Dr. Patrizio Vanella Epidemiology Department Helmholtz Centre for Infection Research



# PARSIMONIOUS STOCHASTIC FORECASTING OF INTERNATIONAL AND INTERNAL MIGRATION ON THE NUTS-3 LEVEL – AN OUTLOOK OF REGIONAL DEPOPULATION TRENDS IN GERMANY

Seite 1 | Vanella – Parsimonious Stochastic Forecasting of International and Internal Migration – Wittgenstein Centre Conference 2021



HELMHOLTZ Zentrum für Infektionsforschung

## Background

- Migration shapes regions the strongest in the short term
- Impact on the population structure, particularly for smaller regions
- High stochasticity of migration as it is more sensitive toward acute events, such as crises
- High connection between international and internal migration flows (e.g., correlation between international immigration flows 1994-2018 and cross-district (NUTS-3) level migration 1995-2019 is 94.8% for Germany
- Economically weaker regions often suffer persistent net emigration, especially among more educated and skilled sub-populations
- depopulation among young and fertile population leads to low births in the long run
- Important to predict future developments in internal migration to enact appropriate policies early



### **Motivation**

- Migration projections on the NUTS-3 level regularly done deterministically
- Problematic, since migratory processes highly stochastic and especially hard to predict
- Lowers the predictive value of deterministic projections further
- Moreover, consistent concepts of modeling the connection between international and internal migration lacking
- Correlations between regional migration flows not always integrated in forecast models
- We tried to cover all of these aspects in our forecast model



#### **Data and Methods**

#### Data

- NUTS-3 data on annual immigration and emigration in Germany by age group (0-17; 18-24; 25-29; 30-49; 50-64; 65+) 1995-2019: Statistische Ämter des Bundes und der Länder
- NUTS-3 data on end-of-year population estimates by 17 age groups 1995-2019: Statistische Ämter des Bundes und der Länder
- Methods
  - First step: Comparison of competing models by backtesting for 2009-2019: SMAPE
  - Second step: stochastic forecast for 2020-2040 based on winning model



## **Data and Methods**

- Model 1. Naïve Migration Models
  - Model 1a. Naïve Random Walk Model
  - Model 1b. Naïve White Noise Model based on Mean
  - Model 1c. Naïve White Noise Model based on Median
- Model 2. Principal Component Model of Net Migration Flows
  - Long-term trend of first principal component
  - Model 2a. Naïve Random Walk Assumptions for p<sub>2</sub><sup>NM</sup>
  - Model 2b. AR(1) fit with mean asymptote for p<sub>2</sub><sup>NM</sup>
  - Model 2c. AR(1) fit with median asymptote for p<sub>2</sub><sup>NM</sup>
- Model 3. Joint Principal Component Model of Gross Migration Flows
- Model 4. Joint Principal Component Model of (Pseudo) Migration Rates
- Model 5. Naïve Random Walk Model of log-(Pseudo) Migration Rates



#### **Data and Methods**

- Model 5 minimizes SMAPE, therefore used for forecast
- Include stochasticity by assuming first differences of log-(pseudo) migration rate time series to follow multivariate Gaussian distribution with zero mean and empirical covariance matrix:

$$\begin{bmatrix} dlMR_{i,1,1} \\ dlMR_{i,2,1} \\ \vdots \\ dlMR_{e,6,396} \end{bmatrix} \sim \mathcal{N}(\vec{0}, \mathbf{\Sigma})$$

 1,000 random draws from that distribution annually add up to 1,000 trajectories of all (pseudo) migration rates, e.g., log-emigration rate of age group a from district d in trajectory t for year y:

$$lMR_{e,a,d,t,y} = lMR_{e,a,d,2019} + \sum_{\tau=2020}^{y} dlMR_{e,a,d,t,\tau}$$

, i.e.

$$MR_{e,a,d,t,y} = \exp(lMR_{e,a,d,t,y})$$



### **Data and methods**

- Results of simulations are highly dimensional, therefore present summary statistics
- Pseudo net migration rate of age group a, in district d, in year y, and in trajectory t as

$$n_{a,d,y,t} \coloneqq i_{a,d,y,t} - e_{a,d,y,t}$$

with

- $i_{a,d,y,t}$ : pseudo immigration rate of age group *a*, in district *d*, in year *y*, and in trajectory *t*
- *e*<sub>a,d,y,t</sub>: emigration rate with same parameters
- Assuming a stable baseline population for each district and age group, we define the migration-inflicted net growth rate between 2019 and 2040 as

$$g_{a,d,t} \coloneqq \prod_{y=20}^{y=40} (1 + n_{a,d,y,t})$$

• Probability of migration-inflicted depopulation of district *d* in age group *a* over forecast horizon approximated as the share of trajectories with  $g_{a,d,t} < 1$ 



#### **Results**

Figure 1. Probability of Migration-inflicted Depopulation between 2019 and 2040 by District and Age Group





#### **Results**

Figure 2. Median Migration-inflicted Net Growth between 2019 and 2040 by District and Age Group [as %]





#### **Results**

- Depopulation among children based on migration highly unlikely for the vast majority of German districts
  - Isolated effect of migration
  - Does not rule out depopulation in that age group due to low fertility
- Exceptions are larger, less family-friendly, cities such as Berlin
- => Higher tendency of young families to migrate to neighboring regions
- Depopulation in age group 18-24 (study age) for most districts, with large migration-inflicted growth among large university cities, e.g., Berlin, Göttingen, Hannover, Münster
- Depopulation among young labor force in economically weaker regions, such as central Germany and the regions at the north sea
- Less movement of among "older" workers
- Higher tendency of aging in economically weaker regions without strong labor markets



# **Discussion**

- We suggested a stochastic joint forecast model pseudo immigration and emigration rates by age group on the NUTS-3 level in Germany
- Model includes correlations in migration patterns between districts and autocorrelations in the time series
- Model consistently models internal and international migration
- Results show increasing heterogeneity in age structure of districts, with an aging of already economically weaker regions, accelerated by internal migration
- Data limits analysis to six age groups
- No consistent time series by age and gender available, therefore, no disaggregation by sex
- Model only measures effect directly connected to migration; other or connected effects, such as birth trends are not covered => "only the beginning of the story, not the end"
- Obviously, high risk in future international immigration, especially by unpredictable refugee trends => high impact on internal migration, as both are highly correlated

Seite 11 | Vanella - Parsimonious Stochastic Forecasting of International and Internal Migration - Wittgenstein Centre Conference 2021



#### Many Thanks! Questions or Remarks?

#### VANELLA, P.; HELLWAGNER, T.; DESCHERMEIER, P.: PARSIMONIOUS STOCHASTIC FORECASTING OF INTERNATIONAL AND INTERNAL MIGRATION ON THE NUTS-3 LEVEL – AN OUTLOOK OF REGIONAL DEPOPULATION TRENDS IN GERMANY [WORK IN PROGRESS]

Seite 12 | Vanella – Parsimonious Stochastic Forecasting of International and Internal Migration – Wittgenstein Centre Conference 2021

