UNIVERSITÀ DEGLI STUDI FIRENZE

DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Drivers of change in population age structures: a time series analysis

Gustavo De Santis, Giambattista Salinari

DISIA WORKING PAPER
2022/05

Drivers of change in population age structures: a time series analysis

Gustavo De Santis, gustavo.desantis@unifi.it
Dipartimento di Statistica, Informatica, Applicazioni – DiSIA, University of Florence, Italy

Giambattista Salinari, gsalinari@uniss.it
Dipartimento di Scienze Economiche e Aziendali, University of Sassari, Italy

Abstract

While "classical" demography imputes population ageing to low fertility, a recent "revisionist" line of thinking signals the emergence of ageing "from the top" (i.e., due to low mortality), starting slightly after World War II. We join this debate proving that, in the long run, mortality affects the population age structure, and therefore also ageing, more than customarily believed. With data taken from the Human Mortality Database on eight populations located in Europe, North America and Oceania, and for as far back as possible (up to 1820 in some cases), and applying cointegration analysis, we show that most of the historical change observed in the proportions of young, adult and old people in these countries can be derived solely from changes in survival, ignoring fertility and migration.

1. Introduction

Classical demography imputes population ageing to low fertility, not to low mortality (Coale 1956; Keyfitz 1975). When mortality is very high, its decline may even lead to population rejuvenation: this actually happened at the beginning of the demographic transition, characterized by a strong reduction in infant and child mortality (Coale 1972; Chesnais 1990, 1992). Extreme cases aside, Coale (1957) showed that, had fertility remained constant, the age structure of Sweden would have been practically the same in 1860 and 1950, despite strong mortality reduction. Bengtsson and Scott (2005, 2010) updated the exercise and confirmed that, with constant fertility, the proportion of people aged 65 and over in Sweden would have remained practically unchanged between 1900 and 2000. In both cases, so the argument goes, huge mortality progress proves of very small consequence on the population age structure, contrary to intuition.

After World War II, however, with improvements in survival concentrated at old and very old ages (Vaupel 2010) and the emergence of a sort of ageing "at the apex" (Bourgeoit-Pichat 1979), or "from the top" (Preston et al. 1989; Caselli and Vallin 1990), some "revisionist" scholars, as Lee and Zhou (2017) call them with specific reference to Preston and Stokes (2012), have challenged the idea that fertility is always the major driver of population ageing. Murphy (2017), for instance, shows that, in the past 65 years or so, it is essentially low mortality that prompted population ageing in 11 European countries. However, the idea of ageing from the top remains controversial: Lee and Zhou (2017) based on their own counterfactual analysis, maintain that even in "modern" populations, low fertility remains the main cause of population ageing.

While Murphy (2021) warns that the results of counterfactual population projections depend on the base year selected for the analysis and on population momentum, the main qualitative conclusions remain largely the same: fertility decline is at the root of population ageing, although the role of longer survival is no longer negligible nowadays.

In this paper, we show that mortality is, in the long run, a very good predictor of population ageing and, more in general, of the age structure of a population. To do this, we exploit two sets of empirical data, for various countries and several years. Our dependent variable is the share of individuals by age, or current (population) age structure. Our independent variable is the age structure of the *stationary* population associated with the current, cross-sectional life table, i.e. the L_x series: this we will call *reference* age structure. Several scholars question the validity of period life tables as indicators of the "true" evolution of survival, maintaining that cohort life tables should be used instead (e.g., Borgan and Keilman 2019). A theoretical discussion about this would be too long: in this paper, we indirectly contribute to the debate by showing that an empirical and strong relationship exists between these period life tables and the population age structures at (about) the same dates.

Our analysis relies on cointegration (section 2). However, readers who are not familiar with it can nonetheless follow our line of reasoning, which is trivial: we test whether the *reference age structure* (the proportion of individuals of age x in the stationary population associated with the period life table of year t) “attracts” the current age structure of a population, and whether this small, but persistent force of attraction eventually prevails, and shapes population age structures.

To simplify, two opposite outcomes are possible:

1) Our predicting capacity proves limited. In this case, the effect of mortality (or, at least, of period mortality) on the age structure is arguably small and the classical result of demographic analysis holds: something else (fertility arguably, but not mortality decline) is at the root of population ageing;
2) a large part of the age structure dynamics can be explained by the evolution of survival (as described by a succession of period life tables), and this in all the countries under scrutiny (eight, see below) and for the entire period of observation (the last two centuries or so). This outcome may lead scholars to reassess the classic interpretation that fertility guides population ageing.

We will try to convince our readers that our results point in this second direction: period life tables do “explain” a large part of observable population age structures.

The paper is organized as follows. The next section is devoted to the formal description of the test that we will use in this paper. The third section presents the data and some descriptive statistics. The fourth section contains a preliminary analysis of the relationship between the current and the reference age structure: even without great sophistication, the two age structures turn out to be very closely connected. The fifth section confirms these results, using a more refined approach: cointegration and the test introduced in section 2. The sixth and last section discusses our findings and their implications.

2. Testing cointegration between the current and the reference age structure

Let $C_{x,t}$ denote the population age structure expressed as the proportion of individuals of age x in year t

$$C_{x,t} = \frac{P_{x,t}}{P_t}, \tag{1}$$

where $P_{x,t}$ and P_t stand for the population of age x and in total in year t, respectively.

Similarly, the *reference age structure* of the population (i.e., the age structure of the current stationary population) can be defined as:

$$K_{x,t} = \frac{L_{x,t}}{\sum_x L_{x,t}} \tag{2}$$

where $L_{x,t}$ indicates the person-years lived at age x in the life table of year t. However, we will use the log transformations of $K_{x,t}$ and $C_{x,t}$, $k_{x,t}$ and $c_{x,t}$ respectively, for two main reasons: to circumvent one of the limitations of proportions (they are bounded in the 0–1 interval, with the lower limit, 0, particularly disturbing), and to better approximate linearity in the relationship between the two series.

The main goal of this paper is to test whether a long-run linear relationship exists between $c_{x,t}$ and $k_{x,t}$ for any given age x, and for varying t:

$$c_{x,t} = m_x + \gamma_x k_{x,t} + \varepsilon_{x,t}, \tag{3}$$

where m_x and γ_x are the model age-specific coefficients and $\varepsilon_{x,t}$ is the error term. If this relationship exists for a given age x, it becomes possible to predict the proportion of individuals of that age at time t from the reference proportion $k_{x,t}$. If this holds for all ages x_s, the entire age structure of year t can be derived from the life table of that year.

Eq. (3) does not rule out the possibility that also preceding values of k_x affect the current value of $c_{x,t}$. Suppose, for instance, that the "true" dynamic of the age structure is given by the following Autoregressive Distributed Lag (ADL) model:

$$c_{x,t} = \beta_{0,x} + \beta_{1,x}c_{x,t-1} + \beta_{2,x}k_{x,t} + \beta_{3,x}k_{x,t-1} + \varepsilon_{x,t} \tag{3'}$$

It can be proven that this form of dynamics corresponds to a long-run relationship of the type represented by eq. (3) where $m_x = \frac{\beta_{0,x}}{1-\beta_{1,x}}$ and $\gamma_x = \frac{\beta_{2,x}+\beta_{3,x}}{1-\beta_{1,x}}$ (see e.g. Johnston and Dinardo 2007:245. See also Pesaran 2015; Box et al 2016, and, for a simplified approach, Giles 2013).

If a long-run linear relationship between $c_{\mathrm{x},t}$ and $k_{\mathrm{x},t}$ actually exists, we are also interested in assessing whether eq. (3) can be approximated by simpler models such as:

$$c_{x,t} = \gamma_x k_{x,t} + \varepsilon_{x,t}, \tag{4}$$

$$c_{x,t} = k_{x,t} + \varepsilon_{x,t}, \tag{5}$$

Model (4) assumes that $m_{\mathrm{x}} = 0$, while model (5) introduces the additional assumption that $\gamma_{\mathrm{x}} = 1$ for all x's. The interest of eq. (5) lies in its simplicity: it says that the current population age structure of a population can be approximated by simply looking at its reference age structure, i.e. the life table for that year. This makes the estimation of m_{x} and γ_{x} unnecessary, and saves the need to collect and analyse long time series of both sets of data (current and reference age structures).

Unfortunately, assessing the existence of a long-run relationship between two time series of the type described in eq. (3) proves difficult, because spurious correlation must first be ruled out (Granger and Newbold 1974). If the "innovations" that affect these series have permanent effects on them (i.e., the series have a "unit root" in the econometric jargon), their dynamics are non-stationary (i.e., their means, variances and autocovariances depend on time) and the two series may appear to be correlated even if they are independent of each other. Arguably, permanent innovations play a fundamental role in the long-run dynamics of mortality and fertility. Think of vaccines, penicillin and antibiotics, coronary bypass, the formation of antibiotic-resistant bacteria, the spread of smoking and alcohol drinking, etc.: all with permanent effects on the evolution of mortality. Likewise, the reduction of infant and child mortality, the increase in the cost of children, the new role of women (higher education and greater participation in the labour market), etc., have permanently changed the evolution of fertility.

Two series are said to be cointegrated if

1. innovations produce permanent effects (unit roots) and
2. the series are (not spuriously) correlated in the long run, with a functional relation of the type described in eq. (3).

In classical cointegration analysis, therefore, one must first test for the presence of unit roots in the series under consideration, and then, if these are present, test for cointegration. The cointegration test is based on the residuals of eq. 3, not on its coefficients (Engle and Granger 1987). If the residuals turn out to be stationary (their mean, variance and covariance are independent of time), a long-run relationship between the two series is likely to exist. This test is usually performed *via* the Augmented Dickey-Fuller (ADF) unit root test, although other solutions are also possible.

To understand the rationale of this approach, let us define $\hat{c}_{x,t} = m_x + \gamma_x k_{x,t}$ as the log-proportion of individuals of age *x* in year *t* predicted by eq. 3 based on the reference age structure (better: on the reference log-proportion of individuals of age *x* in year *t*). If the current and the reference age structure are cointegrated, the current log-proportion of individuals aged x, $c_{x,t}$, will show a tendency to "revert" to its long-run equilibrium value $\hat{c}_{x,t}$. In practice, cointegration means that $c_{x,t}$ and $\hat{c}_{x,t}$ cannot be too far away from each other, because some "force" pushes $c_{x,t}$ towards $\hat{c}_{x,t}$. In this case, the residuals $\hat{c}_{x,t} - c_{x,t}$ will show a stationary dynamic (they will never diverge too much from zero), and this explains why the classical tests of cointegration focus on residuals.

Unfortunately, the standard test used to check whether time series are stationary often proves inconclusive. This "low power" of the classical cointegration test derives ultimately from the fact that three conditions must be met to prove cointegration:

1 and 2) the test must not reject the null hypothesis of a unit root for each of the two series under scrutiny (this is to confirm that innovations do produce permanent effects in both series);

3) the test must reject the null hypothesis of a unit root when it is applied to the analysis of residuals (this is to prove that one of the series tends to revert on the other).

These conditions are rarely met in practice: not necessarily because the underlying hypotheses are false, but because of "noise", such as data errors and, possibly, other intervening variables. Therefore, other, more powerful approaches have been proposed. Among these, the so-called "bounds test" (Pesaran et al. 2001), which, in a way, verifies the three aforementioned conditions with a single test.

The general idea behind this approach is to test the "reversion towards the long-run equilibrium", that is the "force of attraction" that we mentioned earlier, by estimating an Error Correction Model (ECM). The details of the test are briefly presented in Appendix A. What really matters here is that this test yields a result (a "statistic") called F.

For any given significance level, two critical values are offered: F_U and F_L, upper and lower, respectively (Pesaran 2015:526). This, which incidentally justifies the name, "bounds test", means that three possible outcomes are possible:

1) $F > F_U$ signals the likely existence of a long-run relationship;
2) $F < F_L$ indicates that the long-run relationship is unlikely to exist; and, in between,
3) $F_L < F < F_U$ leads to a "suspension verdict": the inference is inconclusive.

The third outcome of the bounds test may have various possible causes: for instance, it emerges when the series have "a different order of integration", which means that only one of them has a unit root, with innovations inducing permanent effects. In this case, further analyses are needed to verify the possible cointegration between the two series.

If $c_{x,t}$ and $k_{x,t}$ turn out to be cointegrated, one can estimate eq. 3 with OLS and use the $k_{x,t}$ series and the estimated coefficients to predict the evolution of $c_{x,t}$. The proportion of the overall variance of $c_{x,t}$ explained by $k_{x,t}$ is given by the simple R^2 statistics of the regression model.

3. The data

For our analysis, we used data taken from the Human Mortality Database (HMD) on eight populations located in Europe, North America and Oceania (Table 1). We selected these populations based on two main criteria:

1. The data start at least in the 1930s so that the series span at least 85 years;
2. The effects of the two world wars are limited either because the national territory was spared (as in the case of Australia, Canada, New Zealand, Sweden, Switzerland and the USA) or because the data allow us to focus only on the civilian population (as in the case of France and England-Wales).

We used data on exposures to calculate the current proportions of individuals of age *x* in year *t* ($C_{x,t}$), and period life tables to compute their reference counterpart, i.e. the proportions of life-years at age *x* out of the total ($K_{x,t}$). We focused on the ages between 0 and 99 years, by five-year age classes (0–4; 5–9; ...; 95–99).

The countries covered by our analysis have time series of varying lengths. Typically, data for the European populations are available since the 19th century, or even before (Swedish data, for instance, begin in the 18th century). However, we decided to start in 1820, or as early as possible after that, because older data are often scarcely reliable. In non-European countries, the series start typically between 1910 and 1930.

The age structures of our populations changed considerably over time. The mean age, for instance increased from about 25–26 years to about 38–39 years, while the mean age of the corresponding reference populations (i.e., the stationary populations associated with the current period life tables) passed from 30–33 to 39–40 years (Table 1). The relative weights of the youngest and the oldest age groups changed considerably in the last two centuries in both the current and the reference populations, while the relative weights of the central age groups barely varied. At these central ages, the limited variability of our independent variable $k_{x,t}$ reduces the explanatory power of the ECM (eq. A.1); in practice, however, this is less problematic than it seems, because also the dependent variable $c_{x,t}$ barely changes. The case of Sweden, the country that we will systematically use as an example in this paper, is shown in the box plots of Figure 1. Other countries, not presented here for reasons of space, behave similarly.

One of the difficulties that our analysis must face is represented by mortality crises: e.g., the cholera epidemics of the 19th century and the Spanish flu epidemics of 1918. These discontinuities may introduce several forms of distortion. The most problematic are probably those linked to fertility swings, which induce baby booms and busts and create "waves" in the age structure $C_{x,t}$ (but not in the reference age structure $K_{x,t}$, our independent variable) of the subsequent 100 years or so. This reduces the explanatory power of our model, which, however, remains high, as we will see shortly.

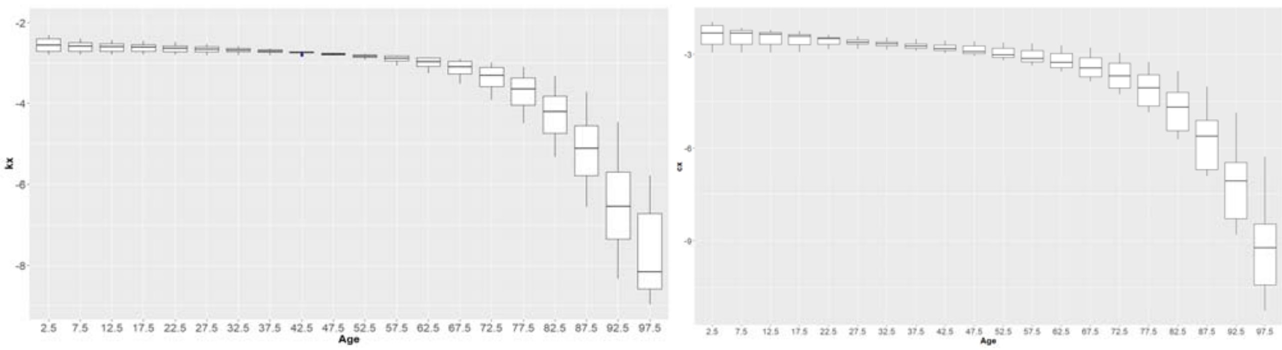
Another factor to keep under control is linearity, which is *assumed* in our two fundamental equations (3 and A.1), but which may not exist in real life. Figure 2 suggests that the linearity assumption generally holds, except for a small, but significant slope change in the passage to older ages (above 60 years). More refined statistical analyses (not shown here) indicate that there is also a slight (but statistically significant) convexity after 60 years, and a concavity before this age. All in all, however, the diagnostics of our ECM model (eq. A.1) reveal that these modest departures from linearity do not affect our results (see Table 3, below).

Table 1 *Summary statistics*

					Mean age of populations			
		Time range of series (years)			Current (C_x)		*Reference* (K_x)	
Country	Code	Start	End	Length	Start	End	*Start*	*End*
Australia	AUS	1921	2016	96	27.8	38.1	*35.5*	*42.1*
Canada	CAN	1921	2016	96	26.7	40.3	*35.7*	*41.9*
England&Wales	ENW	1841	2016	176	25.5	40.0	*31.7*	*41.3*
France	FRA	1820	2016	197	28.3	40.8	*31.4*	*42.0*
New Zealand	NZD	1901	2008	108	25.7	38.2	*34.8*	*41.3*
Sweden	SWE	1820	2017	198	27.4	40.7	*31.3*	*41.7*
Switzerland	CHE	1876	2016	141	28.1	41.5	*30.7*	*42.3*
United States	USA	1933	2017	85	29.6	38.7	*35.0*	*40.7*

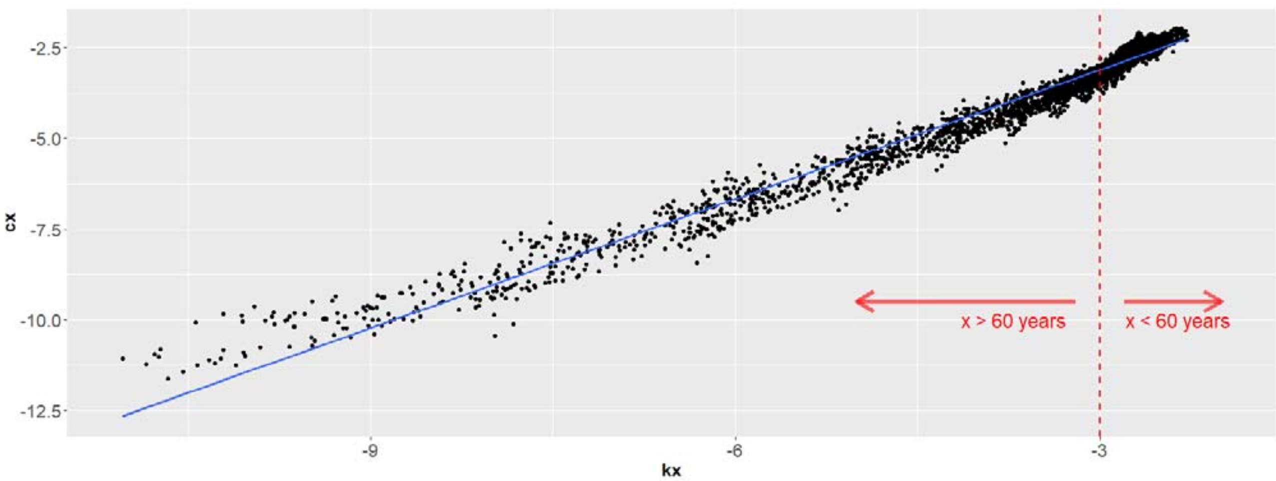
Note: in the longer country codes of the HMD ENW=GBRCENW; FRA=FRACNP; NZD= NZL_NM.
Reference=Stationary.
Source: Human Mortality Database.

Fig. 1 *Median values and variability of k_x and c_x by five-year age classes in Sweden (1820–2017)*



Note: k_x=ln(K_x); c_x=ln(C_x). Five-year ages are classes indicated with their central age.

Fig. 2 *Relationship between k_x and c_x for all the countries and years of Table 1*



Note: k_x=ln(K_x) and c_x=ln(C_x). Small values of K_x and C_x, and therefore lower values of k_x and c_x, (bottom left), characterise older ages.
Source: See Table 1.

4. The long-run relationship between the reference and the current age structure: A naïve analysis

Before tackling the cointegration analysis of section 5, let us present our argument (existence of a long-run relationship between the current and the reference age structure) in the simplest possible way.

Table 2 refers, once again, to Sweden and it shows, at different ages, the estimates of the parameters of eqs. 3 and 4, along with their standard errors and the associated R^2. The last three columns of Table 4 report the mean squared errors (MSE) of three models: the full one (eq. 3), the restricted one (intercept set to zero, eq. 4) and the twice-restricted one, with zero intercept and unitary slope (eq. 5). The general idea behind this exercise is to assess how close, on average, the current age structure is to its reference counterpart, and whether the use of simpler models is statistically justified.

***Table 2** Parameter estimates by age for Sweden (Eqs. 3, 4 and 5)*

	Eq. 3						Eq. 4			Eq. 3	Eq. 4	Eq. 5
Age	**Intercept**	**SE**	**Slope**	**SE**	**R^2**		**Slope**	**SE**		**MSE**	**MSE**	**MSE**
2.5	2.497	0.292	1.909	0.111	0.917		0.937	0.111		0.008	0.033	0.059
7.5	2.451	0.518	1.887	0.193	0.859		0.946	0.193		0.009	0.023	0.042
12.5	2.209	0.600	1.796	0.223	0.809		0.952	0.223		0.010	0.018	0.034
17.5	1.698	0.543	1.603	0.202	0.754		0.958	0.202		0.009	0.013	0.026
22.5	0.16	0.438	1.030	0.165	0.542		0.969	0.165		0.007	0.007	0.014
27.5	-0.969	0.281	0.617	0.105	0.248		0.981	0.105		0.007	0.008	0.010
32.5	-2.007	0.300	0.247	0.110	0.028		0.993	0.110		0.008	0.010	0.010
37.5	-3.894	0.309	-0.429	0.112	0.037		1.005	0.112		0.009	0.013	0.014
42.5	-6.25	2.433	-1.259	0.882	0.078		1.016	0.882		0.013	0.016	0.018
47.5	4.907	4.375	2.793	1.589	0.169		1.030	1.589		0.017	0.018	0.025
52.5	5.833	0.916	3.103	0.318	0.595		1.045	0.318		0.012	0.020	0.036
57.5	3.697	0.996	2.334	0.340	0.740		1.060	0.340		0.012	0.021	0.052
62.5	2.255	0.832	1.824	0.275	0.789		1.072	0.275		0.014	0.022	0.069
67.5	1.583	0.645	1.590	0.203	0.844		1.087	0.203		0.017	0.026	0.100
72.5	0.882	0.543	1.359	0.160	0.881		1.098	0.160		0.020	0.025	0.134
77.5	0.451	0.468	1.227	0.126	0.916		1.107	0.126		0.024	0.026	0.186
82.5	0.105	0.428	1.130	0.101	0.936		1.106	0.101		0.031	0.031	0.239
87.5	-0.049	0.431	1.095	0.087	0.955		1.105	0.087		0.038	0.038	0.337
92.5	-0.327	0.458	1.049	0.075	0.962		1.098	0.075		0.055	0.058	0.478
97.5	-0.972	0.689	0.975	0.086	0.941		1.087	0.086		0.123	0.149	0.699

Note: Five-year age classes are indicated with their central point, as in Figure 1. Eq. 3: full model; Eq. 4: restricted model (no intercept) and Eq. 5: twice-restricted model (no intercept and unitary slope).

Note that because of the non-stationary nature of the $c_{x,t}$ and the $k_{x,t}$ series, the estimated standard errors in Table 2 are likely biased. The R^2 values, on the other hand, may reflect a spurious correlation between the series, although this seems unlikely, because the estimated slopes are positive in 18 cases out of 20 (full model: eq. 3, Table 2). In other words, at almost all ages, the current and the reference proportions of individuals tend to move in the same direction, decreasing or increasing together. The only exceptions are two central age classes (35–39 and 40–44 years) where, however, variability is limited in both $c_{x,t}$ and $k_{x,t}$ (Figure 1), which also leads to very low R^2 values.

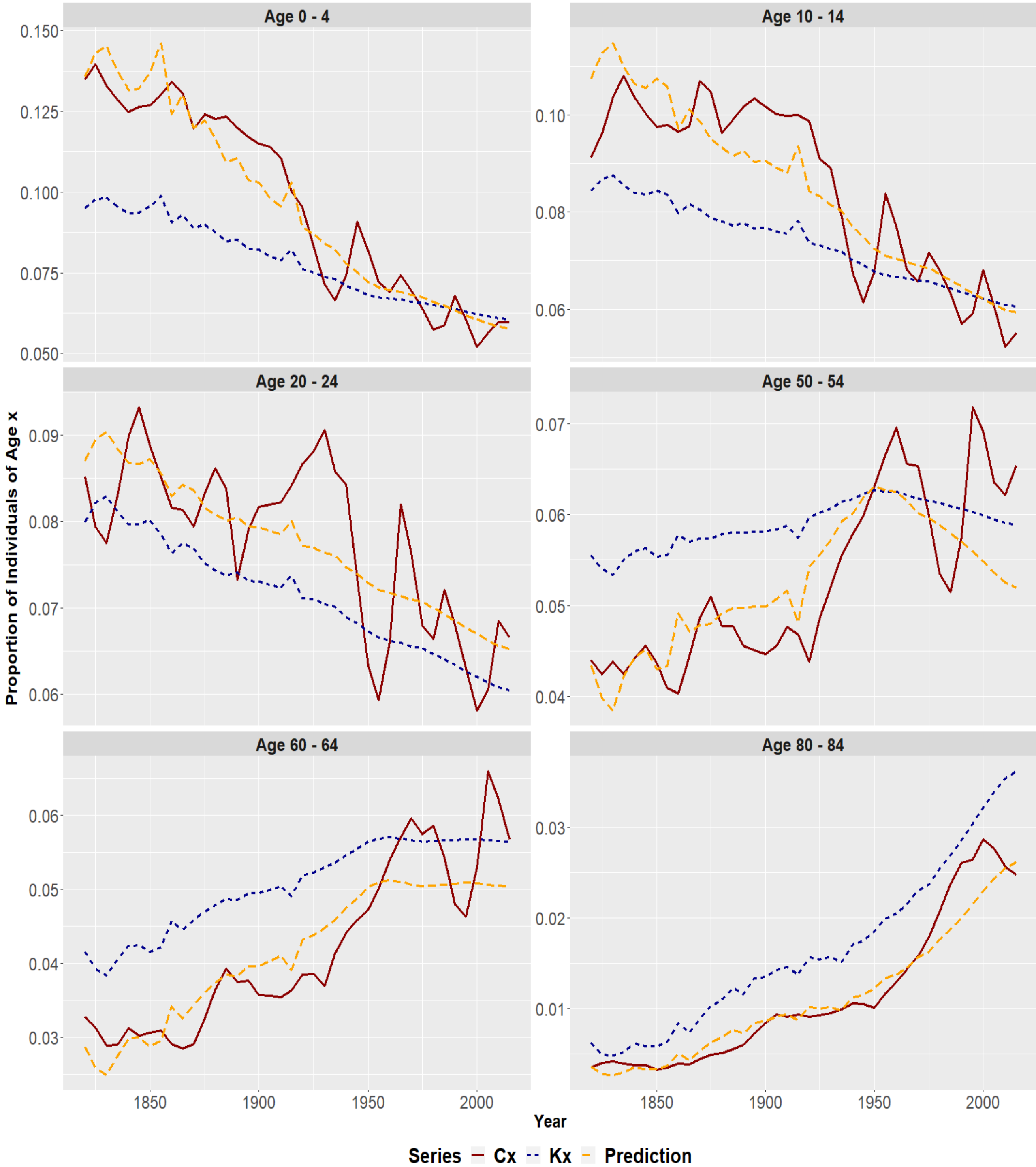
The estimates of Table 2 can be used to model, or "predict", the evolution of the current age structure over time. Figure 3 presents three predicted proportions of individuals by age: the blue dotted line represents the evolution of $K_{x,t}$ (reference age structure and eq. 5); the red, solid line represents the evolution of the current proportion of individuals of age x ($C_{x,t}$), and the orange-dashed line represents the prediction of this proportion ($\hat{C}_{x,t}$) based on the $K_{x,t}$ series and on the estimated age-specific parameters of Table 2 (eq. 3, full linear model).

Our models, based exclusively on the current life table and the associated stationary population, capture remarkably well the general evolution of the age structure in the past two hundred years in Sweden (and in the other countries of Table 1 – not shown here). Of course, our simplified models cannot accurately depict the fluctuations around this trend, i.e. the age-structure effects of a) previous mortality-affecting events such as wars and epidemics, and b) fertility and migration, levels and variations (totally ignored, here).

With the model proportions of Figure 3, we can also estimate the entire age structure of the Swedish population at different epochs (Figure 4). Both Figures, 3 and 4, show that the close correspondence between the current C and the predicted $\hat{C}$ values dates back to very long ago. In other words, the correlation between survival (period life tables) and the shape of the age structure (which includes population ageing) seems to predate not only the strong improvements in old-age mortality that materialized after the 1950s (Murphy 2017), but also the onset of the demographic transition, which started shortly after 1860, in Sweden.

Figures 3 and 4 also show that the twice-restricted version of eq. 3 (i.e., eq. 5 – no intercept, unitary slope) roughly works: the reference proportion of individuals aged x is generally close to its current counterpart ($K_{x,t} \sim C_{x,t}$). The approximation (as measured with the MSE for instance) is generally only slightly worse than in the full model (and it is even better than it, in recent years), and the interpretation is much simpler. As a first approximation, the age structure of the reference population in year t gives a good idea of the current age structure of that population in the same period, which is a way of saying that survival (more precisely: current survival) "explains" most of the current age structure of the population. At young ages (<30), the current proportion of individuals tends to be slightly higher than the reference, while at older ages (>55) the reverse is true. This is not surprising in a population with a four-fold increase in the period (from 2.5 million in 1820 to more than 10 million in 2018). Indeed, this small distortion disappears altogether in recent times (years 2000–2004; last panel of Fig. 4), when the effects of the demographic transition are over. Similar results emerge also for the other countries of our dataset (not shown here)

Fig. 3 *Current (red, solid line), reference (blue, dotted line) and predicted (orange, dashed line) proportion of individuals in selected age-groups (Sweden, 1820–2017)*



Note: predictions based on the full model (eq. 3). Please mind the different scales on the y-axis.

Fig. 4 *Current (red, solid line), reference (blue, dotted line) and predicted (orange, dashed line) proportion of individuals in selected periods (Sweden)*

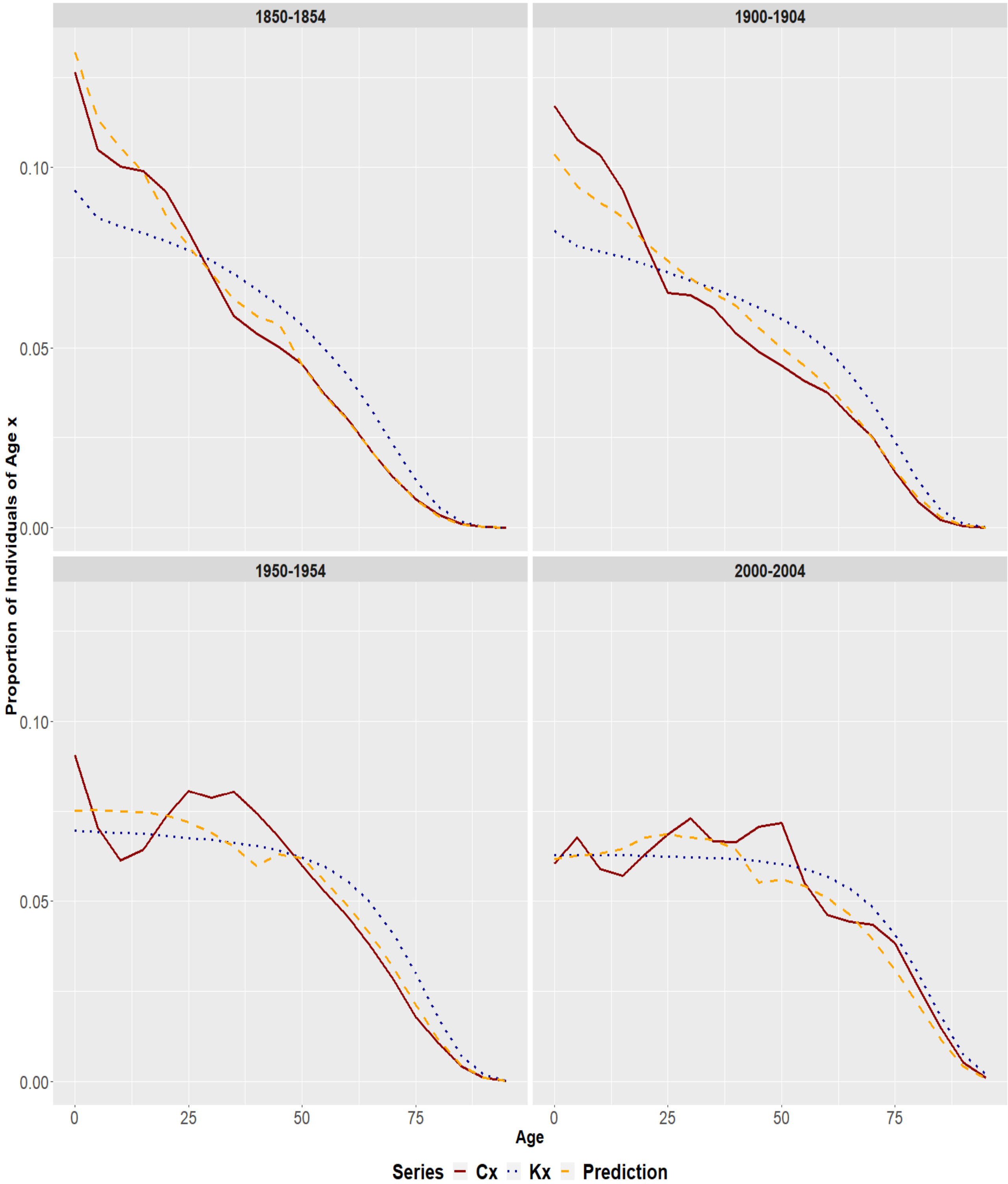
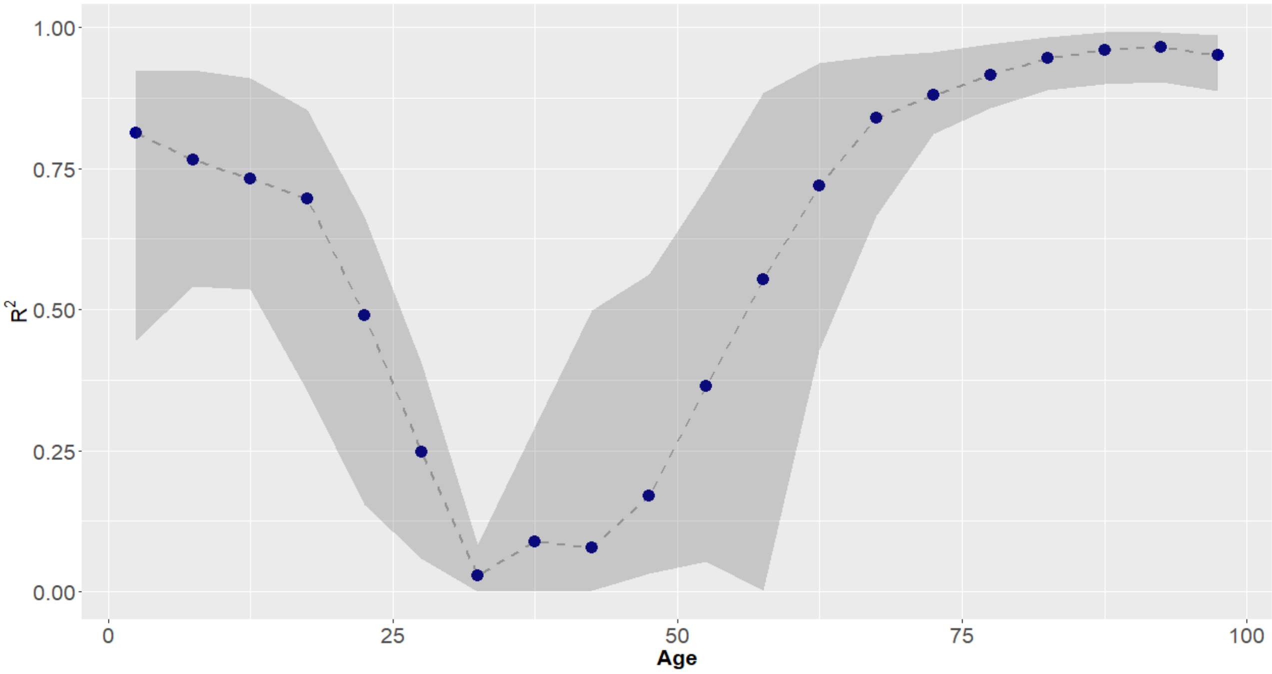


Fig. 5 *Median R^2 (explained variance) of Eq. 3 by five-year age groups (all the countries and years of Table 1)*



Note: The points represent the median explained variance of Eq. 3 by age group (all countries). The shaded area indicates the 95% band.

Figure 5 shows the median R^2 value associated with the estimate of eq. 3 at different ages in all the eight countries included in this analysis. Results are unsatisfactory (low R^2) between 25 and 54 years, but shortcoming is less serious than it may seem, because also the variability of both $k_{x,t}$ and $c_{x,t}$ is very limited at these ages. Instead, where the changes in the age structure of the population are more important, at young and old ages, the part of the variance that our model can "explain" becomes substantial, frequently above 75%, especially past age 60 years. All this is obtained totally ignoring fertility and migration in the current year, and all that happened to *any* demographic "behaviour" (fertility, migration and mortality) in *any* previous year.

5. The cointegration tests

In this section we present the results of the cointegration tests that we carried out considering separately each couple $c_{x,t}$ and $k_{x,t}$ for a fixed age x and varying t. The purpose of these tests is to exclude spurious correlations between the two series, which their stochastic trends make possible. We did this for each of the eight countries of Table 1, and for 20 five-year age groups x (0–4, 5–9, …, 95–99).

Table 3 summarizes the main results. Its first row says that the existence of a long-run relationship between the $c_{x,t}$ and the $k_{x,t}$ series is deemed likely in 107 cases (67% of the total), unlikely in 41 cases (26%) and uncertain in the remaining 12 cases (8%). If the tests are performed at a 5% significance level instead of 10%, a long-run relationship is deemed likely in 60% of the cases.

To assess the reliability of our tests, we performed three kinds of diagnostic on the residuals:

1) The Box-Pierce test of autocorrelation (Box and Pierce 1970);

2) The Shapiro-Wilk test of normality (Royston 1982) and;

3) The score test for non-constant error variance (heteroscedasticity; Cook and Weisberg 1983).

Based on these analyses, we identified a few "problematic" cases (with autocorrelation, non-normality or heteroscedasticity), which are not perfectly fit for this test. Removing them, however, which leaves us with 114 "well-behaved" time series, does not affect our results: in 66% of the cases, a long-run relationship can be identified (Table 3, second row). Overall, the tests results robustly indicate that in about two thirds of the cases, the $c_{x,t}$ are likely cointegrated with the $k_{x,t}$ series. Admittedly, there is still a relatively large proportion of "exceptions", which we impute to the noisiness of our time series (see below), especially in relation with the profound demographic shocks of the period. This notwithstanding, our results seem to us supportive of the existence of a true (not spurious) long-run relationship between the current and the reference age structure.

Our tests can be broken down by country, as we did in Figure 6a. In four of them (Australia, Canada, Sweden and the US) a long-run relationship emerges very frequently, in more than 75% of the 20 age groups considered. The remaining countries (England-Wales, France, New Zealand and Switzerland) do not perform as well, but even here a cointegration relationship appears to be likely in more than 50% of the series (age classes).

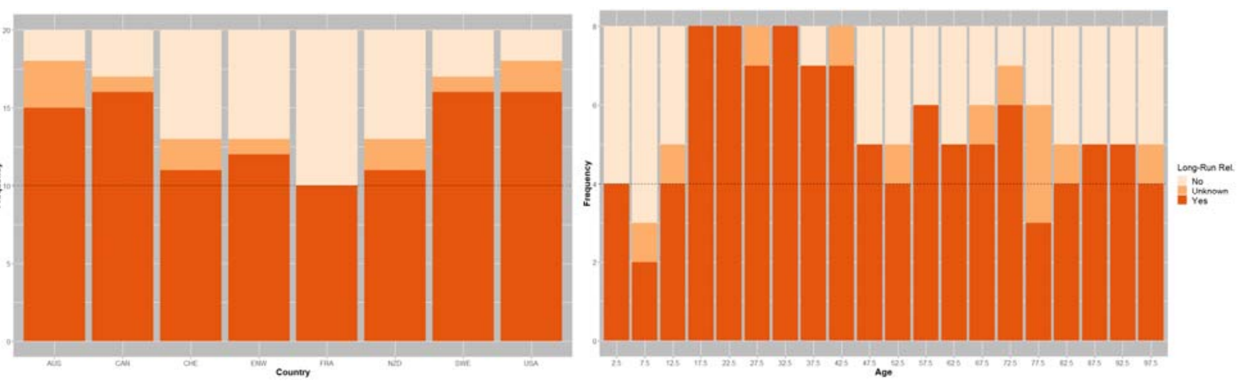
Table 3 *Summary of results on the relationship between $c_{a,t}$ and $k_{a,t}$ (all the countries and years of Table 1)*

Data	No. of tests	A long-run relationship between the $\boldsymbol{c_{a,t}}$ and $\boldsymbol{k_{a,t}}$ series... Likely exists	Likely does not exist	Unresolved	*% Existence*
Whole dataset	160	107	41	12	*66.8*
After diagnostics	114	75	29	10	*65.7*

Note: at 10% significance level.

Fig. 6 *Bounds test*

a. By country *b. By age group*



Note: The 20 (5-year age group) series of each of the eight countries are observed for an unequal number of years (see Table 1).
Source: see Table 1

Figure 6b breaks down our results by age group, which highlights the complexity of the pattern, but confirms that cointegration emerges in the majority – sometimes a large majority – of the cases, despite the several disturbing factors that we could not keep under control (e.g. data quality, epidemics, and the typically low statistical power of cointegration tests).

6. Conclusions and interpretation of results

In this paper, we presented an empirical finding: the evolution of the age structure of the population appears to be cointegrated with the evolution of its "reference" counterpart, i.e., the age structure of the stationary population associated with the period life table. This relationship holds in most cases (meaning, age groups) in the eight countries that we included in our analysis, those with sufficiently long time series in the Human Mortality Database. This means that most of the change observed in the proportions of young, adult and old people in these countries can be derived from the change in survival, and this for a very long time interval, dating back to as much as possible with the available data. The correlation does not seem to be spurious (i.e., due to common stochastic trends) and is rather strong, with the exception of the central ages, where, however, variability is very limited. A simplified version of this finding (which emerges when the intercept of the regression is forced to zero and the slope to one) is that the reference age structure $K_{x,t}$ represents (and has always represented, on average) an acceptable approximation of the current age structure $C_{x,t}$.

Murphy (2017) had already argued that, since mid-20th century, most of the evolution of the age structure in 11 European countries depends on the evolution of mortality. He could not go back in time for more than a century because of the data requirements of his technique, the PHE decomposition (Preston et al 1989). Conversely, we could: our method, less data demanding, allows us to use longer series, up to almost two centuries in the case of Sweden and France, for instance.

Coming to the implications of our findings, let us mention only two. Theoretically, demographers may want to reconsider the relative role traditionally attributed to fertility and mortality in shaping the population age structure: the latter should probably be re-evaluated.

Practically, our findings suggest that population forecasting exercises may be simplified, at least when they are instrumental to other activities, and notably the design of pension systems. Basing them exclusively, or at least prominently, on current and projected survival, without adding the extra uncertainty of fertility and migration forecasts, may simplify the task of policy makers, and possibly pave the way to better solutions of the pension problem.

Funding

This work was partly supported by the Italian MIUR through the JPI MYBL / CREW Project (Joint Programme Initiative: More Years Better Life, 2016 Call. CREW: Care, retirement and wellbeing of older people across different welfare regimes». MIUR Decree: n. 3266/2018; Official Bulletin no. 32 7. Feb 2019).

References

Bengtsson, T., & Scott, K. (2005). Why is the Swedish population ageing? What we can and cannot do about it. *Sociologisk Forskning,* 3, 3–12.

Bengtsson, T., & Scott, K (2010). The ageing population. In T. Bengtsson (Ed.), *Population Ageing –A Threat to the Welfare State* (pp. 7–22). Berlin: Springer Verlag.

Borgan, Ø., & Keilman, N. (2019). Do Japanese and Italian Women Live Longer than Women in Scandinavia?. *European Journal of Population*, 35, 87–99.

Bourgeois-Pichat, J. (1979). La transition démographique : Vieillissement de la Population, in Population Science in the Service or Mankind. Liege : International Union for the Scientific Study of Population, 211–239.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65, 1509–1526.

Box, G. E. P, Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis. Forecasting and Control*. Fifth Edition. New Jersey, John Wiley & Sons.

Caselli, G., & Vallin, J. (1990). Mortality and population ageing. *European Journal of Population,* 6, 1–25.

Chesnais, J.–C. (1990). Demographic transition patterns and their impact on the age structure. *Population and Development Review*, 16, 327–336.

Chesnais, J.–C. (1992). *The Demographic Transition. Stages, Patterns and Economic Implications*. Oxford, Oxford University Press.

Coale, A. J. (1956). The effects of changes in mortality and fertility on age composition. *Milbank Memorial Fund Quarterly*, 34, 79–114.

Coale, A. J. (1957). How the age distribution of a human population is determined. *Cold Spring Harbor Symposia on Quantitative Biology*, 22.

Coale, A. J. (1972). *The Growth and Structure of Human Populations. A Mathematical Investigation*. Princeton University Press.

Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70, 1–10.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica*, 55, 251–276.

Giles, D. (2013) Econometrics Beat: Dave Giles' Blog. ARDL Models – Part II – Bounds Tests (https://davegiles.blogspot.com/2013/06/ardl-models-part-ii-bounds-tests.html)

Granger, C. E. J., & Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2, 111–120.

Johnston, J., & DiNardo, J. (2007). *Econometric Methods*. *Fourth edition*. McGraw–Hill.

Keyftz, N. (1975). How do we know the facts of demography?. *Population and Development Review*, 1, 267–288.

Lee, R., & Zhou, Y. (2017). Does fertility or mortality drive contemporary population aging? The revisionist view revisited. *Population and Development Review*, 43, 285–301.

Murphy, M. (2017). Demographic Determinants of Population Aging in Europe since 1850. *Population and Development Review*, 43, 257–283.

Murphy, M. (2021). Use of Counterfactual Population Projections for Assessing the Demographic Determinants of Population Ageing. *European Journal of Population*, 37(1), 211–242.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics*, 16, 289–326.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford, Oxford University Press.

Preston, S. H., Himes, C., & Eggers, M. (1989). Demographic conditions responsible for population aging. *Demography*, 26, 691–704.

Preston, S. H., & Stokes, A. (2012). Sources of population aging in more and less developed countries. *Population and Development Review*, 38, 221–236.

Royston, P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics*, 31, 115–124.

Vaupel, J. W. (2010). Biodemography of Human Ageing. *Nature*, 464, 536–542.

Appendix A – The bounds test

The general idea behind this approach is to test the existence of a possible tendency of a series to revert towards its long-run equilibrium path by estimating the following error correction model (ECM) for a specific and constant age x, and for varying t:

$$\Delta c_{x,t} = \alpha_{x,0} + \delta_{x,1} c_{x,t-1} + \delta_{x,2} k_{x,t-1} + \sum_{i=1}^{p} \alpha_{x,i} \Delta c_{x,t-i} + \sum_{i=0}^{q} \beta_{x,i} \Delta k_{x,t-i} + \varepsilon_{x,t}, \qquad (A.1)$$

where α_i, β_i and δ_i are the model parameters, $\Delta c_{a,t} = c_{a,t} - c_{a,t-1}$, $\Delta k_{a,t} = k_{a,t} - k_{a,t-1}$, and p and q are lags, to be discussed shortly.

Eq. A.1 presents several advantages. First, it can be estimated with ordinary least squares (OLS).

Second, the lags p and q do not need to be predetermined: a statistical procedure, based on the Bayesian information criterion (BIC) will suggest the best combination of the two. To determine the best p and q (lags) of our ECM, we started from a 3·4 grid search. In practice, for each country in our analysis we estimated the model 12 times, with different combinations of p = 1, …, 3 and q=0, …, 3. In this case, and deviating from what we did in the rest of our analyses, we decided to work with quinquennial data: therefore, we are considering lags of up to 5·3=15 years. In most cases, the best model (with the lowest BIC value) has p=1 (77% of the cases) and q=0 (74% of the cases).

Third, the parameters of eq. 3 in the main text can be derived from those of eq. A.1, because:

$$m_x = \frac{\alpha_{x,0}}{-\delta_{x,1}} \text{ and } \gamma_x = \frac{\delta_{x,2}}{-\delta_{x,1}}. \qquad (A.2)$$

Fourth, the existence of a cointegration relationship between $c_{x,t}$ and $k_{x,t}$ can be tested with the F statistics on the null hypothesis H_0: $\delta_{x,1} = \delta_{x,2} = 0$. The distribution of this statistic is non-standard, but its critical values, calculated with Monte Carlo simulations, are tabulated in Pesaran et al. (2001, Table CI(iii)). These critical values are generally greater than those employed in the standard *F* test, which makes the rejection of the null hypothesis (of no cointegration) more difficult.