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

***Un Modello a Componenti di Coorte per Previsioni della Popolazione – Il Caso della Germania***

Patrizio Vanella[[1]](#footnote-2) and Philipp Deschermeier[[2]](#footnote-3)

**Abstract** The future development of population size and structure is of undeniable importance since planning in many areas of politics and business is conducted based on expectations about the future makeup of the population. Countries with both decreasing mortality and low fertility rates, as is the case for most countries in Europe, are in urgent need of adequate population forecasts to identify future problems for social security systems and overall macroeconomic development. The labor market will especially be affected by the retirement of the *baby boomer generation* and the resulting shortage of young skilled workers. This contribution proposes a stochastic cohort-component model that uses simulation techniques based on stochastic models for fertility, migration and mortality to forecast the population by age and sex. The results provide detailed insight into the future population structure, disaggregated into both sexes and 116 age groups. Moreover, the uncertainty in the forecast is quantified as prediction intervals for each subgroup. The underlying models for forecasting the demographic components have been developed in earlier studies and rely on principal component time series models.

**Abstract** *Il futuro sviluppo della grandezza e della struttura della popolazione è di importanza innegabile, poiché la pianificazione in molte aree di politica ed economia è condotta sulla base di aspettative sulla composizione futura della popolazione. Paesi con tassi di mortalità e fertilità in calo, fenomeno comune nella maggioranza dei paesi d’Europa, hanno un urgente bisogno di previsioni adeguate della popolazione per identificare i problemi futuri per il sistema d’assicurazione sociale e per lo sviluppo macroeconomico generale. Il mercato del lavoro sarà influenzato in particolare dai pensionamenti della generazione baby boom e dalla risultante scarsità di lavoratori giovani e qualificati. Questo studio propone un modello a componenti di coorte che usa tecniche di simulazione a base di modelli stocastici per fertilità, migrazione e mortalità, per predire la popolazione per età e sesso. I risultati offrono informazioni dettagliate sulla struttura della popolazione nel futuro, disaggregate per sesso e 116 gruppi d’età. Inoltre, l’incertezza della previsione è quantificata attraverso intervalli di previsione per ogni sottogruppo. I sottostanti modelli per predire le componenti demografiche sono stati sviluppati in precedenti studi e usano modelli di serie storiche (time series) con componenti principali.*

**Key words:** Population Forecasting, Stochastic Simulation

**1 Future Development of the Demographic Components**

We propose a population forecast based on a probabilistic cohort-component model developed through the combination of Vanella’s (2017a: 543-549) model for forecasting age- and sex-specific survival rates *(ASSSRs)*, the forecast model for age-, sex- and nationality-specific net migration *(ASNSNM)* by Vanella and Deschermeier (2017: 6-23) and Vanella’s (2017b: 11-24) model for forecasting age-specific fertility rates by sex of newborns *(ASSFR)*. Each partial model generates 10,000 scenarios stochastically by simulation of Wiener processes[[3]](#footnote-4), which represent 10,000 scenarios for the future development of the population in Germany by age (*0-115*) and sex (*binary*).

Vanella (2017b: 21-22) incorporated the effect of international migration on fertility in his fertility forecast model. The inclusion of this fertility model into the population forecast model proposed here provides an opportunity to quantify the effects of international migration on the development of the population not only directly through the migration process itself but also indirectly by altering the number of births due to the different reproductive behavior of migrants relative to that of the native population.

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

Step I:

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

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

Step II:

The number of survivors from $\tilde{P}\_{x+1,y+1,m,t}$ at the end of *y+1* are calculated through multiplication with the age- and sex-specific survival rate (*ASSSR*) $s\_{x+1,y+1,g,t}$ for persons aged *x+1* years of sex *g* in year *y+1* and in trajectory *t*:

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

Step III:

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

$$F\_{x,y+1,w,t}=\sum\_{x=14}^{52}\frac{P\_{x-1,y,w,t}+P\_{x,y+1,w,t}}{2}.$$

This approximation presumes constant migration and mortality during a year because we rely on annual data. This approach is mathematically practical; moreover, we believe the possible error arising from this assumption is extremely small and therefore produces no fundamental error in the future population estimation.

Step IV:

The sex-specific live births$B\_{y+1,g,t}$ are estimated:

$$B\_{y+1,g,t}=\sum\_{x=14}^{52}F\_{x,y+1,w,t}\*f\_{x,y+1,g,t},$$

where $f\_{x,y+1,g,t}$ denotes the ASSFR for females aged *x* years in year *y+1* conceiving babies of sex *g* in trajectory *t*.

Step V:

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

$P\_{0,y+1,g,t}=B\_{y+1,g,t}\*s\_{0,y+1,g,t}$.

In this way, the population by sex and age in year *y+1* in trajectory *t* is obtained.

The algorithm is illustrated in Figure 1.

Figure 1: Process of Annual Population Update



Source: Own design

This process is then used to stochastically forecast the population by sex and age until the year 2040. The future development of the population is forecast by estimating the future course of the demographic components of fertility (through ASSFRs), migration (ASNSNM), and mortality (ASSSRs). The underlying variables are simulated indirectly on the basis of the stochastic trajectories of the PCs, which were derived in earlier studies. A quasi-two PC time series model, as proposed by Vanella (2017a: 543-549), is used to forecast the logistically transformed ASSSRs. The first PC is a general mortality index, similar to the classic Lee-Carter index. The second PC is a behavioral index representing differences in mortality between the two sexes associated with nutritional and smoking behavior. Migration flows are accounted for via the quasi-two PC forecast model for ASNSNM proposed by Vanella and Deschermeier (2017: 15-22). In this case, the first PC is a labor market index and the second PC addresses the migration resulting primarily from humanitarian or economic crises. The fertility model builds upon the results of the mentioned migration model. This component is forecast by the quasi-three PC model proposed by Vanella (2017b: 11-24). In this case, the first PC is associated with the tempo effect in fertility, i.e., the postponement of births from younger ages to older ages. The second PC is a quantum index used to address general trends in the quantity of reproductive behavior in a society. The third PC considers the impact of the migration level on fertility and is derived via econometric specification of the influence through the second PC of the migration model proposed by Vanella and Deschermeier (2017: 12-17).

**2 Population Development in Germany until 2040**

The combination of the resulting trajectories for the demographic components results in a probabilistic cohort-component model for forecasting the age- and sex-specific population for the ages 0-115 years. The initial population for the forecast is the age- and sex-specific population reported by Destatis for December 31, 2015. Population numbers for ages 100 years and over are not available in detail but are rather aggregated into an upper age group. Thus, we approximated the population in this age group through an own population update based on the death counts by age and sex.

In general, we observe greater uncertainty for males. Whereas the retirement-age population can be predicted relatively well, the uncertainty in the future working-age population is especially large for males due to the high uncertainty in the migration forecast for males (Vanella, Deschermeier 2017: 17-22). The uncertainty in the population of persons under 25 years of age arises from the fact that this portion of the population has not been born yet. Specifically, the uncertainty of persons under 25 years of age arises from migration, from the mortality of persons “below 1 year” and from uncertainty due to fertility.

The aging of the population is of high social and economic importance, as was stated earlier. Therefore, in addition to the overall age structure, the median age of the male and female populations is considered as a summary indicator for the future age schedule of the population. The median age of the population can be obtained from the simulation results because it is the exact age that cuts the population in half. This computation for all 10,000 trajectories can be used to extract PIs for the median age.

We observe a rejuvenation effect for the upcoming years due to high net migration during this period. The high net migration rates around the year 2015 combined with the high forecast values for the upcoming years leave a mark in the age structure of Germany. This can be seen in the age structure for the male and female populations in the year 2040. By that time, the majority of the population that immigrated during the high influx phase will be approximately 50 years old, while the baby boomer generation will be in their seventh decade of life. Over the forecast horizon, the median age traces this development by a rejuvenation effect for men and women. The probable decrease in the number of births after the middle of the 2020s and decreasing net migration and mortality (Vanella 2017a: 550) lead to an aging of the population structure, as represented by the increasing median ages after that point. Since a larger portion of migrants is male (Vanella, Deschermeier 2017: 20-22), the rejuvenation is stronger for males than for females.The driving factor behind the presented results on the population development is net migration. Following the net migration in 2015, the first years of the forecast result in high and above average values of net migration. In the long run, net migration is expected to decrease to 212,419 in 2040. Counterbalancing net migration, we take a look at the probabilistic forecast of Germany’s natural growth until 2040, which is simply derived from the birth and death forecasts. Following the trend of the recent past, the number of births will continue rising until the mid-2020s. However, as the peek remains below 900,000 births per year for the median scenario, this development cannot be labeled a second baby boom. In comparison, the number of livebirths in the 1950s and 1960s exceeded 1.1 million annually (GENESIS-Online Datenbank 2018). Due to the shift in the age structure, the number of potential mothers is expected to shrink at some point, resulting in decreasing birth numbers in the long run. The aging of the population leads to a steady increase in the number and rate of older people, who have a higher probability of dying in a given year. This results in a steady increase in the number of deaths until 2040. As the median of the simulation is greater than 900 thousand persons in each year of the forecast, the natural population development until 2040 almost certainly will be negative during the forecast horizon. Therefore, a lack of high net migration would naturally lead to a decrease in population in Germany. The connection between natural population growth and net migration is therefore obvious since an old population with low fertility requires positive net migration to avoid shrinking and overaging. In the median trajectory, the natural decrease in the population will grow annually, stressing the importance of positive net migration, especially in the younger ages, to fill the shortages occurring in the labor market. As we will show by some important measures, our model provides a wide range of detailed analyses targeting specific topics of interest. The forecast results offer the possibility for a wide range of future studies, e.g., analyzing the effects of population changes on social security, the labor market or housing demand.

**References**

*GENESIS-Online Datenbank* 2018: Lebendgeborene: Deutschland, Jahre, Geschlecht. In: www-genesis.destatis.de. URL: https://www-genesis.destatis.de/genesis/online/data;jsessionid=C4FF958DD4F6A3F4DB5C783F4B2D2890.tomcat\_GO\_2\_1?operation=abruftabelleAbrufen&selectionname=12612-0001&levelindex=1&levelid=1515759981915&index=1, 10.01.2018.

*Vanella, Patrizio* 2017a: A Principal Component Model for Forecasting Age- and Sex-Specific Survival Probabilities in Western Europe. In: German Journal of Risk and Insurance 106,5: 539–554 [10.1007/s12297-017-0393-y].

*Vanella, Patrizio* 2017b: Age- and Sex-Specific Fertility in Germany until the Year 2040 – The Impact of International Migration. Hannover Economic Papers 606. Hannover: Gottfried Wilhelm Leibniz Universität Hannover, School of Economics and Management.

*Vanella, Patrizio* 2017c: Stochastische Prognose demografischer Komponenten auf Basis der Hauptkomponentenanalyse. Hannover Economic Papers 597. Hannover: Gottfried Wilhelm Leibniz Universität Hannover, School of Economics and Management.

*Vanella, Patrizio; Deschermeier, Philipp* 2017: Ein stochastisches Prognosemodell internationaler Migration in Deutschland. Hannover Economic Papers 605. Hannover: Gottfried Wilhelm Leibniz Universität Hannover, School of Economics and Management.

1. Patrizio Vanella, Gottfried Wilhelm Leibniz Universität Hannover – Center for Risk and Insurance; email: pv@ivbl.de [↑](#footnote-ref-2)
2. Philipp Deschermeier, Institut Wohnen und Umwelt Darmstadt; email: p.deschermeier@iwu.de [↑](#footnote-ref-3)
3. See Vanella 2017c: 13 for an explanation of Wiener processes. [↑](#footnote-ref-4)