December 2020

Predicting international migratory movements using Google searches

Marcus H. Böhme, German Federal Ministry of Finance

André Gröger, Universitat Autònoma de Barcelona (UAB), Barcelona Graduate School of Economics (BGSE), and Markets, Organizations and Votes in Economics (MOVE)

Tobias Heidland (né Stöhr),, Universidad de Kiel, Institut für Weltwirtschaft (IfW) e IZA Institute of Labor Economics

Internet searches in a specific country for keywords semantically related to migration and the name of the destination country can be used to measure bilateral migration flows. The analysis of search volumes from the Google search engine across a hundred countries of origin correlates positively with the flows of people who have moved, over a 12-year-period, towards 35 destination countries. The results show that this approach outperforms traditional migration flow models substantially in terms of predictive power. It may serve to monitor migratory movements in near real time and thus enable policymakers and practitioners to manage migration more effectively.
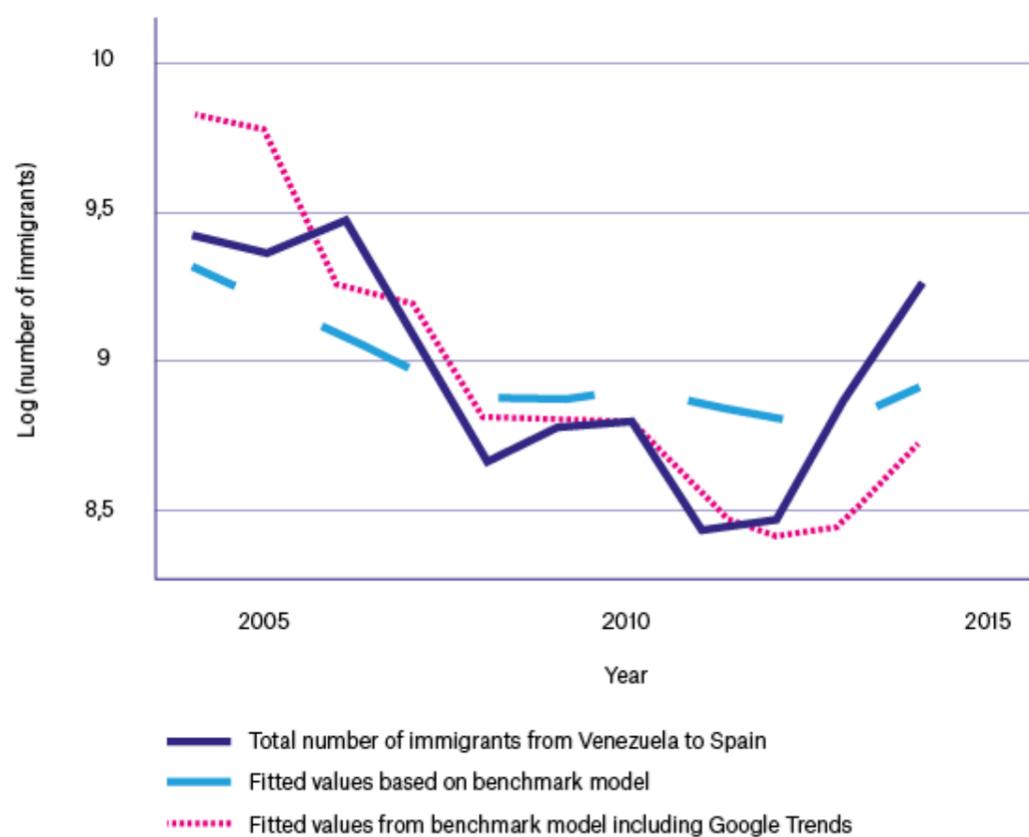
Key points

1 Those aspiring to migrate typically acquire information about migration opportunities online, in their countries of origin, prior to departure. These data can be used to make short-term predictions about numbers of migrants, which may be very useful in situations such as humanitarian crises.

2 Google search data, which represent searches by over 1 billion people worldwide, are free of charge for small queries. Using the data to study international migration can be helpful, especially in developing countries because, to date, information on migration and people's migration intention is scarce or exclusively available to paying users.

3 This approach tends to work better when focusing on countries where internet is more widely used and the Google search engine is more widespread, and where more people speak the languages for which search data are studied.

Google searches can help anticipate future migratory flows

Evolution of immigration from Venezuela to Spain and predictions based on a classical migration flow model as well as an augmented one including the Google Trends Index

Total number of immigrants from Venezuela to Spain
Fitted values based on benchmark model
Fitted values from benchmark model including Google Trends

Source: Compiled by the author based on data from Google Trends, World Bank World Development Indicators, and OECD International Migration Database.

Social Observatory of "la Caixa"

The solid line in the graph represents real arrivals of migrants to Spain originating from Venezuela between 2004 and 2015, according to the International Migration Database compiled by the Organisation for Economic Cooperation and Development (OECD).

The dashed line corresponds to the predictions based on the benchmark model including GDP and population dynamics at both origin and destination as explanatory variables. The dotted line is the result of the previous estimation, augmented with the data from Google searches in Venezuela. The graph shows a sharp increase in immigration since 2012, presumably due to the socio-political and economic crisis in Venezuela. The keywords searched for on Google help predicting this upward tendency more accurately.

Words that people search for online can indicate migrants' intentions

The website, Google Trends, enables us to see how the interest in a keyword evolves over time. The authors extracted this time series for 67 keywords semantically related with migration in three languages, i.e. Spanish, English and French. Internet users who are considering leaving their own country will often search for keywords such as the name of the desired country of destination or combine it with terms such as *immigration*, *passport*, or *salary*. Such internet searches are thus driven by migration intentions and typically lead to increased migration flows.

Graph 1.   **The election of Trump triggered Google searches in Mexico about migrating to the United States**

Evolution of search intensity for the term USA in Mexico between 2016 and 2017.
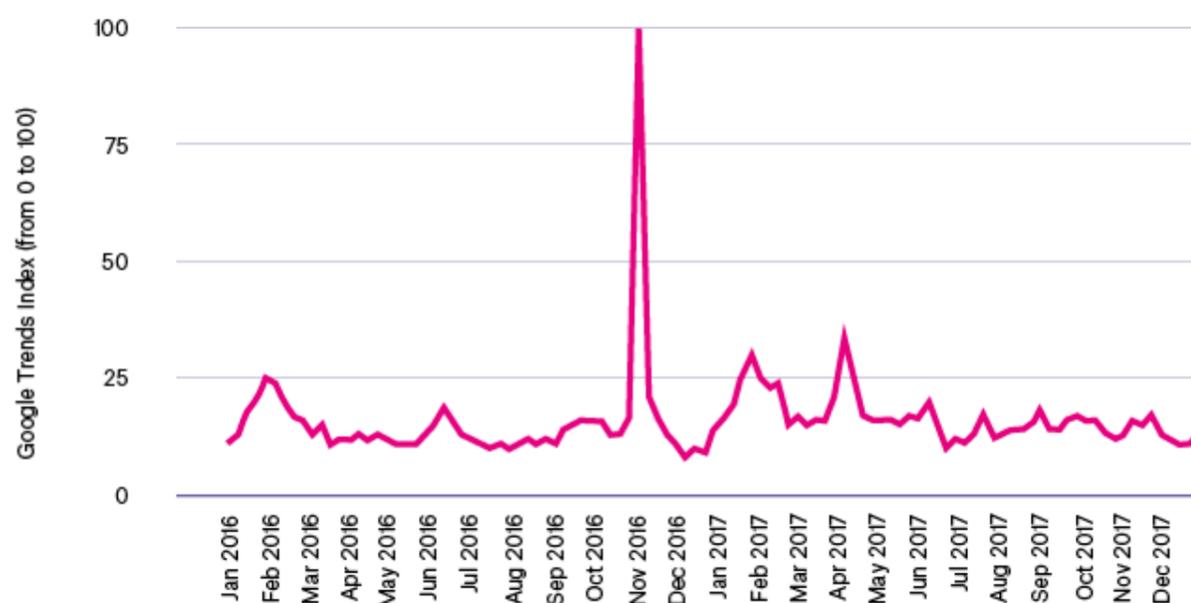
Source: Compiled by the author based on data from Google Trends.

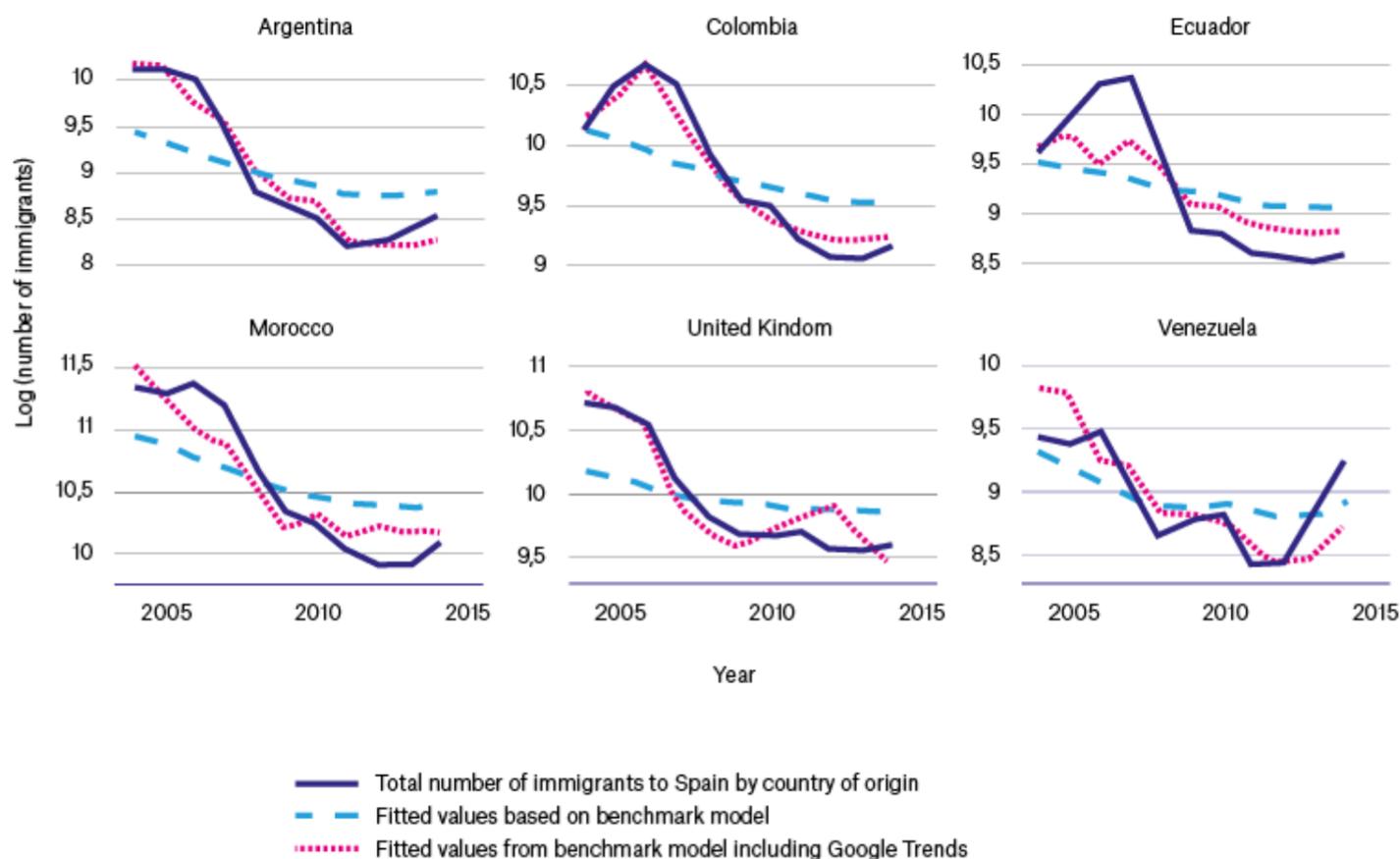Social Observatory of "la Caixa"

In November 2016, Google Trends recorded a sudden peak in search intensity for the keyword "USA". This increase in interest coincided with the date of election of the current president of the United States, Donald Trump. It can be assumed that this partly reflected a certain concern among Mexicans about immigration opportunities in anticipation of Trump's clampdown on immigration. In the face of the fear that entering the United States would become more complicated in the following months, Trump's election victory may have led many Mexican citizens to seek information on migration to the US. Official data from 2017 from the OECD show that there was also a peak in immigration from Mexico to the USA during that year.

1. The portrait of immigration to Spain

The main origins of migratory movements to Spain between 2004 and 2015 among countries that officially or administratively use one of the three languages studied – Spanish, English or French – were Argentina, Colombia, Ecuador, Morocco, the United Kingdom and Venezuela.

Graph 2.   **Google searches help predict migration flows more accurately**

Evolution of immigration from top origin countries to Spain and predictions based on a classical migration flow model as well as an augmented one including the Google Trends Index.

Total number of immigrants to Spain by country of origin
— — Fitted values based on benchmark model
.... Fitted values from benchmark model including Google Trends

Source: Compiled by the author based on data from Google Trends, World Bank World Development Indicators, and OECD International Migration Database.

Social Observatory of "la Caixa"

Predictions based on the model including internet searches follow the actual migration flows quite closely. In all cases, we can observe how the dotted line coincides with marked decreases in immigration to Spain. The key driver behind the decrease in immigration was the economic crisis, which began in 2008 and lasted for several years. The predictions including internet searches also reflect well the slight increase in arrivals towards the end of the study period which corresponds to the rebound of the Spanish economy.

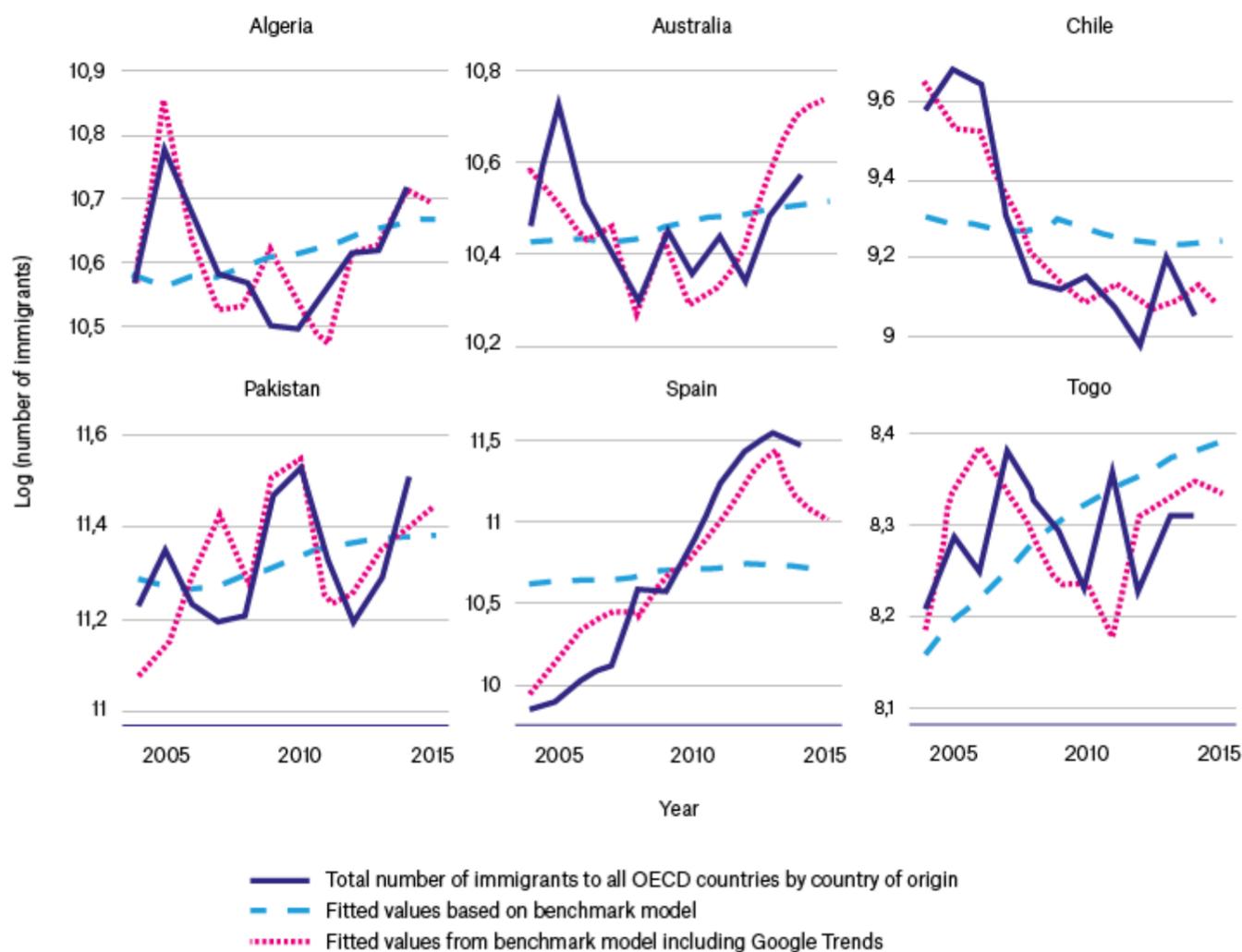# 2. A tool for researchers and practitioners

Migration has become one of the most important and most disputed issues in developed and developing countries alike. Traditional prediction models tend to be static and often perform poorly when it comes to detecting turning points in migration flows. Adding an index that reflects Google searches improves the accuracy of these methods.

Prediction based on keywords uses a free source of data that can help to facilitate migration research, especially in developing countries. Furthermore, the approach holds the promise of allowing short-term predictions. Information on migration flows tends to arrive with considerable delays that can be several years long. This new approach is especially promising with regard to

delivering better results when studying countries or regions that have a relatively high internet penetration.

Graph 3. Towards a tool for global migration flow prediction

Evolution of immigration from selected origin countries to all OECD countries and predictions based on a classical migration flow model and an augmented model including the Google Trends Index.

Total number of immigrants to all OECD countries by country of origin
Fitted values based on benchmark model
Fitted values from benchmark model including Google Trends

Source: Compiled by the author based on data from Google Trends, World Bank World Development Indicators, and OECD International Migration Database.

Social Observatory of "la Caixa"

This graph shows total immigration flows into OECD countries from Algeria, Australia, Chile, Pakistan, Spain and Togo, as well as their predictions based on the benchmark and augmented model. The graph shows that the Google Trends approach (dotted line) not only works for predictions from one country to another, but also from one country to many. The solid line corresponds to official data from the OECD.

3. Towards more finely tuned predictions. Conclusions

Georeferenced internet search data offer new and practically infinite opportunities to measure and predict human behaviour. Innovative applications beyond the context of migration are feasible, provided that demand for specific online information is indicative of subsequent human behaviour. To help identify turning points in migration, predictions of migration patterns can be further fine-tuned to specific corridors or regions where sudden changes are to be expected.

The approach based on Google Trends could be used to predict migration flows close to real time, ahead of official data release. In this way, it can be useful also for giving fast political responses to emergency situations.

4. Reference

BÖHME, M. H., A. GRÖGER and T. STÖHR (2020): "Searching for a better life: Predicting international migration with online search keywords", *Journal of Development Economics,* 142, 102347, https://doi.org/10.1016/j.jdeveco.2019.04.002.