Revisiting forced migration: A machine learning perspective☆

Maja Micevska

Vrije Universiteit Brussel (VUB), Brussels, Belgium

A R T I C L E   I N F O

A B S T R A C T

This study is motivated by an interest in understanding which factors play an important role in forced migration abroad. I use machine learning methods that allow for more effective ways to estimate complex relationships, particularly with highly nonlinear data. There are three key findings of this study. First, I document and test for the importance of the changing nature of conflict in Africa. My estimates show that riots are the most important type of conflict for explaining asylum applications, more important than battles or violence against civilians. Second, this paper is the first macrolevel study to investigate the role of the internet penetration rate in explaining forced migration flows and the first to show the importance of this factor in explaining asylum applications. Third, I find that country fixed effects are of primary importance in explaining forced migration flows. This finding suggests that a considerable variation in forced migration flows still remains to be explained.

1. Introduction

Forced migration is one of the most important issues facing the world today. Hardly anyone doubts that refugees face a difficult decision when contemplating to leave their homes, belongings, and family members in favor of an uncertain future somewhere else. Still, increasing number of people make this decision. According to the UN Refugee Agency (UNHCR), the number of forcibly displaced people (refugees, asylum seekers, and internally displaced people) reached record 70.8 million at end-2018 (UNHCR, 2019).

There are many aspects of forced migration, the humanitarian and political considerations being probably the most important ones. Politicians, researchers, aid workers, and the media often discuss the issue from different perspectives. Images of African refugees crammed into small, unsafe boats crossing over the Mediterranean have become familiar in Europe and around the world. Many observers have been outraged by the stories of sinking boats and drowning refugees, and the issue of how to best help migrants in peril has become hotly debated. At the same time, we have witnessed a split in opinion on the nature of the refugee crisis, i.e., whether it is mainly about forced displacement or economic migration. On one hand, proponents of the forced displacement view concur with the opinion of Grandi (UNHCR, 2019) when explaining the global trends in refugee flows: "What we are seeing in these figures is further confirmation of a longer-term rising trend in the number of people needing safety from war, conflict and persecution." On the other hand, proponents of the economic migration view side with the opinion of populist politicians that majority of the refugees and asylum seekers are actually economic migrants in disguise.

This study is motivated by an interest in understanding which factors play an important role in forced migration abroad. There are a number of studies that explore this issue and in the following section I present a review of the literature. However, previous studies use

traditional estimation methods which are not well-suited when there are nonlinear relationships and interactions in the data. In addition, most studies use rather crude measures of conflict-related violence and ignore the complexity of conflict that could be captured with more nuanced indicators.

The study is different from previous studies in that it applies machine learning techniques for estimation. Compared to traditional estimation methods, machine learning methods allow for more effective ways to estimate complex relationships, particularly with highly nonlinear data (Varian, 2014; Mullainathan and Spiess, 2017). When analyzing forced migration, it is important to consider potential nonlinearities and interaction effects among variables. Let us first consider economic reasons for migration. Poverty is no doubt an important reason for migration but the effect is nonlinear, i.e., poverty may strengthen the desire to migrate but at the same time reduce the ability to do so. Thus, the relationship between economic reasons and migration may change after crossing some specific level of income or other measure of economic reasons.[1] If a measure for economic reasons is modeled in a linear fashion, as most previous studies on forced migration do, then it should not be surprising that the findings of empirical studies have been contradictory.

Now let us turn to conflict in society as a reason for migration. There is no doubt that a high level of conflict-related threat in the country of origin lies at the root of refugee flights. Indeed, most previous studies agree that conflict is the primary determinant of forced migration (Schmeidl, 1997; Davenport et al., 2003; Hatton, 2009). It is less obvious, however, which particular forces matter most since operational measures of conflict in previous research suffer from various weaknesses. Conflict is often captured by a single indicator and possible interaction effects between different types of conflicts are largely ignored. To address this shortcoming, I take advantage of recent developments in rich datasets on conflicts and their characteristics to differentiate among different sources of threat in society.

I use data on forced migration from 45 African countries for the years 1997–2017. I account for traditional and more recent determinants of migration, and additionally implement country and year fixed effects to account for unobserved factors. There are three key findings of this study. First, I document and test for the importance of the changing nature of conflict in Africa, an issue that has been largely ignored not only in the empirical macroeconomic literature but also in policy debates. My estimates show that riots are the most important type of conflict for explaining asylum applications, more important than battles or violence against civilians. Second, this paper is the first macrolevel study to investigate the role of the internet penetration rate in explaining forced migration flows and the first to show the importance of this factor in explaining asylum applications. Third, I find that country fixed effects are of primary importance in explaining forced migration flows. This finding holds after controlling for variables that have been used as determinants of forced migration in previous studies. Apparently, a considerable variation in forced migration flows still remains to be explained.

I proceed as follows. Section 2 provides a survey of empirical research on forced migration. Section 3 presents the data and data sources used in this study. In Section 4, I provide a broad overview of trends in forced migration and conflict in Africa. Section 5 presents and discusses the estimation method and the empirical findings. The final section summarizes the findings, highlights policy implications, and identifies fruitful avenues for future research.

2. Literature review

There are a number of journal articles, working papers, and edited volumes that explore the subject of forced migration. Studies that conduct quantitative analysis to study the subject differ in many regards, including operational measures of main variables, controls included in the analysis, and results obtained.

Let us first consider the dependent variable in this study. UNHCR differentiates among three types of forcibly displaced people: refugees, asylum seekers, and internally displaced people, i.e., people displaced within the same country. Studies focusing on forcibly displaced people who left their home country employ different operational measures of the dependent variable. Many studies (e.g., Schmeidl, 1997; Davenport et al., 2003; Melander and Öberg, 2007) use (changes in) refugee stocks as a dependent variable. Other studies (e.g., Neumayer, 2005; Hatton 2009, 2016) focus mainly on asylum seekers. There are also studies (e.g., Dreher et al., 2019) that consider both refugees and asylum seekers. I also include data on both refugees and asylum seekers in the analysis since these two types of forced migrants might be influenced by different considerations when deciding to leave their country of origin.

Regardless of the measurement of the dependent variable employed, it is important to recognize the difficulties in measuring forced migration flows. First of all, it is often difficult to make a clear distinction between those in need of protection and those migrating for economic reasons, especially when they come from countries undergoing rapid change, crisis or war (Castles et al., 2003). This aggravates the process of assessing validity of asylum applications, which is inevitably reflected in the migration statistics. Further challenges include distinguishing between migration stocks and migration flows, accounting for the different types of forced migration, disentangling asylum seekers from aggregate migration statistics, and updating the statistics to reflect changes in migrants' legal status (Koser, 2010). There are additional challenges with statistics on migration flows explained later in this paper, in the section *Data*.

Many migrants have multiple motivations for leaving their home country. Recognizing the multi-dimensionality of individuals' decisions to migrate, I incorporate variables to account for conflict, economic conditions, political and institutional factors, and other covariates that have been found to be associated with international migration.

---

[1] The nonlinear relationship between income and migration is discussed in more detail in the next section.

2.1. Conflict

Previous research overwhelmingly shows that indicators of conflict are more significant than indicators of economic development as explanatory factors of forced migration flows (e.g., Schmeidl, 1997; Castles et al., 2003; Davenport et al., 2003; Moore and Shellman, 2004; Hatton, 2009). Almost all empirical macrolevel studies on international migration include a measure of conflict, so the importance of this variable is well established by now. An issue that has received less attention is that not all conflicts are alike. Sirkeci and Cohen (2016) define conflict as a continuum of positions ranging from a position where no conflict exists, to one where conflicts lead to violent clashes, wars and life-threatening risks. Major conflicts at the right end of the distribution, such as the recent conflicts experienced by Iraqis and Syrians, are surely more likely to lead to mass outflows of people. Nevertheless, less violent or non-violent conflicts can also create an environment of human insecurity and encourage out-migration. Many people decide to flee not necessarily because of violent conflicts in their country but because of government repercussions after having voiced their opposition or because they are identified with groups that have been targets of human rights abuse.

Even though conflict can come in many different forms, most previous studies ignore the large variations across conflicts. Conflict in society is usually modeled as civil war fatalities or a dummy variable capturing a civil war. There are a few notable exceptions. Davenport et al. (2003) use three distinct indicators of conflict-related violence: state violence, dissident violence, and state-dissident violence. In this study, all the measures of state violence have statistically significant and substantively important effects on migration. Melander and Öberg (2007) account for intensity and geographical scope of armed conflicts and, contrary to expectations, find that migration is more a product of the scope of violence rather than its intensity. Bohra-Mishra and Massey (2011) include a quadratic term for the violence indicator and find that violence has a nonlinear effect on migration, such that rising violence induces out-migration only above a certain threshold. Despite the strengths of these studies, their operational measures of conflict in society are less than optimal. When we consider the concept of conflict, it is clear that not only violence is important but also other factors as well, e.g., the variety of types of conflicts or tensions created by non-violent protests. Taking advantage of recent developments in data collection, I seek to both refine as well as expand the measurement of conflict in society.

Accounting for the large variations across conflicts is particularly important for studies that deal with contemporary Africa. As shown later in this paper, the nature of conflict in Africa has changed dramatically. Civil wars and battles are not the major type of conflict anymore, as it was the case during the 20th century. For more than a decade now, many people from across Africa's urban populations have been taking to the streets to protest and demand change. Branch and Mampilly (2015) document ninety protests in forty African countries during the 2005–14 period. Except for the protests that occurred in four North African countries affected by the Arab Spring in 2011–12, protests and riots in the rest of Africa have unfolded with little international attention by both academic research and policy discussions. Yet, as protests and riots have been sweeping the continent, the need has grown for better understanding of the new types of African rebellion. Protests in Africa do not conform to the Western views on this type of conflict. Protesters have often faced intense state violence, which in turn has led to violent behavior by the protesters, turning non-violent rebellion into violent rioting.[2] This means that the line between violent and non-violent conflict will have to be reconsidered in the case of African uprisings (Branch and Mampilly, 2015).

Forced migration usually occurs when conflict, in combination with other factors, leads people to flee (Castles et al., 2003). Numerous studies have tried to address this issue by testing for interaction effects between a measure of conflict and other variables. Schmeidl (1997) finds a significant interaction between genocides/politicides and energy consumption as well as between civil war and population density. Hatton (2017) interacts battles deaths with year fixed effects to explore if wars have become more important as a reason for asylum applications over time. Dreher et al. (2019) find a significant interaction term between conflict and foreign aid. The issue with including interaction effects in regression analysis is that it is usually difficult to include all the possible combinations of factors. There could be important interaction effects or nonlinearities in the data that are not immediately apparent. Machine learning techniques are more effective methods to discover complex structure that was not specified in advance (Varian, 2014; Mullainathan and Spiess, 2017).[3]

2.2. Economic conditions

In the seminal model of migration by Harris and Todaro (1970), the migration decision is mainly driven by the expected net economic returns. The theory predicts that potential migrants would move from worse economic circumstance to better ones. However, migration, in particular international migration, is costly and often not within the ability of the poorest. Later research describes migration as a household strategy rather than an individual decision (Stark, 1993). When supported by their households, migrants are better able to overcome two major constraints to migrating abroad: shortage of investment funds and lack of risk-mitigating mechanisms. In return, migrants send back remittances which may set in motion development dynamics by lessening the constraints faced by households (Taylor, 2002).

---

[2] In the acknowledgments of Branch and Mampilly (2015), the authors note: "Any book on popular protest today cannot help but be an intervention into a terrain of collective hope but also of significant state violence. For this reason, many people who contributed in fundamental ways to the research and writing of this book cannot be named." Apparently, the fear of repercussions in Africa extends beyond the time of the actual protests.

[3] Another important advantage of machine learning algorithms is that they rely on weaker assumptions compared to traditional econometric methods (Gründler and Krieger, 2016).

In the empirical macroeconomic literature, economic reasons for (forced) migration are usually modeled by including GDP per capita in regression analysis. The findings have been contradictory. For example, in Davenport et al. (2003) GDP per capita does not prove to be significant in explaining forced migration. On the other hand, Neumayer (2005) and Hatton (2009, 2016) find negative effect of GDP per capita in the country of origin, while Dreher et al. (2019) find that refugee flows increase with GDP per capita and infer that richer countries produce a larger stock of refugees. Though these mixed results might seem contradictory, there seems to be a consensus in the literature about the underlying explanation: emigration first increases and then decreases with economic development, following an inverted U-shape pattern (see, e.g., Clemens, 2014 and Dao et al., 2018). This transition takes a long time: Clemens (2014) estimates that, at a real per capita growth rate of 2%, the transition would take 63 years for a country starting from $2000 per capita. But most of the datasets used in macrolevel studies of migration do not cover a sufficient length of time to capture the shape of the transition. Thus, depending on which side of the "migration hump" most of the countries in the sample are located, estimations can produce contradictory results.

As in previous research, the dataset used in this study does not cover a sufficient length of time to successfully reproduce the migration hump. However, the machine learning techniques applied in the paper can identify potential nonlinear effects of GDP per capita.

2.3. Other covariates

As control variables, I include factors that have been suggested in the literature to be important determinants of forced migration. Most macrolevel studies control for the population of the origin country for the reason that the larger the population, the larger number of refugees that a country could produce.[4] Further, the structure of polity may play a role: autocratic regimes are expected to create an environment with higher threats than democratic regimes which could induce more people to flee. Scholars have also pointed to political rights, civil liberties, and political terror as major determinants of forced migration. The role of these traditional factors has been extensively covered in earlier research, so there is not much need for further elaboration.

I also include variables that have recently become a subject of increased scrutiny in empirical research on forced migration. One of these variables is foreign aid. In fact, the idea to use foreign aid to address international migration is by no means new: since the 1980s, donors have steadily used aid to improve living conditions in developing countries and thus to reduce incentives to migrate (Bermeo and Leblang, 2015). This has been in striking dissonance with the policy advice offered by the academic literature.[5] Based on an inverted U-shaped curve that describes the relationship between economic development and migration, most researches have argued that aid given to countries located on the upward-sloping part of the curve will actually increase migration flows (Massey, 1988; de Haas, 2007; Clemens, 2014). Empirically, only a few macrolevel studies have focused on the link between foreign aid and forced migration flows. Vogler and Rotte (2000) find no significant link between aid and asylum seekers in Germany. Dreher et al. (2019) also find no evidence that aid reduces refugee flows in the short run, though they observe long-run effects which appear to be driven by the lagged positive effects of aid on growth. Murat (2020) instead finds that asylum inflows from poor countries are significantly and negatively associated with aid in the short run, while inflows from less poor economies show a positive but non-robust relationship to aid.

Climate change is the next factor that has been at the forefront of the migration and development policy debates. Many individuals, households and communities have been suffering from the consequences of climate change, which has forced them to leave their homes. "Climate migrants" and "environmental migrants" have become common terms in policy discussions to describe people displaced by climate factors. On the other hand, there is a consensus among scholars that there is no such thing as a climate or environmental migrant since population displacements are always the result of a multi-dimensional relationship between environmental, political, economic, and social factors (Castles, 2002; Piguet, 2010). The conceptual framework in Black et al. (2011) is widely used in the study of relationship between climate change and migration. By recognizing that environmental factors have heterogenous impact across countries, they emphasize the need for better understanding of migration drivers in specific contexts. Since the effect of climate change is not homogenous, it is not surprising that many studies conducted in different settings have reached mixed conclusions. For example, Afifi and Warner (2008) and Marchiori et al. (2011) find that variables capturing environmental degradation and climate anomalies are positively correlated with international migration, while Naudé (2008) and Beine and Parsons (2015) find no direct effect but rather a more pronounced indirect impact of environmental factors on migration.

The last variable that I take into consideration is the internet. Internet and social media have become increasingly popular channels of information on which migrants base their decisions on whether and where to migrate (Dekker et al., 2018). Hence, scholarly attention to migrants' use of internet prior to and during migration has grown in recent years. While earlier studies find no evidence of asylum seekers relying on the internet (e.g., Gilbert and Koser, 2006), recent research shows that the majority of asylum migrants have access to internet before and during migration (Dekker et al., 2018; Fiedler, 2018; Emmer et al., 2020). The internet offers a relatively cheap and easily accessible access to information. However, the availability of information via internet is not without risks and does not always mean being better informed. The information can be false or biased depending on the interests of the source of the information. For example, smugglers can make false promises about safety of migration routes, migrants can present an idealized picture

---

[4] Most studies control for population by including it as an independent variable in regression analysis. Hatton (2009, 2016), on other hand, uses (log) asylum applications per thousand population as a dependent variable.

[5] The nature and magnitude of the discord between migration research and policy making when it comes to foreign aid is described in more detail in Castles (2009).

of the life in destination countries, and authorities might present information aimed at deterring migrants. There is certainly a need for further research on how truthful versus false information or migration-encouraging versus migration discouraging information on the internet affects migration behavior. Winkler (2017) is the only empirical macrolevel study that I am aware of that examines the impact of internet on international migration. Using the rate of internet penetration in the home country as an explanatory variable, he finds that the internet and migration are negatively corelated. However, the dependent variable in this study is the migration stock in destination countries, which does not allow for understanding the impact that internet might have on the dynamics of migration movements in the digital age.

2.4. Unobserved factors

Macrolevel studies often suffer from selection biases because it is difficult to control for unobserved country characteristics. Thus the inclusion of many variables and their interaction effects, together with the implementation of machine learning techniques that can rigorously deal with underlying estimation issues, is certainly one of the strengths of the current paper. Nevertheless, there are certainly unobserved variables in this study, with social capital being perhaps the most important intervening factor influencing the migration decision. Social capital refers to resources and information that a network of current or former migrants can provide to a potential migrant (Massey et al., 1993; Palloni et al., 2001). A large body of literature has shown that migrants' networks facilitate further migration of people across national borders (Munshi, 2003; Pedersen et al., 2008; Beine et al., 2011). Nowadays, online networks seem to be as effective as physical networks in providing information and advice that are precious to potential migrants. Information that is shared on social media does not only reach an existing group of (potential) migrants to whom it might be directly addressed, but also others with access to the medium. Future research could focus more extensively on how internet and social media interact with and expand the concepts and reality of social capital and migration networks.

The culture of migration could be another unobserved variable affecting migration flows. Cohen and Jónsson (2011) define the migration culture as a growth in the number of entities (individuals, families, communities) that view migration as a favorable strategy. While the migration culture can develop for many reasons, conflicts are often one of the most important causes. After lengthy periods of conflict-triggered population movements, a culture of migration develops and further reinforces the migration flows (Sirkeci and Cohen, 2016).

Previous studies capture unobserved country characteristics by country effects, conduct statistical analysis using fixed-effects regression methods, and report the F-statistic of joint significance of the country fixed effects. In most of the studies, the reported F-tests suggest a rejection of the null hypothesis that the country intercepts are jointly equal to zero. However, none of the previous studies delves deeper into the country fixed effects and we do not know if there are certain countries that produce unusually large numbers of refugees or asylum seekers that cannot be explained by variables included in the model.

3. Data

I use a panel data set consisting of data on forced migration from 45 African countries for the years 1997–2017 when analyzing refugee flows and 1999–2017 when analyzing asylum applications. This sample is drawn with an eye toward maximizing the countries and years under study, defined only by limits on the availability of data. Four African countries that are important sources of forced migration could not be included because of missing data on population or GDP per capita: there are no data on population for Eritrea, and no data on GDP for Libya, Somalia, and South Sudan. The countries included in the analysis are listed in the Appendix.

The data utilized in this study come from a variety of sources. Data on refugee stocks and asylum applications come from the UNHCR Population Statistics online database. The UNHCR keeps record of the annual estimated stock of refuges in a given country who originated in another country and of the number of returned refugees to the home countries. Because data on refugee flows do not exist, I follow previous work and use the stock data to create a flow measure (e.g., Moore and Shellman, 2004; Melander and Öberg, 2007; Beine and Parsons, 2015; Dreher et al., 2019). I take the first difference in the stock of refugee from an origin country (i.e., the value in year *t* minus the value in year *t*-1) and add the number of returned refugees to the country. The weakness of this measure of refugee flows is that it can be affected by other factors on which there are no data, e.g., naturalization, death, migration to a third country, or non-humanitarian residence permits. As a result, the measure of refugee flows can turn negative. Although this variable is measured with noise, previous literature considers it a reasonable approximation of refugee flows.

The UNHCR data on applications for asylum do not suffer from the weakness of the data on stock of refugees. These are generally first instance applications at the border of the destination country, as reported by governments to UNHCR. Thus, the measure of asylum applications is more accurate than the measure of refugee flows. In addition, I use data on asylum applications from the Eurostat database.[6] This database contains data on asylum applications to EU countries from 2008 onwards. Although the Eurostat database covers fewer countries and years than the UNHCR database, comparison of the two data sets could allow for robustness checks of results and additional insights.

Data on conflict come from the Armed Conflict Location and Event Dataset (ACLED). ACLED collects information on internal political conflict disaggregated by date, location, and actor (Raleigh et al., 2010). The latest version of ACLED differentiates between the following types of conflict: a) battles between armed groups; b) violence against civilians; c) explosions/remote violence in which

---

[6] See: https://ec.europa.eu/eurostat/web/products-datasets/-/MIGR_ASYDCFSTA (accessed June 10, 2020).

the tool for engaging in conflict does not require the physical presence of the perpetrator; d) strategic developments that do not involve active fighting but are within the context of a war or dispute; e) riots as a violent form of demonstration; and f) protests as a non-violent form of demonstration, often against a government institution. ACLED also keeps record of fatalities, though the authors recognize that fatality data are particularly prone to bias and inaccurate reporting (Raleigh et al., 2010).

Real GDP per capita in US$ at constant (2011) prices is taken from the Penn World Table version 9.1. Data on population, individuals using internet (% of population), and net official development assistance and aid received (constant 2016 US$) are available from the World Development Indicators database maintained by the World Bank.[7]

Measure of political regime and institutional characteristics are taken from a variety of sources. Polity IV index is available from the Center for Systemic Peace online database. The version used here is the revised polity score, which is a composite index of the level of democracy. The index is based on an evaluation of competitive and open elections, the political participation of citizens, and checks on executive authority. The range of the polity score is from −10 (strongly autocratic) to +10 (strongly democratic). Freedom House is a non-governmental organization that conducts research on political freedom and human rights around the world. Countries are assigned two ratings: one for political rights and one for civil liberties. Each rating ranges from 1 to 7, with 1 representing the greatest degree of freedom and 7 the smallest degree of freedom. In addition, countries are classified as free, partly free, or not free based on the average of a country's political rights and civil liberties ratings. Political terror index is retrieved from the Political Terror Scale website.[8] The index measures levels of violence carried out by a state against its own citizens based on annual reports on human rights practices that are published by Amnesty International, the U.S. State Department, and the Human Rights Watch report. The series used here is based on reports of the U.S. State Department since these are the most complete data available.

Drawing upon recent empirical research focused on environmental determinants of migration (Beine and Parsons, 2015; Missirian and Schlenker, 2017; Abel et al., 2019), I also include measures of climatic conditions in the analysis. In particular, following Beine and Parsons (2015) I use the number of natural disasters; this variable is obtained from a database maintained by the Centre for Research on the Epidemiology of Disasters.[9] In addition, I use country-level data on precipitation and temperature from the CRU TS 4.02 data set obtained from the Climatic Research Unit of the University of East Anglia.[10] Measuring precipitation and temperature in absolute levels is not appropriate because this measure would fail to adequately capture changes from standard climatic conditions. Instead, as argued by Beine and Parsons (2015), it is better to calculate deviations or anomalies from the corresponding averages. I calculate anomalies as the differences of countries' yearly precipitation and temperature from their averages for the period 1997–2017 divided by the corresponding standard deviations.

4. Trends in forced migration and conflict in Africa

Fig. 1 shows the trends in refugee flows and asylum applications from 45 African countries over the years 1999–2017. Both measures of forced migration show a decreasing trend during the early 2000s and an increase since 2010. The erratic movement of the refugee flows is probably a result of the imprecise measurement of this variable.[11] Asylum applications as reported by UNHCR ascended to a peak in 2002, followed by a reversal until 2007. There were sharp upwards steps in 2008 and 2011, followed by an increasingly steep resurgence since 2013 resulting in a staggering peak of more than 200 thousand asylum applications in 2016.

Fig. 1 also reports the number of asylum applications as reported by the Eurostat database. These numbers include only applications made to EU countries. Obviously, applications to EU countries follow a similar trend as the total applications recorded by UNHCR, though with a lag of about one year. The figure shows that most of the asylum applications from the 45 African countries in the sample were made to EU countries: the share of applications reported by Eurostat in the total number of applications reported by UNHCR ranges from 0.62 in 2011 to 0.92 in 2012. More detailed information on the top 15 sources of asylum applications is presented in the Appendix.

While previous studies have used measures of conflict in society to explain refugee flows and asylum applications, none has seriously considered how variation in types of conflict explain the long-run trends. As illustrated in Fig. 2, the nature of conflict in Africa has changed dramatically over the years and, by ignoring this fact, researchers may misinterpret the relationship between conflict and forced migration. Most previous studies have focused on civil wars or battles in society. However, while battles were the predominant type of conflict in the 1990s and the early 2000s, this type of conflict has become relatively less important over time. The surge in non-violent protests and riots since 2010 is particularly noteworthy. Violence against civilians has also gained in importance.[12] Explosions/remote violence and strategic developments remain insignificant as types of conflict and will be excluded in the analysis that follows. A comparison of the trends in Figs. 1 and 2 suggests that forced migration from Africa seems to bear the imprint of both the changing level and the changing nature of conflict in the region.

To complete the picture, Table 1 reports, by five-year periods, statistics for other indicators that will be later used in estimations. The first two rows report measures of direct threat to safety. Total fatalities, as reported in ACLED, decreased drastically during 2005–9

---

[7] Missing data on the internet penetration rate were interpolated for the following country-year observations: Liberia (2006, 2007), Madagascar (2017), Rwanda (2006), and Sudan (2006, 2008, 2009).

[8] The URL for the website is: www.politicalterrorscale.org (accessed December 24, 2019).

[9] See https://www.emdat.be/(accessed September 30, 2020).

[10] See: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.02/(accessed September 14, 2020).

[11] As elaborated in the previous section, the negative value for refugee flows in 2001 is also most likely due to weaknesses of this measure.

[12] The two peaks in violence against civilians in 2002 and 2008 are associated with violence during presidential elections in Zimbabwe.
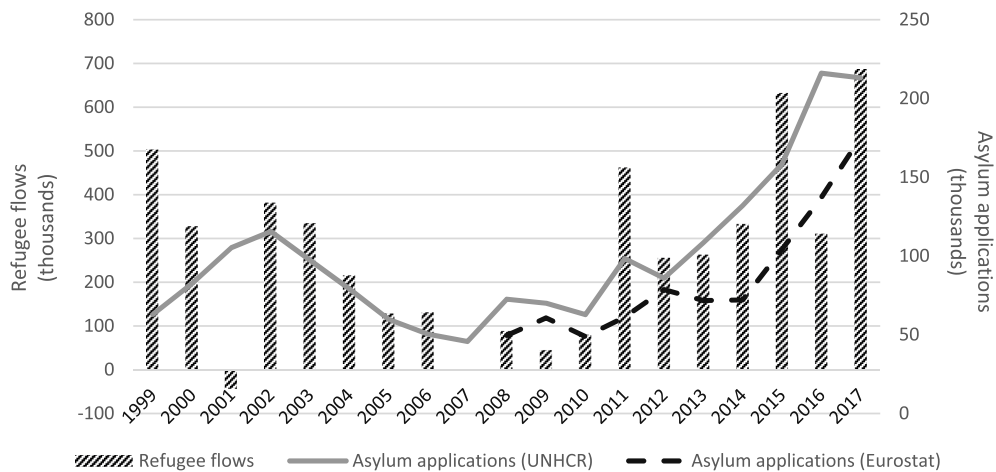
Notes: Results for 45 African countries listed in the Appendix.

Fig. 1. Refugee flows and asylum application from Africa.
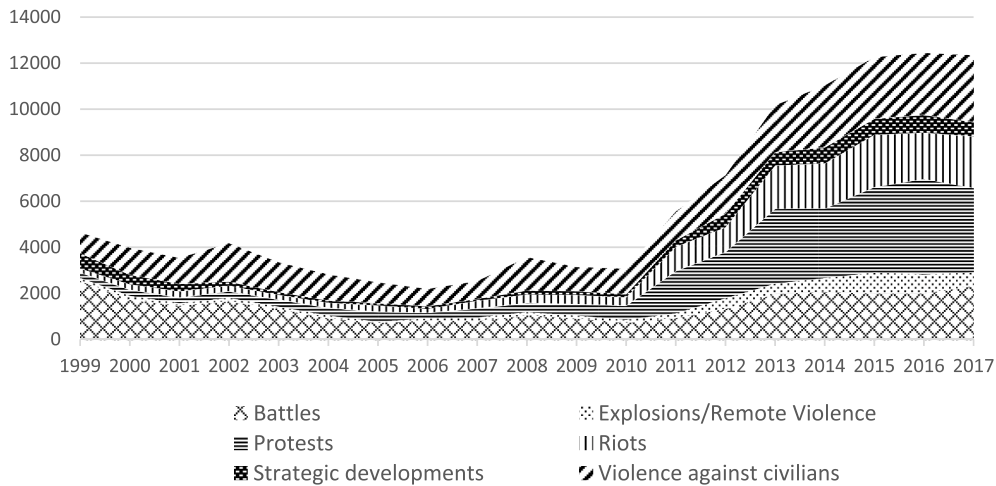
Notes: Results for 45 African countries listed in the Appendix.

Fig. 2. Types of conflict in Africa.

Table 1
Conditions in Africa.

	2000–2004	2005–2009	2010–2014
Total fatalities	120,907	54,821	105,559
Average political terror (scale 1 to 5)	3.05	3.16	2.99
Average total population (millions)	824	931	1058
Average GDP per capita	3767	4497	4869
Average Polity index (scale -10 to 10)	0.53	1.34	1.94
Average political rights (scale 1 to 7)	4.64	4.59	4.67
Average civil liberties (scale 1 to 7)	4.43	4.19	4.35
Proportion "not free" countries	0.37	0.35	0.41
Proportion "partly free" countries	0.47	0.48	0.43
Proportion "free" countries	0.16	0.16	0.16
Average development aid (millions)	569	830	913
Average internet penetration rate	1.15	4.19	11.62
Total natural disasters	494	486	335

Notes: Results for 45 African countries listed in the Appendix.

and then increased during 2010–14, though still remaining below the 1999–2004 level. In contrast to the trend in fatalities, political terror shows a modest rise between 1999–2004 and 2005–9, followed by a decline during 2010–14.

Table 1 reports increasing population in the sample countries, surpassing one billion people in the 2010s. Increasing population is associated with an increasing number of individuals at risk, which might have a positive effect on the number of forcibly displaced people. GDP per capita increased during the period under consideration. As noted earlier, higher average income could make the migration option more viable but less attractive. A measure of democracy derived from the Polity IV database shows a modest trend towards increased democracy. The next five rows in Table 1 are based on the Freedom House measures of political rights, civil liberties, and classification of countries based on their freedom status. These indicators demonstrate hardly any change. Taken together, this indicates that political regime and institutional characteristics probably did not play a significant role in explaining the trends in forced migration. The last three rows in Table 1 show a constant increase in the net official development assistance and aid received and the percent of individuals using internet. The number of natural disasters, on the other hand, shows a downward trend. As previously argued, the recent literature presents mixed evidence about the effect of these factors on international migration: there is little support for direct relationship between environmental conditions, internet penetration or foreign aid and forced migration. Therefore, I have no clear expectation about the overall effect of the trends in these factors on refugee flows.

5. Estimation

5.1. Method

Regression analysis is the most common tool used by existing quantitative research on forced migration. I use instead machine learning techniques that can usefully summarize various sorts of nonlinear relationships and interaction effects. A random forest is an estimation method that builds multiple decision trees on a bootstrapped sample of observations. The random forest is the average of the resulting trees. When building these decision trees, a random sample of predictors is taken at each split to avoid averaging highly correlated trees and thus achieve a larger reduction in variance. The random forest makes the average of the trees less variable and hence more reliable than a single tree. One disadvantage of random forests is that they do not allow for simple summaries of relationships in the data and as such cannot be easily interpreted.[13] However, random forests allow for ranking of variables in importance according to their contribution in the model's prediction accuracy.

Random forests have been recently applied in economic research as an effective way to model complex relationship and uncover hidden patterns in the data. Varian (2014) demonstrates the advantage of the method when analyzing mortgage lending in Boston to see if race played a significant role in determining who was approved a mortgage. In the original study by Munnel et al. (1996), the primary estimation method was a logistic regression where race was included as one of the predictors. The coefficient on race showed a statistically significant negative impact on probability of getting a mortgage for black applicants. By contrast, using a random forest Varian shows that the race variable is relatively unimportant for mortgage approval.

Using a random sample from the 2011 American Housing Survey, Mullainathan and Spiess (2017) compare different estimation methods in predicting house prices. They conclude that random forests can do significantly better than ordinary least squares, even at moderate sample sizes and with a limited number of covariates. In another application of the method, Angrist et al. (2017) use a random forest for classification of economic journal content into fields and styles. The authors claim that the random forest algorithm performs better than other estimation methods such as logistic regression, support vector machines, or the naive bayes algorithm.

Using the results of the random forest analysis as a guidance, I proceed with Lasso (Least Absolute Shrinkage and Selection Operator) regression—a machine learning tool that is probably most familiar to economists. Belloni et al. (2014) present a thorough introduction on how Lasso (and other penalized regression methods) can be used for inference with high-dimensional data. As in the case of the random forest algorithm, Lasso can be a particularly useful tool when dealing with nonlinearities and interaction effects. By setting the regression coefficients of irrelevant variables to zero, Lasso is a relatively straightforward way to select important variables and their interactions.

An important advantage of the random forest method and the Lasso regression is that that these estimators can be computed quite efficiently, and the results are not affected in the presence of highly correlated variables. If I would use the tradition panel data regression, multicollinearity would be a problem: various types of conflicts are highly correlated as are the various measures of political regime or environmental conditions. The multitude of possible interaction effects exacerbate the problem further. Using estimators that produce robust results in the presence of multicollinearity and interactions among variables is important in case of high-dimensional data.

6. Results

I apply the random forest method based on 500 trees with 30–50% of variables sampled for each split of the trees, based on the variance explained criterion. I include both country and year fixed effects. Fig. 3 shows a graphical representation of the variable importance using a parsimonious specification that includes the types of conflict, fatalities, GDP per capita, and population as covariates. The variable importance is computed by recording the total amount by which the mean squared error (MSE) decreases due

---

[13] Recently, there has been some work in the direction of using random forests for inference and estimation of confidence intervals (e.g., Athey et al., 2019).

a) Refugee flows   **b) Asylum applications**
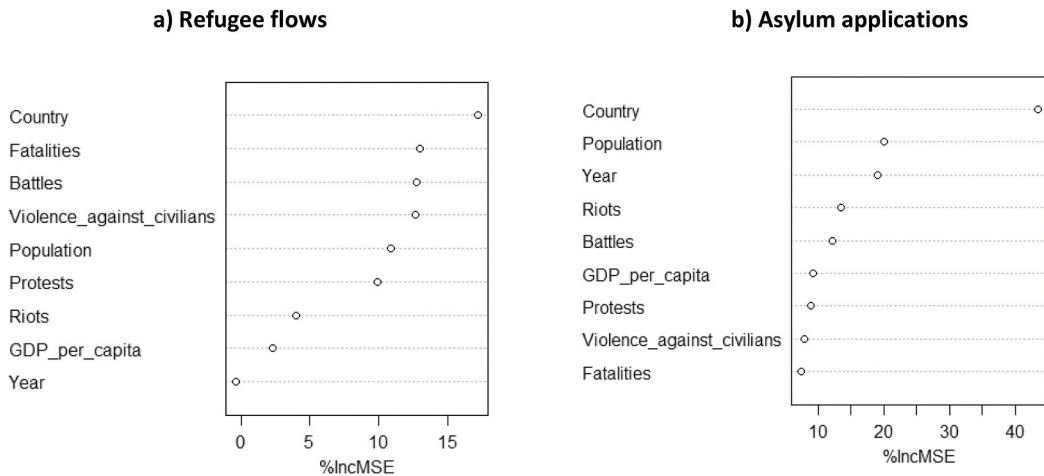
Notes: Results for 45 African countries listed the Appendix. Years 1997–2017 for refugee flows and 1999–2017 for asylum applications.

Fig. 3. Variable importance plot, parsimonious specification.

to splits over a given variable, averaged over all trees. A large value indicates an important variable. The share of variance explained for refugee flows is 18.53 percent and for asylum applications is 81.88 percent. This indicates that the method performs better when estimating asylum applications, which is probably a result of the limitations in measuring the refugee flows discussed earlier.

Fig. 3 shows that country fixed effects are the most important variable in explaining both refugee flows and asylum applications. The fixed effects absorb source country conditions, which seem to be too heterogeneous to be accounted for by the other variables included in the model. The country effect is particularly important in the case of asylum seeking. Fatalities and violent conflicts seem important for refugee flows, but not for asylum applications. While the rising African population is an important driver of the surge in asylum applications, GDP per capita does not seem to play a prominent role in predicting forced migration.

Fig. 4 shows an extended specification that includes traditional covariates which proved statistically significant in previous research: the political terror scale, the Freedom house indices of political rights and civil liberties, and the Polity score of democracy. The addition of these variables leads to a small improvement in the explained variance by the model: 20.87 percent for refugee flows and 82.68 percent for asylum applications. None of the additional variables plays an important role in predicting forced migration.

The results in Figs. 3 and 4 largely agree on the most important variables in explaining refugee flows and asylum applications. In the case of refugee flows, the country effect is the most important variable followed by battles, violence against civilians, and fatalities.
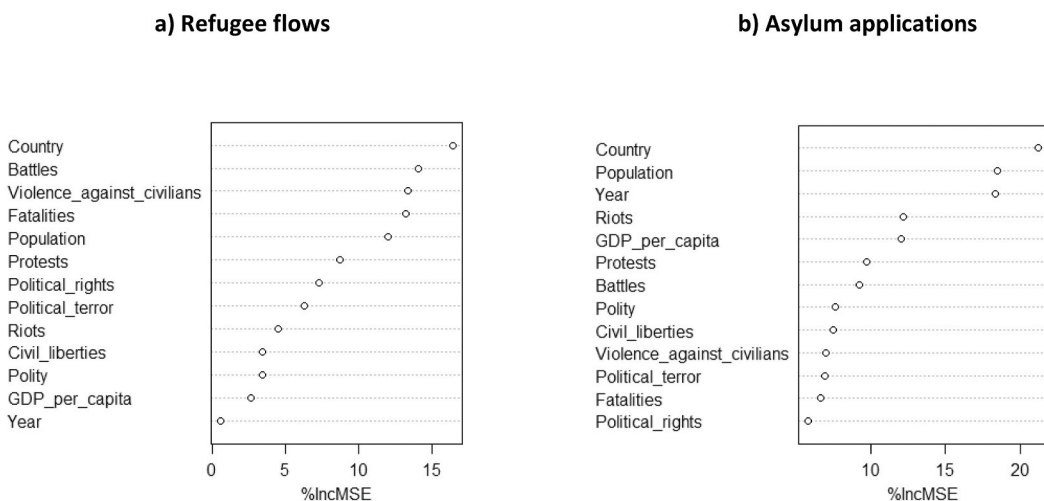
a) Refugee flows   **b) Asylum applications**

Notes: Results for 45 African countries listed in the Appendix. Years 1997–2017 for refugee flows and 1999–2017 for asylum applications.

Fig. 4. Variable importance plot, extended specification.

Given that violent conflicts and fatalities represent the highest level of threat in a society, it should not be surprising that these variables feature prominently in explaining refugee flows. This finding is also consistent with previous research. Note, however, that refugee flows are measured with a lot of noise and the variance explained by the model is relatively low. This might cast some doubt on the validity of this result. In the case of asylum applications, the country effect is again the most important variable followed by the population size, the year effect, and riots. Rioting is usually perceived as less threatening compared to battles, yet this type of conflict plays a role in explaining asylum seeking. While this result is not at odds with previous work, I am unaware of any studies that specifically investigate the important role that riots play relative to other types of conflict.

Fig. 5 presents the full specification including determinants of migration flows that are the focus of recent literature: foreign aid, environmental conditions, and the internet penetration rate. The addition of these variables leads to a further modest improvement in the explained variance by the model: 21.12 percent for refugee flows and 84.01 percent for asylum applications. Except for the internet penetration rate, the newly added variables do not play an important role in predicting forced migration.

To the best of my knowledge, this paper is the first macrolevel study to investigate the role of the internet penetration in explaining forced migration flows and the first to show the importance of this factor in explaining asylum applications. This finding is certainly in line with the recent microlevel literature on the subject based on in-depth interviews with asylum seekers. For many forced migrants the internet and social media have become indispensable channels of information on which they base their decisions on whether to migrate and the destinations where to settle: "… smartphones are lifelines, as important as water and food" (Gillespie et al., 2018, p.1). Yet the internet remains largely absent from the existing body of empirical macroeconomic literature devoted to international migration. In 2017 the average internet penetration rate for the 45 African countries in the sample was 25 percent, ranging from 2.7 percent in Burundi to 62 percent in Morocco. As the number of internet users on the African continent increases in the future, we need a better understanding of the channels through which the internet could affect forced migration.

GDP per capita does not seem to play a prominent role in explaining forced migration (Figs. 3–5). This is consistent with the results documented in the macroeconomic literature showing that conflicts, rather than economic reasons, are the primary determinant of forced migration. However, the result is certainly at odds with the belief held by many individuals in the Western hemisphere that majority of the refugees and asylum seekers are economic migrants in disguise. The result might also seem at odds with evidence from the microeconomic literature (e.g., Aksoy and Poutvaara, 2019), and numerous surveys conducted by the International Organization of Migration and the Mixed Migration Centre showing that irregular migrants from Africa usually cite economic reasons behind their intentions to migrate (e.g., Mixed Migration Centre, 2019). One way to make sense of this seeming contradiction is to consider the changing income distribution, a variable that has been lacking in the recent macroeconomic literature. Rising GDP is usually not uniformly distributed and those who do not benefit from the rising average income or who feel relatively deprived might also feel compelled to leave the country for better living conditions elsewhere (Borjas, 1987; Stark and Taylor, 1991).

Figs. 3–5 show the importance of country effects in explaining forced migration. The country fixed effects capture unobserved time-invariant factors such as persistent conflicts, persistent poverty, geographic position, or proximity to refugee routes. They also capture a culture of migration that might develop over time making migration a favorable strategic option for individuals, families and communities. Perhaps the most important factor captured by the country fixed effects is the social capital as argued before. Other, more local, causes need closer examination. The possible important determinants of forced migration are obviously complex and country-specific.

Using the random forest estimates, I build a Lasso regression model that includes all the variables used in the extended specification above plus interaction effects of the variables identified as most important for explaining asylum applications. By shrinking the co-efficients of irrelevant variables to zero, the Lasso selects the variables (and their interaction effects) that have real information about the response variable. Table 2 shows the top 10 selected variables by their ordinal importance. The tuning parameter in the lasso regressions was chosen by cross-validation.
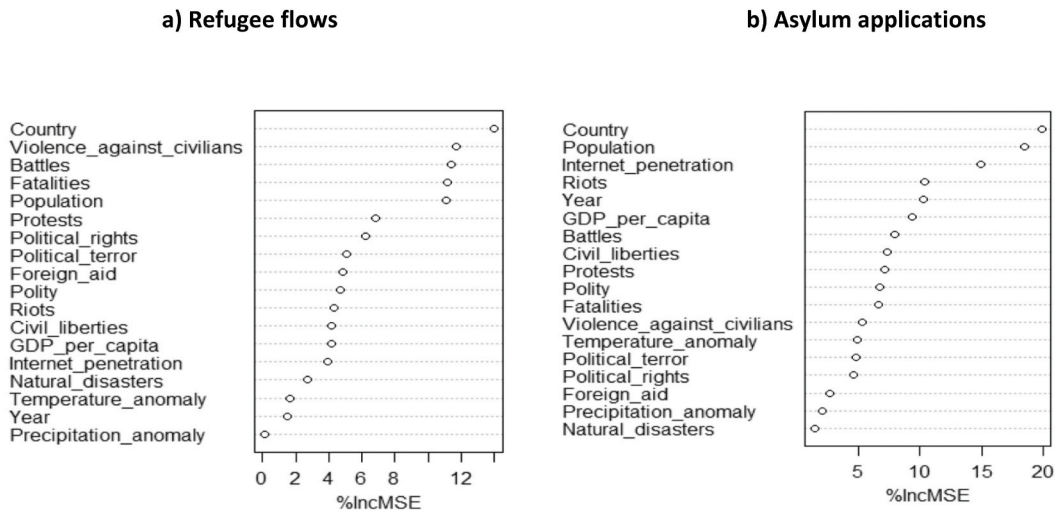
All the countries listed in Table 2 belong to the top 15 sources of asylum applications (see the Appendix). The first column presents the results of a Lasso regression using all the variables included in the random forest estimation presented in Fig. 5 plus country effects interacted with population and riots. Riots seem to play important role in Nigeria and Cote d'Ivoire, while population seems to be an important driver in Guinea, Nigeria, Gambia and Mali. Algeria and the Democratic Republic of the Congo are important sources of asylum applications due to factors not included in the model. This is a curious result since these two countries are the largest by area in Africa. I am not aware of any theory that links the size of country territory to international migration.

The second column in Table 2 includes additional interactions of country effects with internet penetration rates.[14] Riots in Nigeria remain the most important variable, but the rising population and internet penetration in this country also play a role in explaining asylum applications. Except in the case of Nigeria, population does not seem to be an important factor anymore. Instead, increasing internet penetration is important for explaining asylum applications from Gambia, Senegal, Guinea, Cote d'Ivoire, and Mali.

Let us consider asylum applications from Nigeria, the country that is the most important source of asylum seekers because of rising population, riots, and internet penetration. Many observers have cited the Boko Haram insurgency that surfaced in 2009 and intensified in 2014 as the main culprit behind the surge in asylum applications. Others have been skeptical about this view. For example, Abika Dabiri-Erewa—the senior special assistant to the Nigerian President on diaspora and foreign affairs—said that many claims for asylum were spurious: "some who are from the East and West are saying they are running away from Boko Haram."[15] The estimates

---

[14] I have also included interactions of country effects with year effects and GDP per capita, but these interactions never featured among the most important variables.
[15] See: https://www.cfr.org/blog/nigerian-minister-warns-against-nigerian-citizens-seeking-asylum-germany (accessed April 25, 2020).

a) Refugee flows　　　　　　　　**b) Asylum applications**

Notes: Results for 45 African countries listed in the Appendix. Years 1997–2017 for refugee flows and 1999–2017 for asylum applications.

Fig. 5. Variable importance plot, full specification.

Table 2
Results of lasso regressions for asylum applications, 1999–2007 top 10 important variables.

All covariates plus	All covariates plus
- country effects*population	- country effects*population
- country effects*riots	- country effects*riots
	- country effects*internet
Nigeria*riots	Nigeria*riots
Population	Nigeria*population
Algeria	Gambia*internet
Congo, Dem. Rep.	Congo, Dem. Rep.
Guinea*population	Algeria
Nigeria*population	Senegal*internet
Cote d'Ivoire*riots	Guinea*internet
Gambia*population	Nigeria*internet
Mali*population	Cote d'Ivoire*internet
Civil liberties	Mali*internet

Notes: Results for 45 African countries listed in the Appendix. Variables shown by their ordinal importance.

presented in this study seem to provide evidence that it is the riots in urban Nigeria rather than battling Boko Haram in the rural north that mattered more for the steep rise in asylum applications from this country. Further analysis of the Nigerian internal political conflict disaggregated by date, location, and actor could be used to inform future research efforts and policy debates.

6.1. Robustness tests

To check the robustness of the results presented above, I turn to an alternative dataset on forced migration. The Eurostat database contains data on asylum applications to EU countries from 2008 onwards. These data cover smaller number of countries and years than data from the UNHCR database and thus the results are not directly comparable.[16] An advantage of the Eurostat database is that it also contains data on the number of positive decisions on asylum applications. I include the share of positive decisions in the previous year as another explanatory variable in the estimations presented in Fig. 6. The share of variance explained by the model is 75.57 percent.

Fig. 6 shows again the importance of country effects in explaining asylum applications to EU countries. Population and the year

---

[16] The smaller number of observations is probably the reason why estimates based on the Eurostat data showed greater sensitivity to the random seed value.
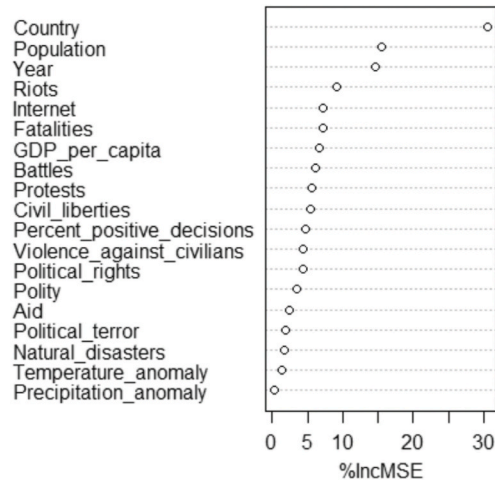
Notes: Results for 45 African countries listed in the Appendix. Years 2008–2017.

Fig. 6. Variable importance plot, Eurostat database.

effect follow in importance. Riots are still the most important type of conflict, but seem to play a lesser role than in the case of global asylum applications. The same applies to the internet penetration rate. Interestingly, the percent of positive decisions on asylum application from the year before does not seem to be important. Overall, the results reinforce the conclusion that important determinants of forced migration are country-specific.

I also tested the robustness of the results in Figs. 3–5 by using a variety of other specifications. In particular, I utilized: averages over three-year periods instead of yearly observations in order to smooth out yearly fluctuations in my baseline results, different lags for independent variables, an aggregate measure of conflict instead of different types of conflict, and limiting the sample to only the top 15 producers of refugees and asylum seekers. The results were qualitatively similar.

7. Conclusion

This study has set to explore which factors play important role in forced migration abroad. I use data on forced migration from 45 African countries for the years 1997–2017. I account for traditional and more recent determinants of international migration, and additionally implement country and year fixed effects to account for unobserved factors. I apply machine learning techniques which are particularly suited for estimation when dealing with nonlinearities and interaction effects in the data. To date, this is arguably the first study that uses machine learning methods to assess the importance of different factors in explaining forced migration.

The study contributes to the literature on forced migration in three ways. First, I document and test for the importance of the changing nature of conflict in Africa, an issue that has been largely ignored not only in the empirical macroeconomic literature but also in policy debates. Civil wars and battles are not the major type of conflict in Africa anymore. For more than a decade now, protests and riots have been a common occurrence and important ways for Africa's urban populations to demand change. My estimates show that riots are indeed the most important type of conflict for explaining asylum applications. As argued by Branch and Mampilly (2019), the current wave of African protests and riots are likely to take an increasingly central role in the future for the reason that there is no end in sight to the conditions giving rise to them. Future research on international migration could benefit from focusing more on different types of conflict in Africa and beyond.

Second, I find that the rate of internet penetration is an important factor for explaining asylum applications. This is a novel finding that deserves further scrutiny in the research of (forced) migration flows. We need a better understanding of the channels through which internet affects migration. An obvious channel would be lowering the communication cost and facilitating the access to information on migration routes and destinations. But there are of course other channels available. For instance, young Africans have been increasingly turning to platforms like Facebook to organize movements demanding more civil liberties, which in turn could lead to government repercussions and increasing migration flows. There is also a pressing need for policy makers to reconsider how they could incorporate the important role played by the internet into programs for migrant outreach. Projects such as InfoMigrants.net launched in 2017 with the aim of providing migrants with verified and balanced information are commendable steps in that direction. Nevertheless, these initiatives are usually designed from the point of view of the host countries. By understanding the channels through which the internet affects migration, the initiatives for migrant outreach could be designed taking into account the point of view of the migrants.

Third, I find that country fixed effects are of primary importance in explaining forced migration flows; an issue that has been ignored by previous work. This implies that there remains a lot to be learned about unobserved factors that are not captured in the existing models of international migration. It also points to a limitation of this study. As in most other studies, the indicators used to explain migration flows are measured at the country level and do not capture within-country differences (e.g., between regions or ethnic groups) in the forces that drive people to abandon their countries. However, there is obviously a great deal of heterogeneity in the circumstances that lead to forced migration that cannot be captured in aggregate-level data. This highlights the need for both researchers and policymakers to focus on countries that are the most important sources of forced migration and scrutinize factors neglected in previous research.

Despite the advantages of machine learning methods in estimating complex relationships and identifying most relevant variables, certain caveats still remain. It is well known that machine learning methods are better suited for prediction purposes rather than causal inference. Considerable work remains to be done to confirm the validity of the conclusions presented here. For example, although this study has clearly demonstrated the importance of internet penetration, this connection might represent only one strand in a much larger and potentially more powerful network of time-varying factors affecting migration which are not included in the estimations.

Given the size and scale of contemporary migration flows and the potential for conflict inherent in diverse societies around the world, policy decisions about international migration will remain among the most important ones in the future. Similarly, sorting out the relative importance of factors affecting migration and integrating them in a model that could be used for both predicting and understanding migration, forced or voluntary, will remain an important empirical task for social scientists. I hope that this study has laid the groundwork for that empirical work.

Declaration of competing interest

I declare no conflict of interest.

Appendix

Table A.1
List of Countries Included in the Study

Algeria□	Congo, Rep.◊	Guinea◊□	Mozambique	Togo
Angola◊	Cote d'Ivoire◊□	Guinea-Bissau	Namibia	Tunisia
Benin	Djibouti	Kenya	Niger	Uganda
Botswana	Egypt◊□	Lesotho	Nigeria◊□	Zambia
Burkina Faso	Equatorial Guinea	Liberia	Rwanda◊	Zimbabwe◊□
Burundi◊	Eswatini	Madagascar	Senegal□	
Cameroon□	Ethiopia◊□	Malawi	Sierra Leone◊	
Central African Rep.◊	Gabon	Mali◊□	South Africa	
Chad	Gambia□	Mauritania	Sudan◊□	
Congo, Dem. Rep.◊□	Ghana□	Morocco□	Tanzania	

Notes: ◊ the country belongs to the top 15 producers of refugees; □ the country belongs to the top 15 producers of asylum seekers as reported by UNHCR.

Table A.2
Top 15 Sources of Asylum Applications 1999–2017

Country	Number of applications
Nigeria	317446
Congo, Dem. Rep.	149082
Algeria	133435
Guinea	109848
Sudan	90383
Ethiopia	79994
Mali	79868
Cote d'Ivoire	76887
Gambia	74073
Cameroon	69083
Egypt	62211
Ghana	56440
Zimbabwe	56313
Morocco	55301
Senegal	54137

Notes: Figures as reported by UNHCR.

Table A.3

Top 15 Sources of Asylum Applications to the EU 2008–2017

Country	Number of applications
Nigeria	228175
Guinea	89585
Congo, Dem. Rep.	76150
Mali	66860
Gambia	65015
Sudan	64690
Algeria	62055
Cote d'Ivoire	59430
Senegal	52175
Morocco	43005
Ghana	42005
Ethiopia	40455
Egypt	33240
Cameroon	26905
Tunisia	26500

Notes: Figures as reported by Eurostat.

References

Abel, G.J., Brottrager, M., Cuaresma, J.C., Muttarak, R., 2019. Climate, conflict and forced migration. Global Environ. Change 54, 239–249.

Afifi, T., Warner, K., 2008. The impact of environmental degradation on migration flows across countries. UNU-EHS Working. Paper 5/2008.

Aksoy, C.G., Poutvaara, P., 2019. "Refugees' and Irregular Migrants' Self-Selection into Europe: Who Migrates where?" IZA Discussion Paper 12800.

Angrist, J., Azoulay, P., Ellison, G., Hill, R., Feng Lu, S., 2017. Economic research evolves: fields and styles. Am. Econ. Rev.: Papers and Proceedings 107 (5), 293–297.

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. Ann. Stat. 47 (2), 1148–1178.

Beine, M., Parsons, C., 2015. Climatic factors as determinants of international migration. Scand. J. Econ. 117 (2), 723–767.

Beine, M., Docquier, F., Özden, C., 2011. Diasporas. J. Dev. Econ. 95 (1), 30–41.

Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. Rev. Econ. Stud. 81 (2), 608–650.

Bermeo, S.B., Leblang, D., 2015. Migration and foreign aid. Int. Organ. 69 (3), 627–657.

Black, R., Kniveton, D., Schmidt-Verkerk, K., 2011. Migration and climate change: towards an integrated assessment of sensitivity. Environ. Plann.: Economy and Space 43 (2), 431–450.

Bohra-Mishra, P., Massey, D.S., 2011. Individual decisions to migrate during civil conflict. Demography 48 (2), 401–424.

Borjas, G.J., 1987. Self-selection and the earnings of immigrants. Am. Econ. Rev. 77 (4), 531–553.

Branch, A., Mampilly, Z., 2015. Africa Uprising: Popular Protest and Political Change. Zed Books.

Castles, S., 2002. Environmental change and forced migration: making sense of the debate. UNHCR Working. Paper 70.

Castles, S., 2009. "Development and migration—migration and development: what comes first? Global perspective and African experiences. Theoria: A Journal of Social and Political Theory 56 (121), 1–31.

Castles, S., Loughna, S., Crawley, H., 2003. States of Conflict: the Causes of Forced Migration to the EU. " Institute for Public Policy Research. Technical Report.

Clemens, M.A., 2014. Does development reduce migration? In: Lucas, R.E.B. (Ed.), International Handbook on Migration and Economic Development. Edward Elgar Publishing.

Cohen, R., Jónsson, G., 2011. Introduction: connecting culture and migration. In: Cohen, R., Jónsson, G. (Eds.), Migration and Culture. Edward Elgar Publishing.

Dao, T.H., Docquier, F., Parsons, C., Peri, G., 2018. Migration and development: dissecting the anatomy of the mobility transition. J. Dev. Econ. 132, 88–101.

Davenport, C.A., Moore, W.H., Poe, S.C., 2003. Sometimes you just have to leave: domestic threats and forced migration, 1964-1989. Int. Interact. 29 (1), 27–55.

de Haas, H., 2007. Turning the tide? Why development will not stop migration. Dev. Change 38 (5), 819–841.

Dekker, R., Engbersen, G., Klaver, J., Vonk, H., 2018. Smart refugees: how Syrian asylum migrants use social media information in migration decision-making. *Social Media + Society* January-March 1–11.

Dreher, A., Fuchs, A., Langlotz, S., 2019. The effects of foreign aid on refugee flows. Eur. Econ. Rev. 112, 127–147.

Emmer, M., Kunst, M., Richter, C., 2020. "Information seeking and communication during forced migration: an empirical analysis of refugees' digital media use and its effects on their perceptions of Germany as their target country. Global Media Commun. 16 (2), 167–186.

Fiedler, A., 2018. The gap between here and there: communication and information processes in the migration context of Syrian and Iraqi refugees on their way to Germany. Int. Commun. Gaz. 81 (4), 327–345.

Gilbert, A., Koser, K., 2006. Coming to the UK: what do asylum-seekers know about the UK before arrival? J. Ethnic Migrat. Stud. 32 (7), 1209–1225.

Gillespie, M., Osseiran, S., Cheesman, M., 2018. Syrian refugees and the digital passage to Europe: smartphone infrastructures and affordances. Social Media + Society 4, 1–12.

Gründler, K., Krieger, T., 2016. Democracy and growth: evidence from a machine learning indicator. Eur. J. Polit. Econ. 45 (S), 85–107.

Harris, J.R., Todaro, M.P., 1970. Migration, unemployment and development: a two-sector analysis. Am. Econ. Rev. 60 (1), 126–142.

Hatton, T.J., 2009. The rise and fall of asylum: what happened and why? Econ. J. 119 (535), F183–F213.

Hatton, T.J., 2016. Refugees, asylum seekers, and policy in OECD countries. Am. Econ. Rev.: Papers & Proceedings 106 (5), 441–445.

Hatton, T.J., 2017. Refugees and asylum seekers, the crisis in Europe and the future of policy. Econ. Pol. 32 (91), 447–496.

Koser, K., 2010. Dimensions and dynamics of irregular migration. Popul. Space Place 16 (3), 181–193.

Marchiori, L., Maystadt, J.-F., Schumacher, I., 2011. The impact of weather anomalies on migration in sub-saharan Africa. J. Environ. Econ. Manag. 63 (3), 355–374.

Massey, D.S., 1988. Economic development and international migration in comparative perspective. Popul. Dev. Rev. 14 (3), 383–413.

Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, J.E., 1993. Theories of international migration: a review and appraisal. Popul. Dev. Rev. 19 (3), 431–466.

Melander, E., Öberg, M., 2007. The threat of violence and forced migration: geographical scope trumps intensity of fights. Civ. Wars 9 (2), 156–173.

Missirian, A., Schlenker, W., 2017. Asylum applications respond to temperature fluctuations. Science 358 (6370), 1610–1614.

Mixed Migration Centre, 2019. 4Mi Snapshot: what Drives Migrants and Refugees to and through Libya? Mixed Migration Centre Report.

Moore, W.H., Shellman, S.M., 2004. Fear of persecution: forced migration, 1952-1995. J. Conflict Resolut. 48 (5), 723–745.

Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31 (2), 87–106.

Munnel, A.H., Tootell, G.M.B., Browne, L.E., McEneaney, J., 1996. Mortgage lending in Boston: interpreting HMDA data. Am. Econ. Rev. 86 (1), 25–53.

Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the US labor market. Q. J. Econ. 118 (2), 549–599.

Murat, M., 2020. Foreign aid, bilateral asylum immigration and development. J. Popul. Econ. 33 (1), 79–114.

Naudé, W., 2008. Conflict, Disasters, and No Jobs: Reasons for International Migration from Sub-saharan Africa. UNU-WIDER Research. Paper 2008/85.

Neumayer, E., 2005. Bogus refugees? The determinants of asylum migration to western Europe. Int. Stud. Q. 49 (3), 389–409.

Palloni, A., Massey, D.S., Ceballos, M., Espinosa, K., Spittel, M., 2001. Social capital and international migration: a test using information on family networks. Am. J. Sociol. 106 (5), 1262–1298.

Pedersen, P.J., Pytlikova, M., Smith, N., 2008. "Selection and network effects – migration flows into OECD countries 1990–2000. Eur. Econ. Rev. 52 (7), 1160–1186.

Piguet, E., 2010. Linking climate change, environmental degradation, and migration: a methodological overview. Climate Change 1 (4), 517–524.

Raleigh, C., Linke, A., Hegre, H., Karlsen, J., 2010. Introducing ACLED: an armed conflict location and event dataset. J. Peace Res. 47 (5), 651–660.

Schmeidl, S., 1997. "Exploring the causes of forced migration: a pooled time-series analysis, 1971–1990. Soc. Sci. Q. 78 (2), 284–308.

Sirkeci, I., Cohen, J.H., 2016. Cultures of migration and conflict in contemporary human mobility in Turkey. Eur. Rev. 24 (3), 381–396.

Stark, O., 1993. The Migration of Labor. Blackwell, Oxford and Cambridge, MA.

Stark, O., Taylor, J.E., 1991. Migration incentives, migration types: the role of relative deprivation. Econ. J. 101 (408), 1163–1178.

Taylor, J.E., 2002. The new economic of labour migration and the role of remittances in the migration process. Int. Migrat. 37 (1), 53–88.

UNHCR, 2019. Global Trends: Forced Displacement in 2018. Technical Report.

Varian, H.R., 2014. Big data: new tricks for econometrics. J. Econ. Perspect. 28 (2), 3–28.

Vogler, M., Rotte, R., 2000. The effects of development on migration: theoretical issues and new empirical evidence. J. Popul. Econ. 13 (3), 485–508.

Winkler, H., 2017. How does the internet affect migration decisions? Appl. Econ. Lett. 24 (16), 1194–1198.