Following a Trail of Breadcrumbs: Using Digital Traces to Improve Migration Estimates

Francesco Rampazzo[*]
Jakub Bijak[†]
Agnese Vitali[‡]
Ingmar Weber[§]
Emilio Zagheni[¶]

Introduction

Measuring international migration is challenging. The lack of timely and comprehensive data about migrants, and varying measures and definitions used by countries are a barrier to reviewing international migration (Bijak 2010; Kupiszewska and Nowok 2008; Willekens 1994). In recent years, Bayesian methods have started to be used to combine different sources of migration data in order to provide a better estimate of the number of migrants (Azose and Raftery 2019; Bijak 2010). In this paper, we aim to complement traditional data sources with social media data. We are proposing to use a Bayesian data assessment model to combine the data from the Labour Force Survey (LFS) and Facebook Advertising Platform to study the number of European migrants in the UK. We are aiming to produce an estimate of European migrants closer to their *true stock* number.

Background

The Integrated Model of European Migration (IMEM) is the Bayesian model that we are aiming to use. This framework has been created by Raymer and colleagues (2013) for combining the flows reported by the sending countries with the flows reported by the receiving countries to estimate a number closer to the true value of the flows. This model has been applied by Disney (2015) to combine multiple migration survey datasets in the United Kingdom (UK), and by Wiśniowski (2017) to combine the LFS data for the case of Polish migration to the UK. The main feature of this model, IMEM, is that it provides a framework which assesses the limitations of the datasets in terms of the definition of migrants used; the bias and the accuracy are also considered to create an appropriate prior distribution which could adjust these data issues.

In this study, we limit our attention to migrants from European countries, since in the British context they are the hardest groups to estimate. Since, at least until 2019, there is no requirement for registration for EU migrants in the UK, survey data are used to estimate the stock of migrants from the EU. The Office for National Statistics (ONS) has been in a process since late 2016 of incorporating more data sources coming from multiple governmental institutions; this process is still ongoing. In addition, the Office for National Statistics has reclassified their estimates as *"experimental statistics"*, as a result, the ONS have been clear to emphasise that the estimates may be inaccurate (ONS 2019a). The aim of the paper is thus to complement existing, but incomplete, official estimates of migrant's stocks' through digital traces. The paper aims to answer the question what does Facebook Advertising data add to ONS data sources?

---

[*]University of Southampton, Max Planck Institute for Demographic Research
[†]University of Southampton
[‡]Università di Trento
[§]Qatar Computing Research Institute
[¶]Max Planck Institute for Demographic Research

Data

Survey-based Migration Data

British migration data is fragmentary: different data sources measure different migrant populations or migration events. In the absence of register data, the UK largely relies on a survey-based system to collect information on its population. The two main sources to estimate international migration to the UK are the International Passenger Survey (IPS), and, secondly, the Labour Force Survey (LFS). As mentioned in the introduction, the ONS is working in complementing multiple data sources since the survey-based system is not enough.

In this piece of research, we are using the second main source of data, which is the LFS, a Europe-wide quarterly household survey, which aims to estimate labour market conditions, including employment and unemployment. Through a boost of this survey, the Annual Population Survey (APS), the ONS collects data on the stocks of foreign born and foreign citizen in the UK at local area level.

New Migration Data

Exploring the digital traces that we leave as queries on search engines and as posts on social media is a new trend in the social sciences (Cesare et al. 2018). In this piece of research, we are using Facebook Advertising data obtained from the Facebook Marketing Application Programming Interface (API) throgh PySocialWatcher (Araujo et al. 2017). We obtain data structure by age and sex for most of the European nationalities (22 countries) living in the UK.

Methodology

The model is constructed as follows. We observe the number of European migrants (stocks) Z^k from country i to the UK and by sex j, from Facebook, F, and from the LFS, L, where k in L, F. The datasets we are using can be described in the form of matrices Z^F for Facebook, and Z^L for the LFS. The value Y_{ij} is the random variable describing the number of the true stock we are aiming to estimate. It is a matrix with dimension $I \times J$. The model is divided into three parts: the Measurement Error Model, the Migration Theory Model, and the Age Dissagregation Model.

Measurement Error Model

The general equation of the measurement error model is:

$$\log \mu_{ij}^k = \log(y_{ij}) + \delta^k + \beta^k + \chi_{ij}^k + \xi_{ij}^k + \epsilon_{ij}^k \tag{1}$$

The equation is composed by four terms, δ^k, β^k, χ_{ij}^k, and ξ_{ij}^k, which are used to convert the data from Facebook and the LFS to comply with the UN definition of international migrant and reduce the underestimation linked to the bias or coverage of the data. In the paper, we will address all the limitations of the different data sources using the preexisting literature.

Theory Based Model

In this part of the model, we are introducing covariates that might help in explaining the true stock of European migrants in the UK.

$$\log y_{ij} = \alpha_0 + \alpha_1 P_{ij} + \alpha_2 I_{ij} + \alpha_3 O_{ij} + \alpha_4 \log G_i + \alpha_5 \log U_{ij} + \epsilon_{ij} \tag{2}$$

The population size of the country of origin P is introduced as a gravity measure, which means the larger is the population the more people can be migrant to the UK. Inflows and outflows, I and O, from and to the European countries are considered as recent trends of immigration and emigration to the UK. The GDP

growth rate, *G*, has the value to give a picture of the wealth of the economy in the home country, as well as the unemployment rate, *U*. Since much of the immigration to the UK from European countries is for working reasons (ONS 2019b), these are the variables to consider in the Theory based model.

Age Disaggregation

The current work is focusing on applying a multinomial-Dirichlet-Dirichlet to the true stock data divided by sex in order to obtain the dissagregation of the estimates by age. In the multinomial-Dirichlet-Dirichlet, we are using alpha values estimated from a Rogers-Castro model. At the moment, we can show some preliminary results on the dissagregation using alpha estimated from a Rogers-Castro on the LFS and Facebook data separately. We are working in combining the two migration schedules obtained from the two data sources.

Results and Next Steps

In Figure 1, we present the results from a first model (figure 1a), in which we estimate the total number of European migrants by countries, while in a second model (figure 1b), we dissaggregate the estimates by sex. The model estimates suggest that there are more European migrants than measured by the official estimates in 2018: our model suggests a 22% undercount, although with considerable uncertainty.
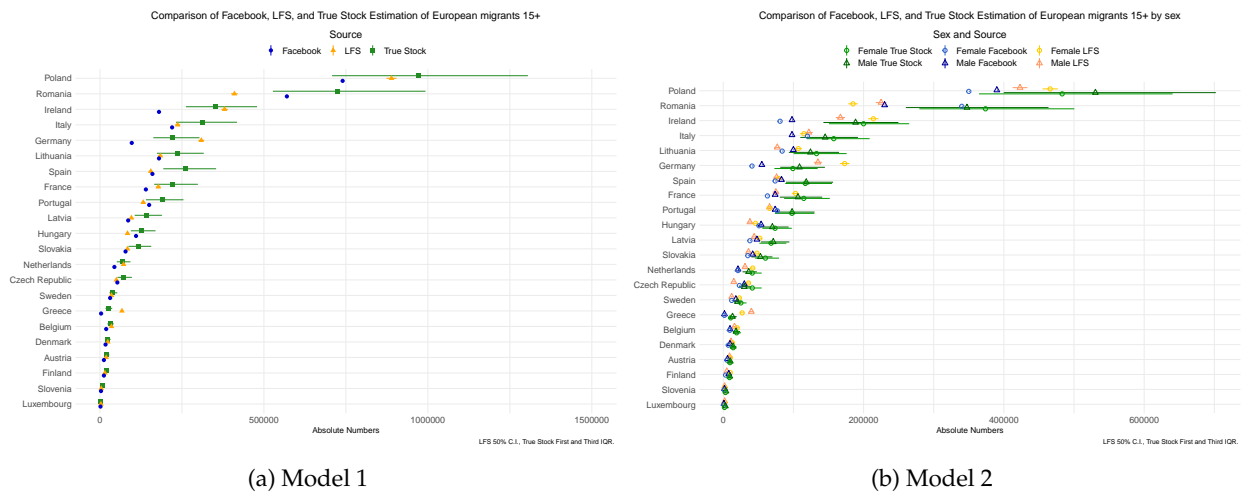

(a) Model 1
(b) Model 2

Figure 1: Comparison of the LFS, Facebook, and True Stock estimates of the total numbers of European migrants disaggregated by sex in the UK in 2018.

In Figure 2, we show the results from the Age Dissagregation model with the level of uncertainty of the estimates. The two migration schedules obtain by the two data sources are different, and highlight a younger population in the Facebook than in the LFS. The final version of the paper will include as well the last 2019 estimates from the Office for National Statistics. Several sensitivity analysis techniques are used to evaluate the quality of the model. In the full paper, we will discuss the advantages and limitations of this approach.
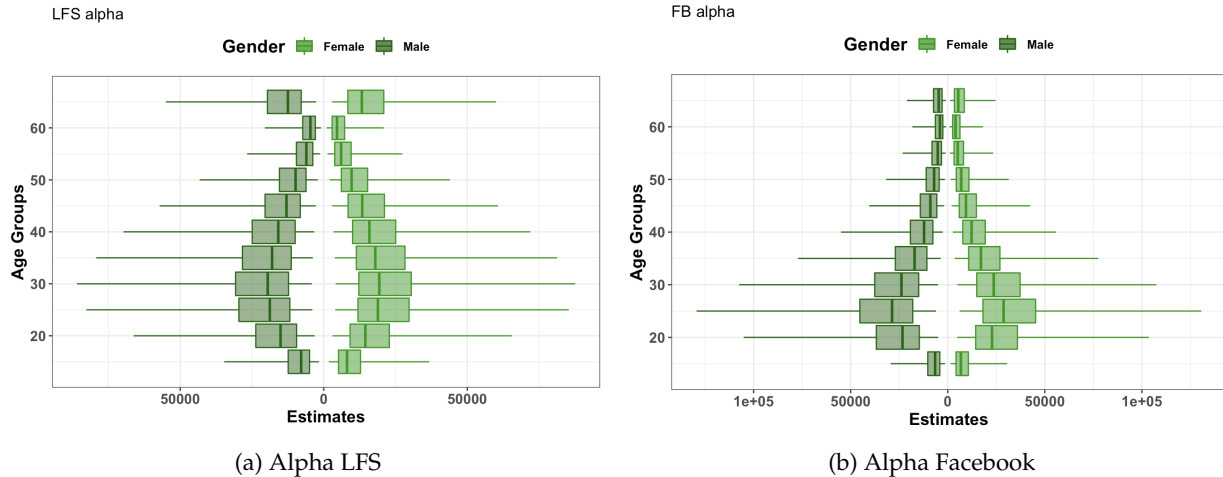
| (a) Alpha LFS | (b) Alpha Facebook |

Figure 2: Population Pyramids of the Italian True Stock estimates of the total numbers of European migrants disaggregated by age and sex in the UK in 2018.

References

Araujo, M., Mejova, Y., Weber, I., and Benevenuto, F. (2017). Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. *Proceedings of the 2017 ACM on Web Science Conference*:253–257. doi:10.1145/3091478.3091513.

Azose, J.J. and Raftery, A.E. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences* 116(1):116–122. doi:10.1073/pnas.1722334116.

Bijak, J. (2010). *Forecasting International Migration in Europe: A Bayesian View*. Springer Science & Business Media.

Cesare, N., Lee, H., McCormick, T., Spiro, E., and Zagheni, E. (2018). Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography* 55(5):1979–1999. doi:10.1007/s13524-018-0715-2.

Disney, G. (2015). *Model-based estimates of UK immigration*. [Phd]. University of Southampton.

Kupiszewska, D. and Nowok, B. (2008). Comparability of statistics on international migration flows in the European Union. In: *International Migration in Europe: Data, Models and Estimates*. John Wiley & Sons: 41–71.

ONS (2019a). Statement from the ONS on the reclassification of international migration statistics - Office for National Statistics..

ONS (2019b). Understanding different migration data sources: August progress report - Office for National Statistics..

Raymer, J., Wiśniowski, A., Forster, J.J., Smith, P.W.F., and Bijak, J. (2013). Integrated Modeling of European Migration. *Journal of the American Statistical Association* 108(503):801–819. doi:10.1080/01621459.2013.789435.

Willekens, F. (1994). Monitoring international migration flows in Europe. *European Journal of Population / Revue européenne de Démographie* 10(1):1–42. doi:10.1007/BF01268210.

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 (Statistics in Society)* 180(1):185–202. doi:10.1111/rssa.12189.