Variation in Migration Flows by Duration Intervals

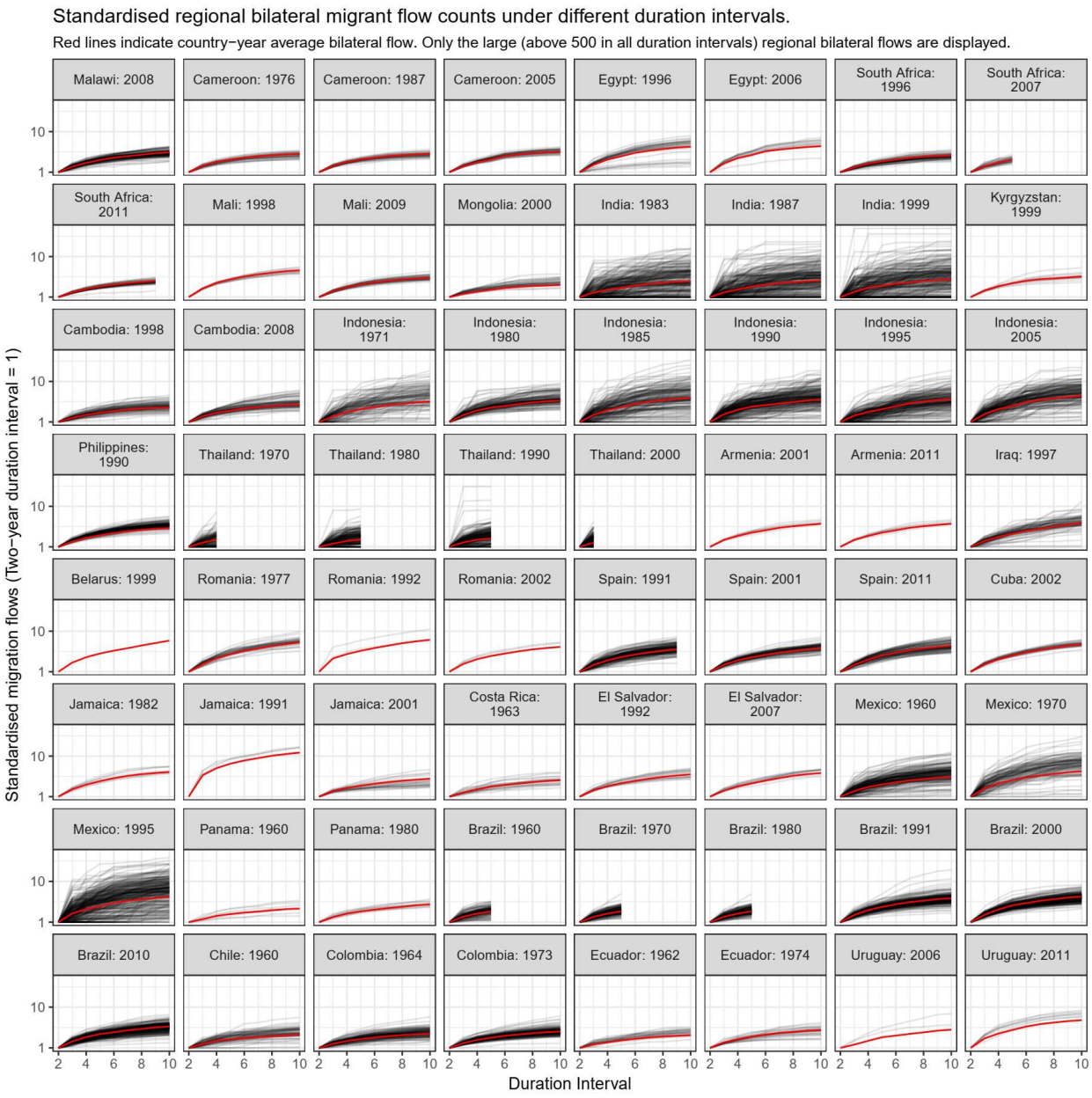
Guy J. Abel, Arkadiusz Wisniowski and Dilek Yildiz

1. Motivation

Migration data are often collected using fixed interval measures in censuses via questions on respondents' region of residence five years ago. However, different applications may require migration data at varying duration intervals; for example, single-year flows are necessary for single-age cohort component population projections. In addition, many countries only collect data based on one duration interval, prohibiting effective comparisons of internal migration systems between countries and international migration flows. Simple conversions of migration measures to equivalent values based on shorter measures are problematic (Kitsul & Philipov, 1980; Rees, 1977; Rogers, 1975). Research into the relationship between migration flows defined with different interval lengths has shown variations in volumes of migration depending on the migration context, such as the region of origin, destination and the migration corridor (Long & Boertlein, 1990; Newbold, 2005; Rogers et al., 2003). The primary aim of this research is to analyse the relationship between migration flow counts and the duration intervals used in their measurement. We then build a predictive model to transform migration flows collected using one measure (e.g., five-year intervals) to another (e.g., one-year intervals).

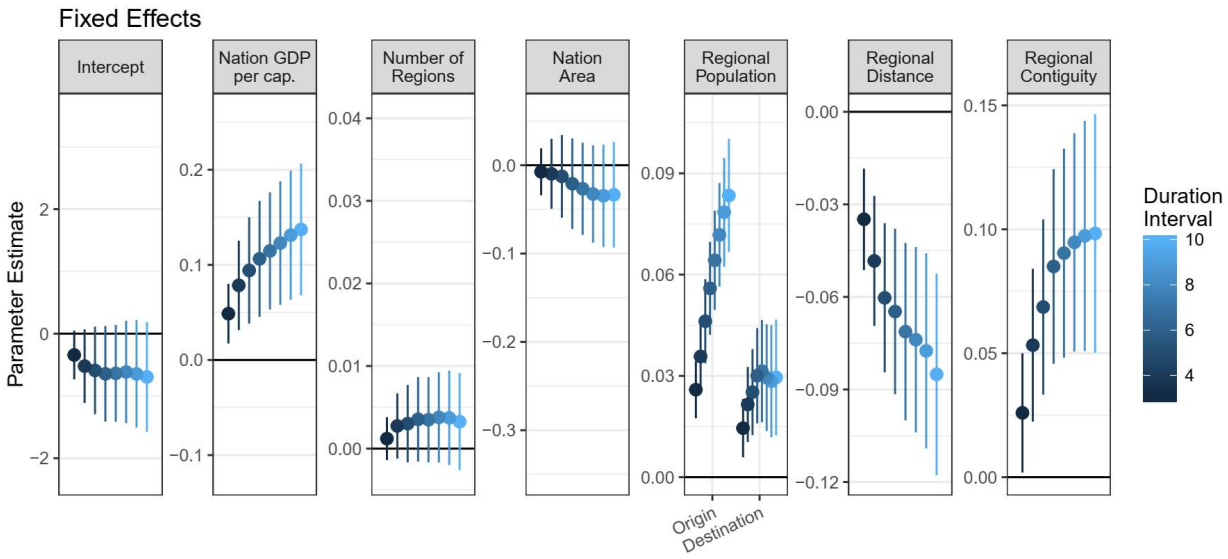
2. Data

The research is based on IPUMS International data (Minnesota Population Center, 2020) from countries with census questions on the duration of stay in the current residence, the locality of the previous residence (same or different region), and the previous region. Multiple flow counts for each migration corridor were derived by aggregating individual responses by their current and previous regions, where the previous region was determined based on varying duration intervals from the duration of stay response. These counts were standardised using a two-year flow as the base, illustrating the comparative rises in flow counts when longer intervals are used. These are displayed in the Figure below for bilateral (origin-destination) migration flow counts for internal migration from 64 censuses across 28 countries. We initially used the two-year flows as a base, as we found some discrepancies in the one-year intervals and the less than one-year intervals in the IPUMS International census samples. Further research will be carried out to clarify how IPUMS has harmonised data at the shortest interval levels.



3. Analysis by Duration Interval

In our initial study, we fitted eight multilevel regression models to the standardised migration flows for each interval. Separate models were fitted for data from each interval, ranging from 3 to 10, where complete duration curves are available (plotted above). Our analysis aimed to investigate differences in factors at the national and regional levels, as well as variations in corridors, countries, and time on the level of migration as the duration interval used to detect migration events is extended.



Key findings from our initial analysis include:

- Countries with higher GDP per capita are associated with larger internal migration flows at longer intervals.

- The number of regions and the geographic area of the country have no definitive association with changing levels of migration as the duration interval increases.

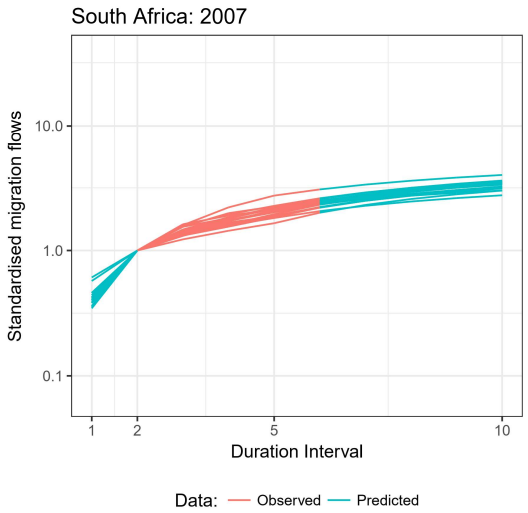
- Origin, destination, and corridor characteristics (distance and contiguity) are associated with different levels of migration as duration intervals increase.

4. Predictive Model

A predictive generalised least square model was developed based on the rise of duration curves (in the previous figure): $y_{ijk} = d_{ijk} - d_{ij(k-1)}$ from all migrant corridors (i, j) at duration interval k in over all available IPUMS International country-year data (c, t):

$$y_{ijkct} = \beta_0 + \beta_1 INTERVAL_k + \beta_1^O POP_{\{it\}} + \beta_1^D POP_{\{jt\}} + \beta_2 DIST\_\{ij\} + \beta_3 CONTIG\_\{ij\} + \beta_4 GDP\_\{ct\} + u_{ijkct}$$

The model was fitted with gls() function in R with AR(1) errors in each corridor $u_{ijkct} = \rho u_{ij(k-1)ct} + \varepsilon_{ijkct}$. The predictive model can be used to extend the standardised duration curves, y_{ijk} below to longer durations (when unknown). For example, a one-year migration flow based on data from the South Africa 2007 census can be derived as $d_{ij1} = d_{ij}\ y_{ij}$. Non-observed flows for longer intervals (k = 7, 8, 9 or 10) can also be derived in a similar fashion.



5. Summary and Future Work

In summary, our initial research reveals smooth, non-linear increases in the number of migrant transition flows as the duration interval increases. The shape of duration curves varies by national and regional characteristics. The developed predictive model enables the estimation of migration flows for various duration intervals where data are required but unavailable.

Further work is required to clean the data for the shorter duration intervals (less than one year and one year). Some of the existing data in IPUMS International for these shorter intervals are inconsistent, so they were dropped at this early stage. Ongoing work is being undertaken to refine a) models to describe the variation in migration levels by duration interval using factors related to broader development (e.g. HDI) rather than economic development and b) machine learning method methods for predicting migration by duration interval.

References

Kitsul, P., & Philipov, D. (1980). *THE ONE-YEAR - FIVE-YEAR MIGRATION PROBLEM* (Issue WP-80-81).

Long, J., & Boertlein, C. (1990). Comparing Migration Measures Having Different Intervals. *Current Population Reports, Series P-23: Special Studies No. 166*.

Minnesota Population Center. (2020). *Integrated Public Use Microdata Series, International: Version 7.3 [dataset]. Minneapolis, MN: IPUMS, 2020*. https://doi.org/10.18128/D020.V7.3

Newbold, K. B. (2005). Spatial scale, return and onward migration, and the Long-Boertlein index of repeat migration. *Papers in Regional Science*, *84*(2), 281–290. https://doi.org/10.1111/j.1435-5957.2005.00018.x

Rees, P. H. (1977). The measurement of migration, from census data and other sources. *Environment and Planning A*, *9*(3), 247–272. https://doi.org/10.1068/a090247

Rogers, A. (1975). *Introduction to Multiregional Mathematical Demography*. Wiley.

Rogers, A., Raymer, J., & Newbold, K. B. (2003). Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science*, *37*(4), 581–601.