

Ali Abbasov

Kamil Aida-zade (Eds.)



BOOK OF SELECTED PAPERS

**6th International Conference on
Problems of Cybernetics
and Informatics
PCI 2025**

**August 25-28, 2025
Baku, Azerbaijan**

Ali Abbasov
Kamil Aida-zade
Editors

BOOK OF SELECTED PAPERS
6th International Conference on
Problems of Cybernetics and Informatics
PCI 2025

August 25-28, 2025
Baku, Azerbaijan

○ Rashid Alakbarov, Oqtay Alakbarov	
Reducing Heat Density and Improving Energy Efficiency in Cooling Systems of Data Centers	106 – 109
○ Elshan Ibayev	
Investigation of the Boundary Functional of the Random Walk Process with a Special Barrier	110 – 112
● Section 7: Information and Control Systems	113
○ Rasim Alguliyev, Afruz Qurbanova	
Evaluation of Term Creation Processes within the Framework of the National Terminology Information System	115 – 119
○ Fuad Mammadov, Huseyn Sultanli	
Development of IELTS Essay Evaluation System with Deep Learning	120 – 122
● Section 8: Natural Language Processing	123
○ Nurzhan Amantay, Yasin Ortakci, Oğuz Fındık	
Primary and Secondary Language Identification in Multilingual Texts Using a Multi-Layer Perceptron and Kolmogorov-Arnold Network	125 – 129
● Section 9: Mathematical Modeling Methods and Their Applications	131
○ Makrufa Hajirahimova, Aybeniz Aliyeva	
Assessing Migration Patterns of Researchers in Azerbaijan Using Digital Trace Data	133 – 136
○ Rashid Alakbarov, Mammad Hashimov	
Reducing Delays and Optimizing Model in Fog Computing-Based SCADA Systems	137 – 140
○ Huseyn Bayramov, Amide Aliyeva, Sevda Akhmedova, Rena Huseynova, Hikmat Asadov	
Issues of Constructing an Integrated Fire Hazard Assessment System	141 – 144
● Section 10: Cybersecurity Methods	145
○ Rasim Alguliyev, Babak Nabiyev, Konul Dashdamirova	
Comparative Analysis of Global Cybersecurity Indices in the Context of the Formation of Cyber Sovereign States	147 – 151
○ Fargana Abdullayeva, Gurban Farajov	
Cloud Cyber Attack Images Classification Using GAN and ViT+ML Algorithms	152 – 156
○ Abdullah Güney Işık, Şeyma Girgin, Hazal Su Biçakci Yeşilkaya	
Machine Learning-Based Credit Card Fraud Detection: Balanced and Imbalanced Cases Evaluation	157 – 161

Assessing Migration Patterns of Researchers in Azerbaijan Using Digital Trace Data

Makrufa Hajirahimova
Department 13
Institute of Information Technology
Baku, Azerbaijan
0000-0003-0786-5974

Aybeniz Aliyeva
Department 13
Institute of Information Technology
Baku, Azerbaijan
0000-0002-1739-1808

Abstract— The process of globalization that is happening all over the world has given rise to international mobility of a skilled workforce with innovative thinking and high knowledge. Understanding the mechanisms behind the mobility of skilled workers is of great importance for the development of science, education and immigration policies. The lack of access to official data stimulates the use of digital trace data obtained from ORCID public records. ORCID profiles contain standardized information education and employment, which allows tracking individuals over time. This study used ORCID data to analyze the mobility of Azerbaijani researchers during 1990-2024. A multivariate nonlinear regression model was proposed for the evaluation of researcher’s migration. The findings of the study may have serious practical implications for the making of migration policies to attract and retain skilled talent.

Keywords—high-skilled migration, scientific migration, ORCID, digital trace data, regression analysis

INTRODUCTION

The globalization of knowledge has promoted to international mobility by helping to transform a skilled workforce and researchers with innovative thinking and high knowledge into key drivers of economic and social development. This has increased performance of researches by facilitating the recombination of knowledge and created the basis for scientific discoveries [1]. As the role of highly skilled researchers in economic development is recognized and countries look to close skills gaps global competition for talent are intensified. Policymakers are actively seeking ways to improve research activity and skills attraction capabilities in their regions in order to achieve economic growth and productivity. However, despite the large economic impact of the scientific achievements of researchers and scientists, the relevance of intellectual migration and the need for a deep understanding of migration processes in order to make sensible policy interventions, the possibilities to measure and track the mobility of scientific researchers remain limited [2]. One of the main reasons for this is the lack of high-quality and accessible data over an extended period of time or the factor that measuring the mobility of highly skilled workers using traditional methods (including surveys, census data and CV analysis) is costly and time-consuming [2, 3].

Understanding the flows of scientists – where they studied, where they moved, how often they moved, how this varies by field, and the type of research university they worked at – is crucial for science and immigration policy.

Recently, the use of non-traditional data sources, in particular bibliometric data [2, 3] and digital trace data [4-6], has become an increasingly researched subject and interest in their application is constantly growing [2, 7-10].

In this study, we monitor the international mobility of Azerbaijani researchers and scholars by proposing the use of ORCID profile data. In particular, we compile the most comprehensive dataset of 1400 international migration of researchers for the period 1990 to 2024 and use it to analyze and estimate the mobility of migration flows of researchers.

PRE-PROCESSING DATA

This section describes the steps taken to retrieve data from ORCID, preparing it for further processing and analysis.

Migration Dataset

Since official data for the analysis of the phenomenon of researchers migration are not available, we extract the necessary migration patterns from the ORCID public data. ORCID (Open Researcher and Contributor ID) is a non-profit organization that provides a unique identifier to each registered user member engaged in research. It contains information on researcher’s works, publications and affiliations. This data is a valuable source of education and employment histories of PhD students, academics, engineers, medical professionals, researchers. The majority of ORCID users have or are in the process of obtaining a PhD. It means that ORCID data captures a subset of high-skilled workers [3, 11]. From ORCID public database we extract migration patterns of registered users, across Azerbaijan in the time horizon spanning from 1990 to 2024 and create a dataset . The data was obtained by analyzing the educational and employment histories of over 13,000 users registered in ORCID. Comparison with independent data sources supports the technical validity and representativeness of this data. Incomplete information on years of employment or education was completed using diaspora and workplace websites, social networking sites such as LinkedIn and Facebook. Upon graduation all students, especially PhD students, have the option to register themselves in the ORCID database. The dataset provides new dimensions of the geography of a subset of highly skilled labor and opens up opportunities to explore hypotheses about intellectual migration. The key variables included in this database and their description are given in Table 1.

TABLE I. VARIABLES AND THEIR DESCRIPTION

Variables	Description
orcid_id	Unique open digital identifier
mig_year	Year in which the researcher left the country
dest_country	Country where the researcher first migrated
has_phd	Does researcher have a PhD?
phd_year	Year in which the PhD researcher received the degree
phd_country	Country where the PhD researcher received the degree

science_field_group	The main field of science in which the researcher is involved
gender	Gender of researcher (female or male)
earliest_degree	Degree when the researcher leaves the country
science_degre_2024	Current degrees of researchers
country_2024	The country where the researcher is currently active
current_position_2024	Current affiliation roles of researchers

Figure 1, displayed Azerbaijani researchers (with ORCID profile) migration over time, from 1990 to 2024. The overall number of researchers shows an increasing trend, with a peak in the year 2018. The total number of migrated researchers also experienced an increasing trend over time from 1990 to 2018, the number of researchers has dropped in 2019 and 2024.

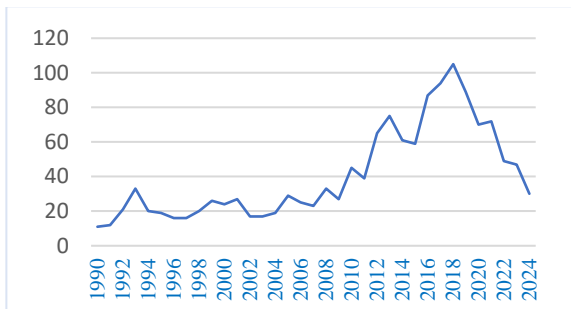


Fig.1. Intellectual migration in Azerbaijan (1990-2024) [6, 12].

The distribution of Azerbaijani researchers and scientists registered in the Orcid database by country is depicted in Figure 2. As can be seen from the graph, the developed countries, like Turkey, Germany, USA and Russia take the leading position in the process of “brain drain” from Azerbaijan.

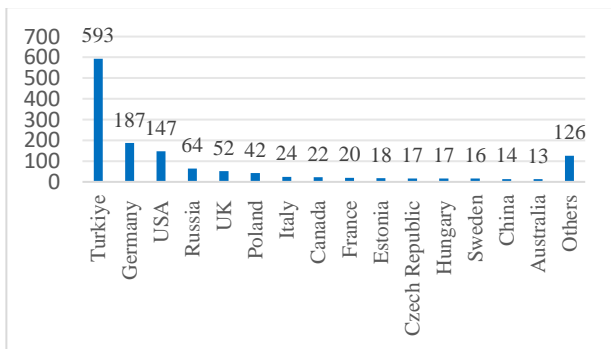


Fig. 2. Distribution Azerbaijani researchers and scientists registered in the Orcid database by countries [6, 12].

In terms of degree, the highest number of migrated researchers registred in ORCID is in the PhD degree (42%) and Master of Science degree (27%) (Fig. 3). The other degrees of Doctor of Science (15%) and PHD Student (16%) have fewer than the number of PhD student and MSc.

Also number of migrated non-PhD researchers registred in ORCID (58%) have more than the number of PhD researchers (42%) (Fig. 4).

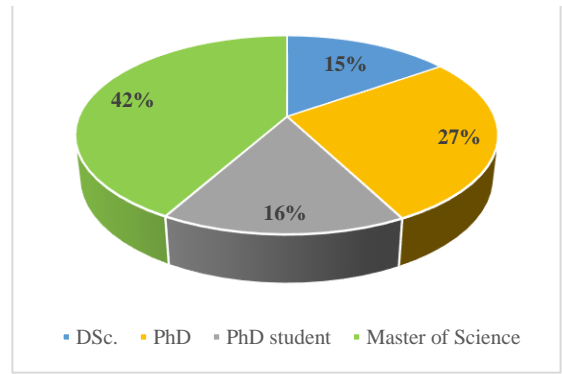


Fig. 3. Number of researchers registred in ORCID in terms of degree (%).

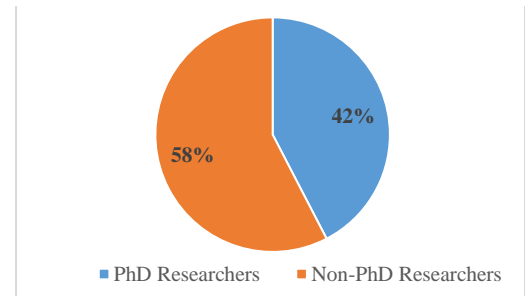


Fig.4. Number of migrated PhD researchers and non-PhD researchers registred in ORCID (%).

METHODOLOGY

This section presents the methodology for analyzing the migration patterns of researchers in Azerbaijan. The methodology includes several stages. The first step is to create dataset for study.

Selection Variables and Creating Dataset

Considering previous research works on migration, we collect a selection of variables we deem to be tied with the phenomenon and interesting from a policy standpoint.

1) *GDP per Capita*. We select GDP per capita as it is one of the most universally adopted and straightforward indicators of the overall economic wealth of a region. This metric is often used to compare economic well-being and living standards between countries. GDP per capita is the total value of a country's finished goods and services (gross domestic product) divided by its total population (per capita) [13].

2) *Education Index*. As an indicator of the overall schooling provision of a region we select the Education Index. This indicator is a component of the Human Development Index and it is defined as the geometric average of mean years of schooling and expected years of schooling [39]. The basis for the data is taken from Global Data Lab [14].

3) *Average monthly wages (\$)*. Averag monthly wages represent the typical earnings of workers in a specific area, influencing various economic indicators. They also represent a standard of living. This data is taken from UNECE Statistical Database, compiled from national and international (OECD, EUROSTAT, CIS) official sources. Gross average monthly wages cover total wages and salaries, before any tax deduction and before social security contributions [15].

We create a dataset taking into account migration data and parameters. The dataset covers intellectual migration data and selected parameter values for the period 1995-2023. An example record from a dataset is shown in Table 2.

TABLE 2. DESCRIPTION OF THE DATASET

Year	Number_mig	GDP_per_capita	Average_monthly_wages	Education_Index	
0	1995	19	1.098	14.2	0.649
1	1996	16	1.096	20.8	0.654
2	1997	16	1.144	35.5	0.658
3	1998	20	1.245	43.5	0.662
4	1999	26	1.324	44.8	0.667

A. Method

There are several statistical approaches for exploring relationships between two or more parameters. Regression analysis model is one of the statistical approaches which recommended for this kind of analysis. In the present study, to make the regressions more flexible, Multiple NonLinear Regression (MNL) models were applied. The general form of a MNL function can be represent (explained) by the following equation [16]:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3^2 + \beta_4 X_4^2 \dots + \beta_n X_n X_j$$

Where, Y is the dependent variable (e.g. migration flow for our study) and X_1, X_2, \dots, X_n are the independent variables. Also, α is the intercept and $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients of the independent variables while n is the number of observations.

One of the major problems of statistical analysis was to establish the appropriate relationship between the dependent variable and a set of independent variables. Different functions including the exponential, power, cubic, logarithmic, quadratic and linear functions are used to identify the best relation.

The MNL analysis was conducted in a Python programming environment using the Scikit-Learn library. The performance of the developed MNL models were assessed by statistical performance measures, Pearson correlation value (r) and the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and Mean Squared Error (MSE). The lower values of the RMSE, MAE, MSE will indicate the better performance of the model.

RESULTS

One key condition for justifying final results In multifactor correlation models is having a requirement of dependencies between factors. We check the Pearson correlation (Fig 7). The Pearson correlation is a statistical measure that assesses the linear relationship between two variables. A correlation matrix which created using Pearson correlation is shown in Table 3. Analysis of the pairwise correlation coefficient matrix shows that the number of migrants (the "Number of migration" dependent variable) has a strong relationship with the factors GDP per capita (0.770659) and Education Index (0.802589). At the same time, there is a strong relationship between GDP per capita (0.918665) and Education Index.

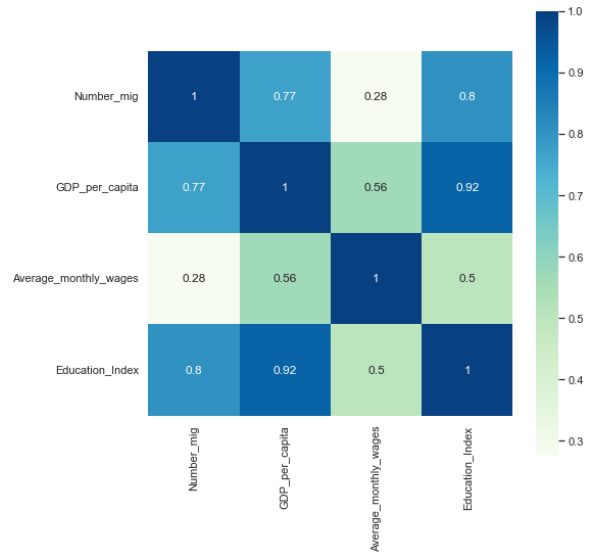


Fig. 7. Pearson correlation.

TABLE III. MATRIX OF CORRELATION COEFFICIENTS BETWEEN VARIABLE

	Number of migration	GDP per capita	Average monthly wages	Education Index
Number of migrations	1			
GDP per capita	0.770659	1		
Average monthly wages	0.276535	0.563194	1	
Education Index	0.802589	0.918665	0.498175	1

The ability of the best MNL model to predict future migration flows has been demonstrated flows in Figure 8.

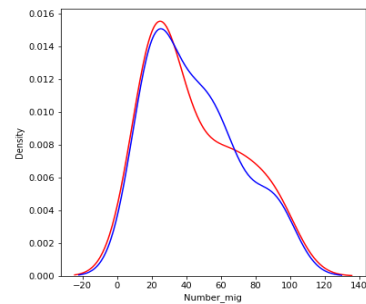


Fig. 8. Comparison between the observed (blue line) and simulated (red line) migration flows.

Some differences can be identified between the observed and predicted (simulated) migration flow. The reason can be that some other indices might have been more influential at that time rather than the selected indices in this study.

The results showed that, our model exhibited the best performance, with an R2 value of 0.893, Adj.R2-Squared value of 0.833, RMSE value of 8.544, a MAE value of 6.644, and MSE value of 73.005, indicating a very close match between predicted and observed values (Table 4).

TABLE IV. THE PERFORMANCE METRICS OF THE MNL MODEL

Metrics	Values
R2-Squared	0.893
Adj.R2-Squared	0.833
RMSE	8.544
MSE	73.005
MAE	6.644

The formation of the multiple non-linear equation can be explained by the following equation where cubic functions are found to be more suitable for the indices based on the correlation values.

$$M = \alpha + \beta_1 AVR^3 + \beta_2 AVR^2 + \beta_3 AVR + \beta_4 GDP^3 + \beta_5 GDP^2 + \beta_6 GDP + \beta_7 EI + \beta_8 AVR * GDP + \beta_9 AVR * EI + \beta_{10} GDP * EI$$

Where, M denotes migration, AVR denotes Average monthly wage, GDP denotes GDP per capita, and EI denotes Education index.

The regression coefficients of the model are obtained using optimization methods such as the least squares method. The least squares method consists in finding such values of the coefficients that minimize the sum of the squares of the forecast errors. Considering the values of the regression coefficients of the model, the predictor model is as follows:

$$M = 383.417 - 0.001AVR^3 + 0.003AVR^2 - 2.326AVR + 4.048GDP^3 - 1.272GDP^2 + 11.244GDP - 558.110EI - 0.659AVR * GDP + 5.339AVR * EI - 38.826GDP * EI$$

By comparing the coefficients obtained for individual factors in this model, it can be seen that the impact of “GDP per capita” on intellectual migration is higher.

CONCLUSION

In this study, non-linear regression model were developed with the combination of different indices (GDP, Average, Education index) to evaluate intellectual migration flows in Azerbaijan. The performance of the developed MNLR model were assessed by statistical performance measures. The results showed that, our model exhibited the best performance, with an R2 value of 0.893, Adj.R2-Squared value of 0.833, RMSE value of 8.544, a MAE value of 6.644, and MSE value of 73.005. The use of digital footprint data allows us to overcome some of the limitations of traditional approaches. The development of this area can contribute to the improvement of existing migration statistics, as well as the creation of new models and the development of new concepts related to population migration. To explore the

better predictive skills for migration forecast more sophisticated models will be included in the extension of this research study.

REFERENCES

- [1] V Bosetti, C Cattaneo, E Verdolini, “Migration of skilled workers and innovation: A European perspective,” *Journal of International Economics*, vol. 96 (2), pp. 311-322, 2015.
- [2] Sanliturk, E. Zagheni, M.J. Danko, T. Theile, and A. Akbaritabar, “Global patterns of migration of scholars with economic development,” *PNAS*, 2023, vol. 120, no. 4. <https://doi.org/10.1073/pnas.2217937120>
- [3] J. El-Ouahi, N. Robinson-García and R. Costas, “Analyzing scientific mobility and collaboration in the Middle East and North Africa.” *Quantitative Science Studies*, vol.2(3), pp. 1023–1047, 2021. https://doi.org/10.1162/qss_a_00149
- [4] S. Orazbayev, “International stocks and flows of students and researchers reconstructed from ORCID biographies,” *MPRA Paper No. 79242*, April 6, 2017. [Online]. Available: <https://mpra.ub.uni-muenchen.de/79242/>
- [5] S. J. Porter, “Measuring ORCID Research Information Citizenship, *Frontiers in Research Metrics and Analytics*,” March 2022, vol. 7:779097, www.frontiersin.org, doi: 10.3389/frma.2022.779097
- [6] J. Bohannon, K. Doran, “Introducing ORCID,” *Science*, 2017, vol. 356(6339), pp. 691-692. <http://dx.doi.org/10.1126/science.356.6339.691>
- [7] H. Kim, “Global mobility of the recent STEM postdoctoral workforce registered in ORCID,” *Quantitative Science Studies*, December 2024, vol. 6, no. 2., pp.1-23. DOI:10.1162/qss_a_00341
- [8] J. Gomez, A. C. Herman and P. Parigi, “Moving more, but closer: Mapping the growing regionalization of global scientific mobility using ORCID,” *Journal of Informetrics*, 2020, vol. 14(3), 101044. <https://doi.org/10.1016/j.joi.2020.101044>
- [9] Subbotin, S. Aref, “Brain Drain and Brain Gain in Russia: analyzing International Migration of Researchers by Discipline using Scopus Bibliometric Data 1996-2020,” *Scientometrics*, vol.126, pp.875–7900, 2021. <https://doi.org/10.1007/s11192-021-04091-x>
- [10] J. Ghirri, M. Mastropietro, S. Vantini, F. Ieva, M. Fontana, “An Evaluation of Researchers’ Migration Patterns in Europe using Digital Trace Data,” *arXiv:2302.07764v1 [stat.AP]* 15 Feb 2023, pp.1-34
- [11] Raghupathi, J. Ren, W. Raghupathi, “Exploring the Nature and Dimensions of Scientific Mobility: Insights from ORCID Database - A Visualization Approach,” *International Journal of Technology Diffusion*, vol. 14, iss. 1, pp. 1-31. January 2023. DOI:10.4018/IJTD.331090
- [12] ORCID: Connecting research and researchers, <https://orcid.org/orcid-search/search?searchQuery=Azerbaijan>
- [13] GDP per capita (currentUS\$)-Azerbaijan, <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations>
- [14] Educational Index, <https://globaldatalab.org/shdi/table/edindex/>
- [15] Gross Average Monthly Wages by Indicator, Country and Year, UNECE Statistical Database, <https://w3.unece.org/PXWeb2015/web/>
- [16] R.I. Eshaa. and M.A. Imteaza, “Non-linear multiple regression analysis for predicting seasonal streamflow using large scale climate mode,” *22nd International Congress on Modelling and Simulation*, Hobart, Tasmania, Australia, 3 to 8 December 2017, pp. 1586-1589. mssanz.org.au/modsim