Comparison of Bayesian and non-Bayesian approaches to correct for measurement error in migration flow data for South America

Andrea Aparicio-Castro, Arkadiusz Wiśniowski, Mark Brown, Francisco Rowe

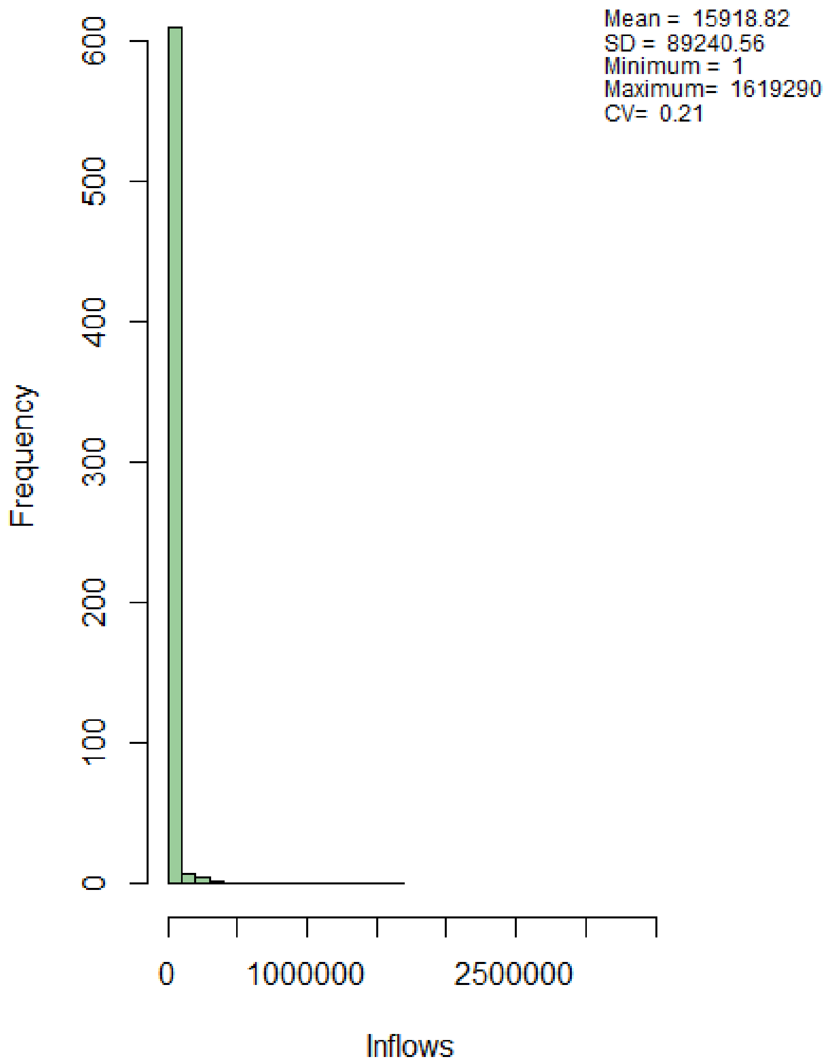
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
North West Social Science Doctoral Training Partnership
MANCHESTER 1824 The University of Manchester

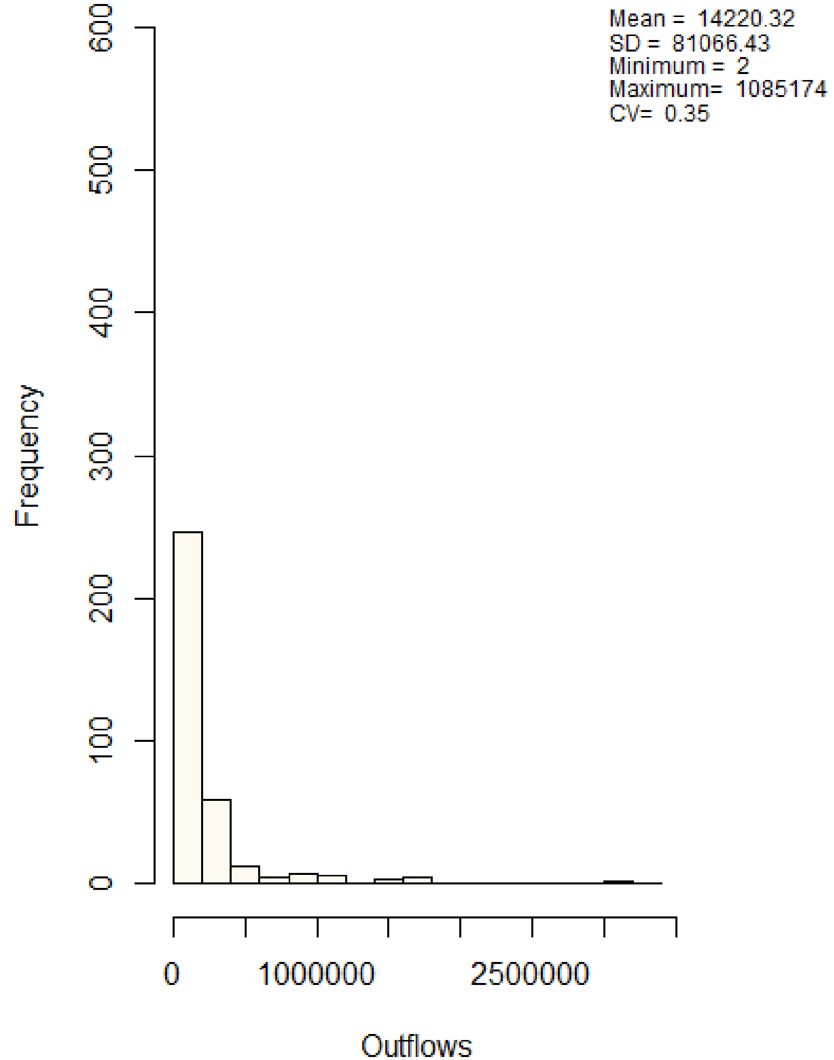
Observed South American countries (in blue)

Data

$$z_{ijt}^{k} \begin{cases} Z_t^S \Rightarrow Z_t^S = \begin{pmatrix} 0 & z_{12t}^S & z_{13t}^S & \cdots & z_{1nt}^S \\ z_{21t}^S & 0 & z_{23t}^S & \cdots & z_{2nt}^S \\ z_{31t}^S & z_{32t}^S & 0 & \cdots & z_{3nt}^S \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1t}^S & z_{n2t}^S & z_{n3t}^S & \cdots & 0 \end{pmatrix} \\ \\ Z_t^R \Rightarrow Z_t^R = \begin{pmatrix} 0 & z_{12t}^R & z_{13t}^R & \cdots & z_{1nt}^R \\ z_{21t}^R & 0 & z_{23t}^R & \cdots & z_{2nt}^R \\ z_{31t}^R & z_{32t}^R & 0 & \cdots & z_{3nt}^R \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1t}^R & z_{n2t}^R & z_{n3t}^R & \cdots & 0 \end{pmatrix} \end{cases}$$

Data: comparison between inflows and outflows in South America





Data

$$z_{ijt}^{k} \begin{cases} Z_t^S \Rightarrow Z_t^S = \begin{pmatrix} 0 & z_{12t}^S & z_{13t}^S & \cdots & z_{1nt}^S \\ z_{21t}^S & 0 & z_{23t}^S & \cdots & z_{2nt}^S \\ z_{31t}^S & z_{32t}^S & 0 & \cdots & z_{3nt}^S \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1t}^S & z_{n2t}^S & z_{n3t}^S & \cdots & 0 \end{pmatrix} \\ Z_t^R \Rightarrow Z_t^R = \begin{pmatrix} 0 & z_{12t}^R & z_{13t}^R & \cdots & z_{1nt}^R \\ z_{21t}^R & 0 & z_{23t}^R & \cdots & z_{2nt}^R \\ z_{31t}^R & z_{32t}^R & 0 & \cdots & z_{3nt}^R \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1t}^R & z_{n2t}^R & z_{n3t}^R & \cdots & 0 \end{pmatrix} \end{cases}$$

$$Y_t = \begin{pmatrix} 0 & y_{12t} & y_{13t} & \cdots & y_{1nt} \\ y_{21t} & 0 & y_{23t} & \cdots & y_{2nt} \\ y_{31t} & y_{32t} & 0 & \cdots & y_{3nt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n1t} & y_{n2t} & y_{n3t} & \cdots & 0 \end{pmatrix}$$

$$\Downarrow$$

$$y_{ijt}$$

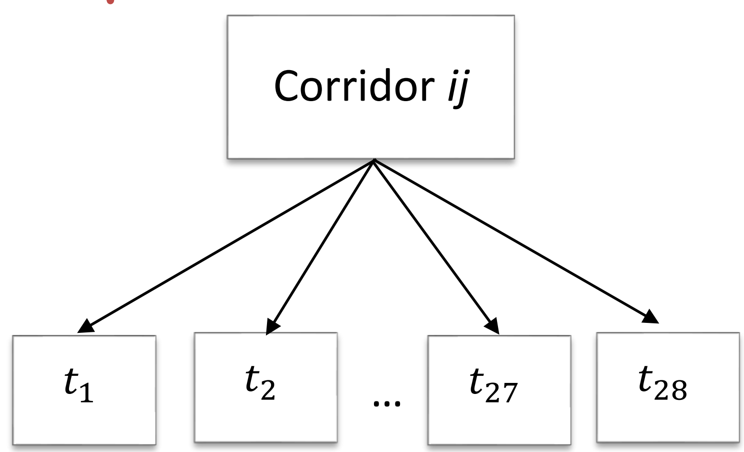
“A person who moves to a country [*j*] other than that of his or her usual residence [*i*] for a period of at least a year (12 months)” (UN, 1998:18)

Data can be seen as …

Origin *i*	Destination *j*	Corridor *ij*	Year *t*	Counts z_{ijt}^{k}
ARG	BOL	ARG -> BOL	1991	…
ARG	BRA	ARG -> BRA	1991	…
ARG	CHL	ARG -> CHL	1991	…
⋮	⋮	⋮	⋮	⋮

Level 2 (90 corridors)

Level 1 (2520 observations)



Adapting data to the UN definition

$$y_{ijt} \quad \Rightarrow \quad \begin{array}{l} z_{ijt}^{S} \sim Poisson(\mu_{ijt}^{S}) \\ z_{ijt}^{R} \sim Poisson(\mu_{ijt}^{R}) \end{array}$$

$$\log(\mu_{ijt}^{k}) = \log(y_{ijt}) + q_{ijt}^{k}$$

$$\log(\mu_{ijt}^{k}) = \log(y_{ijt}) + \boldsymbol{\delta}_{m(k)} + \boldsymbol{\lambda}_{n(k)} + \boldsymbol{\beta}_{q(i)} + \boldsymbol{\beta}_{q(j)} + q_{ijt}^{k}$$

$$\log(\mu_{ijt}^{S}) = \log(y_{ijt}) + \delta_{m(i)} + \lambda_{n(k)} + \beta_{q(i)} + q_{ijt}^{S}$$

$$\log(\mu_{ijt}^{R}) = \log(y_{ijt}) + \delta_{m(j)} + \lambda_{n(k)} + \beta_{q(j)} + q_{ijt}^{R}$$

Adapting data to the UN definition

$$\log(\mu_{ijt}^{S}) = \log(y_{ijt}) + \delta_{m(i)} + \lambda_{n(k)} + \beta_{q(i)} + q_{ijt}^{S}\ ; \qquad q_{ijt}^{S} \sim N\ (0, \sigma_{q(i)}^{2})$$

$$\log(\mu_{ijt}^{R}) = \log(y_{ijt}) + \delta_{m(j)} + \lambda_{n(k)} + \beta_{q(j)} + q_{ijt}^{R}\ ; \qquad q_{ijt}^{R} \sim N\ (0, \sigma_{q(j)}^{2})$$

$$\delta_{m(k)} = \begin{cases} \delta_1 & \text{if criterion is "0 months"} \\ \delta_2 & \text{if criterion is "3 months"} \\ \delta_3 & \text{if criterion is "6 months"} \\ 0 & \text{if criterion is "12 months"} \end{cases}$$

$$\lambda_{n(k)} = \begin{cases} \lambda_{\mathbf{1}} & \text{if k source undercounts immigrants} \\ \lambda_{\mathbf{2}} & \text{if k source undercounts emigrants} \end{cases}$$

$$\begin{aligned} \delta_1 &= \delta_1 + \delta_2 + \delta_3 \\ \delta_2 &= \delta_2 + \delta_3 \\ \delta_3 &= \delta_3 \\ \delta_4 &= -\delta_4 \end{aligned}$$

$$\beta_{q(k)} = \begin{cases} \beta_1 & \text{Good registers} \\ \beta_2 & \text{Less reliable registers} \\ 0 & \text{Excellent register} \end{cases}$$

Non-Bayesian and Bayesian approaches

Non-Bayesian approach => Maximum Likelihood

vs.

Bayesian approach => MCMC

Prior of $\log(y_{ijt})$

$$\log(y_{ijt}) = \alpha_0 + \alpha_1 t + \alpha_2 \log(\mathrm{P_{it}}) + \alpha_3 \log(\mathrm{P_{jt}}) + \alpha_4 \mathrm{B_{ijt}} + u_{0ij} + u_{1ij} t + e_{ijt}$$

$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ 0 & \sigma_{u1}^2 \end{pmatrix}\right]$$

$$e_{ijt} \sim N(0, \sigma_e^2)$$

Variances of non-adapted and adapted estimates

Parameters		Maximum likelihood		MCMC	
		Estimate	95% CI	Mean	95% HDI region
Non-adapted data	σ_{qS}^2	0.665	(0.161, 0.846)	0.423	(0.002, 0.664)
	σ_{qR}^2	49.741	(2.419, 98.098)	50.109	(2.797, 97.073)
Adapted data	σ_{qS}^2	0.602	(0.046, 0.836)	0.4770	(0.074, 0.676)
	σ_{qR}^2	50.099	(2.717, 97.494)	49.837	(3.335, 97.693)

Measurement model parameters of adapted data

Parameters	Maximum likelihood			MCMC		
	Estimate	SE	95% CI	Mean	SD	95% HDI region
δ_1	0.196	0.086	(0.056, 0.395)	0.143	0.067	(0.035, 0.291)
δ_2	0.130	0.076	(0.019, 0.302)	0.090	0.053	(0.0133, 0.215)
δ_3	0.065	0.058	(0.002, 0.221)	0.044	0.039	(0.001, 0.148)
λ_1	4.925	1.180	(2.944, 6.990)	11.112	2.178	(7.271, 13.354)
λ_2	-0.535	30.931	(-62.519, 59.292)	0.263	32.588	(-63.626, 63.548)
β_1	-0.098	0.077	(-0.244, 0.052)	-0.100	0.057	(-0.211, 0.013)
β_2	-0.053	0.046	(-0.142, 0.037)	-0.042	0.039	(-0.118, 0.032)

$\delta_3 = 0.044$ => $\exp(-\delta_3) = 0.9569$

For countries with a minimum duration of stay of 6 months, the true flows represent 95.69% of the observed/reported flows.

Sensitivity of variances to the partial removal of the data (MCMC estimates)

Parameters		Complete data		Partial removal of the data	
		Mean	95% HDI region	Mean	95% HDI region
Non-adapted data	σ_{qS}^2	0.423	(0.002, 0.664)	0.535	(0.115, 0.687)
	σ_{qR}^2	50.109	(2.797, 97.073)	50.115	(2.895, 97.339)
Adapted data	σ_{qS}^2	0.477	(0.074, 0.676)	0.524	(0.081, 0.709)
	σ_{qR}^2	49.837	(3.335, 97.693)	50.325	(2.799, 97.606)

Conclusions

- Despite the differences between MLE and MCMC estimates, the HDIs contain the values obtained in the foremost.

Further work

- Improving the criteria to consider the accuracy of the data sources.
- Including additional measurement error variables to capture other differences (e.g. one-year and five-year definitions).
- Testing non-linear treatments of time and dependency between intercepts and slopes should be tested.

References (part 1)

Alimi, O., Mare, D., & Poot, J. (2015). Does distance still matter for internal migration and, if so, how? Evidence from 1986 to 2006. *Labour, Employment and Work in New Zealand*. Available at https://ojs.victoria.ac.nz/LEW/article/view/3897/3487

Bertoli, S., & Moraga, J. F. H. (2017). Gravity models in the migration and development nexus. *Revue d'economie du developpement*, *25*(3), 69-91. Available at https://www.cairn-int.info/load_pdf.php?ID_ARTICLE=E_EDD_313_0069

Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view* (Vol. 24). Springer Science & Business Media.

Brierley, M., Forster, J., McDonald, J., & Smith, P. (2008). Bayesian estimation of migration flows. In Raymer, J. & Willekens, F. (eds.). *International migration in Europe: Data, models and estimates*, pp. 149-174. England: John Wiley & Sons, Ltd. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470985557.ch7/pdf

Bryant, J., & Graham, P. (2013). Bayesian demographic accounts: Subnational population estimation using multiple data sources. *Bayesian Analysis*, *8*(3), pp. 591-622. Available at: https://projecteuclid.org/download/pdfview_1/euclid.ba/1378729921

References (part 2)

De Beer, J., Raymer, J., Van der Erf, R., & Van Wissen, L. (2010). Overcoming the problems of inconsistent international migration data: A new method applied to flows in Europe. *European Journal of Population/Revue européenne de Démographie*, *26*(4), 459-481. Available at https://link.springer.com/article/10.1007/s10680-010-9220-z

DeWaard, J., Kim, K., & Raymer, J. (2012). Migration systems in Europe: Evidence from harmonized flow data. *Demography*, *49*(4), 1307-1333. Available at https://link.springer.com/content/pdf/10.1007%2Fs13524-012-0117-9.pdf

Disney, G., Wiśniowski, A., Forster, J., Smith, P., & Bijak, J. (2015). Evaluation of existing migration forecasting methods and models. *Report for the Migration Advisory Committee. Southampton: University of Southampton.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/467405/Migration_Forecasting_report.pdf

Newell, C. (1988). *Methods and models in demography*. Guilford Press.

Nowok, B. (2008). Evolution of international migration statistics in selected Central European countries. In Raymer, J. & Willekens, F. (eds.). *International migration in Europe: Data, models and estimates*, pp. 73-87. England: John Wiley & Sons, Ltd.

References (part 3)

Nowok, B., & Willekens, F. (2011). A probabilistic framework for harmonisation of migration statistics. *Population, Space and Place*, *17*(5), 521-533. Available at https://onlinelibrary.wiley.com/doi/epdf/10.1002/psp.624

Raymer, J., Wiśniowski, A., Forster, J., Smith, P., & Bijak, J. (2013). Integrated modelling of European migration. *Journal of the American Statistical Association*, *108*(503),801-819. Available at https://www.tandfonline.com/doi/pdf/10.1080/01621459.2013.789435?needAccess=true

United Nations (UN) (1998). Recommendations on Statistics of International Migration, Revision 1. *Statistical* Papers, Series M, *58*(1), Department of Economic and Social Affairs, Statistics Division. Available at: https://unstats.un.org/unsd/publication/seriesm/seriesm_58rev1e.pdf

Willekens, F. (2008). Models of migration: observations and judgements. In Raymer, J. & Willekens, F. (eds.). *International migration in Europe: Data, models and estimates*, pp. 117-148. England: John Wiley & Sons, Ltd.

Wiśniowski, A. (2017). Combining Labour Force Survey data to estimate migration flows the case of migration from Poland to the UK. *Journal of the Royal Statistical Society: Series A*, *180*(1), 185-202. Available at https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssa.12189

Thanks!

Andrea Aparicio-Castro
andrea.apariciocastro@manchester.ac.uk

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
North West Social Science Doctoral Training Partnership
MANCHESTER 1824 The University of Manchester