Comparative Population Studies – Zeitschrift für Bevölkerungswissenschaft
Vol. 36, 4 (2011): 769-806 (Date of release: 01.11.2012)

Population Development of the Rhine-Neckar Metropolitan Area:

A Stochastic Population Forecast on the Basis of Functional Data Analysis

Philipp Deschermeier

Abstract: Stochastic population forecasts are gaining popularity in these times of demographic change, as compared with the scenario technique frequently used for projections, they provide important additional information: the forecasted population lies within a prediction interval to which a probability of occurrence can be allocated. However, this approach requires long time-series and detailed information about the determinants of population development (fertility, mortality and net migration), which are frequently not available in sufficient depth at the regional level, but are generally subsumed into age groups. Stochastic population forecasts are therefore usually limited to the national level. Nonetheless, methods of functional data analysis enable us to disaggregate the required demographic variables into years of age and to use them as the data basis of a stochastic model, also at the regional level. This essay presents this approach and models based on it using the example of the population development of the Rhine-Neckar metropolitan area in Germany.

Keywords: Rhine-Neckar metropolitan area · Stochastic population forecasts · Functional data analysis

1 Introduction

Demographic change is dramatically transforming the German society: the ratio of young to old and of the gainfully employed to pensioners is shifting in favour of older persons. Both public debate and academic research in Europe are largely focused on the socio-political consequences of demographic change. In particular, the focus is on the problem of financing the statutory pension insurance, which led to the negative sound of the term "aging population". Frequently, the positive aspect of demographic change is forgotten: people have a higher life expectancy and a longer active lifespan (*Schnabel et al.* 2005: 3).

URL: www.comparativepopulationstudies.de
DOI: 10.4232/10.CPoS-2011-21en
URN: urn:nbn:de:bib-cpos-2011-21en4

Rising life expectancy, particularly the above-average drop in the mortality of older people, along with continuous low fertility, brings about far-reaching changes in the age structure. Although at the national level, only a slight decrease of the population is forecasted in the coming 20 years (from 81 million to approx. 77 to 79 million inhabitants), the number of people of working age will drop considerably (*Statistisches Bundesamt* 2009; *Fuchs/Dörfler* 2005a/b; *Börsch-Supan/Wilke* 2009). Hence, in the coming decades, society will be faced with a substantial macro-economic structural transformation affecting all of the important markets: the labour market will lack young workers, the product markets will have to adjust to structurally changed consumer demands and on the capital market, savings behaviour and the demand for productive investments will be changing (*Börsch-Supan* 2007). At the regional level, numbers and structure of future population development diverge to a high degree due to different economic conditions, so that the consequences for demand of goods, the housing market and the supply of workers deviating from region to region will result in a growing need for regional population forecasts as quantitative decision-making bases for regional planning. In particular, stochastic models using methods from time series analysis are becoming increasingly popular. However, these approaches require very long time series for various demographic variables, such as mortality and information on age and gender-differentiated structure, which are frequently not available at the regional level. Therefore, stochastic models can usually only be conducted at the national level. Furthermore, the data required for a population forecast are frequently processed differently by the different statistical offices of the federal states. For instance, the structure of age-specific variables of the federal states deviates, making population development difficult to model, especially for polycentric metropolitan areas which are located across federal state borders.

This article provides a stochastic forecast of the population of the Rhine-Neckar metropolitan area broken down by years of age (up to 90+ years) and gender from 2010 until 2030. This region is situated at the intersection of the three federal states of Baden-Wuerttemberg, Hesse and Rhineland-Palatinate and is a growing economic region. On the one hand, its socio-cultural connections reach far back into the past, when the region formed a territorial unit called the "Electoral Palatinate". On the other hand, since the 1970s economic ties have developed across federal state borders, which are associated with the initiative of the "Rhine-Neckar Triangle" (*Lowack* 2007: 132). The three centrally located regional metropolises of Mannheim, Heidelberg and Ludwigshafen shape the polycentric structure of the region, which, with about 2.4 million inhabitants, is Germany's seventh-largest economic region. The area stretches across 15 rural and urban districts. Metropolitan areas are considered "engines of economic development" (translated by CPoS, *Bundesamt für Bauwesen, Raumordnung und Städtebau* 1995), from which stimuli should emanate to structurally weaker regions. However, due to the future decline in population, they compete with one another, particularly for highly skilled workers (*Klein* 2008: 44). This "war for talent" (translated by CPoS, *Chambers et al.* 1998: 1) illustrates the significance of regional population forecasts that help to identify future bottlenecks early and enable a suitable response, for example through educational and advanced training programmes (*Gans et al.* 2009: 118). The Rhine-Neckar metropolitan area is integrating the future challenges in the regional strategy "Demographic Change" via a network of regional decision makers, the "Lenkungskreis Demographie" (translated by CPoS, *Metropolregion Rhein-Neckar* 2011: 26).

The population forecast presented in the following paper uses functional data for an in-depth analysis of the demographic development of the Rhine-Neckar metropolitan area. This approach is based on the idea that the secondary statistical data basis of the demographic variables individually aggregated to age groups follows a measurable functional context. The objective of this study is to estimate these and utilise them for forecasts. In this way, the differently structured and broken down data sets of the detailed demographic variables of the statistical offices of the federal states of Baden-Wuerttemberg, Hesse and Rhineland-Palatinate can be smoothed to one study region and aggregated to a total region. Time series models are based on functional data analysis (Chapter 3), which reproduce the uncertainty of demographic development through confidence intervals more precisely than the frequently used scenario technique. Chapter 4 illustrates the advantages of this approach with the example of the Rhine-Neckar metropolitan area, and Chapter 5 concludes with an outlook.

2 Population forecasts

It is difficult to foresee the demographic future of a region, as there are very many possible courses it could take. The literature differentiates two methodological approaches for dealing with this challenge. The traditional approaches are deterministic, and although they are frequently used by official statisticians, they portray the development of the population without precise information about the uncertainty of future developments (*Lipps/Betz* 2004: 1). One possibility for this is the use of mathematical extrapolation methods, which analyse the trends of the past and update them using growth functions. The use of growth rates might be suitable for large-scale projections such as that of the global population (*O'Neill et al.* 2001: 207). It is, however, less suitable for modelling at the regional level, where development is marked by a greater dynamic, particularly through migration movements. In order to take these aspects into consideration, conventional deterministic methods use the basic demographic equation as a basis (*Bähr et al.* 1992: 327):

$$B_t = B_{t-1} + G_{t-1,t} - S_{t-1,t} + M_{t-1,t}$$

whereby B_t = population at point in time t;
B_{t-1} = population at point in time t-1;
$G_{t-1,t}$ = births, between t-1 and t;
$S_{t-1,t}$ = deaths, between t-1 and t, as well as
$M_{t-1,t}$ = net migration between t-1 and t.

A possible course of the future population is calculated from an initial population and assumptions about the demographic determinants (fertility, mortality and net migration) (*Statistisches Bundesamt* 2009: 9). The basic demographic equation differentiates various points in time, but views the population only in the aggregation. In order to make the diverse changes in demographic change visible, the basic equation must be differentiated according to age (x = 0, 1, 2, …, k-1, k) and gender (M=men and F=women) of the population. The births and deaths of the individual age groups result from the product of the underlying demographic rates with the population of the previous year. The demographic development of a region is therefore determined by five components: the fertility rate, the male and female mortality rate as well as the gender-differentiated net migration. The cohort component method formalises female population development (index "F") into a matrix model (*Lee/Tuljapurkar* 1994: 1178):

$$B_{t+1}^F = \Omega_t^F B_t^F + M_{t+1,t} \ ,$$

respectively (*Pflaumer* 1988: 136):

$$\begin{pmatrix} B_{0,t+1}^F \\ B_{1,t+1}^F \\ B_{2,t+1}^F \\ \vdots \\ \\ \vdots \\ B_{k,t+1}^F \end{pmatrix} = \begin{pmatrix} 0 & \cdots & vf_{15,t}\alpha_{*,t}^F & \cdots & vf_{49,t}\alpha_{*,t}^F & 0 & \cdots & 0 \\ \alpha_{o,t}^F & 0 & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ 0 & \alpha_{1,t}^F & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & & & & \vdots \\ \\ \vdots & \vdots & & & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & \alpha_{k-1,t}^F & 0 \end{pmatrix} \begin{pmatrix} B_{0,t}^F \\ B_{1,t}^F \\ B_{2,t}^F \\ \vdots \\ \\ \vdots \\ B_{k,t}^F \end{pmatrix} + \begin{pmatrix} M_{0,t}^F \\ M_{1,t}^F \\ M_{2,t}^F \\ \vdots \\ \\ \vdots \\ M_{k,t}^F \end{pmatrix} .$$

The (female) population of a particular year of age x of the ensuing period t+1 is calculated from the sum product of a line of matrix Ω (mortality rate $m_{x,t}$ or fertility rates $f_{x,t}$) with the column vector of the population at point in time t. Parameter v quantifies the percentage of newborn females and $\alpha_{x,t}$ the probability of survival, defined as (1-$m_{x,t}$). This product supplies a vector for natural population development. The number of newborns is calculated from the sum of women of childbearing age multiplied with the fertility rates. This value is distributed over the two genders and then multiplied by the probability of surviving the first six months of life ($\alpha_{*,t}$). The population at time t+1 is the result of adding the vector for natural population development to the vector of net migration.

Since the cohort component method examines the individual components of the basic demographic equation separately and differentiates cohorts (usually years of age), it can precisely indicate those influences that affect only certain age ranges. This approach was used as early as the 1920s by *Whelpton* (1928) to forecast the population of the United States from 1925 until 1975 (*Bähr et al.* 1992: 500). Today, this is the standard method for models of population development and is used by the statistical offices (for example, *Statistisches Bundesamt* 2009). In order to take the uncertainty of the demographic future into account, different progressions are calculated using different assumptions about the demographic rates. These scenarios (*Lutz et al.* 1998a: 140) usually include an optimistic, a neutral and a pessimistic trend. When interpreting these scenarios, a frequent error is to interpret the range of variation between the pessimistic and optimistic progression as a measure of uncertainty (*Keyfitz* 1972: 353). Combining different assumptions for regional forecasts seems unsuitable, since future migration movements in particular have a major impact on the population structure, yet are afflicted with very great uncertainty.[1]

Although the official bodies have been using the deterministic approach for years, it is not unproblematic (*Lee* 1999; *Lipps/Betz* 2003; *Keilman et al.* 2002). With the exception of the range of possible developments, the deterministic approach cannot say anything about the occurrence of the different scenarios. Stochastic approaches, by contrast, enable forecasts that cite the probability with which the result lies within a certain range of variation. This is valuable additional information, for example for policy consulting. Regional planners would like to know, for instance, which of the scenarios is most likely to occur, in order to avoid "unpleasant surprises" and to draw up more concrete plans. Furthermore, the assumption of fixed demographic components contradicts the use of scenarios. For example, the number of live births of a woman is frequently set at 1.4 in Germany, whereby the future values are completely determined by the previous periods. *Lipps* and *Betz* (2003: 4-5) use a simple arithmetic example to show that the development of a demographic component assumed as fixed (perfect autocorrelation) with a time frame of 20 years deviates considerably from a forecast based on time series models, in spite of marginal differences in the individual periods. Furthermore, perfect autocorrelation between the demographic variables also results from the combination of the assumptions made. A possible scenario of "high population growth" arises from the combination of high fertility and low mortality or high life expectancy. In this scenario, a high birth rate implies high life expectancy (*Keilman et al.* 2002: 410). This autocorrelation between the demographic variables produces inconsistencies, as extreme assumptions about one variable need not necessarily result in extreme assumptions about another variable. A population forecast should therefore always demonstrate two components (*Keilman et al.* 2002: 410):

1. a range of possible development and
2. an a probability of occurrence for this range.

In addition, a forecast should be able to take four types of correlation into account: over time, between the different age groups, between the genders and between the demographic components.[2] However, a possible correlation between

---

1 In addition, interested readers can find a detailed portrayal and discussion of various approaches for regional population forecasts in *Rogers* (1985).

2 *Keilman et al.* (2002: 412-414) contains a detailed discussion of the different types of correlation.

the different demographic rates plays merely a minor role in developed countries (*Keilman et al.* 2002: 412). Only the stochastic approach can fulfil these demands on a forecast, because the future values lie within a confidence interval that serves as the measure of the anticipated precision. Three approaches can be distinguished (*Lipps/Betz* 2003: 5):

- analysis of historical forecasting errors
- assumptions by expert groups
- time series models

It was *Keyfitz* (1981) and *Stoto* (1983) who began analysing historical forecasting errors. By comparing earlier forecasts with actual developments, forecasting errors can be derived, from which a forecast interval can be determined for the future (*Keilman et al.* 2002: 415). However, this approach only supplies viable results, if the future deviations of demographic rates are similar to previous ones and if misinterpretations also apply to the future (*Lipps/Betz* 2003: 5). In addition, frequently only a few periods are available from which the error and the construction of the forecast intervals can be derived, which is why the quality of the approaches appears questionable. Furthermore, early forecasts are based on methods that frequently are no longer state of the art and therefore must be classified as less precise than forecasts made using more recent methods. At the regional level, this approach is usually not implementable or difficult to implement due to the lack of forecasts.

Lutz et al. (1996, 1998a/b) use the assumptions of expert groups to project the demographic rates and their uncertainty (*Lutz et al*. 1996). An expert group agrees both on a point estimate and on a confidence interval. In the next step, a projection of population development with a range of variation is calculated using a simulation under a distribution assumption for the demographic rates and assumptions about their correlation. *Keilman et al.* (2002: 415) speak against the use of this approach citing that even experts cannot reliably differentiate between a 95 % and a 99 % confidence interval. In addition, the risk of a serial correlation is high, which leads to inconsistent confidence intervals. *Lee* (1999: 172) discusses that there is no objective way to derive the uncertainty about future demographic developments from expert opinions. Furthermore, suitable expert groups may not be available at the regional level, which may lead to this approach being impossible.

Time series models are based on the assumption that the demographic developments of the past can be explained by using a statistical model and that this context will also be valid in future. Probably, the most well known time series model is that of *Lee* and *Carter* (1992) designed for modelling mortality. The first step of this two-step approach for estimating future trends of age and gender-specific mortality rates matches the model to the time series of mortality rates and then, in a second step, transfers their development into the future. Since the approach is easy to implement and has proven flexible, it has been frequently adapted and developed further (*Booth* 2006). *Hyndman* and *Booth* (2008: 324) provide an overview of other approaches for modelling demographic rates. Time series models are suitable only for short and medium-term forecasts (*Keilman et al*. 2002: 414). A time span that is too long, results in unrealistic forecasts and very broad confidence intervals (*Sanderson* 1995: 274). The main advantages of time series models over the two stochastic alternatives for a regional stochastic population forecast are that the uncertainty can be illustrated consistently and requires neither earlier forecasts nor expert groups. Since a lot of data are frequently only available for aggregated age groups at the regional level, the paradigm of functional data and models based on them appear helpful for a regional population forecast.[3]

This approach enables an effective disaggregation of secondary data. Based on this, the models by *Hyndman* and *Ullah* (2007) as well as *Hyndman et al.* (2011) demonstrate smaller forecast intervals than alternative approaches (*Hyndman/Ullah* 2007: 4953; *Hyndman et al.* 2011: 25). Therefore, these models are the basis for the stochastic population forecast of the Rhine-Neckar metropolitan area.

3 The paradigm of functional data

3.1 Basic idea

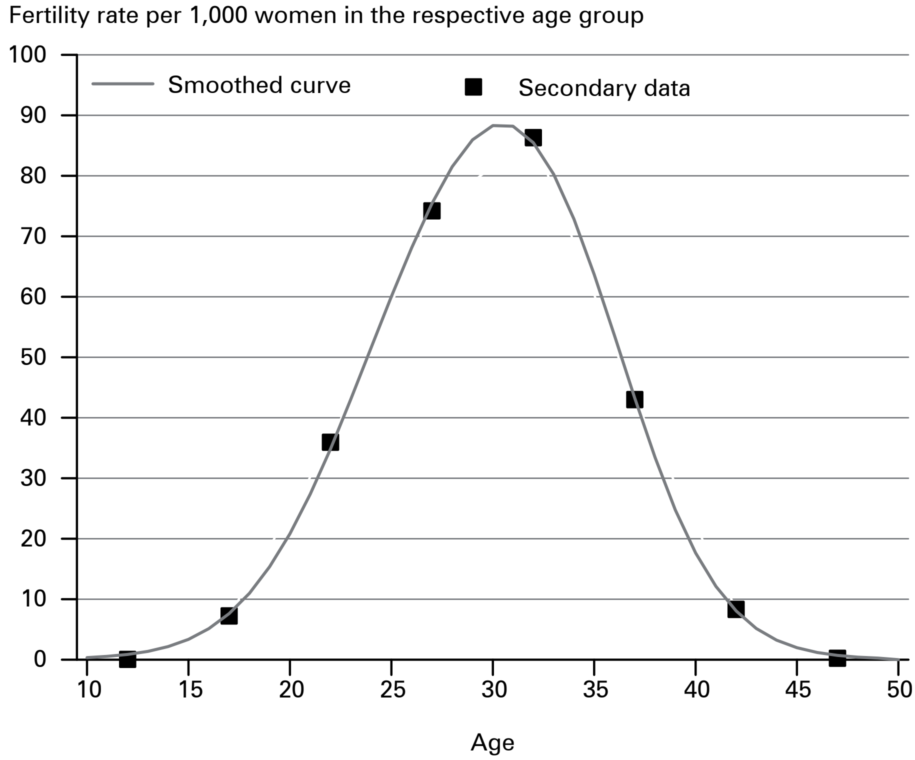
An in-depth analysis of the impacts of demographic change on the population structure requires demographic variables broken down by years of age. When the data situation is not optimal, the use of functional data proves quite helpful, as particularly at the regional level, secondary data are frequently only available differentiated by age groups. The analysis of functional data ("functional data analysis," subsequently referred to as FDA) is an approach for using statistical series, which joins a curve of single data points that represent a connected series (*Ramsay/Silverman* 2001: 5822). The objective of the FDA is to estimate this function and use it as a basis for forecasts. In doing so, each secondary statistical observation $Y_t(x)$ of a demographic component serves as a knot for a smooth function $s_t(x)$. Hence, for a horizon of N points in time, there are N curves. Each observation is subject to a measurement error $\varepsilon_t(x)$ resulting in the following model (*Ramsay* 2008: 5):

$$Y_t(x) = s_t(x)+\varepsilon_t(x)$$

In order to acquire the necessary information about the smooth functions $s_t(x)$ for the FDA, the individual demographic parameters must be disaggregated with their information for age groups to years of age. Statistical smoothing is an approach that attempts to estimate considerable but unavailable patterns of a data set and thereby eliminate any possible noise. According to *Wood* (1994: 27), cubic spline functions deliver the best results for demographic data. The medians of the individual age groups form the supporting points or nodes, of which the first is located at the beginning of the first age group (age=0), the next node at the median

3 Interested readers can find a detailed discussion of the various methodological approaches in *O'Neill et al.* (2001: 210-222).

Fig. 1: Smoothing secondary data using the example of the age-specific fertility rates of one year



Source: own design

of the second group and so forth. The final node is located at the end of the upper age group. The years of age comprising an age group form a section. The smoothing method joins the nodes through different cubic polynomials for the individual sections to a curve, which is always differentiable for all points. The course of the spline functions allows us to read a value for the modelled demographic component for each year of age. Figure 1 illustrates the procedure using the example of a fertility rate: the median values of the individual age groups form the nodes (squares), between which cubic splines run, which compose the smoothed curve.

Functional data, or the individual smooth functions for the N points in time, underlie two variations: amplitude variation and phase variation (*Ramsay* 2008: 3). Amplitude variation is a vertical variation of the functions over time. Hence, the "amount" of the function values is not constant. Phase variation indicates the horizontal variation that, for example, the age-specific characteristics of the demographic parameters show a specific trend over N points in time. For instance, in the past decades the age of women who reached the maximum of the age-specific birth rate rose continuously. These variations in the curves complicate modelling with time series models, as by smoothing data to years of age, models with very many parameters would be needed to portray all of the developments in age and time adequately. For this reason, the FDA uses the so-called "time-warping" approach to illustrate the characteristic variables of a function with a low number of parameters and to minimise their variation through amplitude variation and phase variation (*Ramsay/Silverma*n 2001: 5823): A linear combination of basis functions φ_k (k=1, …, K) corrects the biases:

$$\sum_{k=1}^{K} \beta_k \varphi_k(x).$$

The basis functions contain information about the age and time-specific developments of the smooth function. The more basis functions are used, the more flexible a time series model becomes. The big challenge is keeping the function simple and the number of parameters used low. For the more basis functions are used in calculating the smooth function, the more sensitively the function reacts to measurement errors. This approach is widespread, since the basis functions can be simply estimated by using a principal component analysis, which describes the development of all of the main variables of the curves over the course of time. This alignment of functions over the time frame, also called registration, facilitates further assessments and improves comparability between different points in time. The remaining variation of the curves is merely a result of pure amplitude variation.

3.2 Models

Hyndman and *Ullah* (2007) use functional data to develop a model that can reliably indicate different demographic parameters. For instance, $Y_t(x)$ describes the value of a demographic rate for age x at time t. Based on the basic idea of FDA, the model is based on a smooth function $s_t(x)$, which is observed with one error to a discrete point in time depending on the age. For the (secondary statistical) data points $\{x_i, y_t(x_i)\}$, whereby t = 1, …, n and i = 1, …, "highest age group", therefore:

$$Y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_t(x).$$

The error terms $\varepsilon_{t,x}$ are independently and identically distributed and weighted by the term σ_t depending on the age. This is advisable, because for small function values in certain age ranges, for example the fertility rate for women under 15 and over 45 years of age, the measurement errors are also correspondingly small. Not weighting them could lead to the impression that the model in these sections indicates the data better than in other age ranges. The modelling aims to forecast $Y_t(x)$ for the range of h points in time (t = n+1, …, n+h) and all defined age groups. This approach is based on functional data and attempts to identify and record the amplitude variation and the phase variation as sources of measurement errors:

1. For a population forecast, data are desirable that are broken down by years of age, for this is the only way, for example, to make the diverse effects of demographic change on the age structure of the population visible. However, secondary statistical data are frequently aggregated to age groups. In this case,

the age structure is smoothed using non-parametric smoothing methods for each available point in time, in order to be able to estimate the function that Y_t is based upon. A separate function is estimated for each point in time.

2. Modelling all main characteristic attributes of a demographic variable requires a number of parameters. The lengthy time series data needed are rarely available, particularly at the regional level, due to regional reforms and other circumstances. Instead, basis functions break down the functions estimated in step one into individual elements containing all of the main variables, which, however, have to do with a lesser number of parameters:

$$s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\varphi_k(x) + e_t(x),$$

whereby $\mu(x)$ is the mean of $s_t(x)$ over all observed years, $\{\varphi_k(x)\}$ is a set of basis functions and $e_t(x)$ is a normally distributed error term with mean 0 and variance var(x). The basis functions are the result of a principle component analysis. The mean shows the essential structure of the modelled demographic variable, while the developments deviating in age and time from the mean through the amplitude and phase variation are contained in the basis functions.

3. A univariate time series model is estimated for every coefficient $\beta_{t,k}$ with k = 1, …, K. *Hyndman* and *Booth* (2008: 327) show that the method reacts insensitively to the choice of K, if based on a sufficiently high value. For this reason, it is advisable to form as many basis functions as possible. However, this increases the duration of the model calculation, and the basis functions with a low explanatory content are in some cases difficult or impossible to interpret.

4. The variation caused by the amplitude and phase variation is subject to chronological deviations. The parameter $\beta_{t,k}$ indicates these developments and is forecasted for a time frame of h points in time:

$$\hat{y}_{n,h}(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n,k,h}\hat{\varphi}_k(x),$$

whereby $\hat{\beta}_{n,k,h}$ marks the estimator at time n+h for $\beta_{n+h,k}$.

5. The estimated and forecasted values are the basis for the calculation of future values of $s_t(x)$ from step two, as the forecasts for all age groups are calculated from the point estimators multiplied by the estimated basis functions.

6. The forecast intervals are calculated from the estimated and forecasted variances of the error terms (step two). These confidence intervals are valuable additional information compared with deterministic approaches, as rather than indicating the range of variation of possible developments simply through a variation of scenarios with different assumptions, this approach is based upon statistical methods and distributions.

The advantage of this approach is that it allows for smooth functions, is robust against structural breaks, such as crises, wars or disasters, and supplies a model framework that enables restrictions,[4] yet is simultaneously useable for different (demographic) models. For instance, *Erbas et al.* (2007) forecast the risk of breast cancer in Australian women. The model is a generalisation of the groundbreaking Lee-Carter model (*Lee/Carter* 1992) for estimating and forecasting mortality rates, which in many enhancements can also be used for other demographic variables, for example fertility data (*Lee* 1993). The basic idea of the Lee-Carter model is also reflected in the FDA models, however it differs in some points, for example in that Lee and Carter did not require any smooth functions (*Hyndman/Ullah* 2007: 326).

The approach by *Hyndman* and *Ullah* (2007) estimates the subpopulations of the modelled demographic variable (frequently broken down by male and female) independently of one another. In this case, the forecasted development of both figures can diverge in a lengthy time frame. However, this effect can usually not be explained or founded, or only with difficulty. *Wilson* (2001: 167) observes a global convergence for the mortality rates both between different countries and between the sexes. Approaches that take up and explicitly model this aspect are called "coherent models" (*Li/Lee* 2005: 581) in the literature. Forecasts based on this way of thinking should demonstrate a certain structural correlation. For mortality rates, this was implemented through methodological enhancements of the classical Lee-Carter approach, for example by *Lee* and *Nault* (1993), *Lee* (2000) as well as *Li* and *Lee* (2005). *Hyndman et al.* (2011) used functional data and enhance the approach of *Hyndman* and *Ullah* (2007) in order to avoid a diverged development as the result of imprecise estimations. Since differentiation is made in the stochastic population forecast only according to gender, the following remarks are limited to the case of two subpopulations. The approach also proves very flexible and can be expanded to a larger number of subpopulations.

The secondary statistical values $Y_{t,F}(x)$ of a demographic component are based on a smooth function $s_{t,F}(x)$, which is observed with an error. Among two subpopulations, the following applies to the women (subscript F):

$$Y_{t,F}(x) = \log\left(s_{t,F}(x_i)\right) + \sigma_{t,F}(x)\varepsilon_t(x_i),$$

whereby x_i is the mean of the age group i (i = 1, …, p), $\varepsilon_t(x_i)$ is an independent and identically distributed random variable and $\sigma_{t,F}$ is a term through which the measurement errors vary in age. With the exception of the logarithmic formulation, this approach corresponds to the described approach by *Hyndman* and *Ullah* (2007). The smooth function is observed for both men and women. Modelling a coherent development requires the following two terms: the geometric average (P_t for product

4 The approach allows for qualitative restrictions; for example the assumption that the probability of death rises monotonically from a certain age can be integrated. The restrictions employed for this study are contained in Chapter 4.2.

model) and the square root from the ratio (R_t for ratio model), each for the smoothed figures of a demographic component $s_{t,F/M}(x)$:

$$P_t(x) = \sqrt{s_{t,M}(x) \cdot s_{t,F}(x)}$$
$$R_t(x) = \sqrt{s_{t,M}(x)/s_{t,F}(x)}.$$

The product model describes the age-specific progression of a demographic variable, while the ratio model adapts the gender-specific variation. There are hardly any corrections at values close to one of R_t, while values deviating from one indicate gender-specific deviations in a specific age range. The advantage of this approach is that through the logarithmic formulation, the product in P_t becomes a sum or a difference from the ratio in R_t. *Tukey* (1977) showed that both are practically uncorrelated.

The product model and the ratio model also underlie variations through the amplitude variation and the phase variation. To prevent a diverged development of the demographic rate and also indicate its age and gender-specific change, P_t and R_t are portrayed by the basis functions model for functional data by *Hyndman* and *Ullah* (2007):

$$\log[P_t(x)] = \mu_P(x) + \textstyle\sum_{k=1}^{K} \beta_{t,k}\varphi_k(x) + e_t(x)$$
$$\log[R_t(x)] = \mu_R(x) + \textstyle\sum_{l=1}^{L} \gamma_{t,l}\psi_l(x) + w_t(x)\,.$$

The basis functions $\varphi_k(x)$ and $\psi_l(x)$ contain the variables of P_t or R_t deviating from the mean (μ_p and μ_r). The change of the basis functions over time is shown by the coefficients $\beta_{t,k}$ and $\gamma_{t,l}$. The model errors measure the error terms $e_t(x)$ and $w_t(x)$.

To ensure that the estimators for the forecast are coherent, or do not diverge, the coefficients of the basis functions $\{\beta_{t,k}\}$ and $\{\gamma_{t,l}\}$ for the time series models used must be stationary processes. The forecasted estimators are multiplied by the basis functions. The results are forecasts until time n+h for future $P_t(x)$ and $R_t(x)$ from which the future gender-specific values for the modelled demographic component result:

$$Y_{n,h,M}(x) = P_{n,h}(x) \cdot R_{n,h}(x)$$
$$Y_{n,h,F}(x) = P_{n,h}(x)/R_{n,h}(x)\,.$$

The forecast of $\beta_{n,k,h}$ and $\gamma_{n,l,h}$ for point in time n+h for the individual genders (subscript j marks "male" or "female") is determined by:[5]

$$\log\big[y_{n,h,j}(x)\big] = \hat{\mu}_j(x) + \sum^{K} \hat{\beta}_{n,k,h}\varphi_k(x) + \sum^{L} \hat{\gamma}_{n,l,h}\psi_{l,j}(x).$$

5 The exact procedure for forecasting the individual coefficients is found in *Hyndman et al.* (2011: 6-7).

The presented models based on functional data are suitable for short- and medium-term forecasts of approx. 20 years (*Hyndman*/*Booth* 2008: 339). For a longer time frame, the accuracy decreases and the width of the confidence intervals increases. The forecasts of the demographic variables are based on information about the development in the past, which is why they cannot indicate a future trend reversal. However, both are basically true for the use of time series models with no influence on the choice of model for the forecast of the population. In a direct comparison to alternative specifications, both approaches show greater accuracy (cf. *Hyndman*/*Ullah* 2007: 4953 and *Hyndman et al*. 2011: 17). The gender-differentiated demographic variables (net migration and mortality) are modelled using the approach by *Hyndman et al*. (2011), while the model by *Hyndman* and *Ullah* (2007) is assumed for the fertility rate. Both approaches are particularly suitable for regional population forecasts, as through the use of basis functions the results are very robust against outliers and structural breaks. This is an advantage since at the regional level, for example because of regional reforms, long time series are frequently not available. In addition, social developments such as the possible trend reversal from suburbanisation to reurbanisation can be indicated at a correspondingly lower aggregation level through the basis functions. Furthermore, many variables offered by the statistical offices of the federal states are only broken down by age groups. Using the ideas of the paradigm of functional data, this information can be disaggregated to single years of age. The basis functions also allow us to model demographic trends and structural breaks, which are manifested in the form of phase variation or amplitude variation in the time series. Whether their content can be interpreted or serve merely to correct statistical effects depends on their explanatory content.

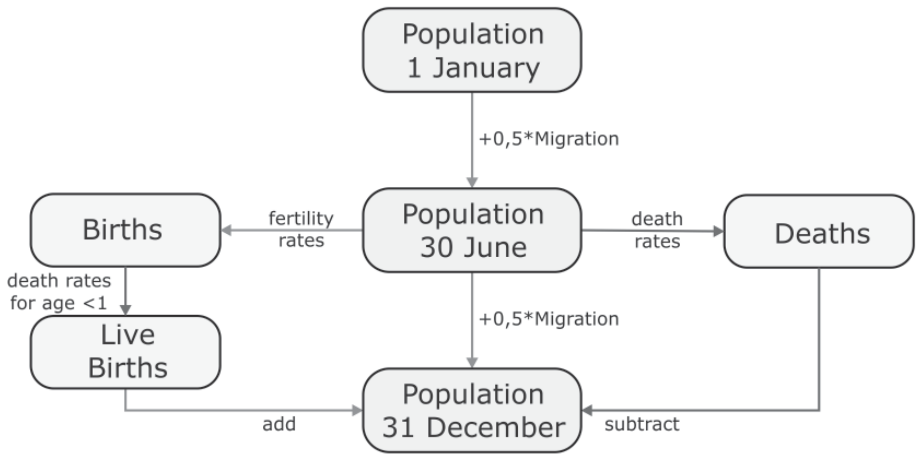
4 The population development of the Rhine-Neckar metropolitan area

4.1 Algorithm

Five demographic components influence population development: the fertility rate, the gender-specific mortality rates as well as male and female net migration. The cohort component method formalises this correlation to a system of equations, however, without precisely specifying the processes within one year. This population forecast is based on the algorithm by *Hyndman* and *Booth* (2008: 340), which is diagrammatically portrayed below (Fig. 2). It begins with the population on 31 December of the previous year. The migration movements do not occur completely at one fixed point in time, but distributed throughout the year. During this time, immigrants can theoretically have children or die. The algorithm takes this aspect into consideration by distributing the number of simulated age-specific net migration changes, half at 1 January and half at 31 December of one year. The population on 30 June of a year corresponds to the population on 31 December of the previous year plus the first half of the net migration. The deaths are calculated from the population on 30 June multiplied by the mortality rates. The number of live births is calculated along these lines, corrected by the infant mortality rate. The newborns are distributed to the genders on the basis of the moving average of the past five years. The female population at time t+1 is calculated:

$$B_{t+1}^{F} = \Omega_{t}^{F}\left[B_{t}^{F} + 0{,}5 \cdot M_{t}\right] + 0{,}5 \cdot M_{t} \, .$$

Fig. 2: Algorithm for the cohort component method



Source: own design according to *Hyndman* and *Booth* (2008)

The population at the end of the year (31 December) is determined through the sum of the population on 30 June of a year, the deaths, the adjusted live births and the second half of the net migration. This population is then used as the starting point for calculating the subsequent year, for which the algorithm is passed through from the beginning. Unlike deterministic models, which project a single population development per scenario each, due to the uncertainty about future development, stochastic approaches require more complicated simulations, since each of the forecasted values are dispersed about one mean each. On the basis of the assumed distributions for the individual demographic rates, a predetermined number of possible trajectories of the future population are simulated and stored in a database. The confidence interval, in which the future population lies, ranges here from (100-α) / 2nd percentile to (100+α) / 2nd percentile (*Keilman et al.* 2002: 416).

Separate simulations of the models from Section 3.2 for the fertility rate, the male and female mortality rates as well as the gender-specific net migration for each year from 2010 until 2030 form the basis for the population forecast. The fertility and mortality rates refer to the population on 30 June (cf. Fig. 2). The births and deaths follow the assumption of the Poisson distribution and are calculated by random drawing from the distribution and summation of these values over the relevant age range. *Hyndman* and *Booth* (2008: 328) provide a detailed overview of the generation of the respective trajectories and their integration into the population forecast.

All of the following calculations were carried out using the statistical program "R" (Version 2.12.0). This free statistical software is available on all prevalent platforms, and the add-on package "Demography" also contains the described approaches for functional data as well as the presented algorithm for the cohort component method. The forecasts require the "Forecast" package. A total of 1,000 progressions of population development were simulated and then exported into an Excel database to calculate the forecast intervals from "R."

4.2 Data: Sources and preparation

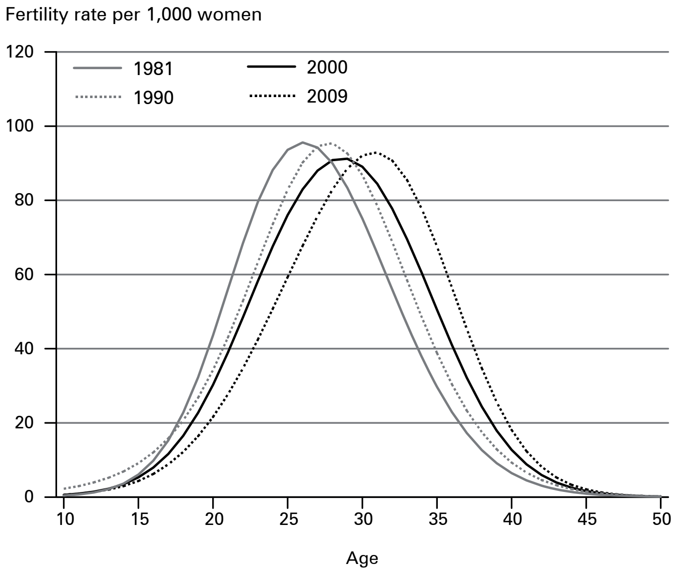
The data used for the population forecast were provided by the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate. All of the aggregated values for the metropolitan area arise from the sum of the individual values from the 15 rural and urban districts. Calculation of the fertility rate requires the number of live births according to age groups of the mothers {10-14, 15-19, …, 40-44, 45-49}. Since the data do not differentiate between the genders of the live births, we assume that per 100 females born there are 105 males born. The mortality rate is calculated from the number of gender-specific deaths per age group {<1, 1-4, 5-9, …, 70-74, 75 and older} to the number of men or women in the corresponding cohort. Both variables are available for the years 1981 to 2009. These variables are referenced to the population data of the corresponding time period, broken down by age groups {<3, 3-5, 6-9, 10-14, 15-17, 18-19, 20-24, 25-29, …, 70-74, 75 and older} and gender in order to calculate the rates. Calculating mortality rates for boys and girls under 10 years of age proved problematic, since the age groups of the deaths and risk population diverged. For this reason, the population data of a group were distributed evenly over the individual years of age and aggregated to the age groups of the deaths. The data on net migration for the years 2002 to 2009 are differentiated according to gender and divided into the age groups {<18, 18-24, 25-29, 30-49, 50-64, 65 and older}. The author's analyses based on the residual method presented above supplement these series by the years 1995 until 2001.

However, these data sets do not fulfil the demands of the models presented here. An analysis of functional data requires smooth (age) functions to base the secondary data upon. Therefore, the median is formed for each age group of the individual demographic rates and used as a node between the smoothed demographic rates. The mortality rates should rise monotonically from a certain threshold (for example years of age greater than 50) (*Hyndman/Ullah* 2007: 4945), because the older a person is, the higher their probability of death. This qualitative restriction can be integrated into the smoothing of the data with weighted penalised regression splines according to *Wood* (1994). This procedure reduces biases in the estimated curves of the upper age cohorts. For the same reason, the data sets of the fertility rates are assumed as concave and prepared with weighted B regression splines according to the method by *He* and *Ng* (1999). Net migration is not a relative value and can be both positive and negative. A weighted locally quadratic regression is carried out (*Moore et al.* 1997) as the smoothing method. These three smoothing methods are contained in the "Demography" package for "R."

Figure 3 shows the chronological development of the age-specific fertility rate for the time range between 1981 and 2009 based on the curves for 1981, 1990, 2000 and 2009. The shift of the maximum of the age-specific rates reflects the rise in the median age of mothers at births of 26.9 years (1981) to 29.6 years (2009). This development reveals the advantage of modelling based on functional data, as the empirical curves for a specific demographic parameter deviate over the range of N points in time from the mean of this parameter. The development corresponds to a phase variation taken into consideration while modelling the time series through the basis functions of the FDA model.

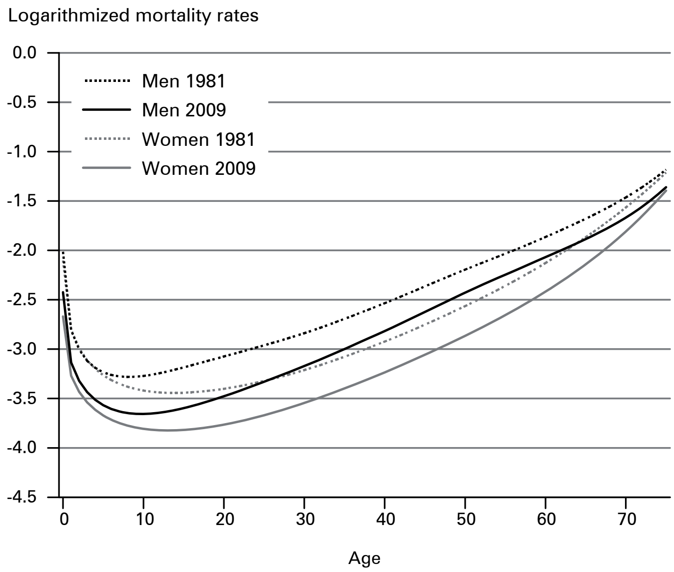
The trend observed in many developed countries and regions with a constantly rising life expectancy is also observed in the mortality rates of the men and women in the Rhine-Neckar metropolitan area (Fig. 4). Women across all years of age have a higher life expectancy and lower mortality rates than men. Between 1981 and 2009, the figures drop equally for both. The main difference is found in the male mortality rates between 35 and 65 years, which run almost linear, while the curves for women show a rather concave progression. The mortality rates between the genders differ the most in this interval. In the higher age groups, the gender-specific curves converge. For all people over 75 years, the (logarithmised) figures could not be smoothed due to the data situation and therefore had to be linearly interpolated. The target values for the age group "90+ years" of the individual years were tak-

Fig. 3: Smooth age-specific fertility rate per 1,000 women for 1981, 1990, 2000 and 2009



Source: Author's calculation of the functional data using B regression splines according to the method by *He* and *Ng* (1999) and on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. 4: Smoothed and logarithmised mortality rates for men and women for 1981 and 2009

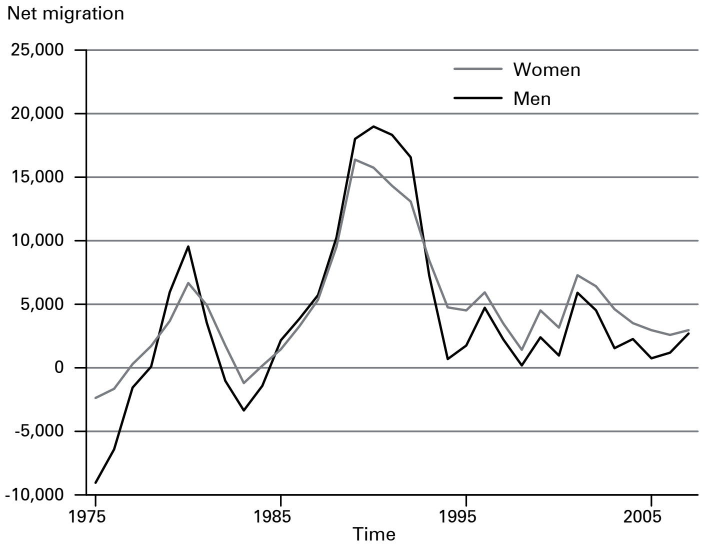


Source: Author's calculation of the functional data using weighted penalised regression splines according to *Wood* (1994) and on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

en from the mortality tables of the German Federal Statistical Office (*Statistisches Bundesamt* 2010: 59). The population in the age from 90 years is contained in the upper age group "90 and older."

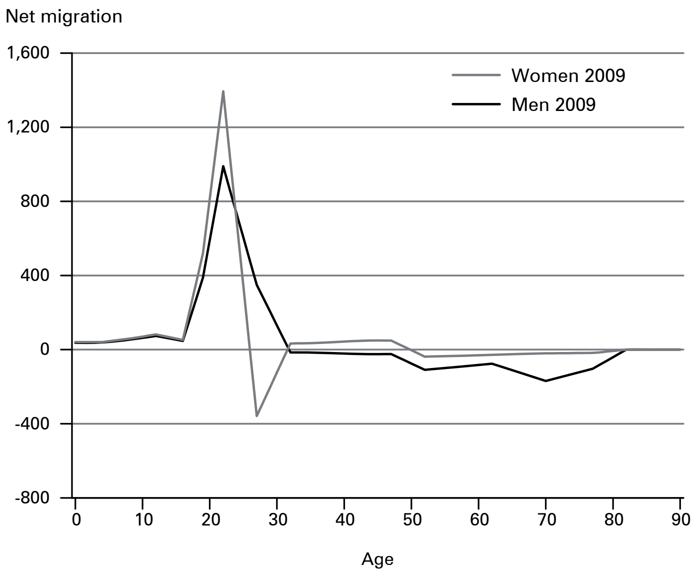
Net migration underlies a variety of influences, which are effective both in the respective regions of origin and destination and which lead to considerably greater chronological fluctuations than in the fertility and mortality rates (Fig. 5). After the influxes and outfluxes connected to German reunification in the early 1990s ebbed, the series stabilised noticeably. And since about the middle of the 1990s, the migration surpluses of women are higher than that of men in the metropolitan area (Fig. 5). The data show that particularly young men and women of the ages between 18 and 25 years migrate into the metropolitan area (Fig. 6), which underscores its significance as an educational and university location. Slight migration losses occur only in the upper age groups.

Fig. 5: Development of net migration (1975-2009)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. 6: Gender differentiated net migration according to years of age for 2009



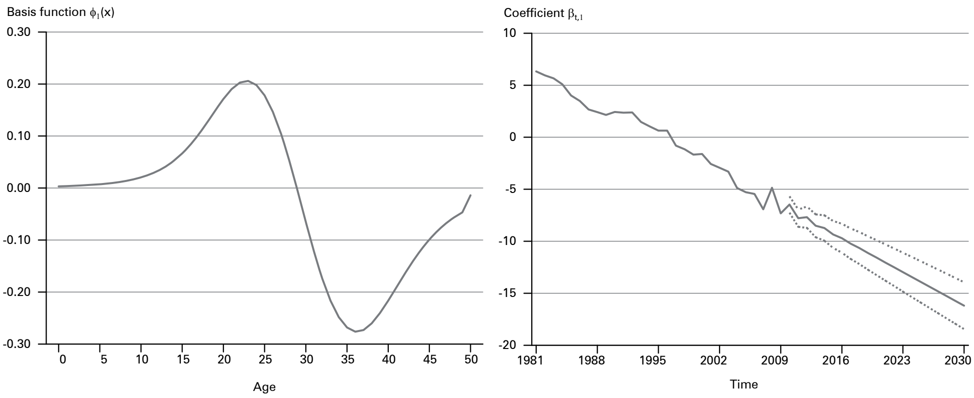
Source: Author's calculation of the functional data using weighted "locally quadratic regression" according to *Moore et al.* (1997) and on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

4.3 The demographic components: model and forecast

The smoothed data of the fertility rate consist of 29 curves for the years 1981 to 2009. A model with K=6 basis functions was estimated for this period. *Hyndman* and *Booth* (2008: 327) showed that the results of the method are independent of the number of basis functions and K=6 is sufficiently large. The basis functions encompass 82.7 %, 11.3 %, 3.3 %, 1.9 %, 0.6 % and 0.1 % or 99.9 % of the variation deviating from the mean. The coefficient of the first basis function (Fig. 7) shows the drop in the fertility rate ($\beta_{t.1}$ falls within the period of observation), which is strongest among women between 20 and 25 years (highest positive function values of the basis function φ_1). The negative values of the basis function for ages between 35 and 40 years show that births increase in this range: The first basis function thus describes the social trend of shifting births to a later point in time. The remaining basis functions as a whole comprise an insufficient share of the variance of the smoothed function to portray the age-specific fertility rates for the period from 1981 until 2009 and are therefore not suitable for contextual interpretation.[6] They counteract systematic statistical biases, which are caused by amplitude variation and phase variation.

The forecast of the fertility rate was conducted for the years 2010 until 2030 (Fig. 8). The maximums of the age-specific fertility rates rise over the entire period, and the individual curves shift over the course of time into higher age groups. The average age of mothers rises from 29.85 (2009) to 31.36 years (2030) and lies within the 80 % forecast interval [31.06; 31.65]. The total fertility rate (TFR) increases con-

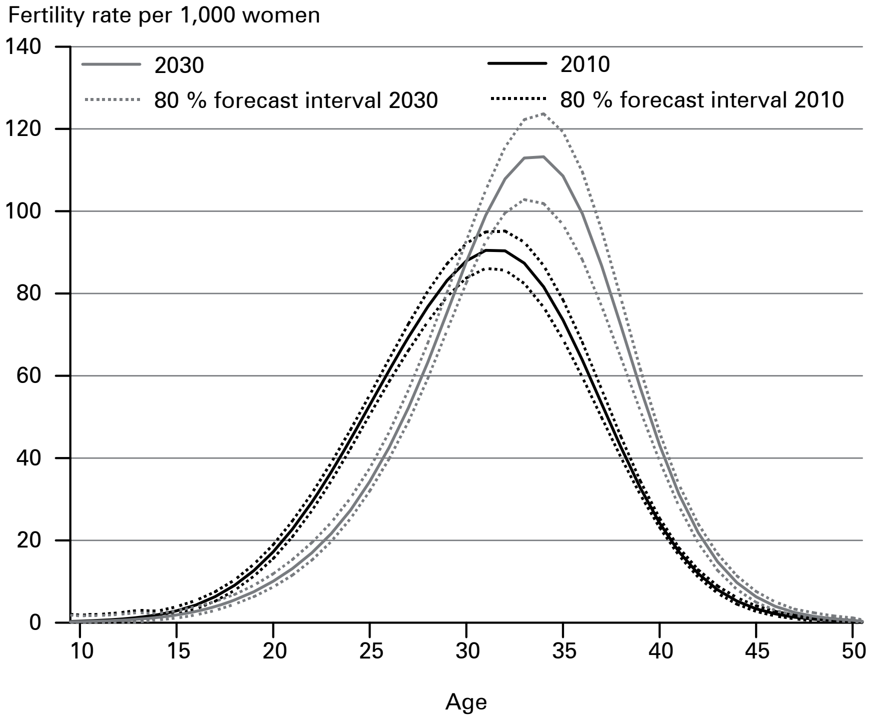
Fig. 7: The first basis function of the fertility model including the forecast of the coefficients $\beta_{t.1}$ for 2010 until 2030 and the 80 % forecast interval (dotted lines)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

6 All basis functions of the models for fertility, mortality and net migration are illustrated in the Appendix.

Fig. 8: Forecast of the fertility rate per 1,000 women for 2010 (black) and 2030 (grey) with the 80 % forecast intervals (dotted lines)

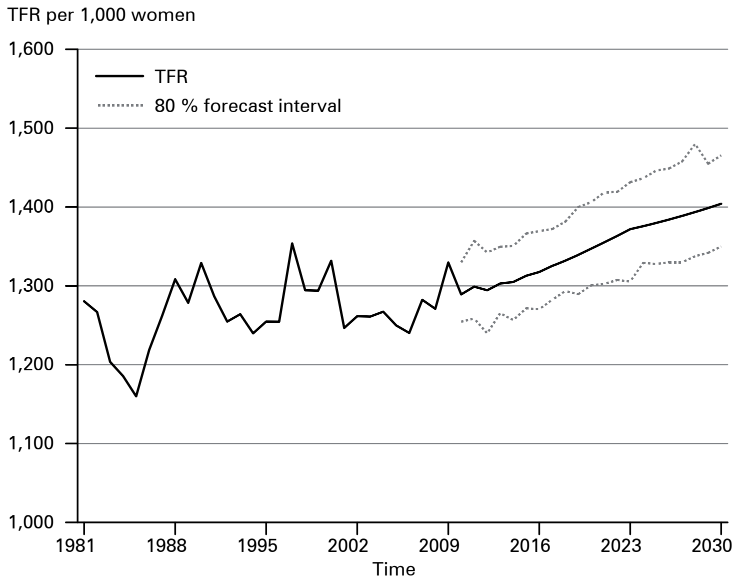


Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

sistently over the forecast range. Figure 9 contains the point estimators and the 80 % forecast intervals. The mean TFR therefore corresponds with expectations of the development of the national TFR with 1.4 births per woman (*Statistisches Bundesamt*, 2009: 27).

Heckman and *Walker* (1990) cite a high correlation between the childbearing behaviour and the income of women as the reason for the shift of births to later in life. This economic interpretation is, however not comprehensive enough. *Ott et al.* (2006) sketch out a three-phase model that takes children into account in life planning only after reaching certain vocational objectives. *Gustaffson* (2001) bases the shift on disadvantages of motherhood on the career of potential mothers. *Börsch-Supan* and *Wilke* (2009) assume that the emerging decrease of people of working age will lead to a growth in female employment, and they will thus approximate the employment behaviour of men. Since the postponement of births is, however, biologically limited, births must be increasingly made up for in later years of age. These effects are reflected in the forecast: the births shift chronologically, the average age of mothers at childbirth rises and the fertility rates of the higher age groups

Fig. 9: Development of TFR per 1,000 women for 1981-2030 with the 80 % forecast interval



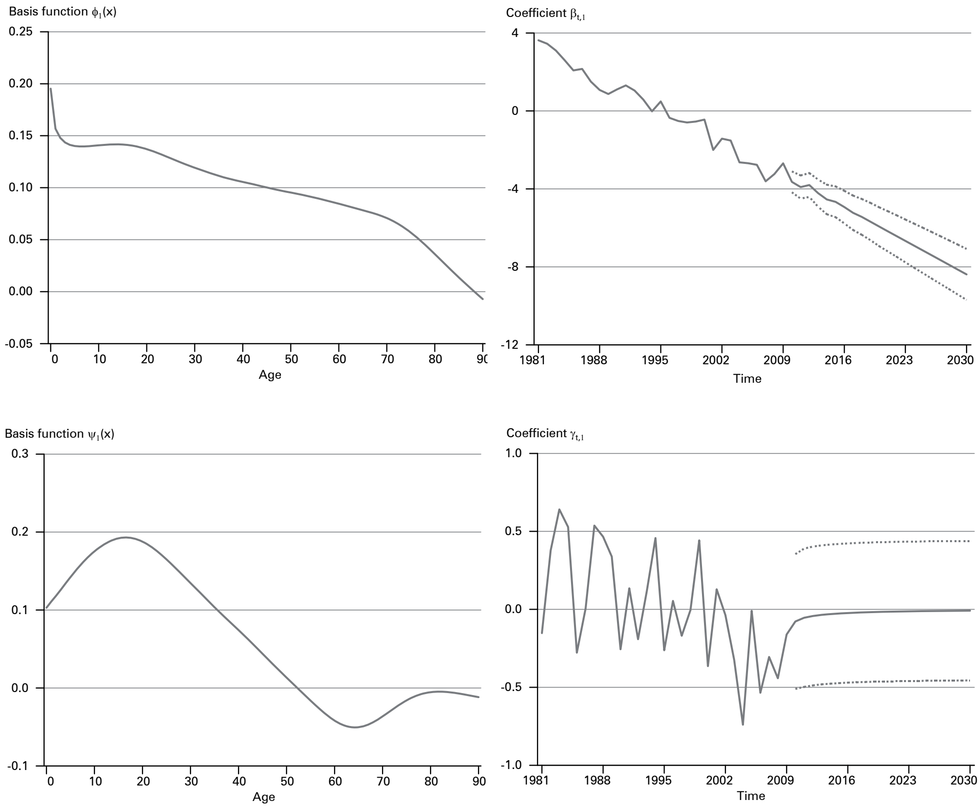
Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

rise considerably. In future, women in the high age groups will contribute most to the modelled TFR (Fig. 9).

To prevent the mortality rates from diverging in the forecast, the model is based on the coherent model of functional data, which contains six basis functions both for the models of the product and for the ratio. The basis functions of the geometric mean of both genders comprise 96.6 %, 1.5 %, 1.0 %, 0.5 %, 0.2 % and 0.1 % of the variation of the smooth functions on which the mortality rates are based. The coefficient $\beta_{t,1}$ of the dominant first basis function (Fig. 10) shows a drop in mortality rates, which is reflected the most in infant mortality (< 1 year) and lessens with age. The basis functions of the ratio model comprise 60.2 %, 17.6 %, 10.2 %, 6.8 %, 3.0 % and 1.8 % of the variation. The first basis function shows that the greatest deviations in the mortality rates between men and women are at about 20 years. This variation is, however, subject to strong fluctuations over time.

The mortality forecast indicates the trend of consistently declining figures, which slows down, however, over the course of time (Fig. 11). The life expectancy (Fig. 12) of women rises until 2030 to 80.55 years and lies within the 80 % forecast interval [80.26; 80.82], while that of the men rises to 78.75 years with the 80 % forecast interval [78.47; 79.06]. *Carnes* and *Olshansky* (2007: 377) cite the obesity of children and young people as well as new infectious diseases as sources that could possibly weaken or even reverse this trend. These possible future influences cannot, however, be included in a forecast based on time series models.

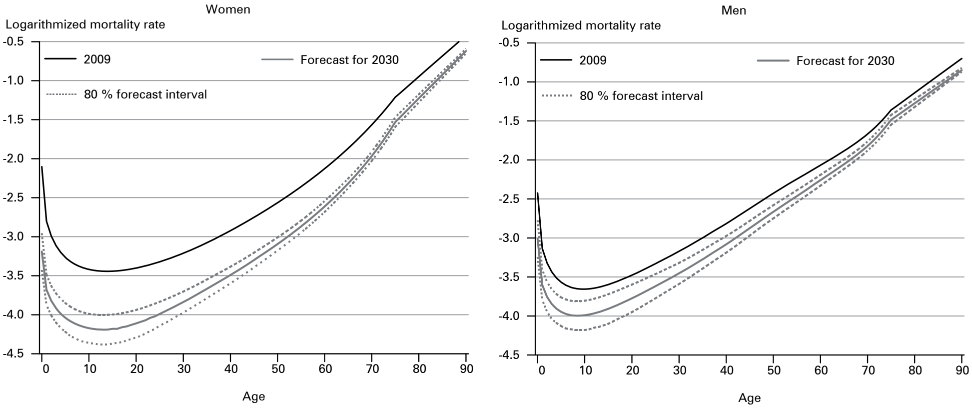
Fig. 10: The first basis functions of the mortality model including the forecast of the coefficients $\beta_{t.1}$ and $\gamma_{t.1}$ for 2010 until 2030 and the 80 % forecast interval (dotted lines)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

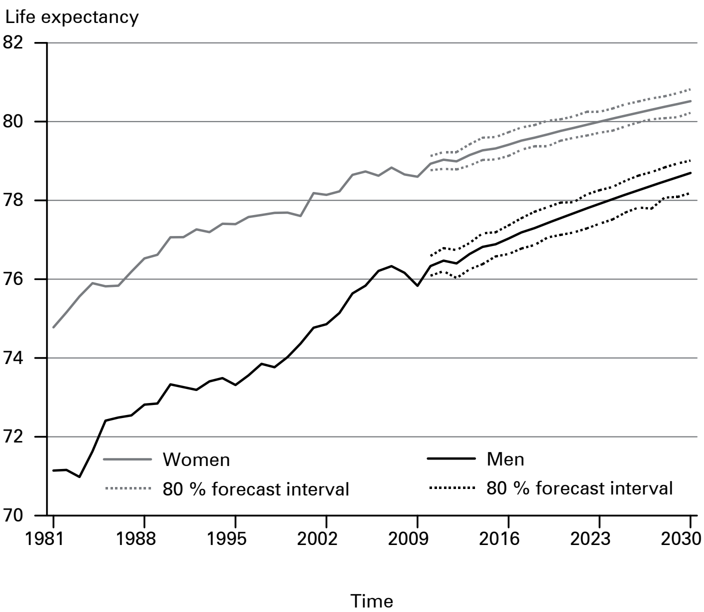
The aggregated net migration (Fig. 5) shows that the values for the two genders run similarly, which justifies modelling using the coherent model. The time series prior to 1995 is subject to great dynamics, which had a negative effect on the model and the subsequent forecast. For this reason, only the values between 1995 and 2009 are included in the model. The high male and female net migration between 18 and 25 years is the dominant effect in the model. The first of the six basis functions of the product model indicate the change over the course of time (Fig. 13) and records 81.6 % of the variation from the mean, the remainder by contrast only 10.5 %, 5.2 %, 1.4 %, 0.7 % and 0.4 %. The first basis function of the ratio model illustrates the gender-specific differences of educational migrants (variance percentage 53.9 %). The coefficient shows that male net migration between 2000 and 2003 was higher but that the trend weakened and even reversed in ensuing years. This

Fig. 11: Age and gender-specific mortality rates for 2009 (black) and the forecast for 2030 (grey) with the 80 % forecast intervals (dotted lines)



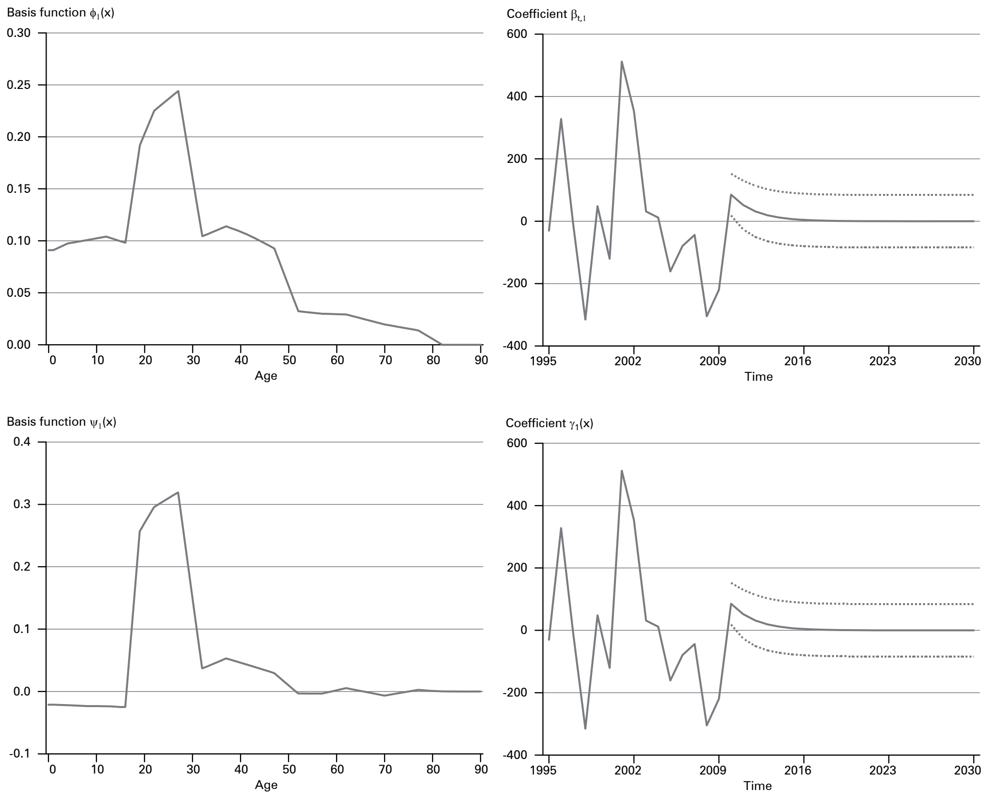
Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. 12: Development of the life expectancy of men (black) and women (grey) for 1981-2030 and the 80 % forecast intervals (dotted lines)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

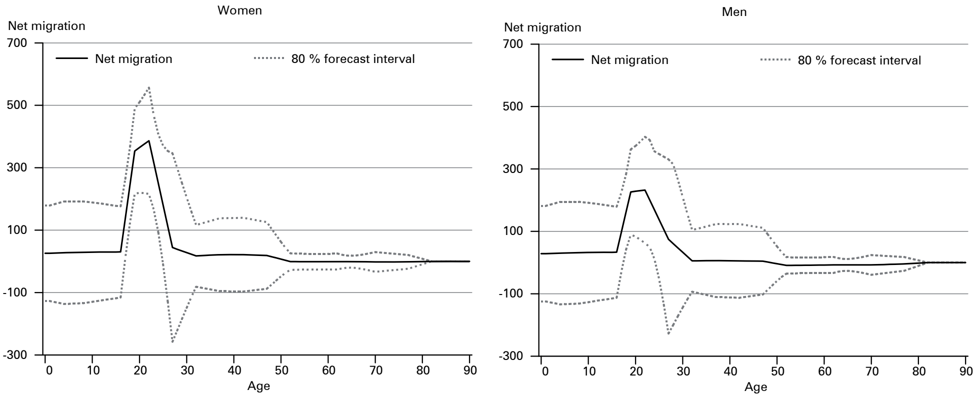
Fig. 13: The first basis functions of the migration model including the forecast of the coefficients $\beta_{t.1}$ and $\gamma_{t.1}$ until 2030 and the 80 % forecast interval (dotted lines)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

development continues in the forecast, as in 2030 (Fig. 14) compared to 2009 (Fig. 6) the values of the women in the age range between 18 and 25 years is considerably higher than that of the men (Fig. 14). In the model, the Rhine-Neckar metropolitan area remains an attractive education or study location for young people. Nonetheless, due to the wide forecast intervals for people under 50 years of age, the forecasts of net migration are the greatest source of uncertainty about future population development.

Fig. 14: Forecast of age distribution of net migration of men and women for 2030 with the 80% forecast intervals (dotted lines)

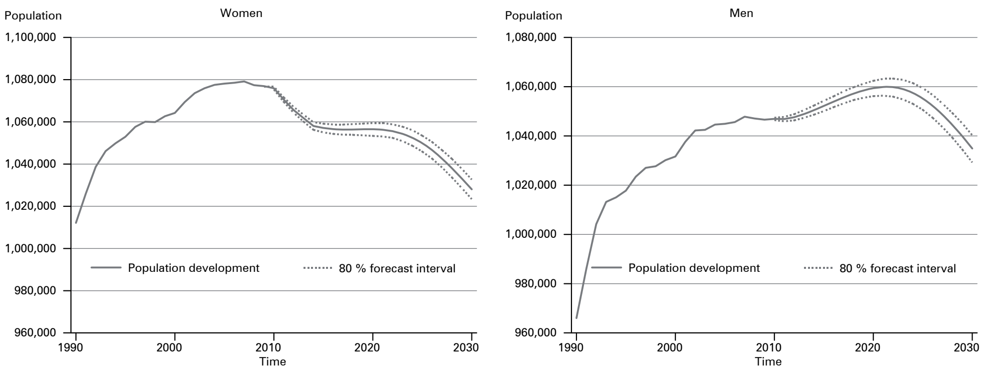


Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

4.4 The population development of the Rhine-Neckar metropolitan area

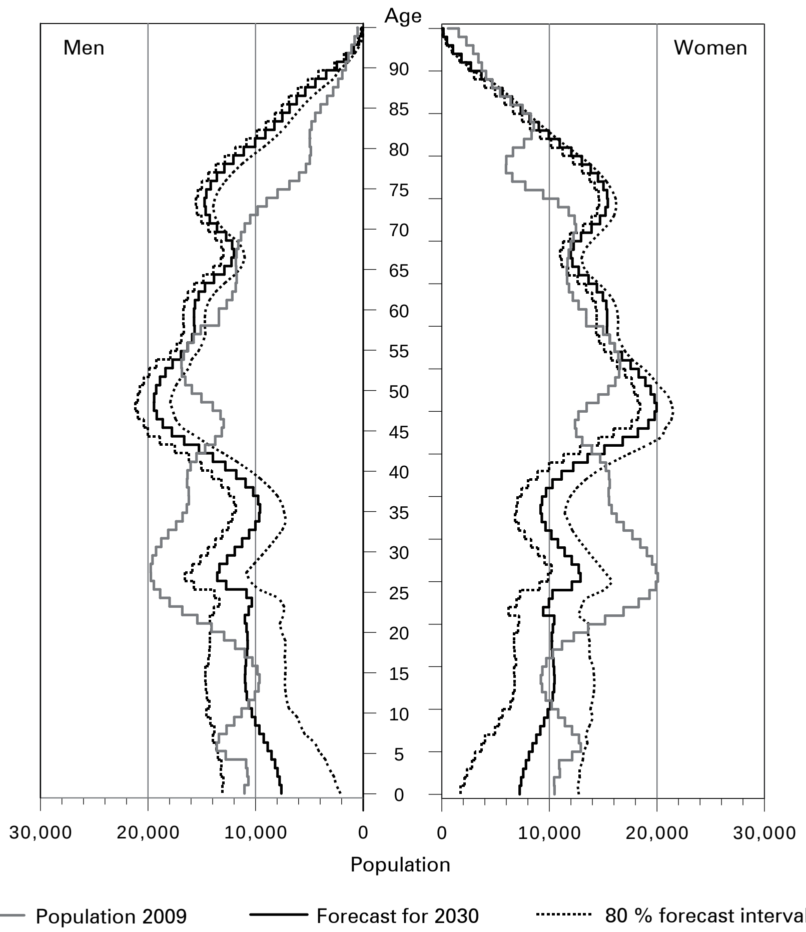
Since the 1990s, the male and female population of the Rhine-Neckar metropolitan area has been growing steadily due to migration gains caused by the political developments in Germany (reunification) and Europe (immigration of repatriates of German origin) (*Gans/Schmitz-Veltin* 2006: 316). Population development for the years 2010 until 2030 is calculated using the algorithm described in Section 4.1 on the basis of the cohort component method and the recorded births, deaths and migrants from the demographic components. The male population will rise slightly until 2022 (Fig. 15) and then drop measured on the median of the simulation by

Fig. 15: Development of the male and female populations between 2010 and 2030 with the 80 % forecast intervals (dotted lines)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. 16: Male and female population according to years of age 2009 (grey) and the 0.5 quantile of the simulation for 2030 (black) with the 80 % forecast intervals (dotted lines)



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

2030 to 1.0349 million inhabitants, which corresponds to a decrease of 2.31 % for the forecast range, and lies within the 80 % forecast interval [1.0292; 1.0403]. The number of women will drop by 2030 by 4.45 percent to 1.0280 million inhabitants (80 % forecast interval [1.0234; 1.0327]). According to this, the female population will decrease somewhat more strongly than the male. The population pyramid in Figure 16 shows the anticipated age structure for the year 2009 as the last year of the time series data and the forecast for 2030 with the 80 % forecast intervals. The largest percentage of the total population in 2009 was the age groups between 20 and 30 years. The fertility rates lie considerably below the level for maintaining the population and even the positive net migration is not sufficient to balance out the mortality surpluses. This makes a distinct decrease in the number of people under 30 years of age apparent. The number of people of working age under 30 years decreases enormously, while the group of the 30 to under 65 year olds will remain at about the level of 2009. This effect arises mainly from the shift of the largest age groups from 2009 to the older age groups in 2030. The forecast interval is greatest in the age range of under 20 years, since the number of these people come from the fertility forecast; while in the year 2030 the people in the age range of over 20 years had already been born by the first forecast year (2010) and therefore are no longer a source of additional uncertainty. The width of the confidence intervals additionally is determined by the variation of net migration, which will distinctly stabilise from 50 years (cf. Fig. 14), whereby the population from 50 years of age can be forecasted with greater accuracy.

5 Conclusions

The demographic future of a region is uncertain. The deterministic models frequently used by the official statistical offices for forecasts are of the "if ..., then ..." character, whereby the demographic determinants develop in different scenarios across predetermined assumptions. These should indicate the breadth of possible developments, however without making statements about their incidence probabilities. The stochastic approach, which uses time series models to calculate both the future population as well as a forecast interval for the demographic parameters, proves more problem-oriented. This approach, however, requires long time series and data broken down by years of age over all of the components of the basic demographic equation. At the regional level, though, such data are frequently not at all available, or only in rough age groups or with insufficient structure. Furthermore, the various statistical offices of the federal states compile the variables in different age groups; aggregation to a cross-regional area such as the Rhine-Neckar metropolitan area, which consists of urban and rural districts in Baden-Wuerttemberg, Hesse and Rhineland-Palatinate, on such a basis is in actuality impossible. The use of functional data proves exceedingly helpful for these problems, as per assumption age-grouped secondary data follow a functional context that can be estimated with the presented FDA models. In this way, data sets are created broken down by years of age that can therefore be aggregated to cross-border functional regions. Hence the paradigm of functional data and the models based on it close data gaps in secondary statistics, particularly at the regional level, and allow for more reliable statements about demographic development than the use of deterministic approaches. Without the approaches from Chapter 3, stochastic models are not implementable at the regional level. Moreover, the use of forecast intervals on the basis of the time series models provide more valuable additional information for regional planners than deterministic forecasts. At the regional level in particular, there is a great need for information about population development to serve as the basis for estimating the consequences of demographic change, for example on the labour supply and particularly on the number of the working population. Furthermore, regional population forecasts serve as the basis for estimating the future demand for living space.

The application of the models for functional data on the Rhine-Neckar metropolitan area illustrates the consequences of demographic change for the seventh-largest economic area in Germany and is a possibility for realizing stochastic population forecasts at the regional level. Although the population will only decrease slightly by 2030, a considerable drop in the number of young people of working age under 30 years will be contrasted by a notable rise in people in the upper age groups and particularly in retirement age. This shortage can be alleviated by immigration. The basic prerequisite for this is to recognise the strengths of the region and to create incentives for young people to move to the region and also to remain there following their studies or training. The impacts of demographic change will be particularly noticeable in the labour market. Regions will be competing for young and well-educated employees. Future need for research will result from this “war for talent”: stochastic population forecasts can be used as a basis for estimating the development of the labour supply. However, many existing forecasts are based on deterministic models since factors relevant for such things as the employment market, such as the unemployment rate, underlie political influences and can therefore not be readily explained by time series models. It is therefore essential that methods be developed that can take these influences into account (for example in generating assumptions) and harmonise the advantages of both approaches. The paradigm of functional data can make a valuable contribution to this objective.

References

Alho, Juha; Spencer, Bruce 2005: Statistical demography and forecasting. New York: Springer.

Bähr, Jürgen; Jentsch, Christoph; Kuls, Wolfgang 1992: Bevölkerungsgeographie. Berlin: De Gruyter.

Birg, Herwig 2005: Grundkurs Demographie – Dritte Lektion: Was taugen Prognosen? In: Frankfurter Allgemeine Zeitung 46, 24.02.2005: 41.

Booth, Heather 2006: Demographic forecasting: 1980 to 2005 in review. In: International Journal of Forecasting 22,3: 547-581 [doi:10.1016/j.ijforecast.2006.04.001].

Börsch-Supan, Axel 2007: Gesamtwirtschaftliche Folgen des demographischen Wandels. In: Geographische Rundschau 2: 48-52.

Börsch-Supan, Axel; Wilke, Christina Benita 2009: Zur mittel- und langfristigen Entwicklung der Erwerbstätigkeit in Deutschland. In: Zeitschrift für ArbeitsmarktForschung (ZAF) 42,1: 29-48 [doi: 10.1007/s12651-009-0006-x].

Brillinger, David 1986: The Natural variability of vital rates and associated statistics. In: Biometrics 42: 693-734.

Bundesamt für Bauwesen, Raumordnung und Städtebau 1995: Raumordnungspolitischer Handlungsrahmen. Bonn.

Büttner, Thiess; Ebertz, Alexander 2009: Quality of Life in the Regions - Results for German Counties. In: Annals of Regional Science 43,1: 89-112 [doi: 10.1007/s00168-007-0204-9].

Carnes, Bruce; Olshansky, Jay 2007: A Realistic View of Aging, Mortality and Future Longevity. In: Population and Development Review 33,2: 367-381.

Chambers, Elizabeth et al. 1998: The War for Talent. In: McKinsey Quarterly 3. [http://www.executivesondemand.net/managementsourcing/images/stories/artigos_pdf/gestao/The_war_for_talent.pdf, 09.06.2011.

De Jong, Gordon 2000: Expectations, gender, and norms in migration decision-making. In: Population Studies 54: 307–319.

Erbas, Bircan; Hyndman, Rob; Gertig, Dorota 2007: Forecasting age-specific breast cancer mortality using functional data models. In: Statistics in Medicine 26,2: 458–470 [doi: 10.1002/sim.2306].

Fuchs, Johann; Dörfler, Katrin 2005a: Projektion des Erwerbspotenzials bis 2050 – Annahmen und Datengrundlage. In: IAB Forschungsbericht 20/2005. Nürnberg: IAB.

Fuchs, Johann; Dörfler, Katrin 2005b: Projektion des Arbeitsangebots bis 2050 – Demographische Effekte sind nicht mehr zu bremsen. In: IAB Kurzbericht 11/2005. Nürnberg: IAB.

Gans, Paul; Schmitz-Veltin, Ansgar 2006: Die Auswirkungen des demographischen Wandels auf den Agglomerationsraum Rhein-Neckar – Ausgewählte regionale Entwicklungsbereiche vor dem Hintergrund sich ändernder Bevölkerungs- und Wirtschaftsstrukturen In: *Gans, Paul; Schmitz-Veltin, Ansgar* (Eds.): Demographische Trends in Deutschland. Folgen für Städte und Regionen. Räumliche Konsequenzen des demographischen Wandels, Teil 6. Forschungs- und Sitzungsberichte der ARL 226. Hannover: 311-335.

Gans, Paul; Schmitz-Veltin, Ansgar; West, Christina 2009: Bevölkerungsgeographie. Braunschweig: Westermann.

Gustafsson, Siv 2001: Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe. In: Population Economics 14: 225-247.

He, Xuming; Ng, Pin 1999: COBS: qualitatively constrained smoothing via linear programming. In: Computational Statistics 14: 315–337.

Heckman, James; Walker, James 1990: The Relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data. In: Econometrica 58,6: 1411-1441.

Höhn, Charlotte et al. 1987: Mehrsprachiges demographisches Wörterbuch. Deutschsprachige Fassung. Schriftenreihe des Bundesinstituts für Bevölkerungsforschung, Sonderband 16. Wiesbaden/Boppard: Boldt.

Hyndman, Rob; Booth, Heather 2008: Stochastic population forecasts using functional data models for mortality, fertility and migration. In: International Journal of Forecasting 24,3: 323–342 [doi: 10.1016/j.ijforecast.2008.02.009].

Hyndman, Rob; Ullah, Shahid 2007: Robust forecasting of mortality and fertility rates: a functional data approach. In: Computational Statistics & Data Analysis 51: 4942–4956 [doi: 10.1016/j.csda.2006.07.028].

Hyndman, Rob; Booth, Heather; Yasmeen, Farah 2011: Coherent mortality forecasting: the product-ratio method with functional time series models. EBS Working Paper [http://robjhyndman.com/papers/coherentfdm].

Keilman, Nico; Pham, Dinh Quang; Hetland, Arve 2002: Why population forecasts should be probabilistic – Illustrated by the case of Norway. In: Demographic Research 6: 409–453 [doi: 10.4054/DemRes.2002.6.15].

Keyfitz, Nathan 1972: On Future Population. In: Journal of the American Statistical Association 67: 347-363.

Keyfitz, Nathan 1981: The Limits of Population Forecasting. In: Population and Development Review 7,4: 579-593.

Klein, Torsten 2008: Kampf um Talente – Das Kind ist (fast) in den Brunnen gefallen. In: HR Performance 6: 44-47.

Lee, Ronald 1993: Modeling and forecasting the time series of U.S. fertility: age distribution, range, and ultimate level. In: International Journal of Forecasting 9: 187–202.

Lee, Ronald 1999: Probabilistic approaches to population forecasting. In: *Lutz, Wolfgang; Vaupel, James; Ahlburg, Dennis* (Eds.): Frontiers of population forecasting. A supplement to Population and Development Review 24: 156 –190 [doi: 10.2307/2808055].

Lee, Ronald 2000: The Lee-Carter method for forecasting mortality, with various extensions and applications. In: North American Actuarial Journal 4,1: 80–92.

Lee, Ronald; Carter, Lawrence 1992: Modeling and forecasting US mortality. In: Journal of the American Statistical Association 87,419: 659–671.

Lee, Ronald; Nault, Francois 1993: Modeling and forecasting provincial mortality in Canada. Paper presented at the World Congress of the IUSSP, Montreal, Canada.

Lee, Ronald; Tuljapurkar, Shripad 1994: Stochastic Forecasts for the United States: Beyond High, Medium, and Low. In: Journal of American Statistical Association 89,428: 1175-1189.

Li, Nan; Lee, Ronald 2005: Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. In: Demography 42,3: 575–594 [doi: 10.1353/dem.2005.0021].

Lipps, Oliver; Betz, Frank 2003: Stochastische Bevölkerungsprognose für West- und Ostdeutschland. MEA Discussion Paper 41. Mannheim: MEA.

Lipps, Oliver; Betz, Frank 2004: Stochastic Population Projection for Germany – based on the Quadratic Spline approach to modelling age specific fertility rates. MEA Discussion Paper 59. Mannheim: MEA.

Lowack, Wolf-Rainer 2007: Die Europäische Metropolregion Rhein-Neckar – Auftrag, Vision und Umsetzung. In: UWF – UmweltWirtschaftsForum 15,3: 130-135 [doi: 10.1007/s00550-007-0042-3].

Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei 1996: World population scenarios for the 21st century. In: *Lutz, Wolfgang* (Eds.): The future population of the world: What can we assume today? London: Earthscan: 361-396.

Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei 1998a: Expert-based probabilistic population projections. In: Population and Development Review 24, Supplement: Frontiers of Population Forecasting: 139-155 [doi: 10.2307/2808054].

Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei 1998b: An expert-based framework for probabilistic national population projections: The example of Austria. In: European Journal of Population 14: 1-17.

Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei 2001: The end of world population growth. In: Nature 412: 543-545.

Metropolregion Rhein-Neckar 2011: Jahresbericht 2010 zur gemeinschaftlichen Regionalentwicklung. In: www.m-r-n.de 9.2.2012. URL: http://www.m-r-n.com/fileadmin/PDF-Downloads/MRN_Jahresbericht_2010_01.pdf.

Moore, Andrew; Schneider, Jeff; Deng, Kang 1997: Efficient Locally Weighted Polynomial Regression Predictions. In: *Kaufmann, Morgan* (Eds.): Proceedings of the Fourteenth International Conference on Machine Learning. San Francisco, CA, USA: 236-244.

O'Neill, Brian et al. 2001: A Guide to Global Population Projections. In: Demographic Research 4: 203-288 [doi. 10.4054/DemRes.2001.4.8].

Ott, Thomas; Schmitz-Veltin, Ansgar; West, Christina 2006: Kinderwunsch – immer später und oft unerfüllt. In: *Institut für Länderkunde* (Eds.): Nationalatlas Bundesrepublik Deutschland, Bd. 12, Leben in Deutschland. Heidelberg/Berlin: Spectrum: 48-49.

Pflaumer, Peter 1988: Confidence Intervals for Population Projections based on Monte Carlo Methods. In: International Journal of Forecasting 4: 135-142.

Ramsay, Jim 2008: Functional Data Analysis. URL: http://gbi.agrsci.dk/~shd/public/FDA2008/FDA_Sage.pdf, 31.05.2011.

Ramsay, Jim; Silverman, Bernard 2001: Functional Data Analysis. In: *Smelser, Neil*; *Baltes, Paul* (Eds.): International Encyclopedia of the Social and Behavioral Sciences. Amsterdam: Elsevier: 5822-5828.

Rogers, Andrei 1985: Regional Population Projection Models. Beverly Hills: Sage.

Sanderson, Warren 1995: Predictability, complexity, and catastrophe in a collapsible model of population, development, and environmental interactions. In: Mathematical Population Studies 5,3: 259-279.

Schnabel, Sabine; Kistowski, Kristin von; Vaupel, James 2005: Immer neue Rekorde und kein Ende in Sicht: der Blick in die Zukunft lässt Deutschland grauer aussehen als viele erwarten. In: Demografische Forschung Aus Erster Hand 2,2: 3-3.

Statistisches Bundesamt 2009: 12. koordinierte Bevölkerungsvorausberechnung – Annahmen und Ergebnisse. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt 2010: Statistisches Jahrbuch 2010 für die Bundesrepublik Deutschland. Wiesbaden: Statistisches Bundesamt.

Stoto, Michael 1983: The accuracy of population projections. In: Journal of the American Statistical Association 78: 13–20.

Tukey, John Wilder 1977: Exploratory Data Analysis. London: Addison-Wesley.

Whelpton, Pascal 1928: Population of the United States, 1925 to 1975. In: American Journal of Sociology 34: 253-270.

Wilson, Chris 2001: On the scale of Global Demographic Convergence 1950-2000. In: Population and Development Review 27,1: 155-171 [doi: 10.1111/j.1728-4457.2001.00155.x].

Wood, Simon 1994: Obtaining Birth and Mortality Patterns From Structured Population Trajectories. In: Ecological Monographs 64,1: 23-44 [doi: 10.2307/2937054].

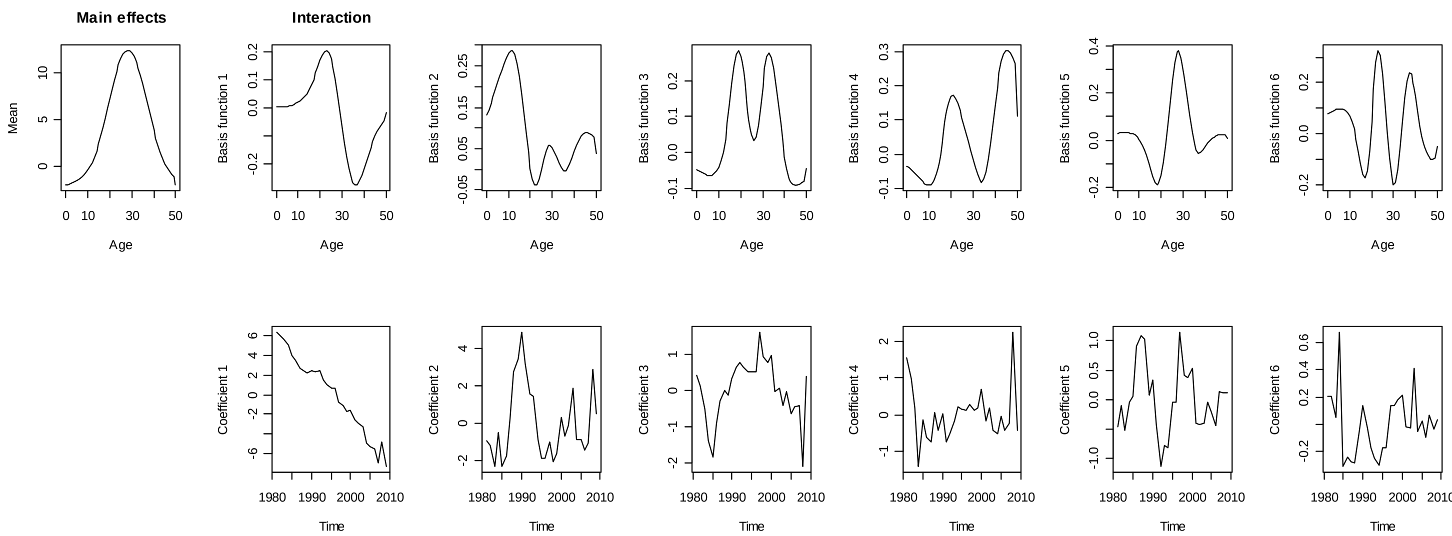
Translated from the original text by the Federal Institute for Population Research, for information only. The reviewed and author's authorised original article in German is available under the title "Die Bevölkerungsentwicklung der Metropolregion Rhein-Neckar. Eine stochastische Bevölkerungsprognose auf Basis des Paradigmas funktionaler Daten", DOI 10.4232/10.CPoS-2011-21de or URN urn:nbn:de:bib-cpos-2011-21de5, at http://www.comparativepopulationstudies.de.

Date of submission: 29.08.2011 *Date of Acceptance: 25.04.2012*

Philipp Deschermeier (✉). Universität Mannheim, Lehrstuhl für Wirtschaftsgeographie, Fakultät für Rechtswissenschaft und Volkswirtschaftslehre. Mannheim, Germany.
E-Mail: philipp.deschermeier@uni-mannheim.de
URL: http://gans.vwl.uni-mannheim.de/1465.0.html

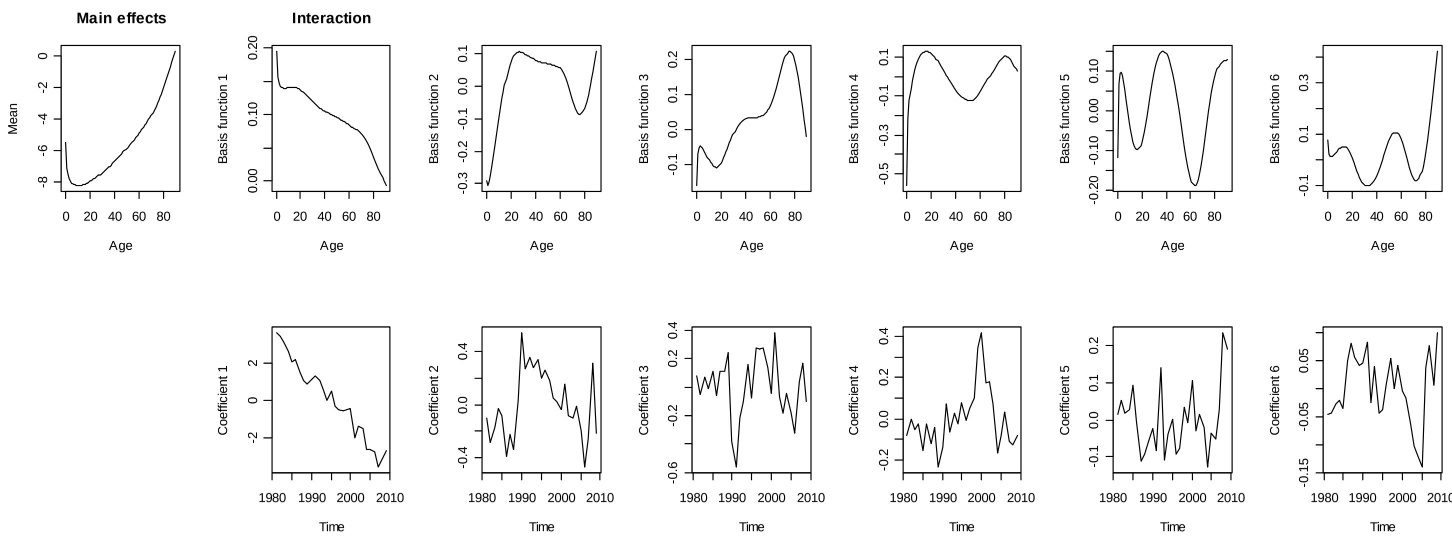
Appendix

Fig. A1: The six basis functions of the fertility models



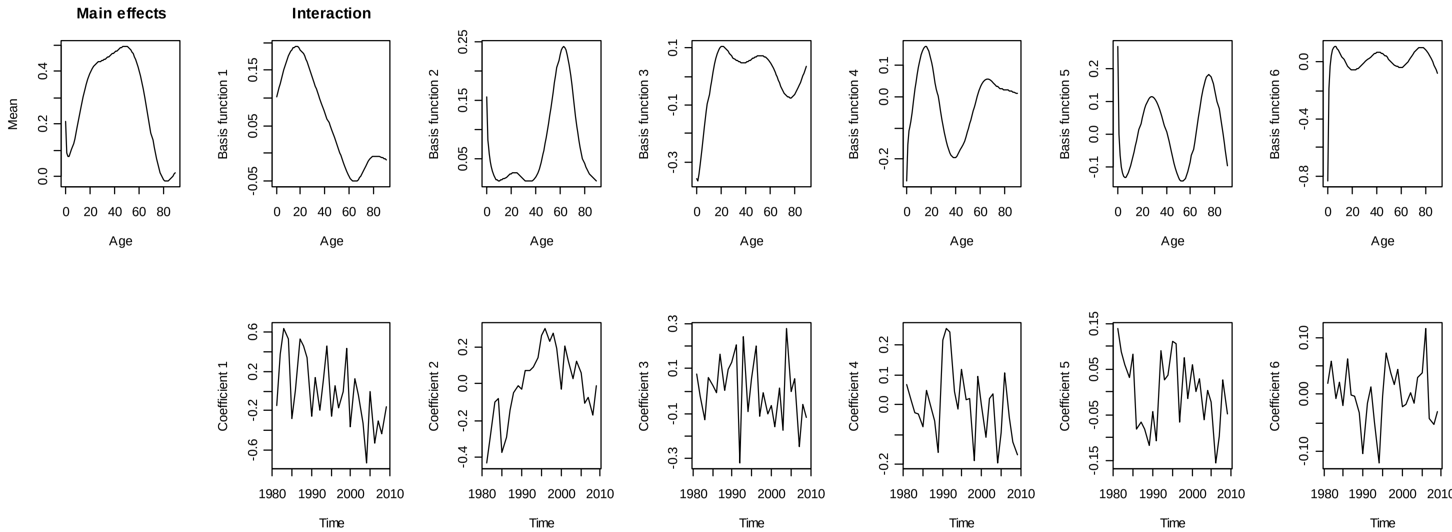
Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. A2: The six basis functions of the product model on mortality



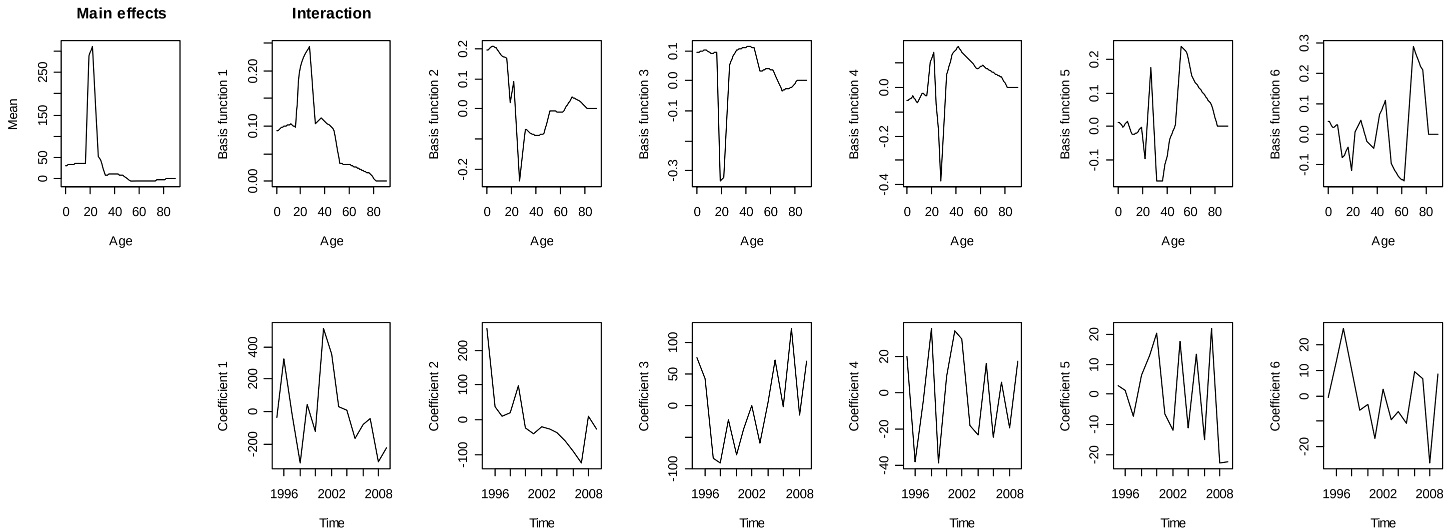
Source: Author’s calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. A3: The six basis functions of the ratio model on mortality



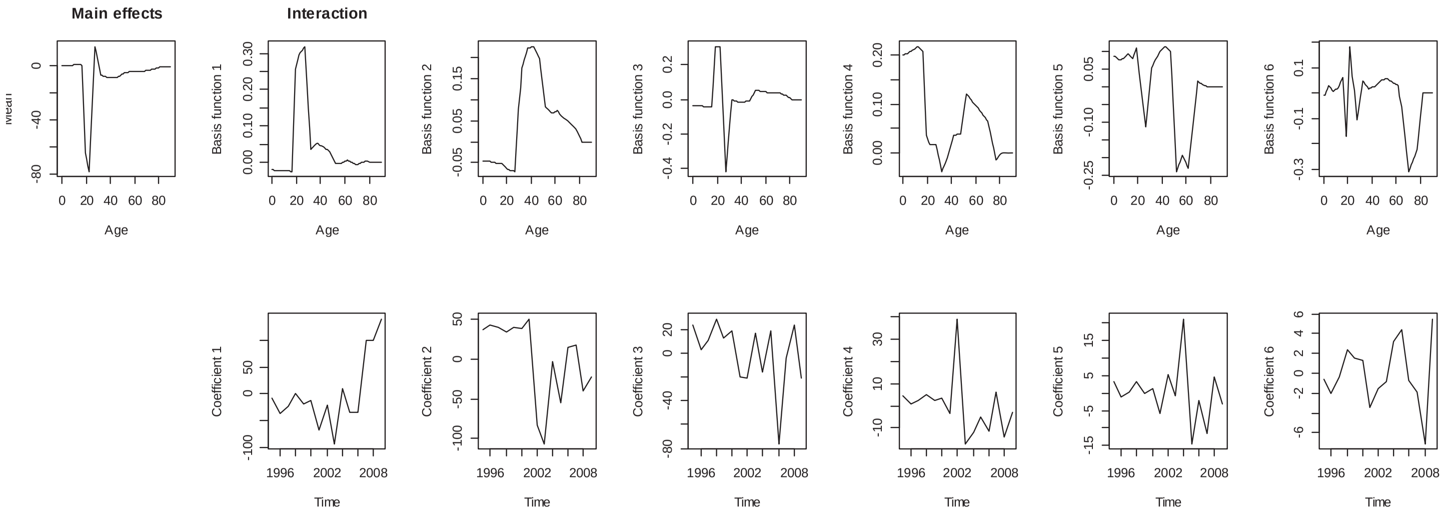
Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. A4: The six basis functions of the product model on migration



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Fig. A5: The six basis functions of the ratio model on migration



Source: Author's calculation on the basis of the data from the statistical offices of the federal states Baden-Wuerttemberg, Hesse and Rhineland-Palatinate.

Comparative Population Studies – Zeitschrift für Bevölkerungswissenschaft

www.comparativepopulationstudies.de

ISSN: 1869-8980 (Print) – 1869-8999 (Internet)

Published by / Herausgegeben von
Prof. Dr. Norbert F. Schneider

Federal Institute for Population Research
D-65180 Wiesbaden / Germany

Managing Editor / Verantwortlicher Redakteur
Frank Swiaczny

Editorial Assistant / Redaktionsassistenz
Katrin Schiefer

Language & Copy Editor (English) / Lektorat & Übersetzungen (englisch)
Amelie Franke

Copy Editor (German) / Lektorat (deutsch)
Dr. Evelyn Grünheid

Layout / Satz
Beatriz Feiler-Fuchs

E-mail: cpos@destatis.de