A Probabilistic Cohort-Component Model for Population Forecasting – The Case of Germany

Patrizio Vanella
Helmholtz-Zentrum für Infektionsforschung
Forschungsgruppe für Epidemiologie
Inhoffenstr. 7, DE-38124 Braunschweig
patrizio.vanella@helmholtz-hzi.de

Philipp Deschermeier
Institut Wohnen und Umwelt
Field of housing markets and housing policy
Rheinstr. 65, DE-64295 Darmstadt
p.deschermeier@iwu.de

Abstract

The future development of population size and structure is of importance since planning in many areas of politics and business is conducted based on expectations about the future makeup of the population. Countries with both decreasing mortality and low fertility rates, which is the case for most countries in Europe, urgently need adequate population forecasts to identify future problems regarding social security systems as one determinant of overall macroeconomic development.

This contribution proposes a stochastic cohort-component model that uses simulation techniques based on stochastic models for fertility, migration and mortality to forecast the population by age and sex. We specifically focused on quantifying the uncertainty of future development as previous studies have tended to underestimate future risk. The model is applied to forecast the population of Germany until 2045. The results provide detailed insight into the future population structure, disaggregated into both sexes and age groups by year. Moreover, the uncertainty in the forecast is quantified as prediction intervals for each subgroup.

Keywords: Demography; Forecasting; Stochastic Simulation; Cohort-Component Method; Principal Component Analysis; Time Series Analysis; Monte Carlo Simulation

1 Introduction

The future development of the population structure is of immense importance since planning in many areas of politics and business is done based on expectations about the future composition of the population. Countries with low fertility and decreasing mortality rates, as is the case for most countries in Europe, particularly need accurate population forecasts since these demographic changes transform the long-term age distribution of the population in favor of older persons. These changes result in widely discussed future problems, e.g., for the social security systems as well as the labor market as a whole. The public discussion about the demographic change in Germany and its challenges is mostly tinged with negative undertones (Deschermeier 2011: 669). Nevertheless, the transformation of a society also represents a very positive aspect: people are getting older while experiencing more healthy and active years of life compared to those in previous generations (Schnabel et al. 2005: 3).

As a result of the European debt crisis, which hit countries in Southern and Eastern Europe especially hard, alongside expansions of the European Union with the inclusion of economically weak countries in southeastern Europe (Brücker et al. 2017a: 3), as well as increasing aggressiveness by the Taliban in Afghanistan (Bundesministerium des Innern 2011: 107; International Organization for Migration 2014: 104), net migration into Germany has increased monotonically since 2009. The net migration was amplified by the so-called *Arab Spring*, which started in 2011 in Tunisia. Since then, Islamists have gained massive amounts of power due to the power vacuum appearing after the end of dictatorships in the affected countries (Council on Foreign Relations 2012). The so-called *Islamic State (IS)* in 2014 had rapid and surprising military success, especially in Syria and Iraq, where they proclaimed a caliphate (Heidelberg Institute for International Conflict Research 2017: 189). Many people subsequently fled from these regions, leading to record refugee migration into Germany in 2015. People from Syria were essentially guaranteed legal refugee status. These changes subsequently motivated many people in places such as Serbia, Albania, Kosovo and Iran to try their chances as refugees as well. Some of these refugees even immigrated illegally using false identification documents to pose as Syrians (Aust et al. 2015; Bewarder and Leubecher 2016; Bundesamt für Migration und Flüchtlinge 2016: 14 – 50; Zeit Online 2015).

Against this background, this paper provides a stochastic population forecast of the year-end population in Germany through the year 2045. The population in each year of the forecast is broken down by sex and age for the range 0 to 102 and 103 years for males and females, respectively. We use stochastic modeling approaches developed in past contributions (Vanella 2017; Vanella and Deschermeier 2018, 2019) to forecast the demographic components of the population development. These forecasts are used to estimate the growth in the age- and sex-specific population, starting from the estimated population on December 31, 2017. In this way, we generate 10,000 sample paths for the future population by simulating a probabilistic cohort-component model by Monte Carlo simulation of Wiener processes of the demographic components.

Stochastic approaches are gaining popularity as an alternative to the common deterministic population projections that use scenarios to address future uncertainty (Keilman et al. 2002: 410). Planners and decision makers need to know which future path is most likely to occur. Stochastic forecasts based on simulations are less prone to subjective decision-making, since the results show a wide range of possible scenarios and quantify them probabilistically. As a result, the risk of personal misjudgment by the modelers is reduced. Our model returns not only the median age- and sex-specific population up to the year 2045 but also quantifies the uncertainty in the forecast, illustrated with 75% and 90% prediction intervals (*PIs*) for each year, age and sex.

The next section presents a condensed historical overview of the evolution of the cohort-component method for population updating and past advances in population projection, starting with the first deterministic models and continuing with improvements to these models through probabilistic forecasting. Our study primarily focuses on Germany; therefore, our overview gives special emphasis to population projections for Germany. In Section 3, we describe the population forecast process in detail by explaining how the demographic components fertility, migration and mortality are forecast and how these individual forecasts are combined into an overall population forecast for Germany via a probabilistic cohort-component model. Section 4 presents and discusses the results, and Section 5 provides an outlook and discusses the limitations of the presented approach.

2 Selected Population Forecasts and Projections with Special Emphasis on Germany

Future population projections are often conducted by deterministic cohort-component models. To the best of the authors' knowledge, this method dates to 1863, when the Census Bureau of England and Wales (1863) ran a projection of the population in England and Wales for the year 1881 by 20-year age groups. Births, deaths and migrations were identified as the components of demographic development. The population was projected by making assumptions about changes in birth rates, mortality rates and net migration for each age group or cohort. Cannan (1895: 508 – 515) further developed the method by taking ten-year age groups and assuming trends in age-specific fertility derived from recent census data. He projected the population in

England and Wales until the year 1951. Whereas Cannan's approach implicitly modeled international migration in combination with deaths, Whelpton (1928: 255 – 270) incorporated expectations of migration in a forecast of the U.S. population by age group, sex and ethnicity until the year 1975, setting the stage for modern cohort-component modeling.

Deterministic methods quantify a limited number of scenarios whose likelihoods of occurrence are not quantified by probability. Therefore, stochastic methods are recommended for population forecasting (Alho and Spencer 2005: 2 – 3; Bomsdorf et al. 2008: 125; Keilman et al. 2002: 410 – 412; Lee 1998: 157 – 170; Lutz and Scherbov 1998: 83).

Ledermann and Breas (1959: 637 – 681) proposed the transformation of age-specific mortality rates *(ASMRs)* into indices through singular value decomposition, which was developed geometrically by Pearson at the beginning of the 20th century (1901: 559 – 563). They were thus the first to use principal component analysis (*PCA*) to reduce the high dimensionality in demographic processes. Le Bras and Tapinos (1979: 1405 – 1449) elaborated on the preliminary work of Ledermann and Breas 20 years later in the first principal component (*PC*)-based population projection for France until the year 2075.

Bozik and Bell (1987) proposed a groundwork for stochastic modeling by applying autoregressive integrated moving average (*ARIMA*) models to forecast age-specific fertility rates *(ASFRs)* in the United States. Bell and Monsell (1991: 156 – 157) applied this method to forecasting age-specific mortality rates *(ASMRs)*. Lee and Carter simplified the Bozik-Bell and Bell-Monsell approaches to forecast age-specific mortality (Lee and Carter 1992: 660 – 668) and fertility rates (Lee 1993: 190 – 199) in the U.S. Since then, various modifications of the *Lee-Carter model* have been proposed (see, e.g., Booth 2006: 554 – 562; Booth et al. 2006: 290 – 304 for an extensive overview), maybe most notably the functional PC approach of Hyndman and Ullah (2007: 4945 – 4952).

Many population projections and forecasts[1] have been made for Germany during the past half-century; the best known is the "*koordinierte Bevölkerungsvorausberechnung"* from the German Federal Statistical Office (*Destatis*). The first version was published in 1966. Since then, 13 updates have been made with improved techniques. The basic principle involves making a set of assumptions about the long-term development of life expectancy, total fertility rate (*TFR*) and net migration (currently two to three alternatives for each) to derive age-specific statistics. These different assumptions are combined to create different realistic scenarios for the course of the future population until the year 2060 (Pötzsch 2016: 37; Pötzsch and Rößger 2015: 7 – 41). The European Union also uses deterministic methods (2015: 14 – 29).

Probabilistic population forecasts for Germany are rare. To the best of the authors' knowledge, the first approach was undertaken by Lutz and Scherbov (1998: 83 – 91). Their idea was to pool a large number of earlier deterministic projections and to approximate the distributions of the parameters by assuming Gaussian distributions. Lutz and Scherbov investigated nine population projections for Germany and derived distributions for the TFR, life expectancy and net migration. On the basis of these summary statistics and assumptions about the distributions of the age-specific rates, they calculated empirical quantiles for the population size via scenario-based simulation to obtain projection intervals through 2050. This method is very attractive when a sufficient statistical basis for inference is lacking but appears rather subjective since it is built upon the scientists' assessment of the future course of the demographic components. Subjective judgment generally has a high potential for error since it is not necessarily connected to statistical data. Furthermore, individuals experience difficulties in translating their qualitative judgment about realistic future scenarios into quantitative probabilities (Lee 1998: 168 – 170).

[1] For further reading on the distinction between forecasts and projections, see, e.g., Bohk (2012: 21 – 25).

Lipps and Betz (2005: 11 – 38) produce separate forecasts for the population in West and East Germany for the period 2002 – 2050, assuming convergence of the mortality and fertility rates in the East towards the levels in the West. They simulate 500 trajectories for a mortality index, the TFR and net migration. The age-specific mortality rates are derived through the classic Lee-Carter index, and the TFR is assumed to follow a *random walk process*[2]. Age-specific fertility rates (*ASFRs*) are deduced from the TFR with a variable Gaussian ASFR distribution. The net migration is modeled as an autoregressive process of order one (*AR(1)*). Age-specific migration is then calculated via a distributional assumption. The simulation of the time series processes produces 500 trajectories with PIs of the age- and sex-specific populations of West and East Germany.

This contribution was a major improvement on previous approaches. A general limitation of models using a fixed age schedule for the ASFR, as assumed by Lipps and Betz, is that they ignore the tempo effect in fertility, which describes the postponement of child-bearing into later points in life (e.g. Vanella and Deschermeier 2019). They assume that the mother's mean age at birth will converge to 31.45 years in the long run. This approach is quite restrictive and, at least from today's perspective, not realistic at 31.45 years[3]. Quantification of the PIs for this statistic seems problematic since the variance in the forecast is apparently constant and has the same value for 2002 and 2050. Uncertainty about the far future is probably greater than that for the near future (Box et al. 2016: 129 – 147).

Bomsdorf et al. (2008: 125 – 128) use ARIMA models to forecast the TFR and the net migration in Germany. They use these summary measures to derive ASFRs and age-specific migration via age schedules, namely, a Beta distribution for the ASFRs. Age- and sex-specific

---

[2] See, e.g., Dickey and Fuller 1979: 427; Vanella 2018: 230 for a definition of a random walk.
[3] The mean age at child birth in 2015 was 31, and long-term increases were nearly linear per annum for almost two decades (see GENESIS-Online Datenbank 2018).

measures for mortality and net migration are obtained from the Lee-Carter model, and 5,000 simulations of the time series models produce empirical PIs. Härdle and Myšičková (2009: 4 – 26) apply the Lee-Carter models for mortality and fertility to estimate these two components for Germany. Furthermore, they forecast immigration to and emigration from Germany with separate AR(1) models to estimate the population in Germany until the year 2057.

Dudel (2014: 95 – 216) non-parametrically forecasts the populations of West and East Germany until 2060 using historical simulation techniques based on 1,000 trajectories. His method, although statistically interesting, has a few caveats. First, the mortality model assumes a perfect correlation between the two genders, which statistically is unlikely (see, e.g., Vanella 2017: 543 – 552). The main trends in mortality reduction result from advances in medicine and better education among the population with regard to health and hygiene, improvements from which females and males both benefit (Pötzsch and Rößger 2015: 34). However, different developments in mortality are evident for both sexes, mostly arising from different smoking (Pampel 2005: 461 – 463; Trovato and Lalu 1996: 31 – 35; Waldron 1993: 458 – 460) and nutritional (Luy and Di Giulio 2006: 1 – 8; World Health Organization 2015) behaviors. Second, Dudel rejects trajectories for the TFR under 1 and over 3, censoring the total density. A pre-specified transformation would have mitigated this problem from the very beginning. Third, the overall migration model can be criticized because it assumes a fixed age schedule (which is unlikely) and the PIs' width remains almost constant over time instead of increasing, which has been identified as a limitation for earlier studies as well.

Deschermeier (2015, 2016) forecasts the total population of Germany until 2035. He uses the model by Hyndman and Ullah (2007) to forecast the ASFRs and applies an advanced version of Hyndman et al. (2013) to forecast ASMRs and net migration. Although the model appears promising, it also underestimates the uncertainty in the forecast. Hyndman's approach smooths the data against outliers, which may be reasonable in some cases to obtain better estimates for the mean prediction. The problem with this method is that this smoothing ignores the probability of future outliers and therefore effectively underestimates the future uncertainty by simply stating that already observed outliers cannot appear again in the future.

The United Nations (*UN*) applies the Bayesian hierarchical model of Raftery et al. (Alkema et al. 2011: 818 – 829; Raftery, Alkema, and Gerland 2014: 60 – 65; Raftery et al. 2013: 780 – 786; Raftery, Lalić, & Gerland 2014: 801 – 806; United Nations 2015: 15 – 33, 2017) for quinquennial life expectancy and TFR projections for all countries. Migration is not addressed stochastically, which results in underestimation of the uncertainty in the population projections in this case as migration is the biggest source of uncertainty. Azose and Raftery (2015: 1631 – 1634) propose a probabilistic global international migration approach based on a Bayesian hierarchical model for the age- and sex-specific net migration rates of the countries. The forecasts are then transformed into net migration numbers of all countries. Azose et al. (2016) integrate that model into the UN population projection model, applying it to probabilistic projections for all countries until the year 2100, giving Germany special mention alongside a small number of other countries.

Fuchs et al. (2018: 44-54) forecast the population until the year 2060 using time series methods for the PCs of the demographic rates. This method is the most complex population forecast in Germany to date, since it is an almost full stochastic model, taking correlations in demographic rates into account and considering the effects of migration on fertility and mortality as well. Nevertheless, the authors appear to underestimate the uncertainty as well, as the PIs of the TFR and net migration remain essentially constant after 2020.

A general problem of many studies is the probable underestimation of the future risk in the population forecasts. Some models quantify risk by qualitative judgment, which is very difficult to translate into mathematical numbers as shown earlier. On the other hand, the presented quantitative studies mostly use the Lee-Carter model for forecasting, which is mostly sufficient for the mean but naturally leads to underestimation of future risk, as this model only considers a small amount of the PCs. The risk explained by the other PCs is thus ignored in the analysis, leading to a systematic underestimation of the future uncertainty. Many models do not quantify the uncertainty in migration at all, which is especially problematic, as international migration is the most uncertain of all demographic components. The overview of the relevant literature shows that approaches for population forecasting for the case of Germany that model all three demographic components by age and sex stochastically do not yet exist, with the exception of Azose et al. (2016). Our contribution is to propose an approach that is not only fully probabilistic but also considers the autocorrelations and cross-correlations of the demographic rates.

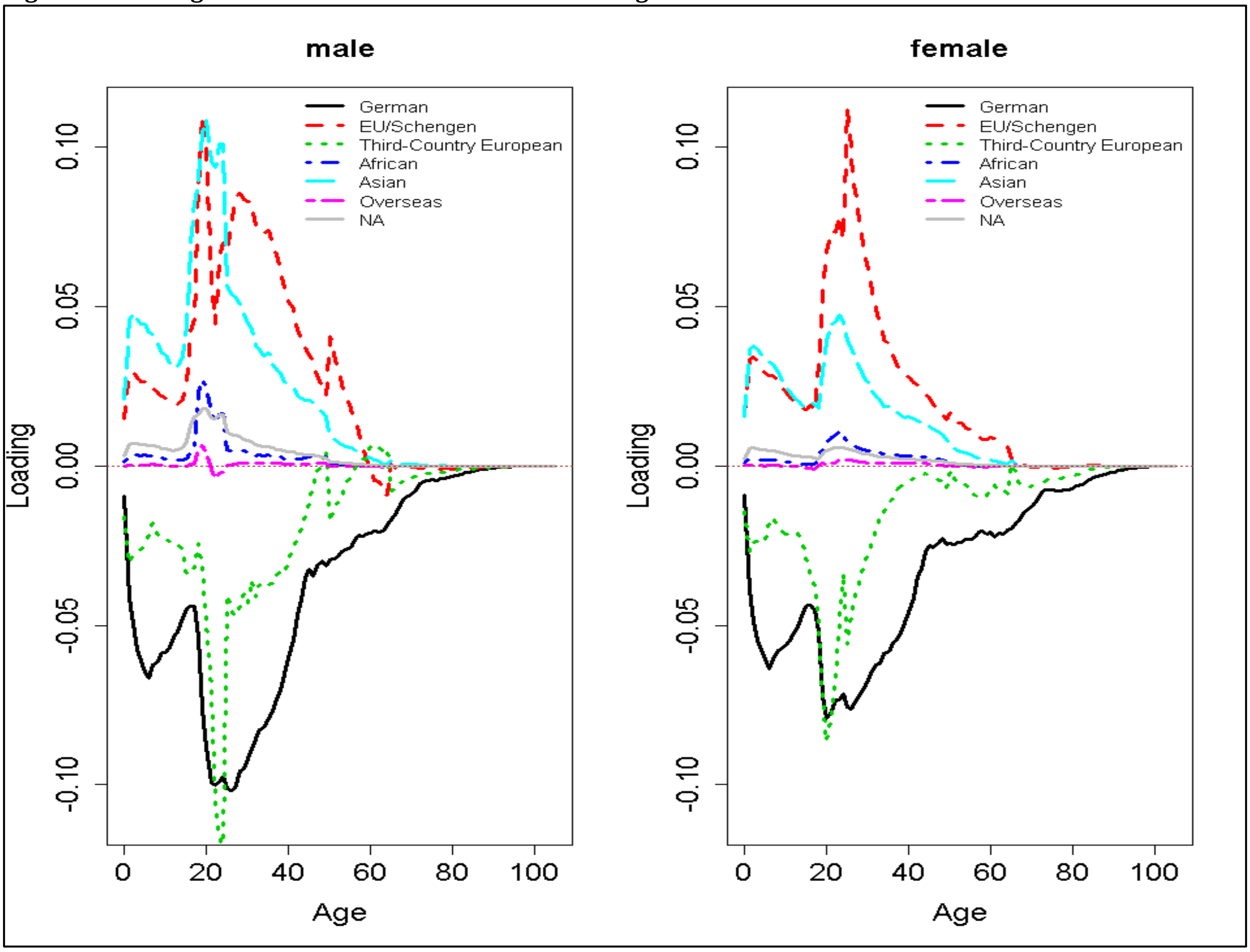
3 Method and Data

In this section, we propose a population forecast based on a probabilistic cohort-component model. The partial models for the demographic components shall be explained shortly.[4] First, the age-, sex-, and nationality-specific net migration *(ASNSNM)* figures are forecast as in Vanella and Deschermeier (2018). The data used are synthetic net migration figures per years of age (0-105), sex (binary) and nationality group, which are estimated by the authors using two data sets provided by Destatis for that study. The nationalities are split into seven groups: Germans, EU- or Schengen-citizens excluding Germany, Third-Country Europeans, Africans, Asians, Citizens from the Americas or Oceania *("Overseas")*, and finally persons with no clear information on their citizenship, either because it is unknown or they have none *("NA")*. The synthetic[5] data used for that study are estimated through two datasets provided by Destatis; the

[4] The original sources serve as a more detailed description of the models and their results.
[5] Our dataset does not exist as such but is rather estimated from different sources used by Vanella and Deschermeier (2018) in their study. Therefore, we call it a *synthetic dataset.*

first includes age-specific migration data by sex, divided by Germans and non-Germans (Destatis 2015, 2016, 2017a, 2018a, 2018f), and the second dataset is disaggregated by nationality and five age groups (Destatis 2017b, 2018b, 2018g).[6] Vanella and Deschermeier (2018: 266 – 267) derived the synthetic dataset used for the analysis from these two provided datasets. The base time period is 1990 – 2017. We run a principal component analysis (*PCA)* on the derived 1,484 ASNSNM figures. The loadings of the first two PCs are for both sexes and the different nationality groups given in Figure 1 and Figure 2.

Figure 1. Loadings of the Labor Market Index of the Migration Model

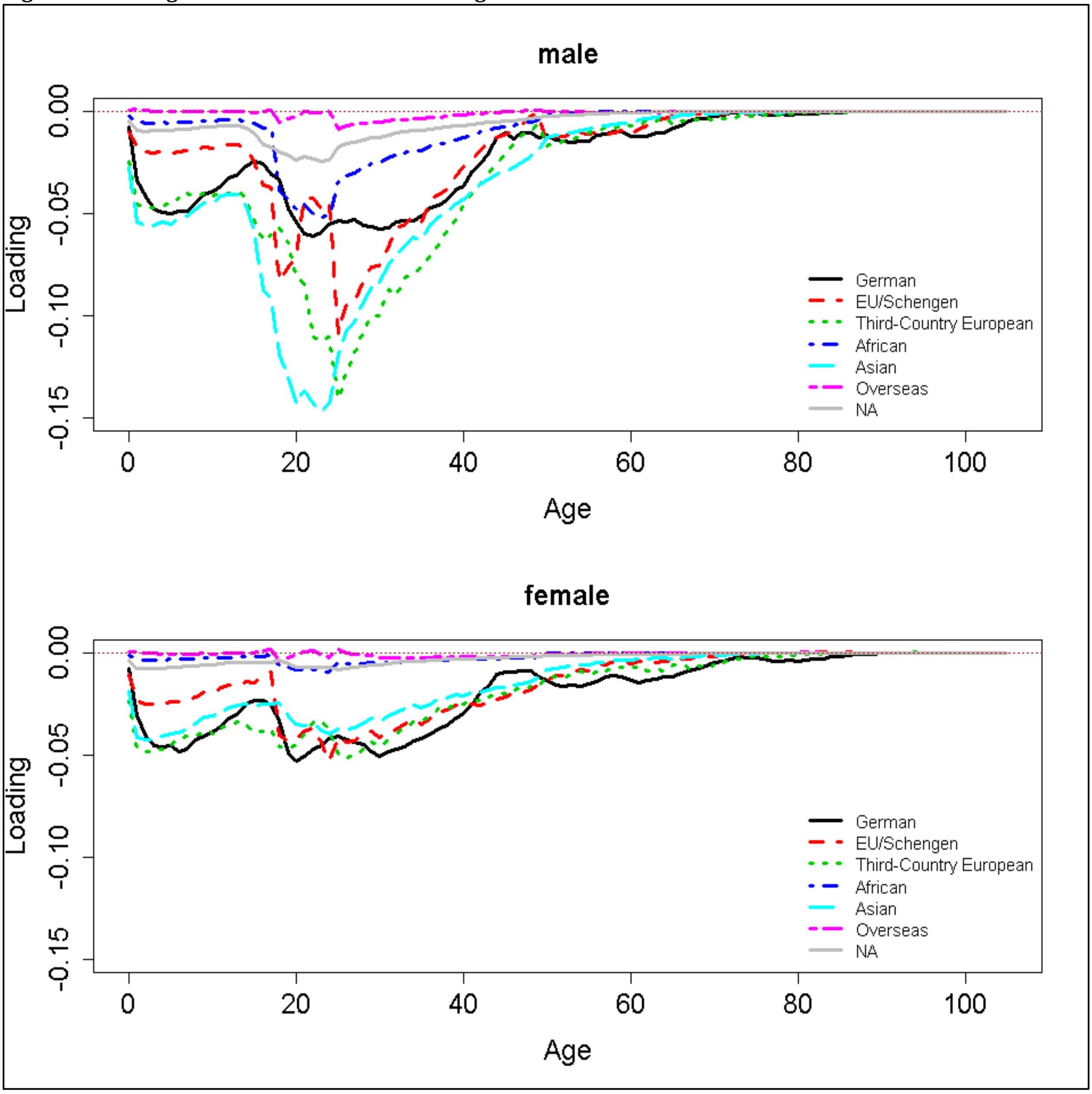


Source: Own calculation and design

[6] The exact method for deriving the synthetic data is outlined in Vanella and Deschermeier (2018: 264-271).

Vanella and Deschermeier (2018: 268) identified the first PC as an index of labor migration due to the high positive loadings on European and Asian net migration alongside high negative loadings on Germans in the working age group.

Figure 2. Loadings of the Crises Index of the Migration Model

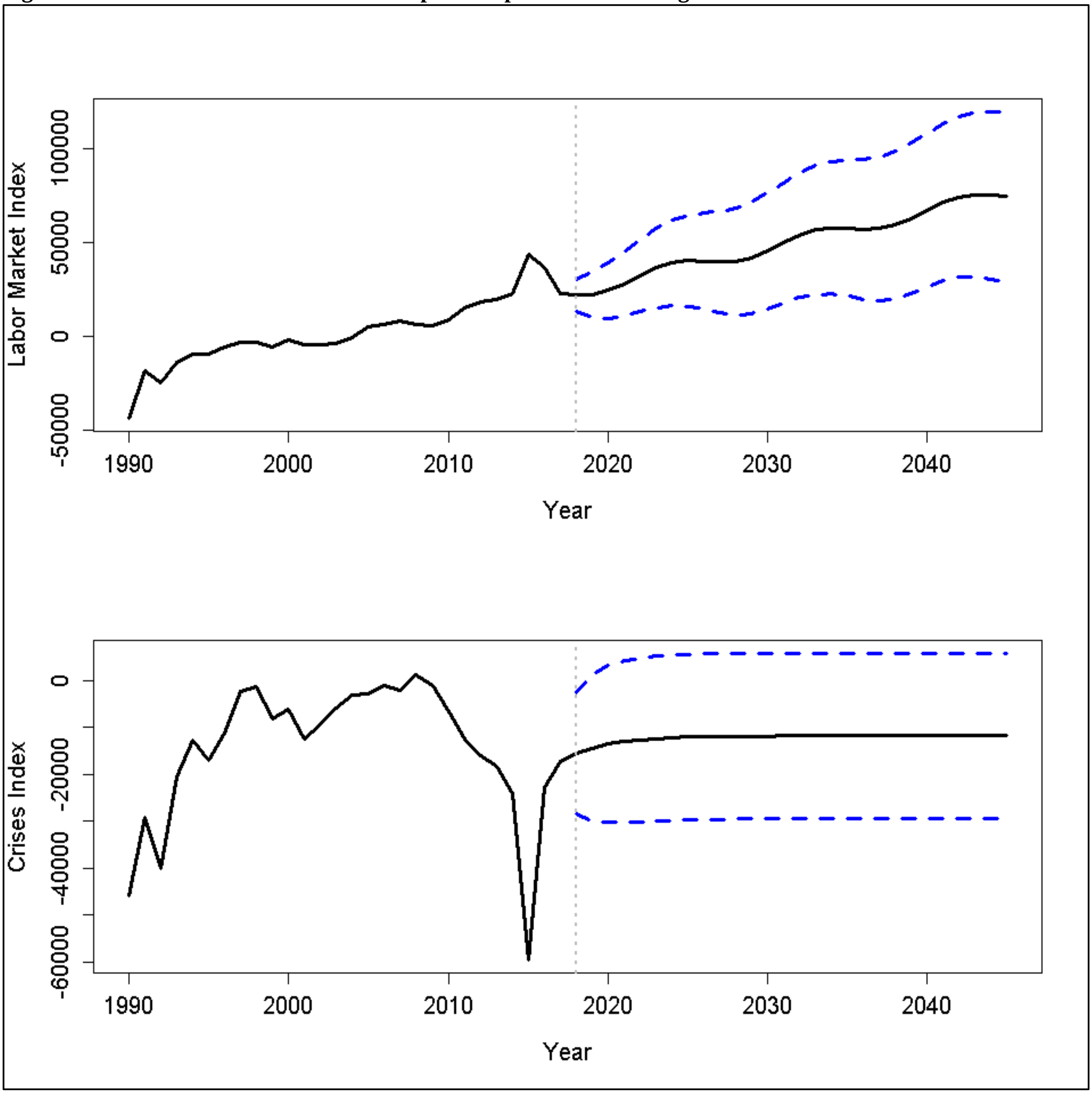


Source: Own calculation and design

The loadings of the second PC are non-positive, thus addressing the overall net migration level. In combination with the historical course, Vanella and Deschermeier (2018: 269 – 270) argue that the absolute value of the PC is especially large in times of significant crises, therefore addressing it as a *Crises Index.*

The historical course together with the forecast of these two variables through 2045 is plotted in Figure 3.

Figure 3. Forecasts of the First Two Principal Components in the Migration Model



Source: Own calculation and design

The Labor Market Index has an increasing long-term trend on average and includes cyclical effects, which are typical for labor markets. The Crises Index is assumed to converge towards its median during the base period due to a lack of better knowledge. The models are fit via ordinary least squares regression in the first step. This serves the derivation of the long-term trend in the index, which is then extrapolated in the forecast. The resulting noise is estimated with ARIMA models, which are then used for future simulation to consider the uncertainty in the forecast. Using this procedure, we can easily assess which trends are stable in the long run and which changes in the PCs result from stochastic elements. ARIMA models are very suitable to reflect the uncertainty in a forecast since future risk accumulates over time and typically becomes greater for more distant points in time. ARIMA models emulate this phenomenon very well.

The remaining 1,482 PCs are assumed to be random walk processes and are simulated accordingly. This assumption accounts for the overall uncertainty in the migration risk. Previous PC-based forecast approaches tend to omit the majority of the PCs from the analysis for simplification. On the downside, this results in a systematic underestimation of future risk since the risk of the omitted variables is ignored. Since our objective is to represent the future uncertainty as realistically as possible, we follow the random walk approach. Random walks are very easy to simulate and represent the true behavior of the 1,482 variables reasonably well. In our case, the Labor Market Index and the Crises Index alone explain 76% of the overall variance in the ASNSNM for 1990 – 2017. Thus, completely omitting the remaining 1,482 PCs would basically systematically underestimate the future risk by 24%. Therefore, we have a reasonable trade-off between simplification of the dimensionality of our forecast problem while still considering the true risk in the forecast. The resulting 10,000 trajectories of the future course of the PCs are transformed back into forecasts of the ASNSNMs through 2045, which are then finally aggregated by sex and age for the net migration forecast, as the final population forecast does not discriminate by nationality. The results of the forecasts are presented in Section 4 among other simulation outcomes.

The model for mortality is based on Vanella (2017). Death numbers by year of age and sex have been provided by Destatis for the years 2000 – 2017 (Destatis 2017c, 2018c, 2019a). Since Destatis does not provide detailed death count data for the age group 100+ for years before 2000, we took the estimated age- and sex-specific death numbers for the period 1956 – 1999 from the Human Mortality Database (2018a, 2018b, 2018c) to produce time series for the period 1956 – 2017 of all age- and sex-specific death variables. For 1956 – 1989, the data are split by West and East Germany and were therefore aggregated. We do not use the data by actual age reached at time of death but, rather, use the difference between year of death and birth year as age. Since the Human Mortality Database *(HMD)* data on deaths in essence originate from Destatis as well (Scholz et al. 2018: 2), we have no conflicting inputs that might result from meshing different data sources.

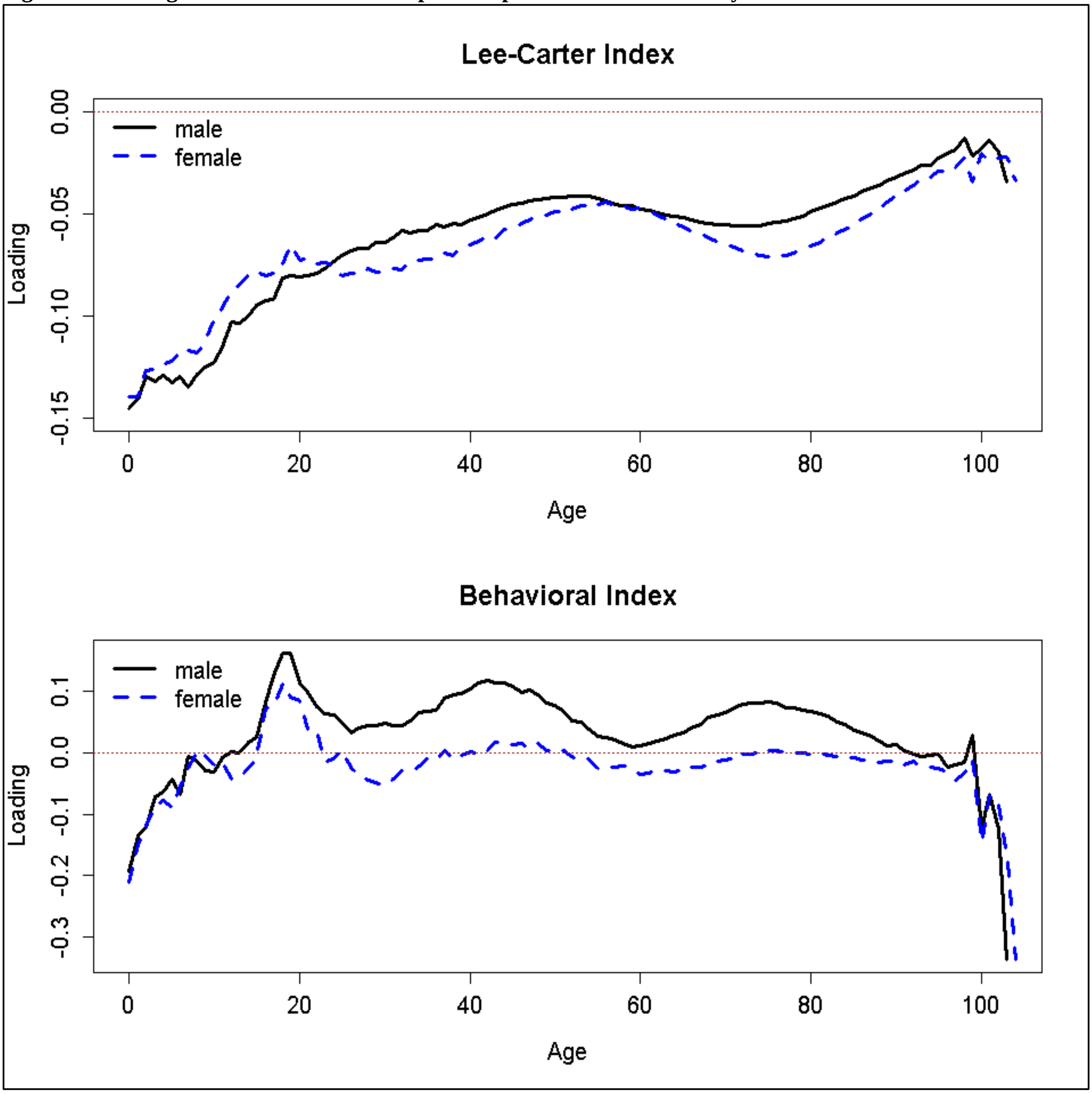
In addition, year-end age- and sex-specific population estimates were downloaded from the HMD for the period 1956 – 2017 (Human Mortality Database 2018d, 2018e, 2018f). The population data from the HMD are based on Destatis data as well but contain estimated values up to 109 years of age. Moreover, the data back to 1989 are adjusted to the 2011 census following Klüsener et al. (2018). The original Destatis data overestimated the old-age population heavily since the previous censuses for the two parts of Germany had been conducted in the 1980s, which led to massive errors in the data due to the German reunification and large errors in the data on international migration (Scholz et al. 2018: 2 – 5). Based on these data, we estimated adjusted age- and sex-specific mortality rates *(ASSMRs)* as follows:

$$m_{x,y,g} = \frac{D_{x,y,g}}{P_{x,y,g} + D_{x,y,g}},$$

with $m_{x,y,g}$ being the ASSMR of persons aged x years of sex g at the end of year y . $D_{x,y,g}$ addresses the corresponding number of people who have died over the course of year y, and $P_{x,y,g}$ is the total population of same age and sex living at the end of year y. In this way, the mortality rates calculated are computed recursively for the period under study instead of based on the population at risk at the end of the period before that. Our way of quantifying the mortality risk has the advantage that migration occurring in the year under study is considered in the overall mortality. Moreover, mortality rates above 1, which may appear in the common definition, are not possible since we base our mortality risk on the hypothetical population that would have been estimated if no deaths had occurred. The timing of migration and births over the year is represented in the data as well; therefore, we use the term *adjusted* mortality rates. Deaths for age group 103+ for males and 104+ for females are aggregated since the population estimates for these age groups are too small and therefore lead to unrepresentative estimates. As a result, the ASSMRs represent mortality for these age groups accordingly. Age- and sex-specific survival rates *(ASSSRs)* result from subtracting the corresponding ASSMRs from 1. A PCA is performed for the ASSSRs. The loadings of the first two PCs are illustrated in Figure 4.

Vanella (2017: 543 – 548) identified the first two PCs resulting from the PCA as a classical Lee-Carter Mortality Index and a Behavioral Index regarding nutritional and smoking behavior, respectively, which explain the gender gap in mortality to some extent.

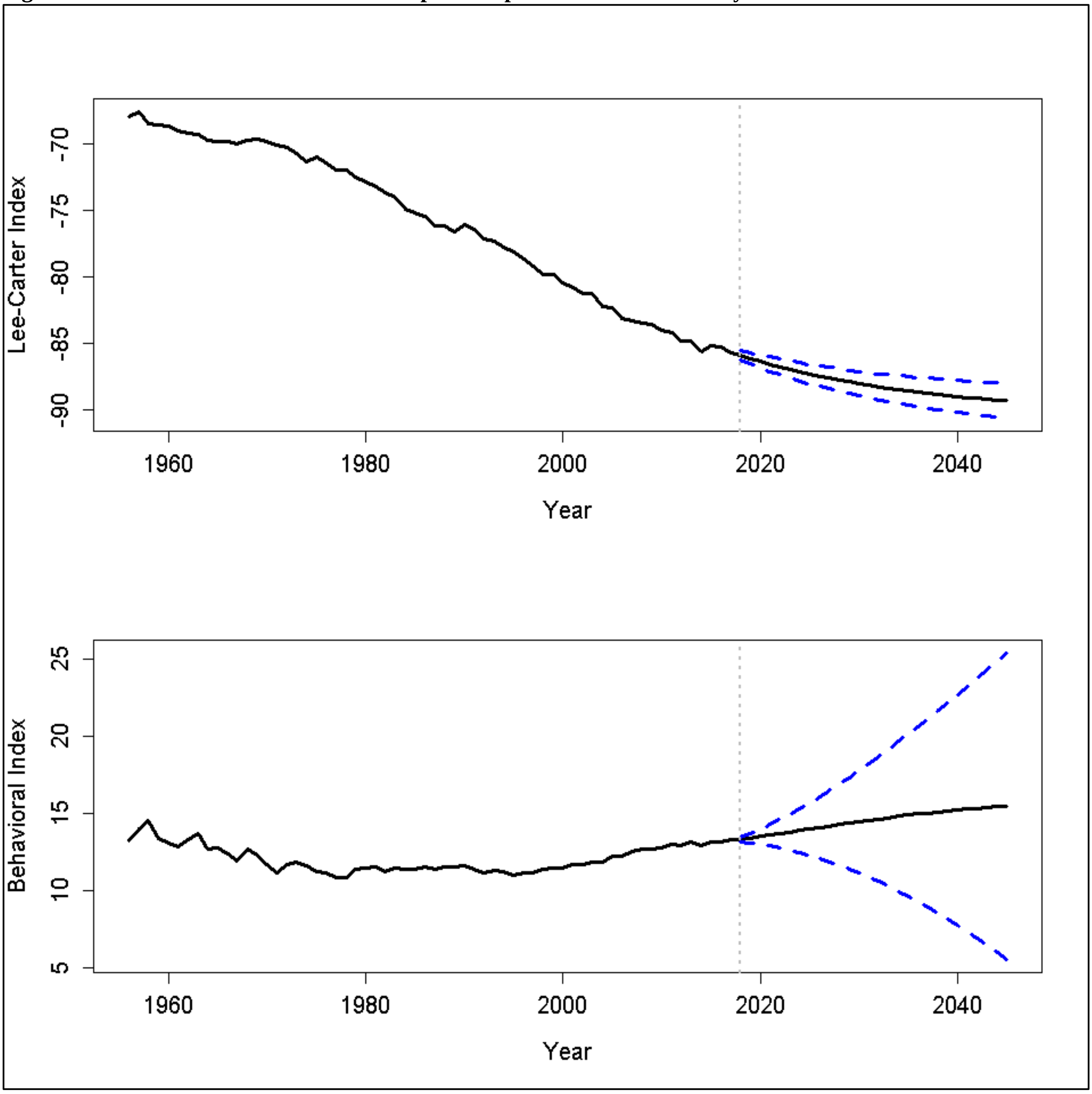
Figure 4. Loadings of the First Two Principal Components in the Mortality Model



Source: Own calculation and design

Figure 5 gives the forecasts of these two indices:

Figure 5. Forecasts of the First Two Principal Components in the Mortality Model



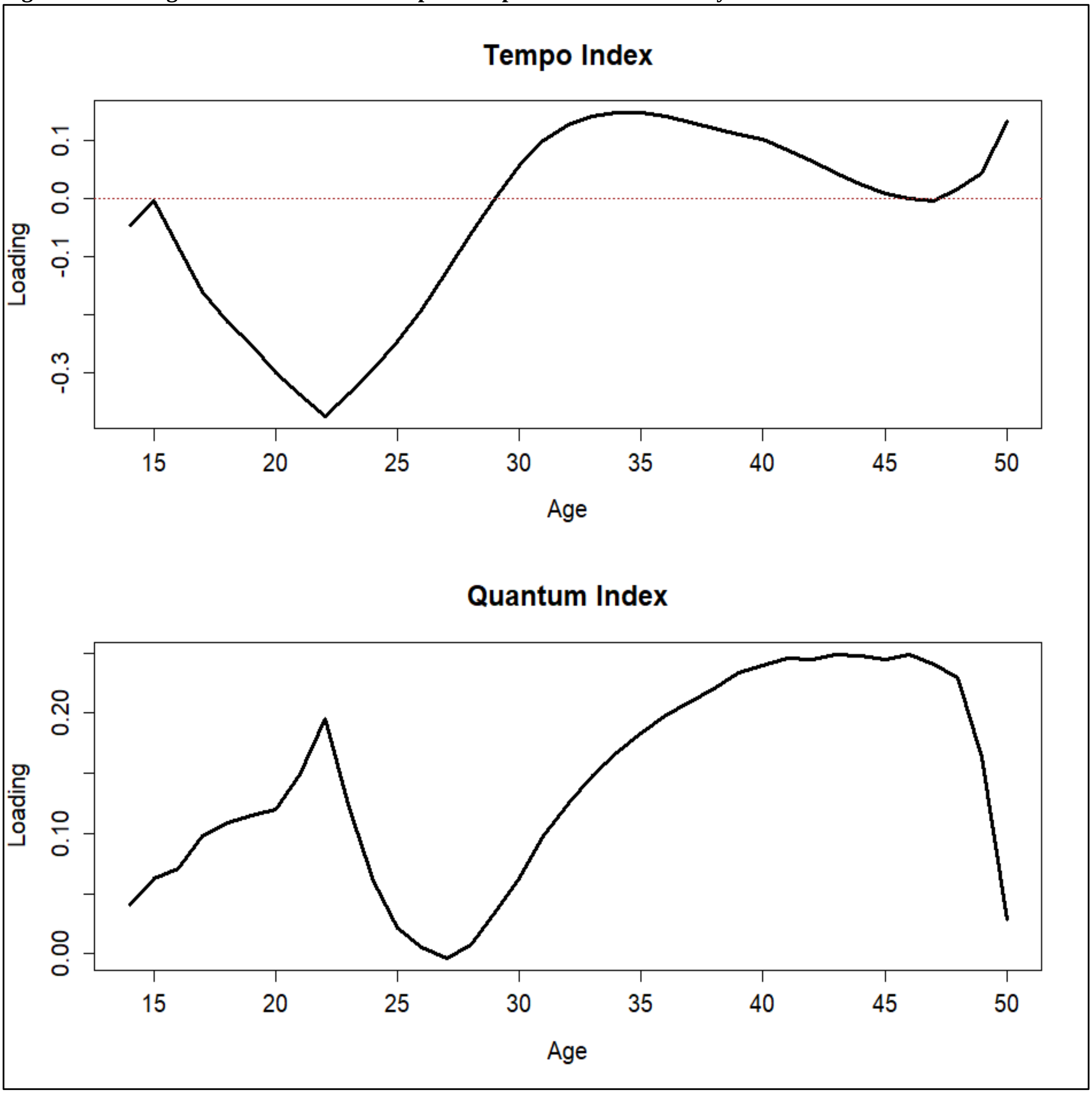
Source: Own calculation and design

The development of the Lee-Carter Index shows a general trend of decreasing mortality over all age groups, which has slowed in recent years. The expected increase in the Behavioral Index reflects convergence in nutritional and smoking behaviors between males and females. The width of the PIs of the Behavioral Index forecast shows the high uncertainty associated with this convergence trend. For both variables, the long-term trends are estimated by logistic models, which are specified by maximum likelihood estimation.

Regarding fertility, we use data on age-specific births among individuals aged 15 to 49 for the years 1968 to 2017 provided by Destatis directly or downloaded from GENESIS-Online (GENESIS-Online Datenbank 2019a; Destatis 2007, 2014a, 2014b, 2018d, 2018e) together with the age-specific data on the female population of reproductive age. Specific birth data on younger or older mothers are not available; therefore, births to mothers under 15 and mothers over 49 years of age are estimated as one age group each. Following Vanella and Deschermeier (2019), we assume 13 as the minimum age at birth and 54 as the maximum. We derive age-specific fertility rates (*ASFRs)* by dividing age-specific births by the corresponding mean age-specific female population for the respective year. As proposed by Vanella and Deschermeier (2019: 86 – 100), we run a PCA on the logistically transformed ASFRs for mothers aged 15 – 49 years and for the age groups 13 – 14 and 50 – 54 for the base period 1968 – 2017[7]. This time horizon was proposed in that paper because it shows fertility developments after the second wave of the feminist movement (Hertrampf 2008). Vanella and Deschermeier (2019: 89 – 95) show that the first PC represents the tempo effect in fertility, whereas the second PC is associated with the general quantum of fertility and is to some extent influenced by family policy. Figure 6 illustrates the loadings of these two PCs.

[7] The authors propose, based on the historical data and further considerations, 1/6 as the upper bound for the ASFRs (Vanella and Deschermeier 2019: 89).

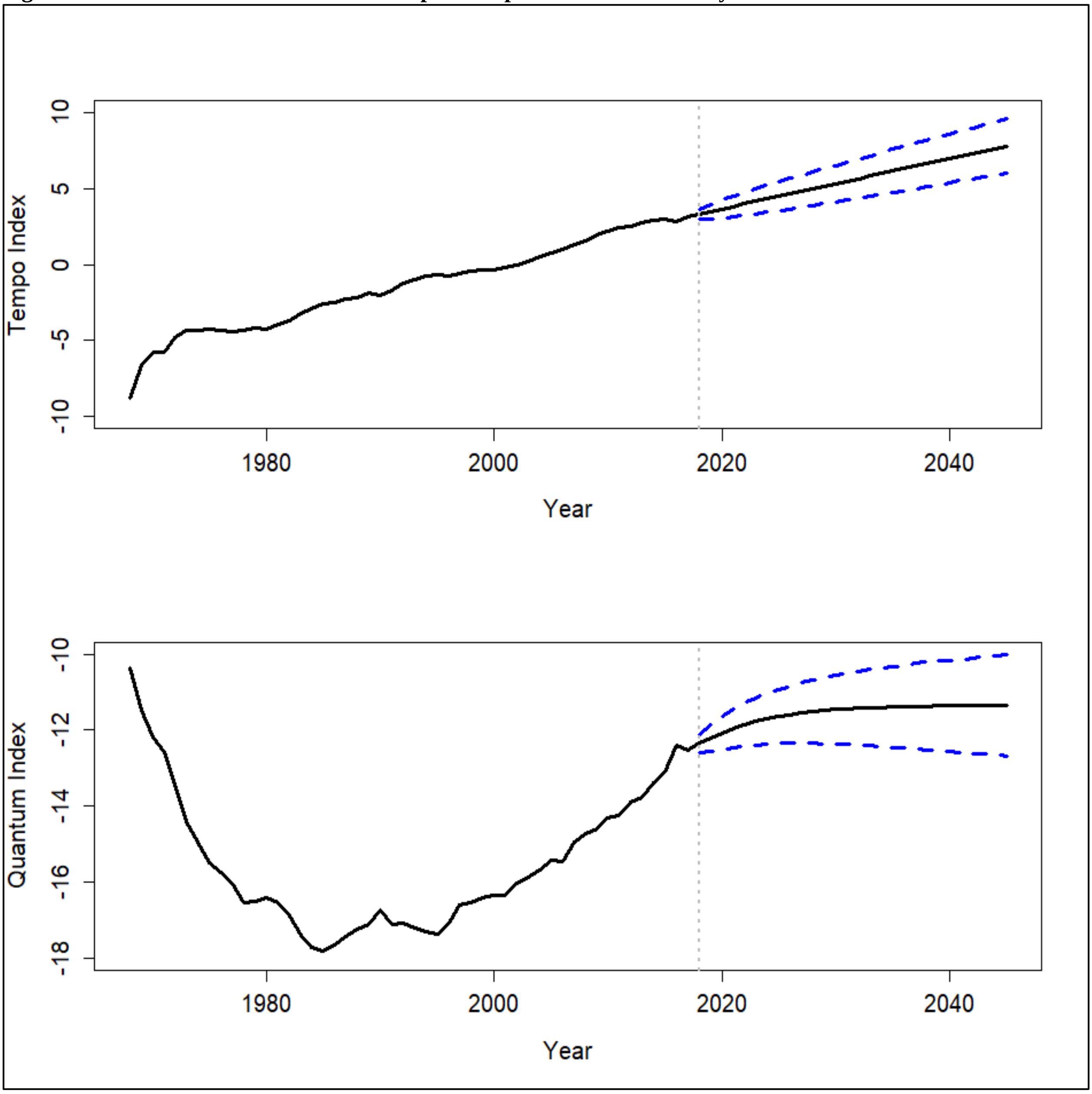
Figure 6. Loadings of the First Two Principal Components in the Fertility Model



Source: Own calculation and design

Figure 7 shows the historical courses of these two variables with the forecast until the year 2045.

Figure 7. Forecasts of the First Two Principal Components in the Fertility Model



Source: Own calculation and design

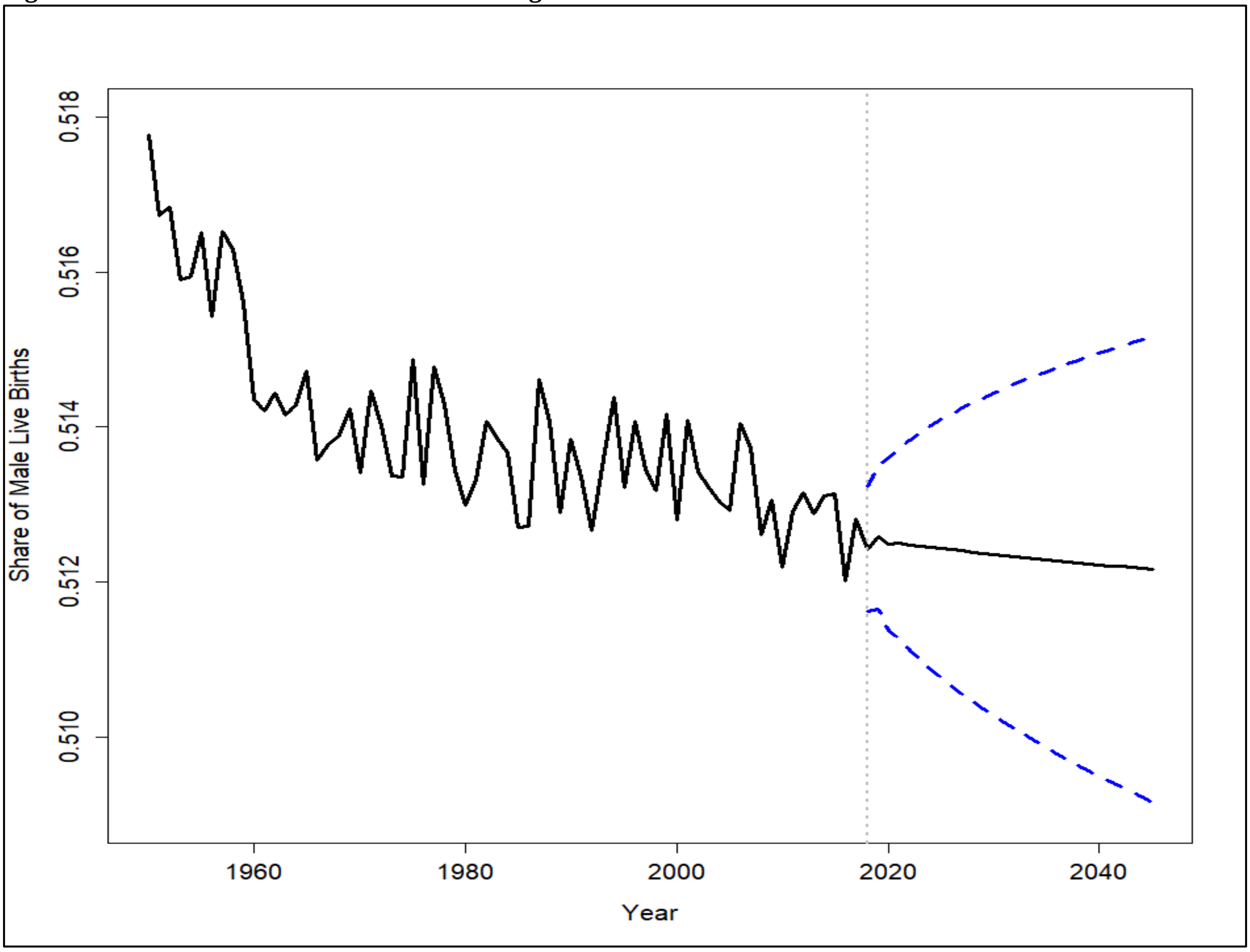
The forecast of the Quantum Index, which addresses the quantum of fertility, is influenced by family policy[8], as Vanella and Deschermeier (2019: 91 – 100) have shown. The authors discuss a variety of models to fit to the Quantum Index. Their backtest for forecasting the TFR results

[8] Several studies show positive effects of family policy on the fertility level; see, e.g., Kalwij (2010) on the effects of parental leave entitlements and daycare opportunities and Gauthier and Hatzius (1997) on the impact of cash benefits on fertility.

in good fits of the so-called *Status Quo Scenario* and *Convergence Scenario*. Whereas the Status Quo Scenario poses specific assumptions on the future statutory financial investment into family policy, the Convergence Scenario extrapolates the Quantum Index by a logistic process. The implicit underlying assumption here is that further investment into family support through financial benefits, parental leave opportunities and daycare coverage is undertaken and that these have a positive impact on fertility but that this impact is not equal to the past increases and weakens over time. The assumptions appear reasonable, and the resulting mean forecast of the TFR resulting from our calculation (1.67 in 2045) leads to similar results as in Vanella and Deschermeier (2019: 99) and are in the interval classified as realistic for the TFR by Destatis (2019b: 15). Therefore, we take the Convergence scenario as the basis of our forecast of the Quantum Index.

The gender of the children is simulated after computing the birth numbers. Therefore, we calculate the ratio of males among all live births annually based on the sex-specific birth numbers in Germany from 1950 to 2017 extracted from GENESIS-Online (GENESIS-Online Datenbank 2019b). We then fit a logarithmic ARIMA model to the data for simulation of the birth ratio until 2045. The ratio's historical course alongside the median forecast and 75% PIs is given in Figure 8.

Figure 8. Forecast of the Share of Males Among Live Births



Sources: GENESIS-Online Datenbank 2019b; Own calculation and design

An apparent trend of a decreasing ratio of male births is evident over the analyzed horizon. This trend can also be observed in other industrialized countries since at least the 1970s (Davis et al. 2007: 941 – 943; James 2000: 1179 – 1182). Although various studies individually report some evidence that environmental factors such as weather (Helle et al. 2008), exposure to toxins (see, e.g., Davis et al. 2007: 941 – 942), and nutritional behavior (Mathews et al. 2008: 1662 – 1666) have some influence on a baby's sex, none of the findings explain the observed trends of decreasing ratios of male births. Considering the apparent basic trend since 1950, assuming that the trend will continue over the forecast horizon is plausible.

All described models are based on principal component time series models and thus include autocorrelations in the time series alongside cross-correlations among the age- and sex-specific demographic rates and numbers.

We now describe the procedure for population forecasting with our model. Let $P_{x,y,g,t}$ denote the population aged *x* years at the end of year *y* for sex *g* in trajectory *t*. The population update is performed through the following step-wise process.

Step I:

The forecast begins with an adjustment of the base population with regards to international migration flows in the first forecast year *y+1***.** The addition of international net migration aged *x+1* years of sex *g* during year *y+1* and in trajectory t ($M_{x+1,y+1,g,t}$) to $P_{x,y,g,t}$ leads to the hypothetical subpopulation $\tilde{P}_{x+1,y+1,g,t}$ at the end of year *y+1* without any deaths:

$$\tilde{P}_{x+1,y+1,g,t} = P_{x,y,g,t} + M_{x+1,y+1,g,t}.$$

Step II:

The actual number of survivors from $\tilde{P}_{x+1,y+1,g,t}$ at the end of *y+1* is calculated through multiplication with the adjusted ASSSR $s_{x+1,y+1,g,t}$ for persons aged *x+1* years of sex *g* in year *y+1* and in trajectory *t*:

$$P_{x+1,y+1,g,t} = \tilde{P}_{x+1,y+1,g,t} * s_{x+1,y+1,g,t}.$$

Step III:

The mean female population by age in *y+1* in the reproductive age group is approximated:

$$F_{x,y+1,w,t} = \frac{P_{x-1,y,w,t} + P_{x,y+1,w,t}}{2}.$$

Step IV:

The live births $B_{y+1,t}$ are estimated:

$$B_{y+1,t} = \sum_{x=u14}^{50+} F_{x,y+1,w,t} * f_{x,y+1,t},$$

where $f_{x,y+1,t}$ denotes the ASFR for females aged *x* years in year *y+1* in trajectory *t*.

Step V:

The numbers of the male $B_{y+1,m,t}$ and female $B_{y+1,w,t}$ live birth numbers are calculated:

$$B_{y+1,m,t} = B_{y+1,t} * r_{y+1,m,t},$$

with $r_{y+1,m,t}$ representing the share of male live births in year *y+1* in trajectory *t*. Subsequently, the female birth numbers are

$$B_{y+1,w,t} = B_{y+1,t} * (1 - r_{y+1,m,t}).$$

Step VI:

The number of survivors among the children born in *y+1* is calculated:

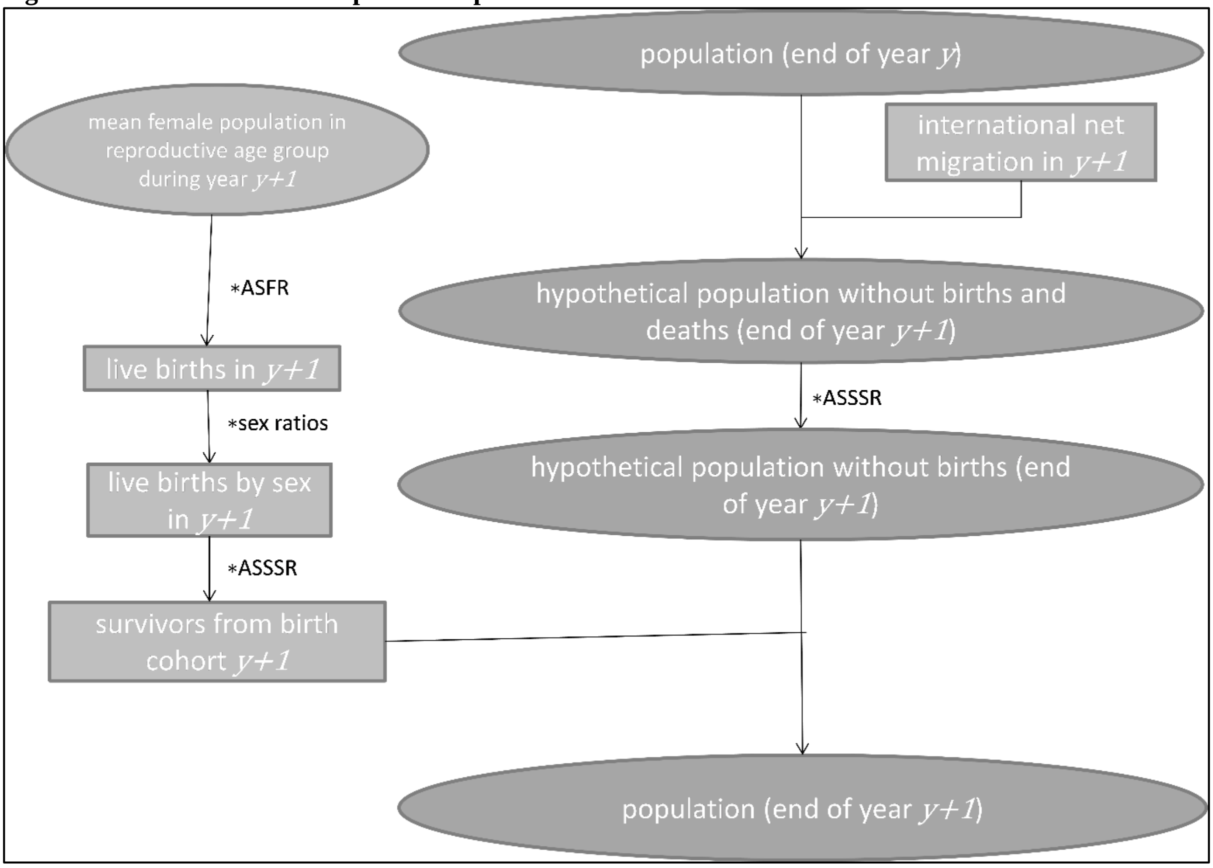
$$P_{0,y+1,g,t} = B_{y+1,g,t} * s_{0,y+1,g,t}.$$

We reiterate that our model includes the timing of births, migration and deaths over the year implicitly as well since the input data are adjusted as such. Assuming that the future timing of these demographic variables is similar to the timing observed in the past data, the timing of the demographic movements is included in the forecast as well. Migrants and newborns enter the population under study distributed over the year. Therefore, they are not at risk of death for the full year in the population under study. We account for this through our adjusted data, as our past data used for the forecast include this information as well through our cohort perspective.

If, e.g., children on average are born in the middle of the year, the ASSSRs in our model indeed do not represent the actual probability to survive the whole year, as would the case for most models, but instead would give the probability of surviving the semi-year in which they lived in Germany until the end of the period, on average.

In this way, the population by sex and age in year *y+1* in trajectory *t* is obtained. This process is then used to stochastically forecast the population by sex and age until the year 2040. The algorithm is illustrated in Figure 9.

Figure 9. Process of Annual Population Update



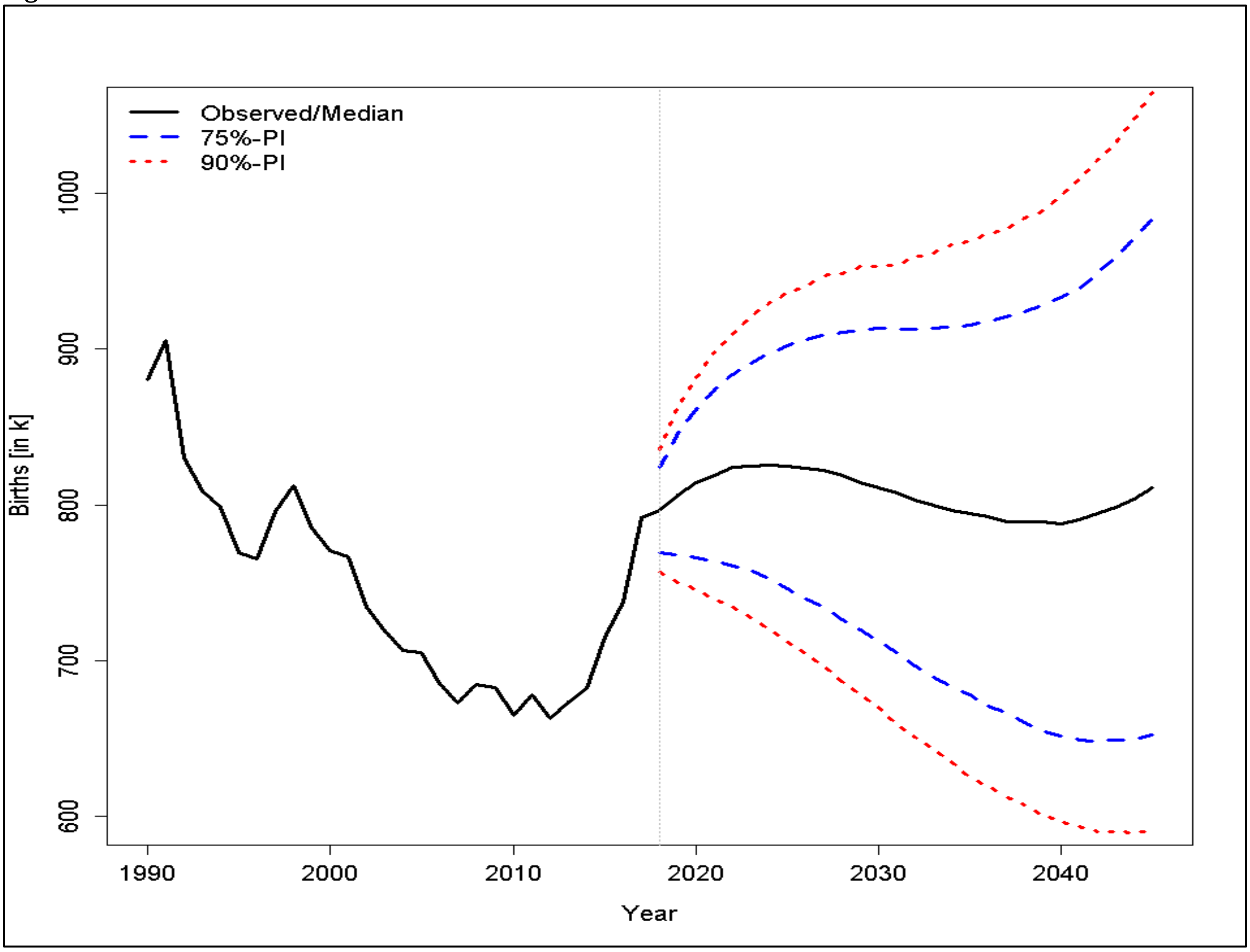
Source: Own design

4 Population Development in Germany until 2045

The combination of the resulting trajectories for the demographic components as explained in Section 3 results in a probabilistic cohort-component forecast of the age- and sex-specific population for the ages 0 – 103 years for males and 0 – 104 years for females. The respective age groups beyond these limits are treated as one age group each. The initial population for the forecast is the age- and sex-specific population estimate from the HMD for December 31, 2017 (Human Mortality Database 2018d).

In Section 3, we described the partial models for forecasting of the demographic components. Now, we provide a selection of the results from the forecast. The overview is kept short since more detailed results for the age-specific measures can be found in previous papers. The fertility model results in 10,000 trajectories for all ASFRs. By multiplication of the ASFRs with the corresponding female population, the birth forecast is completed. The results are given in Figure 10.

Figure 10. Forecast of Birth Numbers



Sources: GENESIS-Online Datenbank 2019b; Own calculation and design

The increasing trend in births, as witnessed since 2012, is expected to continue until 2024. Birth numbers will probably subsequently decrease moderately because most children are born by mothers over 29 years of age, as shown by Vanella and Deschermeier (2019: 101). This decrease can therefore be explained by the decreasing number of births at the beginning of the 1990s, as shown at the left-hand side of the graph. The median increase during the second half of the 2030s stems from a slightly increasing TFR[9] together with almost stagnating birth numbers during the cohorts 2005 to 2011, which by then will be in their reproductive phase.

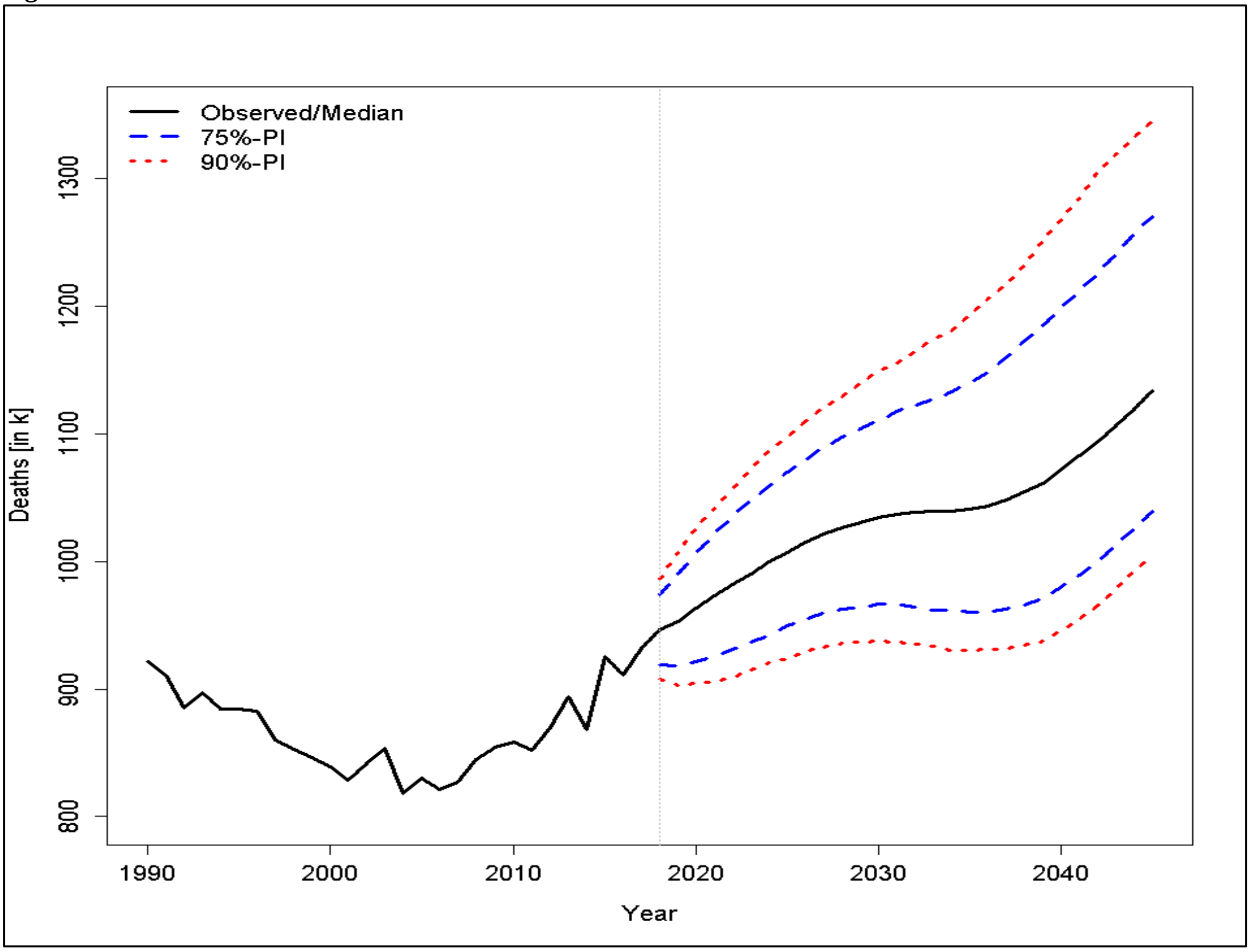
Similarly, the death numbers are derived from the ASSSRs and the population update. As shown in Figure 9, deaths can be derived by simulating the hypothetical age- and sex-specific

[9] Our model predicts a slight increase in the median TFR from its initial value of 1.56 in 2016 to 1.67 in 2045.

population at the end of some period in some trajectory without deaths and then multiplying this number with the respective adjusted ASSMR to derive the actual number of deaths among this group. The resulting death numbers are illustrated in Figure 11.

As a result of the aging of the population, a further increase in deaths, as witnessed since the mid-2000s, is probable until the end of the forecast horizon.

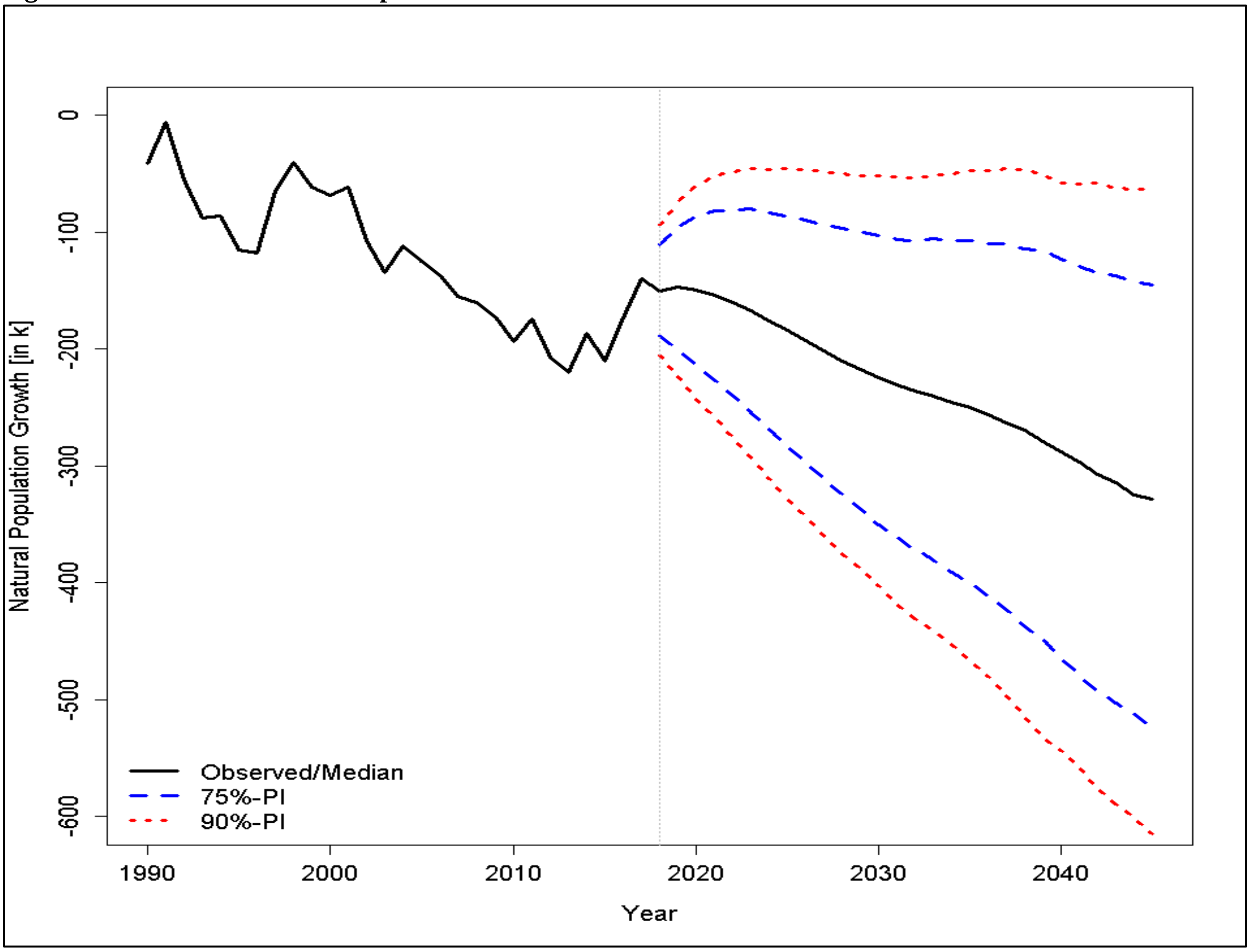
Figure 11. Forecast of Deaths



Sources: GENESIS-Online Datenbank 2019c; Own calculation and design

By subtracting the death numbers from the birth numbers, we calculate the natural population growth, whose forecast can be derived indirectly from the birth and death forecasts as well. The results of the natural population growth forecast are illustrated in Figure 12. A clear negative tendency is probable. The deaths will almost certainly exceed the births over the whole forecast horizon.

Figure 12. Forecast of Natural Population Growth

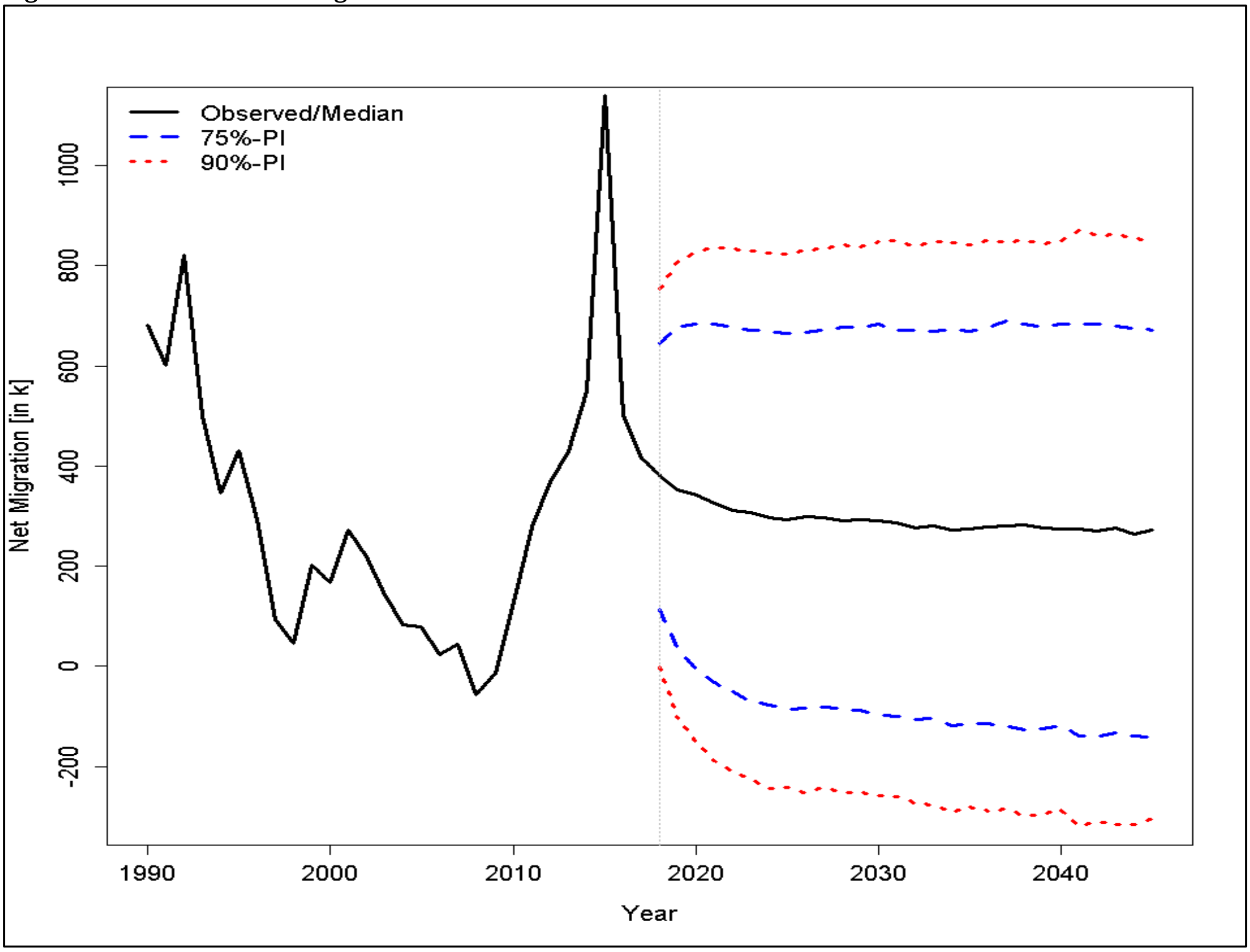


Sources: GENESIS-Online Datenbank 2019b, 2019c; Own calculation and design

Counterbalancing the shrinking population due to natural population decrease is the international net migration. The forecast method for the ASNSNM numbers has been explained in Section 3, the results of the simulation are cumulated into the total net migration for illustration purposes in Figure 13.[10]

[10] More detailed results, although based on the jump-off year 2015, can be found in Vanella and Deschermeier (2018: 274 – 276).

Figure 13. Forecast of Net Migration



Sources: GENESIS-Online Datenbank 2019d; Own calculation and design

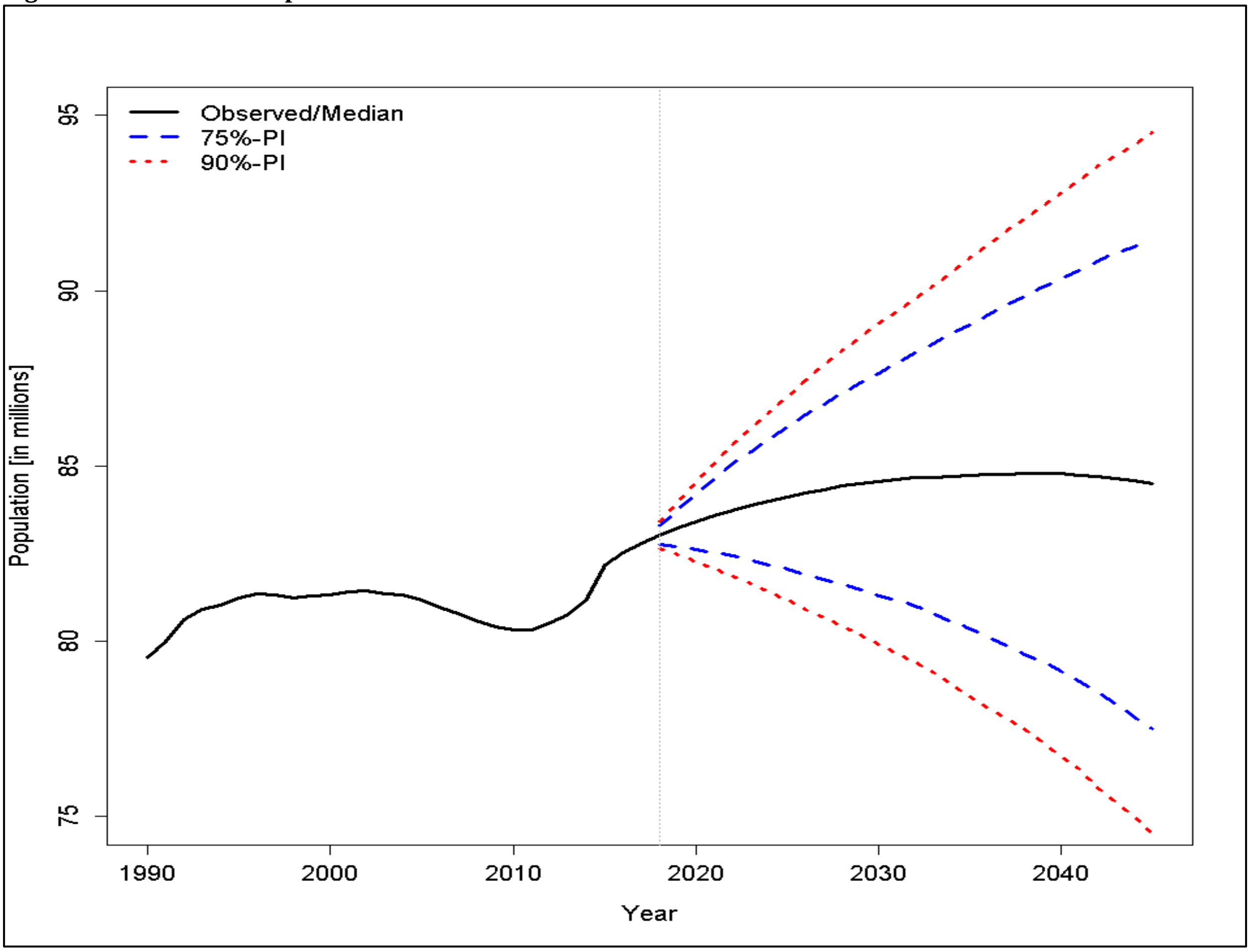
The median scenario gives a slightly decreasing net migration, whereas some cyclic course due to economic cycle is probable. In general, the high uncertainty in migration forecasting is obvious, but in general a positive net migration is very likely. The median of net migration in 2045 is 271,689 persons. This is a higher balance then most previous projections provide, that were calculated before the record influx of 2015. As many bigger cities of origin of the refugees, especially in Syria, are mostly devastated by war (McKenzie 2018; Pleitgen 2017), it seems unlikely that there will be a mass emigration out of Germany in the years to come, as one might expect due to experience from past refugee crises. Furthermore, the results reflect the strong past development of the economy in Germany. This trend is probable to remain stable in the future (OECD 2017: 130 – 133). The attractive labor market is likely to attract more people in the future (Fuchs et al. 2018: 49 – 54), especially within the EU due to the unrestricted free movement of workers (Vanella and Deschermeier 2018: 274 – 277). Moreover, some sending economies in Southern and Eastern Europe still struggle heavily economically since the financial crisis and do not appear likely to recover in the near future (World Bank 2019: 100 – 101). Total net migration in 2045 is estimated to be above zero at 77.62% probability.

The high importance of positive net migration, especially in the younger ages, shall be mentioned to fill the shortages occurring in the labor market due to overaging. We stress that the effect of migration on the labor market and the social security system very much depends on the skill level and education of the immigrants. Especially in cases of refugee migration, where education is often either relatively low or not accepted by German standards, it usually takes a long time for the immigrants to fully integrate into the labor market (Brücker et al. 2017b).

Figure 14 shows the forecast of total population until the year 2045 with 75% and 90% PIs.

In contrast to many earlier studies on Germany (see Section 2), the population is expected to increase moderately over the forecast horizon due to high, yet decreasing, net migration, an increasing TFR and decreasing mortality.

Figure 14. Forecast of Population Size

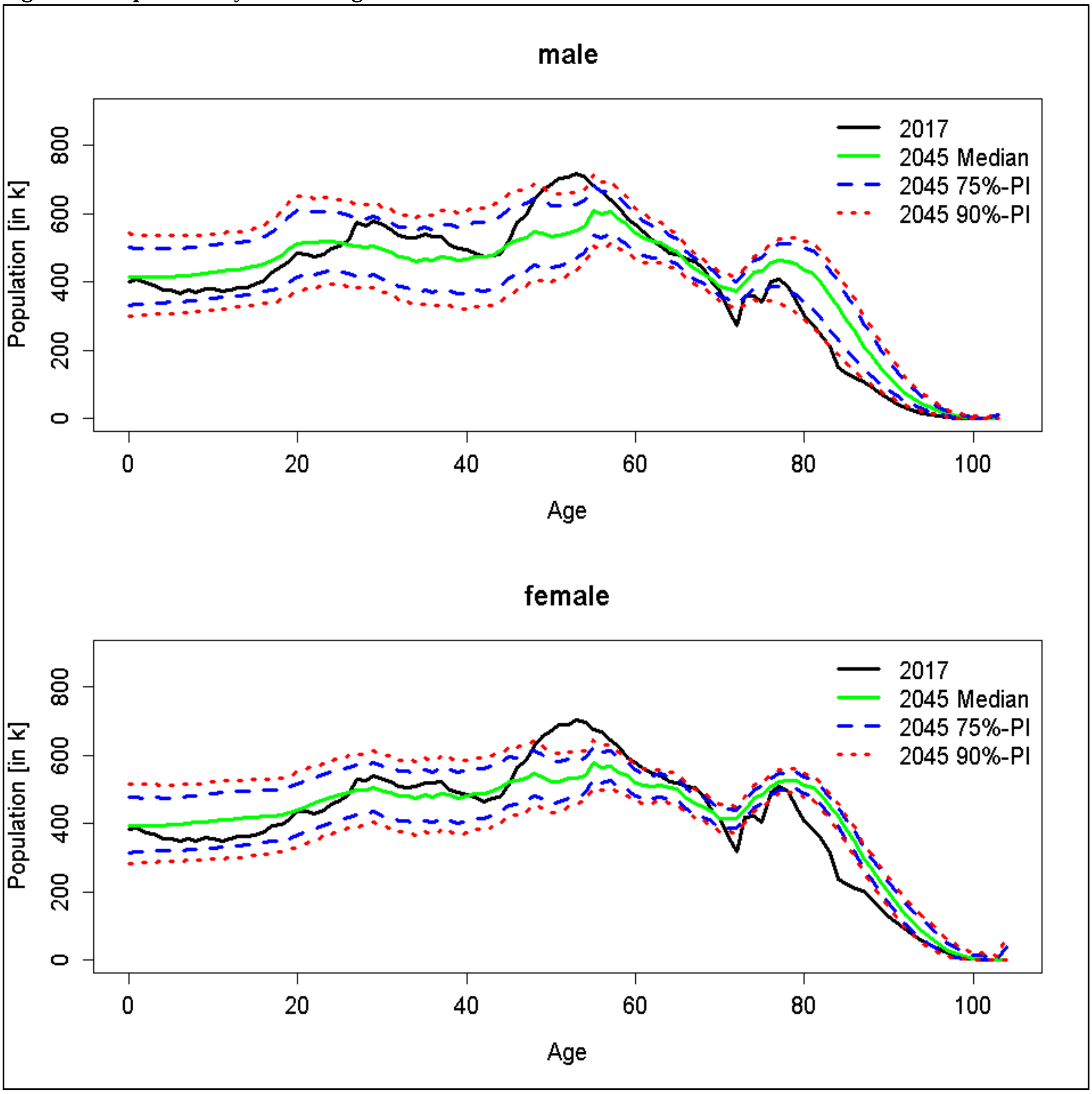


Sources: Human Mortality Database 2018d; Own calculation and design

Our median forecast of the total population is substantially above those of the other studies. Most of the presented studies were conducted before the refugee crisis since 2014 and the above-average net migration since 2010 caused by the European debt crisis. These developments mark significant changes in migration. As such remarkable events often are ignored in migration forecasts, our model considers the probability of future crises, which might lead to high migration influx, through the Crises Index explained in Section 3. The underlying assumption is that crises have a similar probability to occur in the future as they have been observed in the past data. Since our historic data cover two longer periods of international crises, the future risk should be covered appropriately by our model. According to our forecast, the ceteris paribus population in 2045 will be between 74.5 and 94.5 million people at a 90% probability level, with a median outcome of 84.5 million.

In many cases (like in social security), the structure of the population is of higher importance than its size per se. Therefore, Figure 15 gives an overview of the age structure of the population in 2017 compared to the forecast in 2045 with PIs for both sexes.

Figure 15. Population by Sex and Age in 2017 and 2045

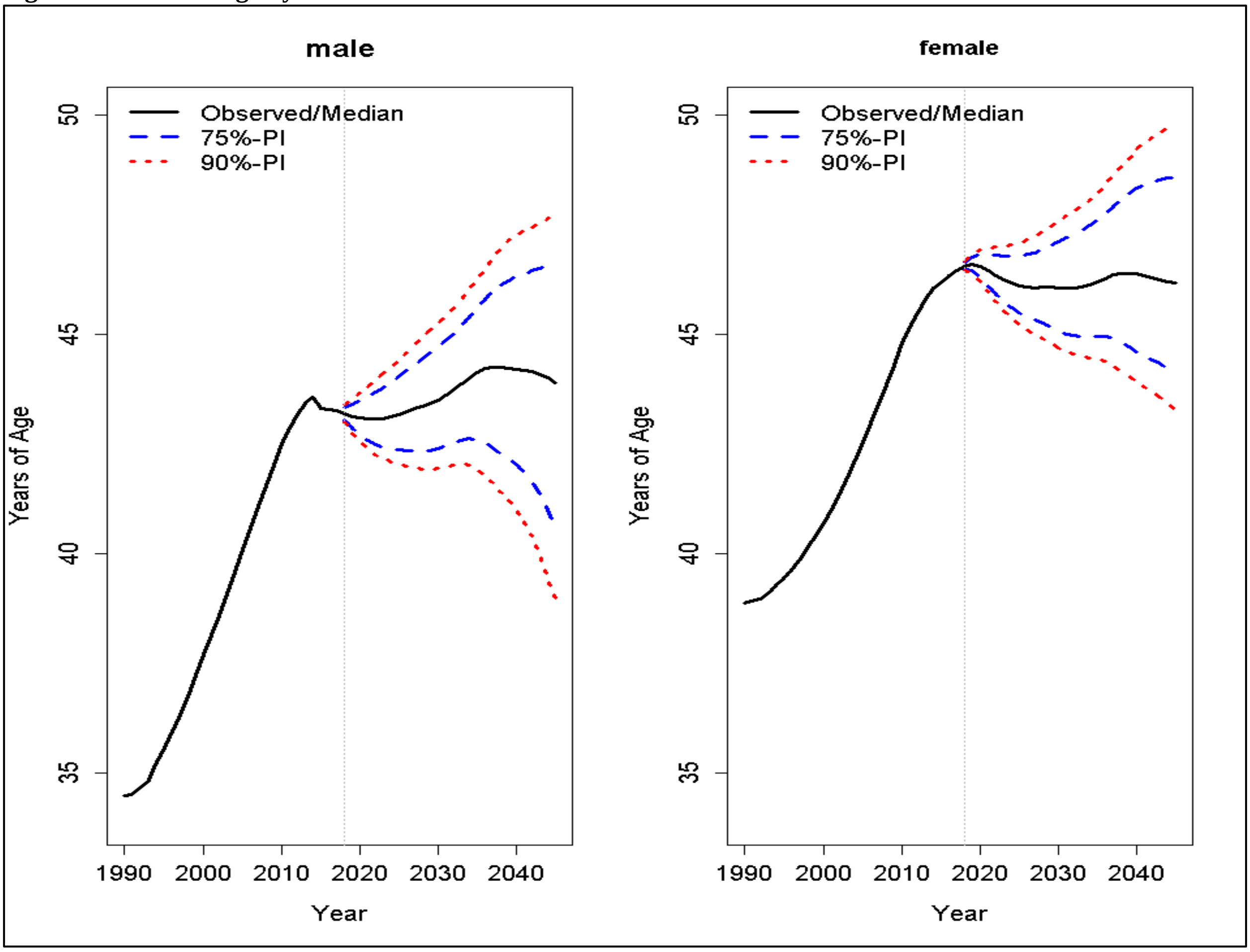


Sources: Human Mortality Database 2018d; Own calculation and design

We observe the almost 30-year shift in the population. In general, there is greater uncertainty for males. Whereas the retirement-age population can be predicted relatively well, the uncertainty in the future working-age population is rather large for males due to the higher uncertainty in the migration forecast for males relative to females (Vanella and Deschermeier 2018: 274 - 277). The uncertainty in the population of persons under 30 years of age mostly arises from the fact that this portion of the population has not yet been born but also from the relatively high uncertainty in international migration.

The aging of the population is of high social and economic importance. Therefore, in addition to the overall age structure, the median age of the male and female populations is considered as a summary indicator for the future age distribution of the population. The median age of the population can be obtained from the simulation results because it is the exact age that cuts the population in half. This computation for all 10,000 trajectories can be used to extract PIs for the median age, similar to the computation of median life span conducted by Vanella (2017: 548 – 552). The results of our analysis are shown for both sexes in Figure 16.

Figure 16. Median Age by Sex until 2045



Sources: Human Mortality Database 2018d; Own calculation and design

We observe a rejuvenation effect for the upcoming years due to the high net migration witnessed during the most recent years, as illustrated in Figure 13, and to increasing birth numbers, as shown in Figure 10. The high net migration around the year 2015 combined with the high forecast values for the upcoming years leave a mark in the age structure of Germany. This can be seen in the age structure for the male and female populations in the year 2045. By that time, the majority of the population that immigrated during the high influx phase will be approximately 55 years old, while the baby boomer generation will be in their eighth decade of life. Over the forecast horizon, the median age traces this development by a rejuvenation effect for men and women. The probable decrease in the number of births after the early 2020s and decreasing net migration and mortality (Vanella 2017: 550) lead to an aging of the population structure, as represented by the increasing median ages after that point. After the mid- to late

2030s, another phase of rejuvenation can be expected due to the increasing number of births resulting from the stronger birth cohorts after 2010 (see Figure 10), which will then enter their most fertile period, and the large increases in death numbers, as the baby boomers will witness high mortality after that period (see Figure 11).

Our model provides a wide range of detailed analyses targeting specific topics of interest, as we have shown based on certain important measures. The forecast results offer the possibility for a wide range of future studies, e.g., analyzing the effects of population changes on social security, the labor market or housing demand.

5 Conclusions, Limitations and Outlook

This paper proposed a probabilistic cohort-component approach for population forecasting by sex and age. It was applied to predict the population of Germany until the year 2045. Germany witnessed a record migration influx in 2015 due to the refugee movement, especially from Syria, Iraq and Afghanistan, in combination with the challenging economic situation in many countries in Southern and Eastern Europe. The record net migration marks a considerable event for Germany's demographic development. The expected strong long-term decrease in the population does not appear to hold based on our findings. The results provide essential data on the consequences of the current trends for decision makers, planners and scientists.

The model predicts the population of Germany by age and sex until the year 2045. The forecast is conducted as a composite of three time series models based on PCA for the three demographic components fertility, international migration and mortality by sex and age. The fertility model is conditional on political intervention as well, considering reforms in family policy to some extent. The method is specified for Germany, but it can also be applied to other countries or regional units, for which sufficiently long time series data for the demographic components are available. Stochastic modeling of the population produced point estimates of the future population in addition to a measure of the future uncertainty via prediction intervals. The results may be disaggregated or aggregated almost arbitrarily regarding sex, age and level of uncertainty.

The model is well suited for regular updating and does not require large amounts of data input since it is restricted to demographic variables and uses official statistics provided by Destatis and the HMD. One interesting result is the detailed reporting and probabilistic quantification of the disaggregated population for all ages and both sexes; therefore, the results offer many possibilities for future forecast studies that require disaggregated population data as inputs, e.g., research on social security, life insurance, the labor market or housing demand.

Our method is restricted to quantitative methods; therefore, past unobserved trends are not considered in the future. Nevertheless, for all demographic variables, the input data span at least as long of a time horizon as is forecast; thus, we believe that all realistic trends that might be observed during the time horizon are included in the model. The addition of expert knowledge would be possible, if the forecaster thinks the past trends insufficiently cover the possible future outcome. The model suffers from a relatively small input time horizon because the migration data are restricted back to the years 1990. Older data is not representative because of the overall very different geopolitical situation in Eurasia back then. Furthermore, fertility is difficult to forecast, since it is significantly influenced by policy as well. We tried to include this effect to some extent into the model as well, following a convergence assumption in family policy to avoid bias to the extent possible. Our forecast horizon is 2045 and not 2060 or 2100, as in other studies, since we do not intend to create misinterpretations for the far future, for which reliable forecasts are not possible with the available data.

A larger forecast period would be interesting but cannot be achieved via responsible statistical modeling. Thus, the future availability of input data suited for model estimation will improve the quality of our models and allow for longer forecast horizons. Even with a forecast horizon that reaches only until 2045, the uncertainty is rather large. Most of the risk stems from the uncertainty about future net migration. Although the net migration model performs reasonably well, a possible extension of the model would be to separately estimate in- and out-migration. Joint estimation of birth rates, survival rates and migration numbers (or rates in the case of out-migration) would represent another possible extension. The forecast model could theoretically discriminate by nationality for the mortality and fertility forecast, resulting in a nationality-specific population forecast. However, this would require detailed time series data on mortality and fertility by nationality alongside data on naturalizations. Moreover, the jump-off population would be needed by nationality as well. All of this information is not available.

Empirical updating might be required if the development in the upcoming years differs from our forecast due to political or economic developments. Those structural breaks are not implemented in our simulation approach.

References

Alho, J. M., & Spencer, B. D. (2005). *Statistical Demography and Forecasting.* New York: Springer Science+Business Media.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Büttner, T., & Heilig, G. K. (2011). Probabilistic Projections of the Total Fertility Rate for All Countries. *Demography,* 48(3), 815 – 839.

Aust, S., Bewarder, M., Büscher, W., Lutz, M., & Malzahn, C.C. (2015). Herbst der Kanzlerin. Geschichte eines Staatsversagens. [Autumn of the Chancellor. History of a State Failure.] *Welt*. https://www.welt.de/politik/deutschland/article148588383/Herbst-der-Kanzlerin-Geschichte-eines-Staatsversagens.html. Accessed 05 March 2019.

Azose, J. J., & Raftery, A. E. (2015). Bayesian Probabilistic Projection of International Migration. *Demography*, 52(5), 1627 – 1650.

Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), 6460 – 6465.

Bell, W. R., & Monsell, B. (1991). Using Principal Components in Time Series Modeling and Forecasting of Age-Specific Mortality Rates. In American Statistical Association (Ed.), *1991 Proceedings of the Social Statistics Section* (pp. 154 – 159). Alexandria: American Statistical Association.

Bewarder, M., & Leubecher, M. (2016). Syrische Staatsbürgerschaft wird massenhaft vorgetäuscht. [Syrian citizenship is faked massively.] *Welt*. https://www.welt.de/politik/deutschland/article156496638/Syrische-Staatsbuergerschaft-wird-massenhaft-vorgetaeuscht.html. Accessed 05 March 2019.

Bohk, C. (2012). *Ein probabilistisches Bevölkerungsprognosemodell: Entwicklung und Anwendung für Deutschland [A probabilistic population forecast model: Development and Application for Germany.]* Wiesbaden: VS Verlag für Sozialwissenschaften.

Bomsdorf, E., Babel, B., & Schmidt, R. (2008). Zur Entwicklung der Bevölkerung, der Anzahl der Schüler, der Studienanfänger und der Pflegebedürftigen: Stochastische Modellrechnungen für Deutschland bis 2050. [Development of population, number of students, newly enrolled and people in need of care: Stochastic simulations for Germany until 2050.] *Sozialer Fortschritt (German Review of Social Policy),* 57(5), 125 – 132.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting,* 22(3), 547 – 581.

Booth, H., Hyndman, R. J., Tickle, L., & de Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research,* 15(9), 289 – 310.

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

Bozik, J. E., & Bell, W. R. (1987). *Forecasting Age specific Fertility using Principal Components* (SRD Research Report Number CENSUS/SRD/RR-87/19). Retrieved from the Bureau of the Census, Statistical Research Division website: https://www.census.gov/srd/papers/pdf/rr87-19.pdf. Accessed 29 November 2018.

Brücker, H., Hauptmann, A., & Sirries, S. (2017a). *Zuzüge nach Deutschland. [Immigration into Germany.]* Retrieved from the Institut für Arbeitsmarkt und Berufsforschung website: http://doku.iab.de/aktuell/2017/aktueller_bericht_1701.pdf. Accessed 29 November 2018.

Brücker, H., Hauptmann, A., & Sirries, S. (2017b). *Arbeitsmarktintegration von Geflüchteten in Deutschland: Der Stand zum Jahresbeginn 2017. [Labor market integration of refugees in Germany: The state at the beginning of the year 2017.]* Retrieved from the Institut für Arbeitsmarkt und Berufsforschung website: http://doku.iab.de/aktuell/2017/aktueller_bericht_1704.pdf. Accessed 19 May 2018.

Bundesamt für Migration und Flüchtlinge (2016). *Das Bundesamt in Zahlen 2015: Asyl, Migration und Integration. [The Federal Office in Numbers 2015: Asylum, Migration and Integration.]* Retrieved from the BAMF website: https://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Broschueren/bundesamt-in-zahlen-2015.pdf?__blob=publicationFile. Accessed 17 May 2018.

Bundesministerium des Innern (2011). *Migrationsbericht 2009. [Migration Report 2009.]* Retrieved from the BAMF website: https://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Migrationsberichte/migrationsbericht-2009.pdf?__blob=publicationFile. Accessed 05 March 2019.

Cannan, E. (1895). The Probability of a Cessation of the Growth of Population in England and Wales During the Next Century. *The Economic Journal,* 5(20), 505 – 515.

Census Bureau of England and Wales (1863). *General Report; with appendix of tables.* Retrieved from the A vision of Britain through Time website: http://www.visionofbritain.org.uk/census/EW1861GEN/5. Accessed 05 March 2019.

Council on Foreign Relations (2012). Islamist Extremism After the Arab Spring. *Interview by Jonathan Masters with Ali Soufan.* https://www.cfr.org/interview/islamist-extremism-after-arab-spring. Accessed 05 March 2019.

Davis, D. L., Webster, P., Stainthorpe, H., Chilton, J., Jones, L., & Doi, R.. (2007). Declines in Sex Ratio at Birth and Fetal Deaths in Japan, and in U.S. Whites but Not African Americans. *Environmental Health Perspectives,* 115(6), 941 – 946.

Deschermeier, P. (2011). Die Bevölkerungsentwicklung der Metropolregion Rhein-Neckar: Eine stochastische Bevölkerungsprognose auf Basis des Paradigmas funktionaler Daten. [Population Development of the Rhine-Neckar Metropolitan Area: A Stochastic Population Forecast on the Basis of Functional Data Analysis.] *Comparative Population Studies,* 36(4), 769 – 806.

Deschermeier, P. (2015). Die Entwicklung der Bevölkerung Deutschlands bis 2030 – ein Methodenvergleich. [Population Development in Germany to 2030: A Comparison of Methods.] *IW Trends – Vierteljahresschrift zur empirischen Wirtschaftsforschung,* 42(2), 97 – 111.

Deschermeier, P. (2016). Einfluss der Zuwanderung auf die demografische Entwicklung in Deutschland. [The Influence of Immigration on Demographic Developments in Germany.] *IW Trends – Vierteljahresschrift zur empirischen Wirtschaftsforschung,* 43(2), 21 – 38.

Destatis (2007). *Lebendgeborene nach dem Alter der Mutter 1990: Neue Länder und Berlin-Ost. [Live Births by Mother's Age 1990: New Federal States and Eastern Berlin.]* Data provided on 13 February 2018.

Destatis (2014a). *Lebendgeborene nach dem Alter der Mütter 1968 – 1990: Alte Bundesländer. [Live Births by Mothers' Age 1968 – 1990: Old Federal States.]* Data provided on 18 June 2015.

Destatis (2014b). *Lebendgeborene nach dem Alter der Mütter 1991 – 2013: Deutschland. [Live Births by Mothers' Age 1991 – 2013: Germany.]* Data provided on 18 June 2015.

Destatis (2015). *Wanderungen zwischen Deutschland und dem Ausland 1991 – 2013 nach Einzelaltersjahren und Geschlecht. [Migration between Germany and abroad 1991 – 2013 by Years of Age and Sex.]* Data provided on 02 September 2015.

Destatis (2016). *Wanderungen zwischen Deutschland und dem Ausland 2014 nach Einzelaltersjahren und Geschlecht. [Migration between Germany and abroad 2014 by Years of Age and Sex.]* Data provided on 13 March 2016.

Destatis (2017a). *Wanderungen zwischen Deutschland und dem Ausland 2015 nach Einzelaltersjahren und Geschlecht. [Migration between Germany and abroad 2014 by Years of Age and Sex.]* Data provided on 23 April 2017.

Destatis (2017b). *Wanderungen über die Grenzen Deutschlands 1990 – 2015 nach ausgewählten Staatsangehörigkeiten/ der Staatsangehörigkeit und Altersgruppen. [German Cross-Border Migration 1990 – 2015 by selected Citizenships / Citizenship and Groups of Age.]* Data provided on 18 April 2017.

Destatis (2017c). *Gestorbene 2000-2015 nach Alters- und Geburtsjahren. [Deaths 2000 – 2015 by Years of Age and Birth.]* Data provided on 15 August 2017.

Destatis (2018a). *Wanderungen zwischen Deutschland und dem Ausland 2016 nach Einzelaltersjahren und Geschlecht. [Migration between Germany and abroad 2016 by Years of Age and Sex.]* Data provided on 13 March 2018.

Destatis (2018b). *Wanderungen über die Grenzen Deutschlands 2016 nach der Staatsangehörigkeit und Altersgruppen. [Migration over the German Borders 2016 by Citizenship and Groups of Age.]* Data provided on 25 April 2018.

Destatis (2018c). *Gestorbene 2016 nach Alters- und Geburtsjahren. [Deaths 2016 by Years of Age and Birth.]* Data provided on 29 March 2018.

Destatis (2018d). *Lebendgeborene nach dem Alter der Mutter 1968 – 1969: DDR. [Live Births by Mother's Age 1968 – 1969: GDR]* Data provided on 08 February 2018.

Destatis (2018e). *Lebendgeborene nach dem Alter der Mutter 1970 – 1989: DDR. [Live Births by Mother's Age 1970 – 1989: GDR]* Data provided on 09 February 2018.

Destatis (2018f). *Wanderungen zwischen Deutschland und dem Ausland 2017 nach Einzelaltersjahren und Geschlecht. [Migration between Germany and abroad 2017 by Years of Age and Sex.]* Data provided on 12 April 2019.

Destatis (2018g). *Wanderungen über die Grenzen Deutschlands 2017 nach der Staatsangehörigkeit und Altersgruppen. [Migration over the German Borders 2017 by Citizenship and Groups of Age.]* Data provided on 12 April 2019.

Destatis (2019a). *Gestorbene 2017 nach Alters- und Geburtsjahren. [Deaths 2017 by Years of Age and Birth.]* Data provided on 19 March 2019.

Destatis (2019b). *Bevölkerung im Wandel. Annahmen und Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung [Population in Transition. Assumptions and Results of the 14th Coordinated Population Projection.]* . Retrieved from the Destatis website: https://www.destatis.de/DE/Presse/Pressekonferenzen/2019/Bevoelkerung/pressebroschuere-bevoelkerung.pdf?__blob=publicationFile. Accessed 13 July 2019.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427 – 431.

Dudel, C. (2014). *Vorausberechnung von Verwandtschaft: Wie sich die gemeinsame Lebenszeit von Kindern, Eltern und Großeltern zukünftig entwickelt. [Forecasting of Relationship: How joint Lifetime of Children, Parents and Grandparents develops in the Future.]* Opladen, Berlin, Toronto: Verlag Barbara Budrich.

European Union (2015). *The 2015 Aging Report: Economic and budgetary projections for the 28 EU Member States (2013 – 2060). European Economy 3/2015.* Luxembourg: Publications Office of the European Union.

Fuchs, J., Söhnlein, D., Weber, B., & Weber, E. (2018). Stochastic Forecasting Labor Supply and Population: An Integrated Model. *Population Research and Policy Review*, 37(1), 33 – 58.

Gauthier, A.H., Hatzius, J. (1997). Family Benefits and Fertility: An Econometric Analysis. *Population Studies*, 51(3), 295 – 306.

GENESIS-Online Datenbank (2018). *Durchschnittliches Alter der Mutter bei Geburt ihrer lebend geborenen Kinder: Deutschland, Jahre, Familienstand der Eltern. [Average Mother's Age by Live Birth: Germany, Years, Parents' Marital Status.]* https://www-genesis.destatis.de/genesis/online/. Accessed 07 March 2018.

GENESIS-Online Datenbank (2019a). *Lebendgeborene: Deutschland, Jahre, Alter der Mutter, Geschlecht der Lebendgeborenen, Familienstand der Eltern. [Live Births: Germany, Years, Mother's Age, Sex of Newborn, Parents' Marital Status.]* https://www-genesis.destatis.de/genesis/online/. Accessed 09 July 2019.

GENESIS-Online Datenbank (2019b). *Lebendgeborene: Deutschland, Jahre, Geschlecht. [Live Births: Germany, Years, Sex.]* https://www-genesis.destatis.de/genesis/online/. Accessed 15 July 2019.

GENESIS-Online Datenbank (2019c). *Gestorbene: Deutschland, Jahre, Geschlecht. [Deaths: Germany, Years, Sex.]* https://www-genesis.destatis.de/genesis/online/. Accessed 15 July 2019.

GENESIS-Online Datenbank (2019d). *Wanderungen zwischen Deutschland und dem Ausland: Deutschland, Jahre, Kontinente. [Migration between Germany and abroad: Germany, Years, Continents.]* https://www-genesis.destatis.de/genesis/online/. Accessed 15 July 2019.

Härdle, W., & Myšičková, A. (2009). *Stochastic Population Forecast for Germany and its Consequence for the German Pension System* (SFB 649 Discussion Paper 2009-009). Berlin: Humboldt-Universität zu Berlin.

Heidelberg Institute for International Conflict Research (2017). *Conflict Barometer 2016.* Heidelberg, DE: Universität Heidelberg, Heidelberg Institute for International Conflict.

Helle, S., Helama, S., & Jokela, J. (2008). Temperature-related birth sex ratio bias in historical Sami: warm years bring more sons. *Biological Letters,* 4(1), 60 – 62.

Hertrampf, S. (2008). Ein Tomatenwurf und seine Folgen: Eine neue Welle des Frauenprotestes in der BRD. [A tossed Tomato and its Results: A new Wave of Female Protest in the Federal Republic of Germany.] *Bundeszentrale für politische Bildung*Retrieved from http://www.bpb.de/gesellschaft/gender/frauenbewegung/35287/neue-welle-im-westen. Accessed 05 March 2019.

Human Mortality Database (2018a). *Germany, Deaths (Lexis triangle).* Provided by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 09 July 2019.

Human Mortality Database (2018b). *West Germany, Deaths (Lexis triangle).* Provided by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 09 July 2019.

Human Mortality Database (2018c). *East Germany, Deaths (Lexis triangle).* Provided by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 09 July 2019.

Human Mortality Database (2018d). *Germany, Population size (abridged).* Provided by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 09 July 2019.

Human Mortality Database (2018e). *West Germany, Population size (abridged).* Provided by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 09 July 2019.

Human Mortality Database (2018f). *East Germany, Population size (abridged).* Provided by University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 09 July 2019.

Hyndman, R. J., Booth, H., & Yasmeen, F. (2013). Coherent Mortality Forecasting: The Product-Ratio Method With Functional Time Series Models. *Demography,* 50(1), 261 – 283.

Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis,* 51(10), 4942 – 4956.

International Organization for Migration (2014). *Afghanistan: Migration Profile*. Kabul: International Organization for Migration.

James, W. H. (2000). Secular movements in sex ratios of adults and of births in populations during the past half-century. *Human Reproduction,* 15(5), 1178 – 1183.

Kalwij, A. (2010). The Impact of Family Policy Expenditure on Fertility in Western Europe. *Demography*, 47(2), 503 – 519.

Keilman, N., Dinh, Q. P., & Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research,* 6(15), 409 – 454.

Klüsener, S., Grigoriev, P., Scholz, R. D., & Jdanov, D. A. (2018). Adjusting Inter-censal Population Estimates for Germany 1987-2011: Approaches and Impact on Demographic Indicators. *Comparative Population Studies,* 43, 31 – 64.

Le Bras, H., & Tapinos, G. (1979). Perspectives à long terme de la population française et leurs implications économiques. [Long-term outlook of the French population and its economic implications.] *Population*, 34(1), 1391 – 1452.

Ledermann, S., & Breas, J. (1959). Les dimensions de la mortalité. [The dimensions of mortality.] *Population*, 14(4), 637 – 682.

Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, 9(2), 187 – 202.

Lee, R. D. (1998). Probabilistic Approaches to Population Forecasting. *Population and Development Review,* 24(Supplement: Frontiers of Population Forecasting), 156 – 190.

Lee, R. D., & Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), 659 – 671.

Lipps, O., & Betz, F. (2005). Stochastische Bevölkerungsprojektion für West- und Ostdeutschland. [Stochastic Population Projection for Western and Eastern Germany.] *Zeitschrift für Bevölkerungswissenschaft (Comparative Population Studies),* 30(1), 3 – 42.

Lutz, W., & Scherbov, S. (1998). Probabilistische Bevölkerungsprognosen für Deutschland. [Probabilistic Population Forecasts for Germany.] *Zeitschrift für Bevölkerungswissenschaft (Comparative Population Studies),* 23(2), 83 – 109.

Luy, M., & Di Giulio, P. (2006). *The impact of health behaviors and life quality on gender differences in mortality.* (MPIDR Working Paper WP 2006-035). Rostock: Max Planck Institute for Demographic Research.

Mathews, F., Johnson, P. J., & Neil, A. (2008). You are what your mother eats: evidence for maternal preconception diet influencing foetal sex in humans. *Proceedings of the Royal Society B: Biological Sciences,* 275(1643), 1661 – 1668.

McKenzie, S. (2018). How seven years of war turned Syria's cities into 'hell on Earth'. *CNN.* https://edition.cnn.com/2018/03/15/middleeast/syria-then-now-satellite-intl/index.html. Accessed 05 March 2019.

OECD (2017). *OECD-Wirtschaftsausblick 2017/2. [OECD Economic Outlook 2017/2.]* Paris: OECD Publishing.

Pampel, F. (2005). Forecasting sex differences in mortality in high income nations: the contribution of smoking. *Demographic Research,* 13(18), 455 – 484.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine,* Series 6(2:11), 559 – 572.

Pleitgen, F. (2017). Inside Aleppo: Residents return to rebuild Syria's shattered city. *CNN.* https://edition.cnn.com/2017/08/15/middleeast/aleppo-syria-rebuilds/index.html. Accessed 05 March 2019.

Pötzsch, O. (2016). (Un-)Sicherheiten der Bevölkerungsvorausberechnungen: Rückblick auf die koordinierten Bevölkerungsvorausberechnungen für Deutschland zwischen 1998 und 2015. [(Un-)Certainties about Population Projections: Review of the Coordinated Population Projections for Germany between 1998 and 2015.] *WISTA – Wirtschaft und Statistik,* 4/2016, 36 – 54.

Pötzsch, O., & Rößger, F. (2015). *Germany's Population by 2060: Results of the 13th coordinated population projection.* Retrieved from the Destatis website: https://www.destatis.de/EN/Publications/Specialized/Population/GermanyPopulation2060_5124206159004.pdf?__blob=publicationFile.

Raftery, A. E., Alkema, L., & Gerland, P. (2014). Bayesian Population Projections for the United Nations. *Statistical Science,* 29(1), 58 – 68.

Raftery, A.E., Chunn, J. L., Gerland, P., & Ševčíková, H. (2013). Bayesian Probabilistic Projections of Life Expectancy for All Countries. *Demography*, 50(3), 777 – 801.

Raftery, A. E., Lalić, N., & Gerland, P. (2014). Joint probabilistic projection of female and male life expectancy. *Demographic Research,* 30(27), 795 – 822.

Schnabel, S., von Kistowski, K., & Vaupel, J. W. (2005). Immer neue Rekorde und kein Ende in Sicht: der Blick in die Zukunft lässt Deutschland grauer aussehen als viele erwarten. [Ever new Records and no End in Sight: A Look in the Future makes Germany look more gray than many expect.] *Demografische Forschung Aus Erster Hand,* 2(2), 3.

Scholz, R. D., Jdanov, D. A., Kibele, E., Grigoriev, P., & Klüsener, S. (2018). About Mortality Data for Germany. Human Mortality Database. https://www.mortality.org/hmd/DEUTNP/InputDB/DEUTNPcom.pdf. Accessed 09 July 2019.

Trovato, F., & Lalu, N. M. (1996). Narrowing Sex Differentials in Life Expectancy in the Industrialized World: Early 1970's and Early 1990's. *Biodemography and Social Biology,* 43(1), 20 – 37.

United Nations (2015). *World Population Prospects: The 2015 Revision: Methodology of the United Nations Population Estimates and Projections* (Working Paper No. ESA/P/WP.242). New York: United Nations, Department of Economic and Social Affairs, Population Division.

United Nations (2017). *Probabilistic projection of total population (both sexes combined) by region, subregion, country or area, 2015-2100 (thousands).* Retrieved from https://esa.un.org/unpd/wpp/Download/Probabilistic/Population/. Accessed 11 September 2017.

Vanella, P. (2017). A Principal Component Model for Forecasting Age- and Sex-Specific Survival Probabilities in Western Europe. *Zeitschrift für die gesamte Versicherungswissenschaft (German Journal of Risk and Insurance),* 106(5), 539 – 554.

Vanella, P. (2018). Stochastic Forecasting of Demographic Components Based on Principal Component Analyses. *Athens Journal of Sciences,* 5(3), 223 – 246.

Vanella, P., & Deschermeier, P. (2018). A stochastic Forecasting Model of international Migration in Germany. In O. Kapella, N. F. Schneider, & H. Rost (Eds.), *Familie – Bildung – Migration: Familienforschung im Spannungsfeld zwischen Wissenschaft, Politik und Praxis. Tagungsband zum 5. Europäischen Fachkongress Familienforschung* (pp. 261 – 280). Opladen, Berlin, Toronto: Verlag Barbara Budrich.

Vanella, P., & Deschermeier, P. (2019). A Principal Component Simulation of Age-Specific Fertility – Impacts of Family Policy on Reproductive Behavior in Germany. *Population Review*, 58(1), 78 – 109.

Waldron, I. (1993). Recent trends in sex mortality ratios for adults in developed countries. *Social Science & Medicine,* 36(4), 451 – 462.

Whelpton, P. K. (1928). Population of the United States, 1925 to 1975. *The American Journal of Sociology,* 34(2), 253 – 270.

World Bank (2019). *Global Economic Prospects. Heightened Tensions, Subdued Investment.* Washington, DC: The World Bank.

World Health Organization (2015). *World report on ageing and health.* Geneva: World Health Organization.

Zeit Online (2015). Deutschland setzt Dublin-Verfahren für Syrer aus. [Germany suspends Dublin procedure for Syrians.] *Zeit*. https://www.zeit.de/politik/ausland/2015-08/fluechtlinge-dublin-eu-asyl. Accessed 05 March 2019.