International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria

Email: permissions@iiasa.ac.at

Young Scientists Summer Program

Bilateral international migration measurement and forecast: an agent-based model

Author: Xinyi Kou
Email: Xinyi.kou@postgrad.manchester.ac.uk

Approved by:

Mentor(s): Guillaume Marois, Dilek Yildiz
Program: Population and Just Societies Program (POPJUS)
Date: September 30, 2022

This report represents the work completed by the author during the IIASA Young Scientists Summer Program (YSSP) with approval from the YSSP mentor.

It was finished by September 30th 2022 and has not been altered or revised since.

Mentor signature:

Table of contents

ZVR 524808900

Abstract

The study of population dynamics has always been at the centre of public policy and planning due to its vital role in human society. Because of the low fertility and mortality in Europe, international migration is becoming an increasingly important factor in shaping population structures. As the most challenging demographic component, international migration studies suffer from the uncertainty at both macro and micro levels. To address the migration uncertainty, this research extends the existing agent-based modelling approach to study the dynamics of international migration flows and its underlying migration mechanisms between three countries, UK, Poland, and Germany from 2002 to 2030. The decision rules are derived from the theory of planned behaviour (TPB), which assumes that the migration intentions are hinged on migration attitude, social networks about migration and perceived behaviour control over migration. Calibrated model parameters are determined by MSE (mean squared error) approach. The resulting parameters are therefore used for migration forecasting until 2030. The model is validated to the existing literatures and calibrated based on two sets of empirical data. This research finds that migration decisions are jointly determined by a combination of migration factors, where the social networks and employment-related migration drivers have a significant effect on final migration decisions. The model shows its ability on studying migration dynamics.

About the author

Example:
Xinyi Kou is PhD candidate at the Department of Social Statistics, University of Manchester, Manchester, UK. (Contact: xinyi.kou@postgrad.manchester.ac.uk)

Acknowledgments

The research was developed in the Young Scientists Summer Program at the International Institute for Applied Systems Analysis, Laxenburg (Austria) with financial support from the Economic and Social Research Council (ESRC) through the North West Social Science Doctoral Training Partnership (NWSSDTP). In particular, thanks to Guillaume Marois and Dilek Yildiz for their overall supervision on the research during my stay, and Michaela Potančoková for her useful feedback.

Bilateral international migration measurement and forecast: an agent-based model

1 Introduction

People's movement, interactions and behaviours have major and important impacts on the society and environment that they are living in. At the same time, changes in these factors also lead to an evolution of the population itself over time. To facilitate strategic decision making and plan developments for the future, it is vital to study and understand changes in the population. Among all components of population change, international migration is the most challenging element to estimate and forecast. It plays a key role in population change and population geographical redistribution, which has impact on the societies, economics and policy response.

International migration studies suffer from uncertainties because of the weak theoretical background of migration measurement and the unpredictable shocks of migration processes (Arango, 2002; McAuliffe & Koser, 2017). The uncertainty increases rapidly as the forecast horizon moves into the future. To better understand the migration uncertainties, Bijak and Czaika (2020) offered a unique typology to study the migration features. They categorized the migration uncertainty into two groups, namely, *epistemic uncertainty* and *aleatory uncertainty*, across four domains, namely, migration drivers, data measurement, analytical methods, and individual decisions.

Epistemic uncertainty stems from the imperfect knowledge about migration. It represents the ignorance about the migration process, which indicates that this type of uncertainty can be obtained if migration process is addressed properly in the research. On the other hand, aleatory uncertainty refers to inherent uncertainty due to the probabilistic variability or other types of randomness (cf Bijak and Czaika 2020). It is fundamentally irreducible due to the unpredictable shocks and makes the migration process 'unknowable' in principle. To address both types of migration uncertainty, a possible solution is to study the possible and approximate theoretical explanations of the underlying migration process through a micro-level approach with scenario-based forecast analysis. Consequently, a micro-level approach, namely, agent-based model is deemed as an appropriate method and proposed for this research to enhance the theoretical micro-foundations in migration studies.

Recently, Klabunde and Willekens (2016) reviewed theoretical frameworks for agent-based models that govern the decision-making rules in migration studies. They concluded two popular behavioural theories in migration studies, which are 1) theory of planned behaviour (TPB) which is derived from psychosocial and cognitive models, and 2) random utility theory which is the decision rule of microeconomic models and experience- or preference-based model. The random utility theory is not only the theoretical foundation of the well-known random utility maximization (RUM) model, but the widely used gravity models which are the major and popular technique to explain and forecast migration flows among migration literatures in the past decades (Ramos 2016, Khan and Fatima 2022). However, the RUM models rely on the rational decision-making rules which are unrealistic in the real-world situation. Another unrealistic assumption is that the model assumes the attractiveness of a destination is not affected by migrations.

Therefore, in this multi-country agent-based model, an extended version of the theory of planned behaviour (TPB) will be utilized as the main theoretical framework. TPB has advantages as a decision-making framework, although Bijak et al. (2021) argued that there is no clear justification for the decision rules in the agent-based model, and TPB suffers for being arbitrary due to its inclusion of infinite number of decisive factors. TPB has the ability to model the features that is relevant to the migration decisions, it allows for modelling the distinction between desired and actual migration behaviour, the social influence, the incorporation of other life events and the uncertainty.

The main contribution of this research is the methodological innovation of agent-based modelling approach in international migration studies with an emphasis on multi-country applications and accommodates the model to a less migration data-driven setting. The model aims to study the international migrations between UK, Poland and Germany from 2002 to 2015 and forecast the migration flows to 2030.

2 Theoretical models and modelling framework

The framework of TPB was originally proposed by Ajzen (1985), which is an extension of theory of reasoned action (TRA). This psychological theory maintains that a particular behaviour action is explained by the individual's behavioural intentions, which consists of three core components, namely, attitude, subjective norms and perceived behavioural control. The schematic representation of TPB within the context of migration is presented in Figure 2.1.

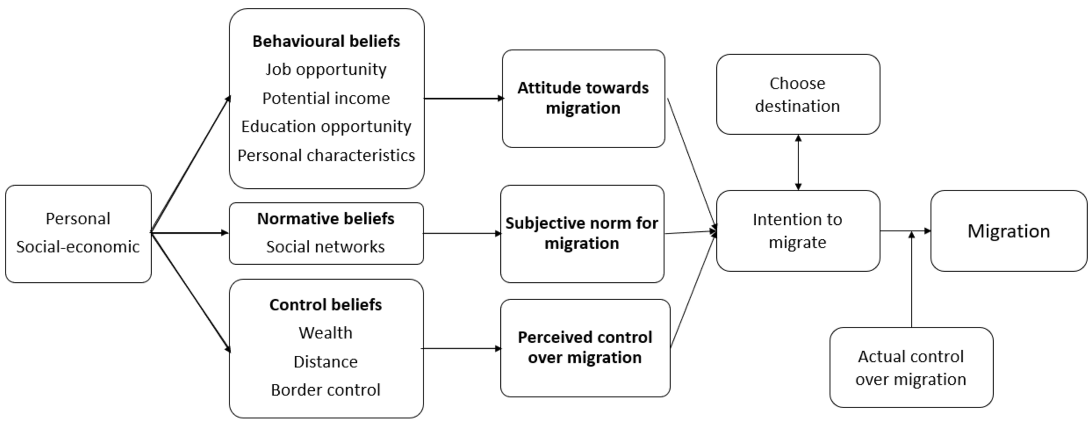

Figure 2.1 The schematic representation of migration decision-making process (source: own elaboration based on Fishbein and Ajzen 2011 and Willekens 2017)

As one of the most popular theory frameworks in agent-based model, the TPB framework has already been applied in agent-based migration models. For example, Kniveton et al. (2011, 2012) utilised the TPB to model the climate change-driven migration in Burkina Faso, where the attitude towards migration was defined solely on the personal background. A subjective norm function was formulated to assign values for social norms and the perceived behavioural control was determined by a person's asset and migration experience. Smith (2014) studied the rainfall-induced migration in Tanzania through a similar approach, where the migration intention was driven by the household's financial ability. Willekens (2017) and Klabende et al. (2017) proposed a multistage stochastics model to model the international migration behaviour with an application on migrations between Senegal to Europe. Their model considered the TPB framework as a process theory, where an actual migration behaviour was controlled by not only intension but also the planning and preparation stage. The duration time at each process was drawn from the common waiting time distribution, namely, exponential distribution. More detailed review can be found in Chapter 2.

In this section, we propose a unique TPB framework for bilateral international migration flows in multiple countries. TPB has the ability to include various components that is relevant to the individual's behaviour, especially it models the difference between desired and actual behaviour, the social influence as well as the uncertainty (Klabunde & Willekens 2016). This is accomplished by its three types of beliefs, which are behavioural beliefs, normative beliefs and control beliefs. In the context of migration, these beliefs are defined as attitude toward migration, subjective norm for migration and perceived behavioural control over migration. These three components together determine the migration intention, and they are influenced by different background factors, including demographic characteristics (e.g., age, gender, marital status, employment status and education status), social-economic factors (e.g., the employment opportunity, education, poverty index and potential income) and environmental factors (e.g., climate change, natural disaster). These background factors indirectly shape the migration intention and actions (Ajzen 1985, Fishbein and Ajzen 2011).

First, the behavioural beliefs, which governs the attitude towards migration, are considered as the reasons of migration behaviour. Czaika and Reinprecht (2020) found there are at least 24 combinations of different migration drivers under the migration process, which are impossible to model all drivers in one single model. Therefore, a number of relative important driving factors of migrations are selected to be involved in the model, in line with the Ockham's razor argument. According to the Castles et al. 2013, there are four major groups of factors determining migrations, which are 1) economic motivation (including seeking employment and reducing poverty); 2) education opportunities; 3) family reunion; and 4) fleeing persecution. In the context of international migration between European countries, the economic and education reasons compose the majority of the migration flows. Therefore, the migration flows due to family reunion and refugee are not considered as the main reason of migration in this model. Despite, the effect of family-driven migrations is addressed by the social networks to some extends. Although it may lose some accuracy of model prediction, it provides the opportunity to clearly explain the pattern of international migration. This behavioural belief is then weighted by the subjective values, the values can be either empirically derived from a dataset or assumed based on the expert opinion. The unique combination of these values defines the agent heterogeneity.

The second belief is the normative belief, which governs subjective norm for migration. It measures the social pressure of their peers. This subject provides the opportunity to model the social networks explicitly. The supports from peers aboard and their destination preferences can positively affect a person's migration choice (Kniveton et al. 2011). In the meantime, the family ties at origin country prevent a person's migration action. The normative belief is weighted by a subjective value as well. Different from the other two components, the value is assigned to each agent because the peer support or peer pressure is distinct across agents.

The third one is the control belief that governs the perceived behavioural control (PBC). This is the final component of the decision-making process regarding migration intensions and actions. This is also the major development of the theory of planned behaviour (TPB) from the theory of reasoned action (TRA). The perceived behavioural control captures the ability to make an actual migration action. For example, a potential migrant might consider not to migrate if they do not have enough savings. The long distance may cause more costs on transportation, and thereafter, prevents the behaviour of an individual's migration. Therefore, the PBC takes all the barriers and obstacles related to the migration behaviour into consideration. If an individual can overcome these barriers, the migration probability increases. The PBC is also weighted by a subjective value.

These three beliefs are indicators of the final migration intensions. The stronger the three components are, the more likely the individual undertakes the migration action. It needs to be noted that the TPB assumes that behavioural intention is the most proximal determinant of human social behaviour. To improve the prediction accuracy, another module layer, the actual behavioural control is introduced by Ajzen (1991). Normally, the perceived behavioural control is used to measure the actual behavioural control (Ajzen 1991, Fishbein and Ajzen 2011), and described in a format of probability. Therefore, a subjectively probability is considered and developed as the actual behavioural control, and it measures whether an individual is actually going to migrate.

3 Model description

In this section, a conceptually dynamic ABM is developed to demonstrate the dynamic of migration flows between countries.

In this model, agents represent persons. Each country stores information about their socio-economic factors, and each person stores information about their attributes and wealth. Each person is allocated a resident country. Persons have the ability to move between countries. Each person makes their migration decisions in two stages. The first one is to assess the migration intentions towards different countries, which is described in section 2. The second stage is the development of actual migration behaviour and destination selection based on the calculated intention scores.

This section starts with the assessment of international migration intentions. As described in section 2, the intention of a person to migrate, hinges on their attitude towards migration (behavioural beliefs, A), subjective norm for migration (normative beliefs, SN) and perceived behavioural control over migration (control beliefs, PBC). All the components are developing according to the time t. Therefore, the calculation of intention scores for each person is based on the following equation:

$$I_{i,j}\ (t) = a * A_{i,j}(t) + b * SN_{i,j}(t) + c * PBC_{i,j}(t) \tag{5.1}$$

where $I_{i,j}(t)$ denotes the migration intention for agent i towards country j. And a, b, c are three parameters for model calibration and fall in the range $[0,1]$.

This simple linear regression of migration intention is developed in line with the assumption of TPB. In the theory of planned behaviour, the attitude, social norm, perceived behavioural control components are independent. Their interaction effects are not a formal part of the TPB (Ajzen 2002; Fishbein and Ajzen 2011).

3.1 Behavioural beliefs: attitude towards migration $\boldsymbol{A_{i,j}(t)}$

The first component in equation (5.1) is the attitude towards migration $A_{i,j}(t)$. In this model, the attitude captures the possible reasons of migration. It is assumed that people migrate for a higher income, a better new life and more job and education opportunities. The personal background, e.g., the employment status, may affect their decisions to migrate for a new job (Kniveton et al. 2011). And persons who are younger and well-educated may have more positive attitude towards migration than older less-educated persons. Except for the personal background, the socio-economic factors in the potential destination country affect the migration decisions in a macro level as well. Therefore, the attitude towards migration is computed by the combination of socio-economic factors in the potential destination countries and the individuals' background:

$$A_{i,j}(t) = \alpha_1 emp_j(t) + \alpha_2 inc_{ij}(t) + \alpha_3 edu_j(t) + \alpha_4 ind_i(t) \tag{5.2}$$

Here, emp_j is the employment opportunity in country j, edu_j is the education opportunity in country j, inc_{ij} is the potential income for agent i in country j, and ind_i is the attributes of agent. The agent's attributes include age, sex, education level and employment status.

The employment opportunity is indicated by the combination of FDI (foreign direct invest) companies (Radosevic et al. 2003, Craigwell 2006, Sanderson and Kentor 2008, Saucedo et al. 2020) and the percentage of employed workers in the potential destination:

$$emp_j(t) = FDI_j(t) + emprate_j(t) \tag{5.3}$$

And the education opportunity is indicated by the number of top 100 universities in the potential destination. All values of $emp_j, inc_{ij}, edu_j, ind_i$ are normalised in the range of $[0,1]$. The parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in [0,1]$ are weight parameters of the relevant socio-economic and personal factors. As there is no previous study exploring the relative values of these factors affecting the migration decisions in specific country context, the α are free parameters and determined by calibration.

3.2 Normative beliefs: subjective norm for migration $\boldsymbol{SN_{i,j}(t)}$

The subjective norm, $SN_{i,j}(t)$, is the social networks of agent i in country j. It plays a vital role in migration decisions (Massey et al. 1993, Epstein 2008, Simon 2019). First of all, social networks can reduce the risk and cost of migration (Massey et al. 1993, Simon 2019). The peers living abroad can provide agents with employment and other kinds of information and help with the accommodations and food when they arrive. Also, when a certain number of migrants is reached, the networks can overshadow some macro-level conditions, e.g., the employment rates and immigration policy, and have a large effect on migration flows (Massey et al. 1993). In this model, the SN is computed as the proportion of agent's peers who are living abord (Klabunde et al. 2017, Simon 2019):

$$SN_{ij}(t) = \beta_i \frac{Net_{ij}(t)}{Net_i(t)} \tag{5.4}$$

where $Net_{ij}(t)$ is the social links of agent i who are living in country j, and $Net_i(t)$ is total social links of agent i. The people living aboard help increase the migration intention of agents. On the other hand, the family ties at origin country prevent agent's migration intention. Parameter β is the weight of the peer support and used as a calibration parameter. In the real application, the social networks are constructed in a static manner. To account for the dynamic effect of social networks on agents, such as marriage or family reunion, each person is assigned a unique value β_i over time for their social norms. The value of β_i is derived from a uniform distribution between 0 and β.

3.3 Control beliefs: perceived behavioural control over migration $\boldsymbol{PBC_{i,j}(t)}$

There are barriers and difficulties before a real migration. The perceived behavioural control over migration, $PBC_{i,j}(t)$, evaluates whether the agent has the ability to overcome the barriers. In this model, it assumes that migration is prevented by the lack of money, the distance between countries and the border enforcement. If the agent cannot afford the migration cost, the agent will not migrate. If the agent can afford the migration cost, a higher wealth can help overcome the cost issue and increase the migration intention, but also can discourage an agent to migrate because an agent cannot gain benefit from the migration action. Therefore, the model assumes an exponential format for the wealth, where the increasing speed of migration intentions decreases when the wealth increases. Another barrier is the distance between countries. A longer distance indicates a higher transportation and psychological cost (Massey et al. 1993) and a lack of information, and thereafter a decrease of migration intention. The last barrier is the border enforcement. This component reflects the visa application restriction and the immigration policy of that country. A higher intense of border control discourages agent from migrating. Therefore, the $PBC_{ij}(t)$ is computed as the following function:

$$PBC_{ij}(t) = \frac{\gamma_1[1 - \exp(-\frac{S_{ij}(t)}{wealth_{ij}(t)})]}{(1 + \gamma_2 dis_{ij})(1 + \gamma_3 border_{ij}(t))} \tag{5.5}$$

Here, the $S_{ij}(t)$ is the saving money after agent i migrates:

$$S_{ij}(t) = wealth_{ij}(t) - cost_{ij}(t) \tag{5.6}$$

where $wealth_{ij}(t)$ is the accumulated wealth of agent i, and $cost_{ij}(t)$ is the migration cost of agent i towards country j. The accumulated wealth is calculated as:

$$wealth_{ij}(t) = wealth_{ij}(t-1) + inc_{ij}(t) - cons_{ij}(t) \tag{5.7}$$

where inc_{ij} is the actual income of agent i, and $cons_{ij}$ is the consumption of agent i.
In formula 5.3, dis_{ij} is the normalised geographic distance between country of agent i and country j, and $border_{ij}(t)$ is the normalised intense of border control. The parameters $\gamma_1, \gamma_2, \gamma_3 \in [0,1]$ are the relative weight of wealth, distance and border control factors that affect the migration intentions and used for model calibration.

Till now, the first step of migration decision process, the migration intention assessment has been finished. To summarize, the following assumptions are derived from the TPB framework to define the agents' migration rules:

a. Agents migrate with the goal to find a job, earn a foreign and higher wage, and pursue a better education. The attributes of agents affect agents' migration decision as well. Other types of migration, such as refugees, are not considered.

b. The social links at origin country and aboard jointly affect agents' migration choice. The family ties at origin country prevent an agent's migration behaviour. While the peer supports from aboard prompt an agent's migration behaviour and affect their destination choice. Networks are seen as the main source of information transfer.
c. Agents' migration decisions are affected by the actual barriers. Three barriers are assumed to discourage the agent from migrating. The first one is the money problem. A lower wealth cannot afford the migration cost while a higher wealth cannot gain benefit from migration. The second barrier is the distance, a longer distance indicates higher cost and thereafter prevents the migration. The final one is the border control. A higher border enforcement intensity discourages an agent from migrating as well.

A summary of parameter settings is shown in Table 3.1.

3.4 Decision-making process

After the calculation of three major components of migration decision-making process. The second step of decision-making process is the development of actual migration behaviour.

During the simulation, all three components in equation (5.1) are updated simultaneously to compute a migration score between 0 and 1 for each person. Each person calculates their own migration intention scores (probabilities) towards the specific destination and stores the score information at each time step. The sum of migration intention scores (including home country) for each person equals 1.

Consequently, the agents select potential destination country with higher intention scores (probabilities) to migrate. The agent chooses to migrate to the country with highest migration intension score at first place. The highest intention score is compared with a random number draw from the uniform distribution between 0 and 1. If the generated random number is less that the highest intention score, the agent decides to migrate the corresponding destination country. If not, the second highest intention score is considered. This country selection process prevents agents from only taking one country with highest migration intention as their destination and creating a biased migration flow.

In line with Cai and Oppenheimer (2013), the model assumes that the proportion of people with migration intentions are the same proportion of people who are migrating. Therefore, the entire population are at risk of migrating out. If a person decides to migrate and migrate eventually, this person becomes the population at risk in the destination country and their wealth decreases because of the cost of migration. The persons who migrate adapt new information about new resident country and notify their peers (links) about their migration decision, the social networks of migrant peers are updated by adding 1 migrant. All the agents move simultaneously within one time step. The model is designed to record migrants and migration flows separately, where migrants denote the number of persons residing in a country different from the home country, and migration flows denote the number of migration event happens within the time step.

Table 3.1 Parameter settings of TPB

Empirical parameters (fixed)	Number of FDI companies fdi_j Percentage of employed workers $emprate_j$ Number of top 400 universities edu_j

	Potential income inc_{ij} Attributes of individuals ind_i Wealth $wealth_{ij}$ Consumption $cons_{ij}$ Migration cost $cost_{ij}$ Distance dis_{ij} Border control $border_{ij}$
Calibration parameters	a; b; c; α_1; α_2; α_3; α_4; β; γ_1; γ_2; γ_3

4 Study Area

Due to the availability of international migration data between European countries, there is a better understanding and a more clarity of international migration patterns in the EU context (before Brexit). Therefore, this model is applied to study the international migration patterns between Poland, UK, and Germany.

Due to the enlargements of European Union, there was a great amount of out migration flows from Poland into other EU countries (Strey et al. 2018). According to Eurostat (2017a, 2017b), Poland sent 183,561 emigration flows (by country of previous residence) into other EU countries in 2015. As one of the first countries that welcomed these immigrants, UK held 86,770 inflows from Poland (based on nationality and previous residence) in 2015, where Poland is the second largest sender country within EU (Eurostat, 2017a; 2017b). Germany, as a country that has shared border with Poland, held 190,800 inflows (based on nationality) from Poland in 2015, where Poland is the second largest sender country within EU (Eurostat, 2017a). In addition, Poland is the predominated sender country of migration stocks (by country of birth) to UK and Germany, with 844,024 migrant stocks (by country of birth) in UK and 1,334,000 migrant stocks in 2016 (Eurostat, 2017c).

Except for the clear migration pattern from Poland to UK and Germany, UK and Germany are important migration flows sender countries for each other because they are the predominated destination countries of intra-EU migrations. For example, UK is the second large destination country of Germany emigrants, with 15,490 emigrants to UK which is a slightly smaller than that in Austria (15,855 emigrants) in 2015 (Eurostat, 2017b). In the meantime, Germany is the six major source country of immigration flows to UK, which is 26,529 inflows (based on nationality and previous residence) in total to UK in 2015 (Eurostat, 2017a; 2017b). With respect to migrant stocks in 2016, the immigrant stocks born in Germany mostly resided in UK (290,545), which is the third large source country of immigrant stocks in UK.

In a qualitative review, Marchand et al. (2019) concluded that there are three major intra-EU migration reasons, which are 1) looking for a job; 2) family reasons and 3) seeking for a better education. Among the migrants residing in UK, there are approximately 50.6% of migrants moving an employment, 34.5% of those moving for family reasons and 10.5% of those moving for studying. These major migration reasons explain 95.6% of the total immigrations to UK. On the other hand, family reasons are the predominated migration drivers for those who migrate to Poland, which represents 65.3% of the total immigrants. Migration for employment purpose and education purpose takes the same proportion of 14.3% of the total immigrants.

There are 93.9% of immigrants in Poland are represented by these three migration drivers. Germany data are not available in the Labour Force Survey (LFS) (Eurostat 2017).

In conclusion, the international migration between Poland, UK, and Germany contributed majority migration flows on intra-EU migrations after the enlargement of EU in 2004. Furthermore, the geographic locations, economic differences between these three countries, as well as the EU enlargement event emphasise their suitability to study international migration drivers and policy changes.

5 Data collection

The model simulates the bilateral international migration flows between UK, Poland, and Germany from 2002 to 2030. Yearly data are collected for all the countries if available.

5.1 Foreign direct investment companies

In the model, the fixed parameter, namely, number of FDI companies, is denoted as the number of jobs created by FDI companies. The data are collected from the Foreign Direct Investment Report (UNCTAD 2016). The report presents the number of jobs created in Europe through foreign direct investment from 2006 to 2016 by country. Between the period 2002 with 2016, UK creates 852627 jobs for Europe, Poland creates 743933 jobs and Germany creates 392865 jobs, as shown in Figure 5.1.

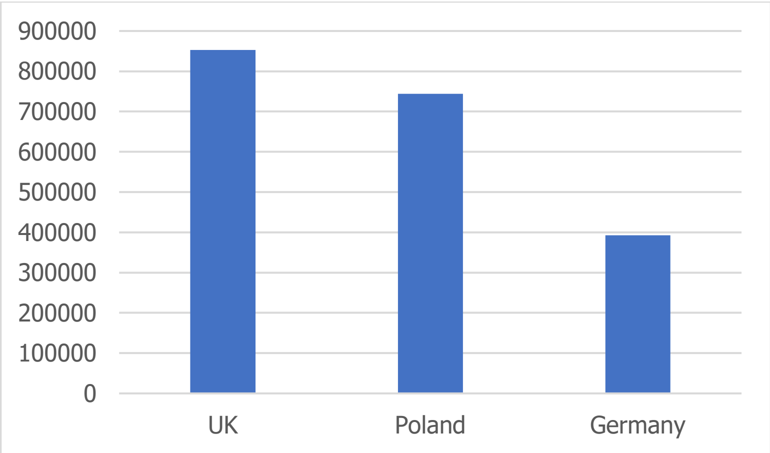

Figure 5.1 Number of jobs created through foreign direct investment projects from 2006 to 2016 by country

This FDI-induced job data is normalised between [0,1] for these countries. Due to the lack of data, the trend of data keeps constant during the simulation.

5.2 Percentage of employed workers

Annually unemployed rates from 2002 to 2021 of these three countries are collected from the OECD dataset. The OECD unemployed rate dataset covered resident population aged 15 years and over living in private or collective households. They collect data from the Microcensus (German Labour Force Survey) from Federal Statistics Office up to 2014 for Germany. Since 2015, the OECD data are collected from European Labour Force Survey from Eurostat for Germany. Regarding Poland, the data come from the Labour Force Survey from Central Statistical Office until 2014, and European Labour Force Survey from Eurostat since 2015. And the data from UK are collected from Labour Force Survey up to 2003 and Annual Population Survey from 2004 onwards. The trends of original data are shown in Figure 5.2.

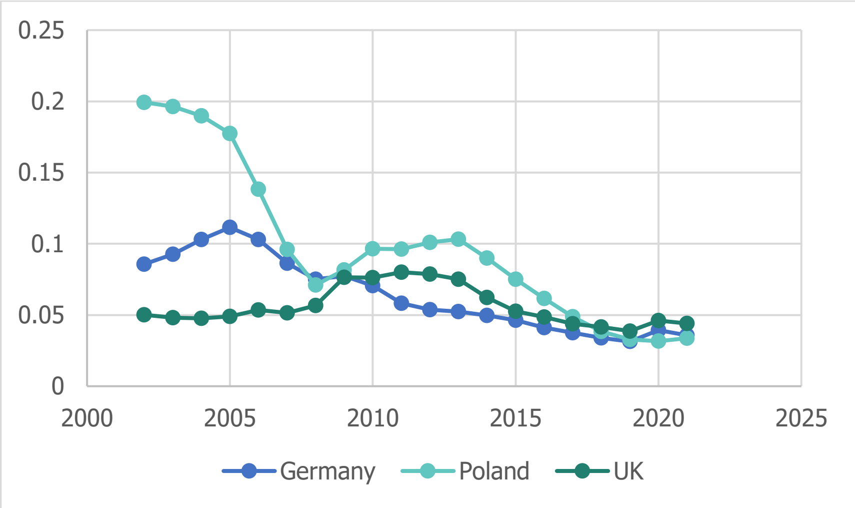

Figure 5.2 Annually unemployment rates of Germany (blue line), Poland (orange line), and UK (grey line) from 2002-2021 (source: OECD dataset).

The annual data collected are annual averages, except for UK, where the annually data are the annual averages of quarterly estimates. Annually unemployed rates data are converted to data of annually employed rates for simulation usage. Future employed rates are kept the same rates as in year 2021 during the simulation.

5.3 Number of top 400 universities

This data collects the number of top 400 universities of all the countries from the World University Rankings of Time Higher Education since 2002, as shown in Figure 5.3. The mean number of top universities from 2002 to 2020 of each country are computed as model inputs. This data are normalised between 0 and 1 and keep constant during the simulation.

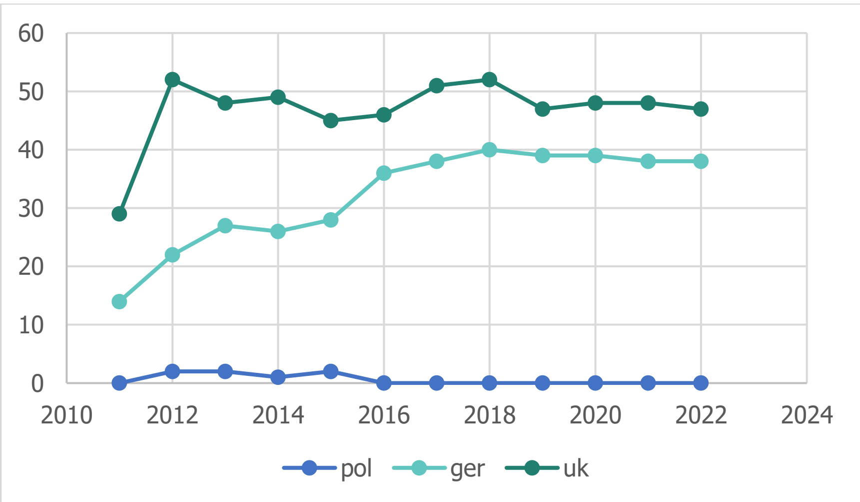

Figure 5.3 Number of world top 400 universities in Poland (blue line), Germany (orange line), and UK (grey line) (source: World University Rankings, THE)

5.4 Potential income and wealth

The mean disposable income and Gini index are collected from the OECD Income Distribution Database. The mean disposable income refers to the cash income regular received after transfers and taxes over the year. The data are provided by German Socio-Economic Panel (Germany), Family Resources Survey (UK), and EU Survey of Income and Living Conditions (Poland) in their local currency. The following plots show the mean disposable income data and Gini index in USD, converted based on Federal Reserve Board average market exchange rate.

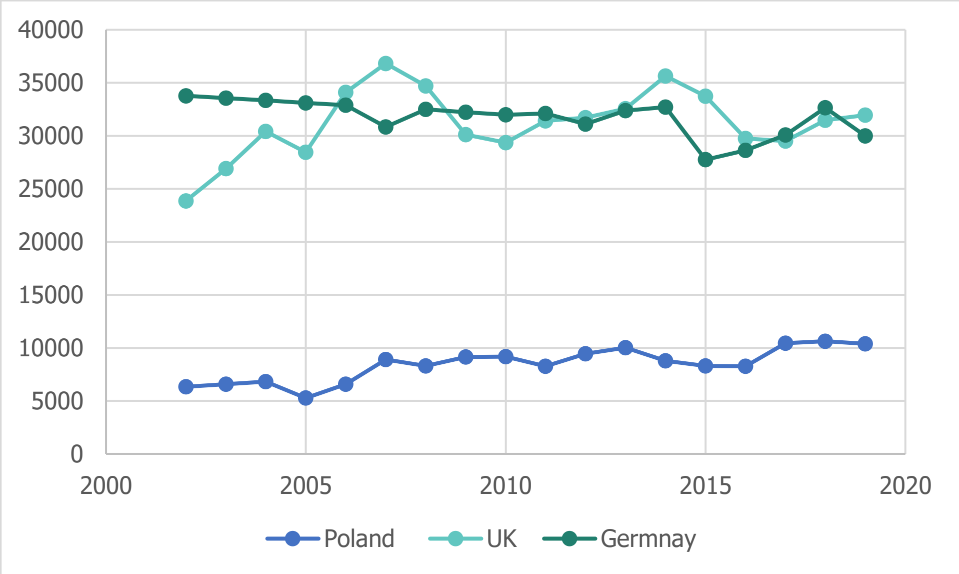

Figure 5.4 Mean disposable income in Poland from 2005 to 2018 (blue); mean disposable income in Germany from 2011 to 2018, and 2008 (grey); mean disposable income in UK from 2002 to 2019 (orange)

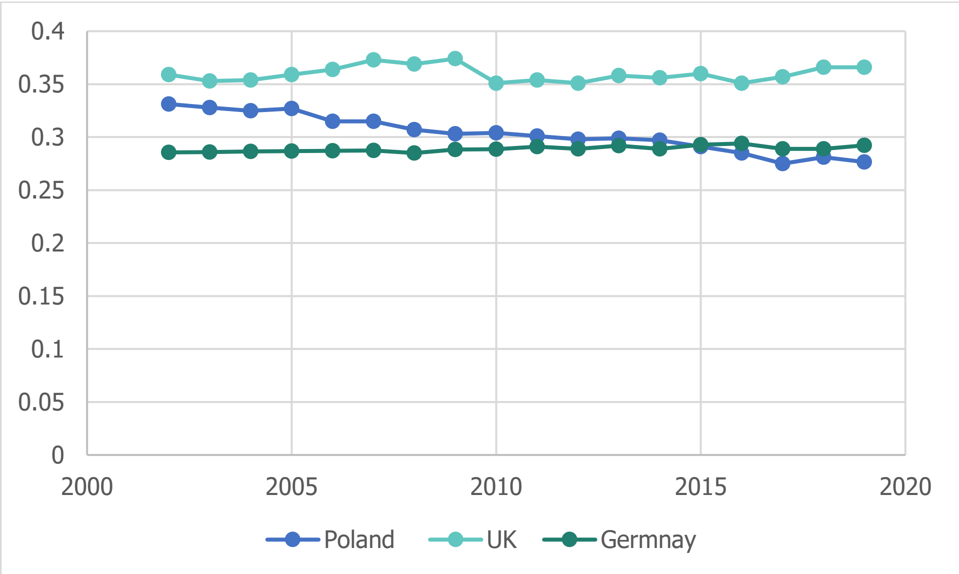

Figure 5.5 Gini Index in Poland from 2005 to 2018 (blue); Gini Index in Germany from 2011 to 2018, and 2008 (grey); Gini Index in UK from 2002 to 2019 (orange)

The missing historical data since 2002 are estimated through a simple linear regression model:

$$y = a * YEAR + b$$

where y is mean disposable income or Gini index, a is the regression coefficient and b is a constant. By fitting the linear regression, the data are complete from 2002 to 2019 as shown in Figure 5.6 and Figure 5.7.

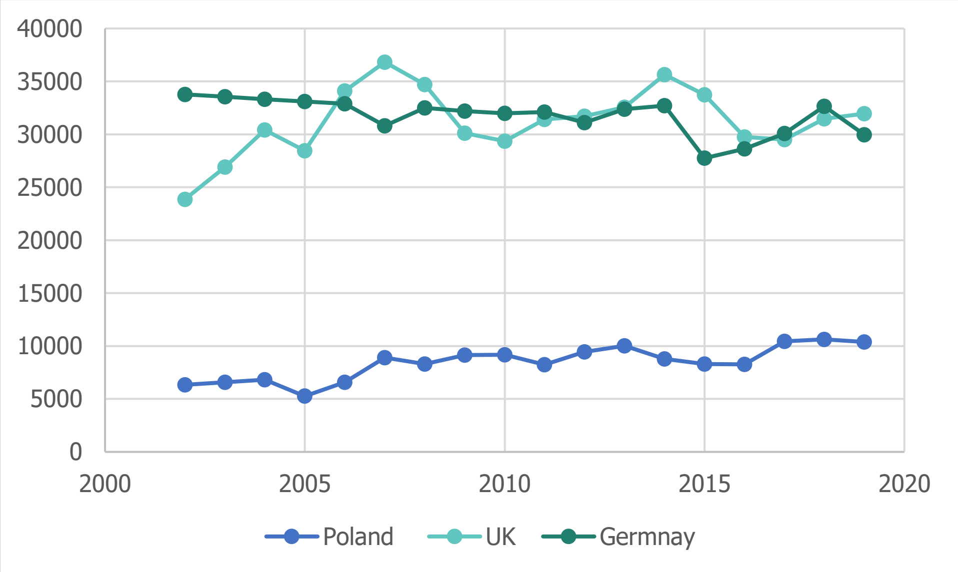

Figure 5.6 Mean disposable income in Poland (blue); Germany (grey) and UK (orange) from 2002 to 2019

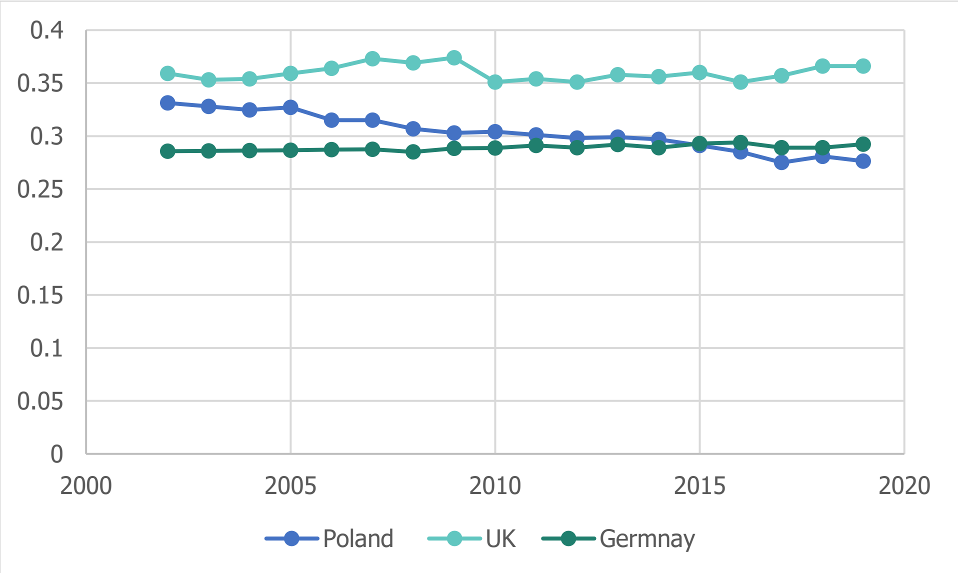

Figure 5.7 Complete Gini index in Poland (blue); Germany (grey) and UK (orange) from 2002 to 2019

The future income patterns are forecasted by the time series ARIMA model. Before determining the type of applied ARIMA model, data stationarities are checked because the ARIMA model assumes the time series data are stationary. If the time series data are non-stationary, a differencing approach is required. The ACF (autocorrelation function) show the coefficient of correlation between two variable values in a time series (see Figure 5.8; 5.9; 5.10).

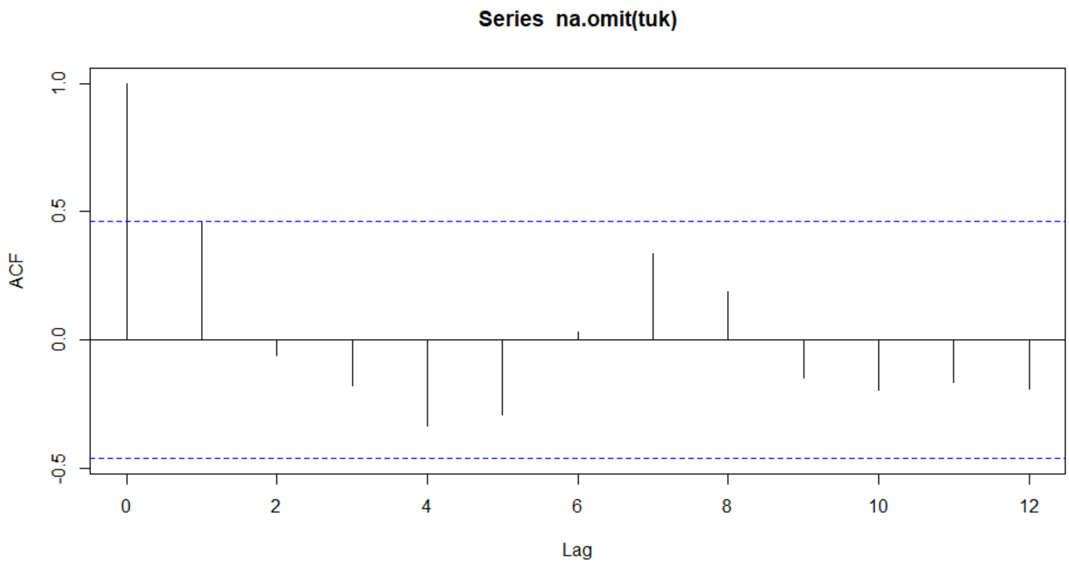

Figure 5.8 The ACF of yearly mean disposable income changes of UK from 2002 to 2019

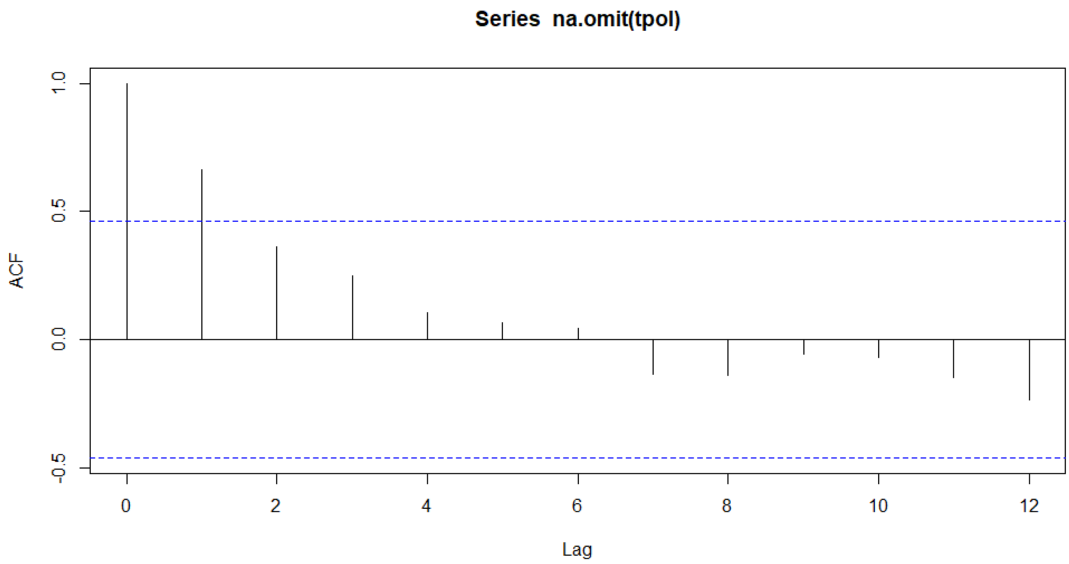

Figure 5.9 The ACF of yearly mean disposable income changes of Poland from 2002 to 2019

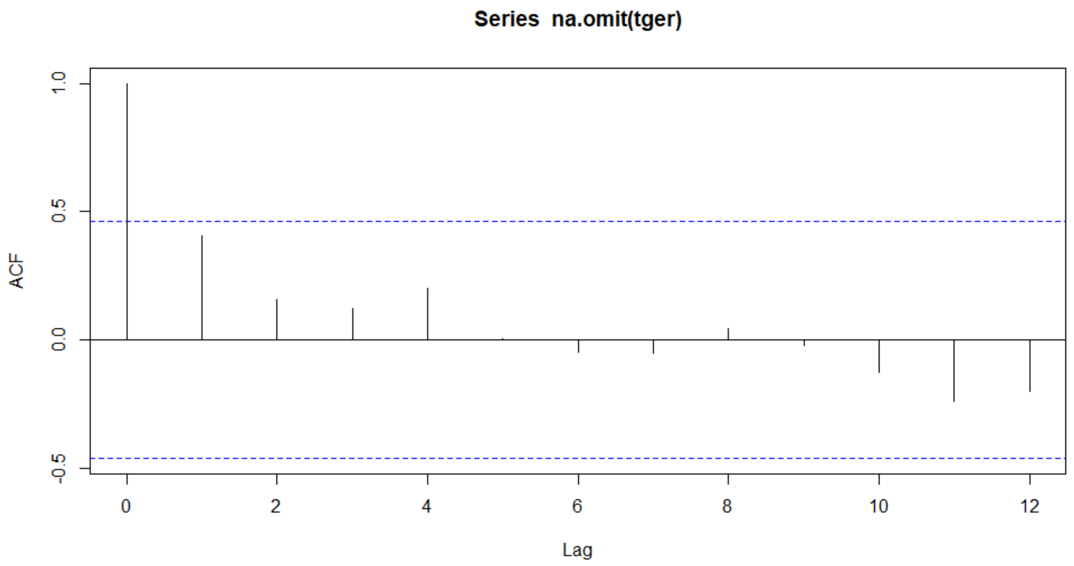

Figure 5.10 The ACF of yearly mean disposable income changes of Germany from 2002 to 2019

The ACF decreases to zero rapidly if data is a stationary time series (e.g., Figure 5.8).
To further check the data stationarity more objectively, the *Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test* (Kwiatkowski, Phillips, Schmidt, & Shin, 1992) is utilized, where its null hypothesis is that the data are stationary. The resulting smaller p-values, 0.002 and 0.03 (less than 0.05) for Poland and Germany respectively, indicate that differencing is required for these data. The forecasted mean disposable income from 2019 to 2030 are shown in Figure 5.11; 5.12 and 5.13.

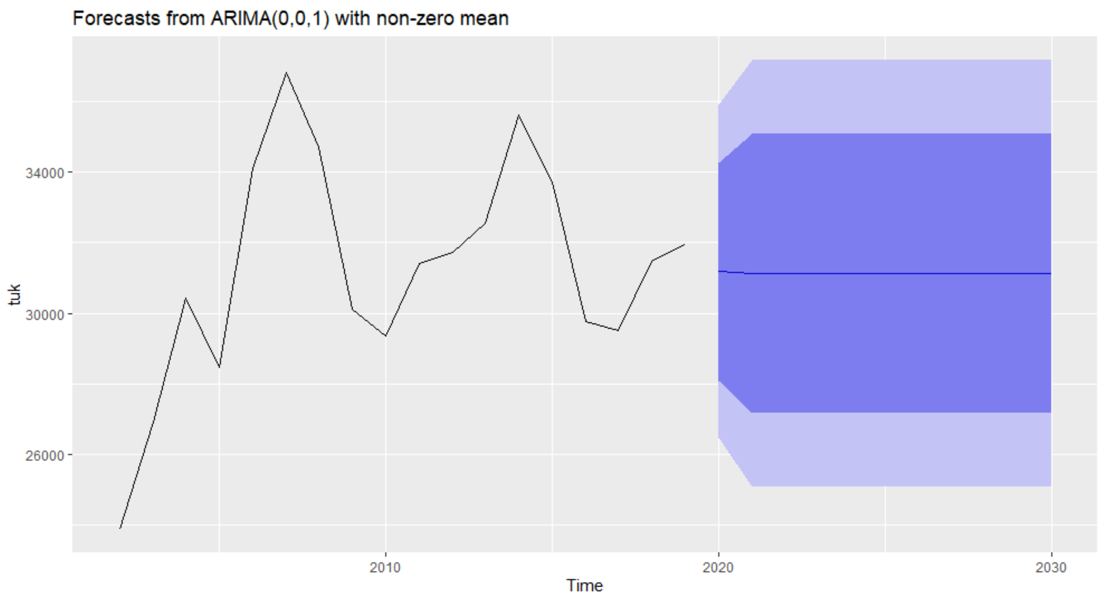

Figure 5.11 Mean disposable income median forecasts (blue line), 80% forecast interval (dark blue) and 95% forecast interval for UK to 2030

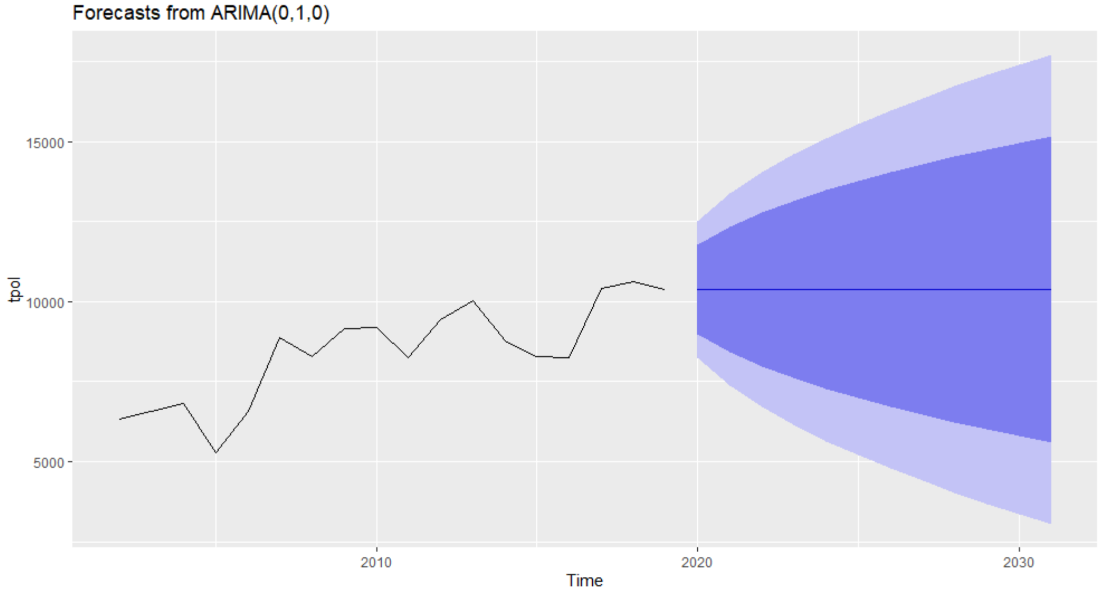

Figure 5.12 Mean disposable income median forecasts (blue line), 80% forecast interval (dark blue) and 95% forecast interval for Poland to 2030

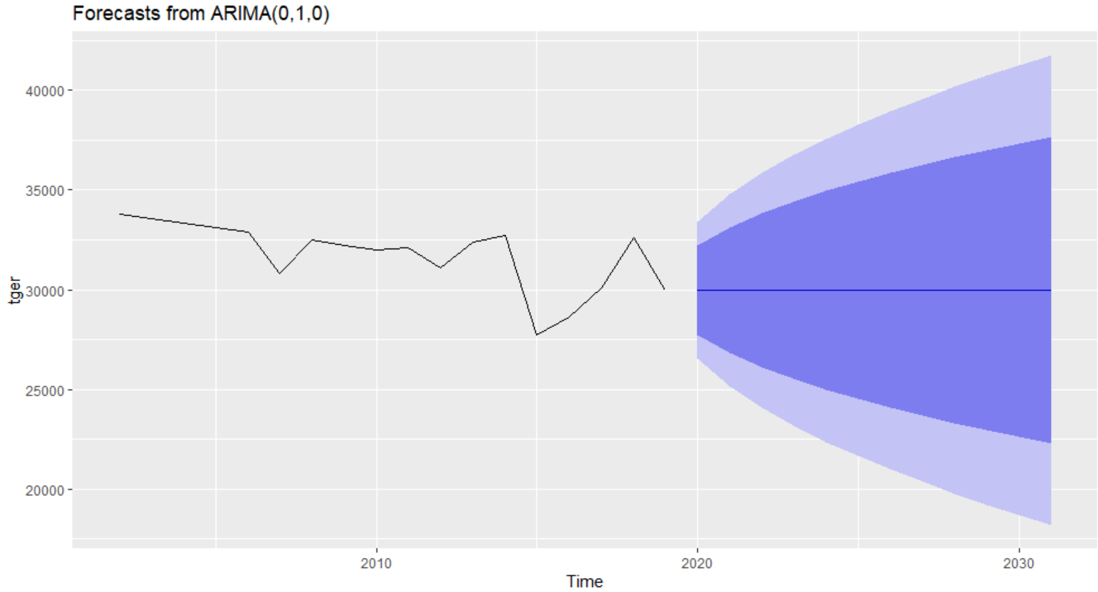

Figure 5.13 Mean disposable income median forecasts (blue line), 80% forecast interval (dark blue) and 95% forecast interval for Germany to 2030

In the ABM model, potential income is assigned to each individual based on a lognormal distribution with mean, the monthly disposable income, and variance, the Gini index. The Gini index are assumed to be static in the future.

The wealth data denotes the initial cumulated wealth at the start of 2002. As there is no available accumulated wealth data, the income data in year 2002 is set as the person's initial wealth. The term 'income' used in this section is also referred as 'wages' in other literatures.

5.4 Micro data

The personal characteristics data are collected from IPUMS micro dataset for UK (Census 2001) and Poland (Census 2002). Due to the data availability, Germany micro data are collected from EU Labour Force Survey 2016. The following characteristics are extracted from the micro dataset:

1. Age (AGE): 0 to 100;
2. Sex (SEX): male or female;
3. Marital status (MARST) (only available for Poland and UK): single, married, divorced and widowed;
4. Education level (EDATTAIN): invalid, less than primary, primary, secondary and university;
5. Employment status (EMPSTAT): invalid, employment, unemployment and inactive;
6. Place of residence one year ago (MIGRATE1) (only available for Poland and UK): non-migrant, inter-regional, intra-regional, inter-province, intra-province and aboard.

The chi-squared test between migration status and other variables for UK and Poland are shown in Table 5.1 and Table 5.2. The results show the immigration event are strongly dependent on the personal characteristics including age, sex, marital status, education level and employment status in UK and Poland.

Table 5.1. Results for the Chi-squared test between each variable and migration status variable for UK. Significance level: '***' p<0.001, '**' p<0.01, '*' p<0.05, '.' P<0.1

Variable	Chi-squared statistic
AGE	8331.2***
SEX	3.6486*
MARST (Marital status)	2699.7***
EDATTAIN (Education level)	7942.5***
EMPSTAT (Employment status)	513.87***

Table 5.2. Results for the Chi-squared test between each variable and migration status variable for Poland. Significance level: '***' p<0.001, '**' p<0.01, '*' p<0.05, '.' P<0.1

Variable	Chi-squared statistic
AGE	1810.7***
SEX	65.879***
MARST (Marital status)	1233.1***
EDATTAIN (Education level)	2419.4***
EMPSTAT (Employment status)	1893.8***

5.5 Consumption

The annual household expenditure per capita data are collected official statistics, as shown in Figure 5.14, and seen as the input of consumption data in the model. The Annual Household Expenditure per Capita is calculated from annual Monthly Average Household Expenditure multiplied by 12 for Germany and Poland. The Annual Household Expenditure per Capita of UK are calculated by multiplying the Weekly Average Household Expenditure by 52. The required data for the annual household expenditure per capita data calculation, i.e., the household expenditure in local currency and average household size, are provided by Office for National Statistics (UK), Statistics Poland (Poland) and Statistisches Bundesamt (Germany). The data currency units are converted into USD based on Federal Reserve Board average market exchange rate.

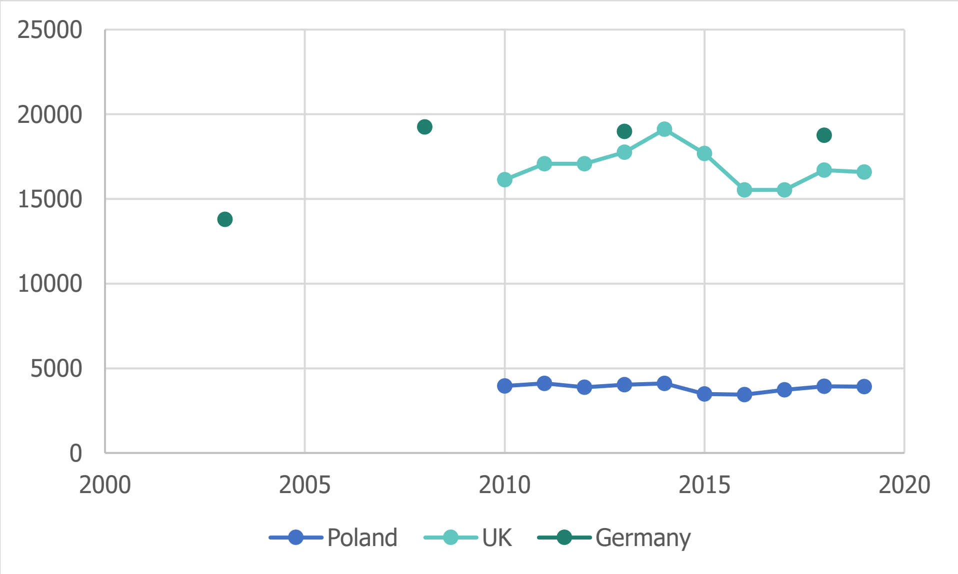

Figure 5.14 Annual Household Expenditure per Capita data for UK (orange) and Poland (blue) between 2010 to 2019; and Germany (grey) for year 2003, 2008, 2013 and 2018.

Similar to the process of predicting income data, the missing consumption data are estimated from linear regression model and forecasted by a time series model (see following figures).

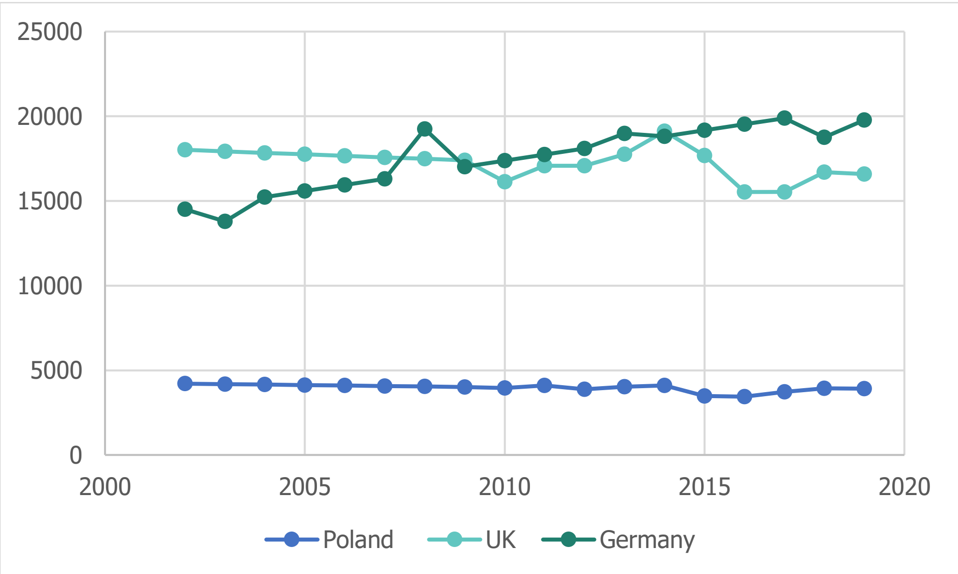

Figure 5.15 Complete Annual Household Expenditure per Capita data for UK (orange), Poland (blue) and Germany (grey) from 2002 to 2019, estimated by linear regression

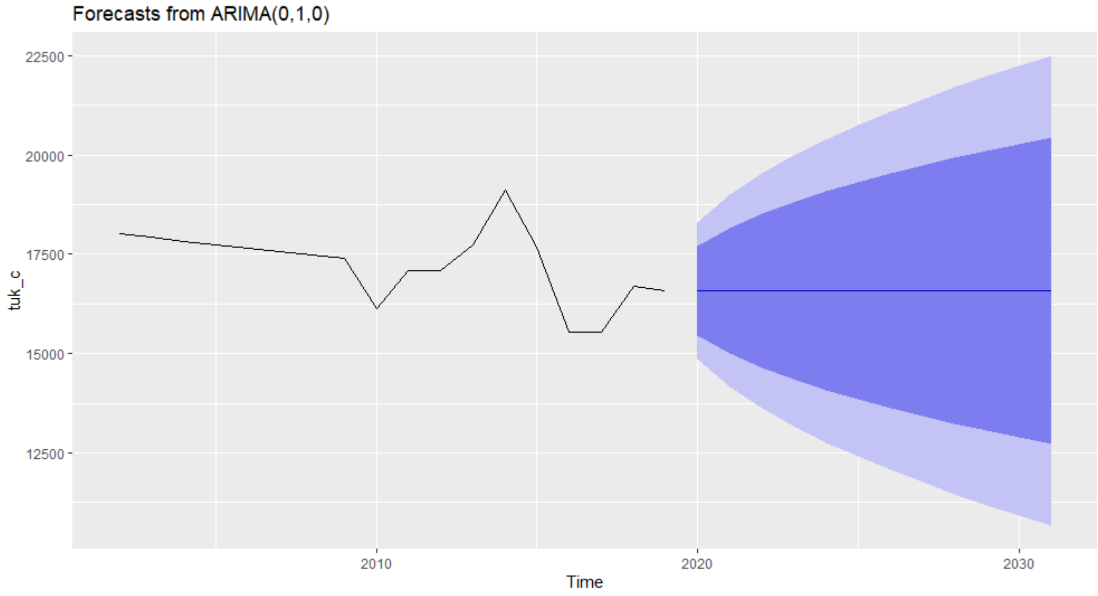

Figure 5.16 Median forecasts (blue line), 80% forecast interval (dark blue) and 95% forecast interval (light blue) of UK Annual Household Expenditure per Capita data until 2030

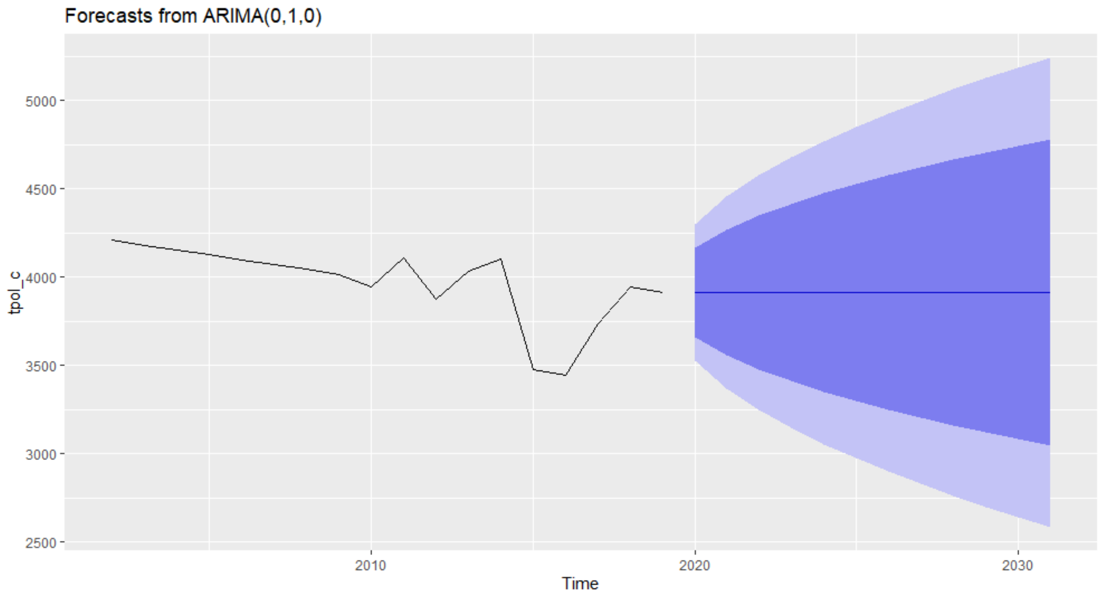

Figure 5.17 Median forecasts (blue line), 80% forecast interval (dark blue) and 95% forecast interval (light blue) of Poland Annual Household Expenditure per Capita data until 2030

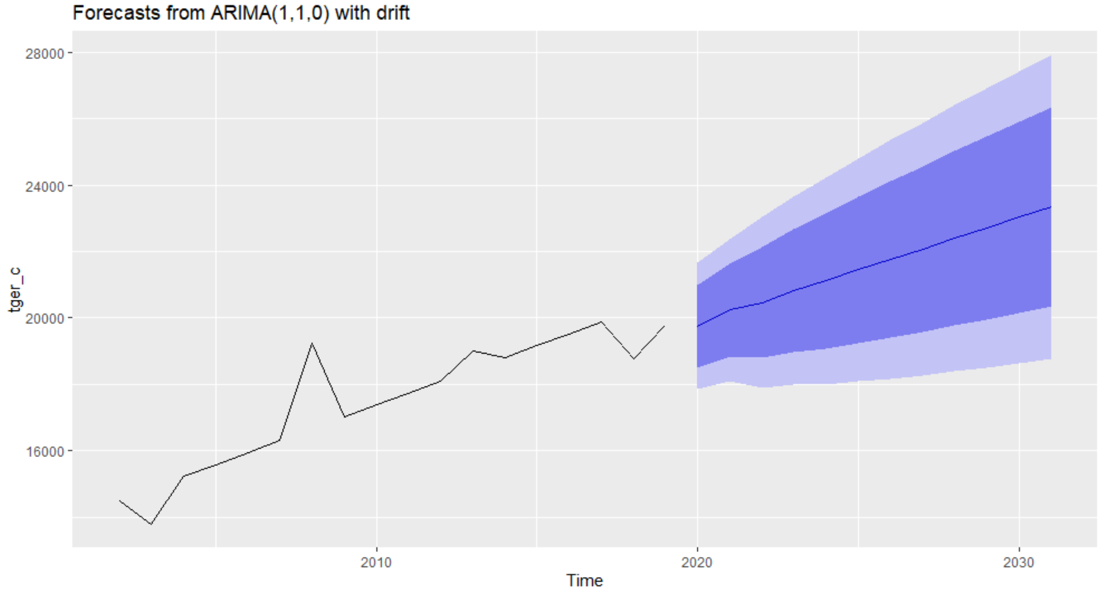

Figure 5.18 Median forecasts (blue line), 80% forecast interval (dark blue) and 95% forecast interval (light blue) of Germany Annual Household Expenditure per Capita data until 2030

5.6 Migration cost

In this model, the migration cost is computed as the sum of consumption and forgone income for one month for the migration preparation. The data collected and predicted from section 5.4 and 5.5 are used in this section.

5.7 Distance and border control

The geographical distances between countries are collected and normalised in the range of 0 and 1. The distance is a critical determinant of migration decision because a longer distance indicates a higher cost of transportation and psychological cost of migration (Massey et al. 1993).

The border control refers to the border enforcement between countries. According to the Schengen Agreement, there are open borders between most member states of EU. Therefore, the border control between UK and Germany is 0 since 2002 until UK left EU on 31 January 2020. The border control between Poland and Germany, and between Poland and UK become 0 after Poland joined EU on 1 May 2004. Except for that, the border control between countries is set as 0.3 during the simulation. The data for border control are under fully assumptions.

5.8 Normalization

Except for potential income and consumption data, which are normalised by logarithm, all other category of data are normalised in the same measurement, which is the so-called 'min-max' method. The variables are normalized according to the following formula:

$$x = \frac{x - \min(x)}{\max(x) - \min\ (x)}$$

where x are the variable values. The resulting values fall in the interval between 0 and 1.

6 Experimental set-ups

The model is implemented in the ABM-focused specific software called Netlogo (Wilensky, U. 1999). According to the population distribution in year 2002, the model is initialised with 1000 person agents in Poland, 1527 person agents in UK and 2156 person agents in Germany. Personal characteristics of each agent are sampled from the micro dataset described in Section 5 and remain static during the simulation. Social-economic environment are loaded from exogeneous dataset and relatively variable information are stored for each country. The information about socio-economic variables are thereafter assigned to each person who are currently residing in this location. A person agent has different income and consumptions at different locations.

The simulation is run for 29-time steps, where each step represents one year from 2002 to 2030. The input data are collected annually, which assumes that the agents have the same variable values during the entire year. Therefore, agents make migration decisions every year, and their attributes change due to the movement between geographic locations. The country-level social-economic information are updated based on the time series exogeneous dataset (described in section 5).

The initial social structure is build based on the Erdős–Rényi Model (Erdős and Rényi 1959), also called the Random Graph Model. The networks are generated randomly depending on probability distributions. There are two variants of the random graph model: $G(n, M)$ and $G(n, p)$. In the $G(n, M)$ model, an undirected network with n nodes and M edges are generated uniformly at random. For example, a $G(3,2)$ is a network with 3 nodes and 2 edges. There are three possible 2-edge networks for a 3 nodes situation, so that the possibility to have a pair connection is $\frac{1}{3}$. On the other hand, a $G(n, p)$ model generates a pair connection between two agents independently with probability p. The expected number of edges in $G(n, p)$ is $C_2^n * p$. As n increases, a $G(n, p)$ is similar to a $G(n, M)$ model with $M = C_2^n * p$ according to the law of large number. Therefore, these two models can be used interchangeably, where the probability p accounts for more uncertainties of the edge numbers, compared with determined edge numbers M. As the number of social networks of each person are not always the same, this ABM model adopts a $G(n, p)$ model to account for various social structures. Specifically, a $G(4683, 0.0036)$ model is proposed for social networks construction where a pair of connections is generated with probability 0.0036 for all 4683 person agents in the simulation. A fully connected social network is built under this model, and there are approximately 16 social links for each person agent (including all kinds of networks, such as family links and friends). This social structure remains static during the whole simulations.

7 Experimental Design

The simulation is running based on 11 computation-based parameters, also known as free parameters, as shown in Table 3.1. All parameters are assumed to be fell in the range of 0 and 1. It is impossible to run every single combination of all 11 parameters values between [0,1] to find the simulated outcomes and empirical value due to its large parameter dimensions. In such case, we keep utilizing the gaussian process emulator (Kennedy and O'Hagan 2001) to predict the simulator outputs of parameter combinations that have not been run from the simulator. First, the simulator (ABM model) is run with several combinations of all parameters based on the predefined design points for each parameter to train the emulator, as shown in Table 7.3.

Table 7.3. Description of parameters and their predefined design points for the simulator run

Parameter	Description	Design points
a	Parameter of A	0.005; 0.1
b	Parameter of SN	0.005; 0.1
c	Parameter of PBC	0.005; 0.1
α_1	Parameter of job opportunity (A)	0.005; 0.5
α_2	Parameter of potential income (A)	0.005; 0.5
α_3	Parameter of education opportunity (A)	0.005; 0.5
α_4	Parameter of personal characteristics (A)	0.005; 0.5
β	Parameter of social norms (SN)	0.005; 0.5
γ_1	Parameter of financial situation (PBC)	0.005; 0.5
γ_2	Parameter of distance (PBC)	0.005; 0.5
γ_3	Parameter of border control (PBC)	0.005; 0.5

Even if there is only two design points for each parameter, there are still 2048 (2^{11}) combinations of design points to run due to the large parameter dimensions. This meant a great computational burdens, which took one day to run on a standard PC. The simulating outputs and their corresponding design points are served as the training data for the construction of an emulator of the ABM simulator.

Therefore, the possible simulator outputs based on various parameter configurations (i.e., design points which have not been run from the ABM simulator) are predicted from the resulting gaussian process emulator. For the unseen design points, a computer experimental design called Latin Hypercube Design (LHD) (Morris and Mitchell 1995) is utilized to generate the parameter configurations, instead of traditional sequential points due to the large experiment regions.

The Latin Hypercube Design divided each parameter interval [0,1] into n equally spaced intervals and only one design point is sampled from each interval. The design points of LHD are generated in a non-linear way to maximize the minimum distance between points. The main advantage is that LHD could generate least design points to represent maximum information from the design interval [0,1].

8 Results

The ABM model has the potential to produce both migration flows and migrants. However, there is a definition issue in Poland on migrants. In general, a person's country of birth is seen as the best way to measure international migrants because citizenship can change but birthplace cannot. However, the birthplace changed in the Poland context because nation borders had been redrawn after World War II. A group of people may be officially recorded as born in Germany, but actually reside in the modern part of Poland (REMINDER project). This meant an increase in the number of migrants.

Therefore, this model only produces the international migration flows because it is the most suitable data to measure the dynamics of international migration. As an indicative approach, the expected outcomes are migration rates (i.e., the simulated migration flows divided by origin population at the start of each time step) for the six migration corridors between three countries, rather than the exact number of migration flows.

8.1 Calibration result

For calibration, the parameters are explored to minimize the mean squared error (MSE) between the simulated outputs and empirical migration flows in 2015 derived from Eurostat. Due to data availability, three migration corridors, which are from Poland to UK and Germany and from Germany to UK, are used for calibration. Table 8.4 shows the best parameter configurations with a minimum MSE of 0.00024502 and a standard deviation of 0.001531173.

Table 8.4. Calibrated parameters based on the minimum MSE

Parameter	Description	Value
a	Parameter of A	0.029517
b	Parameter of SN	0.086686
c	Parameter of PBC	0.013379
α_1	Parameter of job opportunity (A)	0.385432
α_2	Parameter of potential income (A)	0.025073
α_3	Parameter of education opportunity (A)	0.297783
α_4	Parameter of personal characteristics (A)	0.000998
β	Parameter of social norms (SN)	0.8456
γ_1	Parameter of financial situation (PBC)	0.630504
γ_2	Parameter of distance (PBC)	0.05961
γ_3	Parameter of border control (PBC)	0.455503

This set of parameters is chosen to run the final model. The emigration rate in 2015 shown in Figure 8.1 indicate that the simulation model is possible to explain the dynamic migration patterns, where there were most people migrating from Poland to Germany and least people migrating from Germany to UK in 2015.

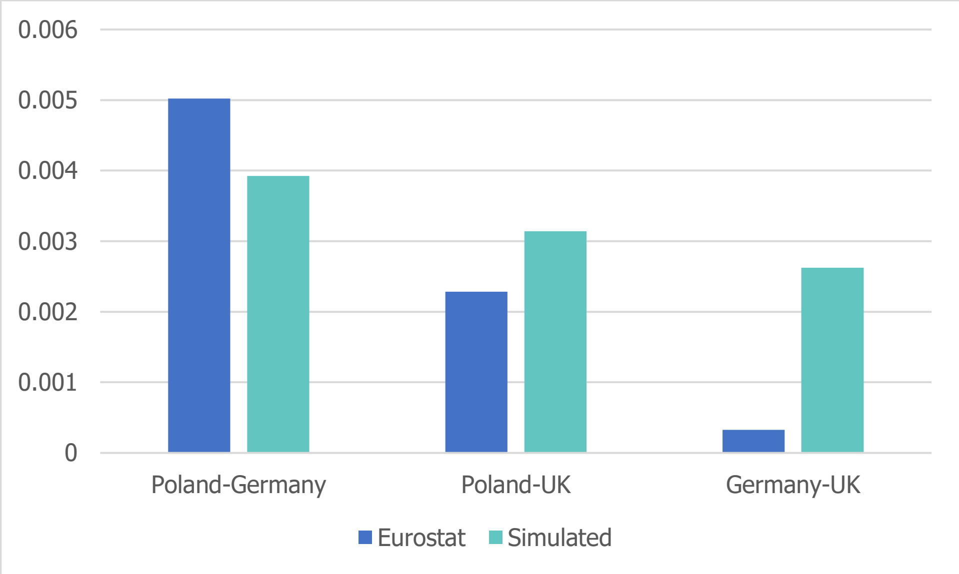

Figure 8.1. Comparison of simulated emigration rate (orange) and Eurostat (blue) of three migration corridors, Poland to Germany, Poland to UK and Germany to UK, in 2015 (Source: own calculation based on migration flows and population data from Eurostat)

One possible future international migration flow projected from 2022 to 2030 is shown in Figure 8.2 and Figure 8.3 based on the scenario with median forecast of personal income and consumptions (described in section 5). The projection results show that the migration corridor from UK to Poland has the highest emigration rate in the future years until 2030, followed by migration corridors between Germany and UK.

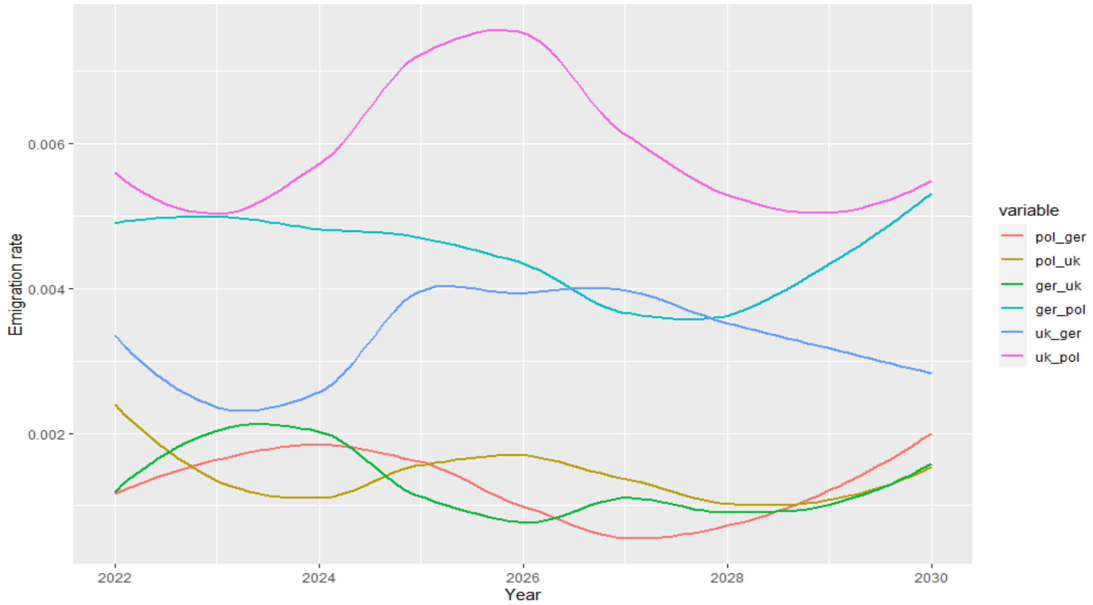

Figure 8.2. Forecasted international migration flows between Poland, UK, and Germany until 2030 based on median personal income and consumption scenario.

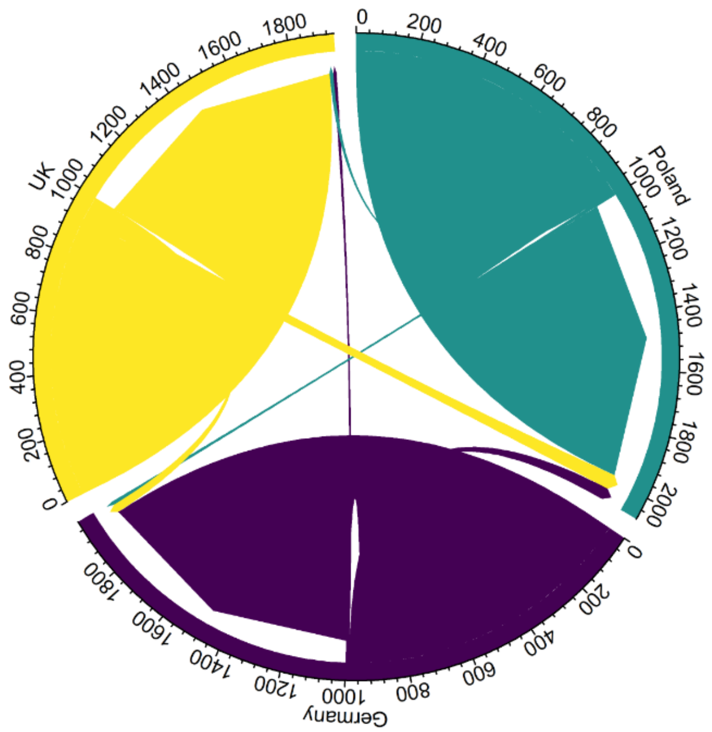

Figure 8.3 Chord diagram of simulated migration flows per thousand, 2022-2030

8.2 Sensitivity analysis

Together with calibration, sensitivity analysis is crucial for model validation as well. A sensitivity analysis was performed to study what proportion of model output uncertainty can be interpreted by the different source of uncertainty in model inputs (Saltelli et al. 2008), and therefore, understand which parameters have influence on the migration behaviours.

In line with Sobol (1993), the sensitivity analysis relied on the Sobol, also known as the FANOVA decomposition, where the total variances of outputs are decomposed as variances of each input and their interactions. In our case, input refers to the parameters in the model and output refers to the simulated emigration rates. Consequently, the global influence of a parameter onto the final output can be measured as the division of parameter variance and total output variance. This ratio is also called as Sobol indices (see Sobol 1993 and Saltelli 2000 for more details).

The results shown in Figure 8.4 indicate that the emigration rates do not depend on parameters solely. For single parameter influences, the emigration rates depend mostly on parameter β and α_1, which represent social networks and job opportunities respectively.

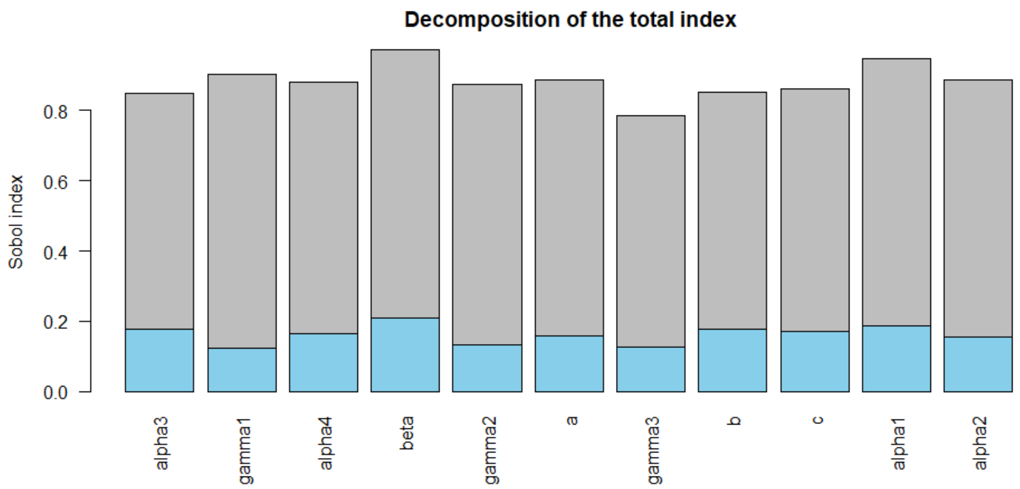

Figure 8.4 Sobol index. For each input variable, the decomposition of the total index (first order index (blue) + sum of the interactions with all other inputs (grey)) is visualized with a box split into two sub boxes, which respective heights give the index values.

8.3 Further calibration result

Due to the data deficiency and their uncertainty, various methods have been proposed to estimate the quantity of international migration flows, which may produce different results between methods. Figure 8.5 shows the difference between Eurostat estimation and Abel and Cohen's (2019) estimation. Abel and Cohen's (2019) estimated international flows for 200 countries and their estimated bilateral international migration database have been recognised and used as formal database for World Bank.

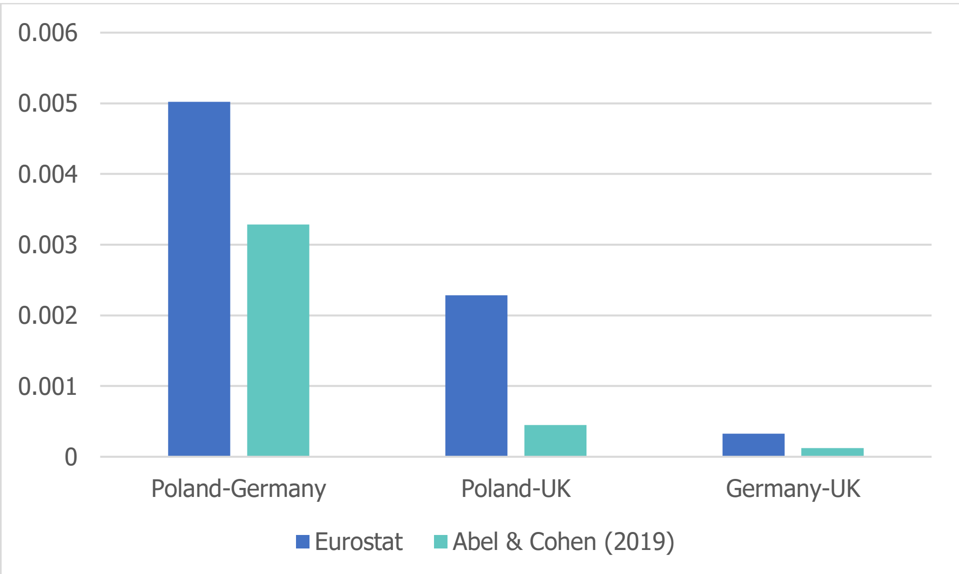

Figure 8.5 Comparison of Eurostat estimation and Abel and Cohen's (2019) estimation on emigration rates from Poland to Germany, Poland to UK and Germany to UK in 2015 (Source: own calculation based on migration flows data from Eurostat and Abel and Cohen (2019))

Abel and Cohen's (2019) estimated less migration flows than Eurostat, although they are sharing similar migration patterns. Further calibration is performed based on Abel and Cohen's (2019) estimation in 2015 for three migration corridors, from Poland to Germany, Poland to UK, and Germany to UK.
Table 4.5 shows the best calibrated parameter configurations with a minimum MSE of 0.0001841361 and a standard deviation of 0.005228758.

Table 8.5. Calibrated parameters based on the minimum MSE for both Eurostat and Abel and Cohen

Parameter	Value (Eurostat)	Value (Abel and Cohen)
a	0.029517	0.001966
b	0.086686	0.029517
c	0.013379	0.002677
α_1	0.385432	0.266169
α_2	0.025073	0.04393
α_3	0.297783	0.133211
α_4	0.000998	0.049925
β	0.8456	0.10809
γ_1	0.630504	0.581876
γ_2	0.05961	0.2758
γ_3	0.455503	0.287923

This set of parameters is used to run the model and produce simulated emigration rate from 2002 to 2030. The results (see Figure 8.6) show a similar migration pattens over three approaches in 2015. It needs to be noted that the lack of simulated data of Germany-UK corridor is because the ABM model produced 0 migration flow.

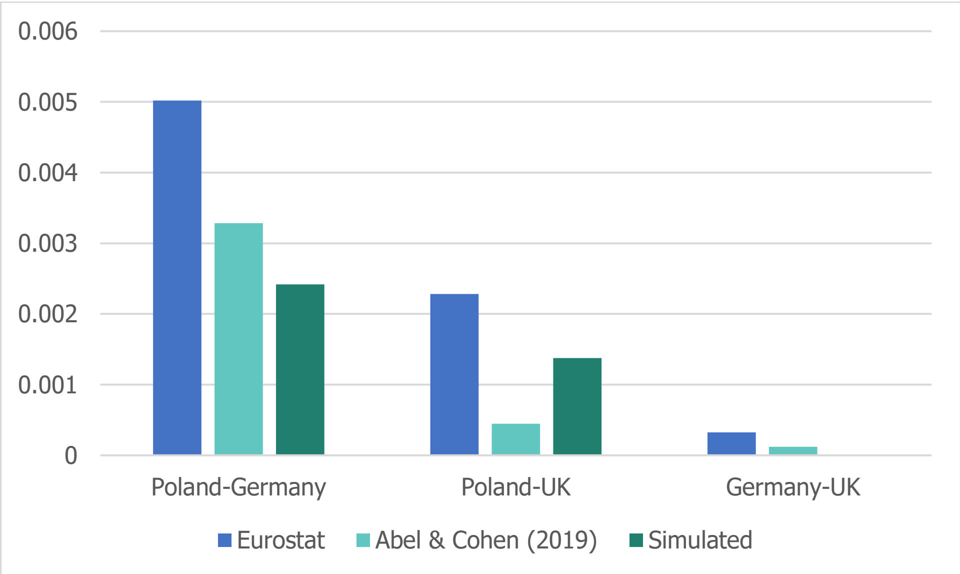

Figure 8.6 Comparison of simulated data and Eurostat estimation and Abel and Cohen's (2019) estimation on emigration rates from Poland to Germany, Poland to UK, Germany to UK in 2015 (Source: own calculation based on migration flows data from Eurostat and Abel and Cohen (2019))

On the other hand, the forecast results show the Germany-Poland migration corridor has the largest migration flows while the others are nearly zero (see Figure 8.7).

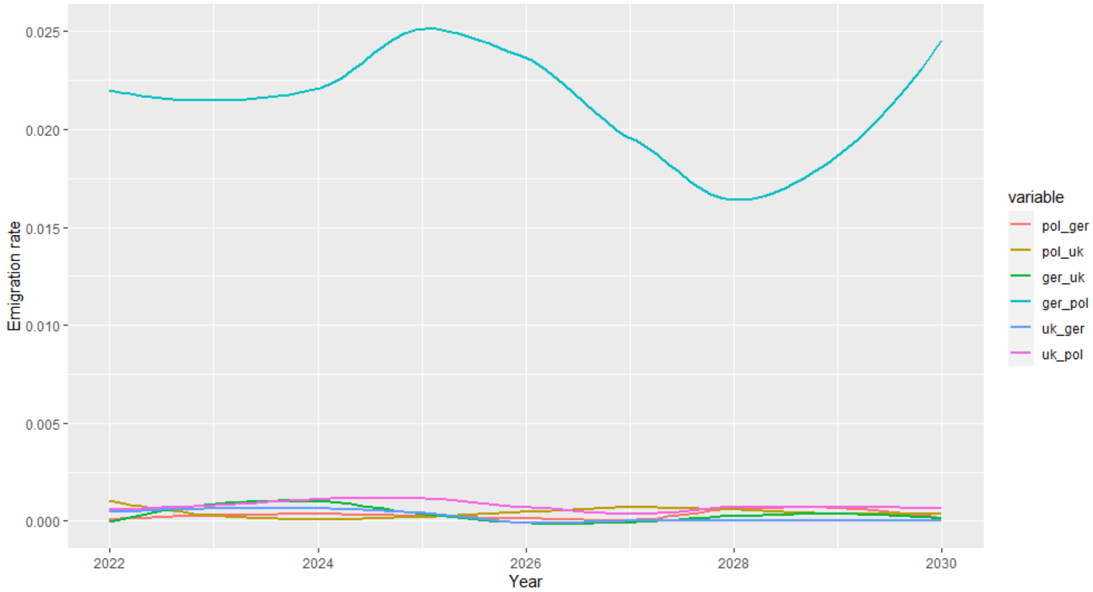

Figure 8.7 Forecasted international migration flows between Poland, UK, and Germany until 2030 based on median personal income and consumption scenario, where parameters are calibrated based on Abel and Cohen (2019).

9 Discussion

As the migration intentions are measured as a linear combination of all three components, including attitudes of migration, social norms about migration and perceived behaviour control over migration (see equation 5.1), the calibrated parameters can be interpreted as the corresponding effect of each component on migration behaviours.

The results indicate that the social network is the most important component for migration behaviours. As the model does not account for households, the family effects on migration behaviours are interpreted by part of the social network effects. Under the component of attitude, the parameter values suggest that employment- and education-related factors are leading migration drivers, which is in line with Marchand et al. (2019). In general, wage (denoted as income in our model) and employment opportunities are deemed as the main drivers of intra-EU migrations. However, our results suggest that the potential income does not have much effect on migration decisions. Similar results are also shown in Arenas-Arroyo et al. (2019), where they found that the differences between minimum wages are not major intra-EU migration drivers. For the perceived behaviour control component, we conclude that the financial availability to afford a migration behaviour is the main obstacle over an actual migration, followed by the problem of getting a visa. The distance between origin and destination does not prevent migration decisions significantly.

The similar migration patterns between simulated emigration rates and Eurostat data emphasis the ability of ABM approach to study dynamic patterns of international migration flows. The overpredictions of Poland-UK and Germany-UK migration corridors owe to the smaller initial agent persons. The emigration rate is computed as the migration flows divided by total population of origin country, and the empirical emigration rate from Germany to UK is below 0.0004. On the other hand, the smallest emigration rate (i.e., one migration from Germany) is 0.00046 which is larger than the empirical rate already. The increase of population caused by immigration does not make a difference because the number of migration event is small as well.

The scenario-based forecast is presented based on median forecast of personal income and consumption. The results show that persons who are currently residing in UK are more likely migrate out to Poland, with a peak in year 2026. One possible explanation for the highest emigration rate from UK is Brexit. The return migration after 'Brain Drain' may be another explanation. The term Brain Drain indicates the emigration of talented professional.

For sensitivity analysis, we find that the migration decision is not determined by single migration driver. All migration factors are jointly affecting the final migration decisions. This finding is in line with the main findings of REMINDER project (Role of European Mobility and its Impacts in Narratives, Debates and EU Reforms, 2015-2018) and Bijak and Czaika (2020). They found that the migration decisions are made from a unique combination of migration factors which are highly interrelated and difficult to isolate one from the other.

In addition, further calibration results based on different empirical data indicate that the estimation differences caused by different method do not have significantly effect on explaining the underlying migration mechanisms, as long as they share the same migration patterns. However, they do have an influence on prediction accuracy of migration flows. This is because the simulated results are calibrated to match the empirical data, different empirical data lead to different predictions. Therefore, a more reliable data should be utilised for calibration purpose. In addition, the underprediction of Abel and Cohen's (2019) estimation over Eurostat leads to an even smaller emigration rate, which causes a few 0s during the simulation. This collapses the reliability of future migration forecasts.

10 Conclusion

This research aims to study the dynamic migrations between Poland, UK, and Germany with respect to the both explanation and prediction of migration flows based on the agent-based modelling approach. This research fills the gap of the lack or formal modelling on aleatory migration uncertainty measurement by studying the underlying migration mechanisms through a micro-level approach. The major contribution is the methodological innovation on the extension the agent-based modelling approach to study international migrations between multiple countries.

For explanation purpose, the research found that individual migration decisions are jointly determined by a combination of migration factors. Social networks are the main contributor of migration behaviours, followed by the component of attitude towards migration. The results show that employment- and education-related migration factors are the leading factors on migration decision-making process. For the perceived behaviour control component, the results show that the actual migration behaviour is significantly prevented by the lack of financial ability to afford the migration. The difficulty of getting a valid visa is the second obstacle that prevents the migration behaviour. For prediction purpose, the model shows its ability on studying the dynamics of international migration flows and has the potential to perform scenario-based forecast. However, the low prediction accuracy limits the understanding of migration flows in a quantitively way. This issue is mainly owing to the smaller population size. The rarely happened migration event also prevents our understanding on disaggregated migration flows, for example, migration flows by education. This suggests that further research should be performed on a larger population size. However, a larger population size heavily increases the computational burden. The problem should be solved by proposing a better calibration method with cheaper computational cost and running the model on a High-Performance Computer (HPC).

Further improvement on the model could be an inclusion of return migration mechanisms as return migration is a major source of international migration flows. In addition, another migration obstacle on language issues could be considered because language is a common issue for international migration. it's challenging to transfer language barrier for each individual into numbers. Unreliable assumptions could add another layer of uncertainty.

Therefore, a promising further step is to improve the prediction accuracy to increase the robustness of the ABM model.

References

Abel, G. J., & Cohen, J. E. (2019). Bilateral international migration flow estimates for 200 countries. Scientific data, 6(1), 1-13.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), Action control. From cognition to behavior (pp. 11–39). Springer.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Arango J (2002) Explaining migration: a critical view. International Social Science Journal, 52(165), 283–296.
Arenas-Arroyo, E., Rienzo, C., & Vargas-Silva, C. Minimum Wages, Earnings and Mobility in the EU.

Bijak J, Higham PA, Hilton J, Hinsch M, Nurse S, Prike T, Reinhardt O, Smith PWF, Uhrmacher AM, and Warnke T (2021) Towards Bayesian Model-Based Demography. Agency, Complexity and Uncertainty in Migration Studies. Methodos Series, vol. 17. Cham: Springer. ISBN: 978-3-030-83039-7. Available via DOI: 10.1007/978-3-030-83039-7.

Bijak, J., & Czaika, M. (2020). Assessing uncertain migration futures: A typology of the unknown. Changes, 1(5).

Cai, R., & Oppenheimer, M. (2013). An agent-based model of climate-induced agricultural labour migration. Presented at the Agricultural and Applied Economics Association's 2013 AAEA Annual Meeting, Washington DC. August 4–6, 2013

Castles, S., De Haas, H., & Miller, M. J. (2013), The Age of Migration: International Population Movements in the Modern World,Macmillan International Higher Education.

Craigwell, R. (2006). Foreign direct investment and employment in the English and Dutch-speaking Caribbean. ILO Subregional Office for the Caribbean.

Czaika, M., & Reinprecht, C. (2020). Drivers of migration: A synthesis of knowledge. IMI Work. Pap. Ser, 163, 1-45.

Epstein, G. S. (2008). Herd and network effects in migration decision-making. Journal of Ethnic and Migration Studies, 34(4), 567-583.

Erdős, P., & Rényi, A. (1959). Some further statistical properties of the digits in Cantor's series. Acta Math. Acad. Sci. Hungar., 10, 21-29.

Eurostat. 2017. "European Union Labour Force Survey: Description of the Dataset". http://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey.

Eurostat. (2017a). Immigration by age group, sex and citizenship (Datafile). Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_imm1ctz&lang=en

Eurostat. (2017b). Immigration by age group, sex and country of previous residence (Datafile). Eurostat. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_imm5prv&lang=en

Eurostat. (2017c). Population on 1 January by age group, sex and country of birth (Datafile). Eurostat. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_pop3ctb&lang=en

Fishbein, M., & Ajzen, I. (2011). Predicting and changing behavior: The reasoned action approach. Psychology press.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425-464.

Khan, M. A., Fatima, Z., & Fatima, S. (2022). Revisiting the Gravity Model of Migration. Foreign Trade Review, 00157325221088707.

Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: state of the art and challenges. European Journal of Population, 32(1), 73-97.

Klabunde, A., Zinn, S., Willekens, F., & Leuchter, M. (2017). Multistate modelling extended by behavioural rules: An application to migration. Population studies, 71(sup1), 51-67.

Kniveton, D. R., Smith, C. D., & Black, R. (2012). Emerging migration flows in a changing climate in dryland Africa. Nature Climate Change, 2(6), 444-447.

Kniveton, D., Smith, C., & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. Global Environmental Change, 21, S34-S40.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. Journal of econometrics, 54(1-3), 159-178.

Marchand, K., Fajth, V., & Siegel, M. (2019). Relevant data to understand migration in the EU.
Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. Population and development review, 431-466.

McAuliffe, M., & Koser, K. (2017). A long way to go: Irregular migration patterns, processes, drivers and decision-making. ANU Press.

Morris, M. D., & Mitchell, T. J. (1995). Exploratory designs for computational experiments. Journal of statistical planning and inference, 43(3), 381-402.

Radosevic, S., Varblane, U., & Mickiewicz, T. (2003). Foreign direct investment and its effect on employment in Central Europe. Transnational Corporations, 12(1), 53-90.

Ramos, R. (2016). Gravity models: a tool for migration analysis. IZA World of Labor.

Saltelli, A. (2002). Sensitivity analysis for importance assessment. Risk analysis, 22(3), 579-590.

Sanderson, M. R., & Kentor, J. (2008). Foreign Direct Investment and International Migration: A Cross-National Analysis of Less-Developed Countries, 1985—2000. International Sociology, 23(4), 514-539.

Saucedo, E., Ozuna, T., & Zamora, H. (2020). The effect of FDI on low and high-skilled employment and wages in Mexico: a study for the manufacture and service sectors. Journal for Labour Market Research, 54(1), 1-15.

Simon, M. (2019). Path dependency and adaptation: The effects of policy on migration systems. Journal of Artificial Societies and Social Simulation, 22(2).

Smith, C. D. (2014). Modelling migration futures: development and testing of the Rainfalls Agent-Based Migration Model–Tanzania. Climate and Development, 6(1), 77-91.

Sobol, I. M. (1993). Sensitivity analysis for non-linear mathematical models. Mathematical modelling and computational experiment, 1, 407-414.

Strey, A., Fajth, V., Dubow, T. M., & Siegel, M. (2018). Determinants of migration flows within the EU. Paper published as part of the Reminder Project. www. reminder-project. eu.

Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL. http://ccl.northwestern.edu/netlogo/

Willekens, F. (2017). The decision to emigrate: a simulation model based on the theory of planned behaviour. In Agent-based modelling in population studies(pp. 257-299). Springer, Cham.