SSOAR

Open Access Repository
www.ssoar.info

Adjusting Inter-censal Population Estimates for Germany 1987-2011: Approaches and Impact on Demographic Indicators - Online Appendix

Klüsener, Sebastian; Grigoriev, Pavel; Scholz, Rembrandt D.; Jdanov, Dmitri A.

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der

Leibniz-Gemeinschaft

Online Appendix

Adjusting Inter-censal Population Estimates for Germany 1987-2011: Approaches and Impact on Demographic Indicators[*]

Sebastian Klüsener, Pavel Grigoriev, Rembrandt D. Scholz, Dmitri A. Jdanov

1 Overview of the available data

Data type	1988-1990	1991-1999	2000	2001-2013
Population estimates by age and sex (90+) (1 January[1])	Federal states, West Berlin, East Berlin	Federal states, West Berlin, East Berlin	Federal states, West Berlin, East Berlin	Federal states, West Berlin[e], East Berlin[e]
Deaths by age, birth cohort, and sex	Germany, West Germany, East Germany	Germany, West Germany, East Germany	Germany, West Germany, East Germany	Germany, West Germany, East Germany
Deaths by age and sex	Federal states, West Berlin, East Berlin	Federal states, West Berlin, East Berlin	Federal states, West Berlin, East Berlin	Federal states, West Berlin[e], East Berlin[e]
Births	Federal states, West Berlin, East Berlin	Federal states, West Berlin, East Berlin	Federal states, West Berlin, East Berlin	Federal states, West Berlin[e], East Berlin[e]
Migration to and from other states in Germany[2,3]	West Germany (incl. West Berlin)	Germany[4], Federal states[5], West Berlin[5], East Berlin[5]	Federal states, West Berlin[e], East Berlin[e]	Federal states, West Berlin[e], East Berlin[e]
International Migration[2]	West Germany[6], East Germany[7]	Federal states, West Berlin, East Berlin	Federal states, West Berlin[e], East Berlin[e]	Federal states, West Berlin[e], East Berlin[e]
Implied migration balance (derived from population estimates at the beginning and the end of a year; and births and deaths by birth cohort and sex during the year).	Germany, West Germany, East Germany	Germany West Germany, East Germany	Germany, West Germany, East Germany	Germany, West Germany, East Germany

[e] Estimates

[1] The official data uses 31 December of the preceding year as the reference point.

[2] If not mentioned otherwise, migration data are available by single-year ages at least up to ages 90+.

[3] In principle, we have available for 1991-2011 the full migration matrices that provide us with information on combinations of moves by the federal state of origin and the federal state of destination. We decided, however, not to make use of this additional information in our adjustment procedure, as we felt that the additional complication would not be justified by the potential gains from using an even more data-intensive approach.

[4] Data for this period are available by single-year ages, but only up to ages 75+.

[5] Only available for broad age categories (<18, 18-24, 25-29, 30-49, 50-64, 65+).

[6] Includes migration to and from East Germany up to 3 October 1990. Data are available by single-year ages, but only up to ages 90+, and only up to ages 75+ for 1988 and 1989.

[7] Only available for both sexes and broad age categories.

2   Deriving migration estimates for East and West Berlin (2000-2013)

As we explained in the main text, because Berlin implemented a reform of its administrative division in the early 2000s, we cannot distinguish between the former territories of East and West Berlin based on published statistical data from 2000/2001 onwards. The specific problem is that the two newly formed city districts of Mitte and Kreuzberg-Friedrichshain straddle territories that formerly belonged to East and West Berlin. To derive estimates for the migration counts that occurred in the territories of East and West Berlin between 2000 and 2013, we obtained migration counts by age and sex for the 12 city districts of Berlin for this period. In splitting up the counts, we use the highly reliable population estimates by age and sex for East and West Berlin that were derived based on the methodology described by *Scholz et al.* (2017). This approach allows us to identify the share of individuals by single year of age and sex living in the parts of the two city districts of Mitte and Kreuzberg-Friedrichshain that formerly belonged to East Berlin. In notation form, this share is derived as follows:[1]

$$\pi_{y,a} = \frac{E^{Mi+KF,East}(y,a)}{E^{Mi+KF}(y,a)},$$
A2.1

with π denoting the share as a number between zero and one and $E(y,a)$ representing the mean population in a given year y at age a obtained by summing up the population on 1 January of the given year and the population on 1 January at the beginning of the next year and dividing it by two. The superscript *Mi+KF* denotes the population of the city districts Mitte and Kreuzberg-Friedrichshain, while $E^{Mi+KF,East}$ includes only the shares of the population of these two city districts who were living in territories of these city districts that are located in former East Berlin.

   In obtaining the estimates for the migration counts that occurred in East and West Berlin, we first considered using information on the differences in migration intensities in the city districts that are either solely in East Berlin or solely in West Berlin. This approach seemed appealing in part because East Berlin has lower migration intensities to and from other countries and other federal states than West Berlin. However, such an approach turned out not to be feasible, as the two city districts with overlapping areas are situated in the centre of Berlin, where the migration intensities are even higher than in the city districts that are solely located in West Berlin. Hence, it would be wrong to assume that the migration intensities in these central districts are composed of "West Berlin-type" migration intensities in the areas that belonged to former West Berlin and "East Berlin-type" migration intensities in the other areas. We therefore decided to use a simple, straightforward approach that is based on the assumption that there are no differences in the migration intensities between the people who live in the two city districts of Mitte and

---

[1]   All of the calculations described in section 2 of this online appendix are carried out separately by sex.

Kreuzberg-Friedrichshain depending on whether they are residing in territories that were formerly part of East or West Berlin. Thus, we have chosen to simply divide up the migration counts by year, type, and age by the population shares derived in equation A2.1. In notation form, this is formulated as follows, with M denoting the migration counts considered (these include migration events to and from another country and to and from another German state):

$$M^{East}(y,a) = M^{East-Mi-KF}(y,a) + \pi(y,a)M^{Mi+KF}, \qquad\qquad \text{A2.2}$$

where $M^{East\text{-}Mi\text{-}KF}$ represents all of the migration events in those city districts that solely cover territories which were part of East Berlin before 1990. The migration events for West Berlin are estimated as follows:

$$M^{West}(y,a) = M^{West-Mi-KF}(y,a) + \big(1 - \pi(y,a)\big)M^{Mi+KF}. \qquad\qquad \text{A2.3}$$

3   Splitting abridged migration data into single-year ages

For the years 1988-1989 and 1991-1999, the data on internal migration in Germany (1988-1989: West Germany) are only available by single-year ages up to ages 75+. In order to derive estimates on migration events by single-year ages up to ages 90+, we take as a reference for the years 1988 and 1989 the proportional distribution of migration counts by single-year ages for ages 75-90+ in 1990.[2] For the period 1991-1999, we use the proportional distribution of migration counts by single-year ages for ages 75-90+ in 1990 (West Germany only) and 2000 (Germany) as a reference, and assume that the proportions by age obtained in 1990 had linearly shifted to the proportions by age obtained in the year 2000 over that decade. We face the limitation that for 1990 the available data for Germany cover West Germany only. Fortunately, however, in 1990 almost 82 percent of the German population aged 75+ were living in West Germany.

Migration data for East and West Germany (as well as for the federal states and East and West Berlin) for the 1991-1999 period are available in abridged format for the following age categories: <18, 18-24, 25-29, 30-49, 50-64, and 65+. To obtain estimates for migration by single-year ages, we assume that the composition of single-year ages within a broad age category (weights) derived on the basis of in-migration data by single-year ages for Germany as a whole can be applied to abridged migration data for East and West Germany (as well as the federal states and East and West Berlin). The application of this simple procedure has returned plausible results that were confirmed by the visual examination of the obtained migration patterns, as well as by comparing them with official data on migration patterns by single-year ages from 2000 onwards.

---

[2]   All of the calculations described in section 3 of this online appendix are carried out separately by sex and region (15 federal states without Berlin, and East and West Berlin).

4   Estimation of migration events by cohort from migration events by age

The number of migration events M at age x in year t for cohort c (i.e., by Lexis triangles) is estimated as follows:

$$M_{t,x,c} = M_{t,x,.} \frac{P_{t,x} + P_{t+1,x+1}}{(P_{t,x} + P_{t+1,x+1}) + (P_{t,x-1} + P_{t+1,x})},$$ A4.1

where $M_{t,x,.}$ represents the number of migration events at age x in year t, while $P_{t,x}$ is the population on 1 January of year t at age x.[3] The number of migration events by cohort c in year t is calculated as the sum of the respective Lexis triangles:

$$M_{t,.,c} = M_{t,x,c} + M_{t,x+1,c}.$$ A4.2

5   Equations of the inter-censal adjustment approaches

Basic approach[4]

In this adjustment approach, we rely solely on the birth and death data during the inter-censal period (see also *Wilmoth et al.* 2007). Below we provide a description of the method we use, which we have taken from the HMD Methods Protocol (*Wilmoth et al.* 2007) and implemented for Germany. In our case, both censuses (as used in our calculations) "occurred" at the beginning of the year. For cohorts born before the beginning of the inter-censal period, we rely on information on the size of the cohort at the start of the inter-censal period (in our case, 1 January 1988) to "estimate" the size at the end of the inter-censal period (1 January 2012) using the following equation:

$$\hat{C}_2(x + N) = C_1(x) - \sum_{i=0}^{N-1}[D_U(x + i, 1988 + i) + D_L(x + i + 1, 1988 + i)].$$ A5.1

In this formula, N denotes the length of the inter-censal period (in our case, 24 years), $C_1(x)$ is the census count for persons aged $[x, x + 1]$ at the beginning of the inter-censal period (1 January 1988), and $\hat{C}_2(x + N)$ is the estimate of the cohort size

---

[3]   All of the calculations described in section 4 of this online appendix are carried out separately by sex and region (in the prior-cleaning adjustment: 15 federal states without Berlin, and East and West Berlin; in the inter-censal adjustment: East Germany, West Germany, and Germany as a whole).

[4]   For all of the approaches apart from the basic approach, we use migration data and/or inter-censal population estimates that we adjusted for prior cleaning. All of the calculations described in section 5 of the online appendix are carried out separately by sex and region (East Germany, West Germany, Germany as a whole; the population-size adjusted approach is also applied to data for the 16 federal states and East and West Berlin).

at the end of this period (1 January 2012). Meanwhile, 1988 is the beginning of the inter-censal period. $D_U(x, t)$ and $D_L(x, t)$ are related to deaths that occurred in year t at age x in the cohorts born in year $t - x - 1$ or, respectively, $t - x$; i.e., in the upper or lower Lexis triangles.

The estimated population size P on 1 January of each year is derived as follows:

$$P(x + n, 1988 + n)$$
$$= C_1(x)$$
$$- \sum_{i=0}^{n-1} [D_U(x + i, 1988 + i) + D_L(x + i + 1, 1988 + i)] + \frac{n}{N} \Delta_x, \quad \text{A5.2}$$

with n representing the years since the beginning of the inter-censal period, and Δ_x denoting the total estimated migration (including errors) during the inter-censal period for the cohort aged x at the beginning of this period; i.e.:

$$\Delta_x = C_2(x + N) - \hat{C}_2(x + N). \quad \text{A5.3}$$

For the cohorts who were born in the inter-censal period, it is relevant to take into account that the exposures in the first year of life are affected by the fact that the size of a new-born cohort was zero at the beginning of the year. Let K be the age of a new-born cohort at the end of the inter-censal period (1 January 2012) and $B(t)$ be the number of births in year t. We derive an estimate of the cohort born in $2011 - K$ at the end of the inter-censal period using the following equation:

$$\hat{C}_2(K) = B(2011 - K) - D_L(0, 2011 - K)$$
$$- \sum_{i=1}^{K} [D_U(i - 1, 2011 - K + i) + D_L(i, 2011 - K + i)], \quad K < N. \quad \text{A5.4}$$

The population size P on 1 January of each year from birth until the second census is estimated as:

$$P(k, 2011 - K + k)$$
$$= B(2011 - K) - D_L(0, 2011 - k)$$
$$- \sum_{i=1}^{k} [D_U(i - 1, 2011 - K + i) + D_L(i, 2011 - K + i)]$$
$$+ \frac{2k + 1}{2K + 1} \Delta'_K, \quad k = 0, \dots, K, \quad \text{A5.5}$$

with:

$$\Delta'_K = C_2(K) - \hat{C}_2(K). \qquad\qquad\qquad \text{A5.6}$$

Migration-adjusted approach

In this adjustment approach, we also trust in the reliability of registered in-migration events (from other countries and from other states within Germany), and believe that the occurrence of out-migration events (abroad or to other German states) over time by age and sex is informative for the distribution of the prior-cleaning-adjusted accumulated error. For cohorts born before the census, we modify the equation (A5.1) as follows:

$$
\begin{aligned}
\hat{C}_2(x + N) = C_1(x) &- \sum_{i=0}^{N-1} [D_U(x + i, 1988 + i) + D_L(x + i + 1, 1988 + i)] \\
&+ \sum_{i=0}^{N-1} [EMI_U(x + i, 1988 + i) + EMI_L(x + i + 1, 1988 + i)] \\
&+ \sum_{i=0}^{N-1} [IMI_U(x + i, 1988 + i) + IMI_L(x + i + 1, 1988 + i)]
\end{aligned}
\qquad \text{A5.7}
$$

with *EMI* denoting the external in-migration events across national borders, while *IMI* represents the internal in-migration events across federal state borders. The meaning of the subscripts U and L is similar to that of the deaths: they denote the number of events that occurred in the upper and lower Lexis triangles in a respective year at a respective age. As we mentioned above, the obtained data on migration events are by age only. We therefore derive the estimates for the upper and the lower Lexis triangles using equations (A4.1) and (A4.2) (see section 4 of this appendix). We are only able to derive these migration data by cohort up to age 87. For reasons provided in section 5.1 of our paper, we assume that there was no migration above that age.[5]

The estimated population size on 1 January of each year is then derived as follows:

---

[5] For East Germany, we have no information on internal migration across the borders of the East German federal states for the 1988-1990 period (these states only came into existence in 1990). For migration out of East Germany – which at that time was mostly to West Germany – we decided to use data on implied migration, as we can obtain these data by single-year ages (see data section 3 of the paper for more details). This choice affects our calculations for East Germany and Germany as a whole.

$$P(x + n, 1988 + n)$$

$$= C_1(x) - \sum_{i=0}^{n-1} [D_U(x + i, 1988 + i) + D_L(x + i + 1, 1988 + i)]$$

$$+ \sum_{i=0}^{n-1} [EMI_U(x + i, 1988 + i) + EMI_L(x + i + 1, 1988 + i)] \qquad \text{A5.8}$$

$$+ \sum_{i=0}^{n-1} [IMI_U(x + i, 1988 + i) + IMI_L(x + i + 1, 1988 + i)]$$

$$+ \sum_{i=0}^{n-1} w_{x,i} \Delta_x,$$

with w representing a weight accounting for temporal variation in out-migration events:

$$w_{x,n} = \frac{EMO_U(x + n, 1988 + n) + EMO_L(x + n + 1, 1988 + n)}{W_x}$$

$$+ \frac{IMO_U(x + n, 1988 + n) + IMO_L(x + n + 1, 1988 + n)}{W_x}, \qquad \text{A5.9}$$

with:

$$W_x = \sum_{i=0}^{N-1} [EMO_U(x + i, 1988 + i) + EMO_L(x + i + 1, 1988 + i)]$$

$$+ \sum_{i=0}^{N-1} [IMO_U(x + i, 1988 + i) + IMO_L(x + i + 1, 1988 + i)]. \qquad \text{A5.10}$$

In these equations, *EMO* represents the external out-migration events across national borders, while *IMO* denotes the internal out-migration events across German federal state borders.

For the cohorts born in the inter-censal period, the equation (A5.4) in the basic approach is replaced by the following equation:

$$\hat{C}_2(K) =$$
$$B(2011 - K) - D_L(0, 2011 - K) + EMI_L(0, 2011 - K) + IMI_L(0, 2011 - K)$$

$$- \sum_{i=1}^{K} [D_U(i - 1, 2011 - K + i) + D_L(i, 2011 - K + i)]$$

$$+ \sum_{i=1}^{K} [EMI_U(i - 1, 2011 - K + i) + EMI_L(i, 2011 - K + i)] \qquad \text{A5.11}$$

$$+ \sum_{i=1}^{K} [IMI_U(i - 1, 2011 - K + i) + IMI_L(i, 2011 - K + i)],$$

$$K < N.$$

The population size P on 1 January of each year from birth until the end of the inter-censal period is estimated as:

$$P(k, 2011 - K + k)$$

$$= B(2011 - K) - D_L(0, 2011 - k) + EMI_L(0, 2011 - k)$$

$$+ IMI_L(0, 2011 - k)$$

$$- \sum_{i=1}^{k} [D_U(i - 1, 2011 - K + i) + D_L(i, 2011 - K + i)]$$

$$+ \sum_{i=1}^{k} [EMI_U(i - 1, 2011 - K + i) + EMI_L(i, 2011 - K + i)]$$

$$+ \sum_{i=1}^{k} [IMI_U(i - 1, 2011 - K + i) + IMI_L(i, 2011 - K + i)]$$

$$+ \sum_{i=0}^{k} w_i' \Delta_K', \quad k = 0, \dots, K,$$

A5.12

with Δ'_K defined in (5.6) and:

$$w_k' = \frac{EMO_U(k, 2011 - K + k) + EMO_L(k, 2011 - K + k)}{W_K'}$$

$$+ \frac{IMO_U(k, 2011 - K + k) + IMO_L(k, 2011 - K + k)}{W_K'}, k = 1, \dots$$

A5.13

$$w_k' = \frac{EMO_L(k, 2011 - K) + IMO_L(k, 2011 - K)}{W_K'}, k = 0,$$

with:

$$W_K' = EMO_L(k, 2011 - K) + IMO_L(k, 2011 - K)$$

$$+ \sum_{i=1}^{K} [EMO_U(i - 1, 2011 - K + i) + EMO_L(i, 2011 - K + i)]$$

$$+ \sum_{i=1}^{K} [IMO_U(i - 1, 2011 - K + i) + IMO_L(i, 2011 - K + i)].$$

A5.14

Simplified migration-adjusted approach

For the *simplified migration-adjusted approach,* we also consider both variation in total migration intensities over time and variation in migration intensities by age. But the variation in migration intensities by age is obtained from a reference year in the middle of the inter-censal period for which we have readily available data; the reference year we chose is 2000. Unlike in the second approach, in this simplified approach we consider in- and out-migration data to be equally reliable, and we use these data to derive the migration intensities (using information on migration both across national borders and across federal state borders within Germany). Thus, we can implement our inter-censal adjustment based on the published official population estimates, which are, however, adjusted for prior cleaning.[6] In calculating the weights, we consider only ages up to age 87. Because migration intensities are low at very old ages, we assume that no additional error emerged after age 87. We first derive the intensities by age (at the end of the year) for our reference year:

$$MI^{model}(x,.) = \frac{TM_U(x-1,2000) + TM_L(x,2000)}{NC(x,2000)}, \qquad x = 0,1,\dots,87.$$

A5.15

Here, $TM_L(x, 2000)$ and $TM_U(x, 2000)$ denote the sum of the registered external and internal in- and out-migration events at ages x and $x - 1$ in cohorts born in $2000 - x$ in the year 2000, while $NC(x, 2000)$ is the average size of the cohort who reached age x during the year (derived by taking the mean of the cohort size at the beginning and at the end of the year). We then apply this variation in intensities across ages at the end of the year (i.e., across cohorts) to all years y, and transfer the intensities back into counts by multiplying the intensities with the population data:

$$TM^{model}(x,y) = MI^{model}(x,.) \cdot NC(x,y),$$

A5.16

where

$$TM^{model}(x,y) = TM_U^{model}(x-1,y) + TM_L^{model}(x,y), \quad x > 0,$$

A5.17

and

$$TM^{model}(0,y) = TM_L^{model}(0,y).$$

A5.18

For cohorts $NC(x,y)$, we again take the mean of the cohort size at the beginning and at the end of the year y. From these counts, we derive by cohort proportional

---

[6] See section 4.1 of the paper.

weights on variation in migration events over the years, which are determined by variation across ages as derived from our reference year:

$$w^{age}(x,c) = \frac{TM^{model}(x, c+x)}{\sum_{x_{min}}^{x_{max}} TM^{model}(i, c+i)},$$

A5.19

$$x_{min} = \max(0, 1988 - c), \quad x_{max} = \max(87, 2011 - c).$$

As a next step, we derive the intensities by year:[7]

$$MI^{model}(., y) = \frac{\sum_{x_{min}}^{x_{max}} TM^{model}(x, y)}{\sum_{i=1}^{NC} \sum_{x_{min}}^{x_{max}} TM^{model}(x, i)}.$$

A5.20

We then turn these intensities by cohort into proportions over time:

$$w^{period}(x,c) = \frac{MI^{model}(., c+x)}{\sum_{x_{min}}^{x_{max}} MI^{model}(., c+i)}.$$

A5.21

This allows us to derive the overall weights as an average of weights defined in (A5.19) and (A5.21):

$$w(x,c) = \frac{w^{age}(x,c) + w^{period}(x,c)}{2}.$$

A5.22

Finally, the population estimates are calculated as follows:

$$P(x,y) = \hat{P}(x,y) + \sum_{i=x_{min}}^{x} w(i, y-x)\,\hat{\Delta}(y-x),$$

A5.23

where $\hat{P}(x,y)$ are the official population estimates adjusted for prior cleaning (post-censal estimates based on the censuses of 1981 and 1987), and $\hat{\Delta}(y-x)$ is the difference between $\hat{P}(x,y)$ and the population estimates based on the 2011 census as of 1 January 2012:

$$\hat{\Delta}(y-x) = \hat{P}(y - x - 2012) - C_2(y - x - 2012).$$

A5.24

---

[7] Unlike in our calculations of the rates by age, we consider here all migration events, including the small proportion of migration events after age 87, to derive the intensities by year.

Population-size-adjusted approach

In this approach, we also do not need to derive our own estimates of the population during the inter-censal period. Instead, we take the prior-cleaning-adjusted[8] official post-censal population estimates based on the censuses of 1981 and 1987, and account in the distribution of the accrual of the accumulated error only for changes in cohort size over time. While this approach is similar to the migration-adjusted method, it uses a different definition of weights and data up to ages 89+ instead of up to ages 88+ (see section 5.1 of the paper for details). More precisely, in this approach population estimates are calculated using (A5.23) – (A5.24), with the weights being derived as follows from the population data for Germany as a whole:

$$w(x,c) = \frac{\hat{P}(x, c+x)}{\sum_{x_{min}}^{x_{max}} \hat{P}(i, c+i)}. \qquad \text{A5.25}$$

As we explained in section 5.1 of the paper, the weights obtained for Germany as a whole are then also used for the sub-territories East and West Germany (as well as for the 16 German states and East and West Berlin, for which we provide data in the online data appendix).

References

Scholz, Rembrandt et al. 2017: About Mortality Data for Germany (Last Update: April 21nd, 2017). Berkeley/Rostock: University of California, Berkeley/Max Planck Institute for Demographic Research [http://www.mortality.org/hmd/DEUTNP/InputDB/DEUTNPcom.pdf, 19.02.2018].

Wilmoth, John R. et al. 2007: Methods Protocol for the Human Mortality Database: Version 5. Berkeley/Rostock: University of California, Berkeley/Max Planck Institute for Demographic Research.

---

[8] See section 4.1 of the paper.

Sebastian Klüsener (✉). Max Planck Institute for Demographic Research. Rostock, Germany; Vytautas Magnus University. Kaunas, Lithuania.
E-mail: kluesener@demogr.mpg.de
URL: https://www.demogr.mpg.de/en/institute/staff_directory_1899/sebastian_kluesener_1644.htm

Pavel Grigoriev, Rembrandt D. Scholz. Max Planck Institute for Demographic Research. Rostock, Germany.
E-mail: grigoriev@demogr.mpg.de, scholz@demogr.mpg.de
URL: https://www.demogr.mpg.de/en/institute/staff_directory_1899/pavel_grigoriev_1578.htm
URL: https://www.demogr.mpg.de/de/institut/mitarbeiter_1899/rembrandt_d_scholz_663.htm

Dmitri A. Jdanov. Max Planck Institute for Demographic Research. Rostock, Germany; National Research University Higher School of Economics. Moscow, Russia.
E-mail: jdanov@demogr.mpg.de
URL: https://www.demogr.mpg.de/en/institute/staff_directory_1899/dmitri_a_jdanov_516.htm