# A comparison of bilateral migration counts by time intervals

Guillermo Vinué and Guy Abel

Wittgenstein Centre (IIASA, VID/ÖAW, WU),
Vienna Institute of Demography of the Austrian Academy of Sciences

British Society for Population Studies 2016 Conference
The University of Winchester

12-14 September 2016

# Outline

# Introduction (I)

- Inconsistent data is a common problem in cross national comparisons in migration research.
- One source of inconsistency is the timing criteria to define counts of migrants.
- Models to harmonise data based on differences in the timing criteria has tended to be limited to a single country (e.g. Courgeau (1973)) or to synthetic data (e.g. Nowok and Willekens (2010)).

# Introduction (II)

- Main goal: To develop a statistical methodology to explain the differences between counts of region to region migration over several time intervals, controlling for some other factors.

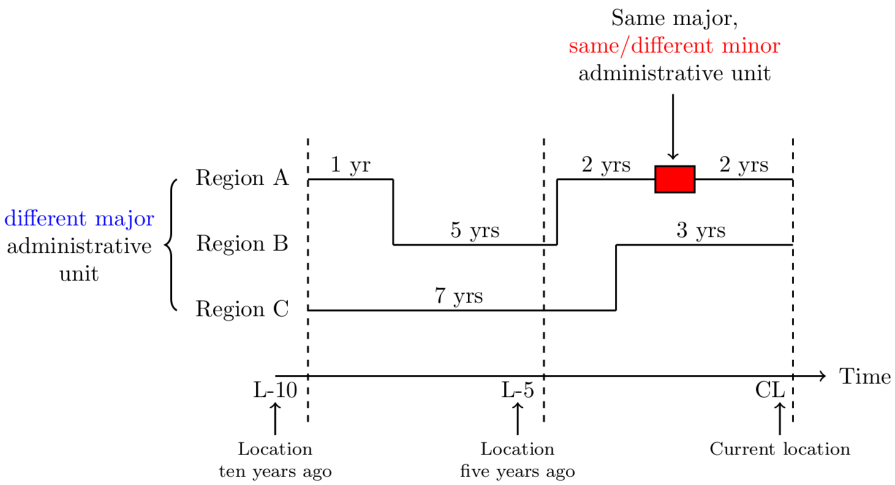
- Our approach builds on previous work by Garcia et al. (2014) and shares many of its advantages.

- Furthermore, our model should allow us to predict bilateral migrant flow counts where data are available for only a single timing criteria (such as one year or five years).

# Data

- IPUMS census microdata provides high quality anonymized samples of the underlying census data for each country.
- We study IPUMS data using combination of questions on:
  1. Length of stay at current residence (*migyrs1: Years residing in current locality*).
  2. Previous residence in a different region (*migratep: Migration status, previous residence*). Categories of *migratep* are as follows:
     - Same major administrative unit.
       - Same major, same minor administrative unit.
       - Same major, different minor administrative unit.
     - Different major administrative unit.
     - Abroad.
  3. District of previous residence.
- Generate flow tables based on origin one, two, three, ..., ten years ago.

# Some issues with migratep



# Countries and input data (I)

- Countries: Armenia (ARM) 2001 and 2011; Belarus (BLR) 1999; Haiti (HTI) 1971; Jamaica (JAM) 1982, 1991 and 2001; Panama (PAN) 1960 and 1980; Romania (ROU) 1992; Slovenia (SVN) 2002; Spain (ESP) 1991, 2001 and 2011.
- Dependent variable: the observed migration flows between regions in different countries.
- Socioeconomic predictors: Variables known to have relationships to migration, such as:
  - Primary and secondary education completed.
  - Proportion of people economically active.
  - Proportion of males.
  - Median age.
- Geographical predictors: IPUMS shapefiles to calculate region areas, distance and contiguity between regions.

# Countries and input data (II)

- Input data frame looks like this:

| country | orig | dest | flow | year | popi | malei | medagei | activei | comp_primi | comp_secondi |
|---|---|---|---|---|---|---|---|---|---|---|
| ARM | aragatsotn | ararat | 220.00 | 2001 | 143830.00 | 0.49 | 28 | 0.49 | 0.87 | 0.54 |
| ARM | aragatsotn | armavir | 300.00 | 2001 | 143830.00 | 0.49 | 28 | 0.49 | 0.87 | 0.54 |
| ARM | aragatsotn | kotayk | 130.00 | 2001 | 143830.00 | 0.49 | 28 | 0.49 | 0.87 | 0.54 |
| ARM | aragatsotn | lori | 20.00 | 2001 | 143830.00 | 0.49 | 28 | 0.49 | 0.87 | 0.54 |
| ARM | aragatsotn | shirak | 120.00 | 2001 | 143830.00 | 0.49 | 28 | 0.49 | 0.87 | 0.54 |
| ARM | aragatsotn | syunik | 70.00 | 2001 | 143830.00 | 0.49 | 28 | 0.49 | 0.87 | 0.54 |

| popj | malej | medagej | activej | comp_primj | comp_secondj | dist | cont | areai | areaj | interval |
|---|---|---|---|---|---|---|---|---|---|---|
| 281810.00 | 0.49 | 29 | 0.59 | 0.85 | 0.53 | 0.79 | 0 | 0.31 | 0.23 | 1 |
| 277140.00 | 0.49 | 29 | 0.53 | 0.88 | 0.56 | 0.32 | 1 | 0.31 | 0.14 | 1 |
| 284500.00 | 0.48 | 29 | 0.38 | 0.86 | 0.57 | 0.56 | 1 | 0.31 | 0.21 | 1 |
| 303960.00 | 0.48 | 31 | 0.39 | 0.86 | 0.57 | 0.65 | 1 | 0.31 | 0.38 | 1 |
| 288950.00 | 0.48 | 30 | 0.38 | 0.88 | 0.58 | 0.47 | 1 | 0.31 | 0.32 | 1 |
| 149360.00 | 0.49 | 31 | 0.88 | 0.88 | 0.59 | 2.30 | 0 | 0.31 | 0.47 | 1 |

- Summary statistics for the variables considered:

| Variable | Mean | Std.Dev. | Minimum | Maximum |
|---|---|---|---|---|
| flow | 655.4 | 2397.19 | 1 | 94917 |
| popi | 729757 | 1049576.41 | 8240 | 6390240 |
| malei | 0.49 | 0.009 | 0.46 | 0.56 |
| medagei | 35.08 | 6.63 | 15 | 48 |
| activei | 0.4 | 0.06 | 0.25 | 0.96 |
| comp_primi | 0.56 | 0.11 | 0.35 | 0.99 |
| comp_secondi | 0.29 | 0.12 | 0.0007 | 0.78 |
| areai | 0.86 | 0.64 | 0.001 | 5.67 |
| popj | 733221 | 1049330.79 | 12410 | 6390240 |
| malej | 0.5 | 0.009 | 0.46 | 0.55 |
| medagej | 35.03 | 6.62 | 12 | 48 |
| activej | 0.4 | 0.06 | 0.25 | 0.96 |
| comp_primj | 0.56 | 0.11 | 0.35 | 0.99 |
| comp_secondj | 0.29 | 0.12 | 0 | 0.78 |
| areaj | 0.76 | 0.64 | 0.001 | 5.67 |
| dist | 4.33 | 4.29 | 0.08 | 23.69 |
| cont | - | - | 0 | 1 |

# Countries and input data (III)

- This map shows the migration flows from one particular origin region (coloured in blue) of Slovenia. The warmer the region color (red, yellow, white), the more migrants go from that region to the destination region.
- This is the movement data for two years living in that blue region in the 2002 census data.

# Statistical methodology (I)

- Statistical tool: Gravity-type spatial interaction models including random effects.
- The gravity model is summarized as $flows_{ij} = (pop_i^{\alpha} pop_j^{\beta})/dist_{ij}^{\gamma}$
- From this equation, we have the standard gravity model, which expresses migration flows as a linear predictor of populations and distances.
- We expand the basic model with the variables mentioned previously.

# Statistical methodology (II)

- Counting the number of migrants → discrete counts over an unbounded positive range → Poisson distribution.
- Because individuals are nested/grouped within countries, there exits a hierarchical (also called multilevel) structure in our data.
- We are also interested in understanding the impact that grouping (countries) can have on the number of migrants individuals, i.e., how much the variation in the amount of flows can be attributed to the country of residence.
- In statistical terms, this means to use random effects. We will also consider as random effects the origin and destination regions and the census year.

# Statistical methodology (III)

- Random effects are represented by the matrix $U \sim N(0, \sigma^2)$.
- Common variables are represented by the matrix $X$ and the associated vector of coefficients $\beta$.

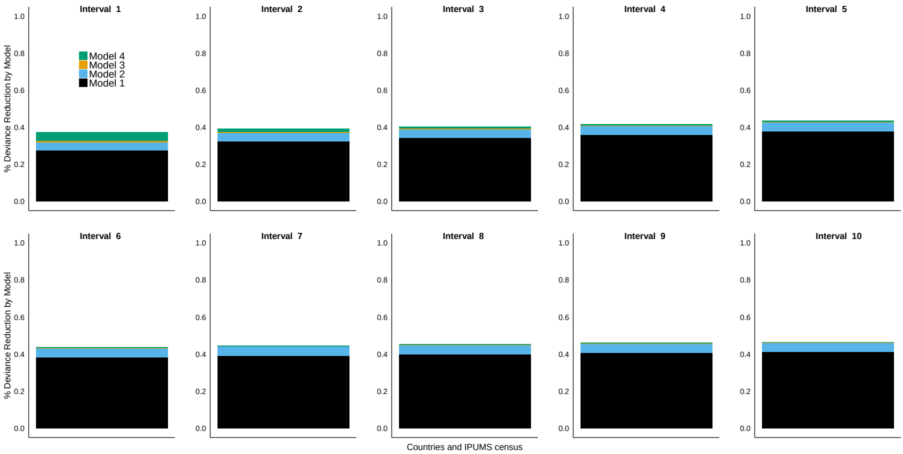
$$log(\widehat{flows_{ijk}}) = \beta_0 + \beta_1 log(pop_{ik}) + \beta_2 log(pop_{jk}) + \beta_3 log(dist_{ijk}) + \beta X + U$$

where $i$ refers to origin, $j$ to destination and $k$ to the interval $(k = 1, \ldots, 10)$.

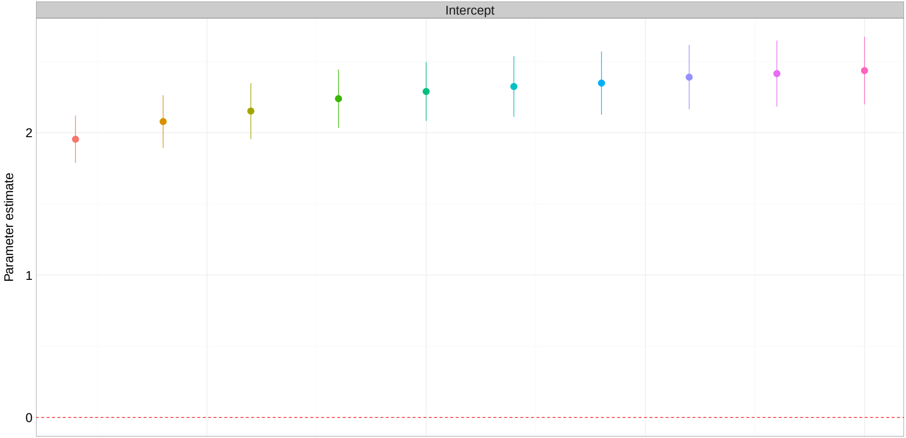
- Procedure: For each interval $k$, we run five models:

| Model | Model description |
|---|---|
| 1 (Null) | Only random effects |
| 2 | Model 1 + $\log10(pop_{ik})$ + $\log10(pop_{jk})$ + $\log10(dist_{ijk})$ |
| 3 | Model 2 + $cont$ + $area_{ik}$ + $area_{jk}$ |
| 4 | Model 3 + $male_{ik}$ + $male_{jk}$ + $medage_{ik}$ + $medage_{jk}$ |
| 5 | Model 4 + $active_{ik}$ + $active_{jk}$ + $comp\_prim_{ik}$ + $comp\_prim_{jk}$ + $comp\_second_{ik}$ + $comp\_second_{jk}$ |

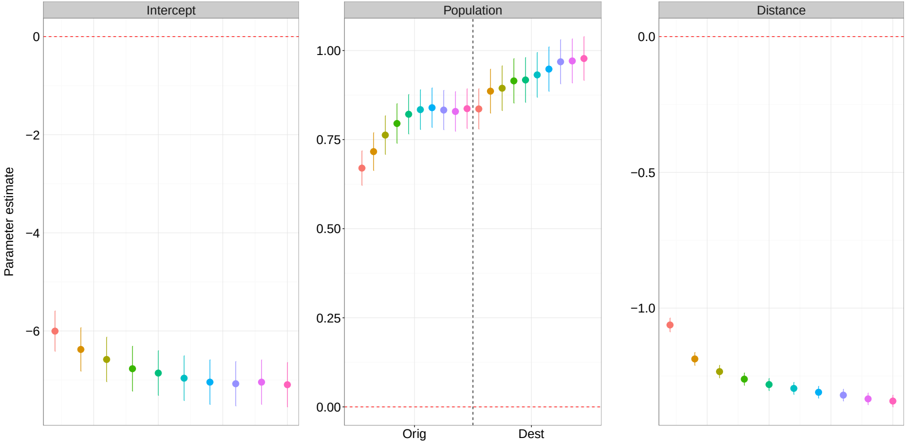
# Results (I): Explanatory ability of the models



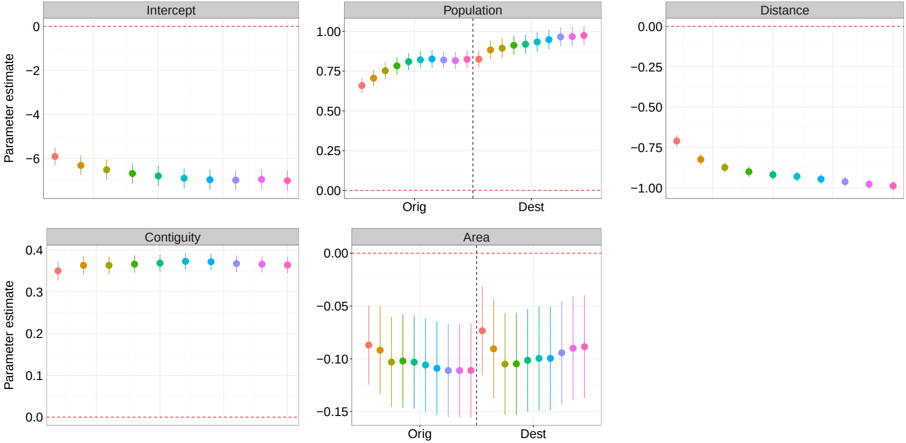
# Results (II): Fixed effects (Model 1 (Null model))



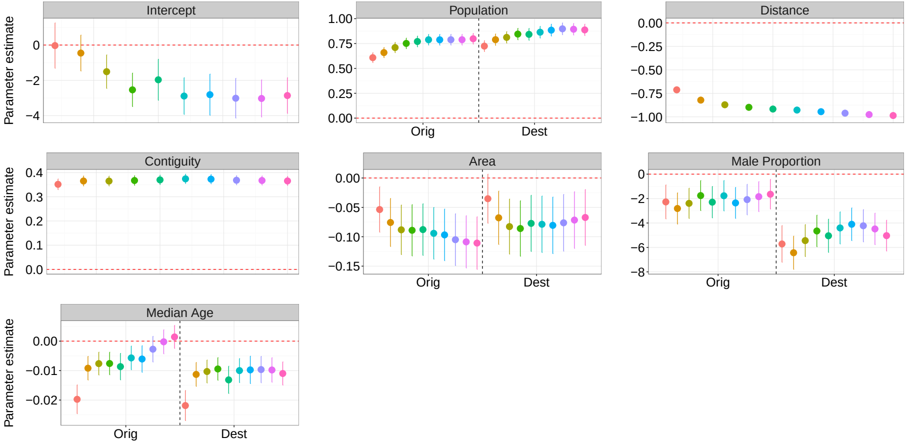
# Results (II): Fixed effects (Model 2)



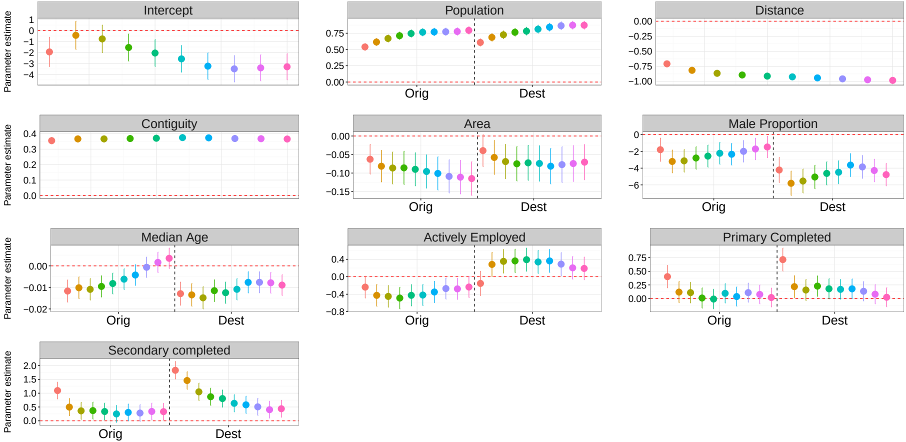
# Results (II): Fixed effects (Model 3)



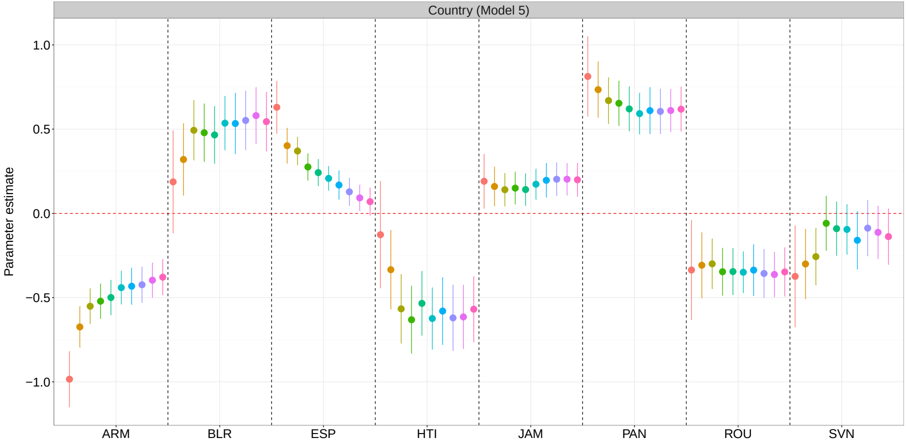
# Results (II): Fixed effects (Model 4)



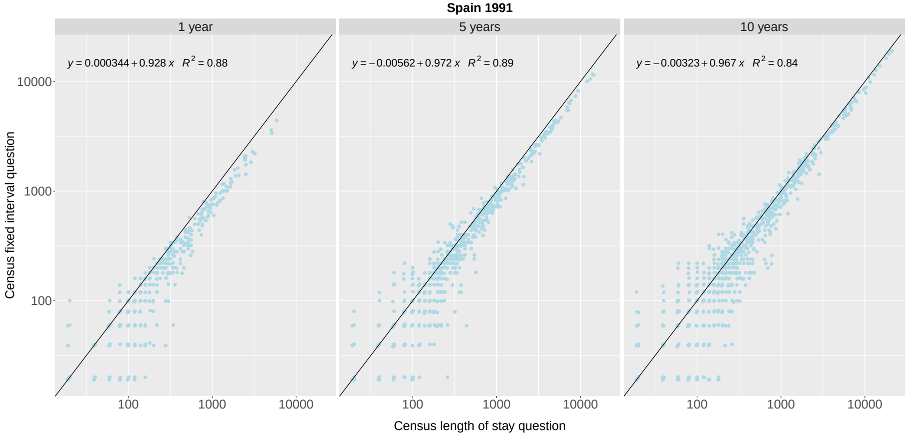
# Results (II): Fixed effects (Model 5)



# Results (III): Random effects



# Results (IV): Estimated flows vs actual flows



# Software tools

- To visualize results about migrations, we have developed this app http://bayes2.ucd.ie:3838/gvinue/AppCountsMigr2/
- It allows us to display several plots, such as a heatmap and a chord diagram, to describe migration flows in an interactive way.
- This app has been created with **shiny**, the web application framework for **R**.

# Conclusions and future work

- Conclusions:
  - The resulting models allow us to compare synthetic bilateral migrant flow counts over a range of timing criteria.
  - Web app to visualize relations between the changes in residences of the migrants.
  - R code for the app available on request.
- Future work:
  - Adding as many countries as possible.
  - Imputing missing predictors variables.

# References


- Courgeau, D. *Migrants and migrations*. Population (selected papers n.3, translated), pages 1-35, 1973.
- Garcia, A., Pindolia, D., Lopiano, K., and Tatem, A. *Modeling internal migration flows in Sub-Saharan africa using census microdata*. Migration Studies, pages 1-22, 2014.
- Nowok, B. and Willekens, F. *A probabilistic framework for harmonisation of migration statistics*. Population, Space and Place, pages 1-13, 2010.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. https://www.R-project.org/.

THANK YOU FOR THE ATTENTION

Guillermo Vinué
www.uv.es/vivigui