Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

FDZ Data description: RWI-GEO-POP-FORECAST: Small-Scale Population Projection V2

November 2023

Philipp Breidenbach, Patrick Thiel, Thorben Wiebe

Impressum

Herausgeber:

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstraße 1–3 | 45128 Essen, Germany

Postanschrift:
Postfach 10 30 54 | 45030 Essen, Germany

Fon: +49 201-81 49-0 | E-Mail: rwi@rwi-essen.de
www.rwi-essen.de

RWI Datenbeschreibung

Schriftleitung: Prof. Dr. Dr. h. c. Christoph M. Schmidt
Gestaltung: Daniela Schwindt, Magdalena Franke, Claudia Lohkamp

FDZ Data description:
RWI-GEO-POP-FORECAST: Small-Scale Population Projection V2
November 2023
Philipp Breidenbach, Patrick Thiel, Thorben Wiebe

Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

FDZ Data description: RWI-GEO-POP-FORECAST: Small-Scale Population Projection V2

November 2023

Philipp Breidenbach, Patrick Thiel, Thorben Wiebe

Contents

List of Figures

List of Tables

0. Abstract

Existing nationwide population projections focus on highly aggregated spatial units. The most disaggregated nationwide study is provided by the German Federal Institute for Building, Urban Affairs and Spatial Research (BBSR) and forecasts population at the county level until 2040 (BBSR, 2021). The Federal Statistical Office's projections - the benchmark for German population projections - are limited to the entire country or to the federal states. Our novel data set offers a population projection with a resolution of one square kilometer up to 2060. Since many local socio-economic characteristics show great spatial heterogeneity, as documented by RWI-GEO-GRID (RWI - Leibniz Institute for Economic Research, 2022), the need for a small-scale population forecast becomes apparent. We closely follow the assumptions on fertility, mortality and (out) migration of the German Federal Statistical Office, so that our forecast reflects those estimates at the national level, but provides detailed information about local communities, including gender and age structure. The generated data set is made available to researchers.

1. Introduction

Germany's demographic structure is changing due to low fertility rates and increased life expectancy. Based on natural development, Germany faces a shrinking and aging population. Migration flows are expected to mitigate the decline. These developments are captured by the Federal Statistical Office (FSO) in its nationwide projection for the years up to 2060 FSO (2019). Depending on the underlying assumptions, the population will shrink to about 74 million people in 2060. These expected developments have significant implications for several key areas of economic policy. Public finances, as well as the labor market and the demand for public services and infrastructure, are strongly affected by the decline and aging of the population. By definition, these nationwide projections do not take into account the diversity of demographic change at lower regional levels, as these projections are only available at administrative levels such as states or counties. The demographic structure at lower levels has shown itself to be quite heterogeneous (RWI - Leibniz Institute for Economic Research, 2022), which means that larger aggregations are too broad to adequately capture regional differences. In particular, for services such as child care, elderly care, medical care, or local transportation, areas below the county level provide a better delineation to describe local needs. Yet, there is a lack of such data that can provide a consistent, nationwide demographic projection based on reliable data sources.

To meet the need for such data, we develop a demographic projection based on grid data. The grids have a size of one square kilometer and cover Germany equally. Synthetic grid data has several advantages over conventional administrative data: (1) Grid data can be easily transformed into administrative regions. Aggregation to any other spatial definition is also easy. (2) Spatial grids are time-independent, unlike, for example, German counties or postal codes, whose boundaries may change over time. (3) Grid data can be efficiently stored, analyzed, and linked with other information. (4) Grids provide high spatial resolution to reveal heterogeneities in less densely populated areas.

This report describes the underlying methodology for our grid-based population projection, which covers Germany from base year 2020 to the final year 2060. We follow the main assumptions of FSO (2019), which include assumptions on fertility, mortality, and migration. The projection is based on grid-level population data provided by RWI-GEO-GRID (RWI - Leibniz Institute for Economic Research, 2022).

The paper is organized as follows. Section 2 provides an overview of the methodology used and the underlying assumptions. It also discusses the input data and its sources, as well as the steps taken to prepare the data. Section 3 describes the final output of the generated population projection and highlights some interesting findings. Section 4 validates our projection by comparing it with other data sources, such as information from the FSO and the World Bank, which also provide national projections. Section 6 provides information on data availability and how to access the population forecast.

2. Strategy

The FSO uses a multitude of different assumptions regarding the components fertility, mortality, and migration while forecasting Germany's population, folded into so-called variants FSO (2019). Each of the three components gets, broadly speaking, modeled for three different developments: (1) low or declining growth rate, (2) moderate or stable growth rate, (3) high or rising growth rate. Each variant includes further assumptions about specific components to achieve these trends. Mortality, for example, might be assumed to be declining (Assumption $L1$), which leads to a new level of life expectancy at birth of 82.5 and 86.4 for males and females, respectively.

In order to obtain results that are consistent with the median demographic assumptions, we choose the median variant as our benchmark, variant G2-L2-W2, and thus assume that all components develop according to the second trend assumption (2). This choice does not explicitly determine how we forecast the components, but only provides benchmarks that should be met to achieve a similar forecast at the national level.

Based on the standard population projection literature (for an overview see Preston et al., 2000), there are two ways to gain population and two ways to lose population: The population can increase due to births and immigration, but it can decrease due to deaths and emigration. As shown in the balancing equation Equation (1) past years components — Births B_{t-1}, Deaths D_{t-1} and Net Migration M_{t-1} — and the past year's population N_{t-1} determine the amount of people living in the country in year t.[1] This expression can be disaggregated from national level using equation Equation (2), which divides the population into age-gender-cohorts within grid g and year t. The population within these cohorts is calculated in a similar way to the national level, but using the cohort-specific components instead. In this so-called Cohort Component Method (Cannan, 1895; Whelpton, 1928, 1936), the population of gender s at any age $a > 0$ in year t is calculated based on the population at age $a-1$ in the previous year $t-1$ and the corresponding mortality and net migration (see Equation (3)). Newborns ($a = 0$) are a special case, since they do not enter their age cohort by aging, but rather by being born. We model this with equation Equation (4), see Section 2.2 for more details. In addition, we split net migration into two parts that are modeled separately, Inner (IN) and Outer (OM) Migration (see Equation (5) and Section 2.4).

$$N_t = N_{t-1} + B_{t-1} - D_{t-1} + M_{t-1} \tag{1}$$

Define

$$N_t = \sum_{a,s,g}^{A,S,G} N_{a,s,g,t}, \tag{2}$$

where

$$N_{a,s,g,t} = N_{a-1,s,g,t-1} - D_{a-1,s,g,t-1} + M_{a,s,g,t-1}, \tag{3}$$

with

$$B_{s,g,t} = N_{0,s,g,t} = \delta_s * \sum_{a=16}^{49} N_{a,female,g,t} * P^a_{g,t} \tag{4}$$

and

$$M_{a,s,g,t} = OM_{a,s,g,t} + IM_{a,s,g,t} \tag{5}$$

The following sections describe the input data we use for each component. It also describes the assumptions involved and the steps taken with respect to the different data sources.

[1] For notational clarity no symbolic distinction between measured and predicted entities is being made.

2.1. Grid Population (N)

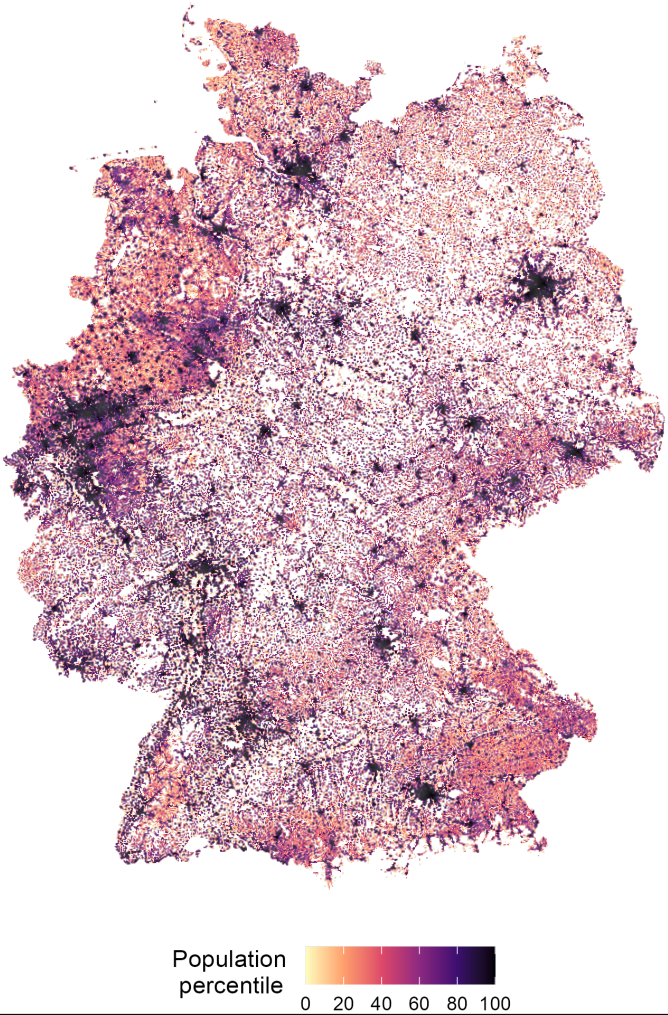
Our starting point is the population on a 1km^2 grid level offered by RWI-GEO-GRID (RWI - Leibniz Institute for Economic Research, 2022). The data set contains population disaggregated information by gender and age groups.[2] A detailed description can be found in Breidenbach and Eilers (2018). While there are 384,185 grids in Germany, as defined by the ETRS89 Lambert Azimuthal Equal-Area projection coordinate reference system (Annoni, 2003), only 223,140 are populated and 161,045 are unpopulated. This results in an unpopulated share of 41.92%.

We plot the total population on the grid level in 2019 as population percentiles in Figure 1 with the highest populated grid at upper end of the distribution.[3] The white spots represent unpopulated grids. The color scheme is dominated by densely populated grids in urban areas.

[2]These are: [0,2], [3,5], [6,9], [10,14], [15,17], [18,19], [20,24], [25,29], [30,34], [35,39], [40,44], [45,49], [50,54], [55,59], [60,64], [65,74], [75, ∞).

[3]We use population percentiles to generate the map instead of absolute population counts because otherwise, the color scheme would be dominated by densely populated grids in metropolitan areas, making differences less visible.

Figure 1
German population on the grid level in 2019
Population counts represented as percentiles



Notes: The figure shows the population counts as percentiles across Germany for 2019 at the 1km^2 grid level. Higher percentiles reflect higher population densities and vice versa.
Source: Authors' graph. The original data is given by RWI-GEO-GRID (RWI - Leibniz Institute for Economic Research, 2022).

Theoretically, the upper limit of the age groups is undefined. For simplicity, we assume that the age distribution is truncated at age 100. The population in grid g, given as $N^i_{g,s,2019}$, is defined as the population of age interval i in grid g, separated by gender s. We then turn the 17 age intervals into a population cohort for every age-year FSO (2019). This assumes that the distribution of age-years of the 17 age intervals is equal within Germany. To transform this information into the required population in age-years $N^a_{g,s,2019}$, we use the following equations:

Propose that

$$N^{i}_{g,s,2019} = \sum_{a}^{i} N^{a}_{g,s,2019} \tag{6}$$

and define

$$w^{a}_{g,s,2019} = \frac{N^{a}_{g,s,2019}}{N^{i}_{g,s,2019}} \tag{7}$$

then

$$N^{a}_{g,s,2019} = w^{a}_{g,s,2019} \times N^{i}_{g,s,2019} \tag{8}$$

(9)

2.2. Births and Fertility (B)

The number of births in a grid g in year t is strongly related to the number of women and their age distribution in the respective grid. Calculating regionalized newborns, we strongly follow the assumptions on the mean age of giving birth and the Total Fertility Rate (TFR) applied in the G2-L2-W2-scenario by the Statistical Offices FSO (2019). By definition, the nationwide projection does not give an intuition for the development of regional heterogeneities between the regional TFR. We assume that these differences remain constant but develop along the total development of TFR and the mean age of giving birth.

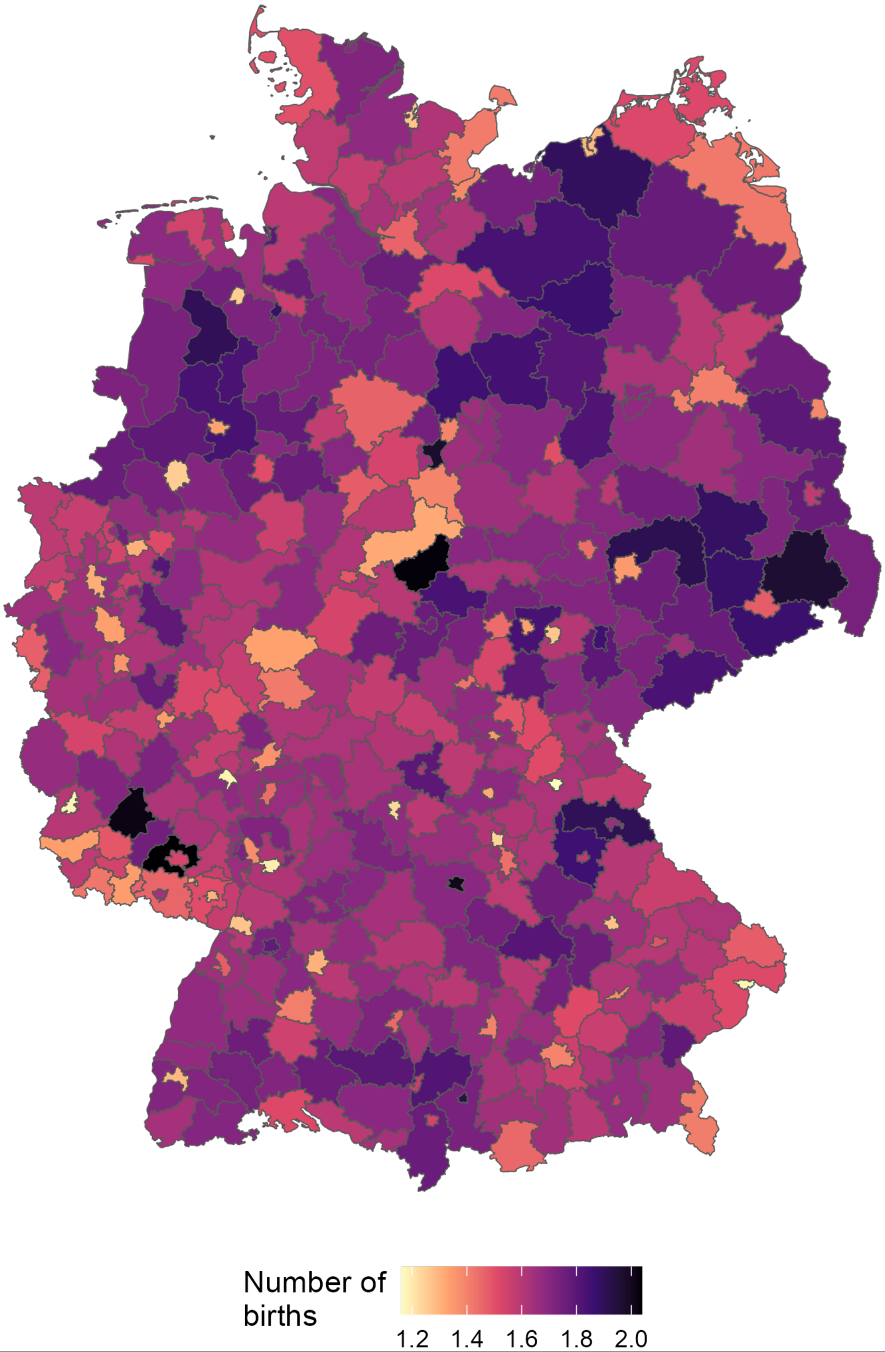
The number of birth is calculated as follows

$$B_{s,g,t} = N_{0,s,g,t} = \delta_s * \sum_{a=16}^{49} N_{a,female,g,t} * P^{a}_{g,t}, \tag{10}$$

where the projected number of newborns $B_{s,g,t}$ is based on the age-specific fertility rates $P^{a}_{g,t}$ of the female population aged between 16 and 49 in grid g and year t (following Federal Statistical Office (FSO), 2022a). δ_s represents the fraction of births of gender s in the base year.[4] These fractions are assumed to be constant for the forecast horizon.

[4] Based on Federal Statistical Office (FSO) (2022b), 51.32% of newborns are born male while the remaining 100% - 51.32% = 48.68% are born female.

Figure 2
Total Fertility Rate at the county level
Average number of children per woman at the country level in 2019



Notes: The figure shows the Total Fertility Rate (TFR) which is defined as the average number of children born per woman at the county level in 2019.
Source: Authors' graph. The data is given by Federal and State Statistical Offices (FSOS) (2022).

Figure 2 illustrates the TFR, representing the average number of children born per woman during her lifetime, for Germany at county level in 2019. TFR varies between 1.16 and 2.03 children born per woman, with an indication of spatial variation. The former East German States show, on average, higher rates of fertility than Western states.

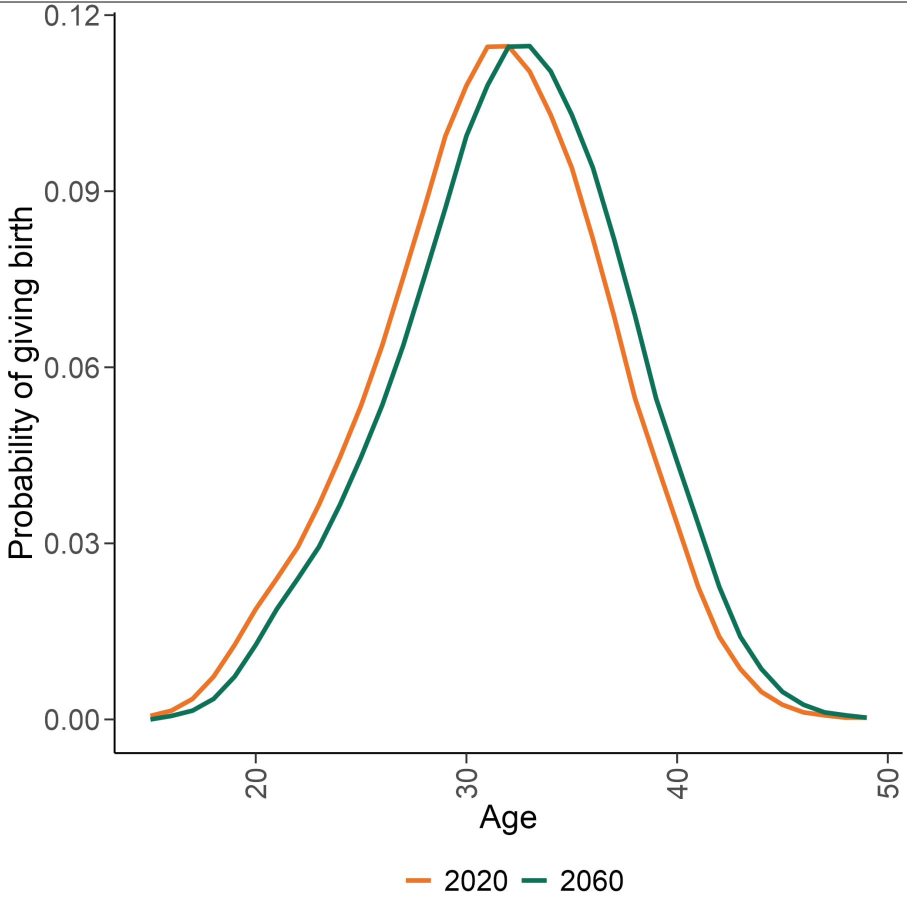
To account for these underlying trends, we combine national with district-level data. We use age-specific fertility rates to capture the overarching national trends (Federal Statistical Office (FSO), 2022a) and combine them with county-specific fertility rates to account for regional heterogeneity at the lowest level available (Federal and State Statistical Offices (FSOS), 2022). Ages on the former are calculated as the difference between the child's and the mother's year of birth ("year-of-birth method"). In the latter case, data is only available for absolute births in the county by age intervals. To dissolve the intervals into individual ages, the age-specific contributions within the interval are calculated, based on the country-level fertility rates (Federal Statistical Office (FSO), 2022a). These contributions are then used to reverse the aggregation and finally split the data into county-level age-specific fertility rates. In case of missing county-level rates, we substitute the country-level fertility. To then combine the information, we first calculate the raw number of children being born at the county level (B_{county}) and at the country level fertility ($B_{country}$) separately. Subsequently, we adjust the number of newborns from mother-age a by all newborns at the county level, i.e., $a \in (16, 49)$, as well as those, who would have been born, if country-level fertility was used instead:

$$B^a = \frac{B^a_{county}}{\sum_{a=16}^{49} B_{county}} * B_{country} \tag{11}$$

Another assumption of variant G2-L2-W2 FSO (2019) addresses the development of the mean age of women at birth, which is assumed to increase from 31.2 years old in 2017 to 32.6 years old in 2060, respectively. In a similar vein, the annual birth rate is scheduled to decrease moderately over the same forecast horizon, from 1.57 to 1.55 children.

To model similar growth, we linearly and sequentially shift the age-specific fertility rates one year to the right, spread over the entire forecast horizon, thus fixing the annual birth rate to the 2019 initially observed 1.54 children. This process is illustrated in Figure 3. Over time, the highest age-specific fertility rates transition from age 32 to 33, thereby also increasing the mean age of women at birth, which reaches 32.6 in 2060, in line with FSO (2019).

Figure 3
Distribution of the probability of giving birth
For women aged 15-49 in 2020 and 2060



Notes: The figure shows the probability of giving birth for women aged 15-49 in 2020 (orange line) and 2060 (green line).
Source: Authors' graph. The data is given by Federal Statistical Office (FSO) (2022a).

2.3. Deaths and Life Expectancy (D)

To model mortality and deaths, information on life expectancy information is used in the form of country-level periodic mortality tables (Federal Statistical Office (FSO), 2022d). Specific death rates on lower regional levels are not used in the projection. Spatial variations in death rates are assumed to stem from worse health conditions in (former) employment and former differences in the medical care system. We assume that these differences in death rates will vanish over time. Starting in 1993, the data contain age-gender-specific mortality rates ($q_{s,t,a}$) calculated on the basis of a (hypothetical) birth cohort of 100,000 newborns, whose survival is modeled year by year (for details see Federal Statistical Office (FSO), 2022c). We consider all mortality rates from 2002 to the base year since, prior to 2002, the modeling is done only up to a maximum age of 89. This matrix of age-gender-specific mortality rates is fed into the popular Lee and Carter (1992) method, upon which the actual forecast is build:[5]

$$ln(q_{s,t,a}) = \alpha_{s,a} + \beta_{s,a} * \gamma_{s,t} + \epsilon_{s,t,a}, \quad (12)$$

where $ln(q_{s,t,a})$ is the natural logarithm of the (deterministic) age-gender-specific mortality rate at time t.[6] $\alpha_{s,a}$ represents the average mortality rates across age a of the mortality schedule of gender s. $\beta_{s,a} * \gamma_{s,t}$ combines the time-varying index, i.e., the overall level of mortality ($\gamma_{s,t}$), with the rate of change in response to changes in that index ($\beta_{s,a}$). A random walk with drift is assumed for $\gamma_{s,t}$. The error term $\epsilon_{s,t,a} \sim N(0, \sigma_\epsilon^2)$ reflects uncaptured irregular influences, such as age- or gender-specific deviations from the general trend of mortality rate development in response to pandemics (Lee & Carter, 1992).

2.4. Migration (M)

People are not fixed in one place but can move across space. Therefore, migration affects the local population pattern. To model migration, we consider two types of migration: On the one hand, people can move across international borders. We call this outer migration (OM), which is defined as net migration, i.e. the difference between immigrants to and emigrants from Germany. On the other hand, migration can also take place within a country. This inner migration (IM) allows movement across grid boundaries within Germany. The overall total net migration for each grid is therefore given as the sum of negative outward and positive inward migration, as summarized in Equation (5). Both types of migration are described in more detail below.

2.4.1. Outer Migration (OM)

We rely on FSO (2019) for a projection of net outer migration up to 2060. The data distinguishes between male and female migrants for four age groups. We convert the age group information to age-years by assuming a uniform distribution within each age group. Each age-year therefore contributes equally to the total number of migrants within each age group. After separating the age-years, we allocate the net migrants to each grid based on the existing share of people with the same age-year and gender. Distributing net migrants based on local conditions takes into account the local population composition. An alternative would be to distribute net migrants equally, so that each grid receives the same share for a given age-gender combination. However, this distribution scheme would not take into account the local distribution and could, for example, assign a higher share of young people to a grid with a relatively old population. This seems unlikely, as migrating to regions with similar age and gender distributions seems more reasonable.

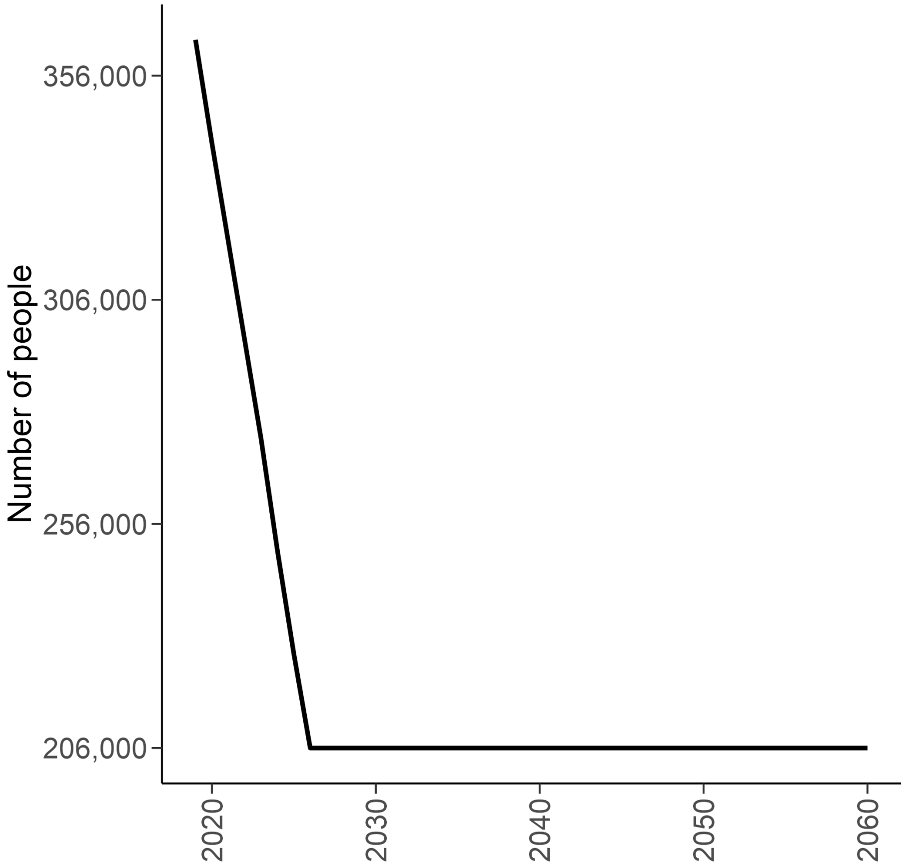
Furthermore, we follow FSO (2019) and assume a medium level of external migration, which decreases steadily until 2026 and stagnates thereafter. Figure 4 shows outer migration at the national level from 2019 to 2060, starting with 364,000 net migrants in 2019. This relatively large number decreases over time to 206,000 in 2026 and beyond, so that the overall average of net migrants is 221,000 people.[7]

[5]The estimation is performed by the R-package *forecast* (R. Hyndman et al., 2023; R. J. Hyndman & Khandakar, 2008).

[6]We assume that $q_{s,t,a} > 0$ for all combinations of s, t, and a.

[7]The number of 221,000 net migrants per year corresponds to the average of net migration between 1955 and 2018 and therefore, assumes a historical stable trend of migration (see FSO (2019)).

Figure 4
Outer migration 2019-2060
Net outer migration for Germany 2019-2060



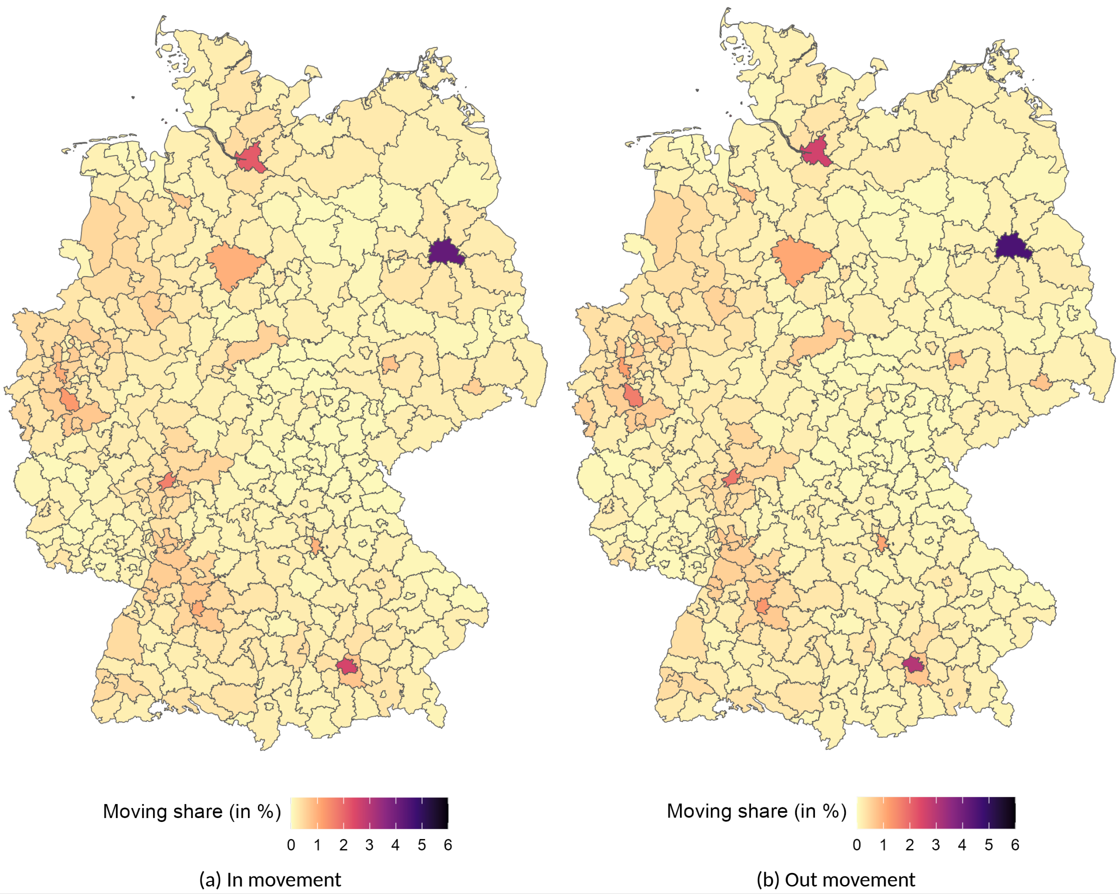
Notes: The figure shows the net outer migration for Germany between 2019 and 2060.
Source: Authors' graph. The original data is given by FSO (2019).

2.4.2. Inner Migration (IM)

Our second migration component, inner migration, cannot make any use of standards from FSO (2019) since the nationwide projection has no need for any assumptions on inner-country migration. In the absence of concrete migration patterns on the grid level, we use county-level information provided by Federal Statistical Office (FSO) (2019b). It provides information on the number of people moving in and out of a given county within Germany, by age group (six in total) and gender. As with all input data, we use information from the year 2019. Similar to out-migration, we convert age groups to age-years by assuming a uniform distribution of proportions within each age group for each gender. Dividing each age-gender combination by the total number of people moving in and out, respectively, yields a share of people moving for each county. We hold these shares constant throughout our time horizon, assuming that the share of people moving for a given county remains at the 2019 level through 2060, our final year of projection. Using these shares, we make the total number of people moving across the county line year-specific by calculating the total number of people living in each county for each age-gender combination and multiplying it by its share. Finally, we calculate net inner migration by subtracting in-migration and out-migration. This net inner migration figure is then distributed to each grid cell within the county so that each grid receives the number of net migrants that depends on the existing share for that age-gender pair. The reasoning is the same as for outer migration, where we want to account for local conditions.

Figure 5 shows the shares of people moving across county lines for 2019 - either moving in (Panel A) or moving out (Panel B). As described above, these shares remain fixed throughout the projection period. Both maps show that large metropolitan areas such as Berlin, Hamburg, or Munich have high rates of migration. Many people seem to move in and out of these areas, which makes sense since they are also areas of high economic and social activity.

Figure 5
Movement in and out at the county level
Share of people moving into the county (Panel A) and out of the county (Panel B) for 2019

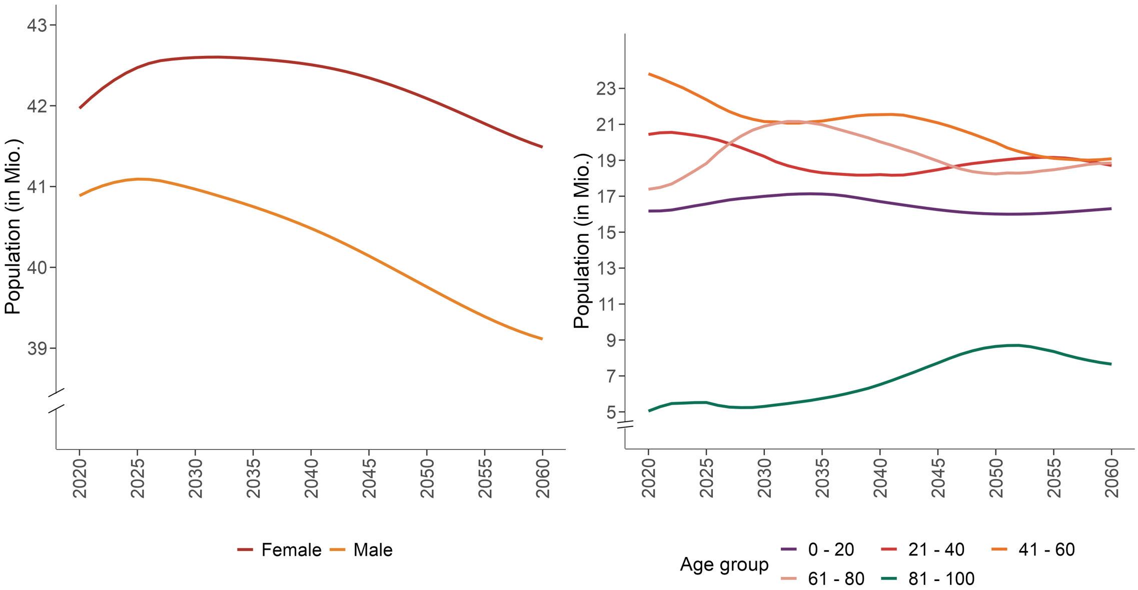


Notes: The figure shows the shares of people moving in (Panel A) and out (Panel B) at the county level for 2019.
Source: Authors' graph. The original data is given by Federal Statistical Office (FSO) (2019b).

3. Results

Development over time. First, we show the time trend for our population forecast. In 2020, there are around 83.4 million people living in Germany. This number decreases to 80.7 million in 2060, a decrease of about 2.7 million. Figure 6 shows that the population is roughly evenly split between males and females, with females having a slightly larger share. The overall time trend is similar for both genders. However, the total population for females shows a stable trend until 2040 and then decreases. The number of males starts to decrease as early as 2025. The overall declining trend can be explained by the large proportion of older citizens and the large birth cohorts of the boomer generation (born between World War II and the 1960s), which begin to pass away around this time. In addition, lower fertility rates amplify the general pattern. A longer (average) life expectancy for women could explain their initially stable trend.

Figure 6
Population trends by gender and age groups
Population over time by gender (Panel A) and by age groups (Panel B)

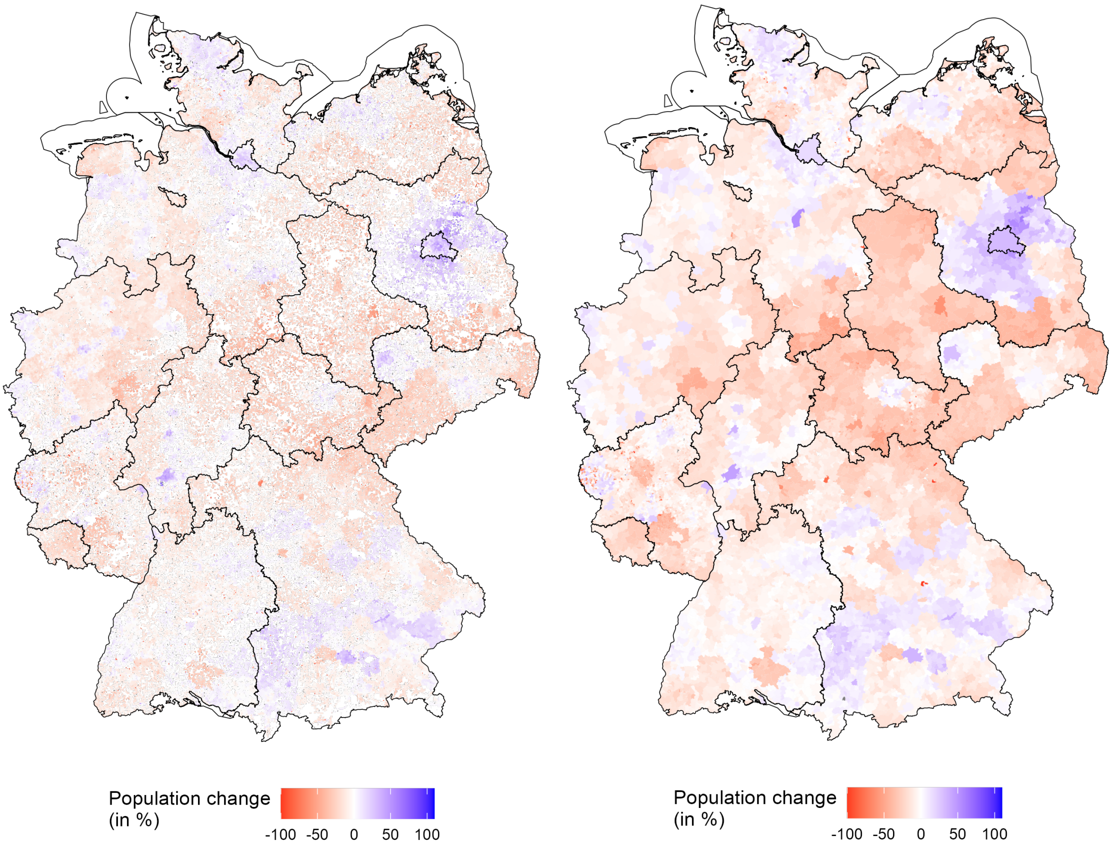


(a) Population by gender (b) Population by age groups

Notes: The figure shows the total number of women and men (Panel A) and the total number of people for five different age groups (Panel B) in Germany between 2020 and 2060.
Source: Authors' graph.

Regional development. Our population forecast not only allows us to draw conclusions about the temporal pattern of population change, but about the spatial development in individual regions. Regardless of the regional level, both maps in Figure 7 indicate that urban areas will have a larger population in 2060 than in 2020. The regions around Berlin will face the strongest growth, but areas in Bavaria and in the north around Hamburg and in Schleswig-Holstein will also gain population. At the other end of the spectrum, less densely populated areas and rural regions will lose people over time since the older population passes away (given the migration and fertility patterns) and the areas become less attractive. This is especially true for areas in central Germany, including Saxony-Anhalt, Saxony, and Thuringia, which are experiencing the largest population declines.

Figure 7
Population change between 2020 and 2060
Population change on the grid level (Panel A) and municipality level (Panel B)



(a) Population on the grid level (b) Population on the municipality level

Notes: The figure shows the population change at the grid level (Panel A) and at the municipality level (Panel B) in Germany between 2020 and 2060. Red colors indicate a decrease in local population, while blue colors represent population growth.
Source: Authors' graph.

The case of Berlin. Since the main advantage of our population projection is its high spatial resolution, we also show a brief case study regarding the city of Berlin. Figure 8a shows the population change between 2020 and 2060. Almost all grid cells increase in population. There are only a few areas that remain at the same level or even lose population. Such exceptions of declining grids within growing cities like Berlin are based on the fact that these grids are already comparatively old today. Therefore, they have fewer births (due to fewer women of childbearing age), more deaths (due to many old people), and fewer in-migrations (due to the distribution of in-migrations by age groups).[8] This example also illustrates a limitation of our projection. It cannot represent future spatial equilibria. Grids with declining population numbers have, for example, a larger housing supply in the future making them more attractive for in-migration. Such response patterns cannot be captured by a static projection.

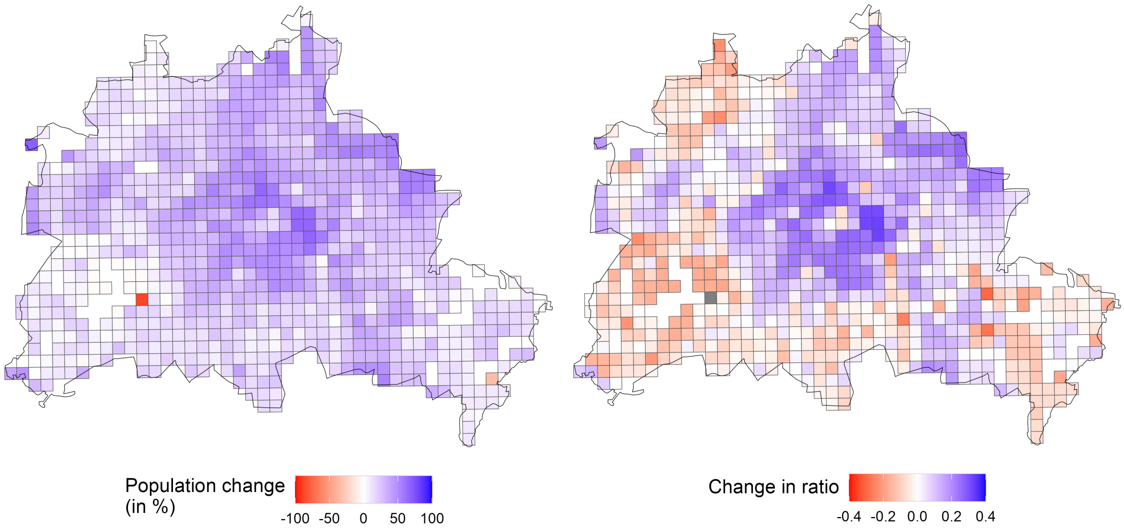
In general, the broad population growth is consistent with other sub-national population change

[8] Because older people have a much weaker migration behavior, the grids with many old people are also less treated by in-migration.

patterns, where Berlin and its neighboring municipalities are the main beneficiaries of the changing population scheme in Germany. It also seems reasonable that Berlin, a hot spot for working and living today, will continue to be a focal point and attract new residents in the future.

Besides the total growth, Berlin faces an aging population. Figure 8b plots the change in the ratio of people 65 and older to the number of people under that age between 2020 and 2060. The map shows that the city center and the northeastern grids, in particular, are populated by a higher proportion of older people in 2060 compared to the situation in 2020. The southwestern regions, on the other hand, will become younger as the ratio of older to younger people decreases.

Figure 8
Population change and change in old-young ratio in Berlin
Population change (Panel A) and change in the old-young ratio (Panel B) between 2020 and 2060 at the grid level for Berlin



(a) Population change (b) Old-young ratio

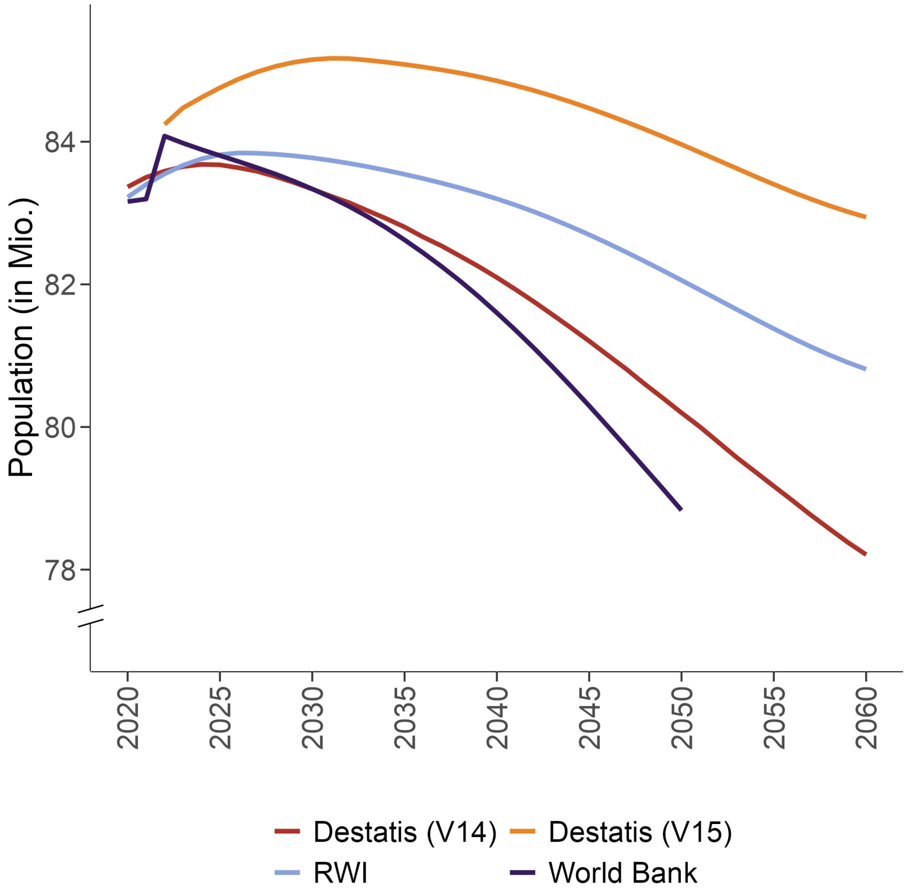
Notes: The figure shows the population change (Panel A) and the change in the old-young ratio (Panel B) at the grid level for Berlin between 2020 and 2060. The change in the old-young ratio is defined as the ratio of the number of people aged 65 and above to the number of people below that age, comparing 2020 and 2060 values.
Source: Authors' graph.

4. Validation and Comparison with Other Sources

While the proposed population projection is unique at the regional level, it is critical for credibility that our data product is consistent with other aggregated projections. We compare our population projections to those of the German Federal Statistical Office (FSO) and the World Bank. The FSO produces population projections at the state and national levels approximately every two years. The latest version (V15) was published in 2021. We compare our estimates to the last two FSO projections, with V14 (published in 2019) forecasting the population to 2060 and V15 forecasting the population to 2070 (see FSO (2019); FSO (2021)). To provide a comparison with a non-German source, we rely on data from the World Bank, which provides population figures for most countries up to 2050 (World Bank, 2023).

Figure 9 shows the population development over time according to the considered sources. All projections show a decreasing population trend in Germany. However, the sources differ in the speed of the decrease and the overall magnitude. The V15 projection of FSO (2021) (in orange) shows the largest total population for Germany independent of the year. The FSO adjusted their projection between the two versions, as the V14 (in red) shows a much lower overall projection over time. The RWI population forecast (in light blue) lies between the two. The World Bank projects the lowest population, although it is at a similar level to the others in the early 2020s. They also project the fastest population decline, which accelerates in 2035.

Figure 9
Projected population 2020 to 2060
Comparison of projections by the Federal Statistical Office (Destatis), the World Bank, and RWI.



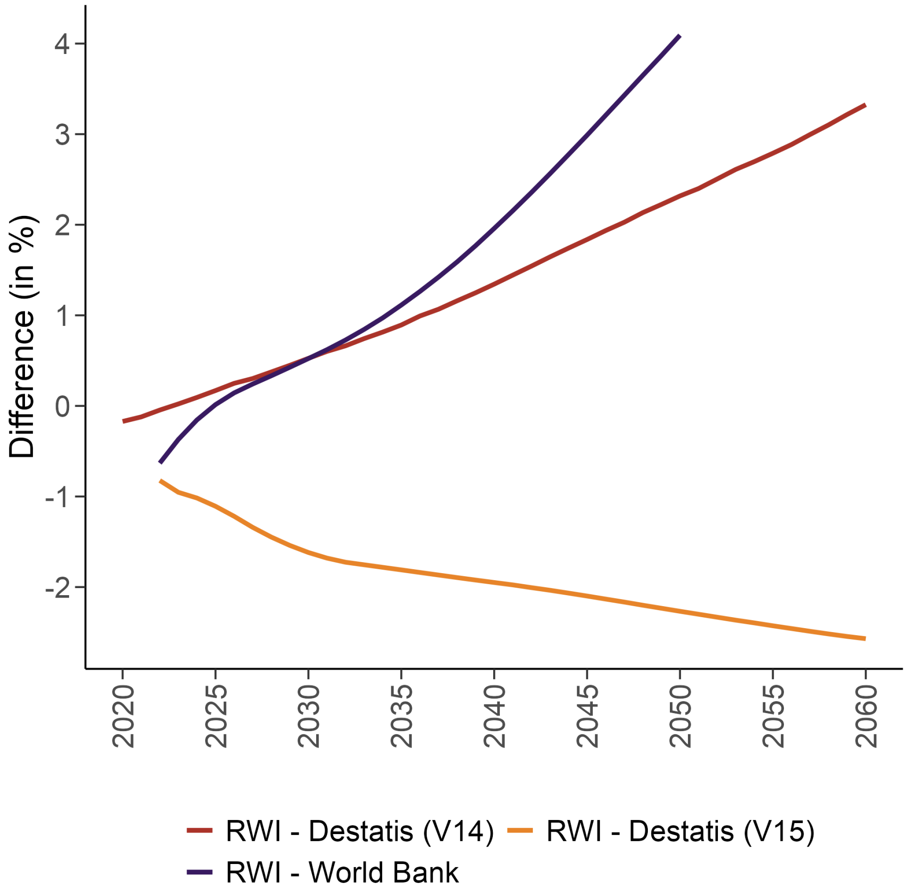
Notes: The figure shows the population projections over time provided by the Federal Statistical Office (Destatis), the World Bank, and the RWI.
Source: Authors' graph. The data is given by FSO (2019), FSO (2021), World Bank (2023), and RWI - Leibniz Institute for Economic Research (2023).

For a better comparison, we also plot the percentage difference between the RWI projection and the three sources (Destatis/ FSO V14 and V15 and the World Bank). Figure 10 shows that the deviation between RWI and Destatis (V14) and the World Bank projection increases over time. Starting in 2020, we have a deviation of 0.2% (≈ 167 tsd. people) compared to Destatis/ FSO V14 and 0.6% (≈ 500 tsd. people) compared to the World Bank. This difference increases to 3.3% and 4.1%, respectively, by the end of the forecast period. The average deviation between our forecast and the two sources is 1.4%.

Since Destatis/ FSO (V15) always forecasts a larger population than we do, the deviation between the two estimates is always negative. In 2020, the difference is 0.8% (≈ 667 tsd. people) and increases to 2.6% (≈ 2.1 million people) in 2060. The mean deviation is 1.9%.

The increasing deviation over time also highlights the uncertainty in forecasting the population. The structure implemented and the assumptions made reinforce each other over time. This path dependency leads to larger deviations over time.

Figure 10
Difference between various population forecasts
Comparison of projection by the Federal Statistical Office (Destatis), the World Bank, and RWI.



Notes: The figure shows the deviation between different population projections over time.
Source: Authors' graph. The data is given by Federal Statistical Office (FSO) (2019a, 2021), World Bank (2023)

5. Scope and Limitations

The proposed data set projects the German population up to 2060 and offers a high spatial resolution of 1km^2 grids. This is the first national projection that considers spatial heterogeneity at such a detailed level. However, due to the assumptions described above and the underlying input data, the data set has some limitations.

We generally follow the assumptions of variant G2-L2-W2 (Federal Statistical Office (FSO), 2019a), which assumes a medium scenario for births, life expectancy, and migration. Other scenarios are possible, and indeed the Federal Statistical Office provides data for a different set of assumptions. We have chosen the middle case because it provides a well-rounded specification without emphasizing any particular component.

Other limitations arise from the lack of more detailed information. For example, we assume that people do not move within their county of residence, but only across county borders. This limitation is due to data availability, as information on entries and exits is only available for cross-border movements. Furthermore, we assume that migrants (internal and external) settle according to the existing population distribution. Existing population patterns are thus facilitated, so that, for example, sparsely populated areas today will have low population densities in the future.

The data set can be merged with all data sets containing precise geographic information, and it is also possible to add further (grid-level) information, e.g., by merging RWI-GEO-GRID (RWI - Leibniz Institute for Economic Research, 2022), allowing detailed analysis. Furthermore, the data can be aggregated to any higher regional level, e.g., postcodes or municipalities.

6. Data Structure and Data Access

The final data set is structured as follows: Each predicted year is stored in a separate CSV file, allowing the researcher to load only the required years and avoid cluttering up memory space. In addition, each file contains five columns as described in Table 1.

Table 1
Variable description
Name, type, and description of all included variables

Variable name	Variable type	Description	Example value
idm	Character	ID of the grid cell	4031_3109
year	Integer	Current year of the file	2035
sex	Integer	Gender of the population count where 1 represents females and 0 represents males	1
age	Integer	Age of the population count (between 0 and 100 years)	85
population	Numeric	Actual absolute number of people, not a percentage	3.3642

Notes: The table describes all the included variables.
Source: Author's table.

The data are available as Scientific Use File (SUF) for scientific, non-commercial research (https://doi.org/10.7807/pop:forecast:suf:v2). The data are available as CSV files. The data are also available for individual years, where each row of the data set specifies the population for a particular age-gender combination.

Data access requires a signed data use agreement. Only researchers from scientific institutions are eligible to apply for data access to the SUF. The SUF can be used at the user's workplace. Upon request, we can provide snapshots of the data or data at higher aggregated regional levels as a Public Use File (PUF) for non-researchers. Data access is provided by the Research Data Center Ruhr at the RWI - Leibniz Institute for Economic Research (FDZ Ruhr). Applications for data access can be submitted online at https://www.rwi-essen.de/en/research-advice/further/research-data-center-ruhr-fdz/data-access. The application form contains a short description and title of the project, the cooperation, the department, the expected duration of the data use as well as the participants in the project.

7. References

Annoni, A. (2003). European Reference Grids - Proposal for a European Grid Coding System. *Proceedings and Recommendations of the European Reference Grids Workshop*, 27–29.

BBSR. (2021). Die Raumprognose 2040 nach dem Zensus. https://www.bbsr.bund.de/BBSR/DE/veroeffentlichungen/analysen-kompakt/2021/ak-18-2021.html

Breidenbach, P., & Eilers, L. (2018). RWI-GEO-GRID: Socio-economic data on grid level. *Jahrbücher für Nationalökonomie und Statistik*, (6), 609–616. https://doi.org/10.1515/jbnst-2017-0171

Cannan, E. (1895). The probability of a cessation of the growth of population in england and wales during the next century. *The Economic Journal*, *5*(20), 505–515.

Federal and State Statistical Offices (FSOS). (2022). Lebendgeborene nach Alter der Mütter - Jahressumme - regionale Tiefe: Kreise und krfr. Städte [12612-93-01-4]. [Data set. Accessed: July 2023].

Federal Statistical Office (FSO). (2019a). Bevölkerung Deutschlands bis 2060 - Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung - Hauptvarianten 1 bis 9. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsvorausberechnung/_inhalt.html#_qeixhpo4b [Data set. Accessed: July 2023]

Federal Statistical Office (FSO). (2019b). Zu- und Fortzüge (über Kreisgrenzen) nach Geschlecht und Altersgruppen - Jahressumme - regionale Tiefe: Kreise und krfr. Städte. [Data set. Accessed: May 2023].

Federal Statistical Office (FSO). (2021). Bevölkerung Deutschlands bis 2070 - Ergebnisse der 15. koordinierten Bevölkerungsvorausberechnung. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsvorausberechnung/begleitheft.html?nn=208696#bundesl%C3%A4nder [Data set. Accessed: July 2023]

Federal Statistical Office (FSO). (2022a). Geburtenziffern (Lebendgeborene je 1000 Frauen): Deutschland, Jahre, Alter [12612-0008]. [Data set. Accessed: July 2023].

Federal Statistical Office (FSO). (2022b). Lebendgeborene: Deutschland, Jahre, Geschlecht [12612-0001]. [Data set. Accessed: July 2023].

Federal Statistical Office (FSO). (2022c). *Qualitätsbericht zur laufenden Berechnung von Periodensterbetafeln für Deutschland und die Bundesländer* (tech. rep.).

Federal Statistical Office (FSO). (2022d). Sterbetafel (Periodensterbetafel): Deutschland, Jahre, Geschlecht, Vollendetes Alter [12621-0001]. [Data set. Accessed: July 2023].

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., & Yasmeen, F. (2023). *forecast: Forecasting functions for time series and linear models* [R package version 8.21.1]. https://pkg.robjhyndman.com/forecast/

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *26*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, *87*(419), 659–671.

Preston, S., Heuveline, P., & Guillot, M. (2000). Demography: Measuring and modeling population processes. 2001. *Malden, MA: Blackwell Publishers*.

RWI - Leibniz Institute for Economic Research. (2022). RWI-GEO-GRID: Socioeconomic Data on Grid Level (Wave 12). https://doi.org/10.7807/microm:v12 [Data set].

RWI - Leibniz Institute for Economic Research. (2023). RWI-GEO-GRID-POP-FORECAST: Small-Scale Population Projection (Version 2). https://doi.org/10.7807/pop:forecast:suf:v2 [Data set].

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

Whelpton, P. K. (1936). An empirical method of calculating future population. *Journal of the American Statistical Association*, *31*(195), 457–473.

World Bank. (2023). Population Estimates and Projections. https://databank.worldbank.org/source/population-estimates-and-projections [Data set. Accessed: July 2023].

Leibniz-Institut für Wirtschaftsforschung

Das RWI wird vom Bund und vom Land Nordrhein-Westfalen gefördert.

Leibniz Gemeinschaft