Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 159 (2016) 338 – 347

Procedia Engineering

www.elsevier.com/locate/procedia

Humanitarian Technology: Science, Systems and Global Impact 2016, HumTech2016, 7-9 June 2016, Massachusetts, USA

Prediction of Approved Asylum Seekers in European Countries

Rong Zhang[a,*] , Mingyue Fan[b]

[a]*Department of Statistics, University of Pittsburgh, 1800 Wesley W. Posvar Hall, Pittsburgh, 15260, the United States*
[b]*School of Information Sciences, University of Pittsburgh, 135 N. Bellefield Ave., Pittsburgh, 15260, the United States*

Abstract

This paper is to develop a statistical model that predicts the number of approved asylum seekers using profiles of 32 European countries in a panel data setting. The ordinary least squares (OLS) linear regression, the fixed effects and the random effects models are explored and compared. In addition, the clustering results in 2014 are compared with manually generated clusters. The evaluation results show that the fixed effects model (with "country" and "year" effects) wins out. The k-means clustering and the hierarchical clustering with complete-link have a better performance within a classification on the number of approved asylum seekers. Our study finds the related country profiles, which build a bridge to the study of refugee problem.

Peer-review under responsibility of the Organizing Committee of HumTech2016

Keywords: Approved asylum seekers; country profiles; OLS linear regression; fixed effects; random effects; clustering; refugee problem

1. Introduction

The refugee problem is one of the most common and perennial social phenomena that has been existing throughout human history. It has recurrently occurred in a vicious manner in Europe where the migration trend is reshaped by political upheaval in the Middle East, Africa, and South Asia. The number of illegal border-crossing detections in the European Union (EU) started to surge in 2011. Migrants and refugees streaming into Europe from Africa, the Middle East, and South Asia have presented European leaders and policymakers with their greatest challenge since the debt crisis (Park [10]). How many refugees each country will take has been a complex and debatable problem. Although the EU has launched refugee quota plans, the EU countries seem not to take the quota well. The report from Nolan [9] shows that the September EU plan has been described as a waste of time by Péter Szijjártó, the Minister of Foreign Affairs and Trade of Hungary.

One of the many reasons that the compulsory quota system is unsatisfactory is that many more people have entered the European Union since the original debate on the distribution of 120,000 immigrants. Since the crowd is a mixture of refugees and migrants, and it is difficult to separate them, acceptance of the quota may cause a wave of massive migration.

An asylum seeker is defined as a person fleeing persecution or conflict, and therefore seeking international protection under the 1951 Refugee Convention on the Status of Refugees; a refugee is an asylum seeker whose claim has been approved. The United Nations also considers migrants fleeing war or persecution to be refugees, even before they officially receive asylum (Syrian and Eritrean nationals, for example, enjoy prima facie refugee status). An economic migrant, by contrast, is a person

* Corresponding author. Tel.: +1-412-877-6558.
E-mail address: roz16@pitt.edu

Peer-review under responsibility of the Organizing Committee of HumTech2016
doi:10.1016/j.proeng.2016.08.203

whose primary motivation for leaving his or her home country is economic gain. The term "migrant" is seen as an umbrella term for all three groups (Park [10]).

On November 13th, 2015, Paris, the capital of France, was attacked by terrorists, one of whom was later revealed as a registered Syrian refugee. This tragedy resulted in 129 deaths and over 200 wounded, and has heavily impacted the EU's already fragile migration strategies. However, the report from Kingsley [6] has suggested that instead of a "close border strategy," a more logical response is to create an organized system of mass-resettlement from the Middle East itself.

How many refugees each country will take is one of the most challenging problems for policy makers and European leaders today but its currently proposed solution is under sever debate and seldom taken.

Moraga and Rapoport [8] propose a comprehensive system of tradable refugee-admission quotas with a matching mechanism in the context of the Syrian refugee crisis. The proposal has three main components: a distribution of refugee quotas across countries in terms of some criterion, a matching mechanism allowing both refugees and receiving countries to choose their preferences, and a permission to trade quotas assigned by the European Commission to make it less costly for the countries to host refugees with more benefits. The first component is the focus of our study. The authors have also stressed its importance as a prerequisite for the countries to implement the coordinated policy. There are quite a few studies trying to assess the "fair" share of refugees each EU country should take with respect to its "capacity." Thielemann et al. [12] suggest that more than one third of refugees should have been moved to other countries in the EU in 2010 based on a "combined capacity index" assessment for asylum related costs and pressures between the receiving countries. Wagner and Kraler [13] compared seven quota distribution rules based on a different key (a weighted addition of country profiles like population size, GDP, unemployment rate and geographical area) applied to the mean number of asylum applications in the EU 28 from 2009 to 2013. In addition, the European Commission ([2], [3]) introduces an eighth quota distribution rule, which is basically a replication of rule number 1 that replaces the geographical element with the average number of spontaneous asylum applications and the number of resettled refugees per one million inhabitants from 2010 to 2014. The previous studies take the country profiles into consideration in determining a refugee quota for each EU country; however, the number of country profiles is too small and all of them are taken as "weights" or "percentages," neglecting the quantitative side. Also, the quotas are derived from the number of asylum applications rather than a model with the number of approved asylum seekers or the number of refugees directly.

A data-driven prediction of refugee acceptance with data mining procedure proposed in this study is expected to be an improvement to the first step of the new system and a new reference to policy makers and European leaders. With this study, we intend to develop a statistical model to predict how many refugee applications would be approved and explore the relationship between refugee acceptance and country profiles based on a panel data setting over the time period from 2008 to 2014 for 32 European countries including the 28 EU members, as well as Iceland, Liechtenstein, Norway and Switzerland.

2. Data description

We obtain the number of approved asylum applications from 2008 to 2014 for the 32 European countries from *Eurostat: http://ec.europa.eu/eurostat/data/database.*, the website of the statistical office of the European Union, which provides statistics in European countries. The data of country profiles is obtained from *the World Bank Data: http://data.worldbank.org/*. Each country contains 60 attributes, including population, economics, education, health and environment.

2.1. Data merge

The two data sets are combined to create a new data frame. The dependent variable is the number of approved asylum application and the independent variables are the attributes of country profiles.

A location variable is also included as geographical area element to complement the country profiles. We divide the 32 countries into three categories: frontier, second-frontier and rear countries. The frontier countries are those border upon refugee countries, for example, Turkey, Greece and Egypt. The second-frontier countries are most central European countries which border on the frontier countries. The rear countries are those farthest to the refugee countries, like most northern European countries. Location variable is numbered with 1 to 3 to represent frontier, second-frontier and rear countries respectively.

2.2. Data pre-processing and missing data implementation

Some variables of country profiles have much larger values. In order to avoid the influence of different units, the unit of "total population" is transformed from "person" to "million people." Moreover, the units of "GNI using Atlas method (current US $)," "GNI on purchasing power parity (current international $)," "GDP (current US $)," "personal received remittances (current US $)" and "foreign direct investment net inflows (current US$)" are changed to "current billion US $" and "current billion international $." In addition, the units of "GNI per capita using Atlas method (current US $)" and "GNI per capita on purchasing power parity (current international $)" are transformed to "current thousand US $" and "current thousand international $."

A covariate with abundant missing entries is lack of information to show the relationship with the dependent variable. We delete the variables in which more than 20% of all (more than 45) are missing values. Therefore, there are 36 covariates left.