DeZIM Institut

DeZIM Briefing Notes +

DBN Issue #04 Berlin, 19 August 2020

The Future of Migration to Germany: Assessing methods in migration forecasting

Dr. Sulin Sardoschau

The Future of Migration to Germany: Assessing methods in migration forecasting

Dr. Sulin Sardoschau

1. The demand for migration forecasts

In the midst of the Covid19 pandemic, forecasters in all fields become painfully aware of how difficult it is to predict and prepare for major events. In the realm of migration forecasting, the crisis demonstrates that these shocks have important implications for the international movement of people, certainly changing but potentially even reversing existing global migration trends and patterns.

Accelerated by the influx of refugees to Europe in 2015, politicians' and policymakers' interest in ways to predict future migration flows has been mounting. Various stakeholders including think tanks, international organizations and research institutions rushed to meet this demand. Consequently, there is a large pool of migration forecasts using very different qualitative and quantitative approaches and therefore come to vastly different results. This briefing note attempts to shed light on and assess the different quantitative migration forecasting methods and emphasize the importance of transparency for both consumers and producers of such forecasts.

This briefing note is structured as follows. First, four types of uncertainty in migration forecasting are discussed. Then, three major forecasting tools (Bayesian Statistical Modelling, Gravity Models, and Structural Equation Models) and their strengths and weaknesses are briefly introduced. Afterwards, the differences of these models and their according outcomes are illustrated by three examples for migration forecasts for Germany. The briefing note concludes with a call for a careful usage of forecasts by users and a call for transparency among producers of forecasts. For this purpose, a short users' guide is provided.

2. Sources of uncertainty in migration forecasting

Complexity of migration determinants

Migration determinants are highly complex. Behavioural, social, cultural, political, economic and many other factors are at play, interacting with one another in multifarious ways. So far, there is no unified migration theory across disciplines and different fields of research naturally focus on different aspects of migration. Economists, for instance, have been focusing on the economic returns to migration in a dual labor market framework, others have paid more attention to migration as a phenomenon arising from complex social systems or migration as a result of socio-psychological mechanisms. Forecasts therefore depart from very different theoretical foundations on the macro-, meso- and microlevel and accordingly produce very different results.

Implicit assumptions

In order to deal with this complexity, the quantitative models used for forecasts have to rely on simplifying assumptions. Some models simplify to the extent that the only determinant of future migration is past migration, while other models include multiple determinants (e.g. international networks, migration policies, climate change). The inclusion or omission of certain variables is itself an assumption about what is expected to influence migration and what is not. For example, the estimated outcomes of migration is forecasts differ depending on whether and to what extent the model uses climate change as a migration-determinant. This means, the outcomes are highly sensitive to small changes in these assumptions. In addition, forecasts can be based on other forecasts (e.g. demographic change or using the same example, the severity of climate change in the future), which incorporate various assumptions themselves.

Insufficient data

Since migration forecasts rely heavily on data on past migration, the insufficient quality of migration data (low frequency, low geographic resolution, low accuracy) introduces more uncertainty into forecasting. The three main types of data sources are administrative data, survey data and big data, which all come with their specific drawbacks. Administrative data (e.g. in the form of tax records or visa permits) has the potential to cover the whole population over time, but it may well be that individuals are counted twice or not at all and oftentimes they do not include socio-economic characteristics of the individuals. Survey data provides more detailed information on individuals (e.g. socio-demographic and economic information, but also migration-specific information like type of migration status, reasons for migrating etc.) and groups of interest can be targeted better. However, most surveys suffer from typical survey biases and are limited in sample size, which restricts reliable statements about sub-groups. The usage of big data is a way to overcome the problem of insufficient administrative data in certain countries and usually a large coverage can be achieved. Nevertheless, social-media usage is highly self-selective and the potential inaccuracy of IP-addresses and phone locations foster opacity about what is actually measured.

The lack of information exchange within and across countries is another impediment to comparable und detailed migration statistics. The most extensive attempt was the construction of bilateral migration matrices between 1960 and 2000 by the World Bank (Özden et al. 2011); a data set that is used exhaustively in migration economics. Despite these large scaled efforts to harmonize census and population registers, there remain large gaps in terms of migration frequency (in this case only every 10 years) and difficulties to harmonize data for countries with lower state capacity and data maintenance capabilities.

Future shocks

The strongest impediment to accurate migration forecasts is the inherent inability to foresee or predict important events or major shifts in economics, politics, technology, climate or other major drivers of migration. Since these changes also interact with one another in complex ways and migration policy responses may in turn respond to these changes, it is very difficult to predict their consequences in a credible way. Quantitative models are in a statistical way not necessarily limited in the timespan they can cover and could therefore predict very far into the future. However, over longer time horizons, the uncertainty of these forecasts increase substantially and the range of possible outcomes becomes so large that it is difficult to make any dependable claims on the future of migration. In principle, all levels of migration become possible if the time frame is just long enough. Additionally, it is difficult for data driven methods to distinguish between a unique event (such as the shutdown of international travel and borders in the Covid19 pandemic) and a change in trends of migration patterns.

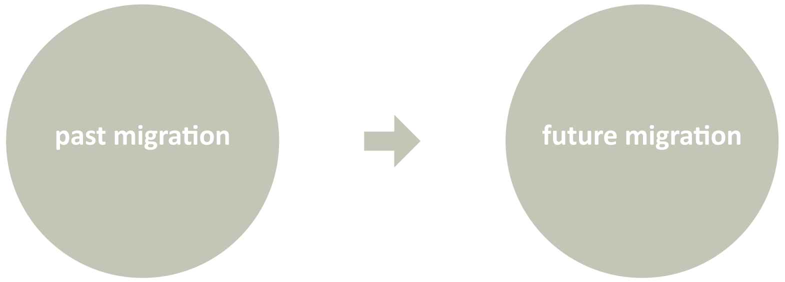
The so-called "refugee crisis" in 2015 in Germany serves as an illustrative example that future shocks cannot be foreseen well in advance. Bijak (2016) used pre 2015 data to predict the future migration to Germany and although the model gives space to uncertainty, the actual migrant inflow of 2015 lies well outside the indicated range of uncertainty. Azose and Raftery (2015) further demonstrate the sensitivity of quantitative models to these shocks. They show that the same statistical model produces substantially different results depending on weather the dataset includes the 2015 influx or not. Accordingly, their estimate for the median net migration to Germany in 2100 roughly doubles (from 400,000 to almost 900,000).

3. Quantitative forecasting methods

Bayesian statistical modelling

In Bayesian models the only influencing factor of future migration is past migration. Hence, this method is considered as a purely data driven approach and does not make any claims about the determining factors of migration There is an important trade-off between a strong theoretical foundation and low data requirements. Bayesian models do not need data other than past migration and therefore do not have to rely on assumptions about the future. But this also means that important migration-drivers cannot be included in the analysis.

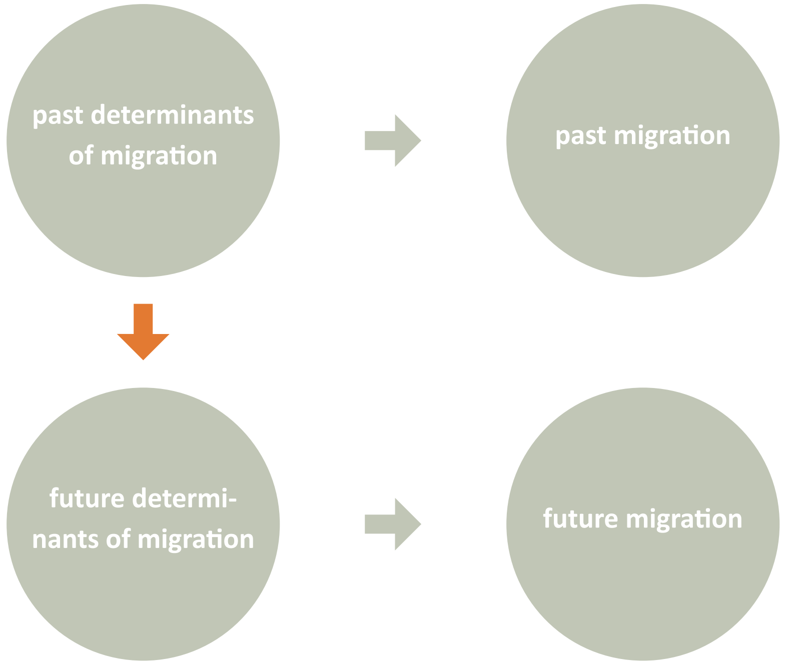
Figure 1



Gravity models[1]

Gravity models focus on the determinants of migration. They can include, for instance, environmental, political, sociological, micro/macro-economic, geographical data. On the one hand, this means that more nuanced hypotheses and theories on what drives migration can be tested. On the other hand, the data requirement increases and strong assumptions about the future have to be made. In principle, gravity models make assumptions about what are important factors that determined migration in the past. They take these lessons and make assumptions about how these determinants develop in the future to make predictions on the change in migration patterns in the future.

Figure 2

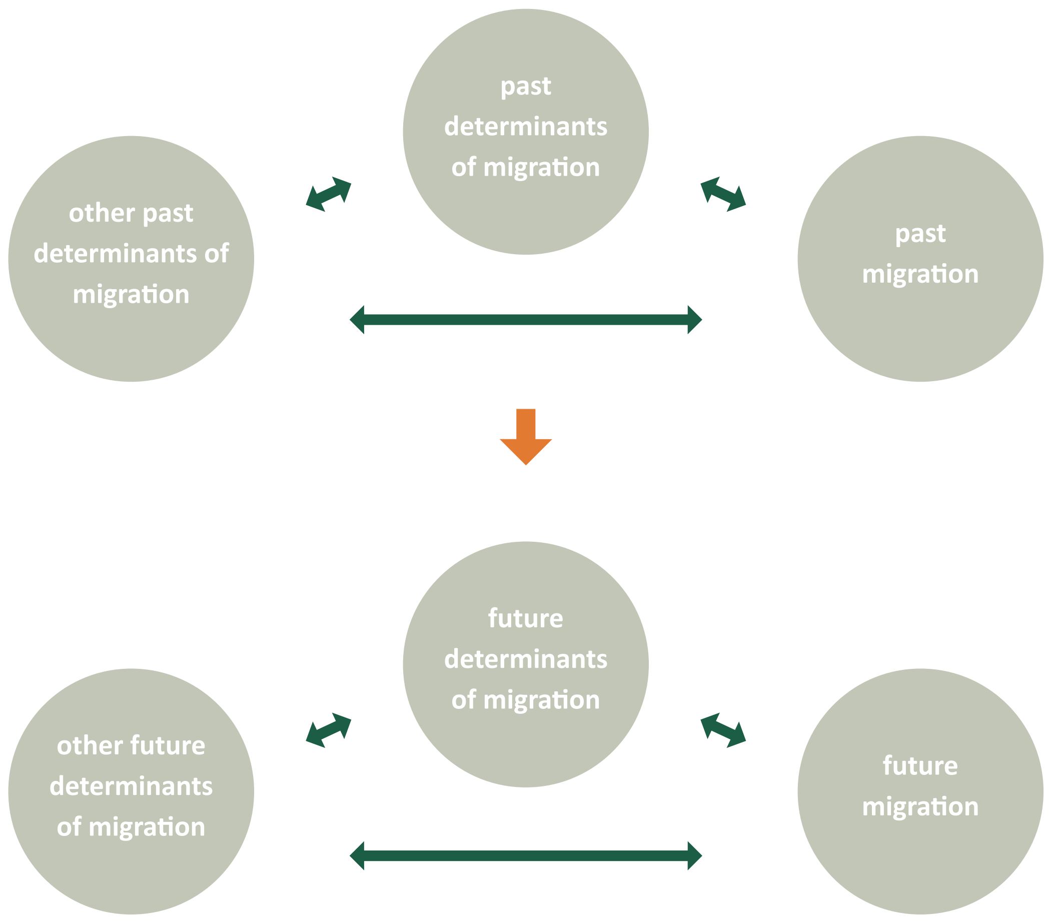


Structural equation models (SEMs)

While gravity models have one model in mind about how migration is impacted, SEMs use multiple hypotheses about different (interactional) relationships between variables and bind them together in a complex mathematical framework. SEMs are theory driven empirics relying on a theoretical model that determines how and whether certain variables are related to one another. These models can incorporate not only unidirectional, but also reverse relationships between variables. SEMs try to get at the causal mechanism between variables and are useful in analysing complex systems with many interdependencies. But SEMs require many observations to increase the precision of the estimates – even more than gravity models. Consequently, both the estimation method and the underlying theory of structural models require a substantial amount of assumptions and data, which can in turn introduce various biases and inaccuracies.

[1] Gravity models follow the basic mechanism that describes gravity as a trade-off between size and distance, which is re-interpreted in migration economics as pull factors (size, for instance, GDP) and migration costs (distance, for instance, geographic distance or language barriers). The scope of gravity models is to build a framework, which can represent the migration decision of a 'representative migrant'.

Figure 3



Strengths and weaknesses of quantitative methods

Table 1 provides an assessment of Structural Equation Models, Gravity Models and Bayesian Models along four dimensions: theoretical foundation, transparency of assumptions, data requirements and predictiveness of the model.

- **Theoretical Foundation:** this dimension assesses in how far the model makes explicit through which channels future migration will be affected and how different factors interact with one another. A strong theoretical foundation requires a guiding theory about migration and its functioning.
- **Transparency of assumptions:** this dimension assesses how the guiding theory is translated into a quantitative estimation of future migration flows. A high level of transparency presents the underlying assumptions of the model in an open and comprehensive manner.
- **Data requirements:** this dimension assesses the scope and level of granularity required for the estimation strategy. High data requirements can pose a hurdle to a precise forecast of future migration flows, as only high volumes of data allow for decreasing errors and confidence intervals.
- **Predictiveness:** this dimension assesses whether the model is predictive, explanatory or descriptive in the design. High predictiveness models include time series models which are designed to extrapolate into the future rather than describe or explain current or past migration.

Table 1 Strengths and Weaknesses of Quantitative Migration Forecasting Models

Model Type	Theoretical Foundation	Transparency of Assumptions	Data Requirements	Predictiveness
Structural Models	strong	high	high	medium
Gravity Models	medium	medium	medium	low
Bayesian Models	weak	medium	low	high

4. Migration forecasts for Germany

In order to make transparent how uncertainties and methodological differences materialize in migration forecasts, Table 2 presents Germany-specific forecasts from the three described methods[2]. We use the three main reference papers for each forecasting method in the field of demography and economics, that is Azose et al. (2016), Hanson and McIntosh (2016) and Burzynski et al. (2019). The table also reports the main data bases used for the estimation as well as the proposed mechanism behind emerging migration patterns. Additionally, the table indicates whether the unexpected influx of refugees between 2013 and 2016 was included in the data used to make to forecast. The column "prediction" shows the estimated net migration flow to Germany between 2020 and 2040.

The range of predicted outcomes across the three models lies at roughly 7 million. If we compared the 20-year span between 2020 and 2040 for all three papers, Azose et al. (2016) predict a 5.8 million net migration flow, Hanson and McIntosh (2016) predict approximately -1.5 million (a net decrease in migration flows) and Burzynski et al. (2019) predict 1.5 million. All of the estimations reveal largely diverging patterns for the next two decades.

Table 2

study	method	database	mechanism	2015 influx[3]	prediction[4]
Azose, Sevcikova & Raftery (2016)	Bayesian Model	UN World Population Prospects 2019	Future migration depends on past migration and long-term migration	yes	+5.8 mio.
Hanson & McIntosh (2016)	Gravity Model	UN World Population Prospects 2017, DIOC data base & IMF GDP forecast of 2018	Migration due to differences in labor supply, resulting from changes in fertility	no	-1.5 mio.
Burzynski, Deuster & Docquier (2019)	Structural Equation Model	UN World Population Prospects 2019; WDI data, DIOC data base	Migration due to differences in wages, consumption and schooling cost	no	+2.5 mio.

2 We thank the authors for extracting Germany specific forecast from their models.
3 Data used include refugee influx to Germany in 2015
4 Net migration flow to Germany between 2020 and 2040

5. Ways forward

As the demand for and the supply of migration forecasts has increased over the last years, policymakers have to navigate an ever-increasing maze of methods and predictions. A comparative analysis of different methods becomes more difficult and forecasters can help to increase transparency, since diverging predictions may cause confusion.

In order to contextualise migration forecasts, researchers and experts can preface their analyses with a sheet that makes the analysis more transparent and can serve as a user's guide. The guide could cover seven dimensions (see box below), which are necessary to understand, interpret and compare the respective model: model type, theory and assumptions, determinants and mechanisms, data, time horizon and frequency, prediction and uncertainty, scenarios and sensitivity. Also, different scenarios should be provided and the quantitative prediction should be re-examined using different assumptions, different data, time frames or definitions.

As for the users of forecasts, it is important to interpret forecasts with caution and become familiar with the fundamental structure of the models and the underlying concepts of the estimates in order to understand any specific number that the forecasting method produces. The user of these forecasts should consider whether theory-based models or purely data-driven models are more appropriate. If theory based models are more attractive, then it should be investigated whether the model assumptions, the determining variables used and the mechanisms are convincing. Attention should be drawn to how changes in assumptions result in changes in predictions. Additionally, uncertainty and different scenarios should be considered, knowing that the forecast is more likely to provide inaccurate rather than accurate results. Lastly, users should interpret these forecasts with an eye on the policies or strategies that will be informed by these forecasts; are they compatible with the large uncertainties involved in making these predictions?

Migration forecasts have become and will remain a main staple of basic migration research and new data and statistical tools promise great improvements in forecasting in the future. For the moment, however, they should be regarded as a window to understanding the overarching concepts and trends, dynamics and mechanisms of migration, rather than a window to the future.

Guiding questions for users and producers of migration forecasts

- **Model type:** to which family of forecasting models does your approach belong to?
- **Theory and assumptions:** what are the theoretical foundations of the model that affect future migration? Which are the assumptions underlying the method (statistical assumptions) and which are the assumptions introduced by the researchers (theory assumptions)?
- **Determinants and Mechanisms:** what are the main determining variables used in the model and what are the mechanisms through which the determinants affect future migration?
- **Data:** which data sources are being consulted for all variables included in the model?
- **Time horizon and Frequency:** what is the time horizon of the forecast and why was a specific time span and were specific time intervals chosen for the forecast?
- **Predictions and Uncertainty:** what is the estimated stock or flow of migrants and how large are the uncertainties? If forecasts are based on other forecasts, how are the respective uncertainties incorporated?
- **Scenarios and sensitivity:** can you provide forecasts for different scenarios? How does the model react to tweaking assumptions and theory? How does the model react to different use of data? How does the model compare to other forecasts and why do they differ?

6. Summary

This brief assesses the strengths and weaknesses of quantitative migration forecasting models, highlighting the various sources of uncertainty and presenting the three main quantitative approaches used in demography and economics. In collaboration with leading researchers in the field, Germany-specific forecasts are presented and compared. Depending on the underlying forecasting method, estimations on the net migration flows to Germany between 2020 and 2040 vary from -1.5 Million to + 5.8 million, uncovering substantial heterogeneity in migration forecasts for Germany. The brief concludes with transparency guidelines for producers and consumers of migration forecasts.

LITERATURE

Aazose, Jonathan J. and Adrian E. Raftery, "Bayesian Probabilistic Projection of International Migration", Demography, 52 (2015), 1627–1650.

Azose, Jonathan J., Aana Ševčíková, and Adrian e. Raftery, "Probabilistic population projections with migration uncertainty", Proceedings of the National Academy of Sciences of the United States of America, 113 (2016), 6460–6465.

Bijak, Jakub,"Migration forecasting: Beyond the limits of uncertainty," Data Briefing Series IOM GMDC ISSN 2415-1653 (2016).

Burzynski, Michal, Christoph Deuster, and Frédéric Docquier, "Geography of skills and global inequality", Journal of Development Economics (2019).

Hanson, Gordon and Craig McIntosh, "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run", Journal of Economic Perspectives, 30 (2016), 57–82.

Özden Çağlar, Christopher R. Parsons, Maurice Schiff, and Terrie L. Walmsley, "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960–2000," The World Bank Economic Review, 25 (2011), 12–56.

THE AUTHOR

Dr. Sulin Sardoschau is an Assistant Professor in economics at the Humboldt University Berlin. Her research areas include applied econometrics, migration economics, political economy and development. She was an economist at the German Center for Integration and Migration Research ((DeZIM-Institute), after obtaining her PhD from the Paris School of Economics in 2018.

This Briefing Note was produced with excellent research assistance by **Lena Mann**.

Published by

DeZIM | Institut

Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM-Institut)

German Center for Integration and Migration Research (DeZIM-Institute)

Mauerstraße 76
10117 Berlin
+49 (0)30 804 928 93
info@dezim-institut.de
www.dezim-institut.de

Directors
Prof. Dr. Naika Foroutan, Prof. Dr. Frank Kalter

Author
Dr. Sulin Sardoschau

Layout
neonfisch.de

ISBN
978-3-948289-05-8

The DeZIM-Institute is a research institution funded by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth. Its central mission is continuous, methodically sound research that transfers into politics, the public and civil society. Alongside the DeZIM research community, it is one of two pillars of the German Center for Integration and Migration Research (DeZIM).

funded by

Bundesministerium für Familie, Senioren, Frauen und Jugend