A Cointegration Analysis of Population Aging

By Gustavo De Santis and Giambattista Salinari

Abstract

While “classical” demography imputes population ageing to low fertility, a recent “revisionist” line of thinking signals the emergence of an ageing “from the top” (i.e., due to low mortality), starting slightly after World War II. We join this debate with two contributions. On the one hand, we try to assess and put in perspective the way counterfactual analysis has generally been used in this domain. On the other, we show that, in the long run, mortality impacts on the population age structure, and on ageing, more that it is customarily believed. With data taken from the Human Mortality Database (HMD) on 13 populations located in Europe, North America and Oceania, we show that a cointegration relationship exists between the actual age structure in year *t* and what we call the reference age structure, that is the age structure of the stationary population associated with the period life table of year *t*. This means that most of the change observed in the proportion of young, adult and old people in these countries can be derived solely from the change in survival, ignoring fertility and migration, and this for a very long time interval, dating back to as much as the data allows, up to two centuries.

Keywords

Age structure, ageing, stationary population, cointegration.

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). However, even leaving this extreme case aside, Coale’s (1957) famous counterfactual analysis showed that, had fertility remained constant, the age structure of Sweden would have been practically the same in 1860 as in 1950. Bengtsson and Scott (2005, 2010) updated the exercise to the period 1900 to 2000, and confirmed that, with unchanged fertility, the proportion of people aged 65 and over in Sweden would have been almost the same. In both cases, contrary to intuition, huge changes in mortality proved of very small consequence on the population age structure.

Recently, things have slightly changed: improvements in survival are concentrated at older ages (Vaupel 2010), and a sort of “ageing from the top”, due to more and more people surviving to old and very old ages, has emerged (Preston et al. 1989; Caselli and Vallin 1990; Preston and Stokes 2012). Some “revisionist” scholars, as Lee and Zhou (2017) call them, have therefore challenged the idea that fertility is always the major driver of population ageing. Murphy (2017), for instance, shows that mortality has been the most important determinant of population ageing in 11 European countries in the past 65 years or so. However, the idea of an ageing process “from the top” remains controversial: Lee and Zhou (2017), for instance, on the basis of their own counterfactual analysis, contend that even in “modern” populations, low fertility remains the main cause of population ageing.

We join this debate with two main contributions. On the one hand, we try to assess and put in perspective the way counterfactual analysis has been used in this domain, suggesting a possible alternative. On the other, we show that, in the long run, mortality impacts on the population age structure, and on ageing, more that it has customarily been acknowledged until now.

As for the first point, counterfactual analysis is a more delicate instrument than it is commonly believed to be. Scholars nowadays prefer to place counterfactuals within the larger frame of causal analysis, which means that they must be based on a "functional causal model" and inserted into a system of (structural) equations describing the relationships between the variables of interest (Pearl et al. 2016). "Processing counterfactuals is the hardest task [in causal analysis] because it requires some information about the functional relationships and/or the distribution of the omitted factors" (Pearl 2009:38).

A closer inspection of the counterfactual analyses usually conducted in this field of research, starting from the first (Coale 1957), reveals that an essential (if only tacitly introduced) assumption is that the main components of population dynamics – mortality, fertility and migration – are independent of each other. Indeed, there is no formal connection between them, and, in simulations, each of them is allowed to vary freely, without affecting the other components, which, typically, are assumed to remain constant, or to follow the path they actually followed (but in a totally different demographic context).

However, populations where mortality, fertility and migration are independent of each other do not seem to have ever existed. Davis (1963), for instance, conjectured that fertility decline was caused by mortality decline and that a positive long-run relationship between mortality and fertility characterized the entire demographic transition (see also Galloway, Lee and Hammel 1998; Billari and Dalla Zuanna 2013). If this assumption holds, mortality can influence the age structure of a population in two different ways: directly, because a change in mortality translates into a change in the age structure, everything else equal; and indirectly, because a change in mortality causes a change in fertility (and migration), and this, in turn, affects the age structure.

We will not be able to distinguish between the two paths, but we will show that mortality is, in the long run, a good predictor of population ageing (and, more generally, of the age structure of a population). To do this we will use period life tables, which, as we will show shortly, prove useful in this type of application, even if some scholars question their validity as indicators of the "true" evolution of survival, and prefer cohort life tables (e.g., Borgan and Keilman 2019).

Our analysis has, admittedly, a few technically difficult passages (due to the notion of cointegration: see section 2), but it essence is trivial: we will try to prove that the *reference age structure* (in our terminology, the proportion of individuals of age x in the stationary population associated with the period life table of year t) "attracts" the current (actual) age structure of a population, and that this small, but persistent force of attraction eventually prevails, and shapes the age structures as we see them. In practice, we will try to prove that, knowing the *reference age structure,* one can predict the evolution of the actual (current or future) age structure of a population.

Note that three outcomes of our analysis are possible. The first is that our predicting capacity proves limited. In this case, the overall effect of mortality (or, at least, of period mortality) on the age structure is arguably small and the classical result of demographic analysis holds: fertility decline, not mortality, is the main driver of population ageing. The second possibility is that our prediction becomes good only after a certain, relatively recent, date, e.g. 1950: this will lend support to the "revisionist approach", that is to the hypothesis of an ageing process "from the top" that emerged only recently, but that did not operate before. Finally, the third possibility is that 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 (13, see below) and for the entire period of observation (the last two centuries or so).

We will try to convince our readers that our results point in this direction (period life tables "explain" a large part of observable population age structures) and argue that this finding may call for a reassessment of the relative importance of mortality among the drivers of population aging.

The paper is organized in four remaining sections. The next section will be devoted to the formal description of the test that we will use in the paper. The third section will present the data and some descriptive statistics. The fourth session will present the results. The last session will be devoted to a discussion of our findings.

2. Testing Cointegration between the Actual and the Reference Age-Structure

In the following, we will indicate with $C_{x,t}$ the actual age structure, i.e., 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 the total population in year *t*, respectively.

Similarly, the *reference age structure* of the population (i.e., the age structure of the current stationary population) will 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*. Let $k_{x,t}$ and $c_{x,t}$ denote the log transformation of $K_{x,t}$ and $C_{x,t}$, respectively. The log-transformation has two main purposes: it circumvents one of the limitations of proportions (they are bounded in the 0-1 interval, with the lower limit, 0, particularly disturbing), and it helps to better approximate linearity in the relationship between the two series, $c_{x,t}$ and $k_{x,t}$.[1]

The main goal of this paper is to test the existence of a cointegration relationship between $c_{x,t}$ and $k_{x,t}$ (see Section 2.1 for the details). To do this, two different strategies can be followed. The first, a more classic approach, is to focus on a specific couple of series ($c_{a,t}$ and $k_{a,t}$, for a specific age *a* and various years *t*), and test whether these two are cointegrated. The second strategy is to pool together all the series $c_{x,t}$ and $k_{x,t}$, at all available ages *x* and years *t*, and perform what is known as a "panel cointegration test". As both lines of research have their pros and cons, we decided to try them both. The essential elements to understand these two tests are described in the next two sub-sections.[2]

2.1 Time series approach to cointegration

If we focus on a single couple of $c_{a,t}$ and $k_{a,t}$ (for a specific and constant age *a*, and for several years *t*), the question is whether a long-run (cointegration) relationship exists between them, of the type:

$$c_{a,t} = m + \gamma k_{a,t} + \varepsilon_t, \tag{3}$$

where m and γ are the model coefficients and ε_t is the error term. Unfortunately, in this case standard regression methods cannot be used, because of the possibility of spurious correlations (Granger and

---

[1] With other types of transformation (e.g., Box-Cox's lambda; Wei 2006:83), or with no transformation at all, the results (not shown here) do not differ in any substantial way from those presented in this paper, but the relationship between the two series $c_{x,t}$ and $k_{x,t}$ becomes slightly worse ("less linear"), and the "noise" stronger.

[2] The next two sub-sections are based on three textbooks on econometrics modelling: Box, Jenkins, Reinsel and Ljung (2016); Jonston and DiNardo (1996); Pesaran (2015).

Newbold 1974). Indeed, if both $c_{a,t}$ and $k_{a,t}$ are non-stationary (their means, variances or autocovariances depend on time), as they actually are in the populations under scrutiny (see Section 3), even if the two series are independent of each other they may seem to move together (spurious correlation). For this reason, in classical cointegration analysis the test for cointegration is based on the residuals of Eq. 3, and not on its coefficients (Engle and Granger's 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 with the Augmented Dickey-Fuller (ADF) unit root test, although other solutions are also possible. To understand the rationale of this approach, let us define $\hat{c}_{a,t} = m + \gamma k_{a,t}$ as the log-proportion of individuals of age a in year t predicted by Eq. 3 on the basis of the reference age structure (better: of the reference log-proportion of individuals of age a in year t). If the actual and the reference age structure are cointegrated, the actual log-proportion of individuals aged a, $c_{a,t}$, will show a tendency to "revert" to its long-run equilibrium value $\hat{c}_{a,t}$. In practice, cointegration means that $c_{a,t}$ and $\hat{c}_{a,t}$ cannot be too far away from each other, because some "force" pushes $c_{a,t}$ towards $\hat{c}_{a,t}$. The residual-based cointegration test proposed by Engle and Granger aims to detect this (possible) "force". Unfortunately, the test used to check whether time series are, or are not, stationary typically suffers from a low statistical power, and may prove inconclusive. Therefore, other approaches have been proposed, and in the following we will use one of them, the so-called "bounds test" (Pesaran et al. 2001). The general idea behind this approach is to test the "reversion towards the long-run equilibrium" (i.e., the "force" we mentioned earlier) by estimating the following (so-called "conditional") error correction model (ECM):

$$\Delta c_{a,t} = \alpha_0 + \delta_1 c_{a,t-1} + \delta_2 k_{a,t-1} + \sum_{i=1}^{p} \alpha_i \Delta c_{a,t-i} + \sum_{i=0}^{q} \beta_i \Delta k_{a,t-i} + \varepsilon_t, \tag{4}$$

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. 4 presents several advantages. First, this model can be estimated with ordinary least squares (OLS).[3] Second, the lags p and q do not need to be predetermined: a statistical procedure, based on the BIC (Bayesian information criterion) will suggest the best combination of the two. Third, the parameters of Eq. 3 can be derived from those of Eq. 4, because:

$$m = \frac{\alpha_0}{-\delta_1} \text{ and } \gamma = \frac{\delta_2}{-\delta_1}. \tag{5}$$

Fourth, the existence of a cointegration relationship between $c_{a,t}$ and $k_{a,t}$ can be tested with the F statistics on the null hypothesis $H_0\colon \delta_1 = \delta_2 = 0$. 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. For any given significance level, Pesaran et al. (2015) propose two critical values, F_U (upper) and F_L (lower – this explains the name "bounds test"), which 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

3) $F_L < F < F_U$ leads to a "suspension verdict": the inference is inconclusive (Pesaran 2015:526).

---

[3] To estimate Eq. 4 we used the *dynlm* R package.

If $c_{a,t}$ and $k_{a,t}$ are cointegrated, one can estimate Eq. 3 with OLS and use the $k_{a,t}$ series and the estimated coefficients to predict the evolution of $c_{a,t}$. The proportion of the overall $c_{a,t}$ variance explained by the $k_{a,t}$ series is gauged, as usual, by the R^2 statistics of the regression model.

Although higher than in the two-step procedure of Engle and Granger (1987), the statistical power of the bounds test remains relatively low, especially if the series are "disturbed", e.g. by wars or epidemics. This is why we resorted also to the "panel cointegration test", rapidly described in the next sub-section.

2.2 Testing Cointegration in Panel Data

As both the current and the reference age structure can vary across age and time, our dataset presents a panel structure. In panel terminology, ages $x \in \{0, 1, \ldots, X\}$ represent the "individuals", whereas calendar years $t \in \{1, 2, \ldots, T\}$ represent the time dimension. Our dataset can therefore be thought of as a set of $2X$ time series ($c_{x,t}$ and $k_{x,t}$), of length T, pooled together. To test for cointegration in this case, one needs to estimate the model:

$$c_{x,t} = m_x + \gamma_x k_{x,t} + \varepsilon_{x,t} \ \forall x. \tag{6}$$

As in Eq. 3, if the series $c_{x,t}$ and $k_{x,t}$ are non-stationary, a spurious correlation between the two can bias the estimate, and, once again, an analysis of the model residuals is necessary. If the residuals turn out to be stationary, the hypothesis of a "true" (non-spurious) linear association between the two series is probably tenable.

To ascertain this, we adopted a two-step procedure. First, we ran the IPS (Im, Pesaran and Shin 2003) panel unit root test on the $c_{x,t}$ and the $k_{x,t}$ series to verify their non-stationarity,[4] an essential precondition for cointegration to exist. Next, we estimated model (6) with OLS and, on the residuals of this model, we ran another IPS test.[5] If the actual (c) and the reference (k) age structures are cointegrated, these residuals must be stationary.

The IPS test is a so-called "first generation" panel unit root test (Pesaran 2015:817), which relies on the critical assumption that the time series within the panel ($c_{x,t}$ or $k_{x,t}$) are cross-sectionally independently distributed. For this reason, we also checked for possible cross-sectional disturbances by performing a so-called "second generation" panel cointegration test. [6]

3. The Data

For our analysis, we used data taken from the Human Mortality Database[7] (HMD) on 13 populations located in Europe, North America and Oceania (Table 1). We used data on exposures to calculate the actual 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).

---

[4] If the series are stationary, virtually all of the problems discussed in this methodological section disappear, and the estimation procedure can be done very simply, for instance with OLS. This, however, is not our case (see Section 2).

[5] We estimated the model with the *pvcm* function, and we ran the unit root test with the *purtest* function, of the *plm* R package.

[6] We first estimated model (6) with the Common Correlated Effects Mean Group Estimator (CCEMG) and then, on the residuals of this model, we computed the CIPS test (Cross-sectionally augmented IPS, Pesaran 2007), following the example given by Pesaran (2015 p. 844). Our software, also in this case, is the *plm* R package, with its *pmg* and *cipstest* functions.

[7] Last accessed in June 2019.

The countries covered by our analysis have time series of varying lengths. Typically, European populations have series that date back to the 19th century, or even before (Swedish data, for instance, date back to the 18th century). However, we decided to start in 1820, or as early as possible after that, because older data are generally considered less 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 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 box plots of Fig. 1 show that the relative weights of the youngest and the oldest age groups in Sweden[8] changed considerably in the last two centuries (in both the actual and the reference population), while the relative weight of the central age groups varied very little. The limited variability of our "independent" variable $k_{x,t}$ reduces the explanatory power of the ECM (Eq. 4). In practice, however, this is less problematic that it seems, because in these central ages also the dependent variable $c_{x,t}$ displays very limited variability (Fig. 1b).

One of the difficulties that our analysis must face is represented by mortality crises: e.g., the cholera epidemics of the 19th century, the Spanish flue epidemics of 1918 and the two world wars of the 20th century. These sudden discontinuities may introduce several forms of distortion. The most problematic are probably those linked to the ups and downs of fertility, which induce "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 6), but which may not exist in practice. However, this does not seem to be the case, as Fig. 2a shows for all the countries, age classes and years of our dataset. Actually, a closer look at Fig. 2a reveals a small, but significant slope change in the passage to older ages (above 60 years; see also Figures 2b and 2c), and, besides, more refined statistical analyses (not shown here) indicate that there is a slight (but statistically significant) convexity above 60 years, and a concavity below this age. All in all, however, the diagnostics of our ECM model (Eq. 4) reveals that these modest departures from linearity do not influence the "general picture" that emerges from our analysis.

---

[8] In this paper, we will systematically use Sweden as an example, which we preferred over others because it has the longest time series of data. The other countries or regions, not shown here for reasons of space, behave similarly.

Table 1 *Summary statistics*

					Mean age of populations			
		Time range of series (years)			**Actual**		**Reference**	
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
Denmark	DNK	1835	2016	182	27.1	40.7	31.5	41.1
England&Wales	ENW	1841	2016	176	25.5	40.0	31.7	41.3
Finland	FIN	1878	2015	138	26.2	41.7	31.8	41.4
France	FRA	1820	2016	197	28.3	40.8	31.4	42.0
Netherland	NLD	1850	2016	167	27.0	41.0	31.8	41.4
New Zealand	NZD	1901	2008	108	25.7	38.2	34.8	41.3
Norway	NOR	1846	2014	169	27.1	39.0	33.6	41.6
Scotland	SCO	1855	2016	162	25.9	41.2	31.8	40.4
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; SCO=GBR_SCO

Source: Human Mortality database (last accessed in June 2019)

Fig. 1 Variability of k_x *[ln(K_x)] and* c_x *[ln(C_x)]* by five-year age classes in Sweden (1820-2017)

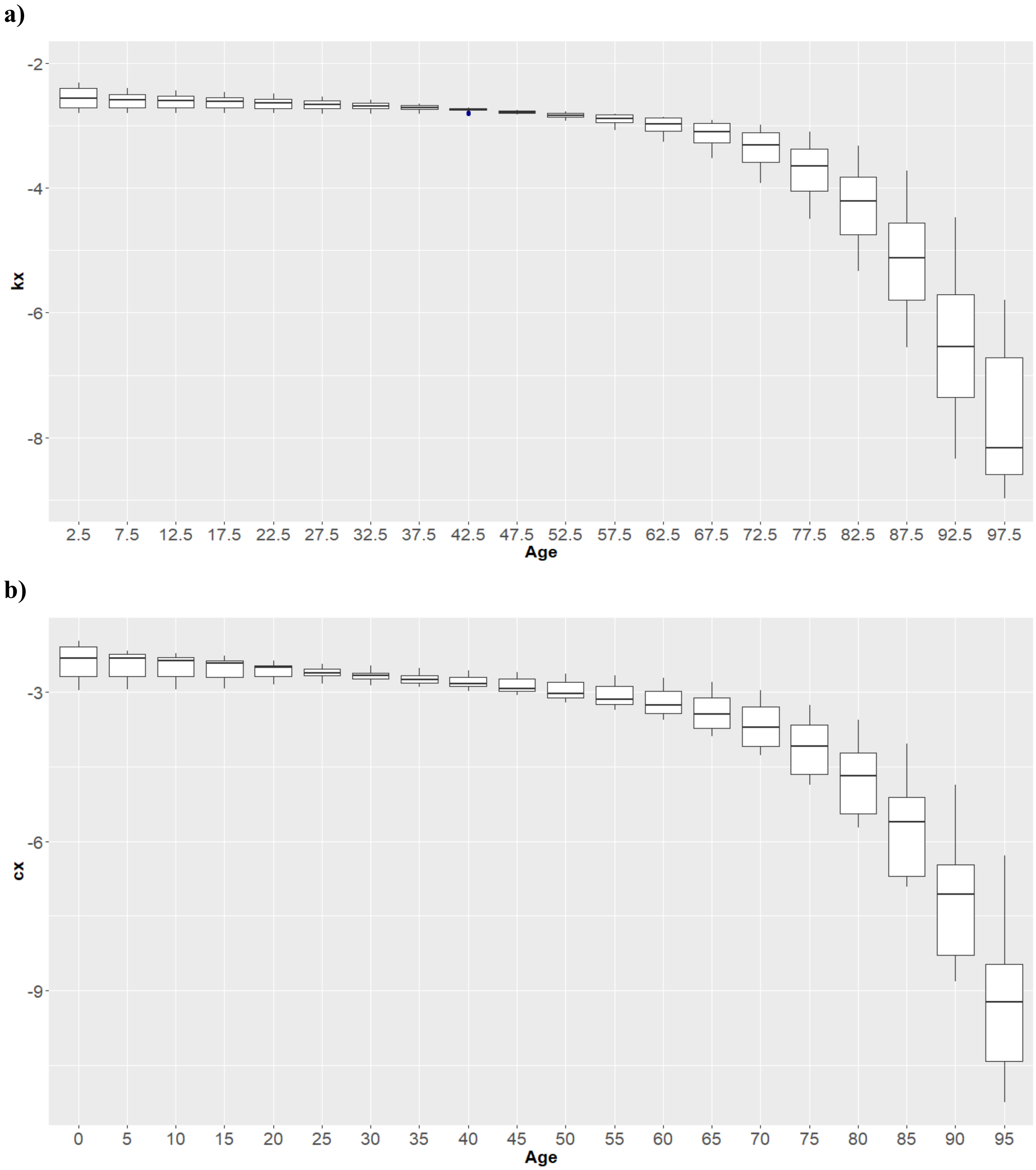
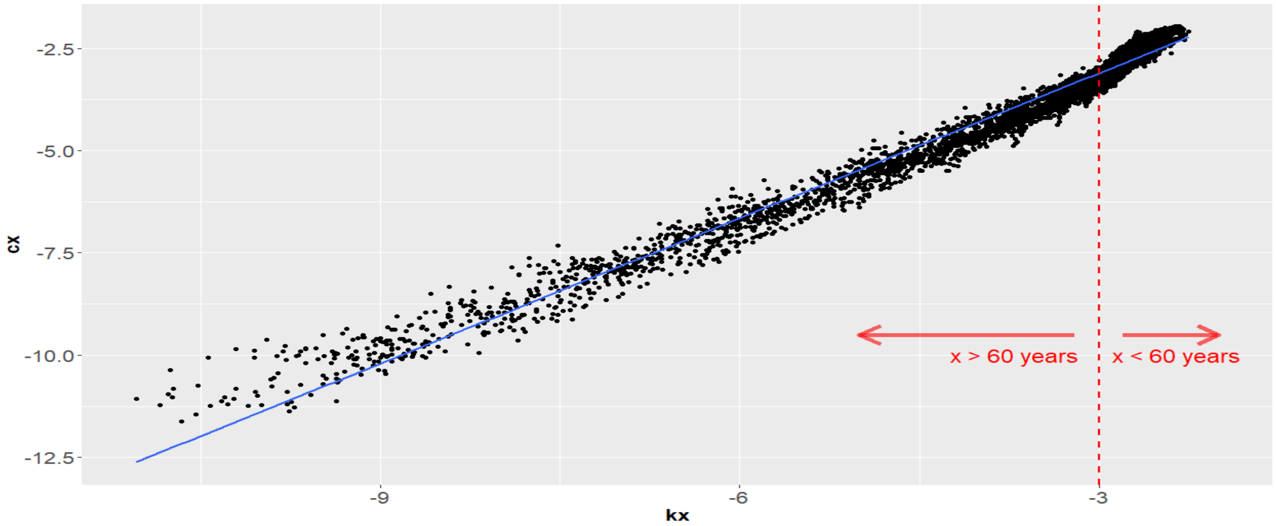
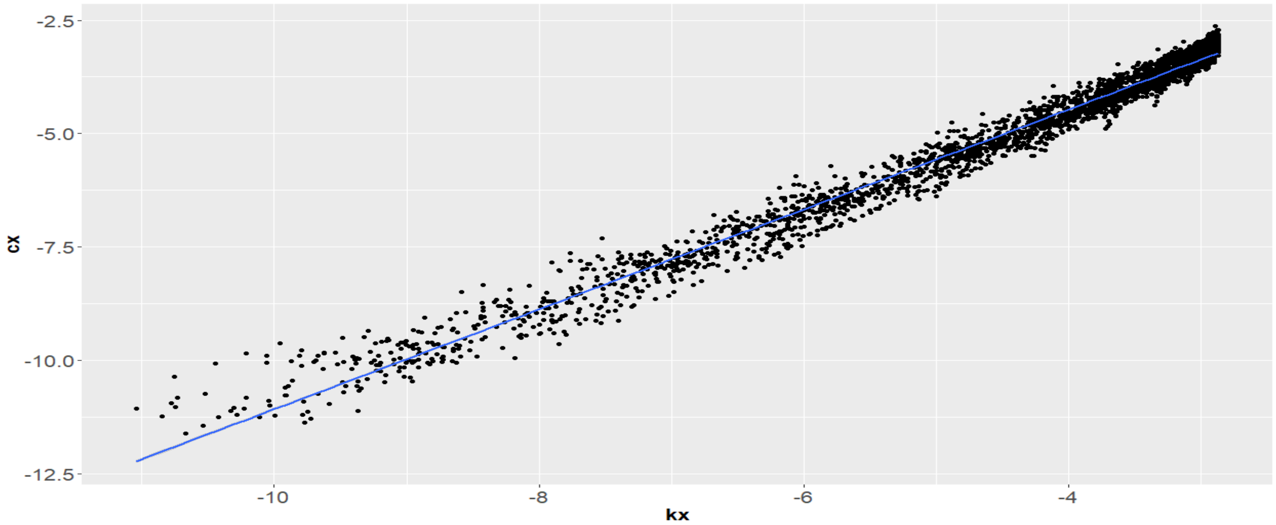


Fig. 2 *Relationship between k_x [ln(K_x)] and c_x [ln(C_x)] (all the countries and years of Table 1)*

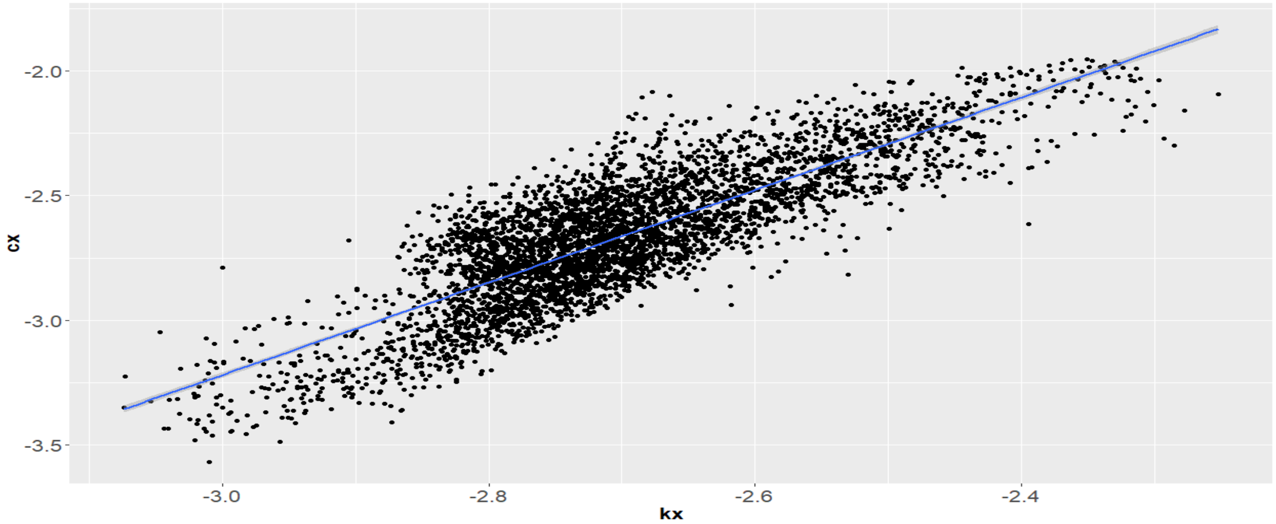
a) All ages



b) Ages 60 and over



c) Ages below 60



Source: See Table 1.

4. Results

Let us first present the results of the cointegration tests performed with the "time series" approach of section 2.1, i.e. considering separately each couple $c_{a,t}$ and $k_{a,t}$. We did this for each of the thirteen countries of Table 1, and for 20 quinquennial age groups x (0-4, 5-9, …, 95-99). To determine the best p and q (lags) of our ECM (Eq. 4) we started from a 3x4 grid search. In practice, we estimated each model 12 times, with different combinations of p = 1, …, 3 and q=0, …, 3 (with quinquennial data, i.e., we are considering lags of up to 5x3=15 years). In most cases, the best model (with the lowest BIC value) has p=1 (82% of the cases) and q=0 (78% of the cases). On this, we ran the Pesaran et al. (2001) bounds test with a significance level of 0.1. To sum things up, we ran 13x20 = 260 tests based on the estimation of 260x12 = 3640 ECMs. The main results are summarized in Table 2.

Let us first consider its first row: the existence of a long-run relationship between the $c_{a,t}$, and the $k_{a,t}$ series (a=0, 1, ..., 19) is deemed likely in 154 cases (60% of the total), unlikely in 89 cases (34%) and uncertain in the remaining 17 cases (6%).

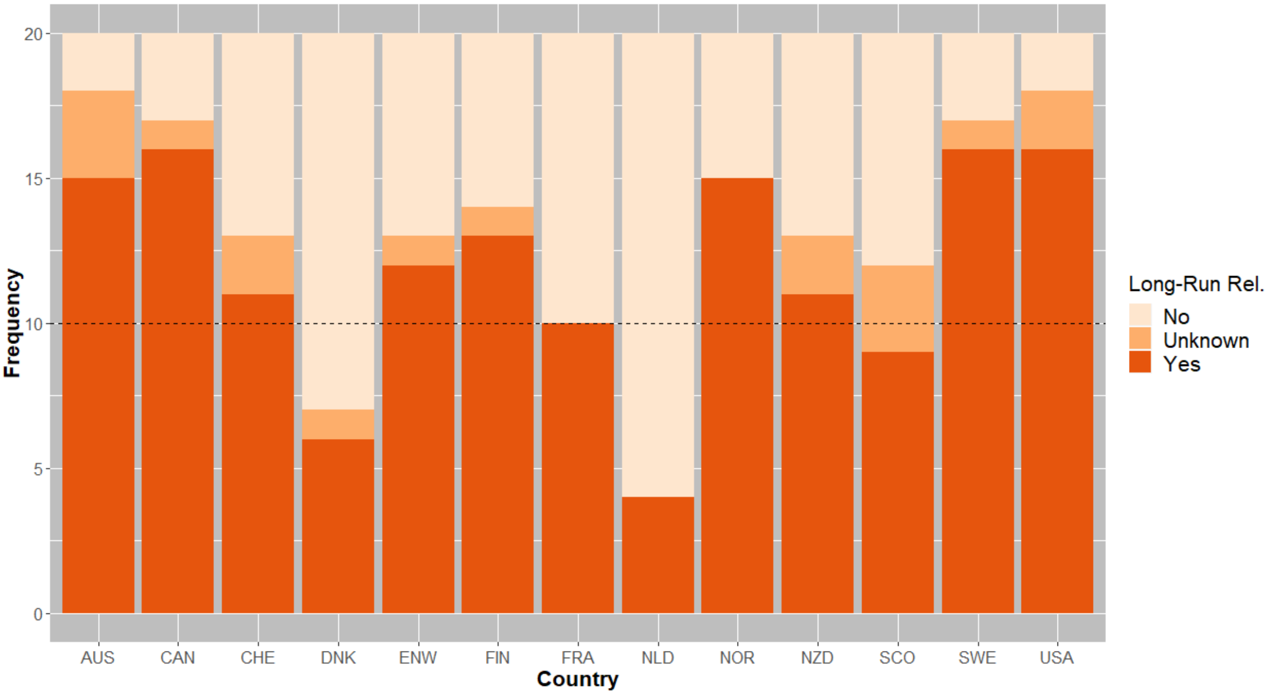
Breaking down the outcome of our tests by country (Fig. 3a), we find that in five of them (Australia, Canada, Norway, Sweden and the US) a long-run relationship emerges very frequently, in more than 75% of the age groups considered. At the opposite extreme, two countries (Netherlands and Denmark) perform very poorly, with a long-run relationship that can be identified in less than 30% of the age groups. Four of the five best performing countries (Australia, Canada, Sweden and the US) were comparatively less affected by the two world wars: this suggests that sudden mortality (and fertility) crises, with their long lasting "wave" effects on the age structure, may negatively affect the relationship between the two series, $c_{a,t}$ and $k_{a,t}$.

Table 2 *Summary of results on the relationship between* $c_{a,t}$ *and* $k_{a,t}$ *(all the countries and years of Table 1)*

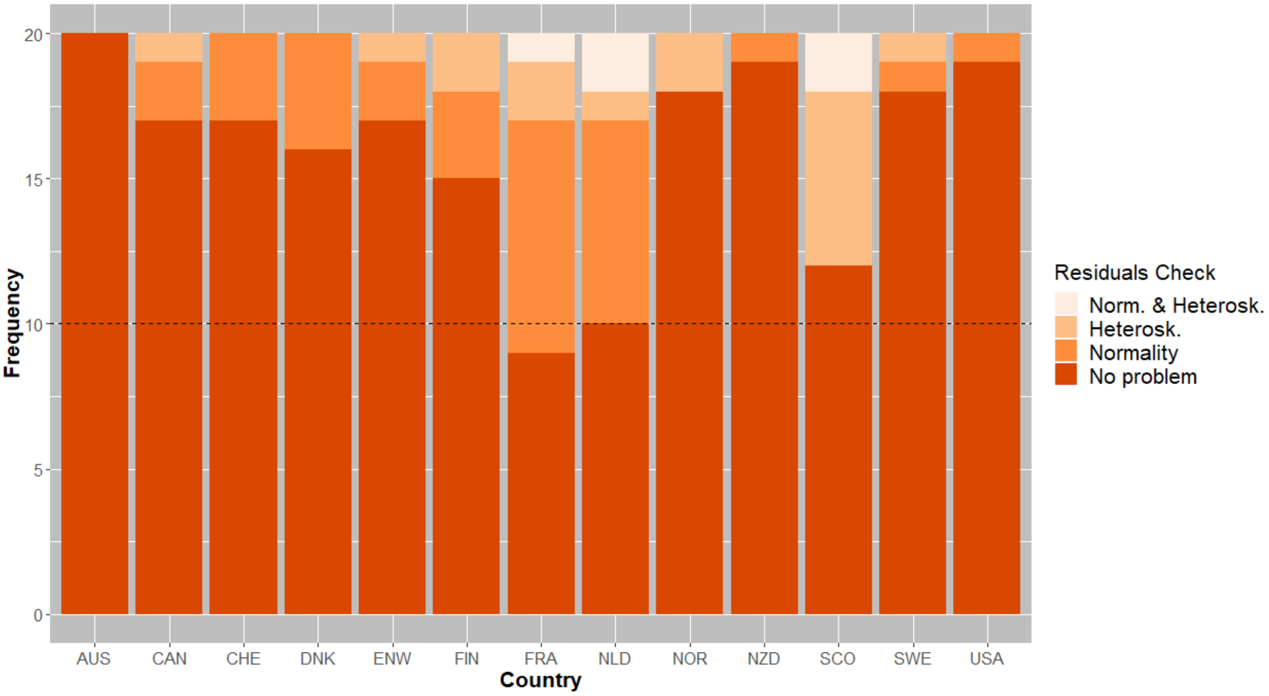
		A long-run relationship between $c_{a,t}$ and $k_{a,t}$series			
Data	**N tests**	Likely exists	Likely does not exist	Unresolved	*% Existence*
Whole dataset	260	154	89	17	*59.2*
After diagnostics	175	109	54	12	*62.3*
After diagnostics and excluding age groups 45-59 years	146	94	41	11	*64.4*

Fig. 3 *Results by country*

a. Bounds test



b. Model diagnostic



Note: there are 20 (5-year age group) series for each country, where, however, they are observed for a different number of years (see Table 1).

Source: see Table 1

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

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

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

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

Luckily, problem number 1 (correlation between the residuals) affects only seven of our estimates (out of 260). Therefore, in Fig. 3b we ignored it and focused only on the remaining problems: non-normality or eteroskedasticity, or both.

Note that the five "best" countries (where the existence of a cointegration relationship appears to be more frequent) are also those where the models diagnostic performs better. Three regions instead – France, Netherland and Scotland – prove to be particularly problematic, both in terms of non-normality and eteroskedasticity.

If we remove the "problematic" cases from our dataset (those with autocorrelation, non-normality or eteroskedasticity), we are left with 175 "well-behaving" time series: on this subset, a long-run relationship can be identified in 62.3 % of the cases (Table 2, second row).

Let us now break down our results by age-group. What emerges is a complex pattern (Fig. 4a). In the age range between 15 and 44 years a cointegration relationship can be found in most of the series scrutinized. Furthermore, a cointegration relationship can be found in more than 50% of the cases between 55 and 74 years, and over 85 years. Conversely, in the remaining age classes (0-14, 45-54, and 75-89) the proportion of cointegrated series generally drops below 50%.

The central age groups, 45-49, 50-54 and 55-59 present a comparatively smaller variability of the $k_{x,t}$ series (see Fig. 1) which translates into a smaller explanatory power of the ECM at these ages. Indeed, the median R^2 of these models[9] drops to about 30% for the central age groups (Fig. 4b), while it is (well) above 40% for the others. If we exclude them from our analysis (Table 2, row 3), the proportion of time series where a long-run relationship can be detected rises to 64.4%.

As for the age classes 75-89, the low proportion of "successful" cointegration may depend on the high noisiness of the mortality series at these ages, especially in past populations, which is the reason why the HMD staff decided to smooth the life tables starting from age 80 (see the "full protocol" of the HMD). Finally, it is more difficult to explain why the proportion of cointegrated series drops in the first age groups, up to 15 years. In this case, too, the problem may be linked to the noisiness of the series, but we cannot prove this conjecture.[10]

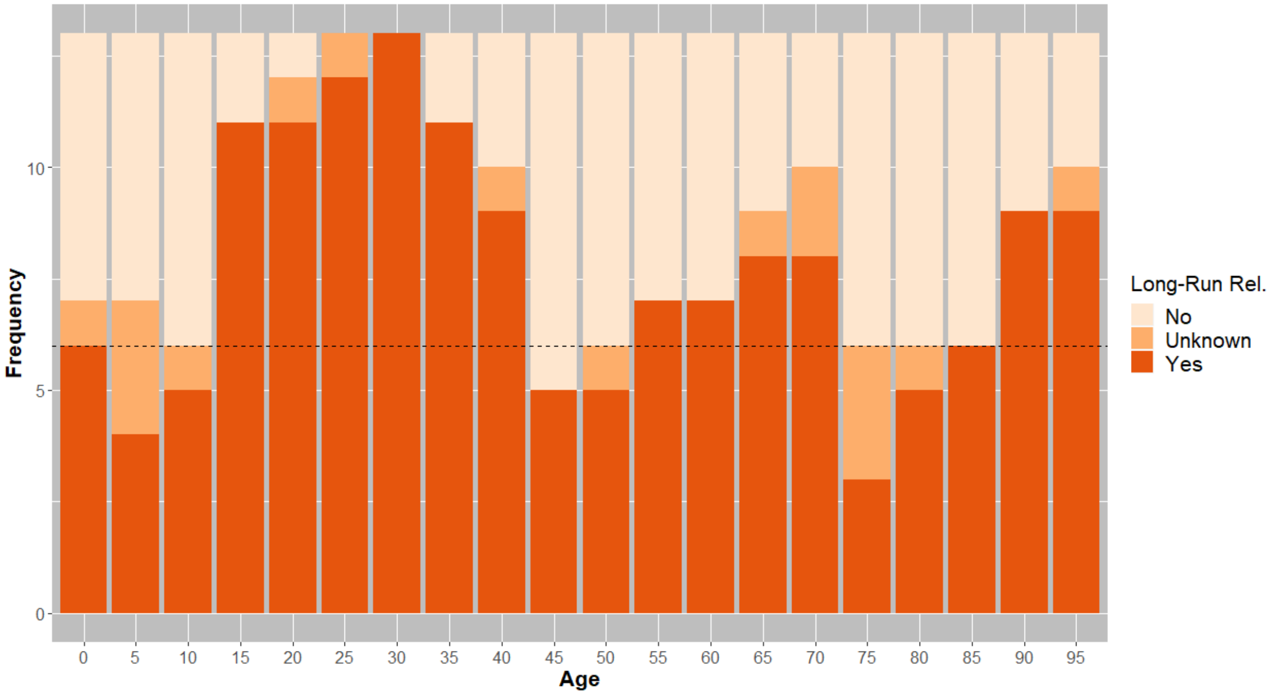
---

[9] The proportion of explained variance of the ECM models (Eq. 4) should not be confused with the explained variance of the cointegration relationship (Eq. 3): in the former the dependent variable is $\Delta c_{a,t}$, whereas in the latter the dependent variable is $c_{a,t}$.

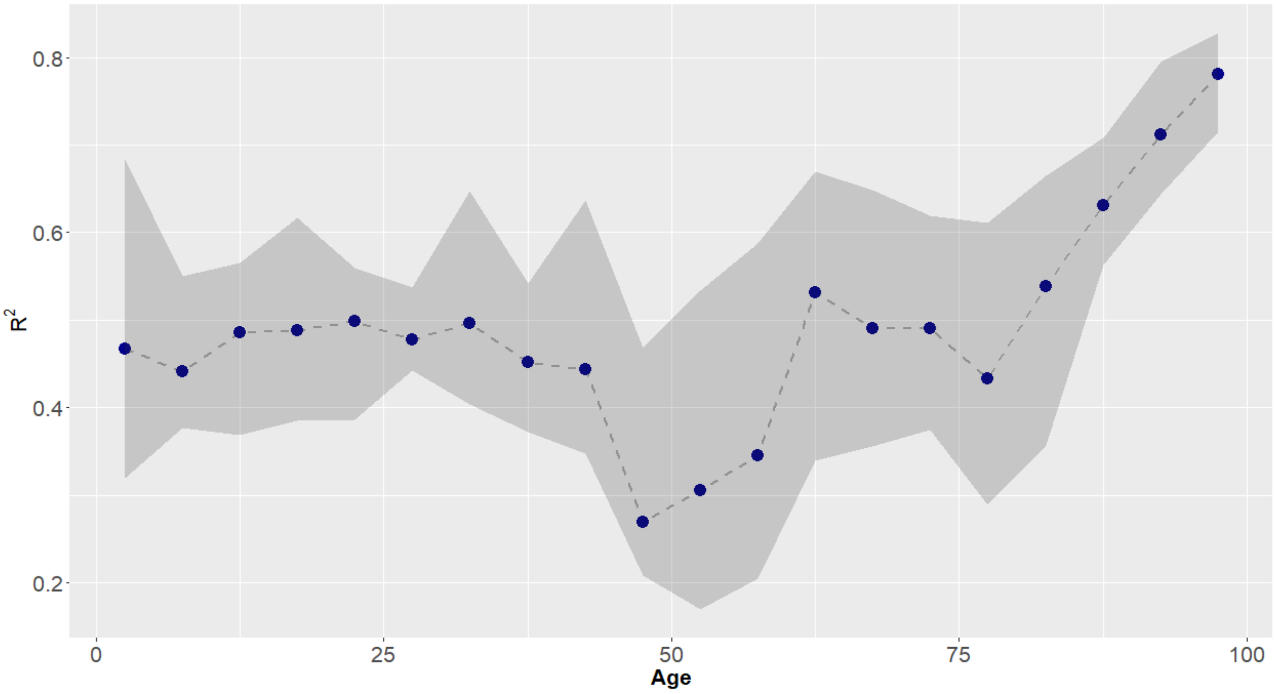
[10] We also tried to split the first age group, 0-4 in the two sub-groups 0-1 and 1-4 (not shown here), but the proportion of cointegrated series detected by our tests raises only marginally.

Fig. 4 Results by five-year age groups (indicated by their central point) for the 13 countries of Table 1

a. Bounds test



b. Median R^2 (explained variance) of the ECMs by age-group



Source: See Table 1

Let us try to summarize our findings, thus far. Once the data have been purged of the "worst" series (those with the most evident statistical problems), a long-run relationship between $c_{a,t}$ and $k_{a,t}$ can be detected in slightly less than two thirds of the cases (down to 56%, if we use the more demanding significance level of 0.05). We take this as a first, admittedly not decisive, indication in favor of the hypothesis that the actual and the reference age structure are cointegrated, especially if we consider the several disturbing factors that we could not keep under control (e.g. data quality, wars, and the typically low statistical power of cointegration tests).

To improve the statistical power of our tests we performed also a panel cointegration analysis (based on Eq. 6) on the entire age structure of the population (Table 3). The first four columns of Table 3 report the P values of the panel unit root test on our $c_{x,t}$ and $k_{x,t}$ series. These tests have been performed both via the IPS test (which does not take into account the effect of cross-sectional dependence) and via the CIPS test (which, instead, does). The IPS test supports (better: does not reject) the null hypothesis of non-stationarity in all cases except one (the $k_{x,t}$ of New Zealand). The CIPS test, instead, supports the null hypothesis of non-stationarity *in all cases*, without exceptions.

As for the analysis of residuals, the IPS test rejects the null hypothesis of non-stationarity in all cases except Netherland (P value of 0.08), which, incidentally, is in agreement with the previous analysis of the bounds (Netherland was the country with the lowest proportion of cointegrated single couple of series; see Fig. 3a). However, even for Netherland, the CIPS test finds that the actual age structure $c_{x,t}$ is cointegrated with the reference age structure $k_{x,t}$.

Summing things up, according to the IPS test, a cointegration relationship between the actual and the reference age structure can be identified in 11 out of the 13 countries analyzed; according to the CIPS test, this is true in *all* of them.

The results of Table 3 indicate that a set of age-specific linear functions (Eq. 3) likely connects the long-run dynamic of the actual age structure with the dynamic of the reference age structure, which implies, among other things, that the former can be predicted if one knows (or imagines) the evolution of the latter.

Fig. 5 shows the median R^2 value associated with the estimate of Eq. 3 at different ages in all the countries covered in this analysis. This figure shows that between 25 and 54 years the evolution of the age structure cannot be accurately predicted from the evolution of mortality. As we remarked earlier, however, this is not much of a problem, because the variability of $c_{x,t}$ is very limited here, which means that there is almost nothing to predict. Where instead the changes in the age structure of the population are more important, at younger and older ages, the part of the variance that our model can "explain" becomes substantial: about 75% in the first quinquennial age groups and well above this value at 60 years and over. And this, totally ignoring what happened to fertility and migration.

Table 3 *Panel cointegration analysis (all the countries and years of Table 1)*

Country	**IPS $c_{x,t}$ P value**	**IPS $k_{x,t}$ P value**	**CIPS $c_{x,t}$ P value**	**CIPS $k_{x,t}$ P value**	**IPS residuals P value**	**CIPS residuals P value**
Australia	1	0.30	>0.1	>0.1	<0.01	<0.01
Canada	1	0.95	>0.1	>0.1	<0.01	<0.01
Denmark	1	1	>0.1	>0.1	<0.01	<0.01
England & Wales	1	1	>0.1	>0.1	<0.01	<0.01
Finland	1	1	>0.1	>0.1	<0.01	<0.01
France	1	0.97	>0.1	>0.1	<0.01	<0.01
Netherland	1	1	>0.1	>0.1	0.08	<0.01
Norway	1	0.98	>0.1	>0.1	<0.01	<0.01
New Zealand	1	0.02	>0.1	>0.1	<0.01	<0.01
Scotland	1	1	>0.1	>0.1	<0.01	<0.01
Sweden	1	1	>0.1	>0.1	<0.01	<0.01
Switzerland	1	1	>0.1	>0.1	<0.01	<0.01
United States	1	1	>0.1	>0.1	<0.01	<0.01

Note: The table reports the P value of the IPS and CIPS tests. In the first four columns, we test the null hypothesis that the $c_{x,t}$ and the $k_{x,t}$ are non-stationary. If these series are cointegrated, the test must not reject the null hypothesis (P value > 0.1). In the last two columns, we test the null hypothesis that the residuals of Eq. 6 are non-stationary. If the series are cointegrated, the test must reject the non-stationarity hypothesis (P value < 0.1). In the IPS test we set the maximum number of lags to 3 times unit (15 years in the present case). In the CIPS test the lag order of the Dickey-Fuller augmentation has been set to 2.
Source: See Table 1.

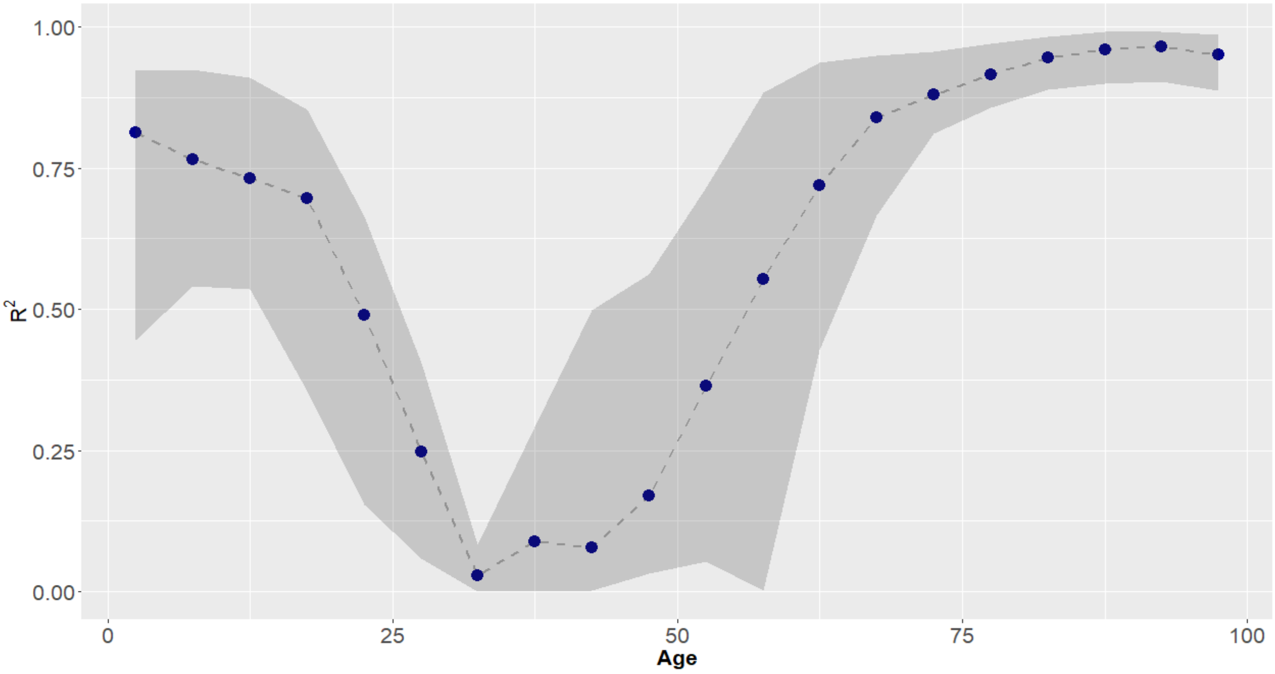
Fig. 6 presents the predicted proportions of individual of age x according to our linear equations in Sweden over the period 1820-2017 (mind the different scales on the y-axis). In this figure the red, solid line represents the evolution of the actual proportion of individuals of age x ($C_{x,t}$), the blue dotted line represents the evolution of $K_{x,t}$, while the orange-dashed line represents the prediction of this proportion ($\hat{C}_{x,t}$) based on the $K_{x,t}$ series and on the estimates of the age-specific parameters of Table 4. Our predictions are able to capture remarkably well the general evolution of the Swedish age structure in the last two hundred years, basically because the general trends observed in the $K_{x,t}$ ("reference") series are mirrored in the evolution of the $C_{x,t}$ (actual) series. Of course, we cannot accurately depict the fluctuations around this general trend, which depend on the legacy on the age structure of events such as wars, epidemics and baby booms and busts.

Fig. 6, and Fig. 7 below, show that the connection between the actual C and the predicted $\hat{C}$ values dates back to very long ago. In other words, the connection between survival (period life tables) and the shape of the age structure (which includes population ageing) seems to predate not only the strong improvement in mortality (which materialized after the 1950s; Murphy 2017), but also the onset of the demographic transition, which in Sweden started slightly after 1860.

Fig. 7 shows how accurately the entire age structure of a population at very different epochs can be reconstituted on the basis of the reference age structure (and, importantly, ignoring fertility and migration). This open the way to several empirical applications. For instance, the estimated age-specific cointegration relationships (Table 4) can be applied to mortality forecasts, e.g. of the Lee-Carter type (Lee and Carter 1992; Lee and Miller 2001) to predict the entire age structure.

In Table 4 we show the estimates at different ages of the parameters of Eq. 3 (“Eq. 3-unrestricted”) that have been employed to produce the predictions of Figures 6 and 7 along with their standard errors and the associated R^2. In this table we also present the estimates of Eq. 3 without intercept, which can be employed to assess how close in the long run the actual proportion of individuals of age x is to its reference counterpart. Finally, the last three columns of Table 4 show the mean squared errors (MSE) of our unrestricted model (Eq. 3), for the model with the intercept set to 0 and for the restricted model with zero intercept and unitary slope. The general idea behind this exercise is to assess how close, on average, the actual age structure is to its reference counterpart.

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.

Table 4 *Parameters estimates by age for Sweden (Eq. 3, restricted and unrestricted versions)*

	Type A					Type B		Type A	Type B	Type C
Age	**Intercept**	**SE**	**Slope**	**SE**	**R2**	**Slope**	**SE**	**MSE**	**MSE**	**MSE**
0	2.497	0.292	1.909	0.111	0.917	0.937	0.111	0.008	0.033	0.059
5	2.451	0.518	1.887	0.193	0.859	0.946	0.193	0.009	0.023	0.042
10	2.209	0.6	1.796	0.223	0.809	0.952	0.223	0.010	0.018	0.034
15	1.698	0.543	1.603	0.202	0.754	0.958	0.202	0.009	0.013	0.026
20	0.16	0.438	1.030	0.165	0.542	0.969	0.165	0.007	0.007	0.014
25	-0.969	0.281	0.617	0.105	0.248	0.981	0.105	0.007	0.008	0.010
30	-2.007	0.300	0.247	0.110	0.028	0.993	0.110	0.008	0.010	0.010
35	-3.894	0.309	-0.429	0.112	0.037	1.005	0.112	0.009	0.013	0.014
40	-6.25	2.433	-1.259	0.882	0.078	1.016	0.882	0.013	0.016	0.018
45	4.907	4.375	2.793	1.589	0.169	1.030	1.589	0.017	0.018	0.025
50	5.833	0.916	3.103	0.318	0.595	1.045	0.318	0.012	0.020	0.036
55	3.697	0.996	2.334	0.340	0.740	1.060	0.340	0.012	0.021	0.052
60	2.255	0.832	1.824	0.275	0.789	1.072	0.275	0.014	0.022	0.069
65	1.583	0.645	1.590	0.203	0.844	1.087	0.203	0.017	0.026	0.100
70	0.882	0.543	1.359	0.160	0.881	1.098	0.160	0.020	0.025	0.134
75	0.451	0.468	1.227	0.126	0.916	1.107	0.126	0.024	0.026	0.186
80	0.105	0.428	1.130	0.101	0.936	1.106	0.101	0.031	0.031	0.239
85	-0.049	0.431	1.095	0.087	0.955	1.105	0.087	0.038	0.038	0.337
90	-0.327	0.458	1.049	0.075	0.962	1.098	0.075	0.055	0.058	0.478
95	-0.972	0.689	0.975	0.086	0.941	1.087	0.086	0.123	0.149	0.699

Note: type A=Eq. 3, unrestricted; type B=Eq. 3 restricted, with intercept=0; type C= Eq. 3 restricted, with intercept=0 and slope=1

Fig. 6 *Actual (red, solid line), reference (blue, dotted line) and predicted (orange, dashed line) proportion of individuals in selected age-groups (Sweden, 1820-2017)*

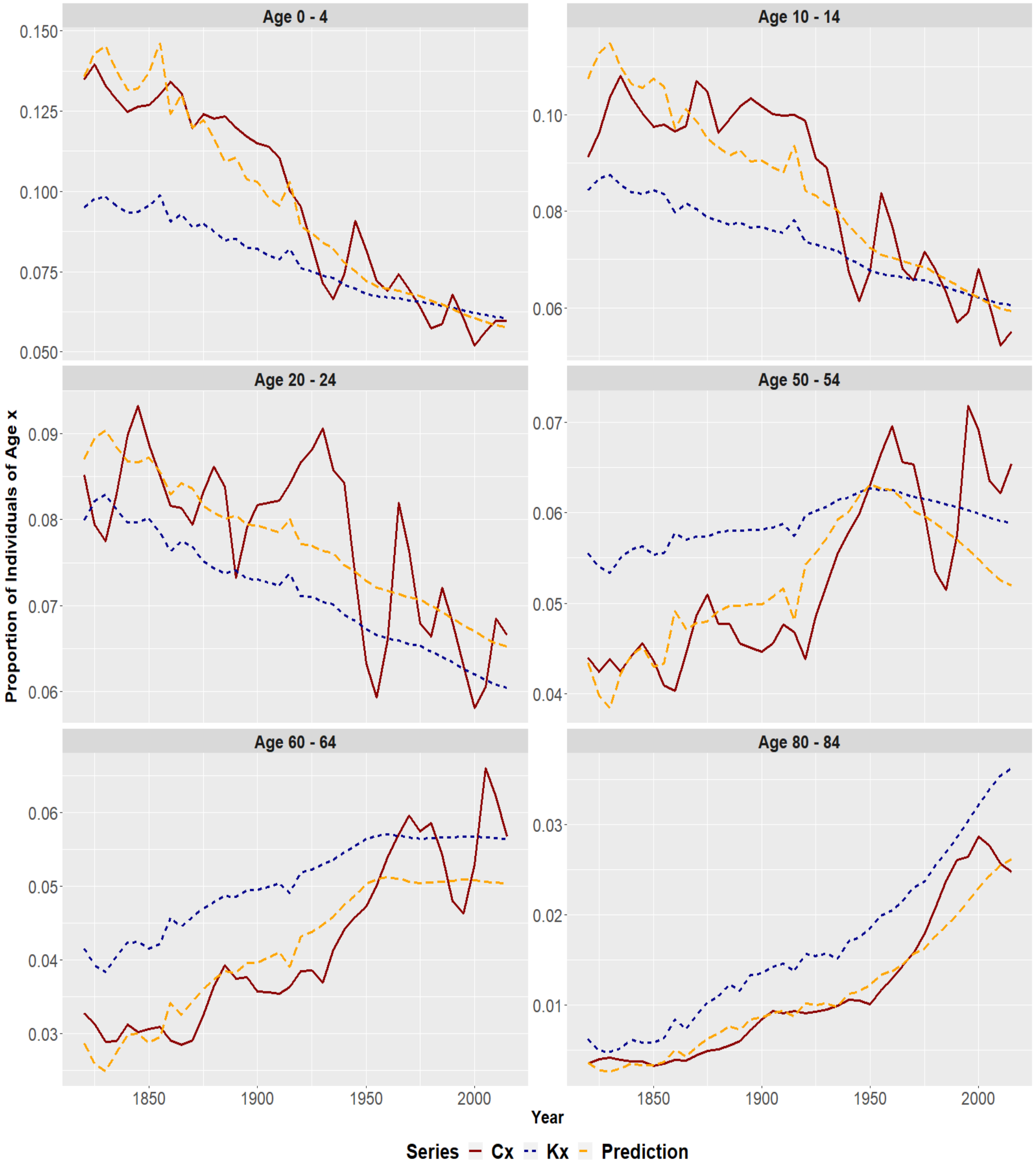
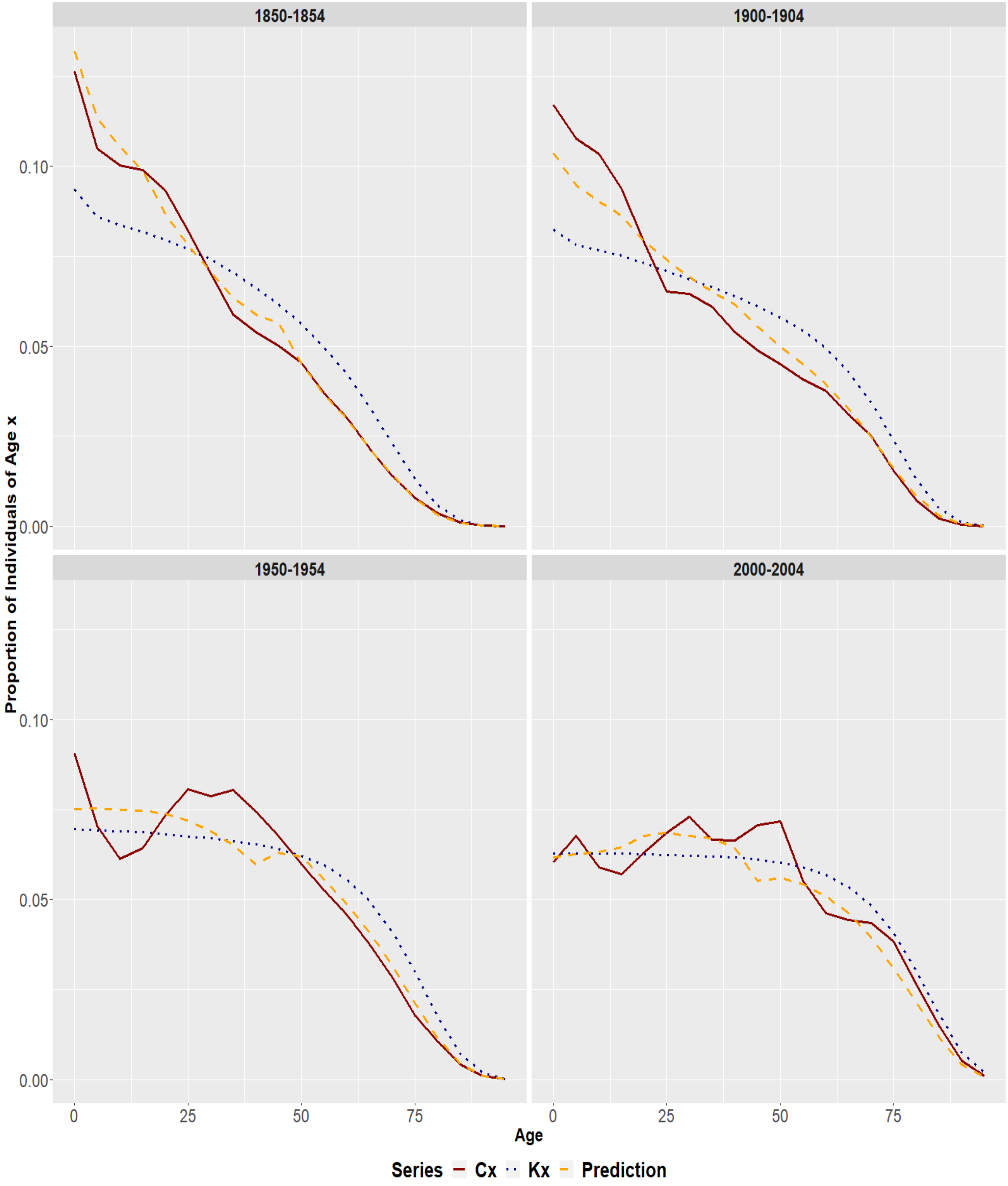


Fig. 7 *Actual (red, solid line), reference (blue, dotted line) and predicted (orange, dashed line) proportion of individuals in selected epochs*



The twice-restricted version of Eq. 3 (no intercept, unitary slope) basically shows that the reference proportion of individuals of age x is generally close to the actual proportion ($K_{x,t} \sim C_{x,t}$; see also Fig. 7). The worsening of the approximation (as measured, e.g., with the MSE) is not dramatic, while the interpretation is much simpler: as a first approximation, the age structure of the reference population in year *t* gives an idea of the actual age structure of that population in year *t*. At younger ages (<30) the actual proportion of individuals tends to be slightly higher than in its reference proportion, while at older ages (>55), the actual proportion is slightly lower (5-10%): this is normal in a population with a four-fold increased (from 2.5 million in 1820 to more than 10 million in 2018). Not surprisingly, this small distortion disappears altogether in recent times (last panel of Fig. 7), when the effects of the demographic transition are over.

5. Conclusions

In this paper, we presented an empirical finding. For 13 countries, we collected evidence that supports the conjecture that the evolution of the age structure of their population is cointegrated with the evolution of their "reference" counterpart, i.e., the age structure of the stationary population associated with the period life table. This means that most of the change observed in the proportion of young, adult and old people in these countries can be derived from (and, we submit, depend on) 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 is not spurious (i.e., due to common stochastic trends) and is rather strong (with the exception of the central ages, whose variability is very limited, however). 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 $K_{x,t}$ may constitute an acceptable approximation of the actual age structure $C_{x,t}$. These results are obtained ignoring fertility and migration altogether: this suggests that survival, both directly and indirectly (e.g. acting through changes in fertility and migration) is an important driver of the changes in population age structure and therefore also of aging.

Murphy (2017) had already found that since mid-20th century, most of the evolution of the age structure in 11 European countries depended 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, Himes and Eggers 1989). But we could: our method, less data demanding, allowed us to go much back in time, up to almost two centuries in the case of Sweden and France, for instance.

Our findings, as well as those of Murphy, complement rather than contradicting the traditional conclusion of "classical" demography, based on simulations. It is true that, *if mortality and fertility were independent of each other*, fertility would impact on the age structure more (possibly, much more) than mortality. However, the assumption may not hold: fertility (and migration) tend to adapt to the evolution of societies and, in particular, to the evolution of survival, to the point that, knowing survival (i.e., period life tables), one can safely predict the shape of the population pyramid, even if this is (heavily) dependent on fertility and migration. This happens, we tentatively submit, precisely because fertility and mortality, in the medium-long run, "adapt" to what survival dictates - or at least have done so in the past 200 years or so for the countries for which we have good data.

Acknowledgements

Gustavo De Santis gratefully acknowledges financial support from the 2016 JPI-MYBL (Joint Programme Initiative - More Years Better Lives), CREW Project ("Care, retirement and wellbeing *of older people across different welfare regimes"), MIUR Decree: n. 3266/2018 (MIUR: Ministry of Education, University and Research).*

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.

Billari, F. C., & Dalla Zuanna, G. (2013). Cohort Replacement and Homeostasis in World Population, 1950–2100, *Population and Development Review*, 39, 563–585.

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

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.

Davis, K. (1963). The Theory of Change and Response in Modern Demographic History. *Population Index*, 29, 345-366.

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

Galloway, P. R, Lee, R. D., & Hammel, E. A. (1998). Infant Mortality and the Fertility Transition: Macro Evidence from Europe and New Findings from Prussia. In M. R. Montgomery & B. Cohen (Eds.), *From Death to Birth. Mortality Decline and Reproductive Change*. Washington D.C., National Academy Press.

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

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, 115, 53–74.

Johnston, J., & DiNardo, J. (1996). *Econometric Methods*. McGraw-Hill.

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

Lee, R., & Carter, L. (1992). Modeling and forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87, 659–671.

Lee, R., & Miller, T. (2001). Evaluating the performance of the Lee-Carter approach to modeling and forecasting mortality. *Demography*, 38, 537–549.

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.

Pearl, J. (2009). *Causality. Models, Reasoning and Inference*. Second Edition. Cambridge, Cambrige University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics. A Primer*. Wiley.

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. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22, 265–312.

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.

Wei, W. W. S. (2006). *Time Series Analysis. Univariate and Multivariate Methods*. Second Edition. Boston Pearson.