Akira Soto Nishimura and Mathias Czaika

Migration Drivers Database

Deliverable 5.6

QuantMig has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

History of changes

Version	Date	Changes
1.1	16 December 2021	Internal review
1.2	20 January 2022	Submission to EC
1.3	11 February 2022	Dissemination to public

Suggested citation

Soto Nishimura, A. and M. Czaika (2022) Migration Drivers Database, QuantMig Deliverable 5.6, Danube University Krems, AT

Dissemination level

PU: Public

Acknowledgments

This document reflects the authors' views, and the Research Executive Agency of the European Commission are not responsible for any use that may be made of the information it contains. We are very grateful for insightful comments and advice from Jakub Bijak and Heidrun Bohnet on earlier versions of the database and documentation.

Cover photo: iStockphoto.com/Guenter Guni

Table of Contents

List of Tables

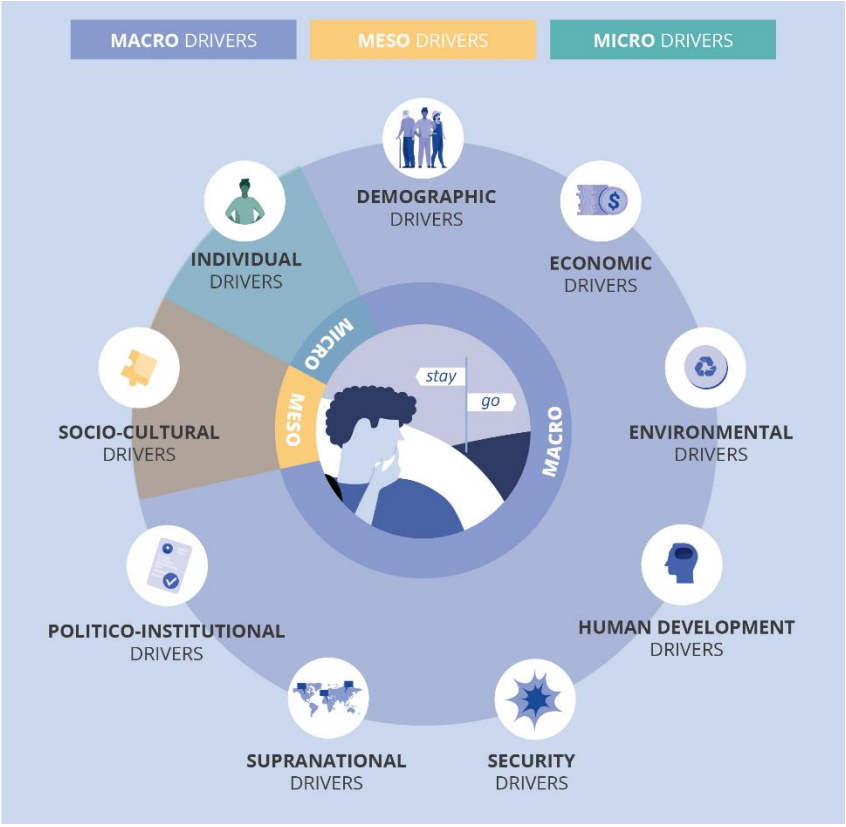
1. Introduction

This document is the accompanying documentation to the QuantMig driver database now available online at http://www.quantmig.eu/migration_driver_inventory/. The primary aim of the QuantMig Migration Driver Database is to provide a relevant resource for users to see not only what drivers of migration have been considered and operationalized but what data sources have been used to operationalize the drivers of migration. This database will also show users what countries and what years a dataset covers as well as other information such as how accessible the data is. A secondary benefit of this database is the direction and significance ($p<.05$) of specific variables on migration that were analysed in the articles were also recorded. This information can be found in the "Effect" column where a variable is marked with one of five outcomes "positive", "negative", "significant positive", "significant negative", and "NA". Positive meaning the variable increased the likelihood of migration. Therefore, users can gain a snapshot of what are the most common drivers of migration studied and what effect direction and with what frequency are these drivers found to be significant. To our knowledge this is the first database that focuses on drivers of migration and provides the aforementioned benefits.
Migration drivers are factors that influence migration decisions, and on aggregate, shape broader population movements by enabling, facilitating, triggering, constraining, or preventing migration. Migration drivers affect not only the likelihood of migration as a behavioural choice, but also inter alia the salience of certain migration routes, and the desirability of destinations. Migration drivers rarely operate in isolation but in combination with other structural factors, which *conjointly* create complex migration driver *environments* (Czaika and Reinprecht 2020).

Over the past decades, research has produced solid evidence on the forces that initiate and perpetuate migration. Researchers aim to understand (i) the relevance, relative importance and interactions of certain drivers, (ii) which driver configurations prevail under what circumstances, and (iii) to identify which drivers are most susceptible to be shaped by policy interventions. A review of existing theoretical and empirical research on migration shows that a substantive body of research has been accumulated and provides a solid understanding on the functional relationship between some specific migration drivers and migration outcomes, even though most studies struggle to grasp the migratory process in its entirety and to fully understand the main structural, macro-level drivers of migration, and in particular, how they interact (Czaika and Reinprecht 2020). Micro-level studies that usually focus on migration intentions, decisions and actual behaviour are often limited as they usually do not address under what conditions do people develop aspirations to migrate and are able to realize them.

In a recent review of the rapidly growing migration driver literature, Czaika and Reinprecht (2020) distinguish nine broader migration driver dimensions, namely demographic, economic, environmental, human development, individual, politico-institutional, security, socio-cultural, and supranational drivers operating at micro-, meso-, and macro-levels (Figure 1).

Figure 1 Fundamental migration driver dimensions



Source: IOM (2021), after Czaika and Reinprecht (2020), reprinted by a kind permission and available at https://www.migrationdataportal.org/themes/migration-drivers

These broad driver dimensions are further disaggregated into 24 driver factors (Czaika and Reinprecht 2020). Based on the review of hundreds of empirical migration drivers analyses which are covered and categorised in the QuantMig Migration Driver Database, we have identified more than 150 specific drivers in this large body of studies (Table A-1). The QuantMig Migration Driver Database therefore provides a unique resource of information about alternative driver specifications as employed in empirical (quantitative) studies exploring the relevance and effects of a large range of driving factors and specific drivers on migration outcomes at individual and aggregate levels. The concrete methodology applied for establishing the QuantMig Driver Database is introduced in the next section.

2. QuantMig Migration Driver Database: methodology

The articles considered for inclusion in this database come from an initial selection done by Czaika and Reinprecht (2020) where they assessed 660 English-language research documents related to migration drivers. Czaika and Reinprecht (2020) write "A key selection criterion was that these studies present novel empirical evidence or were influential in the migration studies field and/or come from respected organisations". These documents appear in peer-reviewed journals, books, reports, and working papers and were identified using various search engines (such as Google Scholar and Scopus), authors' literature databases, documents' cross-references, and through an expert workshop. From this initial collection all articles published from the year 2000 onward were considered for inclusion into the database. Articles were excluded if they were either theoretical only or did not use any large-N quantitative datasets, or the data used came from small sample

sizes as is often the case for qualitative studies that rely on interviews and focus groups. While a strict rule for a specific sample size limit was not used, generally the sample size for surveys needed to be in the hundreds to perform quantitative analysis. Additionally, articles that did some statistical (regression) analysis on the effects of drivers were prioritized over articles that only did a correlation or percentage breakdown. This is not to say that articles were strictly excluded if no statistical analysis was done. The datasets indicated in the database are reflective of what was used in the articles but not always an exact representation of what was used. The database reflects the most recent version of a dataset. For example, an article may have used the Correlates of War 3.0 dataset, but the entry in the database will reflect the characteristics of the more recent Correlates of War 4.0 dataset.

This database includes datasets that have been used in 176 articles published between 2000 and 2019[1]. While many datasets are included in this database and reflect what has been used in the past this database is not fully comprehensive of all datasets that could be used to operationalize a driver of migration. This database could have attempted to include every possible dataset that could contain information on a possible driver of migration but chose instead to collect information on past migration driver analyses. This decision was both a practical one on the authors' side but also one for users of the database. On the author' side the amount of time that would have been required to find and include hundreds of datasets was prohibitive. On the users' side the number of datasets could easily become overwhelming. For instance, the World Development Indicators[2] database alone contains over 900 variables, and the Quality of Governance[3] dataset contains over 1900 variables from more than 100 datasets.

The decision to only include datasets used in past migration driver analyses means there are some datasets that are not included in this database but bear mentioning. Migration policy is a driver that has been included in several articles. However, several migration indexes such as the Migrant Integration Policy Index[4], Civic Integration Policy index[5], Migration Policy Index[6], Immigration Policies in Comparison[7], and Multiculturalism Policies in Contemporary Democracies[8] are not in the database as they were not used in the articles reviewed. Similarly, there are numerous indices of democracy that are not captured in this database. By far the most popular index of democracy used in the articles was the Polity dataset from the Center for Systemic Peace despite serious criticism of this index (Mchenry & Mady, 2006) and evidence that there is considerable variation among the popular democracy indexes (Elff, M., & Ziaja, 2018; Högström, 2013). This is potentially a serious issue if the theoretical drivers of migration such as democracy are not accurately captured by their operationalization. Any conclusion drawn about democracy may be inaccurate if

[1] One article was published in 2020 but was available in 2019
[2] https://databank.worldbank.org/source/world-development-indicators
[3] https://www.gu.se/en/quality-government/qog-data
[4] https://www.mipex.eu/
[5] Wallace Goodman (2010) Integration Requirements for Integration's Sake? Identifying, Categorising and Comparing Civic Integration Policies, *Journal of Ethnic and Migration Studies*, 36:5, 753-772, DOI: 10.1080/13691831003764300

[6] https://users.ugent.be/~sastanda/MPI/MPI.html
[7] http://www.impic-project.eu/data/
[8] https://www.queensu.ca/mcp/about

the index is not reflective of democracy but something else such as neoliberalism.

Just as this database does not include every conceivable dataset, it does not contain every conceivable driver of migration.[9] Some variables were excluded from the database in cases where the variables were highly specific to the context of the analysis such as a dummy variable for living in a specific city. While we have some confidence that the important variables of migration appear in this database there may be variables that appear less frequently than their relative importance or variables that can operationalized in new ways. Anti-immigrant attitudes is an example of a variable that appears infrequently in the database. The health care infrastructure of a country provides another example where in the articles reviewed the number of doctors was the main method of operationalization, but an alternative operationalization that did not appear are the number of hospital beds.

Articles often contain multiple models where the significance and effect direction of a variable can differ across the models. For each article only one value was recorded per variable. The values recorded were either from the main model as indicated by the author of the article or the model with the most variables. If a variable did not appear in the main model but did appear in a different module such as in a robustness check, then the value recorded would be from that model. In some cases, the analysis was split by gender, country, or race/ethnicity. In such cases the results from the male sample were recorded while for country and race/ethnicity splits the value recorded was for the group whose regression table appeared first or in some cases where there were many split analyses the majority value was taken. In articles with multiple dependent variables such as internal and international migration or migration intention and migration behaviour the value recorded was always international over internal and behaviour over desire. In some cases, the dependent variable did not specify internal or international migration and in these cases, they were labelled as international migration. In the database a variable linked to a particular article may appear more than once because an article may have used multiple data sources to construct the variable.

Binary variables should be read that the name of the variable would represent the non-reference group. If the binary variable is "own home" and the effect is positive, this means owning a home has a positive effect on the likelihood of migration. For non-binary variables the direction should be read that an increase in the variable leads to an increase in the effect. For dyadic variables such as linguistic distance the variable should be read similarly in that an increase in linguistic distance leads to increase in the effect. In some cases, recording an effect direction or significance value did not make sense.

3. The Design of the Migration Driver inventory

In the following, we list the categories (columns in the database) as they are browsable in the database and the definitions of these categories:

[9] Also, the database does not include any interactions or transformed variables. In the case of lagged variables, they were not included either when the non-lagged variable was also included in the analysis. If only a lagged version of the variable was included in the analysis, then the lagged variable was included in the database.

3.1 Migration driver typology

1. *Driver dimension*- The broadest category is the nine driver dimensions *demographic, economic, environmental, human development, individual, politico-institutional, security, socio-cultural, and supranational*. For a comprehensive description of these categories see Czaika and Reinprecht (2020).
2. *Driver factor*- The driver dimensions are further categorized into 24 driver factors. See table A-1 for the which driver factor belongs to which driver dimension. Again, for comprehensive description of these categories see Czaika and Reinprecht (2020).
3. *Specific Driver*- This column is meant to create labels more specific than the driver factor but more manageable than the one thousand plus variables. In total there are 162 drivers that group variables similar in concept. To illustrate, in the *Variable* column there are several concepts related to population such as population growth rate, population 20-29, population 20-34, population 15-64, etc. These are all given the same label of population in the *Driver* column. See table A-1 for the which specific driver belongs to which driver factor.
4. *Driver ID*- A code that identifies the driver dimension with a single digit (1-9), the driver factor with a letter and specific driver with a number (001-162)
5. *Variable*- The labels in the variable column are often, but not always a one-to-one representation of the variable in the article. Where the *Driver* column has the label dependent children the *Variable* column may have a more detailed label such as dependent children under 12. Changes were made to the names of some variables to give very similar variables in concept but with different names the same name.
6. *Variable perspective*- The variable used in the study was either an individual, origin, destination, or dyadic level variable. For example, the effect of GDP could be the effect of origin country GDP, destination country GDP, or a dyadic measure such as the difference in GDP between origin and destination. An example of an individual level variable would be owning a business.

3.2 Basic characteristics of migration data

7. *Type of migration* – Was the dependent variable concerned with international, internal, or return migration?
8. *Variable scale type*- Is the variable numeric, binary, or categorical?
9. *Data aggregation level*- Is the data available at the micro (individual), subnational, or country level? In some cases, data is available at the micro-level but difficult to obtain. In these cases, the label given depended in part on whether the article used micro-level data or not?
10. *Data type*- Three data types (survey, index, administrative) and 'unclear' are the possible labels in this column. Differentiation between the three types of data is not always clear. The nature of the variable, dataset, data provider, and data source were all considered in the selection of the label.

3.3 Data provider and source

11. *Dataset name*- Name of dataset as described by the data provider or citing article. It is not always possible to determine from the article the data used. In such cases the dataset name is marked as unclear and the name of the authors of the article.
12. *Data provider*- What institution provides the data?
13. *Data source*- If the providers and the source are not the same, where the data providers get

the raw data from. In some cases, the data source is either so numerous or ambiguous enough that they are marked as unclear.

3.4 Coverage

14. *ISO3*- Which countries are covered? In some cases, a country or territory does not have an ISO3 code and in these cases the name used in the dataset is kept.
15. *Year* - The temporal coverage in terms of for which years data has been collected.
16. *Country coverage*- Number of countries covered grouped into the categories single (one), ten or less, eleven to one hundred, more than one hundred, unclear.
17. *Availability* – Whether the data is *open* meaning available for free and readily available, *free application* meaning an account is required but the creation of the account is free and quick, *application* meaning that either a fee or extensive application for use is required, *unclear* meaning that the data used was not found or availability is not clear. This is often the case with smaller survey data where obtaining the data may depend on the whims of the researcher who collected the data.
18. *Update status* - Whether the data is still updated or not with data of more recent years.

3.5 Additional information

Additional columns contain information on the year the article was published, the reference for the article, links to the article, links to the dataset used, dataset documentation, and notes from the creators of this database.

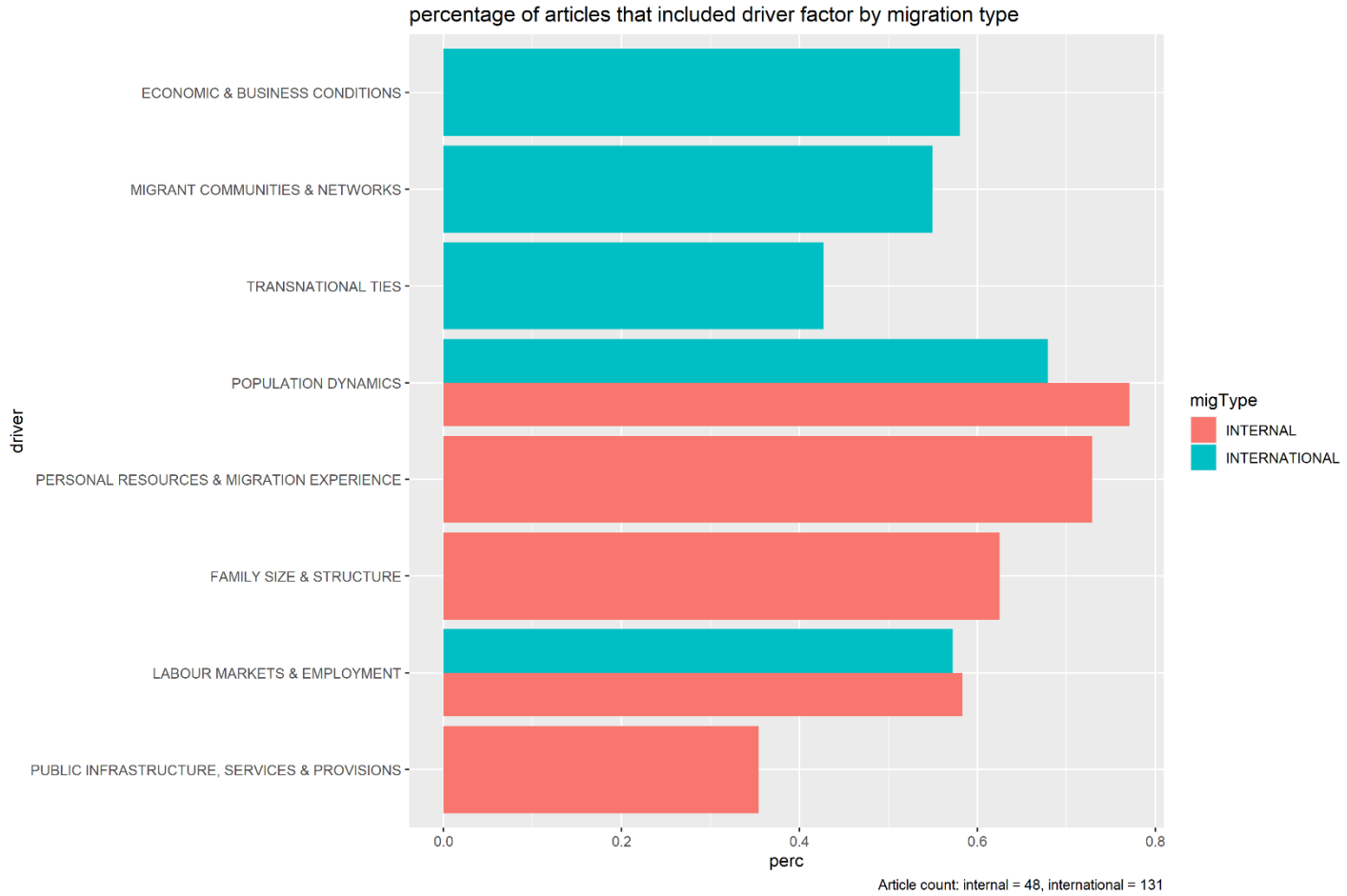
3.6 Direction and Significance of Driver Effect

The "Effect" column provides information on whether in the analysis done in the cited study the variable in question was found to have a positive, positive significant, negative, or negative significant effect. A variable was considered significant at $P<.05$. Positive means an increase in the likelihood of migration. In this column the label NA is used when denoting effect direction and significance was not possible.

4. A descriptive analysis of migration drivers

Figure 2 shows how common driver factors are by type of migration. Population dynamics is a common driver factor in articles that analysed international migration and ones that analysed internal migration. That driver factor was found in over 60% of the articles whereas the driver factor labour markets and employment were relatively common in both types of migration but less so than population. Outside of these two drivers factors no other driver factors are in the top five in appearance for both types of migration.

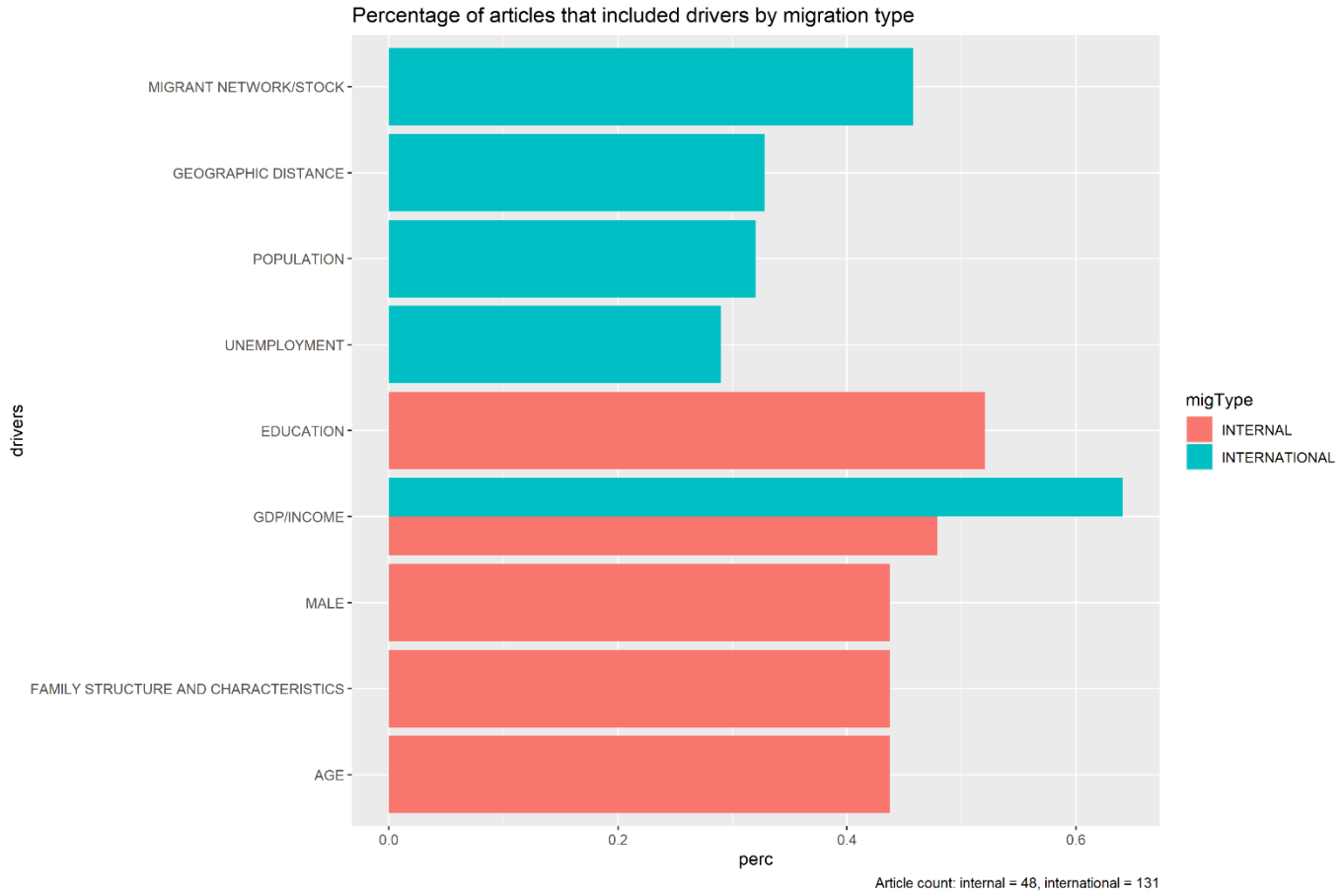
Figure 2 Relative frequency of driving factors in N=176 studies



Source: Author's own calculation based on Soto Nishimura (2022)

Regarding more specific drivers, Figure 3 shows the relative frequency of the most prominent drivers as used in studies on internal and international migration, respectively. Drivers related to GDP and income were common in both internal and international studies. It was the most driver in international migration which makes since as it as driver that can operate on individual level as well as destination, origin, and dyadic level. Moreover, GDP data specifically is easily attainable for many countries and over a long time period.

Figure 3 Relative frequency of drivers in N=176 studies



Source: Author's own calculation based on Soto Nishimura (2022)

Table 1 shows how often driver factors were found to be significant. For the category of driver factors focusing on the direction does not make sense as many variables with different expected directions fall under the same driver factor such as the factor Transnational ties where a common legal origin would be expected to have a positive effect while linguistic distance would be expected to have a negative effect. However, some insight can be gained by focusing on significance. The number of appearances means how many times did this driver factor appear. The second row in the table shows that the driver factor "Urban / rural development & living standards" appeared 24 times in studies on internal migration was found to be statistically significant (P<.05) 63% of the time. It is important to note that this number is not weighted by the number of appearances by study meaning that one study could account for 9 of the 24 appearances and another study account for only 2.

Table 1 Migration drivers and their robustness in showing significant effects on migration

Driver factor	Internal (# of appearances)	Internal [%]	International (# of appearances)	International [%]
Urban / rural development & living standards	24	63	9	56
Transnational ties	4	75	159	75
Public infrastructure, services & provisions	36	50	44	55
Poverty & inequality	6	67	18	61
Population dynamics	98	52	183	66
Political situation, repression & regime transitions	2	100	23	57
Personal resources & migration experience	109	47	122	48
Natural disasters & environmental shocks	4	50	26	15
Migrant communities & networks	5	40	107	76
Migrant aspirations & attitudes	46	52	45	56
Labour markets & employment	47	34	115	50
Globalisation & (post)colonialism	3	33	45	49
Gender relations	5	20	4	100
Family size & structure	100	50	64	41
Education services & training opportunities	11	73	12	67
Economic & business conditions	20	35	138	50
Cultural norms & ties	17	47	11	55
Climate change & environmental conditions	43	44	26	27
Civil & political rights	2	50	31	48
Migration policy & other public policies	0	0	70	67
International relations & geopolitical transformations	0	0	9	78
Health services & situation	0	0	9	78
Conflict, war, & violence	0	0	50	44

Source: Author's own calculation based on Soto Nishimura (2022)

Focusing on the driver factors that appeared at least fifty times, the driver factors transnational ties and migrant communities & networks for international migration were significant 75 and 76 percent of the time, respectively. The driver factor conflict, war, & violence was significant less than half of the time. For internal migration only the driver factors population dynamics, personal resources & migration experience, and family size & structure appeared over 50 times. Of these three factors, population dynamics was significant 52 percent of time and that was the highest value of the three factors. To speculate, the relative infrequency of significance for some of the driver factors may indicate weak theoretical justification for their inclusion or the much more likely case is the

operationalization of the variable does not correspond sufficiently to the theoretical construct. It is surprising to see that the driver factor Conflict, war, & violence is significant only 44 percent of the time. The theory behind including variables related to this driver factor is obvious. There does appear to be a need for improvement in practice on the part of researchers of migration drivers to better match operationalization with theory.

More in-depth analyses will be provided in a separate background paper accompanying the migration driver database.

References

Czaika, M., & Reinprecht, C. (2020). Drivers of migration. A synthesis of knowledge. *IMI (International Migration Institute) working paper No. 163*.

Elff, M., & Ziaja, S. (2018). Method factors in democracy indicators. *Politics and Governance*, *6*(1), 92-104.

Högström, J. (2013). Does the choice of democracy measure matter? Comparisons between the two leading democracy indices, Freedom House and Polity IV. *Government and Opposition, 48(2), 201-221.*

IOM (2021) Migration Data Portal, *https://www.migrationdataportal.org/themes/migration-drivers*

Mchenry Jr, D., & Mady, A. F. (2006). A critique of quantitative measures of the degree of democracy in Israel. *Democratization*, *13*(02), 257-282.

Soto Nishimura A (2022) Migration Drivers Data Inventory Records, 2000-2019: Data Inventory. Online resource, available at https://www.quantmig.eu/migration_driver_inventory/

5. ANNEX

Table A-1 List of Specific drivers

Driver Dimension	Driver Factor	Specific Drivers
Demographic	Population Dynamics	Urban/Density, Education of Population/Enrolment, Ethnicity/Race/Religion, Replacement Rate, Military Personnel, Population, Old Age Dependency Ratio, Male, Fertility Rate, Demographic Pressure, Age
	Family Size & Structure	Single/Divorced/Widowed, Family Structure and Characteristics, Married/Partnered, Dependent Children, Other Family Dependents, Education Other Household Members, Acceptance of Partner Living Apart
Economic	Economic & Business Conditions	GDP/Income, Taxes, Socio-Historical Context of Community, Agriculture Production, Economic and Labour Expectations, Outstanding Household and Business Loans To GDP, Inflation Rate, Economic Conditions/Activity, Devaluation of Currency, Credit Information Distance, Labour Market and Occupations, Market Openness/Competing Index, Capital Per Worker, Business Cycle
	Poverty & Inequality	Economic Poverty/Inequality/Deprivation
	Labour Markets & Employment	Unemployment, Individual Labour Status/Occupation Others, Individual Labour Status/Occupation, Union Coverage, Stock of Employment by Country, Industry Growth/Size, Private Returns to Schooling, Economic and Labour Expectations, Labour Market and Occupations, Labour Market Index, Labour Market Gap, Labour Force, Replacement Rate, Employment Growth, Taxes
	Urban / Rural Development & Living Standards	Standard of Living, Consumption, Undernourishment, Media Access, Societal Problems, Urban/Density, Recreation Score, Ratio of Origin Community Attractiveness to National, Percent Males 20-24 Living at Home, Housing Values, Cost of Living, Agriculture Production
Environmental	Climate Change & Environmental Conditions	Climate, Irrigation/Soil, Topographic Characteristics, Environment Quality, Agriculture Production
	Natural Disasters & Environmental Shocks	Natural Disasters
Human Development	Education Services & Training Opportunities	Education Quality, Education Institutions and Characteristics, Education Expenditure, Literacy, Education of Population/Enrolment, Education Cost of University
	Health Services & Situation	Health Facilities/Doctors, Life Expectancy at Birth, Infant Mortality Rate, Human Development Index
Individual	Personal Resources & Migration Experience	Individual/Household Material Assets, Individual Labour Status/Occupation, Work Experience, Access To Social Services, GDP /Income, Education, Education Area Of Study, Migration Experience/Experience Traveling, Family/Friends Migration Experience/Experience Traveling, Information About Destination, Economic/Social Status, Social Capital, Skill Gain From Migration, Economic Poverty/Inequality/Deprivation, Citizenship Status/Country Of Birth, Media Access, Ties To

	Migrant Aspirations & Attitudes	Destination, Language Skills, Natural Disasters, Consumption, Health Facilities/Doctors, Health Attitude/Personality, Economic Prospects, Migration Intention, Prospects Miscellaneous, Reduce Household Risk, Reason for Choosing Destination, Lifestyle Prospects, Migration Rate, Discrimination, Democracy, Confidence in Government and Government Institutions
Politico-Institutional	Public Infrastructure, Services & Provisions	Social Expenditure, Urban Development Aid, Confidence in Government and Government Institutions, Rural Development Aid, Distance/Access to Amenities/Roads, Government Effectiveness/Quality, Migration Costs, Crime/Policing, Corruption, Political Institutional Quality, Impartible Division of Land, Dispersion of Social Transfers,
	Migration Governance & Infrastructure	Political Institutional Quality, Legal Migrant
	Migration Policy & Other Public Policies	Migration Policy, Social Expenditure, Visa Policies and Practices, Asylum Policy, Employment Protection, Dual Citizenship, Deterrence Index,
	Civil & Political Rights	Democracy, Rule of Law, Political Regime Type, Civil and Political Rights
Security	Conflict, War, & Violence	Government Stability, Far Right Votes, Political Terror Score, Political Stability, Political Security Index, Percentage of Cabinet Portfolios Held by Left-Wing Parties
	Political Situation, Repression & Regime Transitions	Conflict/Violence, Safety, Terrorism Deaths and Events
Socio-Cultural	Migrant Communities & Networks	Ties to Destination, Ties to Origin, Asylum/Refugee Population/Applications, Migrant Network/Stock, Remittances, Migration Rate
	Cultural Norms & Ties	Migration Norm, Involvement in Origin Community, Migration Norm, Parent Wants Child to Continue Education, Discrimination
	Gender Relations	Gender Discrimination and Gender Roles, Female Labour Force Participation, Female Seats in Parliament,
Supranational	Globalisation & (Post)Colonialism	Colonial Link, Tourism, Market Openness/Competing Index, Neighbouring Countries in Top 10 Sending Countries, Longitude Difference, Landlocked, International Non-Governmental Organization Membership, Distance to Equator, Decolonization, Cultural Diversity, Anglophone
	Transnational Ties	Borders/Shared Regions, Trust Distance, Geographic Distance, Trade, Foreign Aid/Investment, Similar Legal Origin/Governments, Linguistic Distance/Shared Language, Shared Currency/Economic Community, Remittances, Religious Distance, Institutional Distance, Genetic Distance, Financial Development Distance, Exchange Rates, Cultural Distance, Migrant Network/Stock, Business Ties, Anglophone, Alliance,
	International Relations & Geopolitical Transformations	Country Size, EU (European Union) History,

Table A-2 Driver Dimension ID codes

Driver dimension	ID
Demographic	1
Economic	2
Individual	3
Socio-cultural	4
Politico-institutional	5
Security	6
Environmental	7
Human development	8
Supranational	9

Table A-3 Driver factor ID codes

Driver factor	ID
Civil & political rights	A
Climate change & environmental conditions	B
Conflict, war, & violence	C
Cultural norms & ties	D
Economic & business conditions	E
Education services & training opportunities	F
Family size & structure	G
Gender relations	H
Globalisation & (post)colonialism	I
Health services & situation	J
International relations & geopolitical transformations	K
Labour markets & employment	L
Migrant aspirations & attitudes	M
Migrant communities & networks	N
Migration governance & infrastructure	O
Migration policy & other public policies	P
Natural disasters & environmental shocks	Q
Personal resources & migration experience	R
Political situation, repression & regime transitions	S
Population dynamics	T
Poverty & inequality	U
Public infrastructure, services & provisions	V
Transnational ties	W
Urban / rural development & living standards	X