The Total Fertility Rate in Germany until 2040 – A Stochastic Principal Components Projection based on Age-specific Fertility Rates

ABSTRACT

Demographic change is one of the greatest challenges faced by Germany as well as a large part of Europe today. One of the main drivers of this change is the low fertility level, often referred to as the Total Fertility Rate (TFR), since the early 1970s. Therefore, on the one hand, while the total population is expected to decline, on the other hand, the relative share of the elderly in the total population is expected to increase. This poses a great challenge for the society in a wide range of aspects, most notably the statutory pension fund.

Therefore, it is important to gain an understanding about the future demographic development, in our case, the course of the TFR. Official forecasts often assume that the TFR would remain at a low level of 1.4 in the long run, which already was proven wrong with the publication of the 2014 data, which shows a TFR of 1.47. However, separate analyses of age-specific fertility lead to expected increases of the future TFR.

This study presents a stochastic projection of the TFR based on econometric-statistical modeling of age-specific fertility rates over principal components. Simulation techniques not only generate the expected future TFR until the year 2040, but also provide point-wise prediction intervals, which cover the future TFR with a probability of 95% annually, based on the current data set.

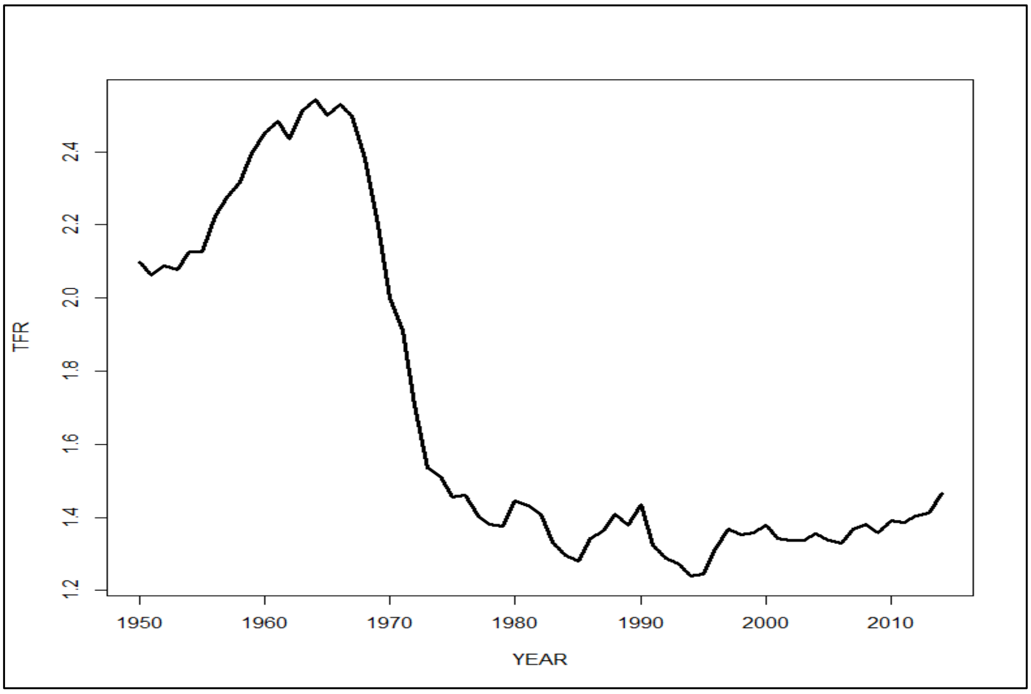
The age-specific structure of the modeling procedure gives a detailed insight on the future development of the reproductive behavior for women in Germany, and therefore, is very informative with regard to possible political intervention with the scope of increasing the TFR. Moreover, the flexible structure of the model allows more sophisticated estimation of future outcome of certain political measures.

Keywords: Total Fertility Rate; Demographic Modeling; Population Projection; Principal Component Analysis, Time Series Analysis

1 INTRODUCTION

One of the major obstacles confronted by Germany as well as large parts of the European continent is the expected decrease of the population along with an aging of the age structure. On April 28, 2015, the German Federal Statistics Office, *Destatis*, published its 13th coordinated population forecast. According to this forecast, the total population is expected to decrease and range between 67.6 million to 73.1 million by 2060 (Statistisches Bundesamt 2015: 15). Taking into account the expectation that the share of the elderly population among this population will increase further (Statistisches Bundesamt 2015: 18), demographic change is one of the biggest challenges the statutory pension fund, as well as a wide range of other fields, is confronted with.[1] Because of the pending future problems, which may arise from the demographic change, it is highly important to visualize the future demographic structure and take appropriate actions at the right time to deal with those changes, especially in politics. The first step in such an undertaking should be the appropriate modeling and projection of the demographic change using mathematical and statistical methods. Since the beginning of the 1970s, the main driver of demographic change can be identified as the low fertility rate (Hyndman; Booth: 323-324). An important statistic in this context is the Total Fertility Rate (*TFR*), which measures the hypothetical number of children a woman will bear on average during the course of her reproductive phase given the current age-specific fertility rates *(ASFR)* would remain stable on their current level. Low fertility rates imply small numbers of births, and thus, few young people adding to the population. This leads to two results. First, the share of the elderly population will increase, since mortality gets lower and therefore people get older, whereas few people are born. Second, the death rate is higher than the birth rate, which results in a ceteris paribus *(c.p.)* decline of the population (Bujard 2015: 136-140). This underlines the undeniable importance of the level of birth rates in demographics. Figure 1 shows the development of the TFR in Germany from 1950 to 2014:

Figure 1: Time Series of TFR in Germany



Source: Own calculation and design based on Destatis and UN data

---

[1] Bujard 2015: 145-154 gives a nice overview.

It can be seen that the TFR rose to an approximate level of 2.5 in the early 1960s. At the end of this decade, it subsequently decreased. The TFR reached its minimum in 1994, where it fell to 1.24. Since then, it established itself at a level of around 1.4 with a slightly positive tendency in the current millennium.

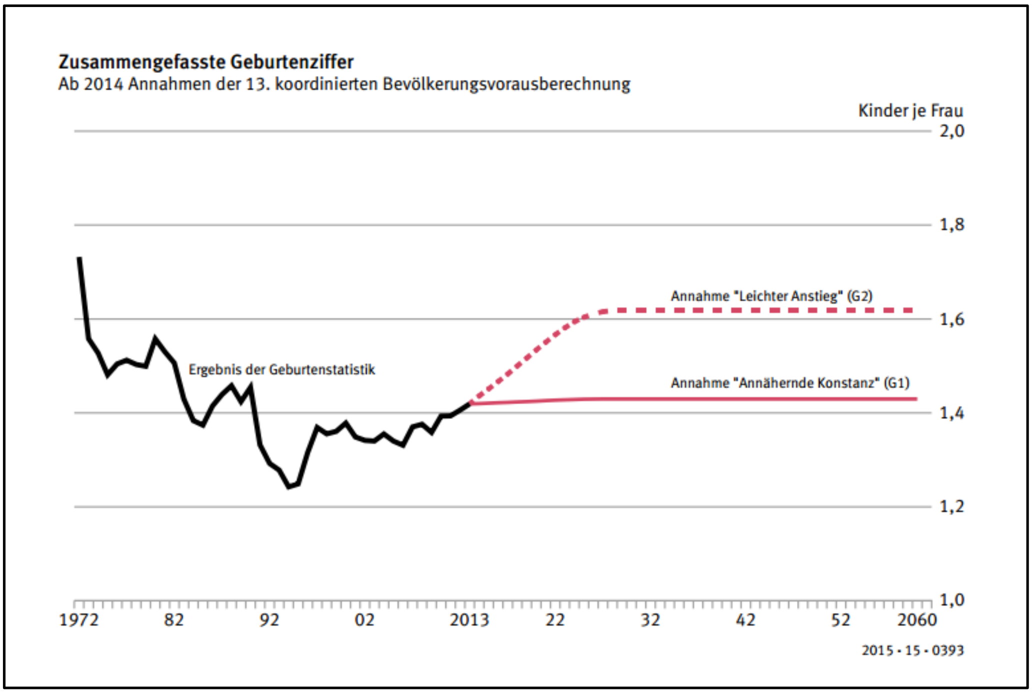
The importance of family politics in increasing the TFR is undeniable, although the effects of certain measures cannot be identified definitely (Bujard 2011: 7-19). The German government realized the issue of low fertility and the importance of family politics for dealing with this issue early. Moreover, it has been trying to motivate women to conceive more children for decades by continuously increasing child benefits as well as passing new laws since the 1980s, guaranteeing a financial compensation for the damage parenting time means for the household, such as the so-called *Elterngeld* and *Betreuungsgeld* (which has been discontinued lately). Since August 2013, there exists a guarantee for a day-care nursery for every child over three years of age, which came into force in the so-called *Kinderförderungsgesetz (KiföG)* in 2008. This shows the increasing awareness of the importance of giving both parents the flexibility of being at work relatively quickly after childbirth, and therefore, the high priority of coping with low fertility in family politics.

The main goal of this study is to present a stochastic projection of the TFR for Germany until 2040 through statistic modeling based on empirical data. Since the TFR is the composition of age-specific birth rates, it can be estimated more accurately by modeling the age-specific fertility rates. This approach will lead to different forecasts whilst giving more information on different groups in the population.

2 GERMAN AND INTERNATIONAL FERTILITY FORECASTING

Common population forecasts in Germany are usually based on deterministic models. The most common one is the cohort-component method by Cannan (1895), Bowley (1924), and Whelpton (1928), which defines fertility, mortality, and migration as the three components that change the population from year *t* to year *t+1*, taking into account the age-specific mortality rates and the aging of the surviving persons by one year from *t* to *t+1*. Assume that the population at year-end 2014 is *P* with the number of deaths during 2015 being *D*, the number of births being *B,* and the number of net migrants being *M*. Then, the population at year-end 2015 will be equal to the value given by *P + B + M - D*. Therefore, it is possible to either state assumptions for the development of *B, M,* and *D* directly or make assumptions on the rates based on subpopulations discriminated by sex and age, for example. Regarding fertility, the TFR can be forecast directly and the number of births can be derived from the TFR. Destatis states two assumptions for the future development of the TFR (as well as for the net migration and the life expectancy) for scenario analysis purposes. The base assumption G1 states that the TFR will remain at its current approximate level of 1.4 in the long run, based on the relatively stable (which is not entirely true, as we have seen in section 1) TFR during the near past. Furthermore, it is assumed that the reproduction rates for women under 30 years of age will decrease further, whereas those for women over 30 will increase with upper boundaries equal to the current highest fertility rates in Europe for these age groups, which can be found in Sweden. These assumptions are supposed to pan out at the current level cumulated over all age groups. Destatis forecasts G1 as the probable development. Assumption G2 states a more optimistic scenario with the TFR increasing linearly to an ultimate level of 1.6 in 2028 and remaining at this level after that point in time. Although the official forecasts do not provide any empirical evidence for this increase, the increase can be categorized as a possible, yet quite improbable alternative. However, during the course of this paper, we will see that this scenario is not as improbable as it is being thought of. The average child-bearing age is assumed to increase to 31.8 in G1 and to 31.4 in G2 in 2028, and remain constant thereafter in both scenarios (Statistisches Bundesamt 2015: 31-33). Figure 2 illustrates the course of the TFR for the two alternatives.

Figure 2: TFR forecast until 2060 by Destatis

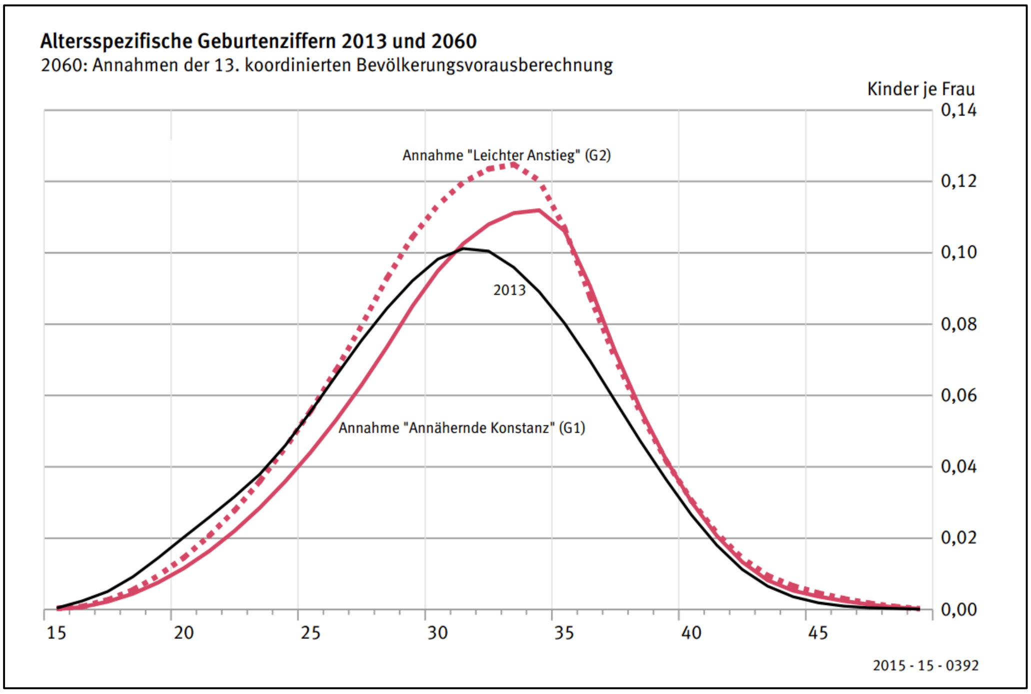


Source: Statistisches Bundesamt 2015: 32

Since the population preserving TFR is estimated at about 2.1 (Espenshade; Guzman; Westoff 2003: 575), both scenarios imply a significant decline in population size, still depending on the mortality and migration levels.

As indicated earlier, there will be a shift in fertility from the younger age groups to the older ones, which is of high importance with regard to the demographic structure in Germany. Since the absolute numbers of women in the older age groups will be relatively high compared to those in the younger age groups, fertility rates among women over 30 may increase, possibly indicating a shift in childbearing to a later point in life. The estimated ASFR by Destatis in 2060 are illustrated for the two scenarios in figure 3:

Figure 3: Age-specific fertility rates 2013 and 2060

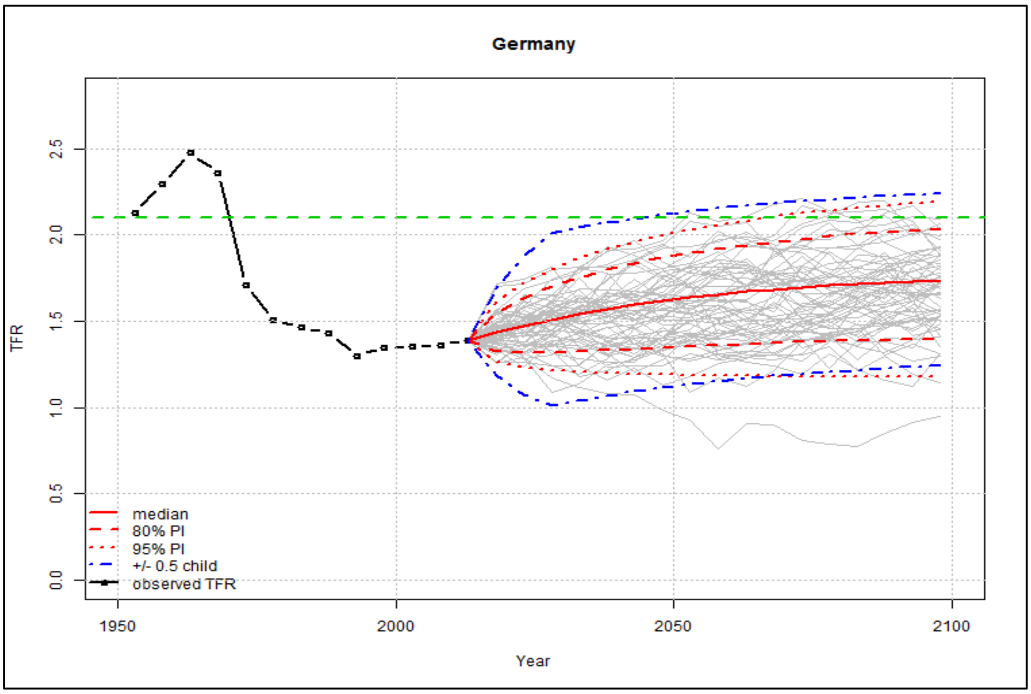


Source: Statistisches Bundesamt 2015: 33

In both scenarios, it is expected that there will be a fertility shift to older ages as well as more births centered between ages 30 and 35, with higher peaks for the birth rates compared to 2013 and a steeper and narrower distribution pattern. Today, since women on average conceive their first child later than they did in the past, there is a great discussion on whether high female education and income have a negative influence on the average number of children (Birg 2001: 33). This question is fundamental because of the constantly increasing share of women seeking academic education after school (Bujard 2012: 3). The alternative theory assumes that these developments only shift the average age in which women conceive their first child to a higher age, but do not influence the total number of children. Based on a time series analysis with EU data, Engelhardt could reject the hypothesis of a negative correlation between female labor rates and fertility rates since the mid-1980s (Engelhardt 2009: 252-260). Using an own estimation model and investigating micro census data, Bujard was able to show that the age effect for women with higher education indeed exists, especially in West Germany (Bujard 2012: 9-10). One might wonder whether it is possible influencing the TFR at all through political intervention. Dorbritz in this context analyzed data from the Population Policy Acceptance Study carried out in 2004 (*PPAS*). Among the German subpopulation aged 20-39, asked for the reasons for not wanting a or another child, more than one fifth of the women declared their state of health wouldn't allow it, around 40% indicated their job being incompatible with having more children. Around 40% of the women also classified the high marginal costs of more children as an important reason (Dorbritz 2008: 585). This gives a strong indication that a big part of the women might be convinced of conceiving more children through better health conditions, which can be expected to become better through ongoing medical advances or better nutritional awareness (Statistisches Bundesamt 2015: 34-35), better job conditions for parents or more statutory financial aid. This indicates that a future increase in fertility is realistic and could be supported by statutory measures and better labor conditions. Bujard (2011) made use of a multivariate analysis for 28 OECD countries and found strong statistical evidence for a positive effect of different family political measures on the TFR.

Returning to the discussion on which kind of model to use for forecasting purposes, it is important to consider the advantages and disadvantages of two main categories: deterministic and stochastic models. Deterministic models have the advantage of easy implementation and calculation, which allows us to gain a very quick idea about the (possible) future development of the population. Despite this advantage, it appears that using stochastic modeling is recommendable. Both approaches rely on the assumptions made, but the main disadvantage of deterministic models is that they cover only specific scenarios that have low occurrence probabilities. Stochastic models, on the other hand, not only quantify expected scenarios, but also identify a broad range of possible outcomes with their respective probabilities (Raftery et al. 2012: 13915). This opens up the option to specify a complete distribution of each possible scenario at some point in the future, including its probability. On an international level, these kinds of modeling approaches have been used for some time. For example, the United Nations (UN) provide an online tool that shows projections for the TFR for a great range of countries, including Germany. Such projections are based on stochastic simulations using a Bayesian Hierarchical approach for creating assumptions regarding the future development of fertility rates until 2100 with 80% and 95% prediction intervals (PI).[2] A 95%-PI indicates the range of future outcomes which c.p. can be expected for the TFR with 95% probability in a certain year. Figure 4 illustrates the UN projection for Germany:

Figure 4: UN Projection for TFR in Germany



Source: United Nations 2015

The underlying model for the UN tool assumes three phases for the fertility development of every country. Phase I is defined as the pre-transition fertility phase. The TFR is very high in this phase and on average is assumed to remain stable on its current level. In phase II, the so-called fertility transition, a steady decrease of the TFR towards the international replacement-level fertility of 2.1 or below is assumed. The process of decline may vary for every country by shape and velocity. Phase III, the post-transitional phase, finally assumes a recovery from the TFR back towards 2.1. Phases II and III are modeled based on five-year observations from 1950 on for 196 countries. Since phase I assumes a

---

[2] The tool is provided by the United Nations Population Division via http://esa.un.org/unpd/wpp/Graphs/.

constant development it is not estimated further. Phase II and III however are estimated by specific parametric functions for all countries making use of the country-specific data in a Markov Chain Monte Carlo *(MCMC)* simulation (Alkema et al. 2011: 818-828). The functional approach itself shall not be explained further at this point. It is obvious that, based on these assumptions, Germany finds itself in the post-transitional phase right now. It follows that the TFR should recover slowly from its low actual level towards 2.1, which can be seen by the slow increase in figure 4. The prediction bands are based on an autoregressive model of order one, denoted AR(1), for the error term which moments are estimated by the MCMC algorithm and normality assumptions (Alkema et al. 2011: 827-828). AR models will be explained in the next section. The European Commission projects a quite similar course of the TFR to that of the UN, but with a steeper increase. It expects the TFR to be over 1.5 by 2030 and to reach a level of 1.6 in 2050 (European Commission 2014: 10).

Since the TFR is a non-negative entity, a logical alternative to modeling it directly is using its logarithmic transformation. The inverse function of a log function is an exponential function, which always takes positive values on the vertical axis. This transformation ensures that no negative values can be predicted for the TFR. A classic approach for stochastic projection of the log-TFR was proposed by Lee, which is based on the time series approach introduced by Box and Jenkins in 1976.[3] Lee fitted an autoregressive moving average model *(ARMA (1,1))* to American data for an indirect projection of the TFR until 2065 based on forecasting the first principal component *(PC)* of the matrix containing the ASFR. He assumed a fertility distribution among the age groups as well as a long-term average and an upper and a lower long-term bound for the TFR (Lee 1993: 194-199). The approach by Lee has major limitations, because it needs very strong assumptions about the long-term average, the upper and lower bound, and the age structure of the mothers, which makes one wonder, whether a statistical approach of this kind has major advantages over simple deterministic approaches at all. Furthermore, it is assumed the TFR will move toward a specific long-term average asymptotically. This means possible long-term divergence trends cannot be captured. Therefore the approach by Lee has been developed over the years by various statisticians. Keeping it short, there will be mentioned only a few of them, which are important for the context of this paper. Hyndman and Ullah (2007) generalized the Lee-Carter model to a robust functional forecasting model. Therefore they used a non-parametric smoothing approach for the original data (logged five-year grouped fertility rates for Australia during the time span 1921-2000) whilst taking outliers in the historic time series due to wars or other extraordinary events into account. Based on the adjusted data they calculated the PC, which they used for projecting the ASFR in Australia through 2020. Hyndman and Booth (2008) developed this approach further by generalizing the log-transformation through a Box-Cox transformation on the one hand and making use of the first six PC for forecasting instead of the first three like seen in the former model. This way they improved the forecasting accuracy. For Germany, Lipps and Betz used Box-Jenkins-modeling for separate projections for the TFR in West and East Germany (Lipps; Betz 2005: 3-44). Härdle and Myšičková used a similar approach with newer data for Germany as a whole (Härdle; Myšičková 2009: 9-15). Similar to Destatis, they, in expectation, c.p., project a constant TFR until 2060 (Härdle; Myšičková 2009: 15), not taking into account the dynamics during the last couple of years, which, as can be seen later, propose a quite different future TFR development. Bomsdorf et al. (2008) fitted a Beta distribution for the normalized ASFR in Germany for the ages 14-50. They assumed the TFR to behave as a random walk process without drift.[4] Essentially, this is the stochastic analogue to the assumption of a constant TFR. In expectation, c.p., it is assumed that the TFR remains constant on

---

[3] The method will be explained in more detail in section 4.

[4] The random walk process will be explained in section 4.

its current level and only fluctuates around the expectation due to a stochastic error term. Dudel used a non-parametric estimation approach separated for East and West Germany with data for West Germany from 1950 to 1989 and for the Federal Republic accumulated from 1990 to 2008 (quite similar to the data used in this paper). Based on this he ran 1,000 simulations for the TFR until 2059, regarding the median scenario as the expected TFR and identifying 90%-PI. He estimated the expected TFR in 2040 at 1.38 and at 1.4 in 2059. Moreover, Dudel estimated the TFR to fall within the interval [1.12;1.86] in 2040 and the interval [1.08;2.05] in 2059 with a 90% probability (Dudel 2014: 165-175). Those results are quite obsolete based on the actual trends, as we will see later. Major weaknesses of the model are the relatively small number of simulations and the restriction of the TFR to the interval [1;3]. This means he discarded simulated outcomes for the TFR under 1.0 and over 3.0 as implausible, which decreases the number of simulations indeed used for calculation of the PI even further. Furthermore the setting of such limits is quite arbitrary, too. Another alternative in forecasting the future TFR are micro simulations. Since the fertility is influenced heavily by exogenous variables, as mentioned above, we could estimate these influences over regression analysis and forecast the explanatory variables and simultaneously the TFR by simulation techniques (see e.g. Leim 2008 and Pavetić 2009). This approach appears quite sophisticated, but the major downside is the massive need for data for this. Furthermore we would need to project all explanatory variables in such a model, which would lead to a huge stochasticity due to the explanatory variables. Therefore the author does not see a major advantage in these models for forecasting purposes, although they are very interesting from an explanatory point of view.

Most official forecasts cover a long time horizon. As mentioned earlier, Destatis forecasts TFR until 2060, whereas the UN predicts it until 2100. However, the time horizon for the model used in this study will end as early as 2040. The reason for this is the relatively short time series for the underlying data. ASFR for Germany can only be tracked back to 1950, and even during the short time period from 1950–2014, structural breaks can be identified in the reproductive behavior for women in a certain age. There are no general rules on the minimum number of past observations for a given projection horizon (Hyndman; Kostenko 2007: 15). It appears reasonable not to choose a projection horizon longer than the underlying past data itself, however. Since there have been structural breaks due to the German reunification not all trends can be tracked before 1991, the author only considers the shorter time horizon through 2040. It does not appear plausible forecasting the TFR beyond that point reliably.

One might wonder whether the TFR is an appropriate measure for estimating the real reproductive behavior of the female population at all. The critics of the TFR argue that it consisted of two effects, namely the quantum and the tempo effect. The tempo effect describes the shifts in childbearing ages. The quantum component would be the observed TFR in the event there didn't exist tempo effects. The TFR is criticized for decades for being vulnerable with regard to distortions caused by the timing effects and could therefore be manipulated by shifts in timing of births (Bongaarts; Feeney 1998: 272-275), which e.g. could cause a low TFR in year t but instead a high one in t+1, because there have been many delayed births in t+1. The most prominent alternative to the TFR is the completed fertility rate (CFR). The CFR tracks the birth cohorts of the mothers, so a CFR of e.g. 2.0 would tell us, that the specific cohort on average indeed conceived 2 children over the course of its reproductive phase. Although the CFR is an observed statistic which tracks actual trends directly, it has the major disadvantage that it can only cover the older cohorts, which already have completed their reproductive phase. Therefore actual trends can be observed with a huge delay only (Bongaarts; Feeney 1998: 271), which renders intervention obsolete. Since the CFR in theory covers the tempo and quantum effect, some researchers think of derivations from the CFR as statistics. Schmertmann et al. (2014) developed a sophisticated Bayesian modeling approach for predicting the CFR, based on principal components[5]. They used past and already observed CFR on the one hand as well as data and valuations on cohorts currently in their reproductive phase on the other hand for estimation of the future CFR for the younger generations.

Nevertheless, since the TFR is most dominant statistic because of its easy interpretation and applicability as well as with regard to the obvious weaknesses inherent of the CFR, the author prefers a forecast of the TFR. Moreover, as we will see, the method presented in this paper covers tempo as well as quantum effects due to the age-specific modeling structure quite well. The approach proposed in the upcoming section makes use of principal components modeling as well as time series methods, like the approaches by Lee and Hyndman. Instead of the Hyndman model, parametric models are used for fitting and forecasting of the first three PC. No smoothing of the original data will be done and no robust method is used, since the author thinks that even outliers are important for correct projection of the range of possible outcomes through PI. Since the trends in fertility have different time horizons, the underlying history for PC fitting will be chosen individually for every PC. Other approaches use the same time horizon for all PC, which may lead to biased forecasts. The model proposed in this paper eliminates the effects of structural breaks possibly leading to wrong results.

3 DATA OVERVIEW AND METHODOLOGICAL ASPECTS

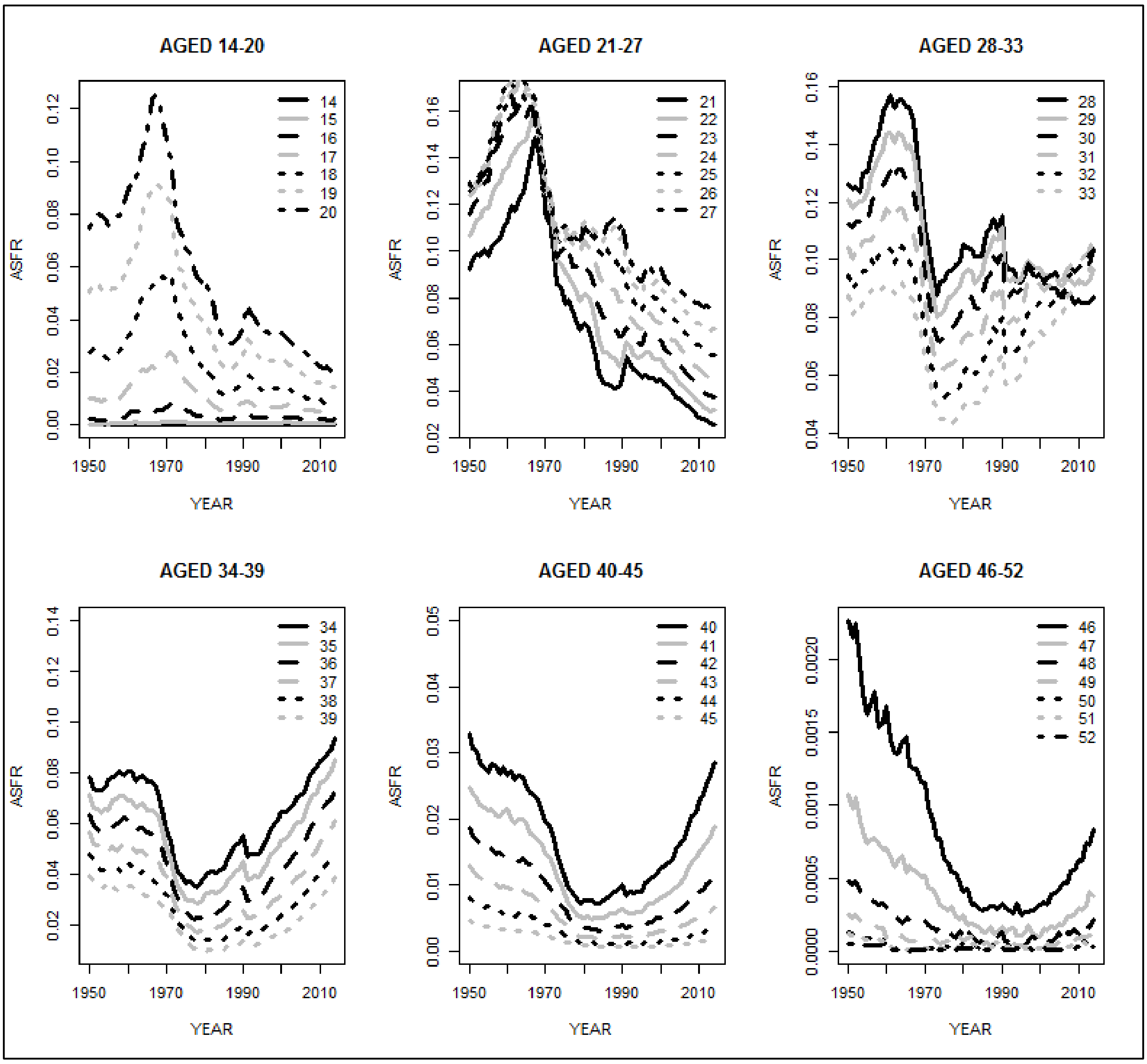
This study analyzes data provided by Destatis and the United Nations Population Division for the period 1950–2014. It should be noted that the data for 1950–1990 was restricted to West Germany only, whereas that after 1991 includes Germany as a whole. Thereafter, the age-specific female population was extracted for each year (United Nations 1950: 145; Statistisches Bundesamt 2015a; Statistisches Bundesamt 1953: 42; Statistisches Bundesamt 1954: 40; Statistisches Bundesamt 1955: 40; Statistisches Bundesamt 1956: 40; Statistisches Bundesamt 1957: 42; Statistisches Bundesamt 1958: 38; Statistisches Bundesamt 1959: 39; Statistisches Bundesamt 1960: 47; Statistisches Bundesamt 1961: 46; Statistisches Bundesamt 1962: 44; Statistisches Bundesamt 1964: 44; Statistisches Bundesamt 1965: 41; Statistisches Bundesamt 1966: 37; Statistisches Bundesamt 1967: 37; Statistisches Bundesamt 1968: 34; Statistisches Bundesamt 1969: 34; Statistisches Bundesamt 1970: 35; Statistisches Bundesamt 1971: 35; Statistisches Bundesamt 2015b). As mentioned earlier, many studies regard West and East Germany by themselves. By making this cut in the early 1990s, one might wonder, whether the data will lose its power. As we will see, this is not the case. Working with the ASFR cancels the absolute dimension. Furthermore, structural breaks were dealt with accordingly by using binary variables or using shorter historic data. We will see, that the reunification did not cause shocks which would account for long-term changes in the reproductive behavior in all cases. Furthermore, information on the number of births ordered by the age of the mother was used to calculate the ASFR (United Nations 1950: 303; Statistisches Bundesamt 1952: 38-39; Statistisches Bundesamt 1953: 60-61; Statistisches Bundesamt 1954: 58-59; Statistisches Bundesamt 1955: 58-59; Statistisches Bundesamt 1956: 56-57; Statistisches Bundesamt 1957: 58-59; Statistisches Bundesamt 1958: 52; Statistisches Bundesamt 1959: 50-51; Statistisches Bundesamt 1960: 64; Statistisches Bundesamt 1961: 64; Statistisches Bundesamt 2014; Statistisches Bundesamt 2015c).

---

[5] Since the method used in this paper is based on principal components too, the basic concept will be explained in section 3.

Based on this data, the ASFR were computed by dividing the age-specific number of births by the number of women in the specific age. Since there is no detailed information on the exact bearing-age of women over 49 but only on all mothers aged 49 or older cumulated, the missing data was extrapolated assuming the number of births for this group to be poisson distributed with λ=1 over the ages of the mothers. This seemed reasonable, since this distribution has a monotone decay and ensures lower birth numbers for bigger ages. Figure 5 illustrates the course of the ASFR for women aged 14–52.

Figure 5: Age-specific fertility rates in Germany, 1950–2014



Source: Own calculations and design based on data from Destatis and the UN

It is obvious that there is a general negative trend among almost all age groups from the late 1960s to the late 1970s, which can be identified by the strong decline in figure 1. The interesting part is the huge difference in the reproductive behavior among different age groups after that. In the young ages under 30, there still seems to be a moderate negative trend in reproduction opposed to the substantial decrease until the mid-1980s. For women in the age group around 30, the fertility rates seem quite stable since the early 1970s. For women over 30, there even exists a clear positive trend. Thus far, these trends support the assumptions for future development made by Destatis.

In the following part, the recent fertility trends for the age groups will be quantified simultaneously, and from these models, the probable future trends will be extracted subsequently. This will be the foundation on which the risk in the birth rates will be quantified and used for a simulation study targeting at modeling the future development of the TFR, in expectation and with point-wise prediction intervals, which should cover the future TFR with a certain probability.

Two problems attached with the projection of the future ASFR need to be addressed first, cross-correlation and autocorrelation. Taking a look at figure 5, we see a very similar reproductive behavior of the women in the age group 34-39, i.e. This means there exists an obvious correlation among these fertility rates. This issue is addressed by the term cross-correlation. This issue needs to be covered appropriately, for correct simulation of the future behavior of the ASFR. It would lead to biased results making assumptions about the future reproductive behavior of 34-year-old women ignoring the 39-olds, since there may be underlying effects, which influence both groups. The superior approach for dealing with this is the principal component analysis (PCA). This method makes use of an orthogonal transformation of the original variables into their principal components. These are new variables, which are uncorrelated. The second main advantage of a PCA is the fast identification of principal components, which can be rejected in the model because of their low relevance for the target variable. Therefore, models with many variables can be reduced to a small amount of dimensions, resulting in more simplicity of the model (Chatfield; Collins 1980: 57). The method shall be explained in short here. Let's say we have a bunch of ASFR we would like to transform into their principal components, which shall be denoted by PC. Each PC can be written as a linear combination of all the ASFR, like this:

$$PC_i = \sum_{j=1}^{J} a_{ij} \cdot ASFR_j \qquad \textbf{(1)}$$

With the a_{ij} denoting the coefficient of the jth ASFR on the PC under study. We call the vector $\overrightarrow{a_i}$ the first eigenvector of the covariance matrix, which corresponds to the ith PC. In matrix notation, the transformation can be denoted as:

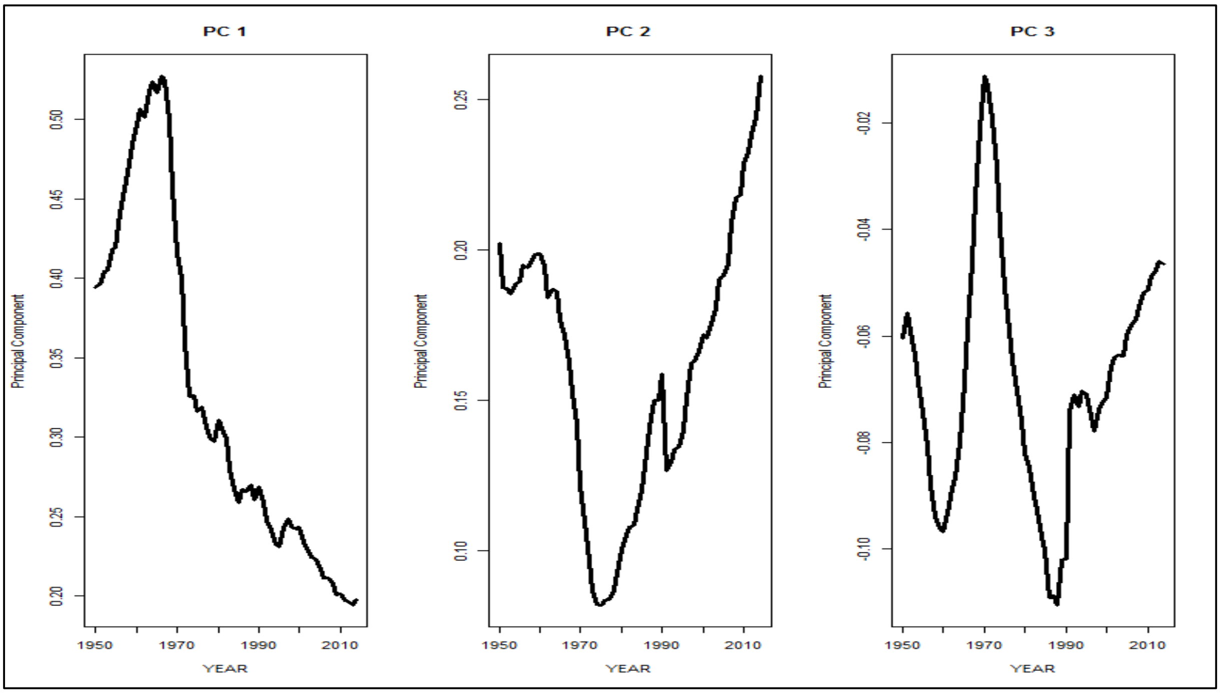
$$\boldsymbol{C} = \boldsymbol{F} \times \boldsymbol{A} \qquad \textbf{(2)}$$

In **(2)**, ***C*** denotes the matrix of the principal components, ***F*** is the matrix of the ASFR for some time interval and ***A*** means the matrix of the eigenvectors of the covariance matrix of the ASFR. The PCA chooses the PC decreasing by the amount of total variation in the original variables they explain. So in the first step, the vector $\overrightarrow{a_1}$ is chosen, such that it explains the maximum amount of variation in the ASFR. PC_2 is determined by the $\overrightarrow{a_2}$, which maximizes the explained share variation in the original variables under the condition that PC_2 is uncorrelated to PC_1 and so on (Chatfield; Collins 1980: 58-63). Using this procedure, we receive the 39 principal components for our original variables, descending by importance. The choice of how many of the PC to label as sufficient to describe the system is purely subjective. The author in this case calculated the cumulative share of the variance explained by the PC. Since the first three PC explain about 99% of the variance of the original variables, using the first three principal components for modeling should suffice for a good forecast. The associated eigenvectors of the first three PC are considered for better understanding:

$$\boldsymbol{a}_1 \approx \begin{pmatrix} .000 \\ .000 \\ .007 \\ .044 \\ .115 \\ .198 \\ .270 \\ .321 \\ .362 \\ .375 \\ .366 \\ .338 \\ .296 \\ .249 \\ .198 \\ .150 \\ .107 \\ .074 \\ .054 \\ .044 \\ .039 \\ .038 \\ .043 \\ .047 \\ .050 \\ .048 \\ .044 \\ .039 \\ .031 \\ .023 \\ .015 \\ .009 \\ .004 \\ .002 \\ .001 \\ .000 \\ .000 \\ .000 \\ .000 \end{pmatrix}, \qquad \boldsymbol{a}_2 \approx \begin{pmatrix} .000 \\ -.002 \\ -.016 \\ -.053 \\ -.088 \\ -.107 \\ -.101 \\ -.096 \\ -.090 \\ -.086 \\ -.078 \\ -.061 \\ -.026 \\ .027 \\ .091 \\ .164 \\ .234 \\ .301 \\ .341 \\ .358 \\ .355 \\ .336 \\ .307 \\ .261 \\ .215 \\ .170 \\ .130 \\ .090 \\ .058 \\ .035 \\ .020 \\ .010 \\ .005 \\ .002 \\ .001 \\ .000 \\ .000 \\ .000 \\ .000 \end{pmatrix}, \qquad \boldsymbol{a}_3 \approx \begin{pmatrix} .001 \\ .005 \\ .034 \\ .134 \\ .263 \\ .347 \\ .348 \\ .291 \\ .176 \\ .027 \\ -.127 \\ -.242 \\ -.307 \\ -.319 \\ -.287 \\ -.229 \\ -.155 \\ -.076 \\ -.009 \\ .055 \\ .102 \\ .129 \\ .145 \\ .144 \\ .131 \\ .116 \\ .092 \\ .066 \\ .043 \\ .028 \\ .015 \\ .008 \\ .004 \\ .002 \\ .001 \\ .000 \\ .000 \\ .000 \\ .000 \end{pmatrix}$$

These vectors contain the coefficients in equation **(1)** or equivalently the columns in **A**. The first eigenvector has relatively big positive components between row 5 (18-year-old women) and 13 (30-year-old women). So we might interpret the first PC as a variable which is strongly positive correlated with young women at the beginning of her working life. The second PC is quite strongly positively correlated with women between age 28 and 40, roundabout the age group of women in the working age after leaving the university. The third PC is negatively correlated with young women under 22 and becomes negatively correlated for women in the mid-20s. It might therefore cover trends over different life-courses balancing each other. The hypothetical historic time series of the PC are plotted in figure 6:

Figure 6: Historic course of PC



Source: Own calculations and design

The first PC visualizes the general degressive negative trend, which especially for the younger age groups can be observed since the late 1960s. The second PC shows the increasing timing effect, which in detail shows the positive trend mentioned earlier with more and more women conceiving their children at later stages in their life course. We observe a structural break in 1991 caused by the reunification and change in the data base. This only causes a general lowering in the total level, but does not influence the long-term trend itself. The third PC shows a real structural break in the early 1990s though, therefore only the data from 1991 on will be used for projection of this principal component.

The second issue to be addressed is the so-called autocorrelation, describing the correlation between a time series and its own past observations. Autocorrelation may be dealt with using Box-Jenkins time series models. The two basic types, which will be of importance later, shall be introduced here. The easiest model type is the so-called Moving Average model of order q, abbreviated MA(q):

$$x_t = \omega_t + \sum_{i=1}^{q} \theta_i \omega_{t-i} \qquad \textbf{(3)}$$

with $\omega_t \sim \mathcal{NID}(0, \sigma^2)$ being stochastic shocks and θ_t being time-specific coefficients for the regressors (Shumway; Stoffer 2011: 90). In this model the variable x is regressed on the last q (and the present) unobserved errors, which makes the model hard to understand at first and not to estimate directly. The second model class to be introduced here are autoregressive processes of order p, also AR(p):

$$x_t = \omega_t + \sum_{j=1}^{p} \phi_j x_{t-j} \qquad \textbf{(4)}$$

with ω_t being the error in period t as before (Shumway; Stoffer 2011: 84). We see, that in this modeling approach the variable x is regressed on its past observations, therefore the term autoregressive. Both approaches can be combined to the so-called Autoregressive Moving Average models of orders p and q, short ARMA(p,q) (Shumway; Stoffer 2011: 92):

$$x_t = \omega_t + \sum_{i=1}^{q} \theta_i \omega_{t-i} + \sum_{j=1}^{p} \phi_j x_{t-j} \tag{5}$$

The idea behind this is to transform the original time series into a so-called stationary time series. We could explain the concept of stationarity in more detail, but this would exceed the goal of this paper. We want to leave it at the interpretation that a stationary time series is no more serially correlated and therefore can be estimated at time t independently of its observation in time s.

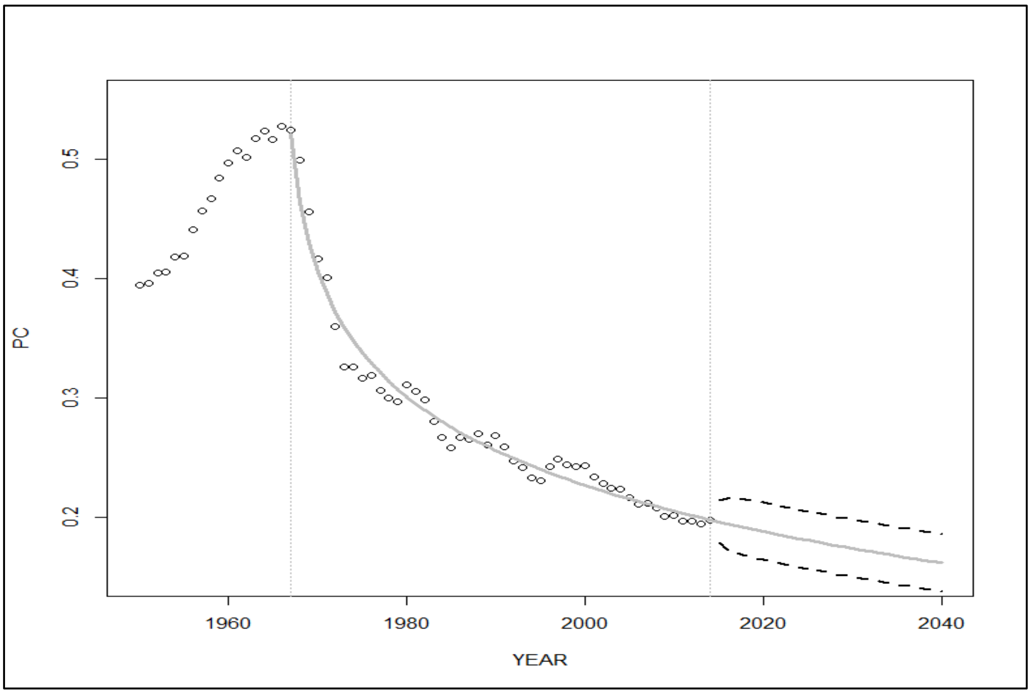
4 FERTILITY PROJECTION FOR GERMANY

The approach chosen for the projections is a parametric time series approach based on the principal components. The first step is the determination of an appropriate parametric function describing the theoretical course of the principal components. We observed a degressive decrease for the first principal component (see the first graphic in figure 6). This might be described appropriately using a logarithmic function. Taking into account the autocorrelation, we need an ARMA-part in our model. Based on Akaike's Information Criterion (AIC) we choose an AR(1) for this, which leads to a stationary model. Based on an Ordinary Least Squares (OLS) approximation, the resulting model is

$$PC_t^1 \; = .5217 - .0836 \ln(t) + .6834\, PC_{t-1}^1 + \omega_t \tag{6}$$

Figure 7 illustrates the data and the fitted model, including a 95%-PI:

Figure 7: Projection of PC 1

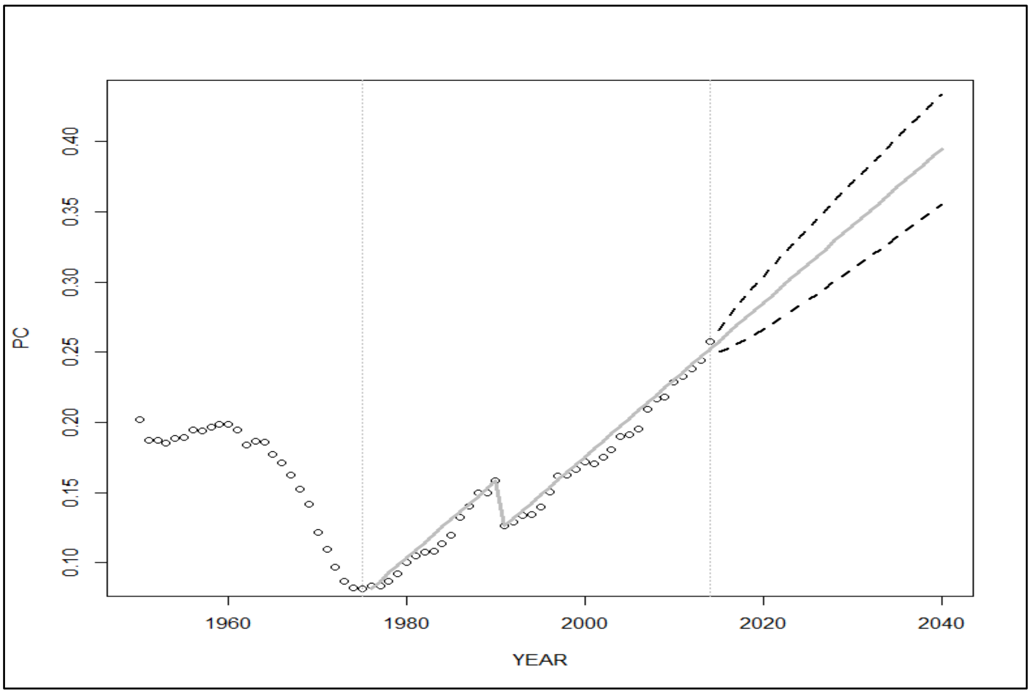


Source: Own calculations and design

The original data in this case are the dots, the gray line is the fitted function in **(6).** There are two dotted vertical lines. The left one marks the point in time, from which on the historical data has been used for estimation, in this case 1967. Data before 1967 has been ignored, since the functional behavior before 1967 has been different and is of no use for future forecasting. The right dotted line marks the last year of observation, 2014, after which the forecast starts. Based on the graphical analysis (and the test results) we conclude that the model has fitted the data well. We see the expectation for the future values of the first principal component. The dashed lines mark the limits of the 95%-PI in each year, which result by making use of the standard errors of the forecast in each year. Note that the t in **(6)** takes the value 1 in the first underlying year 1967.

Figure 8 illustrates the fit for the second PC, along with its projection until 2040:

Figure 8: Projection of PC 2



Source: Own calculations and design

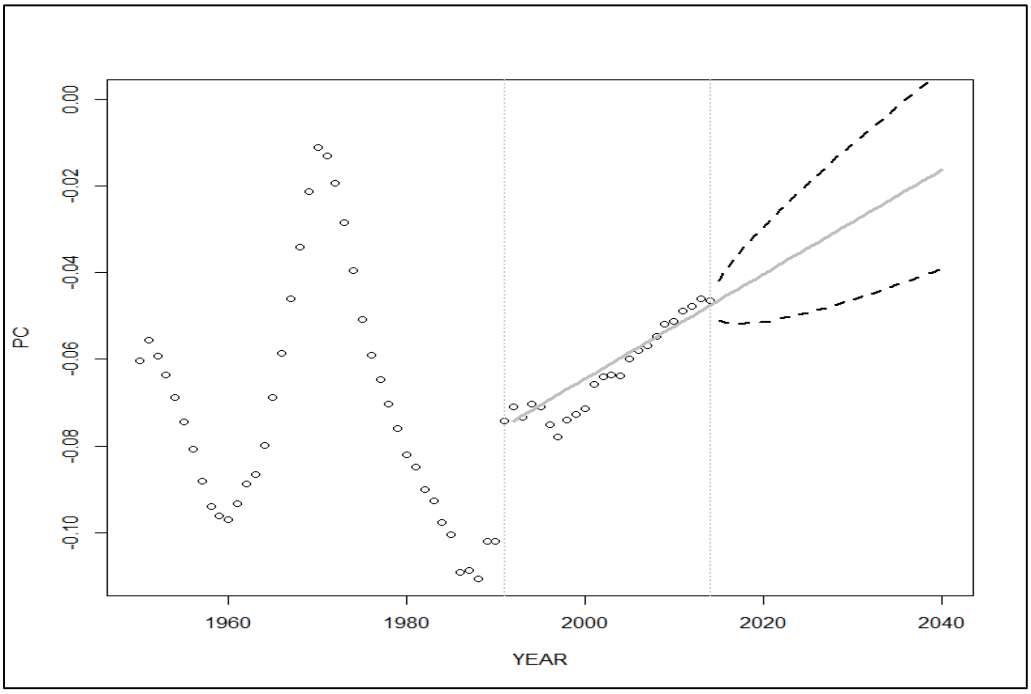
The underlying data for the projection are the observation from 1975 to 2014. Like mentioned earlier, there is a structural break in 1991 in the time series because of the reunification and the change in the underlying data base. The author believes this to affect to total level of the PC via a negative vertical shift only. The long-term trend appears to be a positive linear one, which was not affected by the reunification. This is modeled by the following function:

$$PC_t^2 = PC_{t-1}^2 + .0055 + .0376\, d + \varepsilon_t \qquad \textbf{(7)}$$

The time series has been identified as a random walk with drift. This means in expectation, c.p., the time series in time t takes the observation in time t-1 plus a constant by which it increases constantly each period. In this case the drift takes the value .0055. The next summand is a dummy variable, which was used to create the vertical shift in 1991. d takes the value 1 before 1991 and the value 0 after 1990. Therefore it does not affect the forecast further. ε_t means the error term in this model similar to ω_t in model **(6)**.

Finally, figure 9 shows the approximation and projection of the third PC:

Figure 9: Projection of PC 3



Source: Own calculations and design

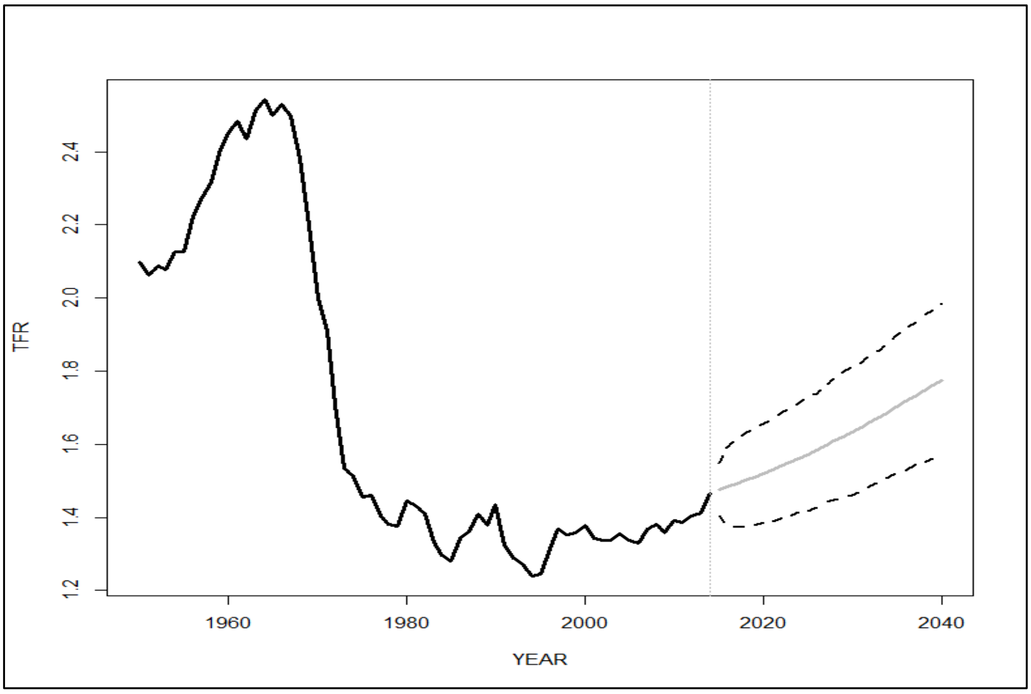
In this case we see a pure structural break after reunification. Therefore only the data after 1990 has been made use of for estimation purposes. In general we can interpret this as a reinforcing trend of younger women conceiving less children and women over 30 conceiving more children, which can be observed after the reunification in a stronger fashion. We observe a PI which becomes relatively broad over time. This is because of the stronger growing uncertainty due to the smaller sample used here. It is obvious we cannot predict the future behavior of this trend as certain with a history of 24 years opposed to 40 or 50 years.

The main goal of this paper is the projection of the TFR, which takes place at this point. The approach chosen here is a simulation study. First of all, based on the theoretical annual distributions 10,000 paths for each PC have been generated by computer simulations. From **(2)** it follows that

$$\boldsymbol{F} = \boldsymbol{C} \times \boldsymbol{A}^{-1}$$

This means we can derive 10,000 simulations for the future paths of the ASFR directly from the simulated outcomes of the PC. Since the TFR can be calculated directly as the sum of the ASFR, this leads us to the simulated values of the future TFR. The outcome of these simulations are illustrated in figure 10:

Figure 10: Projection of TFR



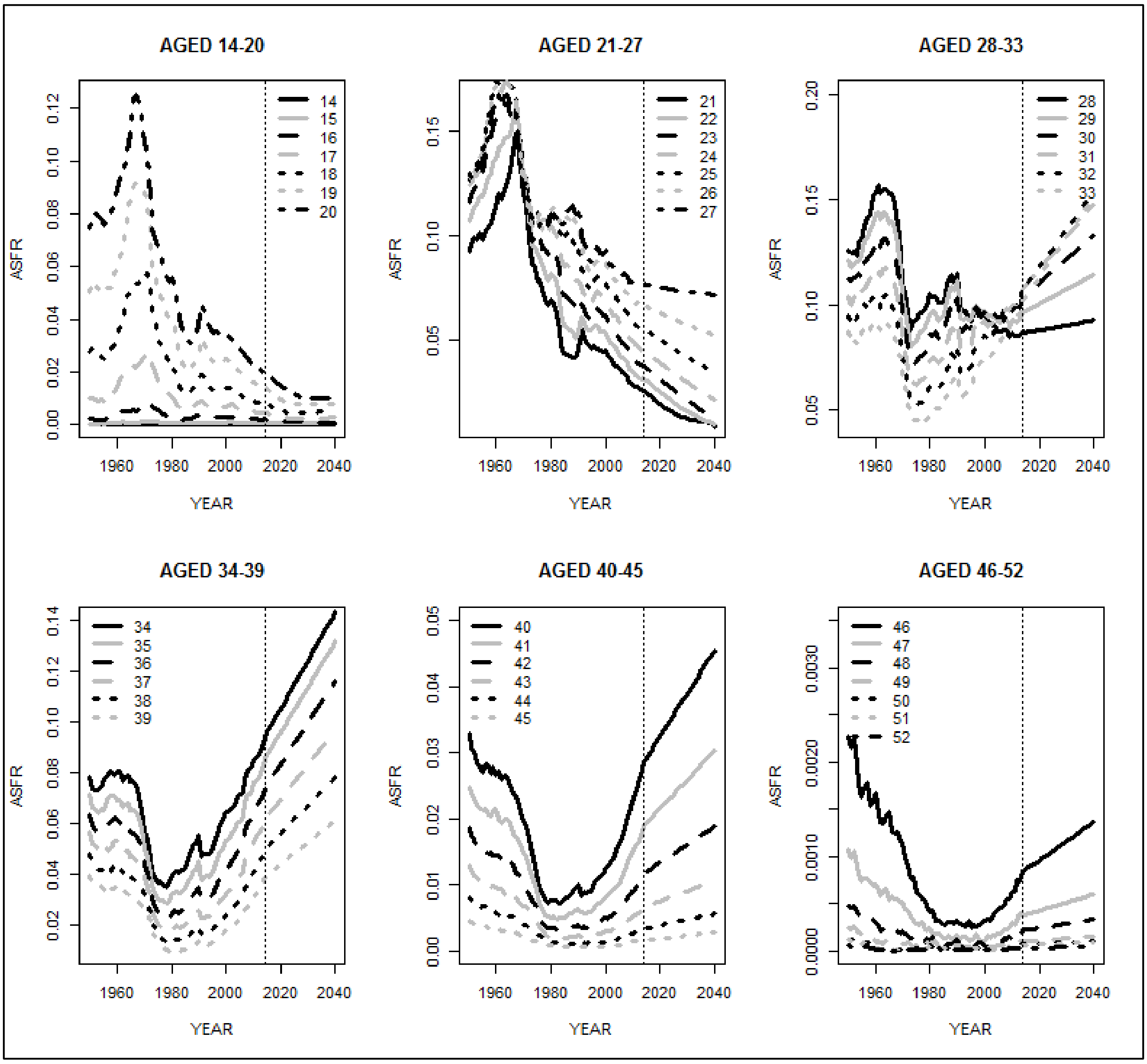
Source: Own calculations and design

We observe the TFR in expectation, c.p., growing from its level of 1.47 in 2014 to about 1.77 in 2040. With a 95% probability we expect the TFR in 2040 to vary between 1.57 and 1.98. Interestingly, this modeling approach leads to a higher expectation in the TFR than the aforementioned approaches by the UN, the EU or Destatis. One positive result of the projection is that we, opposed to common belief, almost certainly will see an increasing TFR until 2040. On the downside the TFR most certainly won't reach the replacement-level fertility. This means the current trend of more people dying than being born will not stop, for the next 25 years at least. This also means a positive net migration will be needed further for refreshment of the population and preventing it from decreasing in absolute number.

Since our modeling approach indirectly models the ASFR, we can catch a glimpse at the expected future course of the ASFR. The model does indeed create PI for each age group as mentioned earlier, for a better overlook we will restrict the illustration to the expected future ASFR, though.

Figure 11 approves the common belief by fertility researchers of a further decrease in births by young mothers and an ongoing higher fertility for mothers over 30, our forecast even expects increasing fertility rates for women aged 29 and higher.

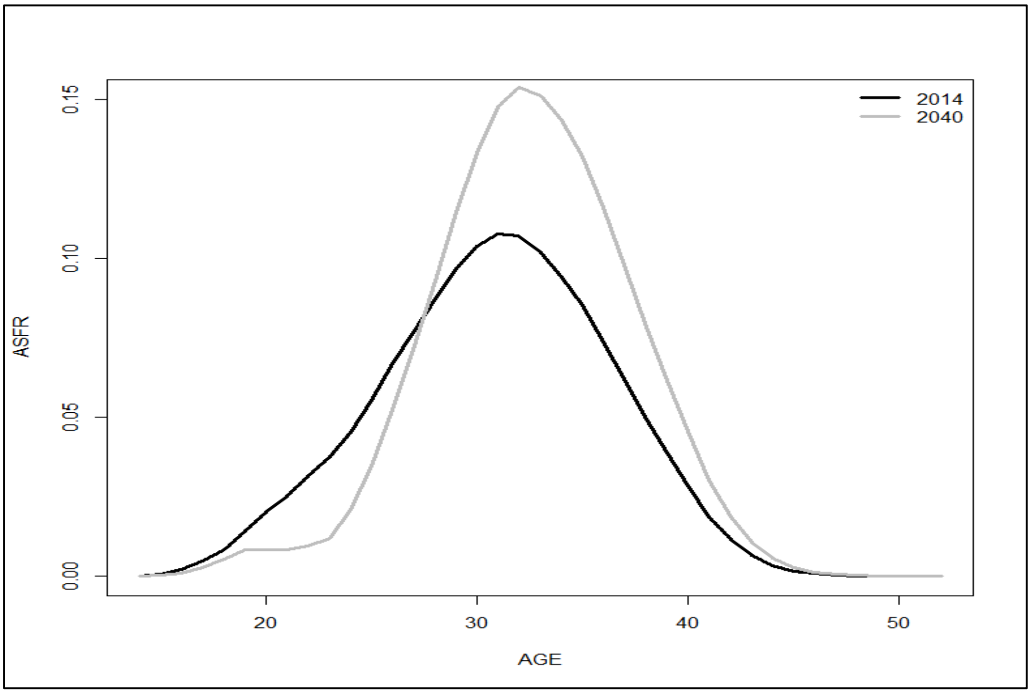
Figure 11: Expected future ASFR



Source: Own calculations and design

To get a better idea, how fertility among the age groups will develop, figure 12 shows the expected shift of the “fertility distribution” from 2014 to 2040. Based on the modeling results, in expectation, c.p., the distribution will slightly shift to the right and become bigger overall. The modal age for giving birth to a child will therefore increase from 31 years of age to 32, a development also expected by Destatis.

Figure 12: ASFR 2014 and 2040



Source: Own calculations and design

5 CONCLUSION AND OUTLOOK

It is well known that the classical population forecasts are based on deterministic mathematical models, including those by the Federal Statistics Office. These models assume a certain future behavior of the components of demographical change, including the TFR. ASFR are in these models commonly derived from the assumed TFR and assumptions for the distribution of the ASFR. In such cases, uncertainty about the future shall be incorporated into the model by scenario analysis in which alternative model outcomes are produced by simply assuming an alternative development of the components. The basic assumption by Destatis is that the TFR will remain at its perennial level of 1.4, whereas the alternative simulates an increasing TFR to a level of 1.6 in 2028. From this point on it would remain constant on this level. The latest data already proved the basic assumption wrong, since the TFR in 2014 has been at about 1.466. Deterministic forecasts are highly vulnerable with regard to their accuracy because they simulate only one outcome, which is, by itself, highly unlikely relative to the infinite theoretical outcomes. This study therefore promotes the use of stochastic projection, which can predict a broad range of possible outcomes and quantify them. Several authors have already proposed statistical models in demography and especially fertility on an international level with Rob Hyndman and Ronald Lee being two important persons in this field, just to name a few. For Germany specifically, the UN creates stochastic projections for the demographic development in general and the TFR in particular. Since the UN approach is based on quite strong assumptions regarding the three phase-structure whereas it uses relatively little country-specific data on Germany itself and instead simply assumes convergence towards 2.1, the author believes that these projections might be highly error prone. This to a certain point is covered by the quite large projection intervals, which might leave the user of these results with little information after all, since ranges that big do not tell us too much about the future behavior of the TFR. This might make it difficult to deal with them with regard to political intervention.

The main purpose of this study was to create a stochastic projection for the German TFR. Since very important dynamics are only revealed by a survey of the age-specific fertility rates, the modeling approach was aimed at modeling these rates indirectly over a principal component analysis. This has the advantage of including cross-correlational effects into the simulation. Future trends therefore will affect correlated age-groups accordingly. Furthermore, Box-Jenkins time series methodology was made use of, therefore capturing stochastic shocks, which may have long-term effects on the course of the ASFR.

For the regression analysis, Destatis and UN data for the period 1950–2014 were analyzed and made use of individually and according to current trending behavior. Because of past structural breaks in the time series, and therefore trends being traceable only until certain points in time, the prediction was aimed at a shorter time horizon through 2040, opposed to 2060, as commonly observed in population forecasts. The reason for this was to cover the forecasting period as reliably as possible. The projection resulted in an expected TFR in 2040 of 1.77 with a 95% prediction interval between 1.57 and 1.98. The future fertility structure among the age groups is predicted to have a distribution in general shape quite similar to the official forecasts. Furthermore, the forecast approves the common belief of an ongoing shift from births during the younger reproductive years to the years beyond 30. Since the fertility rates for women under 30 will decrease with weaker rates than the increase rates for women over 30, the author expects overcompensation and therefore a higher TFR in future compared to the TFR over the course of the last decade. Although the results indicate an increasing TFR, it is almost certain it will not reach the replacement-level fertility until 2040, though. This stresses the importance of retaining a positive net migration rate.

The author believes that the results of this projection approach are well-suited for application, since they are well-funded on real data for Germany and solid statistical models. Furthermore the results are easy to comprehend and very rich of information on the different age groups. The simulation technique also accounts for uncertainty, which in the most forecasts is not captured sufficiently. The detailed results of the projection could be used as a fundament for possible political intervention with the scope of increasing the TFR, since we gain an idea about the reproductive behavior of single age groups. In this respect, qualitative research along with cross-sectional econometric analysis is of major importance. The age-specific structure of the model makes it very flexible, allowing for more sophisticated and flexible scenario analysis for estimation of possible outcomes of theoretical political measures on the overall fertility.

The approach proposed in this paper in not flawless. Like every statistical model it needs big data sources to be applied optimally based on asymptotics. The data used here individually has been restricted to 24 to 48 observations, which is very little. This is a standard problem of demographic forecasts, since data is scarce compared to e.g. finance data. Furthermore the restriction on actual data and simulations only covers developments, which might be realistic based on the underlying data but does not give the opportunity to include unpredictable developments, which are included in a Bayesian approach. The restriction on observed data makes the model more objective though, as mentioned earlier. Moreover it shall be stressed that the author does not claim the presented model to be universally applicable. It appears appropriate for Germany, though. Slight deviations of the approach presented here might very well be applicable for other low-fertility countries, too. This very well might be investigated in further research works. This has not been the target in this paper, though. We need to consider that the model's validity cannot be proven or rejected, since there is too little data available for valid testing. The goodness of fit is based on graphical analysis and the subjective assessment of the author. The latter beliefs the approach in this paper is appropriate for Germany because of the detailed estimation of the ASFR taking stochasticity into account. Micro simulations may well be good and more sophisticated alternatives for forecasting compared to a macro approach like applied in this paper. These models are extremely data consuming and it is therefore almost impossible updating the forecasts on a regular basis. Moreover, the inclusion of explanatory variables renders the task of predicting the future behavior of the explanatory variables themselves, which leads to a huge amount of additional stochasticity. This leads to more complicated models which might not lead to better forecasting at all.

REFERENCES

Alkema, Leontine et al. 2011: Probabilistic Projections of the Total Fertility Rate for All Countries. In: Demography 48(3): 815-839.

Birg, Herwig 2001: Das demographisch-ökonomische Paradoxon. In: Forschung an der Universität Bielefeld: 23/2001: 32-37.

Bomsdorf, Eckart; Babel, Bernhard; Schmidt, Rafael 2008: Zur Entwicklung der Bevölkerung, der Anzahl der Schüler, der Studienanfänger und der Pflegebedürftigen: Stochastische Modellrechnungen für Deutschland bis 2050. In: Sozialer Fortschritt 57(5): 125-132.

Bongaarts, John; Feeney, Griffith 1998: On the Quantum and Tempo of Fertility. In: Population and Development Review 24(2): 271-291.

Bowley, A.L. 1924: Births and Population in Great Britain. In: The Economic Journal 34(134): 188-192.

Box, George; Jenkins, Gwilym 1976: Time Series Analysis. Forecasting and Control. San Francisco: Holden-Day.

Bujard, Martin 2011: Familienpolitik und Geburten. Ein internationaler Vergleich. Wiesbaden.

Bujard, Martin 2012: Die Kinderzahl von Akademikerinnen. Befunde eines Schätzmodells mit Mikrozensusdaten der Jahre 1982 und 2011. In: Bevölkerungsforschung Aktuell. Mitteilungen aus dem Bundesinstitut für Bevölkerungsforschung 33(5): 2-11.

Bujard, Martin 2015: Consequences of Enduring Low Fertility – A German Case Study: Demographic Projections and Implications for Different Policy Fields. In: Comparative Population Studies 40(2): 131-164.

Cannan, Edwin 1895: The Probability of a Cessation of the Growth of Population in England and Wales during the next Century. In: The Economic Journal 5(20): 505-515.

Chatfield, Christopher; Collins, Alexander 1980: Introduction to Multivariate Analysis. Chapman & Hall.

Dorbritz, Jürgen 2008: Germany: Family diversity with low actual and desired fertility. In: Demographic Research 19(17): 557-598.

Dudel, Christian 2014: Vorausberechnung von Verwandtschaft: Wie sich die gemeinsame Lebenszeit von Kindern, Eltern und Großeltern zukünftig entwickelt. Beiträge zur Bevölkerungswissenschaft 45. Bundesinstitut für Bevölkerungsforschung.

Engelhardt, Henriette 2009: Zum Wandel der Korrelation von Fertilität und Frauenerwerbstätigkeit in Raum und Zeit: Eine empirische Analyse unter Berücksichtigung der Effekte ausgewählter sozialer Indikatoren. In: Zeitschrift für Familienforschung 21(3): 246-264.

Espenshade, Thomas; Guzman, Juan Carlos; Westoff, Charles 2003: The surprising global variation in replacement fertility. In: Population Research and Policy Review 22(5-6): 575-583.

European Commission 2014: The 2015 Ageing Report. Underlying Assumptions and Projection Methodologies. In: European Commission. Institutional Paper 08/14.

Gesetz zur Förderung von Kindern unter drei Jahren in Tageseinrichtungen und in Kindertagespflege. 10.12.2008.

Härdle, Wolfgang; Myšičková, Alena 2009: Stochastic Population Forecast for Germany and its Consequence for the German Pension System. In: Humboldt-Universität zu Berlin. SFB 649: Economic Risk. Discussion Paper 2009-009.

Hyndman, Rob; Booth, Heather 2008: Stochastic population forecasts using functional data models for mortality, fertility and migration. In: International Journal of Forecasting 24: 323-342.

Hyndman, Rob; Kostenko, Andrey 2007: Minimum Sample Size Requirements for Seasonal Forecasting Models. In: Foresight. The International Journal of Applied Forecasting 2007(6): 12-15.

Hyndman, Rob; Ullah, Shahid 2007: Robust forecasting of mortality and fertility rates: A functional data approach. In: Computational Statistics & Data Analysis 51(10): 4942-4956.

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

Leim, Iris 2008: Die Modellierung der Fertilitätsentwicklung als Folge komplexer individueller Entscheidungsprozesse mit Hilfe der Mikrosimulation. Metropolis.

Lipps, Oliver; Betz, Frank 2005: Stochastische Bevölkerungsprojektion für West- und Ostdeutschland. In: Zeitschrift für Bevölkerungswissenschaft 30(1): 3-44.

Pavetić, Monika 2009: Familiengründung und –erweiterung in Partnerschaften: Statistische Modellierung von Entscheidungsprozessen. VS Verlag für Sozialwissenschaften.

Raftery, Adrian et al. 2012: Bayesian probabilistic population projections for all countries. In: Proceedings of the National Academy of Sciences 109(35): 13915-13921.

Schmertmann, Carl; Zagheni, Emilio; Goldstein, Joshua; Myrskylä, Mikko 2014: Bayesian Forecasting of Cohort Fertility. In: Journal of the American Statistical Association 109(506): 500-513.

Shumway, Robert; Stoffer, David 2011: Time Series Analysis and Its Applications: With R Examples. Springer.

Statistisches Bundesamt 1952: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1952. Wiesbaden.

Statistisches Bundesamt 1953: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1953. Wiesbaden.

Statistisches Bundesamt 1954: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1954. Wiesbaden.

Statistisches Bundesamt 1955: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1955. Wiesbaden.

Statistisches Bundesamt 1956: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1956. Wiesbaden.

Statistisches Bundesamt 1957: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1957. Wiesbaden.

Statistisches Bundesamt 1958: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1958. Wiesbaden.

Statistisches Bundesamt 1959: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1959. Wiesbaden.

Statistisches Bundesamt 1960: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1960. Wiesbaden.

Statistisches Bundesamt 1961: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1961. Wiesbaden.

Statistisches Bundesamt 1962: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1962. Wiesbaden.

Statistisches Bundesamt 1963: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1963. Wiesbaden.

Statistisches Bundesamt 1964: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1964. Wiesbaden.

Statistisches Bundesamt 1965: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1965. Wiesbaden.

Statistisches Bundesamt 1966: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1966. Wiesbaden.

Statistisches Bundesamt 1967: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1967. Wiesbaden.

Statistisches Bundesamt 1968: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1968. Wiesbaden.

Statistisches Bundesamt 1969: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1969. Wiesbaden.

Statistisches Bundesamt 1970: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1970. Wiesbaden.

Statistisches Bundesamt 1971: Statistisches Jahrbuch für die Bundesrepublik Deutschland: 1971. Wiesbaden.

Statistisches Bundesamt 2014: Lebendgeborene nach dem Alter der Mütter. Provided on request.

Statistisches Bundesamt 2015: Bevölkerung Deutschlands bis 2060: 13. koordinierte Bevölkerungsvorausberechnung. Wiesbaden.

Statistisches Bundesamt 2015a: Bevölkerung nach Altersgruppen. Früheres Bundesgebiet. Provided on request.

Statistisches Bundesamt 2015b: Bevölkerung: Deutschland, Stichtag, Altersjahre, Nationalität/Geschlecht/Familienstand. URL: https://www-genesis.destatis.de/genesis/online/data;jsessionid=3F657705D120032CDE988BCB51EF019C.tomcat_GO_1_2?operation=abruftabelleAbrufen&selectionname=12411-0006&levelindex=1&levelid=1439296414812&index=1. 18.12.2015.

Statistisches Bundesamt 2015c: Lebendgeborene: Deutschland, Jahre, Alter der Mutter, Geschlecht der Lebendgeborenen, Familienstand der Eltern. URL: https://www-genesis.destatis.de/genesis/online/data;jsessionid=3F657705D120032CDE988BCB51EF019C.tomcat_GO_1_2?operation=abruftabelleAbrufen&selectionname=12612-0007&levelindex=1&levelid=1439296772111&index=7. 18.12.2015.

United Nations 1950: Demographic Yearbook 1949-50. New York.

United Nations. Department of Economic and Social Affairs. Population Division 2015: World Population Prospects: The 2015 Revision. URL: http://esa.un.org/unpd/wpp/, 10.08.2015.

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