DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 54, ARTICLE 23, PAGES 719–762
PUBLISHED 14 APRIL 2026
http://www.demographic-research.org/Volumes/Vol54/23/
DOI: 10.4054/DemRes.2026.54.23

Research Article

Probabilistic population forecasts for small regions

Julius Goes

Henriette Engelhardt

Contents

1	Introduction	720
1.1	Cohort-component method	721
1.2	Bayesian methods	722
2	Data	723
3	Mortality	724
4	Fertility	726
4.1	Direct estimation	727
4.2	Indirect estimation	728
4.3	Out-of-sample comparison	729
5	Migration	731
5.1	Out-of-sample validation	733
5.2	Age-specific migration	734
6	Estimation	737
7	Population forecasts	738
7.1	Out-of-sample validation	738
7.2	Results	739
8	Discussion	742
9	Acknowledgments	744
	References	745
	Appendix	751

Demographic Research: Volume 54, Article 23
Research Article

Probabilistic population forecasts for small regions

Julius Goes[1]

Henriette Engelhardt[2]

Abstract

BACKGROUND
Age-specific population forecasts for small areas or subnational regions are a valuable tool for local governments. However, typical population projection methods based on the cohort-component approach are difficult to apply on a smaller subnational scale.

OBJECTIVE
We introduce Bayesian methods suitable for obtaining reliable age-specific population forecasts for small regions using the cohort-component method.

METHODS
Our approach improves fertility forecasting by extending the Lee–Carter model with an age-region interaction term. We propose to forecast net-migration counts using skewed error terms, and introduce a Dirichlet regression to model migration age patterns as well as age proportions of fertility.

RESULTS
We run our model to produce age-specific population forecasts for a set of 13 heterogeneous regions in Bavaria, Germany. We compare our method with other standard approaches and find that it produces superior out-of-sample forecasts according to both point measures and scoring rules.

CONCLUSIONS
The findings suggest that the proposed Bayesian methods offer good predictive accuracy and are suitable in obtaining precise forecasts of age-specific population for smaller geographical regions.

CONTRIBUTION
We introduce a new method for the probabilistic projection of subnational population that works well and outperforms other current methods.

[1] Institute for Statistics, University of Bamberg, Bamberg, Germany. Email: julius.goes@uni-bamberg.de.
[2] Professor of Demography, Institute for Sociology and State Institute of Family Research, University of Bamberg, Bamberg, Germany. Email: henriette.engelhardt-woelfler@uni-bamberg.de.

1. Introduction

Reliable age-specific population forecasts for small areas are essential for effective urban planning, resource allocation, and infrastructure development. They help local governments anticipate future needs – such as schools, healthcare services, housing, and transportation – ensuring efficient resource distribution as a community grows or changes. Despite these crucial applications, population forecasts at the regional or subnational level have historically received far less attention than those at the national level. Although methods used for national forecasts can usually be adapted for regional use, significant challenges often arise. First, migration flows are far more significant at a regional level. While these flows are less important and often neglected at a national level (Booth 2006), at a subnational level, they are typically larger relative to the population and can become a primary driver of demographic change. Furthermore, regional migration is inherently difficult to predict and can be heavily influenced by external factors such as wars and changes in political structures – for example, the dissolution of the German Democratic Republic, which resulted in significant out-migration to economically stronger regions of West Germany between 1990 and 1995 (Wolff et al. 2022). Second, available data for subnational areas are often shorter, of poorer quality, and more erratic and can include zero cell counts (Wilson et al. 2022). Third, it is more difficult to predict regional populations, leading to higher forecast errors at the regional than at the national level (Tayman 2011). Nevertheless, interest in population forecasts for small areas has grown recently. For example, Wilson (2012); Cameron and Poot (2011); Swanson and Tayman (2014); and Swanson, Tayman, and Cline (2025) each apply different techniques to incorporate uncertainty intervals into existing deterministic regional population forecasts.

Moreover, Alexander and Alkema (2022) propose a probabilistic cohort-component method to predict the number of women of reproductive age. Their method incorporates census data and aims to minimize data requirements. Yu et al. (2023) predict subnational populations in the state of Washington by scaling national input data to obtain regional estimates. Wiśniowski and Raymer (2025) propose a multiregional population projection model that includes interregional migration counts – that is, migration flows between regions, to forecast age-specific population for Australian states. However, such data may not always be available. Moreover, their method is tailored for larger subnational regions. Further work is required to adapt it to small areas where the incidence of events is low or zero (Wiśniowski and Raymer 2025). In this paper, we aim to address the challenge of obtaining reliable, age-specific population forecasts at a small regional level using a probabilistic implementation of the cohort-component method. To this end, we introduce new methods specifically designed to predict age-specific population at a subnational level while accounting for regional dependencies.

The paper is organized as follows: We first describe the cohort-component method and its probabilistic extension and give a brief overview on Bayesian methods in demog-

raphy. Then we introduce the data that is used to create population forecasts for small areas. Next, we outline our proposed methodology for obtaining regional predictions of mortality, fertility, and net-migration. We assume the same model for both sexes to calculate mortality and net-migration, although they are estimated separately. However, for notational convenience a sex-specific subscript is omitted. Lastly, we present the results of the population forecast and end the paper with a discussion.

1.1 Cohort-component method

The most commonly used approach for population projection are cohort-component methods. These are based on the fundamental equation stating that the population (E) at time $t+1$ is equal to the population at time t plus the number of births (B) plus the number of in-migrants (I) minus the number of deaths (D) minus the number of out-migrants (O) (Preston, Heuveline, and Guillot 2000):

$$E_{t+1} = E_t + B_t - D_t + I_t - O_t. \tag{1}$$

The cohort-component method of population projection uses the fundamental equation as its key underlying concept. It takes a baseline population with a specific age structure and projects it forward using age-specific mortality, fertility, and net-migration rates. The process is repeatedly applied to obtain a future population for each period after the base period and was mathematically formalized by Leslie (1945). A more detailed description can be found in Raftery and Ševčíková (2023) or Preston, Heuveline, and Guillot (2000: ch. 6).

Traditional population projections based on the cohort-component method typically define deterministic scenarios based on combinations of mortality, fertility, and migration. However, this approach has several issues. First, it is impossible to define prediction intervals. Second, the components remain fixed in their respective low, medium, or high variants throughout the forecast duration, which can lead to unrealistic assumptions (Booth 2006). Third, a scenario-based approach provides inconsistent indicators of uncertainty (Lee 1998). To overcome these problems, probabilistic cohort-component methods have been introduced. These methods account for uncertainty by assigning a statistical model to each component, and therefore allow for a probabilistic projection of the future population. This provides a more realistic understanding of the uncertainty of the forecast. Probabilistic implementations of the cohort-component projection method are typically implemented within a Bayesian framework (e.g., Alexander and Alkema 2022; Yu et al. 2023). While some authors distinguish between deterministic population projections and probabilistic population forecasts, we use these terms interchangeably in this work to refer to probabilistic predictions of future population size and structure.

In this paper, we propose a probabilistic implementation of the cohort-component method that is similar in style to the approach used by the United Nations (UN) for their official population forecasts for all countries worldwide (see United Nations 2024).

Probabilistic population forecasts can be obtained using a Monte Carlo approach, which involves generating random draws (or trajectories) from the probability distributions of the demographic components: mortality, fertility, and net-migration. Each trajectory represents a stochastic realization of how these components may evolve over time. For each realization, age-specific population values are projected forward in time. The process is repeated across all random draws, producing a range of possible future population outcomes. From these simulations, uncertainty intervals (e.g., 95% prediction intervals) can be derived by calculating empirical quantiles. For our projections, we assume that half of the net-migration occurs at the beginning and the other half at the end of each projection period. The population is projected forward for each sex separately, utilizing sex-specific inputs for mortality and net-migration. The total number of births is divided into male and female births using a sex ratio at birth estimate derived from past data, assumed to be constant over time.

1.2 Bayesian methods

Bayesian methods are becoming increasingly popular in demographic literature (Bijak and Bryant 2016) and have been successfully applied to model subnational estimates for each input of the cohort-component method, that is – mortality, fertility, and migration rates. More precisely, numerous researchers utilize Bayesian hierarchical models to pool information across time and space to model subnational mortality using principal component analysis and singular value decomposition (Alexander, Zagheni, and Barbieri 2017; Dharamshi et al. 2025), the TOPLAS relational model, a spline-based approach, (Schmertmann and Gonzaga 2018; Rau and Schmertmann 2020), or classic random effect models extended with a spatially structured parameter (Congdon 2014; Goes 2024). For subnational fertility, Ševčíková, Raftery, and Gerland (2018) propose Bayesian methods to produce subnational estimates of total fertility rate, while Schmertmann et al. (2013) use empirical Bayesian methods to smooth regional fertility data before applying a Brass relational model to obtain subnational fertility estimates. Bryant and Zhang (2016) introduce a Bayesian method for obtaining subnational migration estimates, and Yu et al. (2023) adapt a Bayesian method by Azose and Raftery (2015), originally designed for national-level data, to generate subnational forecasts.

In this paper, we propose a method employing a Bayesian implementation of the cohort-component method to generate probabilistic forecasts of age-specific populations. Here, we introduce new methods to predict each input of the cohort-component method that are specifically suited for subnational data. More precisely we propose skewed error

terms that allow us to predict net-migration counts more accurately, which are heavily influenced by external factors. In addition, we introduce a novel approach to forecast age-specific fertility rates using a Dirichlet regression. We evaluate our proposed methods on out-of-sample data and compare the results to the standard approaches in the demographic literature.

2. Data

We aim to forecast population counts for small areas or regions. There is no universally accepted definition for what constitutes a 'small area.' For example, Wilson et al. (2022) use the term for areas with populations under 100,000, though they acknowledge this cutoff is somewhat arbitrary. The European Union defines a Nomenclature of Territorial Units for Statistics (NUTS), differentiating various region sizes.[3] Their smallest unit, NUTS-3, typically denotes a region within a state or province. For example, in Germany, NUTS-3 regions are often districts (Landkreise) or independent cities (kreisfreie Städte). The exact definition varies from country to country, but the general principle is that they represent relatively small, geographically defined areas. In the following, we will use the term 'small areas' to refer to NUTS-3 regions.[4]

To illustrate our method, we apply it to a specific set of NUTS-3 regions, within the state of Bavaria, Germany, called Upper Franconia. Though in general, the method that is introduced can be broadly applied to any set of NUTS-3 regions, or even lower-level areas that the European Union refers to as local administrative units (LAU). Bavaria is the largest state in Germany in terms of area and is located in the southeast part of the country. Upper Franconia is one of seven administrative subunits within Bavaria, located in the northeast and consists of $R = 13$ NUTS-3 regions. We believe these to be a typical set of small areas in a sense that they are heterogeneous in age structure, GDP per capita (cf. Figure A-2 in the Appendix), and population size population size, ranging from around 40,000 to 150,000. In addition, several of the regions have populations with a high share of students with a distinct age profile, leading to a specific fertility schedule different from that of other regions. Moreover, due to their differences in GDP, and thus economic activity, some regions experience vastly different net-migration.

The data for evaluating our method are provided by the Statistical Office of Bavaria (Bayerisches Landesamt für Statistik; BLfS). The data are publicly available and can be downloaded from GENESIS, the database of the Bavarian statistical institute. The datasets consist of population counts (BLfS 2025), counts of deaths (BLfS 2022), counts of births (BLfS 2024a), and counts of in- and out-migration (BLfS 2024b). The death and part of the migration data are disaggregated by age and grouped into five-year intervals,

[3] See https://ec.europa.eu/eurostat/web/nuts, accessed 17 July 2025.
[4] This is different from the definition used for a small area community, see, e.g., Rao and Molina (2015).

starting from $[0, 5)$ up to $[80, 85)$ with an open-ended age group of $85+$, resulting in a total of $X = 18$ age groups. The counts of births are also provided in five-year age groups, with an open-ended age group for the youngest (< 20) and the oldest ($40+$). However, these datasets are available for different time periods. Specifically, counts of births are available from 1995 to 2023, while counts of deaths are available only from 2000 to 2017. Unfortunately, after 2017, counts of deaths at the regional level are no longer available in a five-year age grid due to a change in data protection regulations. Thus, to estimate mortality rates we use data from 2000 to 2017, while estimates of fertility can be obtained using a longer time series. Age-specific population counts are available from 2000 to 2024.

For migration analysis, two datasets are available. The Statistical Office of Bavaria provides migration counts, distinguishing between in-migration and out-migration, using a broad age grid (< 18, 18–25, 25–30, 30–50, 50–65, $65+$). These data cover the period from 2000 to 2023 and are used to predict future total migration counts. However, this age grid is too coarse for calculating the migration schedule (see Section 5). Therefore, we contacted the statistical institute to request migration data on a more granular five-year age grid. They kindly provided the necessary data, though it is available only for the years 2011 to 2023. We used these finer-grained counts to calculate the age-specific migration schedule. Notably, this dataset contains 8 missing data points out of a total of 3,042 for both males and females. The missing values do not follow a discernible pattern. Instead of performing a complete case analysis, we chose to impute the missing data points. Given the small number of missing values, we opted for single imputation rather than multiple imputation, assuming ignorability – that is, the distribution of the data is the same for both observed and unobserved data points. Missing values were imputed using the `mice` package (multivariate imputations by chained equations) in `R` (van Buuren and Groothuis-Oudshoorn 2011), specifically employing the CART (classification and regression trees) method.

3. Mortality

To produce forecasts of mortality rates we follow the ideas of Goes (2024) using a Renshaw–Haberman (RH) model (Renshaw and Haberman 2006) extended with a spatially structured regional effect. This is done to capture correlation patterns across regions and thereby increase the model fit and predictive accuracy. Let $D_{x,t,r}$ denote the number of deaths and $E_{x,t,r}$ the population at risk or exposure in age group $x \in \{1, \dots, X\}$, year $t \in \{1, \dots, T\}$, and region $r \in \{1, \dots, R\}$. For the number of deaths, we assume a Poisson likelihood with $D_{x,t,r} \sim \text{Poi}(m_{x,t,r} \cdot E_{x,t,r})$, where $m_{x,t,r}$ denotes the underlying

mortality rate, with

$$\log(m_{x,t,r}) = \alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}\gamma_k + \nu_r + \varepsilon_{x,t,r}$$

and $\varepsilon_{x,t,r} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. Here, the parameter α_x denotes the average log-mortality rate at age x and models the general shape of mortality by age. The parameter κ_t estimates the global change over time and γ_k, the global effect of cohort k. The index k is a function of age and period. If the intervals are of different lengths – that is, if the age intervals are M times wider than the period intervals – the cohort index is given by $k = M(X-x)+t$ with $k \in \{1, \ldots, M(X-1)+R\}$ (Heuer 1997). The parameters $\beta_x^{(1)}$ and $\beta_x^{(2)}$ measure the response to changes of κ_t and γ_k, respectively, at age x. The regional effect ν_r captures spatial dependencies, which is modeled using a BYM2 prior proposed by Riebler et al. (2016), an extension of the famous Besag-York-Mollié (BYM) model (Besag, York, and Mollié 1991). An error term $\varepsilon_{x,t,r}$ accounts for overdispersion. To make the model identifiable we invoke the typical constraints – that is, $\sum_x \beta_x^{(1)} = \sum_x \beta_x^{(2)} = 1$ and $\sum_t \kappa_t = \sum_k \gamma_k = \sum_r \nu_r = 0$.

To demonstrate the validity of our proposed model, we evaluate its predictive performance against a classical benchmark: the standard Lee–Carter model (Lee and Carter 1992), estimated separately for each region. Following the approach of Goes (2024), we fit both models using data from 2000 to 2014. We then generate out-of-sample predictions of future deaths for the years 2015 to 2017 and compare their predictive accuracy against observed values. To obtain future deaths for both models, we forecast age-specific mortality and multiply those forecasts by the observed values of exposure. The resulting value is then used as the parameter of a Poisson distribution from which future deaths are drawn. We denote our proposed model as RH_BYM2 and the standard Lee–Carter model as LC.

We calculate the mean absolute error (MAE), the root mean squared error (RMSE), and the coverage. We refrain from calculating the very popular mean absolute percentage error (MAPE), due to the high amount of observed zero deaths in our dataset. The coverage of a prediction interval is defined as the proportion of time that the true value lies within this interval. Ideally, coverage values should be close to their nominal value, which in this case is 0.95. Furthermore, using Bayesian methods allows us to obtain not just point forecasts but an entire predictive distribution, which can be compared using so-called scoring rules. To compare these predictive distributions, we utilize the log score (LogS) and ranked probability score (RPS). For a detailed explanation, readers are referred to Gneiting and Raftery (2007) and Czado, Gneiting, and Held (2009). Similar to point measures, lower scores indicate a better fit. These metrics can be computed in `R` based on samples from the posterior predictive distribution using the `scoringRules` package (Jordan, Krüger, and Lerch 2019).

The results demonstrate that our proposed model provides superior forecasts of age-specific mortality across all metrics. Specifically, it achieves lower values for both point measures (MAE and RMSE) and scoring rules, for both men and women. Additionally, the coverage values are closer to the nominal value of 0.95. The results are shown in Table 1.

Table 1: Out-of-sample performance of mortality forecasts

Model	MAE	RMSE	Coverage	LogS $\cdot 10^-2$	RPS $\cdot 10^{-2}$
Men					
RH_BYM2	**3.47**	**6.27**	**0.97**	**16.65**	**17.42**
LC	4.09	7.28	0.98	17.42	19.92
Women					
RH_BYM2	**3.61**	**8.00**	**0.97**	**15.22**	**18.12**
LC	4.03	8.55	0.97	16.20	20.32

Notes: Values in bold denote the best of the column. MAE = mean absolute error. RMSE = root mean squared error. Coverage denotes the coverage for a nominal value of 0.95. LogS = log score. RPS = ranked probability score.

4. Fertility

Let $B_{x,t,r}$ denote the number of live births that women have in age group x, year t, and region r. Moreover, let $E^{(F)}_{x,t,r}$ denote the female population at risk in age group x, year t, and region r. We assume that age-specific fertility rates $f_{x,t,r}$ are positive for age groups corresponding to the age interval $[15, 45)$ and zero otherwise. This is due to the following reason. The data of births are available in five-year age groups, except for the lowest (< 20) and highest ($40+$) age groups. To generate future births, we need to multiply the age-specific fertility rate with the respective age-specific female population. Therefore, we assume that all of births of the lowest age group occur within the interval of $[15, 20)$ and all of the births of the highest age-group within $[40, 45)$, resulting in positive age-specific fertility rates for the age interval of $[15, 45)$.

The total fertility rate (TFR) is defined as $TFR_{t,r} = \sum_x 5 \cdot f_{x,t,r}$. In addition to the TFR, we can define the age proportion:

$$\phi_{x,t,r} = \frac{5 \cdot f_{x,t,r}}{TFR_{t,r}}.$$

This represents the fraction or proportion of total fertility occurring in age group x. Note that $\sum_x \phi_{x,t,r} = 1$ for all r and t. There are multiple ways to model and later on forecast

fertility rates. One can either model them directly or indirectly. A typical model for direct estimation of age-specific fertility rates is the Lee–Carter model, originally designed for mortality rates but shortly thereafter adapted for age-specific fertility rates (Lee 1993). Alternatively, one can use the UN's approach, where estimates of TFR and age proportion are back-transformed into age-specific fertility rates (see Ševčíková et al. (2016) for details). We refer to the former as the direct approach and the latter as the indirect approach. Bohk-Ewald, Li, and Myrskylä (2018) compare various methods to forecast cohort fertility rates and find both the direct approach using a Lee–Carter model and the indirect approach to be among the best-performing methods.

4.1 Direct estimation

In the direct approach, the age-specific fertility rate $f_{x,t,r}$ is estimated using an extension of the Lee–Carter model with a Poisson likelihood for the total number of births, thus $B_{x,t,r} \sim \text{Poi}\left(f_{x,t,r} \cdot E^{(F)}_{x,t,r}\right)$ with

$$\log(f_{x,t,r}) = \alpha_x + \beta_x \kappa_t + \delta_{x,r} + \varepsilon_{x,t,r}, \tag{2}$$

where $\varepsilon_{x,t,r} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\varepsilon)$. The parameter α_x denotes an age-specific intercept and can be thought of as the initial fertility level, while the parameter κ_t represents the change in fertility over time scaled by an age-specific factor β_x. The model of Equation (2) is different from the standard Lee–Carter model due to the addition of an age-region interaction $\delta_{x,r}$. This effect is included because varying fertility patterns were observed for different regions. Instead of the above interaction term, one could also think of naturally extending the Lee–Carter model by setting $\delta_{x,r} = \beta^{(2)}_x \nu_r$, where $\boldsymbol{\beta}^{(2)} = \left(\beta^{(2)}_1, \ldots, \beta^{(2)}_X\right)^\top$ denotes another vector of age effects and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_R)^\top$ represents a vector of regional effects. However, this approach introduces an additional $X + R - 2$ parameter (subject to two constraints), while the more general $\Delta \in \mathbb{R}^{X \times R}$ with entries $\delta_{x,r}$ introduces a total of $X \cdot R - X$ parameters (subject a corner constraint, details are given below). Since the latter approach introduces more parameters into the model, it is more flexible. Moreover, we compared both the in- and out-of-sample fit of both models and found the approach given in Equation (2) to be superior. Details are omitted for brevity.

Using an age-region interaction term is different from the approach presented in Section 3, where the RH model was extended by the addition of a spatially structured regional effect. It would have also been possible to extend the Lee–Carter model of Equation (2) with the same effect – that is, to set $\delta_{x,r} = s_r$, though we chose the interaction term for the same reason as stated above. Both the in-sample and out-of-sample performance of Equation (2) was better than that of a model with a spatially structured effect. However,

the flexible approach using the interaction term is possible only due to the relatively few age groups and regions within our dataset. For more age groups (e.g., in the mortality model) and/or more regions, the number of parameters to estimate grows substantially and thus increases computational burden. In such a case, a Lee–Carter model extended with a spatially structured effect may be preferred.

To ensure model identifiability of Equation (2), we impose the following constraints: $\kappa_1 = 0$, $\|\boldsymbol{\beta}\|_2^2 = \sum_{x=1}^{X} \beta_x^2 = 1$, $\kappa_t \geq 0$ for all t, and $\delta_{x,R} = 0$ for all x, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_X)^\top$ denotes the vector of age parameters. Note that $\boldsymbol{\beta}$ is constrained to have a norm of one, diverging from the more common sum-to-one constraint. This choice was made primarily for computational reasons, as we observed this specification to be more stable in `Stan`, the software we use for parameter estimation, than the more common sum-to-one constraint. However, this led to us having to set the additional constraint of $\kappa_t \geq 0$ (see Section 6 for details).

4.2 Indirect estimation

The indirect approach forecasts both TFR and age proportions and then converts them into age-specific fertility rates using

$$f_{x,t,r} = \frac{\phi_{x,t,r} \cdot \text{TFR}_{t,r}}{5}.$$

This method is implemented by the UN for their world population forecasts (Ševčíková et al. 2016). Thus, we need to obtain regional estimates of TFR and age proportions to then back-transform these into age-specific fertility rates.

Ševčíková, Raftery, and Gerland (2018) propose a method to predict subnational TFRs. Similar to them, we assume that the regional TFR follows a Bayesian hierarchical autoregressive model of order one, AR(1). More precisely,

$$\text{TFR}_{t,r} - \mu_r = a_r \cdot (\text{TFR}_{t-1,r} - \mu_r) + \omega_{t,r}, \tag{3}$$

with $\omega_{t,r} \overset{iid}{\sim} \mathcal{N}(0, \sigma_r^2)$, a region specific intercept μ_r, and autoregressive parameter a_r. Both the intercept and autoregressive parameters are assigned hierarchical priors (see Appendix section "Overview on choice of priors" for an overview on the selected priors). In Equation (3), region-specific TFRs are modeled directly, in contrast to Ševčíková, Raftery, and Gerland (2018), where subnational TFR estimates are obtained by scaling country-specific TFR estimates using a regional scaling factor. This scaling factor is proposed to follow an AR(1) model with a mean of one, implying that in the long term,

forecasts of subnational TFRs converge to the respective TFR forecast of the country. For some regions, such as those with a significant population of students, this is an unrealistic assumption. To circumvent this issue, Yu et al. (2023) adjust the TFR data for student populations by scaling past TFR values. However, this requires additional data, and it is unclear if this adjustment performs reliably. We therefore model the regional TFRs directly, assuming a distinct mean for each region.

To forecast age proportions, Ševčíková et al. (2016) propose a linear combination of the past observed trend and the past global trend of all age proportions in the dataset. Again, this method seems unsuitable for regions with college populations, as these areas may exhibit distinct behavior compared to other regions. Moreover, their method is deterministic. That is, they do not specify a probability distribution for the age proportion, making it impossible to generate probabilistic forecasts for $\phi_{x,t,r}$. To circumvent these issues, we propose to model the regional age proportions using a Dirichlet regression (see e.g., Gueorguieva, Rosenheck, and Zelterman (2008)), given their inherent constraints of $\sum_x \phi_{x,t,r} = 1$ and $\phi_{x,t,r} \in [0, 1]$. Let $\boldsymbol{\phi}_{t,r} = (\phi_{1,t,r}, \ldots, \phi_{X,t,r})^\top$ denote a multivariate random variable with $X = 6$ groups, then $\boldsymbol{\phi}_{t,r} \sim \text{Dirichlet}(\psi_{1,t,r}, \ldots, \psi_{X,t,r})$, where

$$\log(\psi_{x,t,r}) = \alpha_x + \beta_x \kappa_t + \delta_{x,r}.$$

The linear predictor $\log(\psi_{x,t,r})$ has the same structure as that of a Lee–Carter model extended with an age-region interaction $\delta_{x,r}$. To make the model identifiable we impose the same constraints as for Equation (2) – that is, $\kappa_1 = 0$, $\|\boldsymbol{\beta}\|_2^2 = \sum_{x=1}^{X} \beta_x^2 = 1, \kappa_t \geq 0$ for all t and $\delta_{x,R} = 0$ for all x. An explanation for their choice is given in Section 6. We tested multiple specifications for the linear predictor of $\psi_{x,t,r}$, such as a random intercept random slope model with the inclusion of an age-region interaction effect. Nonetheless, the Lee–Carter structure performed best according to the Watanabe-Akaike Information Criterion (WAIC; Watanabe 2010) as well as out-of-sample criteria. Results are omitted for brevity. To the best of our knowledge, this is the first attempt to estimate age proportions using a Dirichlet regression.

4.3 Out-of-sample comparison

We now assess the predictive performance of both approaches to ultimately decide which to use in the population projection. We fit both approaches on a subset of the data, specifically for the years 1995 to 2015, and compare their predictive accuracy against the ground truth, which is the observed number of births between 2016 and 2023. Forecasts for births are obtained as follows: For the direct approach, forecasts of age-specific fertility rates are multiplied by known exposure values. We then draw samples from a Poisson distribution, using the product of age-specific fertility and exposure as the distribution's pa-

rameter. For the indirect approach, forecasts of age-proportion and TFR are transformed to obtain forecasts of age-specific fertility rates, which are then multiplied by known exposure values to yield birth forecasts. Additionally, we compare the predictive accuracy of our models with the standard Lee–Carter model, estimated for each region separately, a method implemented by Myrskylä, Goldstein, and Cheng (2013) for predicting cohort fertility rates across multiple countries, showing good predictive performance. As with the evaluation of mortality rates, we calculate the MAE, RMSE, coverage, and both scoring rules for fertility forecasts. Additionally, we include the MAPE, since there are no observed zero births for any specific age group in our dataset.

Examining the out-of-sample performance of all models, we find that the direct approach – more specifically, the Lee–Carter extension with an age-region interaction term described in Equation (2) – is superior in the most categories, albeit by a small margin. This is followed by the indirect approach, and then the Lee–Carter approach (see Table 2). The mean forecast of the direct and indirect approach are very close to each other, with the former being marginally better in terms of MAE and RMSE, while the latter is slightly superior according to the MAPE. The standard Lee–Carter method performs worse based on the point measures. Furthermore, the direct and indirect approaches have coverages of 0.9, which is close to the nominal value of 0.95, and thus they effectively capture the underlying uncertainty. However, the standard Lee–Carter model exhibits the highest coverage values. This can be explained by the fact that the Lee–Carter method allows the error term for each region to have a different variance. In contrast, the error term in the direct approach has the same variance for all regions. Looking at the scoring rules, we find the predictive distributions of the two proposed approaches outperform the standard Lee–Carter model. According to the LogS, the indirect approach performs slightly better, while the direct approach is superior in terms of the RPS. However, the calculation of the LogS based on samples from the posterior predictive distribution is highly variable, whereas the variability is lower for the RPS when calculated using the empirical posterior predictive cumulative distribution function. Therefore, Krüger et al. (2021) recommend the use of the RPS when estimating the scoring rules based on draws from the posterior predictive distribution. Consequently, we place more value on the RPS results and choose to use the direct approach to generate our forecasts of age-specific fertility rates, which will serve as an input for the cohort-component method. Nonetheless, both proposed methods perform better than the classical Lee–Carter model.

Table 2: Out-of-sample performance of the three fertility approaches

Model	MAE	RMSE	MAPE	Coverage	LogS $\cdot 10^{-3}$	RPS $\cdot 10^{-3}$
Direct approach	**14.76**	**22.28**	18.13	0.90	2.59	**6.72**
Indirect approach	14.85	22.89	**17.45**	0.90	**2.57**	6.78
Lee–Carter	17.74	26.80	22.98	**0.92**	2.64	7.58

Notes: Values in bold denote the best of the column. MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95. LogS = log score. RPS = ranked probability score.

5. Migration

In Section 1 we mentioned that forecasts of net-migration by age are needed to produce regional population forecasts. Often, net-migration totals are distributed across age to obtain age-specific net-migration rates. However this does not take into account the different age-distributions of in- and out-migration, and net-migration rates do not have a consistent age pattern, which give biased results (Rogers 1990). Therefore, to estimate subnational migration, we broadly follow the approach of Ševčíková, Raymer, and Raftery (2024). They decompose forecasts of total net-migration counts into future in- and out-migration counts, which are subsequently transformed into age-specific in- and out-migration counts. In general, instead of counts, one can also model net-migration rates and then transform these into age-specific in- and out-migration counts (see Welch, Ševčíková, and Raftery 2024). However, as data for net-migration counts are available, we opted for the former approach.

Germany experienced a steady increase in net-migration in the early 2010s, peaking in 2015 with a net total of 1,139,402 according to the Federal Office for Migration and Refugees (FOMR 2016), which is a 50% increase compared to the previous year, with Syria being the primary country of origin due to the Syrian civil war. The full scale invasion of Russia against Ukraine in 2022 triggered another refugee movement in Europe – the biggest since World War II – resulting in Germany experiencing an increase in net-migration totals of over 400% compared to the year prior, reaching a net total of 1,462,089 according to the Federal Ministry of Interior, Build and Community (FMI) and the FOMR (FMI and FOMR 2024). These two events can be considered as two one-period migration shocks that need to be accounted for when modeling net-migration counts for regions in Germany or more broadly across Europe. As expected, both effects can be seen when plotting net-migration counts of our regions in Upper Franconia with pronounced peaks in 2015/2016 and 2022 (see Figure 1). Furthermore, we observe that the migration shock related to the Syrian civil war was lagged by a year for a single

region, specifically the city of Bamberg. This is due to the so-called AnkER facilities (standing for 'arrival, decision, and return' facility), which were established in parts of Germany in 2016 to centralize asylum application processes before distributing migrants to other areas of the country. The only AnkER facility in Upper Franconia is situated in the city of Bamberg (FOMR 2021).

Figure 1: Net-migration counts for all regions in Upper Franconia for both females and males

Notes: Each of the 13 regions is colored differently. Dashed lines in 2015 and 2022 denote years with unusually high net-migration totals in Germany.

To model total net-migration rates for multiple regions, Azose and Raftery (2015) propose a hierarchical AR(1) approach using Gaussian error terms, which can also be used to model total net-migration counts. However, such error terms are not ideal for handling past migration shocks, as these events primarily manifest as positive outliers. A normal or t-distribution is, however, symmetric, meaning positive and negative migration shocks are equally likely. This is not reflected in the data. Fernández and Steel (1998) introduce a general class of skewed error terms, which can be included into the autoregressive modeling equation of Azose and Raftery (2015). This allows for explicitly

modeling positive outliers and therefore better to describe the observed data while also allowing for the possibility of a future one-time shock event. Consider a univariate random variable with a probability density function $f(\cdot)$ that is continuous, unimodal, and symmetric around 0 (e.g., a normal or a t-distribution). For ease of notation, we assume the scale parameter of $f(\cdot)$ to be one, though in theory the scale parameter of $f(\cdot)$ can be any positive value. An additional parameter $\gamma \in (0, \infty)$ accounts for the degree of skewness, generating the following class of skewed distributions:

$$p(\varepsilon|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \left[f\left(\frac{\varepsilon}{\gamma}\right) 1_{[0,\infty)}(\varepsilon) + f\left(\varepsilon\gamma\right) 1_{(-\infty,0)}(\varepsilon) \right]. \tag{4}$$

Here, $1_{[0,\infty)}(\varepsilon)$ denotes an indicator function that equals one if ε lies in the interval $[0, \infty)$ and zero otherwise. The parameter γ accounts for the degree of skewness, with $\gamma = 1$ indicating no skewness, while $\gamma \in (0, 1)$ indicates left and $\gamma \in (1, \infty)$ right skewness. For further details, we refer to Fernández and Steel (1998). We can use Equation (4) to introduce skewed normal error terms with a heavy right tail, allowing for a higher probability of positive outliers – that is, positive mortality shocks.

To model total net-migration counts we employ the AR(1) model of Azose and Raftery (2015) with a skewed normal error term given by

$$N_{t,r} - \mu_r = a_r \cdot (N_{t-1,r} - \mu_r) + \varepsilon_{t,r} \cdot \sigma_r, \tag{5}$$

where $\varepsilon_{t,r}$ follows a skewed normal distribution given in Equation (4), scaled by a region specific variance parameter σ_r. Instead of using skewed errors, we could also follow the ideas of Goes, Barigou, and Leucht (2025) and introduce a global migration shock parameter, where for a given time period $\tilde{t}$ the migration counts for all regions are simultaneously shifted upwards. However, as evident in Figure 1, not all regions experienced the migration shock of the Syrian civil war in 2015 at the same time, making the method less suitable for the data at hand.

5.1 Out-of-sample validation

To validate our methodology – that is, to see if the skewed normal error terms of Equation (4) provide better out-of-sample forecasts compared with the Gaussian errors – we compare the predictive accuracy of both approaches in conjunction with an AR(1) structure. We estimate the model of Equation (5) with both skewed normal errors as well as Gaussian errors. The former we call Azoze–Raftery skewed while the latter is simply denoted as Azoze–Raftery model. To generate forecasts using Equation (5), we need the ability to

draw random samples from Equation (4). This is achieved by generating random draws from a uniform distribution $u \sim \text{Unif}(0, 1)$ and plugging those draws into the quantile function of Equation (4). Derivation of the distribution as well as the quantile function of Equation (4) can be found in the Appendix.

The models were fitted on data from 2000 to 2018 and future net-migration counts for 2019 to 2023 were forecasted to assess their predictive accuracy. A different out-of-sample window was specifically chosen compared to that in Section 4 to allow the data to return to a more typical level after the high net-migration values observed in 2015 to 2017 (cf. Figure 1). After generating forecasts of net-migration counts for both models, these were compared against the true observed counts. As before, we analyzed the predictive accuracy of the posterior mean using classical point measures and also compared the entire predictive distribution using scoring rules. For men, we observe a comparable performance in terms of point measures, which is unsurprising given that both models share the same underlying structure. However, when examining the coverage and scoring rules, the skewed error term proved more effective at describing the predictive distribution, indicated by lower scores and broader coverage. For women, we observe similar findings: Both approaches yield comparable point forecasts, yet the skewed error terms provide a superior description of future migration shocks. The detailed results are presented in Table 3.

Table 3: Out-of-sample performance of net-migration forecasts

Model	MAE	RMSE	MAPE	Coverage	LogS	RPS $\cdot 10^{-3}$
Men						
Azoze-Raftery skewed	**272.45**	**362.21**	150.37	**0.78**	**485.07**	**13.51**
Azoze-Raftery	280.98	378.13	**149.45**	0.71	493.92	13.88
Women						
Azoze-Raftery skewed	**228.55**	**329.64**	90.67	0.85	**475.81**	**11.52**
Azoze-Raftery	230.14	331.79	**89.23**	0.85	488.54	11.68

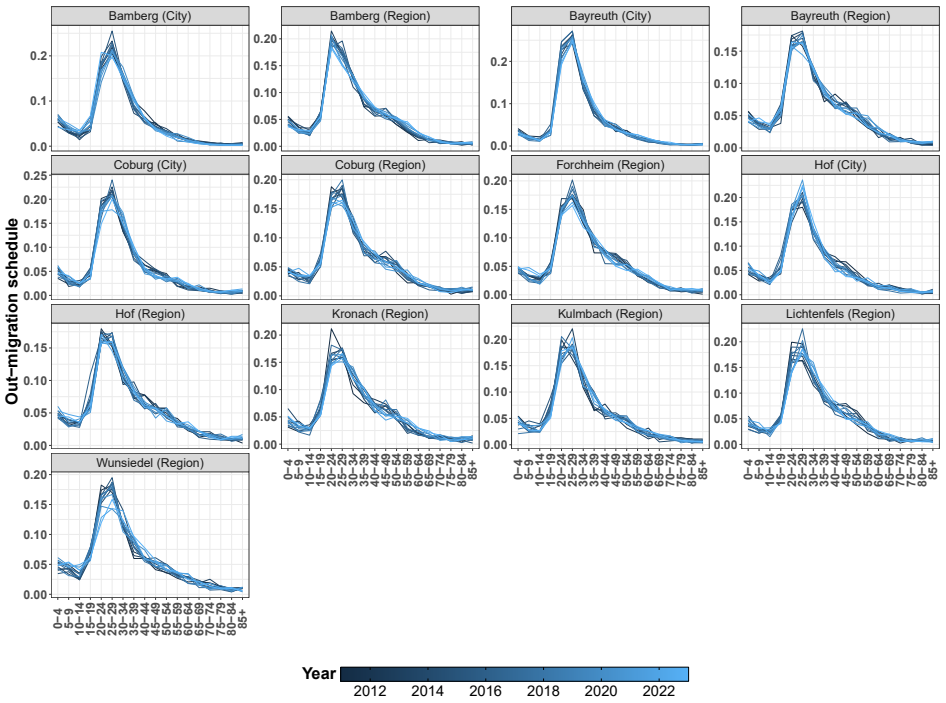
Notes: Values in bold denote the best of the column. MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95. LogS = log score. RPS = ranked probability score.

5.2 Age-specific migration

As mentioned previously, the total migration counts have to be transformed into age-specific counts. This is achieved by multiplying them by a so-called migration schedule, which is defined as the age distribution of migration. Looking at the age-specific in- and out-migration counts from 2011 to 2023, we notice that the amount may change, yet

the age distribution or pattern stays fairly consistent. The age distribution of the out-migration counts for a specific region is calculated by $R_{x,t,r}^{(out)} = O_{x,t,r} / \sum_{x=1}^{X} O_{x,t,r}$, where $\sum_x R_{x,t,r}^{(out)} = 1$ by construction. The age distribution of the in-migration counts is calculated similarly. A graphical representation of the age distribution of the male out-migration schedule, demonstrating a time-invariant pattern, is shown in Figure 2. The same time-invariant pattern can also be observed for females and is shown in the Appendix in Figure A-1.

Figure 2: Out-migration schedule over time of men for all regions in Upper Franconia



The concept of time-constant migration schedules is known in the literature and was first mathematically formalized by Rogers and Castro (1981), who introduced a statistical model to estimate this time-independent migration schedule.

Hence, using such a model, we can multiply migration totals by time-invariant migration schedules to obtain age-specific migration counts.

Ševčíková, Raymer, and Raftery (2024) propose a method to decompose total net-migration counts, $N_{r,t}$, into in-, $I_{r,t}$, and out-migration counts, $O_{r,t}$, using a mixed-effects model, which we will employ as well. These counts are then transformed into age-specific in- and out-migration counts by multiplying the total in- and out-migration counts with estimated migration schedules. More specifically, we calculate by sex

$$\begin{aligned} I_{t,r} &= \beta_r^{(0)} + \beta_1 N_{t,r} + \varepsilon_{t,r} \\ o_{x,t,r} &= O_{t,r} \cdot \hat{R}_{x,t,r}^{(Out)} \\ i_{x,t,r} &= I_{t,r} \cdot \hat{R}_{x,t,r}^{(In)}, \end{aligned}$$

where $O_{t,r} = I_{t,r} - N_{t,r}$, $\varepsilon_{t,r} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{Mixed})$ and $\beta_r^{(0)}$ denotes a region specific intercept, and β_1 a slope parameter. Here, $o_{x,t,r}$ and $i_{x,t,r}$ denote the age-specific out- and in-migration counts, respectively. Moreover, let $\hat{R}_{x,t,r}^{(Out)}$ and $\hat{R}_{x,t,r}^{(In)}$ denote the estimated migration schedules derived from the data, which are assumed to be constant over time. The most well-known model for estimating migration schedules is the so-called Rogers–Castro curve, proposed by Rogers and Castro (1981), with a Bayesian implementation by Yeung, Alexander, and Riffe (2023). While staying within the Bayesian framework, we attempted to implement Yeung, Alexander, and Riffe (2023)'s model using multiple time periods as input data. However, we encountered problems with convergence in `Stan` (see Section 6). We therefore choose to model the age-migration schedule using a Dirichlet regression, due to its implied sum-to-one constraint. Let $R_{x,t,r}$ denote the age pattern of migration, with the index for out- or in-migration dropped for notational convenience.

We assume $\boldsymbol{R}_{t,r} = (R_{1,t,r}, \ldots, R_{X,t,r})^\top$ to be Dirichlet distributed, thus

$$\boldsymbol{R}_{t,r} \overset{iid}{\sim} \text{Dirichlet}(\psi_{1,r}, \ldots, \psi_{X,r}).$$

This implies that each region has its own set of age-specific parameters that remain constant over time. Using the posterior predictive distribution we can generate samples of both $\hat{R}_{x,t,r}^{(Out)}$ and $\hat{R}_{x,t,r}^{(In)}$.

We compare the performance of the Bayesian implementation of the Rogers–Castro model for a single time period with our Dirichlet model. We find their performance to be similar, with our Dirichlet model even showing better in-sample fit. Further details are provided in the Appendix.

6. Estimation

The models are fitted in a Bayesian framework using the probabilistic modeling software `Stan` (Stan Development Team 2024c), a tool for Hamiltonian Monte Carlo (HMC). Stan can be accessed through the interface `rstan` in `R` (Stan Development Team 2024a). Built-in diagnostic measures are used to check for convergence, which are described in Vehtari et al. (2021). The parameters of all models are given weakly informative priors. Details regarding the specific choice can be found in the Appendix. It should be noted that posterior predictive checks revealed a good fit of all models to the data. For the sake of brevity the results are omitted. The code and data for all models is available on GitHub.[5]

As mentioned in Section 4, we impose the vector of age parameters $\boldsymbol{\beta}$ in Equation (2) to have a norm of one. However, this does not lead to a unique solution. More precisely, two sets of parameterizations for $\boldsymbol{\beta}$, which we denote $\mathcal{B}$ and $-\mathcal{B}$, yield the same norm, resulting in a non-unique set of parameters. We therefore have to additionally impose that $k_t \geq 0$ for all t. Despite this, we occasionally encountered convergence issues in `Stan` where estimates of $\kappa_1, \ldots, \kappa_T$ would behave differently in one chain than in another. To mitigate this, we implement another set of constraints and impose that $k_1 < k_{T/2} < k_T$, which alleviated the problem. Still, these constraints are solely due to computational reasons and might not have to be imposed using a different estimation scheme.

In addition, it is important to note that the age pattern of migration introduced in Section 5 can theoretically take on values of zero and one – that is, $R_{x,t,r} \in [0,1]$. However, the probability density function of the Dirichlet distribution is zero if any of its components (i.e., any of $R_{x,t,r}$) is zero. In addition, `Stan` works with log posteriors, which are undefined at zero. Therefore, the components must satisfy $R_{x,t,r} \in (0,1)$ (see Stan Development Team 2024b: ch. 26.1). In our data, the in-migration count for males was zero for a single age group in a single region and single year, leading to a value of 0 in the migration schedule, which we will denote $R_{x^*,t^*,r^*} = 0$. To circumvent this issue, we added a small constant and set this specific value to $R_{x^*,t^*,r^*} = 1 \times 10^{-6}$. Afterward, we rescaled the migration schedule so that it sums to one for this specific year and region.

The computation time for each input – mortality, fertility, and net migration – varies with model complexity and the number of parameters. More specifically, due to the interaction terms, the fertility models take the longest, while the relatively simple AR(1) model has a shorter runtime. Despite these differences, `Stan`'s parallel processing capabilities ensure that the overall runtime remains moderate: The migration model completes in approximately one minute, while the fertility model requires up to 60 minutes on an Intel i5-8365U CPU (1.60 GHz) with 16 GB of RAM. However, we anticipate that computation time will increase with larger datasets, such as those including more regions or years.

[5] See https://github.com/goesj/Population-Forecast.

7. Population forecasts

Before running the full model to obtain population forecasts for all 13 regions in Upper Franconia, we first want to validate our method in the short term. Following this validation, we will present the full population forecasts for all regions from 2024 to 2044.

7.1 Out-of-sample validation

To validate our methodology, we conducted an out-of-sample population forecast and compared the predicted population with the observed one. Since the data of the population is given in age groups of five, our projection period is of the same length. Thus, we can project the population only one projection interval (i.e., five years) at a time. We use the observed data until 2019 and forecast the population for a single projection period, from 2020 to 2024. For this, we estimate the input parameters as follows: The mortality model is trained from 2000 to 2017 (due to a lack of available data, see Section 2) and then used to forecast for the years 2018–2024. Fertility and net-migration are estimated from 1995–2019 and 2000–2019, respectively, with out-of-sample predictions generated for 2020–2024. Afterward, we compare the predicted population for each age group and region to the observed values. Table 4 presents the validation results, including the RMSE, MAE, and coverage.

Table 4: Out-of-sample validation of the age and region-specific population forecast

MAE	RMSE	MAPE	Cov
115.46	207.22	2.83	0.85

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95.

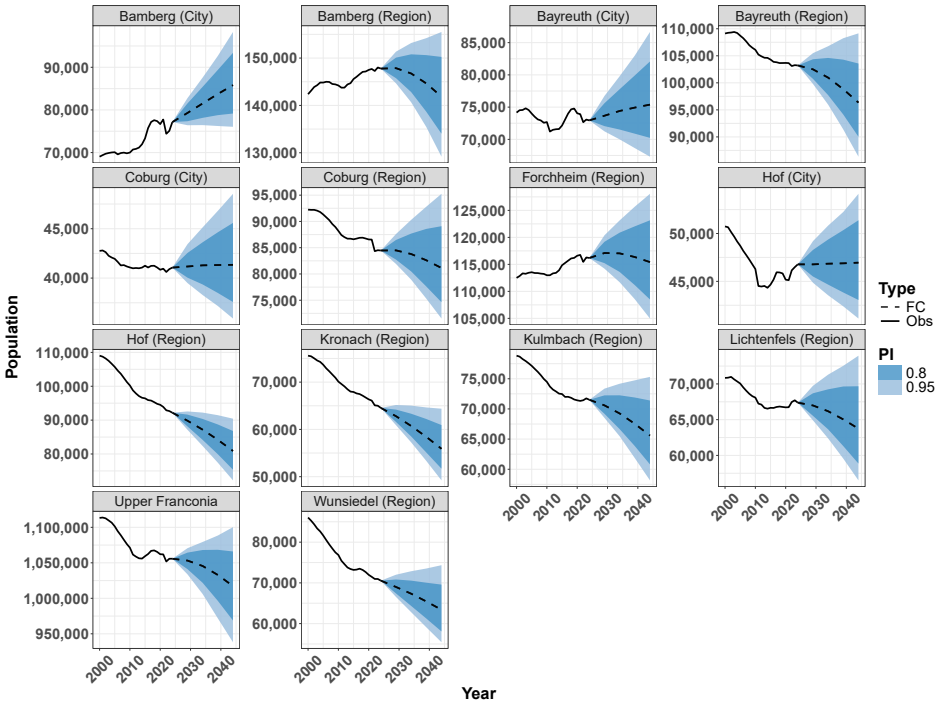
We observe a coverage value of 0.85, which is consistent with the coverage values of the out-of-sample forecasts for net-migration and fertility. Coverages for out-of-sample forecasts are typically smaller than their nominal value due to the increased uncertainty inherent in out-of-sample predictions. However, this coverage remains within a satisfactory range and is comparable to population forecasts by other authors (e.g., Alexander and Alkema 2022). We therefore conclude that the model adequately depicts the uncertainty of the population forecasts. Furthermore, the MAPE value is relatively small, indicating that the forecasted population closely approximates the true observed population. Overall, these results suggest that our model performs well in forecasting the

age-specific population, albeit validated only in the short term. To thoroughly evaluate the performance of long-term age-specific forecasts, additional data are required.

7.2 Results

We run our model and obtain population forecasts, disaggregated by age, region, and sex, for the period from 2024 up to 2044 in five-year intervals. The total population projections, including uncertainty intervals for all regions, are presented in Figure 3. In this figure, the black solid lines represent the observed values until 2024, while the dashed line indicates the median forecasts, with associated uncertainty intervals depicted in varying shades of blue.

Figure 3: Population forecasts for Upper Franconia including all regions separately



Note: The observed data is shown by the solid line, while the median forecasts are given by the dashed line. The 80% prediction intervals are plotted in the darker shade of blue while 95% prediction intervals are shown in the lighter shade of blue.

Starting with a total population of 1,055,758 in 2024, the overall number of people in Upper Franconia is expected to slightly decrease by approximately 4% over the projection period. The median forecast of the total population in 2044 is 1,010,867 with an 80% prediction range of 968,694 and 1,065,963. However, this pattern is not uniform across all regions. Forecast results of the province as well as all regions separately are shown in Figure 3.

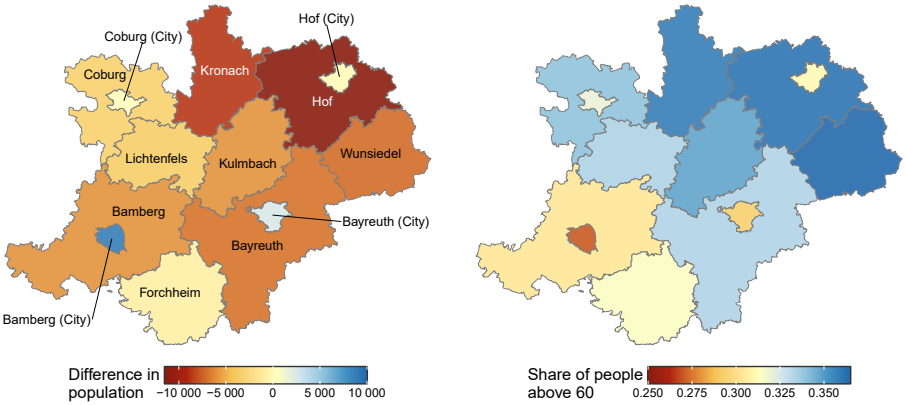
Generally, we observe three distinct types of patterns that appear to associate with region type: long-term decline, long-term population growth, and constant population with the possibility of either growth or decline. More precisely, Bavaria, Germany, comprises of two different types of NUTS-3 regions: First, there are regions where multiple smaller cities and villages share an administrative body, denoted as 'Region' in Figure 3. Second, there are city regions, denoted as 'City' in Figure 3, where a single larger city has its own administrative boundary. We anticipate these city regions to behave differently from their counterparts for two main reasons: First, these city regions can be considered the economic engines of Upper Franconia, as evidenced by their higher GDP values (cf. Figure A-2). We therefore expect them to have a higher number of net-migration. Second, all these city regions are considered medium-sized cities (50,000–100,000 inhabitants), which have generally experienced population growth in Germany over the last 15 years, while rural municipalities have seen a decline (Wolff et al. 2022). Examining the population forecasts for these city regions, we observe either likely population growth (Bamberg, Bayreuth) or a constant population (Coburg, Hof), though with a considerable degree of uncertainty.

For the non-city regions, the pattern is quite different. Most are expected to experience a population decline due to low net-migration and a high proportion of older people. In these cases, the low net-migration cannot counteract the negative difference between the expected number of deaths and births, leading to a shrinking population. To better illustrate the negative relationship between population growth and age structure, we plotted the difference in forecasted median population between 2044 and 2024 alongside the share of people over 60 years in 2024 on a map (see Figure 4). This visualization clearly shows that regions with a high share of older people in 2024 (primarily in the northeast of Upper Franconia) are more likely to experience population decline in the future.

Furthermore, two regions particularly stand out: the region of Bamberg and Forchheim. Both have experienced a fairly constant population growth over the past 20 years, a trend expected to continue in the immediate future. However, this pattern is projected to reverse according to the median forecasts. That is, after a period of growth, a population decline becomes more likely, especially in the region of Bamberg. This is due to an anticipated increase in deaths that is expected to occur in the future. The reason being that even though these regions have a younger age structure than other members of the region group, people aged between 50 and 60 still constitute the highest share of the population.

In 20 years, however, these individuals will be between 70 and 80, leading to an increase in expected deaths that projected net-migration may not be able to counteract, thereby reversing the trend of population growth.

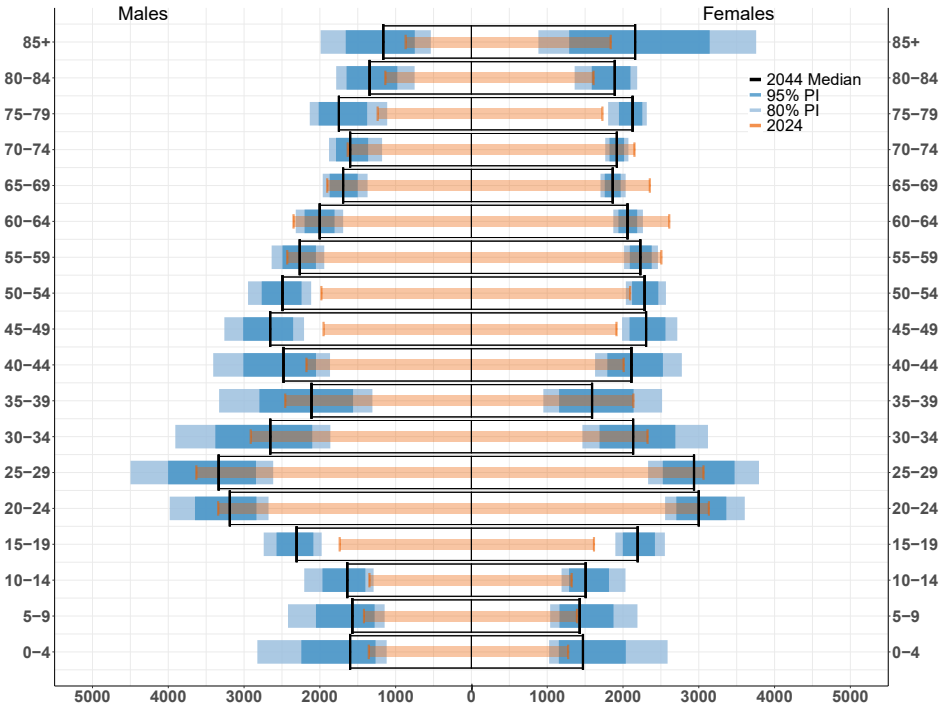
Figure 4: Difference in population between 2024 and median forecasts for 2044



Note: The information (Region) was dropped to shorten the name of certain regions for a more compact visualization.

In addition to serving as an economic hub, several of the city regions – such as Bamberg, Bayreuth, and Coburg – are home to universities, contributing to a significant population of college-aged residents. It is not expected that this population age structure will change. Rather, the share of people between 20 and 30 remains constant over time due to high in-migration counts in early adulthood and high out-migration counts after having graduated university. Looking at the probabilistic population pyramid for the city of Bayreuth (cf. Figure 5), we see that the total amount of people between 20 and 30 remain relatively constant over time. Thus, our migration model effectively preserves the age-specific shape of these regions into the future.

Figure 5: Population pyramid for Bayreuth (City)



8. Discussion

In this article, we have developed a probabilistic forecasting method of age-specific population including age-specific mortality, fertility, and net-migration rates for small areas. We introduced new approaches to forecast both regional age-specific fertility rates and regional net-migration counts. Specifically, we extended the standard Lee–Carter model for age-specific fertility rates by including an age-region interaction term, which resulted in improved predictions across both point measures and scoring rules. Additionally, we incorporated the concept of skewed error terms to better capture and predict future migration shocks. Our findings demonstrate that using skewed error terms led to more accurate predictions of net-migration counts compared to the standard approach that relies on sym-

metric normal errors. Lastly, we introduced a new method to estimate the age-proportion of fertility and the age-specific migration schedule using a Dirichlet regression.

The model was applied to real data to obtain population forecasts for 13 heterogeneous regions in Bavaria, Germany. The long-term predictions (2024–2044) reveal heterogeneous regional trends, differentiating between stable or growing city populations driven by economic factors and a high share of students, as well as declining populations in non-city regions attributed to aging and insufficient net-migration. While we did not compare the out-of-sample performance of our population forecasts with that of other methods, we found that our proposed methods for each input variable of the cohort-component method consistently outperformed other standard approaches. This is of significant importance, as improved accuracy in forecasting individual components directly translates to more precise population forecasts overall, due to the underlying fundamental equation.

The method that we have proposed is very general in a sense that it can be applied to forecast age-specific populations of other small areas – for example, other NUTS-3 regions. In addition, no adjustments are needed when applying it to predict the age-specific population on a national level. Furthermore, it is unclear whether the indirect approach to forecast age-specific fertility rates performs worse, albeit slightly, than the direct approach due to the simple AR(1) model for regional TFRs, the Dirichlet regression for the age proportions, or a mix of both. Adding a regional correlation structure to model regional TFRs might improve predictions (Fosdick and Raftery 2014), though more research should be done to see if the Dirichlet regression is suitable in forecasting age proportions or not. Nonetheless, it constitutes an interesting addition to the demographic literature.

There are two valid points of criticism regarding our model. First, it can be considered relatively complex and data-intensive. This could pose an entry barrier for some users in statistical agencies, as it involves significant production costs – particularly in terms of staff time – and requires familiarity with both statistical methods and software. These issues should be considered when deciding between competing methods, not just their predictive performance (cf. Smith, Tayman, and Swanson 2013: ch. 12). Moreover, some evidence suggests that simpler models can perform just as well as more complex ones when it comes to forecasting (Green and Armstrong 2015). Simpler extrapolation methods – such as autoregressive integrated moving average (ARIMA) models, which are fit directly into the population data – are less data-intensive and therefore provide a valuable alternative. Nevertheless, most of the data needed by our model should be available, at least for Western countries. Furthermore, when predicting population by age group, one must account for shifts in the age structure, which implies that some variant of the cohort-component method is needed (Smith, Tayman, and Swanson 2013). In addition, the cohort-component method is directly linked to the fundamental population equation, the cornerstone of demographic theory. This positions the method at the core of

population studies, providing a solid theoretical foundation. Methods that use the cohort-component model therefore offer theoretical advantages over other approaches that lack this foundation (Burch 2018). Second, our proposed method is computationally demanding, as it requires repeated sampling from a complex posterior distribution. However, this enables the generation of probabilistic population forecasts, including credibility intervals. A quicker and lower-cost alternative based on cohort-change ratios, such as the Hamilton–Perry model (Hamilton and Perry 1962), is deterministic in nature and produces only point forecasts. Moreover, adopting a frequentist approach for parameter estimation in our proposed model would reduce the computational burden, although it would limit the results to point forecasts as well. To obtain prediction intervals in a frequentist framework, a bootstrap procedure would be required, significantly increasing computational costs and thus negating the advantage of faster computation times.

Using Bayesian methods we have accounted for many sources of uncertainty; however, we have not considered model uncertainty. That is, we have not accounted for model misspecification when predicting each input variable for the cohort-component method. One approach to explicitly incorporate model uncertainty into the prediction problem is the stacking approach by Yao et al. (2018), with implementations in demographic literature by Barigou et al. (2023) and Goes (2024). Adding model uncertainty may help both in improving the predictive accuracy of our method and in providing an even more realistic description of the underlying uncertainty.

Finally, we have assumed that all inputs of the cohort-component method – namely mortality, fertility, and net-migration – are independent of each other, which may be a problematic assumption. For example, a significant increase in net-migration in a specific region could impact the fertility pattern of that region in the subsequent years, something not considered in our approach. Additionally, the population is forecasted for each sex separately. Allowing for correlation between sexes – similar to Wiśniowski and Raymer (2025) – or perhaps more importantly between all inputs is an interesting direction of future research.

9. Acknowledgments

The authors would like to thank Chen-Hao Hsu for valuable discussion on the topic of fertility. Moreover, Julius Goes gratefully acknowledges financial support by the Oberfrankenstiftung (grant FP01054).

References

Alexander, M. and Alkema, L. (2022). A Bayesian cohort component projection model to estimate women of reproductive age at the subnational level in data-sparse settings. *Demography* 59(5): 1713–1737. doi:10.1215/00703370-10216406.

Alexander, M., Zagheni, E., and Barbieri, M. (2017). A flexible Bayesian model for estimating subnational mortality. *Demography* 54(6): 2025–2041. doi:10.1007/s13524-017-0618-7.

Azose, J.J. and Raftery, A.E. (2015). Bayesian probabilistic projection of international migration. *Demography* 52(5): 1627–1650. doi:10.1007/s13524-015-0415-0.

Barigou, K., Goffard, P.O., Loisel, S., and Salhi, Y. (2023). Bayesian model averaging for mortality forecasting using leave-future-out validation. *International Journal of Forecasting* 39(2): 674–690. doi:10.1016/j.ijforecast.2022.01.011.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43(1): 1–20. doi:10.1007/BF00116466.

Bijak, J. and Bryant, J. (2016). Bayesian demography 250 years after Bayes. *Population Studies* 70(1): 1–19. doi:10.1080/00324728.2015.1122826.

Bohk-Ewald, C., Li, P., and Myrskylä, M. (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences* 115(37): 9187–9192. doi:10.1073/pnas.1722364115.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* 22(3): 547–581. doi:10.1016/j.ijforecast.2006.04.001.

Bryant, J. and Zhang, J.L. (2016). Bayesian forecasting of demographic rates for small areas: Emigration rates by age, sex, and region in New Zealand, 2014–2038. *Statistica Sinica* 26: 1337–1363. doi:10.5705/ss.2014.200t.

Burch, T.K. (2018). *Model-based demography. Essays on integrating data, technique and theory*. Demographic Research Monographs. Cham: Springer International Publishing. doi:10.1007/978-3-319-65433-1.

Cameron, M.P. and Poot, J. (2011). Lessons from stochastic small-area population projections: The case of Waikato subregions in New Zealand. *Journal of Population Research* 28(2–3): 245–265. doi:10.1007/s12546-011-9056-3.

Congdon, P. (2014). Estimating life expectancies for US small areas: A regression framework. *Journal of Geographical Systems* 16(1): 1–18. doi:10.1007/s10109-013-0177-4.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics* 65(4): 1254–1261. doi:10.1111/j.1541-0420.2009.01191.x.

Dharamshi, A., Alexander, M., Winant, C., and Barbieri, M. (2025). Jointly estimating subnational mortality for multiple populations. *Demographic Research* 52(3): 71–110. doi:10.4054/DemRes.2025.52.3.

Federal Ministry of Interior and Community and Federal Office for Migration and Refugees (2024). Migration report of the Federal Government 2022. Executive summary (Tech. rep. BAMF). Nuremberg: BAMF. doi:10.48570/bamf.fz.kurz.mb2022.en.2024.migrationreport.1.0.

Federal Office for Migration and Refugees (2016). Migration report 2015. Central conclusions (Tech. rep. BAMF). Nuremberg: BAMF. https://www.bamf.de/SharedDocs/Anlagen/EN/Forschung/Migrationsberichte/migrationsbericht-2015-zentrale-ergebnisse.html.

Federal Office for Migration and Refugees (2021). Evaluation of AnkER facilities and functionally equivalent facilities. (Research Report 37). Nuremberg: BAMF. https://www.bamf.de/SharedDocs/Anlagen/EN/Forschung/Forschungsberichte/fb37-evaluation-anker-fg-einrichtungen.html.

Fernández, C. and Steel, M.F.J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93(441): 359–371. doi:10.1080/01621459.1998.10474117.

Fosdick, B. and Raftery, A.E. (2014). Regional probabilistic fertility forecasting by modeling between-country correlations. *Demographic Research* 30(35): 1011–1034. doi:10.4054/DemRes.2014.30.35.

GENESIS-Online (Bayern) (2022). Dataset 12613-108s. Fürth: Bayerisches Landesamt für Statistik. https://www.statistikdaten.bayern.de/genesis/online.

GENESIS-Online (Bayern) (2024a). Dataset 12612-005. Fürth: Bayerisches Landesamt für Statistik. https://www.statistikdaten.bayern.de/genesis/online.

GENESIS-Online (Bayern) (2024b). Dataset 12711-104. Fürth: Bayerisches Landesamt für Statistik. https://www.statistikdaten.bayern.de/genesis/online.

GENESIS-Online (Bayern) (2025). Dataset 12411-007s. Fürth: Bayerisches Landesamt für Statistik. https://www.statistikdaten.bayern.de/genesis/online.

Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477): 359–378. doi:10.1198/016214506000001437.

Goes, J. (2024). Bayesian forecasting of mortality rates for small areas using spatiotemporal models. *Demography* 61(2): 439–462. doi:10.1215/00703370-11212716.

Goes, J., Barigou, K., and Leucht, A. (2025). Bayesian mortality modelling with pandemics: A vanishing jump approach. *Journal of the Royal Statistical Society Series C: Applied Statistics* 74(4): 1150–1182. doi:10.1093/jrsssc/qlaf018.

Green, K.C. and Armstrong, J.S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research* 68(8): 1678–1685. doi:10.1016/j.jbusres.2015.03.026.

Gueorguieva, R., Rosenheck, R., and Zelterman, D. (2008). Dirichlet component regression and its applications to psychiatric data. *Computational Statistics and Data Analysis* 52(12): 5344–5355. doi:10.1016/j.csda.2008.05.030.

Hamilton, C.H. and Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces* 41(2): 163–170. doi:10.2307/2573607.

Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* 53(1): 161–177. doi:10.2307/2533105.

Jordan, A., Krüger, F., and Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software* 90(12): 1–37. doi:10.18637/jss.v090.i12.

Krüger, F., Lerch, S., Thorarinsdottir, T., and Gneiting, T. (2021). Predictive inference based on Markov chain Monte Carlo output. *International Statistical Review* 89(2): 274–301. doi:10.1111/insr.12405.

Lee, R.D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting* 9(2): 187–202. doi:10.1016/0169-2070(93)90004-7.

Lee, R.D. (1998). Probabilistic approaches to population forecasting. *Population and Development Review* 24(Supplement: Frontiers of population forecasting): 156–190. doi:10.2307/2808055.

Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association* 87(419): 659–671. doi:10.2307/2290201.

Leslie, P.H. (1945). On the use of matrices in certain population mathematics. *Biometrika* 33(3): 183–212. doi:10.2307/2332297.

Myrskylä, M., Goldstein, J.R., and Cheng, Y.h.A. (2013). New cohort fertility forecasts for the developed world: Rises, falls, and reversals. *Population and Development Review* 39(1): 31–56. doi:10.1111/j.1728-4457.2013.00572.x.

Preston, S.H., Heuveline, P., and Guillot, M. (2000). *Demography: Measuring and modeling population processes*. Oxford: Blackwell.

Raftery, A.E. and Ševčíková, H. (2023). Probabilistic population forecasting: Short to very long-term. *International Journal of Forecasting* 39(1): 73–97. doi:10.1016/j.ijforecast.2021.09.001.

Rao, J.N.K. and Molina, I. (2015). *Small area estimation*. Hoboken: John Wiley & Sons, 2nd ed. doi:10.1002/9781118735855.

Rau, R. and Schmertmann, C.P. (2020). District-level life expectancy in Germany. *Deutsches Ärzteblatt International* 117(29–30): 493–499. doi:10.3238/arztebl.2020.0493.

Renshaw, A. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38(3): 556–570. doi:10.1016/j.insmatheco.2005.12.001.

Riebler, A., Sørbye, S.H., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research* 25(4): 1145–1165. doi:10.1177/0962280216660421.

Rogers, A. and Castro, L.J. (1981). Model migration schedules. RR-81-030 (IIASA Tech. Report). Laxenberg: International Institute for Applied Systems Analysis. https://pure.iiasa.ac.at/id/eprint/1543/1/RR-81-030.pdf.

Rogers, A. (1990). Requiem for the net migrant. *Geographical Analysis* 22(4): 283–300. doi:10.1111/j.1538-4632.1990.tb00212.x.

Schmertmann, C.P., Cavenaghi, S.M., Assunção, R.M., and Potter, J.E. (2013). Bayes plus Brass: Estimating total fertility for many small areas from sparse census data. *Population Studies* 67(3): 255–273. doi:10.1080/00324728.2013.795602.

Schmertmann, C.P. and Gonzaga, M.R. (2018). Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography* 55(4): 1363–1388. doi:10.1007/s13524-018-0695-2.

Ševčíková, H., Li, N., Kantorová, V., Gerland, P., and Raftery, A.E. (2016). Age-specific mortality and fertility rates for probabilistic population projections. In: Schoen, R. (ed.). *Dynamic demographic analysis*. Cham: Springer International Publishing: 285–310. doi:10.1007/978-3-319-26603-9_15.

Ševčíková, H., Raftery, A.E., and Gerland, P. (2018). Probabilistic projection of subnational total fertility rates. *Demographic Research* 38(60): 1843–1884. doi:10.4054/DemRes.2018.38.60.

Ševčíková, H., Raymer, J., and Raftery, A.E. (2024). Forecasting net migration by age: The flow-difference approach. doi:10.48550/arXiv.2411.09878.

Smith, S.K., Tayman, J., and Swanson, D.A. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-7551-0.

Stan Development Team (2024a). RStan: The R interface to Stan. R package version 2.32.6. https://mc-stan.org/.

Stan Development Team (2024b). Stan functions reference. Version 2.36. http://mc-stan.org/.

Stan Development Team (2024c). Stan reference manual. Version 2.36. http://mc-stan.org/.

Swanson, D.A. and Tayman, J. (2014). Measuring uncertainty in population forecasts: A new approach. In: Marsili, M. and Capacci, G. (eds.). *Proceedings of the sixth Eurostat/UNECE work session on demographic projections*. Rome: National Institute of Statistics, 203–215.

Swanson, D.A., Tayman, J., and Cline, M. (2025). A new approach to probabilistic county population forecasting with an example application to West Texas. *Population Research and Policy Review* 44(4): 43. doi:10.1007/s11113-025-09961-3.

Tayman, J. (2011). Assessing uncertainty in small area forecasts: State of the practice and implementation strategy. *Population Research and Policy Review* 30(5): 781–800. doi:10.1007/s11113-011-9210-9.

United Nations (2024). World Population Prospects 2024: Methodology of the United Nations population estimates and projections. UN DESA/POP/2024/DC/NO. 10. New York: United Nations, Departement of Economic and Social Affairs, Population Division.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software* 45(3): 1–67. doi:10.18637/jss.v045.i03.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis* 16(2): 667–718. doi:10.1214/20-BA1221.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11(12): 3571–3594. doi:10.48550/arXiv.1004.2316.

Welch, N.G., Ševčíková, H., and Raftery, A.E. (2024). Bringing age back in: Accounting for population age distribution in forecasting migration (arxiv preprint). doi:10.48550/arXiv.2403.05566.

Wilson, T. (2012). Forecast accuracy and uncertainty of Australian Bureau of Statistics state and territory population projections. *International Journal of Population Research* 2012: 419824. doi:10.1155/2012/419824.

Wilson, T., Grossman, I., Alexander, M., Rees, P., and Temple, J. (2022). Methods for small area population forecasts: State-of-the-art and research needs. *Population Research and Policy Review* 41(3): 865–898. doi:10.1007/s11113-021-09671-6.

Wiśniowski, A. and Raymer, J. (2025). Multiregional population forecasting: A unifying probabilistic approach for modelling the components of change. *European Journal of Population* 41(11). doi:10.1007/s10680-025-09729-7.

Wolff, M., Haase, A., Leibert, T., and Cunningham Sabot, E. (2022). Calm ocean or stormy sea? Tracing 30 years of demographic spatial development in Germany. *Cybergeo: European Journal of Geography* 2022(1003). doi:10.4000/cybergeo.38031.

Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis* 13(3): 917–1007. doi:10.1214/17-BA1091.

Yeung, J., Alexander, M., and Riffe, T. (2023). Bayesian implementation of Rogers–Castro model migration schedules: An alternative technique for parameter estimation. *Demographic Research* 49(42): 1201–1228. doi:10.4054/DemRes.2023.49.42.

Yu, C.C., Ševčíková, H., Raftery, A.E., and Curran, S.R. (2023). Probabilistic county-level population projections. *Demography* 60(3): 915–937. doi:10.1215/00703370-10772782.

Appendix

Comparison in-sample fit of migration schedule

We compare the fit of the traditional Rogers–Castro curve using the implementation of Yeung, Alexander, and Riffe (2023) with the Dirichlet model as described in Section 5. For the Dirichlet regression having multiple years of migration schedules as input variables helps in reducing the variance of the estimates. The Rogers–Castro model of Yeung, Alexander, and Riffe (2023), on the other hand, uses a single year as the data input but is proven to work well. To be able to compare both methods, we estimated the migration schedule for each region using the mean values of the age-distribution, $\bar{R}_{x,r}$, as input for the model of Yeung, Alexander, and Riffe (2023). Our Dirichlet regression model was trained with all available data – that is, multiple years – and therefore has the advantage of having seen more data. Nevertheless, we generated replicate data using the posterior predictive distribution of both models and compared the in-sample fit of each region for both models in terms of MSE and MAE using the age-schedule of all years as the observed values. While the methods are difficult to compare since they are using different input values, we check if the replicated in-sample data of the Dirichlet regression are on a similar level to that of the Rogers–Castro model. Evaluating the results in Table A-1, we notice two things: First, the performance of both models is very satisfactory, and second, the Dirichlet regression performs on par with the Bayesian implementation of the Rogers–Castro model. Although, the Dirichlet regression is estimated with more data, making the comparison difficult.

Table A-1: Comparison of the in-sample performance of the migration schedule for men

Method	**MAE** $\cdot 10^{-2}$	**RMSE** $\cdot 10^{-2}$	**MAPE**
Out-migration			
Rogers Castro	0.64	0.86	28.96
Dirichlet	0.49	0.71	15.52
In-migration			
Rogers Castro	0.66	0.91	29.38
Dirichlet	0.54	0.82	16.42

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error.

For women, we observe similar results. These are given in Table A-2. We conclude that the Dirichlet model is adequate in describing the age-migration patterns.

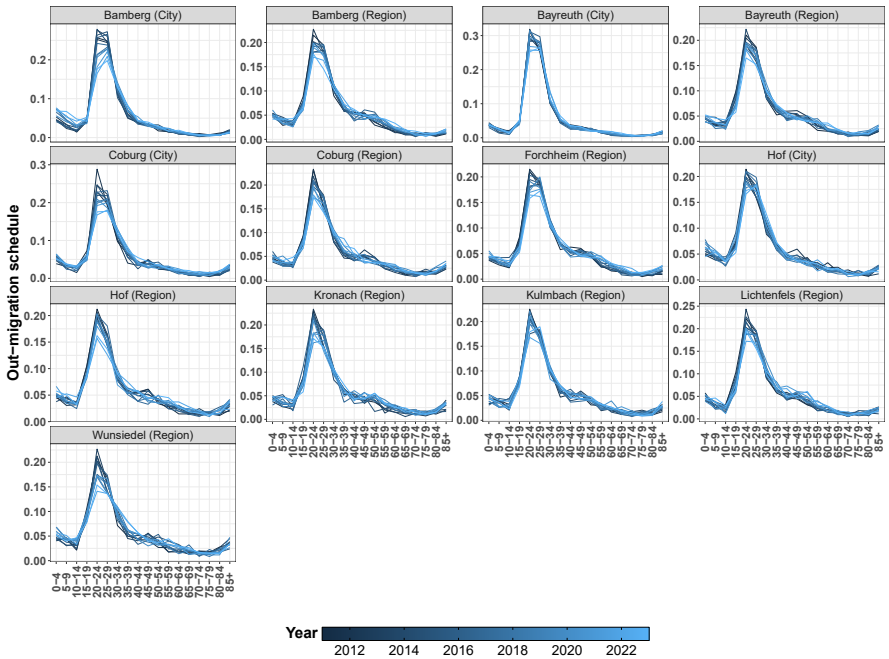
Table A-2: Comparison of the in-sample performance of the migration schedule for women

Method	**MAE** $\cdot 10^{-2}$	**RMSE** $\cdot 10^{-2}$	**MAPE**
Out-migration			
Rogers Castro	0.79	1.02	27.78
Dirichlet	0.56	0.83	14.83
In-migration			
Rogers Castro	0.76	1.04	26.36
Dirichlet	0.59	0.92	15.48

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error.

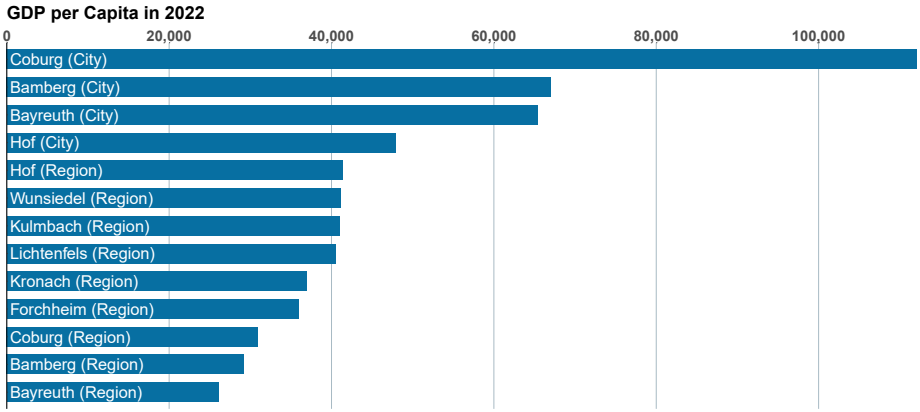
Graphics of migration schedules

Figure A-1: Out-migration schedule over time of women for all regions in Upper Franconia



Additional figures

Figure A-2: GDP per capita for all regions in Upper Franconia in 2022



Additional tables

Table A-3: Comparison of the out-of-sample performance of the mortality models for males disaggregated by age

AgeGroup	Model	MAE	RMSE	Cov	LogS	RPS
0–4	RH_BYM2	**1.02**	**1.30**	**0.97**	**60.33**	**27.10**
0–4	Lee–Carter	1.03	1.31	1.00	60.51	27.23
5–9	RH_BYM2	**0.24**	**0.39**	**1.00**	**19.75**	**5.94**
5–9	Lee–Carter	0.25	0.40	0.97	21.65	6.16
10–14	RH_BYM2	**0.26**	**0.33**	**1.00**	**14.54**	**3.99**
10–14	Lee–Carter	0.32	0.39	1.00	17.31	4.96
15–19	RH_BYM2	**0.61**	**0.71**	**1.00**	**37.71**	**13.66**
15–19	Lee–Carter	0.74	0.87	1.00	41.83	16.17
20–24	RH_BYM2	**1.08**	**1.42**	**0.92**	**65.31**	**30.84**
20–24	Lee–Carter	1.20	1.46	1.00	67.64	30.90
25–29	Lee–Carter	**1.04**	**1.31**	1.00	**61.64**	**27.65**
25–29	RH_BYM2	1.06	1.38	**0.97**	63.78	29.18
30–34	Lee–Carter	**1.33**	**1.62**	0.97	**69.24**	**34.68**
30–34	RH_BYM2	1.38	1.64	**0.95**	71.62	36.53
35–39	Lee–Carter	**1.12**	**1.37**	1.00	**70.27**	**31.68**
35–39	RH_BYM2	1.21	1.51	**0.95**	76.38	35.16
40-44	RH_BYM2	**1.54**	**2.16**	**0.95**	**80.31**	**43.58**
40–44	Lee–Carter	1.72	2.25	0.92	83.14	46.57
45–49	Lee–Carter	**2.71**	**3.57**	0.97	**101.15**	**73.44**
45–49	RH_BYM2	2.75	3.60	**0.95**	102.75	76.48
50–54	RH_BYM2	**3.04**	**3.74**	**1.00**	**106.10**	**82.31**
50–54	Lee–Carter	3.44	4.27	1.00	112.09	95.07
55–59	RH_BYM2	**4.25**	**5.45**	0.95	**119.19**	**117.37**
55–59	Lee–Carter	4.52	5.64	**1.00**	119.40	118.43
60–64	RH_BYM2	**4.46**	**5.65**	0.92	**124.38**	**130.64**
60–64	Lee–Carter	4.59	5.79	**0.97**	127.42	136.13
65–69	RH_BYM2	**5.09**	**6.35**	**1.00**	**127.70**	**141.77**
65–69	Lee–Carter	7.09	8.54	1.00	140.73	189.48
70–74	RH_BYM2	**5.37**	**6.86**	1.00	**129.06**	**152.67**
70–74	Lee–Carter	8.92	11.32	**0.97**	160.71	227.88
75–79	RH_BYM2	**7.31**	**8.96**	**1.00**	**139.35**	**201.56**
75–79	Lee–Carter	11.75	15.29	1.00	172.27	318.95
80–84	RH_BYM2	**8.30**	**10.58**	**0.95**	**146.91**	**231.47**
80–84	Lee–Carter	9.11	11.51	1.00	150.10	242.82
85+	Lee–Carter	**12.74**	**16.60**	**0.97**	**170.13**	**363.88**
85+	RH_BYM2	13.62	18.00	0.90	175.10	383.49

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95. LogS = log score. RPS = ranked probability score.

Table A-4: Comparison of the out-of-sample performance of the mortality models for females disaggregated by age

AgeGroup	Model	MAE	RMSE	Cov	LogS	CRPS
0–4	RH_BYM2	**0.85**	**1.09**	**0.97**	**53.49**	**22.44**
0–4	Lee–Carter	0.93	1.13	1.00	55.22	23.78
5–9	RH_BYM2	**0.19**	**0.32**	1.00	**14.25**	**4.16**
5–9	Lee–Carter	0.20	0.34	**0.97**	16.64	4.51
10–14	RH_BYM2	**0.20**	**0.33**	1.00	**14.74**	**4.25**
10–14	Lee–Carter	0.22	0.33	1.00	14.94	4.33
15–19	RH_BYM2	**0.49**	**0.62**	1.00	**30.28**	**10.59**
15–19	Lee–Carter	0.50	0.64	**0.97**	34.35	11.02
20–24	Lee–Carter	**0.46**	**0.49**	1.00	**28.35**	**9.06**
20–24	RH_BYM2	0.48	0.51	1.00	29.33	9.69
25–29	RH_BYM2	**0.70**	**0.87**	**0.97**	**46.54**	**18.09**
25–29	Lee–Carter	0.74	0.92	0.97	49.70	19.58
30–34	RH_BYM2	**0.87**	**1.09**	**0.97**	**50.35**	**21.87**
30–34	Lee–Carter	0.92	1.15	0.97	54.44	23.40
35–39	RH_BYM2	**0.85**	**1.17**	0.97	**55.51**	**23.26**
35–39	Lee–Carter	0.93	1.25	0.97	58.54	25.28
40–44	RH_BYM2	**0.98**	**1.25**	0.97	**62.75**	**26.15**
40–44	Lee–Carter	1.03	1.37	0.97	66.33	28.52
45–49	RH_BYM2	**1.75**	**2.12**	1.00	**82.79**	**46.10**
45–49	Lee–Carter	1.77	2.24	1.00	84.51	47.70
50–54	RH_BYM2	**2.42**	**3.05**	1.00	**98.81**	**68.37**
50–54	Lee–Carter	2.60	3.37	**0.97**	102.19	72.12
55–59	RH_BYM2	**2.91**	**3.70**	**0.92**	**108.23**	**81.98**
55–59	Lee–Carter	3.36	4.25	0.95	114.29	93.84
60–64	RH_BYM2	**3.39**	**4.40**	**0.95**	**111.82**	**94.45**
60–64	Lee–Carter	3.93	5.17	0.97	119.45	111.31
65–69	RH_BYM2	**4.17**	**5.27**	**0.95**	**121.87**	**116.92**
65–69	Lee–Carter	4.59	5.75	0.97	125.91	127.30
70–74	RH_BYM2	**6.37**	**8.69**	0.87	**148.12**	**188.55**
70–74	Lee–Carter	6.79	9.25	**0.90**	153.14	192.52
75–79	RH_BYM2	**6.72**	**8.27**	1.00	**136.60**	**186.23**
75–79	Lee–Carter	11.20	14.68	1.00	179.10	312.43
80–84	RH_BYM2	**9.09**	**11.96**	0.90	**152.66**	**256.78**
80–84	Lee–Carter	11.37	14.63	**0.97**	162.98	315.10
85+	Lee–Carter	**20.94**	**26.51**	**0.97**	**201.22**	**614.16**
85+	RH_BYM2	22.47	27.91	1.00	203.52	629.86

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95. LogS = log score. RPS = ranked probability score.

Table A-5: Comparison of the out-of-sample performance of the fertility models disaggregated by age

AgeGroup	Model	MAE	RMSE	MAPE	Cov	LogS$\cdot 10^{-2}$	RPS$\cdot 10^{-2}$
15–19	Direct Approach	3.88	4.83	39.16	0.90	3.12	2.87
15–19	Indirect Approach	**3.70**	**4.63**	37.00	0.90	**3.06**	**2.77**
15–19	Lee–Carter	3.84	4.79	**36.62**	**0.93**	3.06	2.80
20–24	Direct Approach	9.74	12.62	14.57	**0.94**	4.16	7.43
20–24	Indirect Approach	**9.11**	**11.80**	**13.73**	0.96	**4.10**	**7.02**
20-24	Lee–Carter	11.85	14.75	17.38	0.97	4.32	8.91
25–29	Direct Approach	24.64	30.13	12.12	0.79	5.22	18.78
25–29	Indirect Approach	**23.38**	**29.50**	**11.28**	**0.82**	**5.09**	**17.77**
25–29	Lee–Carter	30.44	38.10	14.58	0.69	5.50	23.55
30–34	Direct Approach	29.46	37.35	11.07	0.85	5.37	22.13
30–34	Indirect Approach	31.43	39.62	11.60	0.85	5.46	23.77
30–34	Lee–Carter	**28.11**	**35.52**	**10.81**	**0.93**	**5.20**	**20.66**
35–39	Direct Approach	**15.67**	**21.29**	**11.57**	**0.93**	**4.59**	**12.15**
35–39	Indirect Approach	16.42	22.40	11.87	0.92	4.60	12.62
35–39	Lee–Carter	25.65	35.43	24.33	0.98	4.79	15.65
40–44	Direct Approach	5.20	6.44	20.29	**0.97**	3.43	3.88
40–44	Indirect Approach	**5.04**	**6.39**	**19.21**	0.96	**3.42**	**3.84**
40–44	Lee–Carter	6.56	9.93	34.18	1.00	3.56	4.23

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95. LogS = log score. RPS = ranked probability score.

Table A-6: Comparison of the out-of-sample performance of age-specific population for Upper Franconia of our model disaggregated by age and sex

Sex	Age group	MAE	RMSE	MAPE	rel-HW	Coverage
female	0–4	103.18	150.75	6.55	12.77	0.85
female	5–9	34.75	46.27	2.12	5.51	1.00
female	10–14	46.03	52.54	2.87	4.97	0.77
female	15–19	92.01	126.44	5.41	5.89	0.46
female	20–24	146.05	197.86	7.41	10.07	0.77
female	25–29	122.17	222.97	5.09	11.16	0.92
female	30–34	105.41	149.56	4.31	8.73	0.92
female	35–39	134.36	172.67	5.35	5.95	0.54
female	40–44	63.46	71.49	2.91	4.62	0.92
female	45–49	35.56	41.33	1.74	4.20	1.00
female	50–54	33.73	41.48	1.34	3.33	0.92
female	55–59	31.99	40.79	0.97	2.35	0.92
female	60–64	36.91	42.46	1.25	1.91	0.85
female	65–69	34.62	49.40	1.16	1.83	0.77
female	70–74	23.46	29.48	0.87	2.06	1.00
female	75–79	33.59	41.11	1.92	4.17	1.00
female	80–84	30.89	37.23	1.78	6.49	1.00
female	85+	52.04	75.14	2.80	15.47	1.00
male	0–4	101.61	161.53	6.32	11.86	0.85
male	5–9	58.77	79.92	3.48	5.44	0.92
male	10–14	38.76	43.50	2.24	4.73	0.92
male	15–19	96.09	120.80	5.59	5.85	0.54
male	20–24	119.63	176.34	5.42	10.92	0.92
male	25–29	203.95	275.83	7.75	13.45	0.77
male	30–34	133.80	182.50	4.88	11.58	0.92
male	35–39	99.08	173.50	3.46	8.36	0.92
male	40–44	58.06	68.32	2.42	6.60	1.00
male	45–49	54.33	68.31	2.41	5.88	0.92
male	50–54	45.36	57.89	1.64	4.64	0.92
male	55–59	56.06	68.09	1.74	3.08	0.77
male	60–64	52.43	62.45	1.55	2.73	0.92
male	65–69	40.33	54.16	1.52	3.36	0.85
male	70–74	32.06	38.07	1.52	5.05	1.00
male	75–79	44.53	50.42	3.31	7.05	1.00
male	80–84	51.83	64.37	4.07	8.37	0.92
male	85+	95.32	115.86	8.87	13.08	0.85

Notes: MAE = mean absolute error. RMSE = root mean squared error. MAPE = mean absolute percentage error. Coverage denotes the coverage for a nominal value of 0.95. rel-HW denotes the relative half width and is defined as rel-HW = $100 \cdot \frac{(y_u - y_l)/2}{E(y)}$, with y_u and y_l denoting the upper respectively lower prediction interval and $E(y)$ the mean forecast.

Cumulative distribution and quantile function of skewed distribution

Consider a univariate random variable with probability density function $f(\cdot)$ – that is continuous, unimodal, and symmetric around 0: for example, a centered normal or student-t distribution. A parameter $\gamma \in (0, \infty)$ accounts for the degree of skewness that generates the following class of skewed distributions:

$$p(x|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \left[f\left(\frac{x}{\gamma}\right) 1_{[0,\infty)}(x) + f\left(x\gamma\right) 1_{(-\infty,0)}(x) \right].$$

Here, $1_{[0,\infty)}(\varepsilon)$ denotes an indicator function, which equals one if ε lies in the interval $[0, \infty)$ and zero otherwise.

The cumulative distribution function (CDF), which we will denote by $P(x|\gamma)$, can be calculated using integration by substitution. Recall that integration by substitution is given by

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(t)dt.$$

Then to obtain the CDF, we have

$$P(x|\gamma) = \frac{2}{\gamma + 1/\gamma} \cdot \left[\int_{-\infty}^{\min(0,x)} f(u\gamma)du + \int_0^x f\left(\frac{u}{\gamma}\right) 1_{[0,\infty)}(x)du \right].$$

We start by looking at $x < 0$, which results in

$$\begin{aligned} P(x|\gamma) &= \frac{2}{\gamma + 1/\gamma} \cdot \int_{-\infty}^{x} f(u\gamma)du \\ &= \frac{2}{\gamma + 1/\gamma} \cdot \int_{-\infty}^{x\gamma} \frac{f(t)}{\gamma} dt \\ &= \frac{2}{\gamma^2 + 1} \cdot F(x\gamma), \end{aligned}$$

where $F(x)$ denotes the CDF of $f(x)$. Next, for $x \geq 0$ we have

$$\begin{aligned} P(x|\gamma) &= P(0|\gamma) + \frac{2}{\gamma + 1/\gamma} \cdot \int_0^x f\left(\frac{u}{\gamma}\right) du \\ &= P(0|\gamma) + \frac{2}{\gamma + 1/\gamma} \int_0^{\frac{x}{\gamma}} f(t) \cdot \gamma dt \\ &= P(0|\gamma) + \frac{2}{\gamma + 1/\gamma} \cdot \left[\gamma \cdot F\left(\frac{x}{\gamma}\right) - \gamma \cdot F(0)\right]. \end{aligned}$$

Thus,

$$P(x|\gamma) = \begin{cases} \frac{2}{\gamma^2+1} \cdot F(x\gamma) & \text{for } x < 0 \\ \underbrace{\frac{2}{\gamma^2+1} \cdot F(0)}_{P(0|\gamma)} + \frac{2\gamma}{\gamma+1/\gamma} \cdot \left[F\left(\frac{x}{\gamma}\right) - 1/2\right] & \text{for } x \geq 0. \end{cases}$$

The quantile function $G(z|\gamma) = P^{-1}(z|\gamma)$ can be obtained by solving $z = P(x|\gamma)$ for x, with $P(\cdot|\gamma)$ strictly increasing.

Starting with $x < 0$, we get

$$\begin{aligned} & z = \frac{2}{\gamma^2 + 1} \cdot F(x\gamma) \\ \Leftrightarrow \quad & \frac{z \cdot \left(\gamma^2 + 1\right)}{2} = F(x\gamma) \\ \Leftrightarrow \quad & x = \frac{1}{\gamma} \cdot F^{-1}\left(\frac{z \cdot (\gamma^2 + 1)}{2}\right). \end{aligned}$$

Here, $F^{-1}(z)$ denotes the quantile function of $f(.)$. Now we look at $x \geq 0$:

$$
\begin{aligned}
& z = P(0|\gamma) + \frac{2\gamma^2}{\gamma^2+1} \cdot [F(x/\gamma) - 1/2] \\
\Leftrightarrow \quad & \frac{\gamma^2 + 1 \cdot [z - P(0|\gamma)]}{2\gamma^2} = F(x/\gamma) - 1/2 \\
\Leftrightarrow \quad & \frac{\gamma^2 \cdot \left[\left(1 + 1/\gamma^2\right) \cdot (z - P(0|\gamma))\right]}{\gamma^2 \cdot 2} = F(x/\gamma) - 1/2 \\
\Leftrightarrow \quad & \frac{\left(1 + 1/\gamma^2\right) \cdot [z - P(0|\gamma)]}{2} + 1/2 = F(x/\gamma) \\
\Leftrightarrow \quad & x = F^{-1}\left[\frac{\left(1 + \frac{1}{\gamma^2}\right) \cdot [z - P(0|\gamma)]}{2} + 1/2\right] \cdot \gamma \\
\Leftrightarrow \quad & x = F^{-1}\left[\frac{\left(1 + \frac{1}{\gamma^2}\right) \cdot \left[z - \frac{2}{2\gamma^2+2}\right]}{2} + 1/2\right] \cdot \gamma.
\end{aligned}
$$

Thus, the quantile function $G(z|\gamma)$ is given by

$$
G(z|\gamma) = \begin{cases} \frac{1}{\gamma} \cdot F^{-1}\left(\frac{z \cdot (\gamma^2+1)}{2}\right) & \text{for } z < P(0|\gamma) \\ F^{-1}\left[\frac{\left(1+\frac{1}{\gamma^2}\right) \cdot \left[z - \frac{2}{2\gamma^2+2}\right]}{2} + 1/2\right] & \text{for } z \geq P(0|\gamma). \end{cases}
$$

Overview on choice of priors

For the mortality model, we use the following choice of priors:

$$
\begin{aligned}
\alpha_x|\alpha_{x-1},\alpha_{x-2};\sigma_\alpha^2 &\sim \mathcal{N}(2\alpha_{x-1}-\alpha_{x-2},\sigma_\alpha^2) \quad \text{and } \alpha_j \sim \mathcal{N}(0,\sigma_\alpha^2) \text{ for } j=1,2\\
\sigma_\alpha &\sim t_5^+(0,1)\\
\left(\beta_1^{(1)},\ldots,\beta_X^{(1)}\right) &\sim \text{Dirichlet}(1,\ldots,1)\\
\left(\beta_1^{(2)},\ldots,\beta_X^{(2)}\right) &\sim \text{Dirichlet}(1,\ldots,1)\\
\kappa_t|\kappa_{t-1},d;\sigma_\kappa^2 &\sim \mathcal{N}(\kappa_{t-1}+d,\sigma_\kappa^2), \quad \text{and } k_1 \sim \mathcal{N}(d,\sigma_k^2)\\
d &\sim \mathcal{N}(0,2^2)\\
\sigma_\kappa &\sim t_5^+(0,1)\\
\gamma_k &\overset{iid}{\sim} \mathcal{N}(0,\sigma_\gamma)\\
\sigma_\gamma &\sim t_5^+(0,1)\\
\nu_r &= \sigma_\nu\left(\sqrt{1-\rho}v_r^* + \sqrt{\rho}u_r^*\right)\\
u_r|u_{-r};\sigma_u^2 &\sim \mathcal{N}\left(\frac{\sum_{r\neq j} w_{rj}u_j}{\sum_{r\neq j} w_{rj}}, \frac{\sigma_u^2}{\sum_{r\neq j} w_{rj}}\right)\\
\sigma_\nu &\sim t_5^+(0,1)\\
\rho &\sim \text{Beta}(0.5,0.5),
\end{aligned}
$$

where ν_r follows a BYM2 model. Here, σ_ν denotes its standard deviation, and $\rho \in [0,1]$ defines a mixing parameter that shows how much of the effect is spatially unstructured, given by $v_r^* \sim \mathcal{N}(0,1)$, and how much is due to a scaled spatially structured effect u_r^*, which follows a scaled conditional autoregressive model, with symmetric adjacency matrix $\mathbf{W} \in \mathbb{R}^{R\times R}$ with entries w_{rj} denoting one if two regions are neighbors and zero otherwise (see Riebler et al. (2016) for details). In addition $t_5^+(0,1)$ denotes a location scale t-distribution truncated to the interval $[0,\infty)$ with five degrees of freedom, location of zero and scale of one. These are the same priors as used in Goes (2024).

For the direct fertility model, we use the following priors:

$$
\begin{aligned}
\alpha_x &\overset{iid}{\sim} \mathcal{N}(0, \sigma_\alpha^2) \\
\sigma_\alpha &\sim t_5^+(0,1) \\
\beta_x &\overset{iid}{\sim} \mathcal{N}(0, 5^2) \\
\kappa_t|\kappa_{t-1}, d; \sigma_\kappa^2 &\sim \mathcal{N}(\kappa_{t-1} + d, \sigma_\kappa^2) \quad \text{and } k_1 \sim \mathcal{N}(d, \sigma_k^2) \\
d &\sim \mathcal{N}(0, 2^2) \\
\sigma_\kappa &\sim t_5^+(0,1) \\
\delta_{x,r} &\overset{iid}{\sim} \mathcal{N}(0, 10) \\
\sigma_\varepsilon &\sim t_5^+(0,1).
\end{aligned}
$$

To model net-migration counts, the following choice of priors were used:

$$
\begin{aligned}
\mu_r &\overset{iid}{\sim} \mathcal{N}(m_1, s_1) \\
m_1 &\sim \mathcal{N}(0, 200) \\
s_1 &\sim \mathcal{N}^+(0, 100) \\
\varepsilon_{t,r} &\overset{iid}{\sim} \text{Skew-Normal}(\gamma) \\
\gamma &\sim \mathcal{N}(1, 1) \\
\sigma_r &\overset{iid}{\sim} \mathcal{N}^+(m_2, s_2) \\
m_2 &\sim \mathcal{N}(0, 50) \\
s_2 &\sim \mathcal{N}^+(0, 20) \\
a_r &\overset{iid}{\sim} \mathcal{N}(0, 2).
\end{aligned}
$$