is the largest contributor to the eight cohorts born 1837 until 1844. The 1849 cohort is the largest contributor to the six cohorts born 1877 to 1882. However, these three observations are uncertain with these data, and there is a risk it was induced by our own data adjustment described in App. A.4.2.

[5] This very time lag between distributions enables simple models of human renewal to produce period cycles (Wachter 1991).

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

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

[8] We estimate that the fertility of the 1971 cohort was over 99% complete as of 2016.

[9] 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.

List of Tables

List of Figures

References

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

Nina Boberg-Fazlic, Maryna Ivets, Martin Karlsson, and Therese Nilsson. Disease and Fertility: Evidence from the 1918 Influenza Pandemic in Sweden. IZA Discussion Papers 10834, Institute for the Study of Labor (IZA), June 2017. URL `https://ideas.repec.org/p/iza/izadps/dp10834.html`.

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. doi: 10.1073/pnas.1722364115.

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.

Richard A Easterlin. *Birth and fortune: The impact of numbers on personal welfare.* University of Chicago Press, 1987.

Dalkhat Ediev, David Coleman, and Sergei Scherbov. New measures of population reproduction for an era of high migration. *Population, Space and Place*, 20(7):622–645, 2014. doi: 10.1002/psp.1799.

Simon Garnier. *viridis: Default Color Maps from 'matplotlib'*, 2018. URL `https://CRAN.R-project.org/package=viridis`. R package version 0.5.1.

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 Dec., 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 Dec., 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 Dec., 2018).

Hannes Hyrenius. Reproduction and replacement: A methodological study of swedish population changes during 200 years. *Population Studies*, 4(4):421–431, 1951. doi: 10.1080/00324728.1951.10416787.

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

R.R. Kuczynski. *Fertility and reproduction: methods of measuring the balance of births and deaths.* Falcon Press, 1932.

Ronald Lee. The formal dynamics of controlled populations and the echo, the boom and the bust. *Demography*, 11(4):563–585, 1974. doi: 10.2307/2060471.

Juan Antonio Ortega and LA del Rey. Birth replacement ratios in europe: a new look at period replacement. [Unpublished] 2007. Presented at the Population Association of America 2007 Annual Meeting New York New York March 29-31 2007., 2007. URL `https://paa2007.princeton.edu/papers/71949`.

Marius D. Pascariu, Maciej J. Danko, Jonas Schoeley, and Silvia Rizzi. ungroup: An R package for efficient estimation of smooth distributions from coarsely binned data. *Journal of Open Source Software*, 3(29): 937, 2018. doi: 10.21105/joss.00937.

L. Perozzo. Della rappresentazione graphica di una collettivita di individui nella successione del tempo. *Annali di Statistica*, 12:1–16, 1880.

Lotta Persson, Gunnar Andersson, and Edward J Nash. Human fertility database documentation: Sweden. Technical report, Human Fertility Database, January 2018. URL `https://www.humanfertility.org/Docs/SWE/SWEcom.pdf`. Revised by Sebastian Klüsener.

Samuel H Preston and Haidong Wang. Intrinsic growth rates and net reproduction rates in the presence of migration. *Population and Development Review*, 33(4):357–666, 2007. doi: 10.1111/j.1728-4457.2007.00192.x.

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 0.12.32.

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*. AB Allmänna Förlaget, Stockholm, 1969. URL `https://www.scb.se/Grupp/Hitta_statistik/Historisk_statistik/_Dokument/Historisk-statistik-for-Sverige-Del-1.pdf`.

Statistika Centralbyrån. Population by age and sex. year 1860 - 2017. database, Dec. 2018a. URL `http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101_ _BE0101A/BefolkningR1860/`. (data downloaded Dec., 2018).

Statistika Centralbyrån. Population by region of birth, age and sex. year 2018 - 2120. database, Dec. 2018b. URL `http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0401_ _BE0401A/BefolkprognRevN`. (data downloaded Dec., 2018).

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. URL `https://catalog.hathitrust.org/Record/000542891`.

Fil Lic Gustaf Utterström. Some population problems in pre-industrial sweden. *Scandinavian Economic History Review*, 2(2):103–165, 1954. doi: 10.1080/03585522.1954.10407619.

Kenneth W Wachter. Pre-procreative ages in population stability and cyclicity. *Mathematical population studies*, 3(2):79–103, 1991. doi: 10.1080/08898489109525328.

Chris Wilson, Tomáš Sobotka, Lee Williamson, and Paul Boyle. Migration and intergenerational replacement in europe. *Population and Development Review*, 39(1):131–157, 2013. doi: 10.1111/j.1728-4457. 2013.00576.x.