Revisiting Rogers and Castros multi-exponential models for migration estimation

UoROS2020

J. Sebastian Ruiz-Santacruz, ADRI - Shanghai Univesity and Externado de Colombia University
Mabelin Villareal-Fuentes, International Labour Organization

04/12/20

Migration age schedule: Why is important?

- It describes the migration age pattern associated to selectivity as an interaction of push and pull factors in origin-destination areas.
- It allows visual comparisons about the intensity and the migrant popuation general structure.

Rogers and Castro model 1981

- It estimates a migration schedule based on exponential curves that belong to a demographic behaviour.
- Give us some demo-economic parameters.
- Other alternatives are kernel smothing (in case of under-5 child over estimation suspicion) and non parametrical parameters based on the curves.
- Ohter definitions: post-retirement (Wilson, 2020)
- Other options: non-parametric measures based on the form (Bernard, 2013).

Uses

- Some countries still need methods to smooth migration and grasp the mean behaviour of the migration calendar.
- There are opportunities to use it in estimation children under 5 years old. P. e. When empirical data shows it historically. (!!)
- Population projections.

Migration parameters

Multi-exponential model:

$$M(x) = a_1 e^{-\alpha 1 x} + a_2 e^{-\alpha 2(x-\mu 2) - e^{-\lambda 2(x-\mu 2)}} + a_3 e^{-\alpha 3(x-\mu 3) - e^{-\lambda 3(x-\mu 3)}} + a_4 e^{\lambda 4 x} + c$$

- 7,9,11 y 13
- M(x): Standardized age rate x.
- μ_2, μ_3: Location parameters.
- a1, a2, a3, a4: Levels and equation coeficients.
- α1, λ2, α2, λ3, α3, λ4). Demo-economic descriptions.

- **Ratios:**
  - **a1/a2: Dominant labour curve**
  - **a2/a1: Infant dependency indicator**
  - **λ2/α2: Labour curve asimetry**

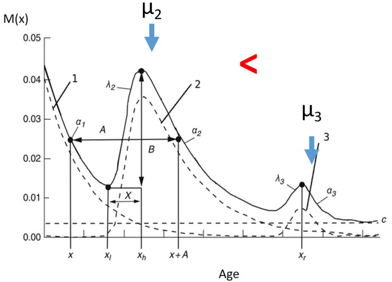
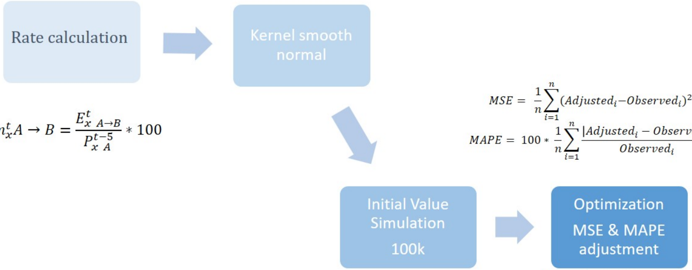


Figure 1: Based in Rogers and Castro, 1981

Procedure



$$m_x^t A \to B = \frac{E_{x\ A \to B}^t}{P_{x\ A}^{t-5}} * 100$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Adjusted_i - Observed_i)^2$$

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n}\frac{|Adjusted_i - Observed_i|}{Observed_i}$$

Mean Absolute Percentual Error

migraR package

elflacosebas Update README.md — 93a2e38 on 3 Sep — 133 commits

.Rproj.user/E68A6BE5	new commit for update and check spelling	3 months ago
R	new commit for update and check spelling	3 months ago
data	16dic update roxigenized	11 months ago
man	new commit for update and check spelling	3 months ago
.Rbuildignore	added a few roxygen pars, lots more to do. attach error not yet handl...	11 months ago
.gitignore	attempt to define well the functions	3 years ago
DESCRIPTION	new commit for update and check spelling	3 months ago
NAMESPACE	delee plotrc	10 months ago
README.md	Update README.md	3 months ago
build.R	new commit for update and check spelling	3 months ago
migraR.Rproj	attempt to define well the functions	3 years ago

README.md

migraR

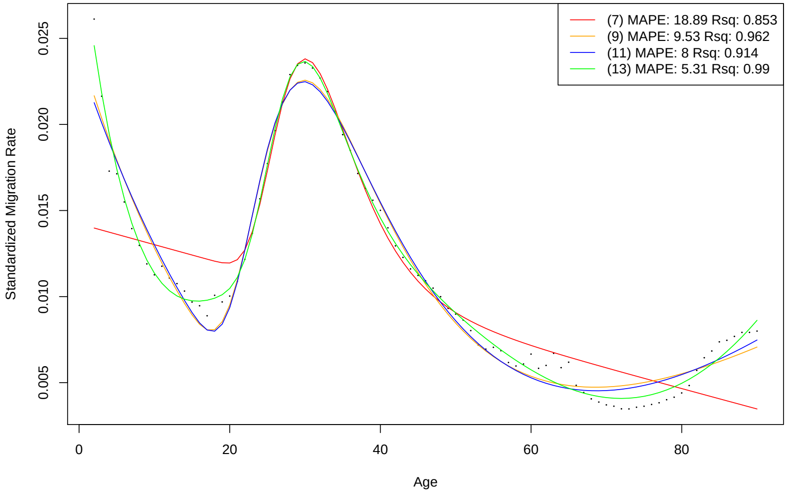
R package for migration analysis focused now on the Rogers and Castro multi exponential model and the estimation of the parameters using a simulation based on uniform a priori distributions for each parameter between 0 and 1, and 0 to 100 in the case of the location parameters corresponding to ages in the migration pattern.

More information: https://github.com/elflacosebas/migraR

Example

```
# Calling packages and dataset.

 library(migraR)
 library(dplyr)
 library(tidyverse)

 data("es_asmr")
 data1 <- es_asmr[-c(1,2),c(1,6)]
 colnames(data1) <- c("x","y")

 # Fitting and Plotting data
 fitted.val.7  <- best_migramod(dataIn = data1, maxite = 200, profile = "seven")
 fitted.val.9  <- best_migramod(dataIn = data1, maxite = 200, profile = "nine")
 fitted.val.11 <- best_migramod(dataIn = data1, maxite = 200, profile = "eleven")
 fitted.val.13 <- best_migramod(dataIn = data1, maxite = 200, profile = "thirteen")

 x11()
 plot(data1, cex=0.1, xlab = 'Age',
     ylab = 'Standardized Migration Rate')
 lines(data1[,1], fitted.val.7$modelClass$value(fitted.val.7$bestParam,data1), col="blue")
 lines(data1[,1], fitted.val.9$modelClass$value(fitted.val.9$bestParam,data1), col="orange")
 lines(data1[,1], fitted.val.11$modelClass$value(fitted.val.11$bestParam,data1), col="blue", lty=3)
 lines(data1[,1], fitted.val.13$modelClass$value(fitted.val.13$bestParam,data1), col="green")
 legend('topright',
        legend = c(paste("(7)", "MAPE:", round(as.numeric(fitted.val.7$bestMAPE),2),
                         "R²:", round(as.numeric(fitted.val.7$bestRcuad),3)),
                   paste("(9)", "MAPE:", round(as.numeric(fitted.val.9$bestMAPE),2),
                         "R²:", round(as.numeric(fitted.val.9$bestRcuad),3)),
                   paste("(11)", "MAPE:", round(as.numeric(fitted.val.11$bestMAPE),2),
                        "R²:", round(as.numeric(fitted.val.11$bestRcuad),3)),
                   paste("(13)", "MAPE:", round(as.numeric(fitted.val.13$bestMAPE),2),
                         "R²:", round(as.numeric(fitted.val.13$bestRcuad),3))),
                  col = c("red",'orange',"blue","darkgreen"), lty = c(2,6,3,5))
```

Example



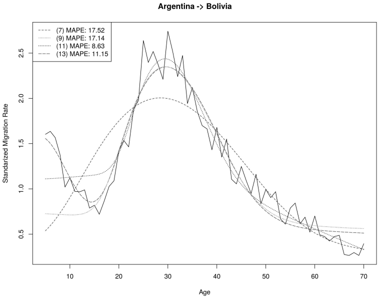
(7) MAPE: 18.89 Rsq: 0.853
(9) MAPE: 9.53 Rsq: 0.962
(11) MAPE: 8 Rsq: 0.914
(13) MAPE: 5.31 Rsq: 0.99
Standardized Migration Rate
0.005
0.010
0.015
0.020
0.025
0
20
40
60
80
Age

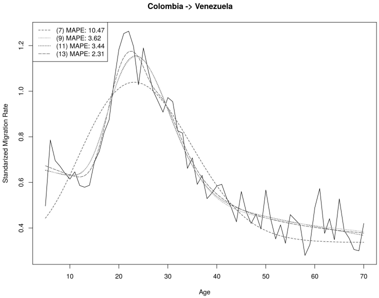
Application Results

- Data on Latin American migration: REDATAM project https://bit.ly/3qf454D
- Most of the models adjusted ended in 11 or 13 parameters (115 out of 139 pairs of origin-destination).
- Many curves do not fit with the pattern because there is a curve between 5 and 15 aprox that we have called: Child migration delay.
- Candidate to be a mechanism of migration (internal or international).
- Worst estimation of the post-retirement migration peak.

Results: Two examples

Model 1: Colombia - Venezuela (Hombres)
Model 2: Colombia - Venezuela (Mujeres)
Model 3: Argentina - Bolivia (Hombres)
Model 4: Argentina - Bolivia (Mujeres)

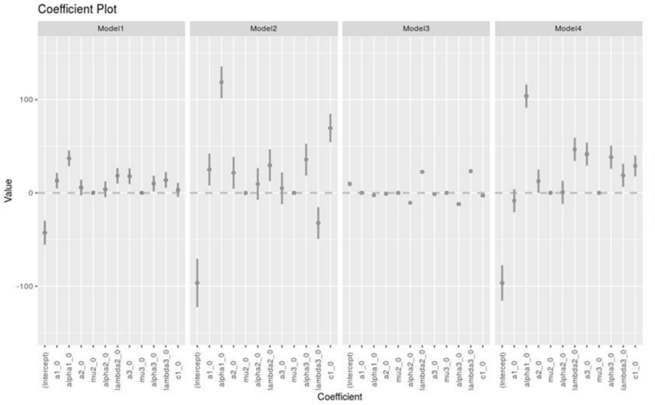




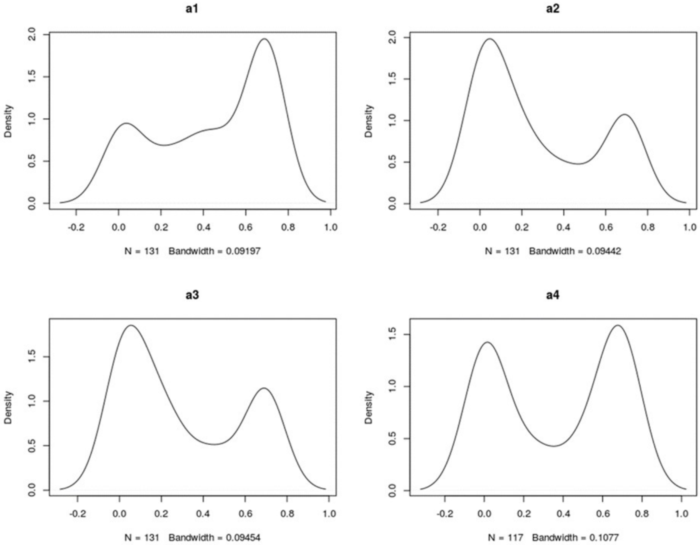
Results Final parameters

	Mujeres: Colombia – Venezuela 2011			Mujeres: Argentina – Bolivia 2001		
Parametros	Valor	5%	95%	Valor	5%	95%
a1_0	0.512	0.225	0.877	0.863	0.224	0.877
α1_0	0.654	0.225	0.878	0.394	0.224	0.878
a2_0	0.058	0.224	0.878	0.398	0.226	0.877
μ2_0	**38.000**	**23.000**	**87.000**	**23.000**	**23.000**	**87.000**
α2_0	0.826	0.227	0.878	0.876	0.224	0.878
λ2_0	0.253	0.226	0.877	0.031	0.226	0.878
a3_0	0.837	0.226	0.878	0.686	0.223	0.877
μ3_0	**27.000**	**23.000**	**87.000**	**21.000**	**23.000**	**87.000**
α3_0	0.806	0.227	0.877	0.751	0.224	0.877
λ3_0	0.534	0.224	0.878	0.008	0.225	0.877
a4_0	0.531	0.224	0.877	0.624	0.227	0.877
λ4_0	0.782	0.225	0.878	0.562	0.225	0.878
c1_0	0.085	0.248	0.975	0.573	0.251	0.976
a1_hat	0.690	0.700	0.700	0.700	0.700	0.700
α1_hat	0.028	0.013	0.127	0.017	0.001	0.179
a2_hat	0.700	0.000	0.700	0.700	0.000	0.700
μ2_hat	**32.481**	**19.879**	**88.000**	**17.854**	**23.229**	**88.000**
α2_hat	0.423	0.058	0.700	0.166	0.034	0.700
λ2_hat	0.124	0.122	0.700	0.000	0.075	0.700
a3_hat	0.700	0.000	0.700	0.700	0.000	0.700
μ3_hat	**28.529**	**19.656**	**88.000**	**46.646**	**22.911**	**88.000**
α3_hat	0.150	0.064	0.700	0.360	0.041	0.700
λ3_hat	0.527	0.136	0.700	0.083	0.081	0.700
a4_hat	0.670	0.619	0.700	0.685	0.700	0.700
λ4_hat	0.000	0.000	0.516	0.000	0.000	0.167
c1_hat	0.690	0.645	0.700	0.685	0.700	0.700
ECM	0.023	0.062	4.618	0.297	3.365	34.645
MAPE	**2.316**	**3.616**	**24.904**	**6.420**	**12.817**	**54.731**

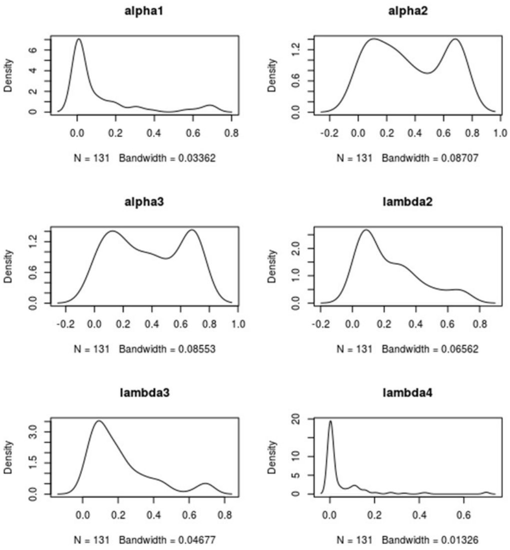
Results: Influence of initial values on the MSE



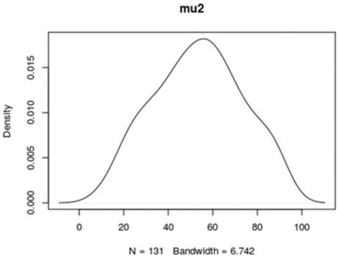
Results: Final parameters

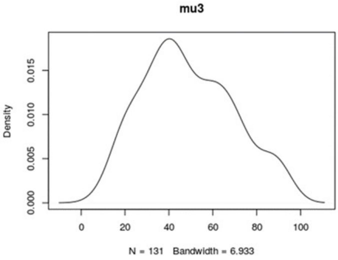


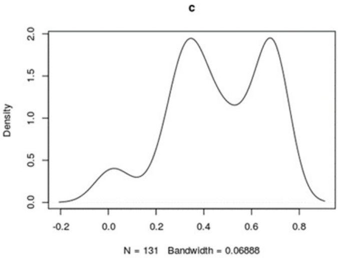
Results: Final parameters



Results: Final parameters

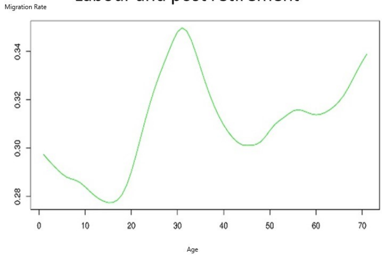




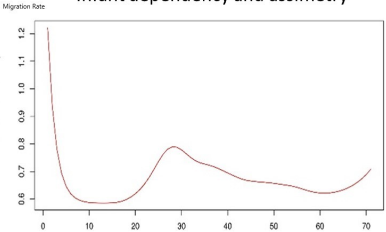


Results: Typology

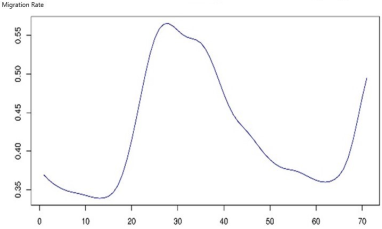
Labour and post retirement



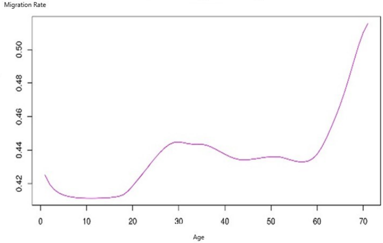
Infant dependency and assimetry



Labour curve more extended among ages

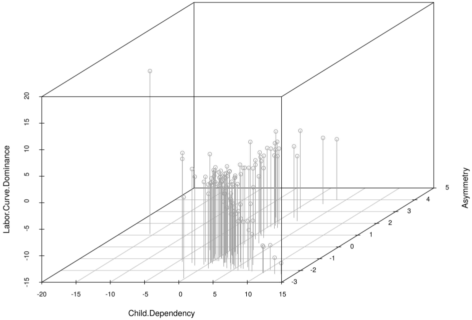


Assimetry and aged migration



Results

- Using the logarithm of the estimated ratios we can say that the latinamerican migration system has a medium-high child dependency, a predominance of labour force curve and a very asymetric labour force curve, are congruent with the reality of Latin American migrations.



- The more child dependency the less is the symmetry.

Why is important for the ILO?

- It is necessary to establish robust historical measures that allow their projection and allow to carry out public policy scenarios.
- Could be part of a country classification according to migration characteristics that lead to better standardization.
- Knowing various estimation methodologies provides feedback for the collection and harmonization data methods.
- Having other aproaches to labour force estimation due to migration with other sources rather than LFS.

Ongoing work

- Detection and analyisis of international emigration using administrative register of migration Colombia (Statistics Colombia).
- Using Spanish administrative data to see a better child curve and defining what we could say about. 15 parameter model (A. Mendoza, Treasury - Colombia and J. Recano CED - Barcelona).
- Bayesian hierarchical model: Incorporate beta priors and some imputs as GDP and others.
- Combining this results with the ones obtained with the ILO bulk information (Advice M. Villareal, ILO)