A Gravity Analysis of Refugee Mobility Using Mobile Phone Data

Michel Beine[1], Luisito Bertinelli[2], Rana Cömertpay[3], Anastasia Litina[4], and Jean-François Maystadt[5]

[1,2,3]*Department of Economics and Management, University of Luxembourg, L-1359 Luxembourg*

[4]*Department of Economics, University of Ioannina, University Campus, 455 00, Ionnina Greece*

[5]*IRES/LIDAM, UCLouvain; Fonds National de la Recherche Scientifique, Belgium*

December 30, 2020

Abstract

The objective of this study consists in analyzing the determinants of the internal mobility of refugees in Turkey. We track down this mobility relying on geolocalized mobile phone calls data and bring these measures to a micro-founded gravity model in order to estimate the main drivers of refugee mobility across 26 regions in 2017. Our results show that the movements of refugees are sensitive to income differentials and contribute therefore to a more efficient allocation of labor across space. Comparing these findings with those of individuals with a non-refugee status, we find that refugees are more sensitive to variations of income at origin and to distance, while less responsive to changes in income at destination. These findings are robust to the way mobility is inferred from phone data and to the choice of the geographical unit of investigation. Further, we provide evidence against some alternative explanations of mobility such as the propensity to leave refugee camps, transit through Turkey, social magnet effects and sensitivity to agricultural business cycles.

Keywords: Refugee Mobility, Gravity Model of Migration, Forced Displacement, Mobile Phone Data, News Media, Poisson Pseudo-Maximum Likelihood

JEL-Classification: J6, 015

1 Introduction

Over the last decade there has been a surge in the number of international refugees in search for a better living outside their country of birth. The civil conflicts in Syria and in other countries have created an almost unprecedented humanitarian crisis leading to about 26 millions refugees at the end of the year 2019 (United Nations High Commissioner for Refugees, 2020). The magnitude of these developments has created a need to better understand the dynamics behind the movements of these individuals in search of durable solutions. On the academic side, economists and other social scientists have started studying key questions related to refugees. As reviewed by Becker and Ferrara (2019) and Maystadt et al. (2019), several papers have addressed a set of important economic questions such as the economic and political consequences of refugees in receiving countries. However, the question of the mobility of these refugees subsequent to their initial settlement has received much less attention form scholars. To the best of our knowledge, both the characterization of the extent to which refugees move after their initial arrival and the understanding of the patterns of this mobility have not really been addressed in the existing economic literature.

The questions regarding the determinants of the mobility of refugees subsequent to their early settlement remain nevertheless of primary importance. Such understanding may be first important for relief operations to better target those in need of assistance. Researchers investigating the consequences of forced displacement in hosting areas have either assumed that forcibly displaced people choose their location in a quasi-random way (Godøy, 2017; Grönqvist et al., 2012) or have overlooked the dynamic nature of such location decision. Anecdotal evidence suggests nevertheless that a significant share of refugees may move multiple times within their country of asylum (Bose, 2013, 2014), suggesting that they follow some systematic patterns. If one aims to uncover these patterns, knowing whether refugees tend to be stuck in their initial place of settlement or, on the contrary, show a high propensity to move to other locations is important for authorities in order to supply the right level of public assistance to these individuals. If refugees do not move, public infrastructures at entry points are likely to be quickly subject to congestion and need to be expanded. On the contrary, if they tend to be mobile, it is important to know where these refugees tend to relocate, which in turn raises the question of the specific determinants of their mobility. To address these questions, one can fortunately rely on a large literature which has looked at the identification of the main determinants of the internal and international movements of economic migrants, especially using gravity models (Beine et al., 2016). Two broad types of factors emerge from this literature, namely factors shaping the level of attractiveness of the potential locations and factors generating frictions to the mobility of individuals between these various potential locations.

The first factor, which is not specific to refugees, is income or wage differentials. The existence of differences in the level of expected income is obviously the main robust determinant of movements of economic migrants, both internationally and internally. It is also a predictor of the selectivity of these migrants, for instance in terms of skills (Grogger and Hanson, 2011). In that respect, the sensitivity of the movements of refugees to income differentials is an important element of knowledge. If refugees are allowed to move and work, and respond significantly to income differentials, their mobility process constitutes a factor of efficiency of the allocation of labor across space since they tend to move from low to high productivity locations. Equally important is the comparison of this sensitivity with other types of workers such as legal economic immigrants.[1] The sensitivity to the second factor, namely migration costs such as distance, is worth being

[1]A second factor of attractiveness, more specific to refugees, involves the level of local aid these individuals can receive. Aid provided upon arrival is a crucial element to overcome the distress the refugees often face when escaping urgent and dangerous contexts. Nevertheless, if refugees tend to be attracted by more generous locations, the provision of aid might undermine the efficiency of the process of labor reallocation across space. This question connects with the literature on the social magnet effect that has looked to what extent migrants develop opportunistic location strategies with respect to the level of public transfers (Razin and Wahba, 2015). In the last part of the paper, we investigate this issue and look at whether refugees tend to be attracted by the level of aid provided by local authorities.

analyzed. The existence of migration costs has been shown to be of primary importance in the literature dealing with economic migration. Accounting for migration costs is key for predicting migration flows as well as the type of migrants settling in each location (Chiquiar and Hanson, 2005). This question is also highly relevant for determining the optimal allocation process of refugees. If refugees are highly sensitive to migration costs, they might be unable to take advantage of attractive locations and could be highly dependent on welfare benefits given by local authorities. This has in turn considerable consequences for public finance. Descriptive evidence provided by the World Bank (2018) tends to show that refugees are likely to move over shorter distances, suggesting a higher sensitivity to migration costs. Sound econometric investigation is nevertheless desirable to confirm this initial piece of evidence.

A primary reason for the absence of existing studies about the determinants of the mobility of refugees lies in the difficulty of relying on the traditional measures of mobility used in the migration literature. Studies usually rely on administrative data to track movements of individuals across and within countries. Data from censuses or population registers are often the main source to identify movements of natives and migrants across locations. Some alternative measures, based for instance on fiscal reports or health data, have also been used to measure movements of households across regions at higher time frequencies (see for instance Hatton and Trani (2005) on British data or Beine and Coulombe (2018) on Canadian data). Due to the elusive status and the unstable situation characterizing refugees, these sources cannot unfortunately be used for tracking consistently their displacement, which calls for alternative creative solutions of data collection. In this paper, we propose to implement such a solution by tracking movements of refugees through the use of cell phone data in Turkey.[2] We exploit geolocalized call

[2]Furthermore, even when administrative data are available, the mobility patterns identified through the use of mobile phone data are found to contrast with the refugee presence revealed by such data. Such a discrepancy should warn researchers about some of the pitfalls in using a one-shot distribution of refugees within a country. Indeed, the movement of refugees has often been used as a natural experiment to assess the impact of migration (Tumen, 2015). Most of these studies have considered the potential

detail records provided by Salah et al. (2018) within the Data for Refugees Turkey (D4R) challenge. More specifically, we look at the location of 100,000 randomly selected mobile transactions (involving 50,000 refugees and 50,000 non-refugees) recorded by cell towers to define likely decisions to move across 26 regions in Turkey. This allows us not only to compute bilateral migration flows at a quarterly frequency and at the provincial level for refugees but also to compare our findings with those obtained on a sample of non-refugees (natives and legal immigrants). While the use of phone data to track individuals is not new (see among others Blumenstock et al. (2016), Wesolowski et al. (2012) and Deville et al. (2014)), we are, to the best of our knowledge, the first ones to use this approach to measure mobility in order to characterize the determinants of internal mobility of refugees.[3]

Based on the mobility measures inferred from phone data, we estimate the determinants of refugee movements across 26 Turkish regions in 2017. Turkey is an interesting case to study the mobility of forcibly displaced people within a country of asylum. The movement of Syrian refugees in Turkey started in 2011 as a result of the Syrian Civil War. Currently the official statistics report approximately 6.6 million refugees (United Nations High Commissioner for Refugees, 2020). The ongoing conflict induced many refugees to remain in Turkey while others moved to farther European countries. Furthermore,

threat of native displacement but the threat of refugee displacement in the country of asylum has rarely been discussed. The nature of the resulting bias will depend on the composition of such re-displaced refugees. The risk has been mostly overlooked since extant studies on the impact of refugees on hosting economies, including in Turkey, mostly rely on registration or administrative data. Such data would not be able to capture such internal patterns of displacement. For Turkey, exceptions are Altindag et al. (2020) and Tumen (2019) who are able to exploit time variation in the concentration of refugees in Turkey. The issue is acknowledged in Akgunduz et al. (2018) but the proposed use of a distance-based IV approach constitutes only a partial solution since it will capture the LATE effect of the initial population shock. The issue is particularly relevant in Turkey since refugees have spread to the rest of Turkey from the second half of 2014 onward (Tumen, 2016; Ceritoglu et al., 2017).

[3]Beine et al. (2019) characterize internal mobility as a measure of integration of refugees in Turkey and focus on the response to news about protests and demonstrations. In contrast to this paper, they do not rely on micro-foundations to derive the gravity equation. There exist major differences on the methodological side: no fixed effects in the specification, regional income proxied by nightlight data from satellites, analysis only at the NUTS 3 geographical level to list the main ones. While they find that refugees are sensitive to distance, they do not find that they move in response to income differentials, a result that contrasts with the findings of this paper.

due to a bilateral agreement reached in 2015, borders between the EU and Turkey have been closed to refugees, which lowers significantly their movements for the purpose of transiting to European countries. According to official figures, in 2015, the population of Syrian refugees was approximately 2.8 million, in 2016 about 3.1 million while in 2017 approximately 3.8 million (United Nations High Commissioner for Refugees, 2017a). The number of refugees living in camps was approximately 250,000 in 2017, i. e., less than 7% (United Nations High Commissioner for Refugees, 2017b) compared to approximately 260,000 in 2016, i. e., less than 8.5% (United Nations High Commissioner for Refugees, 2016) and 270,000 in 2015 (Bahçekapili and Çetin, 2015), approximately 9.6%. Overall, the number of the Syrian population in camps decreases over time, despite the actual increase in the number of refugees, thus suggesting an increasing mobility over time and highlighting the need to further uncover its determinants.

While the Turkish law does not grant relocation rights but only temporary protection status, yet this temporary status is accompanied with the right to apply for a work permit in certain areas and certain professions. As such, internal mobility within the borders of Turkey is rather free for Syrian refugees. In some cases, relocation has even been encouraged in an attempt to close down and relieve some camps (United Nations High Commissioner for Refugees, 2019). The freedom to move, combined with the right to work, are important elements for the purpose of eliciting the determinants of the mobility of refugees and to compare their pattern with the one of the non-refugee population. There is an emerging literature assessing the impact of refugees on the hosting economies in Turkey[4] but, in line with the general literature on refugees, little attention has been paid to the determinants of the mobility of refugees in this country.

In order to estimate the sensitivities of movements of refugees to income and distance, we bring our measures of mobility based on cell phone data to a standard gravity equation.

[4]The related literature has analyzed the impact of refugees on the labor markets (Del Carpio and Wagner, 2015; Tumen, 2016; Ceritoglu et al., 2017), firm entry and performance (Akgunduz et al., 2018; Altindag et al., 2020), consumer prices (Balkan and Tumen, 2016), and high-school enrollment (Tumen, 2019).

The gravity model is itself derived from a Random Utility Model applied to refugees (and non-refugees) in which factors of attractiveness and friction enter in the deterministic component of utility associated to each location. In the benchmark estimations, the model is estimated on data defined at the NUTS–2 level involving 26 Regions.

Our main results can be summarized as follows. First, we find that refugees react to income differentials. Refugees tend to leave relatively poor locations and are attracted by wealthier ones. Second, their sensitivity to income differs in two ways from the one of non-refugees. On the one hand, refugees react less to income levels at destination than non-refugees. One possible explanation might be the lack of information available to refugees. On the other hand, non-refugees do not show any propensity to leave relatively poor areas, which might be due to a higher degree of attachment to their current location. In that sense, allowing refugees to move or incentivizing refugees with reliable information may contribute to a more efficient allocation of the labor force across space. Third, refugees are indeed more sensitive to distance than other individuals, even though the discrepancy is not as high as one could expect (their estimated elasticity is about 35% higher compared to non-refugees). Finally, refugees appear to be sensitive to humanitarian aid and asylum grants. Nevertheless, while we find that humanitarian aid and asylum grants discourage refugees from migrating, we do not find any evidence of a social magnet effect, i. e. a systematic attraction by more generous locations. Our results resist a set of robustness checks. They are robust to alternative procedures through which we map phone calls to mobility measures. The results are broadly similar when we change the geographical definition of our unit of analysis (using larger and smaller regions than in the benchmark analysis). We also provide evidence that the economic motivation of movements that we document for refugees is not confounded by alternative explanations of mobility such as the propensity to leave refugee camps, motivations of transit in Turkey or seasonal moves driven by agricultural business cycles.

The remainder of our paper is structured as follows. Section 2 derives our gravity equation from a Random Utility Model (Section 2.1) that we develop to characterize location choices of refugees and non-refugees (Section 2.2). Section 3 presents the data (Section 3.1) and some descriptive statistics to help understand better the sample of our study (Section 3.2). Section 4 provides the empirical results for our main research question (Section 4.1) and shows that our findings survive various robustness tests (Section 4.2). Section 5 provides some implications for policy and concludes.

2 A RUM model for refugees

The economic literature has for long widely relied on the gravity model to understand migration decisions (Ravenstein, 1985, 1989). Despite its simplicity, the gravity model has shown impressive predictive power, making it an essential tool for forecasting mobility between and within countries (Crozet, 2004; Mayda, 2010; Garcia et al., 2015; Beine et al., 2016). As the main determinants of migration, the gravity model has identified differentials in employment opportunities and income per capita between the areas of origin and destination, together with the geographical and cultural distance, as proxies for migration costs. In this paper, we build on the large literature on the gravity model and apply it to the mobility of refugees across Turkish regions. In that respect we follow two recent strands of that literature.

First, while the applicability of the gravity model to forcibly displaced people is limited to international movements of asylum-seekers to OECD countries (Hatton, 2009, 2016, 2017), it seems that the same set of factors explains the movements of economic migrants and refugees across countries, albeit with different intensities. In particular, geographical and political factors have stronger weight for refugees or asylum-seekers compared to economic ones. Such conclusion is confirmed in the cross-sectional analysis of the gravity model proposed by the World Bank (2018). The applicability of the gravity model to forced displacement shows that the forced nature of population movement should not hide the potential agency played by forcibly displaced people in their migration decision (Ibánez, 2014; Maystadt et al., 2019). However, the cross-country nature of this literature is limited in shedding light on the determinants of mobility in complex emergencies. To the best of our knowledge, we are the first ones to apply the gravity model to the mobility of forcibly displaced populations within a recipient country in conjunction with highly disaggregated phone data that allows to track people in a consistent way.

Second, the recent literature has emphasized the need for using sound micro-foundations to derive gravity equations applied to migration. Anderson (2011) derives a gravity equation in a full equilibrium framework, emphasizing the importance of the concept of multilateral resistance. Grogger and Hanson (2011) and Beine et al. (2011) derive gravity equations from a Random Utility Model (RUM). The use of a RUM model allows to uncover explicitly the various underlying assumptions on which the gravity equation relies. We follow this route and to that aim develop a RUM model from which we derive our gravity equation.

2.1 A RUM Model of location decision

Let us consider the location choice for an individual of type $l = \{R, NR\}$ who has to decide where to locate among K ($k = \{1, ..., K\}$) potential locations over the next period of time t ($t = 1, ..., T$). We denote location i as his current location. For refugees ($l = R$), this location can be seen as the first location in which he has settled when arriving in Turkey from his country of origin. For non-refugees ($l = NR$), location i is the most recent residing place. Suppose that individuals work in their living place (no commuting) and that every individual is allowed to work and to locate freely among all the K potential locations.

The level of utility of a type-l individual associated to staying in his initial location i

is given by the following equation:

$$U^l_{ii,t} = ln(w_{i,t}) + A_{i,t} + \epsilon^l_{ii,t} \tag{1}$$

where $w_{i,t}$ denotes the level of wage prevailing in location i in period t.[5] $A_{i,t}$ captures other factors shaping the attractiveness of area i, including humanitarian aid and asylum grants. It also includes the occurrence of particular events such as outbreaks of violence that could affect their perception of the level of attractiveness of the location. $\epsilon_{ii,t}$ is an error term capturing the stochastic part of the alternative-specific utility and following an iid extreme value distribution of type 1.

If individual l chooses to move from the current location i to another location j, the level of utility associated to this choice is given by:

$$U^l_{ij,t} = ln(w_{j,t}) + A_{j,t} - C_{ij} + \epsilon^l_{ij,t} \tag{2}$$

where C_{ij} denotes the level of the (time-invariant) migration costs between areas i and j that individual has to incur if he chooses to move in that corridor. Given that we consider only locations within Turkey and given that all individuals are allowed to move freely (no visa costs), we capture variation in C_{ij} through the geodesic distance between these locations.

Let N^l_{it} denote the size of population of type l residing in location i at time t. Assuming that $\epsilon^l_{ii,t}$ and $\epsilon^l_{ij,t}$ follow an iid extreme value distribution of type 1 allows us to apply the results of McFadden (1984) and to derive the share of individuals from location i choosing to locate in j at time t. This share is given by solving the following

[5]See Anderson (2011) about a discussion on the type of utility function to include in the RUM. In particular, our utility function implies a degree of risk aversion equal to 1 and implies that only relative incomes between locations matter rather than absolute differences.

maximization problem :

$$Prob[U^l_{ij,t} = Max_k(U^l_{ik,t})] = \frac{N^l_{ijt}}{N^l_{it}} = \frac{exp[ln(w_{j,t}) + A_{j,t} - C_{ij}]}{\sum_{k=1}^{K} exp[ln(w_{k,t}) + A_{k,t} - C_{ik}]} \tag{3}$$

A similar expression can be derived for the share of stayers in the population ($\frac{N^l_{iit}}{N^l_{it}}$), assuming that $C_{ii} = 0$. Combining these expressions, we can obtain the ratio of the number of movers from i to j at time t over the number of stayers in location i as :

$$\frac{[N^l_{ijt}/N^l_{it}]}{[N^l_{iit}/N^l_{it}]} = \frac{N^l_{ijt}}{N^l_{iit}} = \frac{exp[ln(w_{j,t}) + A_{j,t} - C_{ij}]}{exp[ln(w_{i,t}) + A_{i,t}]} \tag{4}$$

Taking logs of expression (4), we get an equilibrium expression of the odd ratio between movers and stayers :

$$ln(\frac{N^l_{ijt}}{N^l_{iit}}) = ln(w_{j,t}) - ln(w_{i,t}) + A_{j,t} - A_{i,t} - C_{ij} \tag{5}$$

2.2 RUM-based gravity equation

We can build on the equilibrium condition (5) to derive a gravity equation that we bring to the data to characterize mobility patterns of both types l. Adding an error term to allow for random variation in observed migration flows N^l_{ijt}, we estimate the following specification for the gravity equation:

$$ln(\frac{N^l_{ijt}}{N^l_{iit}}) = \alpha^l_i + \alpha^l_j + \alpha^l_t + \beta^l_1 ln(w_{j,t}) + \beta^l_2 ln(w_{i,t}) + \beta^l_3 C_{ij} + u^l_{ijt} \tag{6}$$

This equation is estimated on two different samples spanning refugees ($l = R$) and non-refugees ($l = NR$), which in turn allows to compare the β^l coefficients between the two populations. The non-refugees include natives and former legal immigrants. Several remarks need nevertheless to be formulated in order to clarify a set of assumptions and limitations that arise when moving from the theoretical condition of equation (5) to the estimable equation (6).

First, in contrast with most studies using micro-founded gravity equations, our dependent variable in equation (6) is fully consistent with the equilibrium condition (5) but implies that we compute N^l_{iit}.[6] N^l_{iit} is computed from $N^l_{it} - \sum_{k=1}^{K \ni i} N^l_{ijt}$. As an alternative to the inclusion of N^l_{iit}, other studies include origin-time fixed effects (α^l_{it} using the current notations) which would further prevent the inclusion of income at origin.[7]

Second, equation (6) includes separate coefficients between origins and destinations in the levels of wages and other factors of attractiveness, while the equilibrium condition of the RUM model implies similar coefficients ($\beta^l_1 = \beta^l_2$). One reason for this is the discrepancy in the access to information between the current and the potential external locations. The standard RUM model assumes that individuals have similar access to information regarding the key factors across different locations. In reality, individuals have much less and more noisy information on external locations compared to their current one. This is especially true for refugees that discover a country new to them. It implies that the role of conditions at origin and destination can be different, calling for different regression coefficients in equation (6).

Third, our RUM model and our gravity equation both ignore some network effect, i. e. the attraction effect exerted by individuals from their community in other locations. Once again, due to their low variation over time, network effects at the aggregate level are difficult to introduce within a short period of time. Nevertheless, to the extent that networks do not grow much over the period, they could be reasonably well accounted for by the destination fixed effects α_j in equation (6). In general, the location fixed effects α_i and α_j partially capture the role of $A_{j,t}$ and $A_{i,t}$ terms in equation (5).

Finally, since equation (6) relies on a double-log functional form, there are two types of issues that arise in the estimation in the presence of a significant share of zeros

[6] This is made possible because we have an exhaustive choice set for refugees.

[7] In a robustness check, we do that and show that we obtain similar results for β^l_1 and β^l_5.

values for N^l_{ijt}. The first issue is the usual selection problem *à* la Heckmann since observations for which $N^l_{ijt} = 0$ would be dropped from the sample in an OLS estimation of equation (6). This would lead to an obvious bias in the estimated coefficients since this regression would drop corridors that are found to be rather unattractive for individuals. The second issue is more subtle and has been identified by Santos Silva and Tenreyro (2006). In the presence of a significant share of zeros, equation (6) is likely to be subject to an heteroscedastic error term, which in turn creates a dependence between higher moments of this term and the key covariates, generating another type of bias in the estimated coefficients. To solve both issues, Santos Silva and Tenreyro (2006) propose to use the Poisson Pseudo-Maximum Likelihood (PPML) estimator.[8] We follow this recommendation and use the PPML estimator for the estimation of equation (6). We also cluster all standard errors at the origin and destination levels.

3 Data and Descriptive Statistics

We first describe our data in Section 3.1. Second, we provide some descriptive statistics that will help visualize and therefore better explore our sample in Section 3.2 .

3.1 Data

The D4R Challenge and Constraints. The source of our data is the D4R, a non-profit challenge with the aim of improving the living conditions of Syrian refugees currently residing in Turkey. Türk Telekom (TT)[9], in collaboration with the Scientific and Technological Research Council of Turkey and Boğaziçi University, along with other academic and non–governmental organizations, organized an anonymized dataset of

[8]Regarding the first issue, PPML estimator involves an exponential model, which automatically includes the zero values in N^l_{ijt}.

[9]Formerly state-owned, TT is the first integrated telecommunications operator in Turkey. Vodafone and Turkcell are the two other operators in Turkey. As of the fourth quarter of 2016, TT, Turkcell and Vodafone have respectively a subscriber market share of 30%, 45% and 25% (Türk Telekom, 2019).

mobile call detail records (CDRs) of phone-calls and SMS messages of TT customers. The data collected and provided by the company cover the time period between 1 January 2017 to 31 December 2017.

The D4R dataset is collected from a sample of 992,457 TT customers. 184,949 are identified as refugees and 807,508 as Turkish citizens. While not much is revealed concerning the individual characteristics of the customers, we know that approximately 25% of the refugee customers are identified as "female" and the remaining 75%, as "male". Overall it provides three distinct datasets.[10] We employ Dataset 3 (Coarse Grained Mobility) in our analysis in order to construct the refugee mobility measures aggregated at the regional level. Dataset 3 is a randomly selected dataset of 50.000 refugees and 50.000 non-refugees that is being followed throughout the whole year. To ensure privacy the data is provided with reduced spatial resolution, i. e. at the district (rather than the antenna which was the case in the other datasets) level. Each individual is associated with an ID number. The ID of the traced individual starts with a number: 1 if the individual is a refugee, 2 if the individual is not a refugee and 3 if this information is unknown. We thus have detailed information on whether the call belongs to a particular person, being a refugee or not, as well as the day and time of the call.[11]

The D4R challenge is a unique initiative that allows to study various aspects associated with refugee mobility. However, addressing such a sensitive issue is a major challenge and maximum protection of personal data is a prerequisite. To this end the data comes with several restrictions and shortcomings. The main restriction is associated with the fact that the refugee ID is not entirely clear. Analytically, the data provider highlights that the term "refugee" is entailing to asylum seekers, migrants, and any individual that may have a "temporarily protected foreign individual" ID number in Turkey. To give

[10]Dataset 1 (Antenna Traffic) includes one year site-to-site traffic on an hourly basis and it provides information about the traffic between each site for a year period. Dataset 2 (Fine Grained Mobility) randomly chooses a group of active users (who make calls and send SMS) every two-week period and reports cell tower identifiers. Datasets 1 and 2 are described in more detail in the appendix.

[11]See Appendix A.1 for more information on the data provided by the D4R.

the refugee status to a number, three conditions should be satisfied: i) the customers in the database have ID numbers that are given to foreigners and refugees in Turkey; ii) the customer should be registered with Syrian passports; and iii) use special tariffs reserved for refugees (Salah et al., 2018). Moreover, on 3 September 2020, Turkey's Directorate General of Migration Management (2020) state that Turkey hosts 3,612,694 Syrian refugees with temporary protected status plus 93,299 Syrians with legal residency permit and 110,000 with granted citizenship. Consequently, it is highly unlikely that we capture patterns of mobility that are not associated with other groups than Syrian refugees.[12]

Another constraint is that we cannot be entirely sure if the actual caller is indeed a refugee or a non-refugee. While individuals register with their refugee cards in order to connect, there is no guarantee on who is eventually using the phone. It is however more likely that refugees may use non-refugee phones rather than vice versa (refugee contracts have more limitations with respect to the number of calls they can do). Last, we cannot exclude noise in the exact location of the call. In some cases, the antenna location may not be precise as a line might connect to a different antenna due to the capacity of the network. Last, missing data is another concern as in some cases whole days of data may not be reported in the dataset (Salah et al., 2018).

For all the above reasons and given the scope of our research, we use the Nomenclature of Territorial Units for Statistics (NUTS) by Eurostat. Similar to a large literature in regional studies applied to the EU regions (Combes and Overman, 2003; Crozet, 2004; Hirschle and Kleiner, 2014; Fischer and Pfaffermayr, 2018), we conduct our analysis at the NUTS–2 administrative level, i. e. for 26 regions in Turkey.[13] More recently, Mitze

[12] An example would be people with Syrian passports falsely claiming the refugee status to benefit from reduced tariffs or Syrians who initially got the refugee status, later they obtained legal residency and retained their past connections. In both cases the analogy would be much less than 200,000 Syrians with legal residency permit or granted citizenship versus more than 3.60 million Syrian refugees. We thus view quite unlikely that these numbers can systematically influence our mobility patterns at any level, especially at the NUTS–2 level.

[13] Information on the NUTS statistical regions of Turkey is provided in Table B.1 and Figure B.1 in

(2019) finds a stronger explanatory power of local labor markets conditions during the global financial crisis at the NUTS–2 level, compared to NUTS–3 level. We also aggregate the data over time, i. e. we construct quarterly measures of mobility from location measures drawn from the phone call data. This approach mitigates most of the above mentioned concerns. First, it allows us to combine information and construct measures from D4R datasets. Second, it summarizes flows to other NUTS regions irrespectively of who is using the phone (it could thus capture movement of the whole family). Also, aggregating mobility at the NUTS-2 level, would nevertheless reveal systematic patterns of refugee mobility even if less than 100% is composed of Syrian refugees. And last, it resolves imprecise location concerns since the data at NUTS–2 level is very accurate. It also mitigates the concern from the absence of reporting data on a daily basis. This level of aggregation is also in line with our research question. Since we want to capture "internal migration" flows, the desirable property of our geographical unit of analysis is that it is not too small, in which case it could potentially capture commuting flows or regular social exchange patterns. As such we chose NUTS–2 which balances the trade-off between a sufficiently large unit of analysis and a large number of observations. Last, the NUTS–2 level allows us to combine our constructed measure of mobility with high quality administrative data available at quarterly level as well.[14]

A last concern about the D4R data is that while the sample of refugees is representative, it may not be the case for non-refugees who are sampled based on the sample of refugee customers.[15] While it is unclear whether this sampling process generates any systematic bias in the data, we focus our analysis on the refugee population to understand the determinants of their internal migration. While we benchmark the analysis using the sample of non-refugees, we undertake this exercise only for comparison reasons. This

the Appendix.

[14]Our results are robust to defining mobility at NUTS–1 (larger) and NUTS–3 (smaller) regional levels. These results are reported in Section 4.2.

[15]As mentioned in Salah et al. (2018), Turkish citizen customers have been mainly sampled "from the cities with registered refugee presence, to simplify comparisons" (p.4).

is also in line with our research contribution, i. e. the gravity model for the refugee population, since the gravity model for natives has been more analyzed in the relevant literature.

Dependent Variable: Refugee Mobility. Our main variable is the measure of mobility, which we construct using Dataset 3, i. e. the dataset that follows 50,000 refugees and 50,000 non-refugees throughout the whole year. We construct migration rates at the NUTS–2 level and at a quarterly frequency.

The migration rate has the form *Migration Rate_'l'_'i'* where *'l'* refers to the refugee (i. e. *'l'=R*) or non-refugee (i. e. *'l'=NR*) status of the observation, and *'i'* corresponds to the minimum number of calls generated from a given province to characterize the latter as the residence location (i. e. frequency filter of *'i'* calls, in our case, we set *'i'*=10). If there are several calls from different places within a given quarter, we choose the place from which the majority of calls comes from. To increase the likelihood that our measure properly reflects location of residence and not workplace, we restrict our analysis to calls that take place only between 8 pm to 8 am, i. e. hours that are less likely to be working hours, following the usual approach in this literature using phone data (Blumenstock et al., 2016).[16]

To define mobility, we compare the residency (as defined above) between sequential quarters. If residence is the same NUTS–2 area, the caller is a stayer, if not, (s)he is a mover. Subsequently, we compute mobility between quarters based on a migration rate $ln(\frac{N^l_{ijt}}{N^l_{iit}})$ where N^l_{ijt} corresponds to leavers and N^l_{iit} to stayers. As mobility is observed at the quarter frequency, it represents movers between quarters $t-1$ and t. By construction, any explanatory variable in our analysis is therefore measured prior to the mobility at quarter t.

Under Section 4.2, we test the robustness of our analysis using a stricter mobility measure, i. e. with a frequency filter of 20 calls, and using a more flexible mobility

[16]For the sake of following the terminology of the standard gravity model, we frequently refer to the residence place of the refugee as the origin.

measure, i. e. with a frequency filter of 5 calls. We undertake another robustness check, i. e., we use a second filter where we impose the supplementary condition that a minimum number of calls (5, 10, 20) has taken place at least during a number of different days (5, 10, 20) during the quarter.

Standard Gravity Model Determinants. We employ two main sets of determinants of mobility. First, we use the standard gravity model controls, i. e. variables that relate to the attractiveness (resp. repulsiveness) of region j (resp. i) for prospective refugees, the so-called pull (resp. push) factors.

At the NUTS–2 level we have systematic regional GDP data also available at the quarterly level. We obtain data on regional GDP from the Turkish Statistical Institute (TurkStat, 2020a).

Proximity between pairs of NUTS–2 regions is measured using geodesic distances, i. e. the length of the shortest curve between two points along the surface of a mathematical model of the earth, based on the centroid coordinates.[17] Distance here captures practical difficulties of moving across these regions.

3.2 Descriptive Statistics

Our sample is composed of 1,950 bilateral observations for which we have information about all variables in our baseline specification (Table 1).[18] According to our mobility measure, bilateral movements of refugees between NUTS–2 regions are limited, and amount to 0.6%, i. e. on average, 6 refugees per thousands moves to another NUTS–2 region from one quarter to another.

Figure 1 shows the presence of refugees in Turkey in 2017. More precisely, the Directorate General of Migration Management in Turkey provides yearly data on the

[17] The centroid coordinates are based on the WGS 1984 datum and we rely on Vincenty (1975) equations to calculate distances.

[18] The number of observations results from pairing each NUTS–2 region with another NUTS–2 region, given the bilateral nature of mobility. We do so for every quarter of the year 2017. Mobility is constructed from one quarter to another, resulting in 1,950 bilateral observations ($26 * 25 * 3$). Table B.2 in the Appendix provides a detailed description for all the variables we use in our study.

presence of refugees at the regional level. We divide these figures by the measure of the regional distribution of refugees that we have obtained from our mobile phone data. The map indicates that administrative data overestimate the presence of refugees in Southeast Anatolia – at the Syrian border – and underestimate their presence in the northwestern (Istanbul, East Marmara and West Anatolia), Aegean and Mediterranean regions. Figure 2 shows the mobility of refugees in Turkey in 2017 as obtained from the D4R data. The first map corresponds to out-migration of refugees, their origin, while the second map shows in-migration of refugees, their destination. As can be seen from Figure 2, refugees tend to leave regions in the eastern part of Turkey for regions in the northwestern, Central Anatolia, Mediterranean and Istanbul.

For the levels of income at origin and destination, we use data on quarterly GDP per capita from TUIK (Turkish Statistical Institute). Numbers are reported in Turkish Lira. As can be seen from Table 1, based on our study sample, the lowest income corresponds approximately to 328 TRY and this is the income in *Şanlıurfa* region (in first quarter). The highest income is approximately 1,734 TRY in *İstanbul* region (in third quarter). The mean income is approximately 856 TRY and this is equivalent to the income in *Konya* region.

The shortest distance is approximately 96 kilometers and this is the distance between the *Gaziantep* and *Hatay* regions while the longest distance is approximately 1,400 kilometers, between *Van* and *Tekirdağ* regions. The mean distance is approximately 580 kilometers and this is equivalent to the distance between *İstanbul* and *Samsun* regions.

Table 2 offers a comparison between the mobility of refugees and non-refugees in our sample based on the frequency of their moves and the distance they travel. Interestingly, according to our mobility measure, non-refugees move more often and further than refugees. On average, refugees travel 582 kilometers while non-refugees travel 733 kilometers. No refugee in our sample covers a cumulated distance larger than 2,500 kilometers while cumulated distance over the total year 2017 exceeds 3,000 kilometers for some non-refugees. However, the distance of refugee and non-refugee mobility is similar for distances between 0-1000 kilometers and non-refugees move more and further for distances exceeding 1,000 kilometers. Overall, this analysis of distance for refugees and non-refugees is in line with some previous evidence (World Bank, 2018).[19]

4 Results

We first present our benchmark results for the determinants of mobility for refugees and compare those to the ones for non-refugees (Section 4.1). Then we show that our results are robust to the geographical level of analysis and to the filtering procedure of phone data that we adopted to construct our measures of mobility (Section 4.2). We then assess whether our estimates reflect a story of mobility of refugees driven by the attractiveness of incomes at origin or destination or by a set of alternative patterns.

4.1 Benchmark Results

Our benchmark results are based on the estimation of equation (6). In the benchmark results, we measure bilateral movements using a 10-call filtering procedure to capture the respective locations at origin and destination, use calls taking place between 8.00 pm and 8.00 am and define regions based on the NUTS–2 geographical level. Table 3 reports these benchmark results. Column (1) provides our benchmark estimation for the refugees, with a focus on income levels at origin and at destination. These results suggest that refugees tend to follow patterns of mobility emphasized in the empirical literature on internal and international migrants. Like other types of migrants, refugees in Turkey tend to leave poorer locations and to be attracted by richer ones. Income differential across regions seems therefore to matter a lot for their location choice. Like in most gravity estimate of mobility, the elasticity of distance is negative and comprised between

[19]An histogram of the distance covered by refugees and non-refugees is shown in Figure B.2 in the Appendix.

0 and -1. A limitation of the estimation in column (1) is that it does not account for origin-time specific shocks beyond income at origin. The same holds for destination-time specific shocks. To overcome this limitation, columns (2) and (3) include estimation results obtained respectively with origin-time and destination-time fixed effects.[20] Results of column (1) are found to be similar with these ones.

An important dimension of the analysis is the comparison of the mobility patterns of refugees with other categories of individuals. To that aim, column (4) provides the estimates based on the same benchmark specification, but for non-refugees. It should be emphasized that non-refugees represent an heterogeneous group of individuals, composed by both natives and traditional immigrants of Turkey. Having said that, settled immigrants are expected to behave closer to natives than refugees. The comparison between columns (1) and (4) shows that in contrast with refugees, non-refugees do not react at variations of income at origin, reflecting possibly a stronger attachment to their current location. They also react slightly more to changes in income at destination, reflecting maybe better information about outside economic opportunities. Estimations show that the elasticity of distance is about one-third higher for refugees than for non-refugees, in line with the evidence provided by the World Bank (2018). Nevertheless, estimates provide evidence that refugees in Turkey respond to economic opportunities in their location choice and are not excessively constrained by factors hampering their mobility.

4.2 Robustness checks to methodological choices

Since our measures of mobility are inferred and not directly observed, it is desirable to conduct several robustness checks to assess the sensitivity of our results to alternative methodological choices. We consider two main choices, i. e. the level of geographic

[20]On top of accounting for unobserved time varying shocks, these estimations account for issues of multilateral resistance of migration, respectively at origin and destination. See Anderson (2011) and Bertoli and Fernández-Huertas Moraga (2013) on the issue of multilateral resistance.

aggregation to define locations and the filtering procedures of phone data to infer location choices and mobility.

4.2.1 Geographical level of aggregation

First, we look at the sensitivity to alternative choices with respect to the geographical level at which data are aggregated. The choice of the NUTS–2 level to define the locations underlying our measures of mobility can be seen as the result of a trade-off between small and large areas. On the one hand, the choice of excessively large areas would conceal important movements of interest within each area. On the other hand, the use of very precise locations can lead to a confounding effect of commuting or simple visits to close locations. The RUM model and the derived gravity equation should describe mobility choices based on differences in economic attractiveness across locations. Therefore, it is desirable to get rid of mobility patterns based on other motivations (such as shopping). While not preventing totally some movements related to commuting, our choice of locations at the NUTS–2 level should mitigate the concerns compared to the choice of NUTS–3 areas.

To assess the robustness of our results to the geographical definition of a location, column (1) of Table 4 reproduces the estimation from column (1) of Table 3 for the sake of comparison. Columns (2) and (3) show respectively the results from conducting the analysis at regional NUTS–1 and NUTS–3 levels respectively. There are 12 NUTS–1 regions in Turkey while NUTS–3 regions correspond to the 81 Turkish provinces.

As can be seen from Table 4, results are in general robust to performing the analysis at different geographic aggregations. We interpret this result as a support for the absence of strong spatial dependence in our estimations.[21] Columns (2) and (3) respectively indicate that a 10% increase in the GDP decreases the likelihood to migrate of refugees by roughly 5% and 6% while at destination a 10% increase in the GDP increases the

[21]Geographers refer to this issue as the Modifiable Areal Unit Problem (MAUP), which results from relying on arbitrarily chosen areas to represent information and results in statistically biased effects.

likelihood to migrate by roughly 11% and 35%. At a NUTS–1 level, a 10% increase in the distance to be covered decreases the migration likelihood by roughly 5%, and 9% at a NUTS–3 level.

4.2.2 Alternative filters of phone calls

Since our location and mobility measures are inferred from phone calls, it is important to assess the robustness of the results to the way these calls are filtered. First, there is a trade-off between taking a too low minimum call threshold to assess location and a too strict minimum level. If a too low minimum threshold is used, this can lead to noisy measures of locations polluted by individuals giving calls from locations they just visit temporarily. On the other hand, if a too high threshold is imposed, this can lead to the dismissal of many valid observations of location since some individuals do not call that much. The minimum 10-call threshold that we use in the benchmark analysis can be seen as a value taking this trade-off into account.

Table 5 contains results considering different minimum frequency filters to compute the mobility. Column (1) of Table 5 is once again our benchmark. In column (2), refugee mobility is computed such that at least 20 calls in a given NUTS–2 region are used to define the latter as the residency of an individual. In column (3), refugee mobility is computed such that at least 5 calls in a given NUTS–2 region are used to define the latter as the residency of an individual, i. e. any individual characterized by less than 5 calls is dropped from our sample. As can be seen from Table 5, results remain robust to having a more flexible approach as under column (3) or a more restrictive approach as under column (2).

A second filter can also be considered to avoid capturing temporary movements of refugees rather than more permanent moves. If one individual gives, say, more than 10 calls from a location but during one single day, this can indicate a temporary move to that location but not a permanent settlement. A second condition regarding the minimum number of days these calls are spread can be imposed to infer the location of individuals. Once again, if this minimum number becomes too high, this can lead to the loss of many valid observations of locations. To assess the sensitivity of our results to this choice, columns (4) to (6) of Table 5 report the results based on mobility measures obtained with this double filtering procedure, with the three values of the frequency filter. Results are in line with those reported in columns (1) to (3) based on a single filter.[22]

4.3 Alternative explanations to economic attractiveness of locations

Our benchmark results suggest that refugees are likely to respond to the economic attractiveness of locations in their mobility choice. They tend to leave current locations that are less attractive in terms of income and to head to places with higher expected income. While they are more sensitive to factors of friction in their mobility compared to non-refugees, the estimations suggest that this sensitivity does not prevent them to grab attractive economic opportunities. In short, refugees in Turkey behave very much like other categories of workers and their movements from low to high productivity locations can be seen as contributing to a more efficient process of allocation of labor across space. Nevertheless, it is important to check whether alternative mechanisms driving the mobility patterns could be also consistent with our data. In this section we explore various alternative stories.

4.3.1 Westward movements

Since refugees are initially settled in the southeastern part of the country, one may be concerned that the estimated attractive nature of the income per capita at destination is entirely driven by other factors (e. g. proximity to Europe, ...) associated with a

[22]To mitigate concerns that some individuals in the raw data are not present throughout the whole period, we did an additional robustness check where we excluded those individuals who did not make at least 10 calls in all four quarters. This yields a "balanced panel" of individuals. Our results, though not reported in the text, remain qualitatively the same despite the large decrease in the number of individuals entering the construction of the mobility measure.

Western move by refugees. Such a move would simply capture a positive income gradient along the East-West axis. While this is a legitimate concern, we believe that this threat should not be overstated. First, we cover the year 2017, i. e. after the EU and Turkey decided to close their common border to refugees. Therefore, motivations of movements based on pure transit to Europe are lower than before. Second, Figure 2 indicates that beyond the Westwards general pattern, there is more variation than one would expect in terms of in- and out-migration among refugees. For instance, the Southern regions of Adana (TR62), Konya (TR52), and Şanlıurfa (TRC2) feature relatively high in-migration rates. In contrast, other Western regions such as e. g. Tekirdağ (TR21), Balıkesir (TR22), Aydın (TR32), and Manisa (TR33) have high rates of out-migration. Nevertheless, in order to assess the importance of this phenomenon in explaining internal mobility of refugees, we generate a dummy variable indicating whether the move is a westward movement. To do so, we compute the centroid (i. e. longitude and latitude) of every NUTS unit and for every dyad, we then compare longitudes to determine whether it implies a westward movement. Columns (1) and (2) of Table 6 show the results of the benchmark specification augmented with an interaction term between income at origin or income at destination and this dummy variable. While these interaction terms turn out to be statistically significant, their magnitude is very low in absolute terms, suggesting that the income elasticities are similar between refugees going westward and those going in other directions.

4.3.2 Dyadic fixed effects and network movements

Another way of investigating the specificity of westward movements is through the inclusion of dyadic fixed effects in equation (6). The dyadic fixed effects will indeed capture the specific mobility related to westward movements. Nevertheless, this specification is really demanding given the structure of our data. The identification of income elasticities will rely only on the time variation within a corridor. With 3 observations per dyad, this leads to estimation issues. These issues are reflected by the fact that the Poisson maximum likelihood estimation drops a substantial proportion of observations.[23] Furthermore, a drawback of this specification is that the role of distance is no longer estimated while this is one of the key interests of our analysis. Results of the specification with dyadic fixed effects are reported in column (3) of Table 6. The previous results regarding income elasticities are qualitatively similar with this specification. Once again, refugees seem to be attracted by high income regions and tend to leave poorer ones.[24]

A by-product of the inclusion of dyadic effects is that it provides a way to overcome the absence of a network variable in the benchmark specification. Networks at destination could be a confounding factor of income at destination since networks are likely to be located in attractive places. In other terms, it could be that refugees tend to settle in wealthy locations not only because of the expected high income but also because of the presence of a large network in these locations that will help them to settle and integrate. In contrast with Blumenstock et al. (2019) and in absence of information about the dyadic nature of specific calls, our data do not allow us to recover the topology of the network, preventing us to estimate the role of strong ties between individuals (Giulietti et al., 2018). Given the limited representativeness of our data for refugees, it is also difficult to have a good measure of the aggregate network, i. e. to estimate the role of weak ties in the network.[25] Nevertheless, since our investigation period is short (only one year) and that networks of refugees are already substantial at the end of the year 2016, one can expect that networks will not vary too much over time, even in locations receiving net inflows of refugees. If this assumption is correct, the inclusion of dyadic fixed effects

[23]On the issue of dropping some observations, see Correia et al. (2020) for a discussion about the so-called separation problem in Poisson models with a rich set of fixed effects. Another issue is that standard errors are no longer double-clustered at origin and at destination.

[24]The estimation of this specification for non-refugees (not reported here for the sake of brevity) also gives similar results than the benchmark ones. Furthermore, the comparison between refugees and non-refugees gives rise to the same conclusions regarding their relative sensitivity to income at origin and at destination.

[25]In a previous version of this paper, we used the total calls from refugees in a given location. Nevertheless, this measure was very noisy and it is unknown to what extend it correlates with the size of the network at destination.

can account for the impact of networks. The results show that the high income elasticity that we obtain in the benchmark specification is not driven by the absence of a role for network. If any, and with all the reservations regarding the different samples and the non causal interpretation of our estimates, the elasticity of income at destination is found higher in column (3) of Table 6 compared to the benchmark specification.

4.3.3 Accounting for contiguity

A limitation of the benchmark specification is that frictions in the mobility patterns are only captured through distance. Refugees have been reported to travel on short distances (World Bank, 2018). A question is whether they move mainly to areas that are very close and to what extend distance plays a significant role. In other words, it could be that the negative elasticity of distance reflects only that refugees move to contiguous provinces to the extend these exhibit higher income. To address this question, we supplement our benchmark specification with a contiguity dummy (taking 1 if the origin and destination share a common border, 0 otherwise). Columns (4) and (5) of Table 6 report the results of this extended specification, respectively for refugees and non-refugees. The results confirm that refugees and non-refugees tend to move more to contiguous locations, emphasizing the role of frictions in their mobility. Nevertheless, this specification confirms that refugees are still sensitive to distance and overall more sensitive to distance than non-refugees.

4.3.4 Accounting for agricultural business cycles

The evidence of high in-migration rates in locations such as Adana (TR52) and Şanlıurfa (TRC2) presented in Figure 2 raises the attention to the role of demand in agriculture. These regions are indeed known to generate a high demand for seasonal workers in Turkey, including Syrian refugees. The specificity of this labor demand could confound the estimations of the income elasticities at destination to the extend that these regions benefit from an increase in income during the harvesting season.[26] To tackle this issue, we augment our benchmark specification with an interaction term capturing the increase in income during the harvesting period (third quarter of 2017) in an agricultural province. The definition of an agricultural location relies on the share of agriculture in its provincial GDP (TurkStat, 2020b). We use two thresholds: 30% and 50% of agricultural output in the provincial GDP. Based on a threshold of 30%, we obtain 20 agricultural regions. Based on a threshold of 50%, 14 regions out of the 26 regions are classified as agricultural locations. Columns (6) and (7) of Table 6 provide the results with each threshold. The results obtained using this extended specification show that our estimated income elasticities are robust to the inclusion of agricultural business cycles generating a specific labor demand for refugees.

4.3.5 Refugee camps

Another version of the West-East different gradient would be that refugees being resettled primarily in refugee camps on the Eastern part of the country would be more or less keen to leave these regions, whether the areas are poor or rich. On the one hand, refugee camps provide some basic infrastructure and host large communities of refugees on which these individuals can rely on. On the other hand, many refugees clearly prefer to live outside these camps as their facilities do not match their expectations. These considerations suggest that refugees living initially in these camps can exhibit different sensitivities to income at origin. Furthermore, refugee camps tend to be mostly located in poor regions although there are a few exceptions.[27] The inclusion of origin fixed effects account for this last aspect. It might nonetheless be desirable to investigate the possible heterogeneity of the income elasticities with respect to the camps.

[26] It could also affect the estimation of the elasticity of income at origin for refugees already settled in these agricultural provinces.

[27] In 2017, there are 25 refugee camps in Turkey and these are located in 6 NUTS–2 regions: Adana, Gaziantep, Hatay, Malatya, Mardin and Şanlıurfa. Among them, Adana, Gaziantep and Hatay are among the middle-income regions on Turkey while Malatya, Mardin and Şanlıurfa are poorer regions. See Figure B.1 in the Appendix.

To that aim, we carry-out regressions getting rid of the phone data generated in the NUTS–3 region containing a camp. This means that we still consider refugees in the NUTS–2 regions hosting a camp but only those living in areas without camps. Column (8) of Table 6 reports these estimates. The estimated income and distance elasticities are in line with the benchmark specification. Income elasticities are found to be slightly higher in absolute terms, although the magnitude of the difference is not substantial. The same holds for the sensitivity to distance. All in all, the results point to a modest heterogeneity in the behavior of refugees located in areas with and without camps. Further investigation based on individual data would be welcomed to better understand this heterogeneity.

4.3.6 Social Magnet

Another pattern of mobility is related to the provision of aid given to refugees. The literature on the social magnet effect has investigated to what extent migrants develop opportunistic location strategies with respect to the level of public transfers (Razin and Wahba, 2015). A similar type of motivation could be theoretically expected for refugees regarding their choice of location within the country of settlement. Aid provided upon arrival is a crucial element to overcome the distress refugees often face when escaping urgent and dangerous contexts. Nevertheless, if refugees also tend to be attracted by more generous locations, the provision of aid might undermine the efficiency of the process of labor reallocation across space. It could also be the case that more attractive locations in terms of expected level of wage are also more (or less) generous in the level of transfers provided to refugees. If this is the case, this can affect the quality of estimation for the income elasticities.

To deal with the aspect related to the social magnet channel, we extend the benchmark specification with two types of aid targeted to refugees, both at origin and at destination. Based on the Global Database of Events, Language and Tone (GDELT) dataset, we create variables for news that are relevant to the refugee population. In particular we have chosen the following categories: humanitarian aid and asylum grants. We aggregate these 2 types of events quarterly and at the NUTS–2 level.[28] Events related to humanitarian aid are therefore related to the literature showing that welfare benefits may attract or retain potential migrants (Razin and Wahba, 2015). The news for asylum grants are directly linked with policy considerations that have a direct impact on the decisions of refugees and their ability to integrate and to move freely around the country. Events related to the granting of asylum status can also be directly interpreted as a possible change in expectations (Cortes, 2014) and therefore, local integration at origin or destination. We should acknowledge that the interpretation given to these hypothesized drivers are subject to discussion and that a lack of evidence may also be due to measurement errors. Nonetheless, the extension of the gravity model to political factors allows us to compare our results to a recent and growing literature on international migration.

We look at the effect of humanitarian aid (column (9) of Table 6) and the effect of asylum grants (column (10) of Table 6). We find that refugees are moderately sensitive to humanitarian aid and asylum grants. An increase of the provision of these services tends to decrease their probability of moving out of their current location. In contrast, we do not find any evidence of a social magnet effect through which refugees would favor locations providing higher levels of these transfers. These conclusions should nevertheless be drawn with much caution due to the non causal estimation of the effect of such transfers. In particular, the level of transfers at origin and destination could definitely be dependent on the level of (unobserved) attractiveness of these locations in equation (6). Further econometric investigation is therefore needed before drawing more clear-cut conclusions about the (internal) social magnet effect of aid to refugees. Nevertheless, the income elasticities at origin and destination when accounting for the aid to refugees are found to be rather similar to those estimated in the benchmark specification.

[28]A detailed description of GDELT and the variables construction are provided in the Appendix A.2.

4.3.7 Reaction to news about protest

Finally, one could hypothesize that refugees will also move in reaction to specific events such as protests and demonstrations against immigrants. Based on GDELT, we create a variable for news related to violence and protests. In particular, we have chosen the category: violent protests. We aggregate this type of events quarterly and at the NUTS–2 level.

As the occurrence of such events is not randomly distributed across potential locations, this type of events could also confound the estimation of income elasticities. We capture such events at origin and destination and supplement equation (6) with these variables. Results (column (11) of Table 6) do not support a role of protests and demonstrations in our framework. Once again, caution is needed about the conclusions to be drawn due to the non-causal nature of these estimates. Income elasticities remain basically unaffected by the inclusion of these variables.

5 Policy Implications and Conclusion

In this paper, we look at the determinants of the internal mobility of refugees after their early settlement in Turkey. Turkey is home of more than 3 million refugees that are allowed to move and work freely. It therefore provides an ideal context to address this topic. A good understanding of the patterns of refugees' mobility is key for an optimal provision of aid and support by the hosting authorities. Evaluating to what extent location choices of refugees respond to the factors of economic attractiveness is also important to know whether their movements contribute to an efficient allocation of labor across space.

The existing empirical literature on the mobility of refugees is scarce, especially due to the absence of reliable data that track the movements of this category of immigrants. Due to the elusive status and the unstable situation characterizing refugees, one cannot rely on traditional measures of mobility based on administrative data. This calls for alternative ways of measuring movements of refugees. In this paper, we use a unique dataset of mobile phone data to measure internal movements of refugees across Turkish regions over the year 2017. An additional appealing feature is that we can compute similar measures for non refugees, which allows to make useful comparisons between the two categories.

This big data approach allows us to conduct a traditional gravity approach applied to migration and to identify the main determinants of their movements as well as to compare these to the non refugee population. Although we do not exploit exogenous changes in those determinants, the risk of confounding factors is minimized through the use of a large set of combined fixed effects and the short nature of our period of investigation. We find that refugees are highly sensitive to distance, in line with the literature on economic migration showing that this sensitivity is increasing in the skill level of immigrants. Refugees also tend to move more often, but on shorter distances.

We also find that refugees respond to income differences between regions. They are likely to leave poor areas and are attracted by high-income regions. This contrasts with the patterns of non refugees who do not display any sensitivity to income at origin. Our conclusion regarding the sensitivity to income differential is robust to possible alternative explanations of mobility, including the propensity to cross Turkey from East to West, the propensity to leave refugee camps and the attraction to agricultural areas during the harvest season. Finally, we find that refugees are sensitive to humanitarian aid and asylum grants. An increase in the provision of these services tends to decrease their probability of moving out of their current location. Nevertheless, we do not find any evidence of a social magnet effect through which refugees would favor locations providing higher levels of these services.

Further investigation on the nature of mobility of refugees would be welcome. First, the non-experimental nature of our estimates calls for further studies based on a causal identification strategy to confirm our main conclusions. Second, while Turkey is one of the major location of refugees, similar analyses relying on other sources of data would be welcome to evaluate the degree of external validity of our findings. Finally, the access to individual data on calls allowing to recover the topology of the personal network would be valuable to capture more consistently the role of networks in explaining location choices of refugees.

References

Akgunduz, Y. E., van den Berg, M., and Hassink, W. (2018). The Impact of the Syrian Refugee Crisis on Firm Entry and Performance in Turkey. *World Bank Economic Review*, 32(1):19–40.

Altindag, O., Bakis, O., and Rozo, S. (2020). Blessing or burden? Impacts of refugees on businesses and the informal economy. *Journal of Development Economics*, forthcoming.

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, 3(1):133–160.

Bahçekapili, C. and Çetin, B. (2015). The Impacts of Forced Migration on Regional Economies: The Case of Syrian Refugees in Turkey. *International Business Research*, 8.

Balkan, B. and Tumen, S. (2016). Immigration and prices: Quasi-experimental evidence from syrian refugees in turkey. *Journal of Population Economics*, 29(3):657–686.

Becker, S. O. and Ferrara, A. (2019). Consequences of forced migration: A survey of recent findings. *Labour Economics*, 59:1 – 16.

Beine, M., Bertinelli, L., Cömertpay, R., Litina, A., Maystadt, J.-F., and Zou, B. (2019). Chapter 22. Refugee Mobility: Evidence from Phone Data in Turkey. In Salah, A.A., Pentland, A., Lepri, B., Letouze, E. (Eds.),. *Guide to Mobile Data Analytics in Refugee Scenarios*, pages 433–449.

Beine, M., Bertoli, S., and Moraga, J. F.-H. (2016). A Practitioners' Guide to Gravity Models of International Migration. *The World Economy*, 39(4):496–512.

Beine, M. and Coulombe, S. (2018). Immigration and Internal Mobility in Canada. *Journal of Population Economics*, 31(1):69–106.

Beine, M., Docquier, F., and Ozden, C. (2011). Diasporas. *Journal of Development Economics*, 95:30–41.

Bertoli, S. and Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102(C):79–100.

Blumenstock, J., Chi, G., and Tan, X. (2019). Migration and the Value of Social Networks. CEPR Discussion Paper DP 13611, CEPR.

Blumenstock, J., Eagle, N., and Fafchamps, M. (2016). Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters. *Journal of Development Economics*, 120:157–181.

Bose, P. S. (2013). Building Sustainable Communities: Immigrants and Mobility in Vermont. *Research in Transportation Business & Management*, 7:81–90.

Bose, P. S. (2014). Refugees in Vermont: Mobility and Acculturation in a New Immigrant Destination. *Journal of Transport Geography*, 36:151–159.

Ceritoglu, E., Yunculer, H. B. G., Torun, H., and Tumen, S. (2017). The Impact of Syrian Refugees on Natives' Labor Market Outcomes in Turkey : Evidence from a Quasi-Experimental Design. *IZA Journal of Labor Policy*, 6(5):1–28.

Chiquiar, D. and Hanson, G. (2005). International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113:239–281.

Combes, P.-P. and Overman, H. (2003). The Spatial Distribution of Economic Activity in the European Union. *Handbook of Regional and Urban Economics*, 4:2845–2909.

Correia, S., Guimaraes, P., and Zylkin, T. (2020). PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed effects. *Stata Journal*, 20(1):95–115.

Cortes, K. E. (2014). Are Refugees Different from Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the United States. *The Review of Economics and Statistics*, 86(2):465–480.

Crozet, M. (2004). Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model. *Journal of Economic geography*, 4(1):439–458.

Del Carpio, X. and Wagner, M. (2015). The Impact of Syrian Refugees on The Turkish Labor Market. Policy Research Working Paper 7402, World Bank.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel,

V. D., and Tatem, A. J. (2014). Dynamic Population Mapping Using Mobile Phone Data. *Proceedings of the National Academy of Sciences*, 111(45):15888–15893.

Directorate General of Migration Management (2020). Temporary Protection.

Fischer, L. B. and Pfaffermayr, M. (2018). The More the Merrier? Migration and Convergence among European Regions. *Regional Science and Urban Economics*, 72(C):103–114.

Garcia, A. J., Pindoliay, D. K., Lopiano, K. K., and Tatemzz, A. J. (2015). Modeling Internal Migration Flows in Sub-Saharan Africa Using Census Microdata. *Migration Studies*, 3(1):89–110.

GDELT (2014). Event Record Exporter. Retrieved from http://analysis.gdeltproject.org/module-event-exporter.html.

GDELT (2019). GDELT Project. Retrieved from https://www.gdeltproject.org/.

Giulietti, C., Wahba, J., and Zenou, Y. (2018). Strong versus weak ties in migration. *European Economic Review*, 104(1):111–137.

Godøy, A. (2017). Local Labor Markets and Earnings of Refugee Immigrants. *Empirical Economics*, 52(1):31–58.

Grönqvist, H., Johansson, P., and Niknami, S. (2012). Income Inequality and Health: Lessons from a Refugee Residential Assignment Program. Working Paper Series 2012:11, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Grogger, J. and Hanson, G. H. (2011). Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics*, 95(1):42–57.

Hatton, T. (2009). The Rise and Fall of Asylum: What Happened and Why? *Economic Journal*, 119(535):183–213.

Hatton, T. and Trani, M. (2005). Immigration and Inter-Regional Mobility in the UK, 1982-2000. *The Economic Journal*, 115(507):342–358.

Hatton, T. J. (2016). Refugees, Asylum Seekers, and Policy in OECD Countries. *American Economic Review*, 106(5):441–445.

Hatton, T. J. (2017). Refugees and Asylum Seekers, The Crisis in Europe and The Future of Policy. *Economic Policy*, 32(91):447–496.

Hirschle, J. and Kleiner, T.-M. (2014). Regional Cultures Attracting Interregional Migrants. *Urban Studies*, 51(16):3348–3364.

Ibánez, A. M. (2014). Growth in Forced Displacement: Cross-Country, Sub-National and Household Evidence on Potential Determinants. In Lucas, R. E., editor, *International Handbook on Migration and Economic Development*, Chapters, chapter 13, pages 350–387. Edward Elgar Publishing.

Mayda, A. M. (2010). International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics*, 23(4):1249–1274.

Maystadt, J.-F., Hirvonen, K., Mabiso, A., and Vandercasteelen, J. (2019). Impacts of Hosting Forced Migrants in Poor Countries. *Annual Review of Resource Economics*, 11:439–459.

McFadden, D. L. (1984). *Econometric analysis of qualitative response models*, volume 2 of *Handbook of Econometrics*. Elsevier.

Mitze, T. (2019). The Migration Response to Local labour Market Shocks: Evidence from EU Regions during the Global Economic Crisis. *Oxford Bulletin of Economics and Statistics*, 81(2):271–298.

Ravenstein, E. G. (1985). The Laws of Migration. *Journal of the Royal Statistical Society*, 48(2):167–235.

Ravenstein, E. G. (1989). The Laws of Migration – Second Paper. *Journal of the Royal Statistical Society*, 52(2):241–305.

Razin, A. and Wahba, J. (2015). Welfare Magnet Hypothesis, Fiscal Burden, and Immigration Skill Selectivity. *The Scandinavian Journal of Economics*, 117(2):369–402.

Salah, A. A., Pentland, A., Lepri, B., Letouzé, E., Vinck, P., de Montjoye, Y.-A., Dong, X., and Dagdelen, Ö. (2018). Data for Refugees: The D4R Challenge on Mobility of

Syrian Refugees in Turkey. *ArXiv*, abs/1807.00523.

Santos Silva, J. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Türk Telekom (2019). Turk Telekom At A First Glance. Retrieved from https://www.turktelekom.com.tr/en/aboutus/pages/turk-telekom-at-a-first-glance.aspx.

Tumen, S. (2015). The use of natural experiments in migration research. *IZA World of Labor*.

Tumen, S. (2016). The Economic Impact of Syrian Refugees on Host Countries: Quasi-experimental Evidence from Turkey. *American Economic Review*, 106(5):456–460.

Tumen, S. (2019). Refugees and 'native flight' from public to private schools. *Economics Letters*, 181:154–159.

TurkStat (2020a). Gross Domestic Product Per Capita, Value (2009 Base). *Ankara: Turkish Statistical Institute*.

TurkStat (2020b). Share of the Regions by Sectors, Value (2009 Base). *Ankara: Turkish Statistical Institute*.

United Nations High Commissioner for Refugees (2016). UNHCR Turkey: Syrian Refugee Camps and Provincial Breakdown of Syrian Refugees Registered in South East Turkey.

United Nations High Commissioner for Refugees (2017a). 2017 Year-End Report. Operation: Turkey.

United Nations High Commissioner for Refugees (2017b). *Global Trends: Forced Displacement in 2016.* United Nations High Commissioner for Refugees (UNHCR) .

United Nations High Commissioner for Refugees (2019). UNHCR Turkey 2019 Operational Highlights. Retrieved from http://reporting.unhcr.org/node/27380.

United Nations High Commissioner for Refugees (2020). *Global Trends: Forced Displacement in 2019.* United Nations High Commissioner for Refugees (UNHCR) .

Vincenty, T. (1975). Direct and Inverse Solutions of Geodesics on the Ellipsoid with

Application of Nested Equations. *Survey Review*, 23:88–93.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., and Buckee, C. O. (2012). Quantifying the Impact of Human Mobility on Malaria. *Science*, 338:267–270.

World Bank (2018). Moving for Prosperity: Global Migration and Labor Markets. Policy Research Report. Washington, DC: World Bank.

Tables and Figures

The following pages present tables and figures, which we refer to in the main text.

Table 1: Summary Statistics of the Variables.

	Obs.	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Dependent Variables					
Mobility of Refugees	1950	0.006	0.022	0	0.333
Mobility of Non-Refugees	1950	0.012	0.054	0	1.375
Explanatory Variables					
Income Per Capita	1950	855.898	276.491	327.743	1734.066
Distance	1950	580.118	314.576	95.520	1398.486
Humanitarian Aid	1950	2.902	4.732	0	30
Asylum Grant	1950	0.529	0.900	0	9
Violent Protest	1950	3.655	13.154	0	68

Table 2: Average Traveled Distance and Number of Moves of Refugees and Non-Refugees.

		Refugees	Non-Refugees
Average Traveled Distance (km/movers)		581,7	733,2
Number of Moves	0	21645	30197
	1	1006	806
	2	225	645
	3	14	51

Table 3: Benchmark Analysis: Determinants of Refugee Mobility in Turkey.

Variable	(1)	(2)	(3)	(4)
		Refugees		**Non-Refugees**
Log Income at origin	-0.759***	-	-0.759***	-0.156
	(0.219)		(0.219)	(0.212)
Log Income at destination	1.405***	1.405***	-	2.019***
	(0.196)	(0.196)		(0.167)
Log Distance	-0.595***	-0.595***	-0.595***	-0.436***
	(0.111)	(0.111)	(0.111)	(0.088)
Constant	-5.172**	-10.86***	5.170***	-14.85***
	(2.244)	(1.530)	(1.757)	(1.951)
Origin FE	Yes	No	Yes	Yes
Destination FE	Yes	Yes	No	Yes
Time FE	Yes	No	No	Yes
Origin-time FE	No	Yes	No	No
Destination-time FE	No	No	Yes	No
Regions	26	26	26	26
Observations	1,950	1,950	1,950	1,950
R-squared	0.357	0.445	0.425	0.656

Notes: Estimated equation: equation (6) using PPML except columns (2) and (3). 10-call filtering procedure used and time window between between 8 pm and 8 am. Level of regional analysis: NUTS–2. Dependent variable is measured by a quarterly migration rate $\frac{N_{ij,t}}{N_{ii,t}}$, where N_{ij} corresponds to migrants in the ij corridor and $N_{ii,t}$ to stayers. Robust standard errors clustered at the origin and destination are reported in parentheses. *** denotes statistical significance at the 1 percent level ($p < 0.01$), ** at the 5 percent level ($p < 0.05$), and * at the 10 percent level ($p < 0.10$), all for two-sided hypothesis tests.

Table 4: Robustness Tests on the Level of Regional Analysis.

	Dep. Var: Mobility of Refugees		
	NUTS–2	NUTS–1	NUTS–3
Variable	(1)	(2)	(3)
Log income at origin	-0.759***	-0.463**	-0.585*
	(0.219)	(0.271)	(0.357)
Log income at destination	1.405***	1.098***	3.515***
	(0.196)	(0.193)	(0.841)
Log Distance	-0.595***	-0.447***	-0.914***
	(0.111)	(0.129)	(0.131)
Constant	-5.172**	-6.135**	-20.11***
	(2.244)	(2.725)	(6.591)
Origin FE	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Regions	26	12	81
Observations	1,950	396	14,085
R-squared	0.357	0.622	0.151

Notes: Estimated equation: equation 6 using PPML. 10-call filtering procedure used and time window between between 8 pm and 8 am. Dependent variable is measured by a quarterly migration rate $\frac{N_{ij,t}}{N_{ii,t}}$, where N_{ij} corresponds to migrants in the ij corridor and $N_{ii,t}$ to stayers. Robust standard errors clustered at the origin and destination are reported in parentheses. *** denotes statistical significance at the 1 percent level ($p < 0.01$), ** at the 5 percent level ($p < 0.05$), and * at the 10 percent level ($p < 0.10$), all for two-sided hypothesis tests.

Table 5: Robustness checks to alternative filters of phone calls.

	Dep. Var: Mobility of Refugees					
Frequency Filter:	10-calls	20-calls	5-calls	10-calls	20-calls	5-calls
Min. number of calling days:		1		10	20	5
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Log income at origin	-0.759***	-0.873***	-0.668***	-0.832***	-1.060***	-0.856***
	(0.219)	(0.251)	(0.200)	(0.273)	(0.312)	(0.197)
Log income at destination	1.405***	1.063***	1.583***	0.999***	0.890***	1.276***
	(0.196)	(0.222)	(0.168)	(0.210)	(0.226)	(0.221)
Log Distance	-0.595***	-0.561***	-0.532***	-0.628***	-0.732***	-0.590***
	(0.111)	(0.145)	(0.102)	(0.145)	(0.193)	(0.125)
Constant	-5.172**	-2.704	-7.297***	-2.125	0.363	-3.819
	(2.244)	(2.596)	(1.927)	(2.711)	(2.981)	(2.386)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Regions	26	26	26	26	26	26
Observations	1,950	1,950	1,950	1,950	1,950	1,950
R-squared	0.357	0.144	0.529	0.139	0.134	0.282

Notes: Estimated equation: equation 6 using PPML. 10-call filtering procedure used and time window between between 8 pm and 8 am. Dependent variable is measured by a quarterly migration rate $\frac{N_{ij,t}}{N_{ii,t}}$, where N_{ij} corresponds to migrants in the ij corridor and $N_{ii,t}$ to stayers. Robust standard errors clustered at the origin and destination are reported in parentheses. Cols 1 and 4:10-call filtering procedure. Cols 2 and 5: 20-call filtering procedure. Cols 3 and 6: 5-call filtering procedure. Cols 1-3 : minimum number of days of calls : 1. Cols 4-6 : minimum number of days of calls : 10. *** denotes statistical significance at the 1 percent level ($p < 0.01$), ** at the 5 percent level ($p < 0.05$), and * at the 10 percent level ($p < 0.10$), all for two-sided hypothesis tests.

Table 6: Alternative stories of patterns of refugees mobility

	Dependent Variable: Mobility of Refugees (columns (1-4), (6-11)) and Non-Refugees (column (5))										
	Westward movements			Contiguity		Agric. BC		Camps	Soc. Magnet		Protests
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log income Orig	-0.995***	-1.031***	-0.807***	-0.642***	-0.051	-0.759***	-0.759***	-0.799***	-0.726***	-0.636***	-0.673***
	(0.248)	(0.251)	[0.262]	(0.233)	(0.214)	(0.219)	(0.219)	(0.165)	(0.223)	(0.221)	(0.230)
Log income Dest	1.554***	1.591***	3.530***	1.414***	2.025***	1.424***	1.427***	1.595***	1.375***	1.361***	1.313***
	(0.212)	(0.218)	[0.722]	(0.167)	(0.156)	(0.193)	(0.195)	(0.234)	(0.190)	(0.199)	(0.209)
Log Distance	-0.645***	-0.630***	-	-0.303***	-0.197	-0.595***	-0.595***	-0.712***	-0.589***	-0.586***	-0.599***
	(0.107)	(0.107)		(0.147)	(0.122)	(0.111)	(0.111)	(0.113)	(0.112)	(0.113)	(0.111)
West*Log Inc.Orig	-0.065***	-	-	-	-	-	-	-	-	-	-
	(0.025)										
West*Log Inc.Dest	-	-0.064***	-	-	-	-	-	-	-	-	-
		(0.025)									
Contiguity	-	-	-	0.551***	0.487**	-	-	-	-	-	-
				(0.196)	(0.194)						
Agric. Inc (30% GDP)	-	-	-	-	-	0.011	-	-	-	-	-
						(0.045)					
Agric. Inc (50% GDP)	-	-	-	-	-	-	0.013	-	-	-	-
							(0.061)				
Hum aid. Orig	-	-	-	-	-	-	-	-	-208.3*	-	-
									(0.113.9)		
Hum aid. Dest	-	-	-	-	-	-	-	-	71.75	-	-
									(93.94)		
Asyl. Grants Orig	-	-	-	-	-	-	-	-	-	-1307***	-
										(450)	
Asyl.Grants Dest	-	-	-	-	-	-	-	-	-	280.7	-
										(368.0)	
Violent Protests Orig	-	-	-	-	-	-	-	-	-	-	-118,825
											(118,431)
Violent Protests Dest	-	-	-	-	-	-	-	-	-	-	129,722
											(138,288)
Constant	-4.116*	-4.197*	-24.96***	-7.848***	-17.09***	-5.295**	-5.326**	-5.605**	-5.143**	-5.699**	-5.105**
	(2.275)	(2.276)	[6.004]	(2.494)	(2.057)	(2.204)	(2.175)	(2.224)	(2.219)	(2.254)	(2.312)
Origin FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dyadic FE	No	No	Yes	No	No	No	No	No	No	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig-Dest FE	No	No	Yes	No	No	No	No	No	No	No	No
Regions	26	26	26	26	26	26	26	26	26	26	26
Observations	1,950	1,950	1,032	1,950	1,950	1,950	1,950	1,950	1,950	1,950	1,950
Pseudo R^2	0.363	0.144	0.482	0.366	0.655	0.358	0.358	0.362	0.364	0.366	0.364

Notes: Estimated equation: equation 6 using PPML except column (3). 10-call filtering procedure used and time window between between 8 pm and 8 am. Dependent variable is measured by a quarterly migration rate $\frac{N_{ij,t}}{N_{ii,t}}$, where N_{ij} corresponds to migrants in the ij corridor and $N_{ii,t}$ to stayers. Migration rate is for refugees except in column (5) which reports results for non-refugees. Robust standard errors clustered at the origin and destination are reported in parentheses. Robust standard errors are reported in brackets (see footnote 25). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: The Presence of Refugees in Turkey in 2017:
Administrative Data versus D4R Data.

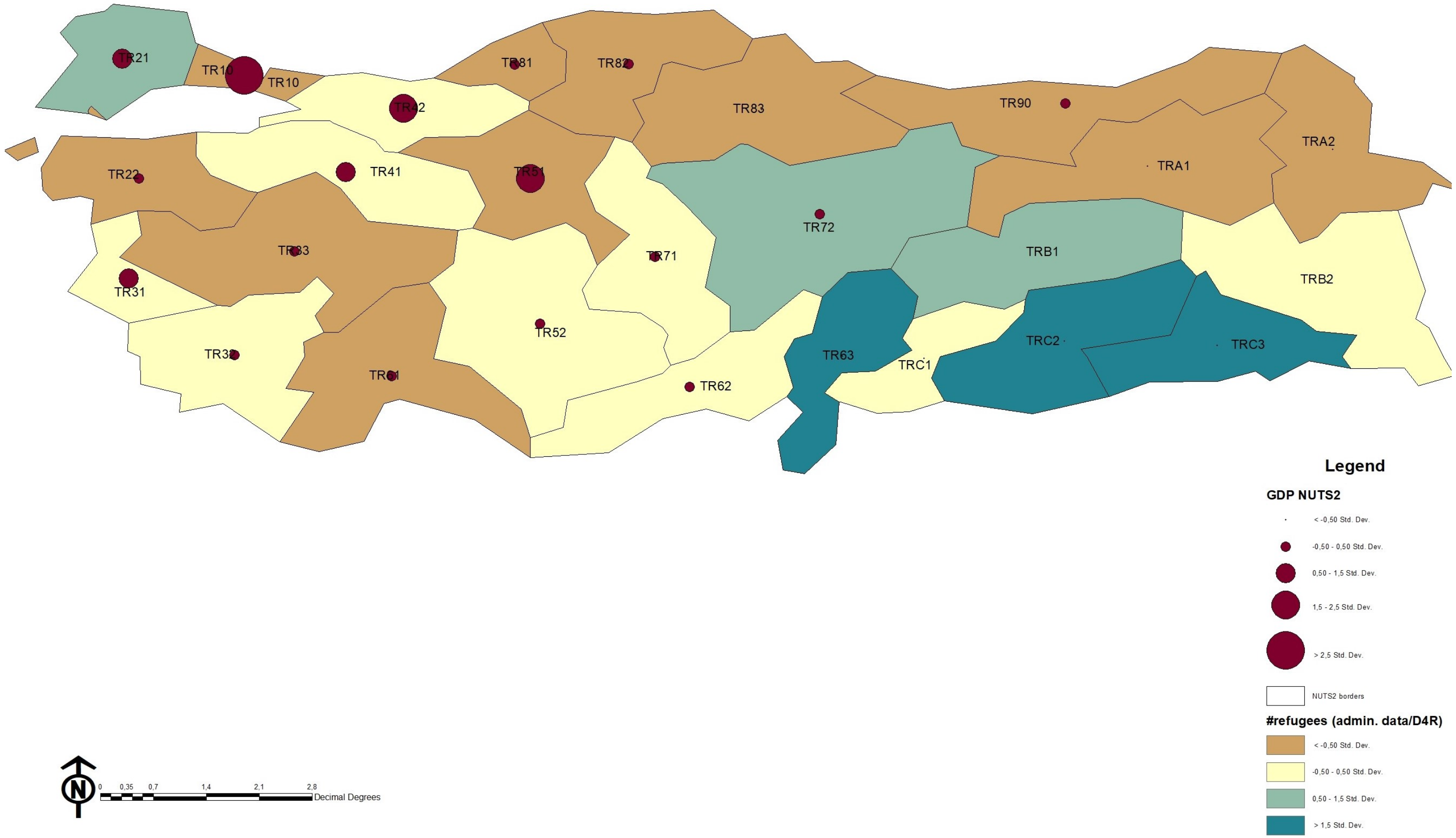
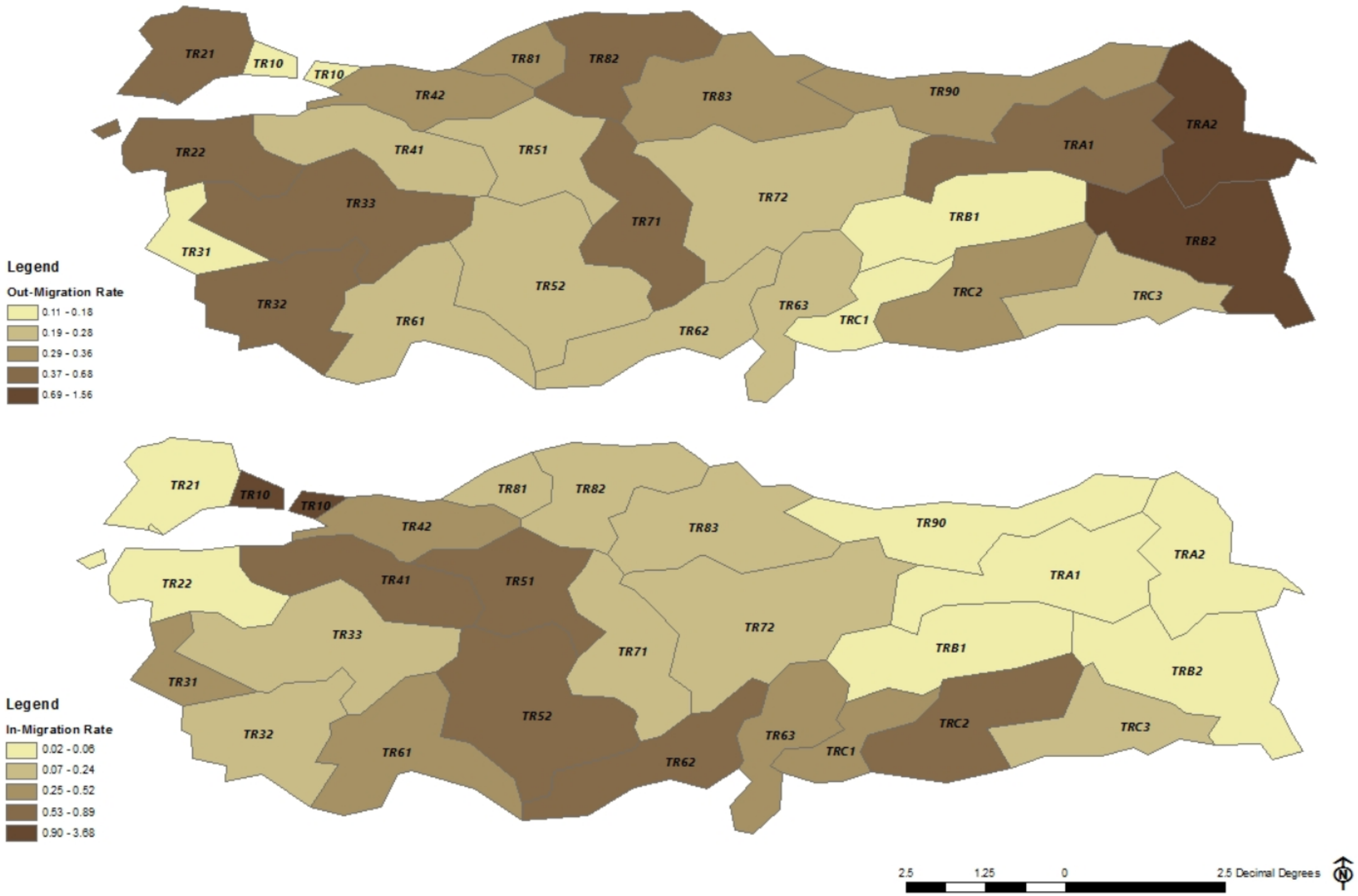


Figure 2: Mobility of Refugees in Turkey in 2017:
Out-Migration versus In-Migration of Refugees.



Separate Appendixes with Supplemental Material for:

A Gravity Analysis of Refugee Mobility Using Mobile Phone Data

December 30, 2020

Abstract

This document contains a set of appendixes with supplemental material.

Keywords: Refugee Mobility, Gravity Model of Migration, Forced Displacement, Mobile Phone Data, News Media, Poisson Pseudo-Maximum Likelihood.

JEL Classification: J6; 015

Appendix A Data

Appendix A.1 The D4R Challenge-Datasets 1 and 2

Türk Telekom (TT) and the D4R Challenge gave access to three distinct datasets providing different types of information. Two additional datasets (beyond Dataset 3 which is the one we use in our analysis) are the following:

Dataset 1 (Antenna Traffic) includes one year site-to-site traffic on an hourly basis and it provides information about the traffic between each site for a year period. A prerequisite is of course that one of the involved parties is registered with TT and in this case the information is available only for the TT customer. Information such as the total number of calls and total duration of the calls are available only in an aggregate format. This information is available both for voice calls and SMS messages. While no personal characteristics are revealed, there is information on the total number and the total duration of refugee calls and SMS by antennae.

Dataset 2 (Fine Grained Mobility) randomly chooses a group of active users (who make calls and send SMS) every two-week period and reports cell tower identifiers. This random-sampling process is repeated for the whole year period. As in Dataset 1, there is no personal information revealed, beyond the refugee or non-refugee status. The dataset provides information about the base station ID, whether the call or SMS is incoming or outgoing as well as the day and hour.

Crucially, as phone call data contain highly sensitive information, there is no possibility to link the three datasets.

Appendix A.2 GDELT

The Global Database of Events, Language and Tone (GDELT) dataset is a world-wide news media platform that is available for over 30 years, in over 100 languages and is updated daily to construct a number of indices related to the incidence of news that could directly or indirectly concern the refugee population (GDELT, 2019). The database consists of over a quarter billion geo-referenced event records in over 300 categories. The platform is open for research and analysis. It uses the Conflict And Mediation Event Observations (CAMEO) system where a code corresponds to a type of event and is defined in a three-level taxonomy. Each observation provides information in several layers. For instance, every observation has information about the location, the involved actors, the impact of the event, the type of action, to mention a few of the available categories. Another element available in GDELT that is essential for our analysis is the tone of the news, i. e. whether it has a negative or a positive connotation for the refugee population. Since the same type of news may have a different effect depending on the tone, we also aggregate the news based on the tone.

With respect to location, each observation provides latitude and longitude, thus the data are being reported at a very fine level. Using geographic information system (GIS), we are able to construct our events variables at the NUTS–2 level, in line with our main analysis. Moreover, the news coverage also has time variation at a fine level and we can thus construct the same measures at the quarterly level between January 2017 and December 2017.

Using the EVENT Record Exporter tool[29] provided by GDELT, we first obtain all events that took place in 2017 in Turkey. We then choose the following categories: humanitarian aid, asylum grants and violent protests. Indeed, GDELT provides information about several types of news events, which are given an id *GlobalEventID* and a variable *EventBaseCode*, which shows to which category this particular event belongs to.

Examples of news in our analysis are e. g., delivery of financial aid and other essential items to patients and other injured in the Syrian war, during a visit to several hospitals in the city of Kilis or the Turkish government setting up new refugee health-care centers

[29] The EVENT Record Exporter allows to export small subsets of data from the GDELT Event Database that match the search criteria. By specifying a set of criteria for the event type and actors involved, along with an optional date range, the system will search the entire GDELT Event Database for all matching entries and export matching records as a CSV file (GDELT, 2014).

and employ Syrian medical staff in various Turkish cities. To explore the effect of "social magnet" (Section 4.3.6) at the NUTS–2 region and the quarterly levels, we aggregate the number of news reporting information on "humanitarian aid, mainly in the form of emergency assistance" and on "grant[ing] asylum to persons". For these events, we focus on those with a positive tone to ensure that selected events relate to actual provisions of aid and asylum. To assess the role of violent protests (Section 4.3.7) at the NUTS–2 region and the quarterly levels, we aggregate the number of news reporting information on "protests forcefully, in a potentially destructive manner", For these events, we focus on those with a negative tone to ensure that selected events relate to actual occurrence of violent protests.

The distribution of these events in 2017 across the NUTS-2 regions in Turkey is shown in Table B.3.

Appendix B Supplementary Tables and Figures

Table B.1: NUTS Statistical Regions of Turkey.

NUTS–1 Regions	NUTS–2 Subregions	NUTS–3 Provinces
Istanbul (TR1)	Istanbul (TR10)	Istanbul (TR100)
West Marmara (TR2)	Tekirdağ (TR21)	Tekirdağ (TR211)
		Edirne (TR212)
		Kırklareli (TR213)
	Balıkesir (TR22)	Balıkesir (TR221)
		Çanakkale (TR222)
Aegean (TR3)	Izmir (TR31)	İzmir (TR310)
	Aydın (TR32)	Aydın (TR321)
		Denizli (TR322)
		Muğla (TR323)
	Manisa (TR33)	Manisa (TR331)
		Afyonkarahisar (TR332)
		Kütahya (TR333)
		Uşak (TR334)
East Marmara (TR4)	Bursa (TR41)	Bursa (TR411)
		Eskişehir (TR412)
		Bilecik (TR413)
	Kocaeli (TR42)	Kocaeli (TR421)
		Sakarya (TR422)
		Düzce (TR423)
		Bolu (TR424)
		Yalova (TR425)
West Anatolia (TR5)	Ankara (TR51)	Ankara (TR510)
	Konya (TR52)	Konya (TR521)
		Karaman (TR522)
Mediterranean (TR6)	Antalya (TR61)	Antalya (TR611)
		Isparta (TR612)
		Burdur (TR613)
	Adana (TR62)	Adana (TR621)
		Mersin (TR622)
	Hatay (TR63)	Hatay (TR631)
		Kahramanmaraş (TR632)
		Osmaniye (TR633)

Central Anatolia (TR7)	Kırıkkale (TR71)	Kırıkkale (TR711)
		Aksaray (TR712)
		Niğde (TR713)
		Nevşehir (TR714)
		Kırşehir (TR715)
	Kayseri (TR72)	Kayseri (TR721)
		Sivas (TR722)
		Yozgat (TR723)
West Black Sea (TR8)	Zonguldak (TR81)	Zonguldak (TR811)
		Karabük (TR812)
		Bartın (TR813)
	Kastamonu (TR82)	Kastamonu (TR821)
		Çankırı (TR822)
		Sinop (TR823)
	Samsun (TR83)	Samsun (TR831)
		Tokat (TR832)
		Çorum (TR833)
		Amasya (TR834)
East Black Sea (TR9)	Trabzon (TR90)	Trabzon (TR901)
		Ordu (TR902)
		Giresun (TR903)
		Rize (TR904)
		Artvin (TR905)
		Gümüşhane (TR906)
Northeast Anatolia (TRA)	Erzurum (TRA1)	Erzurum (TRA11)
		Erzincan (TRA12)
		Bayburt (TRA13)
	Ağrı (TRA2)	Ağrı (TRA21)
		Kars (TRA22)
		Iğdır (TRA23)
		Ardahan (TRA24)
Central East Anatolia (TRB)	Malatya (TRB1)	Malatya (TRB11)
		Elazığ (TRB12)
		Bingöl (TRB13)
		Tunceli (TRB14)
	Van (TRB2)	Van (TRB21)
		Muş (TRB22)
		Bitlis (TRB23)
		Hakkâri (TRB24)
Southeast Anatolia (TRC)	Gaziantep (TRC1)	Gaziantep (TRC11)
		Adıyaman (TRC12)
		Kilis (TRC13)
	Şanlıurfa (TRC2)	Şanlıurfa (TRC21)
		Diyarbakır (TRC22)
	Mardin (TRC3)	Mardin (TRC31)
		Batman (TRC32)
		Şırnak (TRC33)
		Siirt (TRC34)

Table B.2: Definitions of the Variables.

	Definition
Dependent Variables	
Mobility of Refugees	A quarterly migration rate $\frac{N_{ij}}{N_{ii}}$, where N_{ij} corresponds to leavers and N_{ii} to stayers, which is of the form *migr_rate_'R'_'10'* where *'R'* indicates the refugee status of the observation, and *'10'* corresponds to the frequency filter, e. g. the minimum number of incoming and outgoing calls generated from a given NUTS–2 region between 8 pm and 8 am to characterize the latter as the residence location of an individual. **Source.** Dataset 3 – Coarse Grained Mobility, D4R Challenge
Mobility of Non-Refugees	A quarterly migration rate $\frac{N_{ij}}{N_{ii}}$., where N_{ij} corresponds to leavers and N_{ii} to stayers, which is of the form *migr_rate_'NR'_'10'* where *'NR'* indicates the non-refugee status of the observation and *'10'* corresponds to the frequency filter, e. g. the minimum number of incoming and outgoing calls generated from a given NUTS–2 region between 8 pm and 8 am to characterize the latter as the residence location of an individual. **Source.** Dataset 3 – Coarse Grained Mobility, D4R Challenge.
Explanatory Variables	
Income Per Capita	Quarterly GDP in Turkey, weighted by the share and divided by the population of each NUTS–2 region. **Source.** Turkish Statistical Institute
Distance	Geodesic distances (Vincenty, 1975), i. e. the length of the shortest curve between two points along the surface of a mathematical model of the earth, based on the WGS 1984 datum coordinates of the centroids of each NUTS–2 region. **Source.** Eurostat
Humanitarian Aid	Number of news reporting information defined as “Extend, provide humanitarian aid, mainly in the form of emergency assistance”. **Source.** GDELT
Asylum Grants	Number of news reporting information defined as “Provide, grant asylum to persons”. **Source.** GDELT
Violent Protests	Number of news reporting information defined as “Protest forcefully, in a potentially destructive manner”. **Source.** GDELT

Table B.3: Distribution of Events in 2017 across NUTS–2 regions in Turkey.

NUTS–2 Subregions	Humanitarian Aid	Asylum Grants	Violent Protests
İstanbul (TR10)	87	6	107
Tekirdağ (TR21)	0	1	2
Balıkesir (TR22)	5	1	1
İzmir (TR31)	20	3	7
Aydın (TR32)	2	0	0
Manisa (TR33)	5	0	0
Bursa (TR41)	0	0	1
Kocaeli (TR42)	8	0	0
Ankara (TR51)	81	12	150
Konya (TR52)	3	1	1
Antalya (TR61)	11	0	3
Adana (TR62)	20	3	4
Hatay (TR63)	8	4	0
Kırıkkale (TR71)	0	0	0
Kayseri (TR72)	24	0	6
Zonguldak (TR81)	4	0	0
Kastamonu (TR82)	0	0	0
Samsun (TR83)	2	0	5
Trabzon (TR90)	1	0	0
Erzurum (TRA1)	3	0	11
Ağrı (TRA2)	6	0	0
Malatya (TRB1)	1	5	0
Van (TRB2)	3	0	0
Gaziantep (TRC1)	10	2	5
Şanlıurfa (TRC2)	26	0	6
Mardin (TRC3)	3	0	2

Figure B.1: Turkish Administrative Regions:
Nomenclature des Unités Territoriales Statistiques.

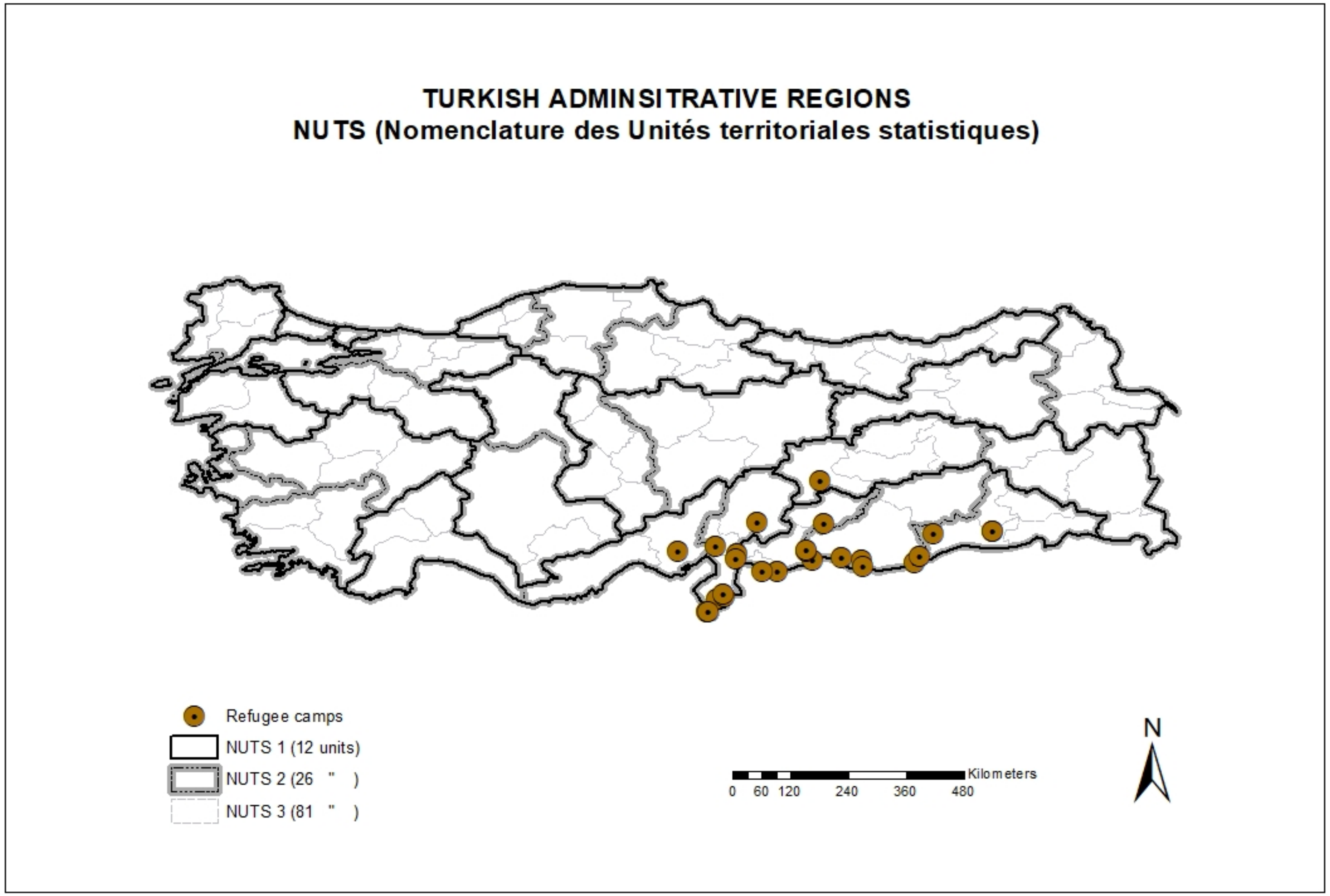


Figure B.2: Histogram of the Distance Covered by Refugees and Non-Refugees.