Segregation and internal mobility of Syrian refugees in Turkey : Evidence from mobile phone data

Simone Bertoli, Caglar Ozden, Michael Packard

Segregation and internal mobility of Syrian refugees in Turkey: Evidence from mobile phone data*

Simone Bertoli[a], Caglar Ozden[b], and Michael Packard[c]

[a] *Université Clermont Auvergne, CNRS, CERDI, IZA and IUF*

[b] *Development Research Group, the World Bank*

[c] *Georgetown University*

Abstract

We use mobile phone usage data to measure the extent of segregation of Syrian refugees in Turkey, and analyze its role in their internal mobility patterns. We construct a range of dissimilarity and normalized isolation indices using the hourly phone call volume of refugees and natives. The richness of the data allows us to compute the indices across different provinces and over time. Segregation levels show high variation across the country, with significantly lower levels of segregation in provinces with a higher share of refugees. Refugee mobility across provinces over time appears to be negatively correlated with segregation at destination, while native mobility is not. Based on data from Istanbul, segregation does not influence intra-province mobility. This is possibly due to the differences in segregation indices across the hours of the day, suggesting that residential segregation is higher than labor market segregation.

Keywords: mobile phones; call detail records; segregation; Syrian refugees; Turkey.

JEL codes: O18; F22.

*The authors are grateful to the Editor-in-Chief Andrew Foster and to three anonymous referees for their comments; this study is performed using the one-year anonymized mobile communication data made available by Türk Telekomünikasyon A.Ş. within the D4R Challenge; the authors are grateful to the participants to the D4R workshop (Boğaziçi University, January 2019), to the 12^{th} Migration and Development Conference (UC3M, Madrid, June 2019) and to the UNHCR/WB Conference on Forced Displacement (Copenhagen, January 2020) for their comments; Simone Bertoli acknowledges the support from the *Agence Nationale de la Recherche* through the program "*Investissements d'avenir*" (ANR-10-LABX-14-01); the findings presented here do not necessarily represent the views of the World Bank's Board of Executive Directors or the governments they represent; the usual disclaimers apply.

"When refugees leave their homes, they enter an informational no-man's-land. Phones become a lifeline [...] It is like the underground railroad, only that it is digital." *The Economist*, February 11, 2017.

1 Introduction

Digital breadcrumbs left by various human activities present a promising and yet mostly untapped data sources for research in developing countries (Cesare et al., 2018). The academic potential of these "big data" sources arises from their high-level spatial resolution, availability in real time and relatively low processing costs even though they are typically collected for unrelated commercial reasons.[1] The value of such digital data is further magnified when specific groups – such as refugees or undocumented migrants – might not be fully enumerated in standard surveys or censuses,[2] or when one is interested in fine-grained outcomes such as residential segregation.

This paper uses a mobile phone dataset to address central questions in the migration literature: What are patterns of segregation of refugees within their host communities? What are the linkages between the refugees' internal mobility patterns and extent of segregation in different locations? Our empirical innovation is the utilization of a unique dataset that consists of hourly phone activity of Syrian refugees in Turkey and a comparable group of Turkish citizens at over thirty thousand cell phone towers. High spatial and temporal disaggregation of the dataset allows us to not only measure the level of different segregation indices with precision, but also to observe their changes over relatively short time periods and narrow geographic areas. Using this information, we, then, investigate the extent of the role played by spatial segregation in the internal mobility decisions of refugees and natives as they move to other regions within the country.

Escaping violence and seeking safety are the main reasons why refugees arrive at the border of another country. In the case of middle- or high income destinations, economic factors enter the picture as refugees cross the border and start moving within the country

[1] Individual-level wealth and poverty status, for example, can be inferred from the analysis of mobile phone activity, instead of the consumption data collected through expensive household surveys (Blumenstock et al., 2015)

[2] Surveys conducted at destination fail to enumerate recently arrived immigrants (Hanson, 2006). The collection of data on the migration histories of household members in the origin countries faces the challenges posed by whole household migration (Ibarraran and Lubotsky, 2007), deliberate misreporting (Hamilton and Savinar, 2015) and household dissolution (Bertoli and Murard, 2020).

to seek places for more permanent settlement. Among the factors that influence internal mobility decisions of refugees are similar to those for economic migrants and natives – availability of appropriate jobs, affordable housing and the presence of other people with similar cultural backgrounds at this location. These diaspora networks provide the social capital, information and support mechanisms to navigate the new economic and cultural environment. While diaspora effect is about the relative size of the refugee population in a given region or a province, segregation is about the distribution of the refugees within that area. An uneven distribution of refugees at a given province might make that location less appealing. In addition to competing for the same jobs or housing, refugees might face discrimination or animosity from the local population as their numbers increase and they live in large segregated and isolated communities, something that might increase their visibility, and hence the salience of their presence of the natives. Our empirical analysis aims to compare these forces and assess the importance of segregation at a location relative to these other pull and push factors identified in the literature in attracting or deterring new arrivals (World Bank, 2018).

Our analysis would not have been possible without the phone usage data. Mobile phones have become critical tools for refugees to keep in touch with their friends and families both in the origin and in the host countries as they find their way to their intended destinations. Phones enable the refugees to perform many daily economic activities, ranging from searching for jobs to finding apartments. Detailed mobile phone data on the precise time, location and duration of the calls can provide valuable academic and policy insights. These data are stored by operators in Call Detail Records (CDRs) for billing purposes. To aid academics in evidence-based policy formulation, Türk Telekom, one of the three leading mobile phone operators in Turkey, has made available anonymized and aggregated information based on the CDRs of the universe of their Syrian customers and of a large sample of comparable Turkish customers within the context of their Data for Refugees (D4R) Challenge. The data covers the time frame from January 1 to December 31, 2017.

The D4R data provide a detailed snapshot of the geographic segregation of refugees in Turkey, which has become the host to the majority of the Syrians who fled during the civil conflict that started in 2011 (World Bank, 2018). High geographic (at the transmission tower level) and time (hourly) variation of our data allows us to explore how the levels of interaction between refugees and the Turkish population changes over time and across different regions.

The data enable us to compare different measures of segregation across regions of Turkey that have been differently exposed to the massive inflow of refugees.[3]

We use the call volume of refugees and natives at a given cell tower (or a geographic area in which the tower is located) as a proxy for the population distribution. First, we construct dissimilarity and (normalized) isolation indices that have been extensively studied in the literature to measure the segregation of minority communities. The richness of the data allows us to calculate these indices over time, across different provinces, even during different hours of the day. We find that segregation levels show high variation across the country, with significantly lower levels of dissimilarity and isolation in provinces with higher share of refugees. The dissimilarity index is generally stable over time, both nationally and in many provinces with large numbers of refugees. The isolation index, however, shows some variation and tends to increase over time.[4]

Once we identify and discuss the patterns of segregation over time and across geographic areas, we estimate gravity models of internal mobility to understand the determinants of refugees' and natives' mobility decisions during 2017. In addition to the standard economic and geographic factors, we ask whether if the extent of segregation in an area can ask as a pull or a push factor and influence the mobility decisions of refugees and natives.[5]

In our inter-province gravity estimation, we find that Syrians tend to move towards provinces that have large overall populations, high income per capita and host a larger number of refugees, indicating the power of economic forces and diaspora externalities. In addition, refugees' location choices negatively correlate with the measured level of dissimilarity at destination, implying that refugees prefer to move to provinces with *lower* measures of segregation. Similar economic factors influence the mobility decisions of natives, but segregation has no effect.

An important share of internal migration takes place within large provinces. With a refugee population of over 1 million and total population of 15 million, Istanbul provides the ideal case study. Running a similar gravity estimation for migration among 41 districts

[3]For example, the Southeastern provinces close to the Syrian border host a disproportionate share of the refugees, and the same is true for big cities like Istanbul which is now home to a quarter of the refugees. (Cavlin, 2020)

[4]The increased mobile phone penetration of Türk Telekom in the population of refugees over our period of analysis might also be at play here, notably for the increase in the isolation index.

[5]Beine et al. (2021) also estimate gravity equations for refugees and natives using the D4R data by Türk Telekom but instead focus on the mobility response of Syrians to political and social events.

of Istanbul, we find that economic factors and diaspora networks again are the key determinants. Segregation in this case has no effect on the intra-province mobility patterns of refugees or natives. Finally, we explore differences in our segregation indices across hours of the day. Both dissimilarity and isolation indices significantly dip during the day, especially in early afternoon and increase rapidly in the evening. These patterns imply that there is large variation between where the refugees live and work. In other words, residential segregation is higher than labor market segregation. This might also be the reason why we do not observe segregation as an important determinant of intra-provincial mobility decisions.

The rest of the paper is organized as follows: Section 2 provides a review of the relevant literature; Section 3 describes the main data source that is used in the analysis, namely the D4R data provided by Türk Telekom; Section 4 introduces the dissimilarity and isolation indices that are used in the analysis and presents the main patterns over time and space. Section 5 presents the results from the empirical analysis, linking segregation measures to mobility patterns. Finally, Section 6 summarizes the main results and concludes.

2 Literature survey

The spatial segregation of immigrant, racial or ethnic groups has been widely and systematically studied by social scientists since at least the 1940s (Jahn et al., 1947; Duncan and Duncan, 1955). Researchers across disciplines have focused on segregation because of its close relationship–both as a determinant and a response–with extreme poverty, inequality, and discrimination. In many contexts, location plays an important role in one's life experiences, influencing the amount and quality of public goods, amenities and jobs available as well as the set of other individuals one interacts with on a daily basis. Therefore, the social composition (based on race, ethnicity or socioeconomic status) of the residents of a location can have vast implications when it comes to economic and social outcomes for individuals and groups. In return, the extent of segregation at a location becomes a key factor in influencing if a person would like to move there, regardless of whether he is a member of the minority or the majority community.

Broadly, the literature on spatial segregation can be grouped into three strands. The first one is methodological: How should segregation be measured and what aspects of segregation do different indices capture? The second strand focuses on the determinants and

magnitude of segregation, and how it varies over space, time, and across different ethnic or racial groups. Finally, the third strand of this literature measures the effects of segregation on different measures of social and economic progress. This strand includes (but is not limited to) the relationship of segregation to economic inequality, access to public goods and services, economic and cultural assimilation, and other measures of social interaction. Our contribution can be best placed within this strand as we explore how segregation influences mobility of the next group of people – both the refugees and the natives – as they make their mobility decisions.

While there has been a lively debate in the sociology literature on the efficacy of various segregation measures,[6] much of the debate had been settled by the late 1980s. The analysis done in this paper is mainly informed by Massey and Denton (1988) who perform a comprehensive analysis of 20 segregation indices using data from the United States. They identify five distinct dimensions of segregation: evenness, exposure, concentration, centralization and clustering. For each dimension, they suggest a measure that best captures each of these dimensions. The two measures used in this paper, the dissimilarity (D) index and the isolation (I) index, are offered as the suggested measures of evenness and exposure, respectively. Following the work of Cutler et al. (1999), and described below, we adjust our isolation index to account for the changing size of the refugee population over our observation period. The D-index measures the share of a minority population that would have to move in order to achieve full integration. Its value ranges from 0 to 1 with 1 implying full segregation and 0 implying full integration. The I-index, a measure of *exposure*, "gives the probability that [a minority group member] shares a unit with another [minority group member]" (Massey and Denton, 1988, p. 288).

Most papers in the literature focus on measuring the segregation of the African-American and Hispanic populations in the United States,[7] though there are also a number of papers focusing on the segregation of immigrants in the United States,[8] and a few measuring segregation of different populations abroad. Most papers find dissimilarity indices ranging between 0.3 and 0.7, with African-Americans experiencing higher levels of segregation than any other

[6] Jahn et al. (1947), Duncan and Duncan (1955), Cortese et al. (1976), Taeuber and Taeuber (1976), and Massey and Denton (1988), just to name a few.

[7] See, for instance, Cutler et al. (1999) and Massey and Denton (1988).

[8] See Catanzarite (2000), Clark and Blue (2004), Cutler et al. (2008a), Hall (2013), Iceland and Scopilliti (2008), Iceland (2004), Iceland and Nelson (2008) and Logan et al. (2002).

group studied.[9]

The literature that has used mobile phone data to study segregation is much smaller. Blumenstock and Fratamico (2013) uses mobile phone data from a large South Asian city to examine the relationship between social and spatial segregation. Järv et al. (2015) and Silm and Ahas (2014) are the papers most closely related to our own. Both papers use mobile phone data from Estonia's capital city of Talinn and study the spatial segregation of the Russian-speaking minority. The papers find that segregation is lower during the day and lower on workdays as compared to weekends, a pattern that we also observe among Syrian refugees in Turkey.

Segregation can have important impacts on the socioeconomic outcomes of minority groups. With respect to native-born minority groups, almost all results point to a negative impact of segregation. Most famously, this has been documented in Massey and Denton (1993), who argue that the segregation in America has contributed significantly to the extreme levels of poverty faced by minority populations. Segregation can cause inequality for a variety of reasons. There is evidence that racial and ethnic segregation leads to lower school quality for minority groups (Bygren and Szulkin, 2010; Mayer, 2002; Goldsmith, 2009), as well as inequality in the provision of other public goods (Trounstine, 2016), can cause disparities in health (Williams and Collins, 2001), and can limit access to high paying jobs for minority workers (Leonard, 1987; Turner, 2008). For immigrants, the results are not so clear. Cutler et al. (2008b) finds that, while immigrants tend to be negatively selected into ethnic enclaves. They experience positive socioeconomic effects from living in an enclave, these results are with consistent research that shows the positive effects of immigrant networks in finding employment opportunities and developing human capital (Borjas, 1992, 1995; Patel and Vella, 2013).

The determinants of the location-choices of refugees in the host countries have not been explored in the literature. We can expect that refugees (as migrants more in general) are more mobile than natives, as they have already paid the psychological cost of moving from home, which has an important fixed dimension as suggested by Sjaastad (1962).[10] Furthermore,

[9] While the literature has identified five distinct dimensions of segregation (and measures that do well to isolate each individual dimension), empirically they tend to be highly correlated, with areas that show high levels of one dimension also showing high levels of the four others.

[10] Parsons and Vézina (2018) provide evidence that a large share of Vietnamese refugees in the United States moved away from the location they were initially assigned within a few years.

they will have less of a personal attachment to specific location within the destination and incur lower emotional costs when they move. The absence of a large literature on this topic is surprising given that the stakes appear to be high for the host countries, as different location-choices have a first-order influence on the employment prospects of the refugees (Bansak et al., 2018). We contribute to fill in this gap through the estimation of a gravity equation on flow data computed separately for Syrian refugees and for Turkish natives from the D4R dataset. Consistently with the literature, we rely on a specification that can be derived from an underlying micro-foundation of location choices through a random utility maximization model *à la* McFadden.[11]

Several papers, by data scientists, sociologists and economists, explore issues related to segregation and assimilation of the refugees as part of the D4R challenge and they were published in the volume edited by Salah et al. (2019). Among the specific papers several are related to our paper. Boy et al. (2019) calculate dissimilarity and isolation indices. Instead of the call volume (from the larger Dataset 1) as a proxy for population, they use Dataset 2 which tracks a different set of people every two weeks. The small sample size creates certain biases, including underestimation of the population sizes when compared to the censuses. This is part of the reason why we use Dataset 1 and implement adjustments to the indices as we discuss in detail in the data and methodology sections. Marquez et al. (2019) also calculates segregation measures and links them to local attitudes towards refugees using Twitter data. Hu et al. (2019) and Sterly et al. (2019) explore mobility patterns, but do not perform causal analysis or link to segregation measures. Finally Beine et al. (2019) and Beine et al. (2021) use a gravity model to explore the impact of political and social event on movements of refugees. Our main departure is to focus on the measures of segregation refugees are exposed to in the Turkish province of destination.

3 Data

Our main database is based on Call Detail Records (CDRs) of incoming and outgoing calls of customers of Türk Telekom, one of the three main mobile phone operators with a market

[11]The estimation with aggregate data is equivalent to an estimation with the underlying individual-level data as we only have dyadic and destination-specific regressors (Guimaraes et al., 2003), and it is also consistent with the presence of unobserved individual heterogeneity in the cost of moving (Bertoli and Fernández-Huertas Moraga, 2015).

share of 24.7 percent in January 2017 (Salah et al., 2019). This dataset is made available to researchers within the framework of the Data for Refugees (D4R) Challenge (Salah et al., 2019), jointly led by Türk Telekom, Bosphorus University and TÜBITAK (the Turkish National Research Council). CDRs tell us whether the caller (initiator) or the callee (recipient) is a native (Turkish) or refugee (Syrian) customer of Türk Telekom. CDRs include the ID number of the tower utilized for the call and the time of the call. Türk Telekom system is able to identify whether a customer is a refugee because a Syrian passport or a refugee ID were used to initially register the mobile phone line and the SIM card comes with preferential rates available only to refugees.[12]

The D4R database includes the phone activity of a sample of nearly one million customers (992,457), out of which 184,949 are tagged as refugees, while the rest is composed of a sample of native customers. The D4R data captures all refugee customers whereas the native customers are sampled from the same spatial distribution (at the province level) as the refugee customers. The time-invariant tag for native and refugee customers is combined with the identifier of the tower to which the caller or the callee was connected during each call. This process generates the count of native *and* refugee calls for *each* tower in the network for *each* hour between January 1, 2017 and December 31, 2017. We will be referring to this as Dataset 1. Türk Telekom also discloses the exact latitude and longitude of each of tower in its network. This information allows us to match the call data in the D4R data and perform the analysis at different levels of Turkish administrative units (provinces, districts) or split the country into equally sized cells, or, as discussed below, restrict the analysis to towers located within urban areas.[13]

The key feature of this database, as mentioned earlier, is its high level of geographic and temporal disaggregation, as the unit of an observation in D4R Dataset 1 is given by the "hour-tower-type of customer" triplet.[14] The dataset includes more than thirty thousand distinct towers and $8,760 = 365 \times 24$ observations for each hour of the year in each tower

[12]More precisely, Türk Telekom offers to Syrian refugees 4 gigabytes of data, 1000 text message and 500 minutes of airtime to any phone number in Turkey, for 19 Turkish Lira a month, i.e., less than 4 USD; see http://www.on5yirmi5.com/haber/guncel/olaylar/226839/turk-telekomdan-turke-ayri-suriyeliye-ayri-fiyat.html (last accessed: April 8, 2019).

[13]See Salah et al. (2019) for a more detailed description of the database; a preliminary version of this chapter can be downloaded from: http://d4r.turktelekom.com.tr/Content/Documents/d4r-proceedings.pdf (last accessed April 8, 2019).

[14]We aggregate incoming and out-going calls in the analysis.

separately for refugees and natives. Figure 1 shows the distribution of number of calls by refugees and natives for each day of 2017. The map in Figure 2 highlights the geographic dimensions of our data, showing the share of the refugee calls among all calls made in each of the 957 districts (*ilçe* in Turkish), indicating the refugees form a smaller share of the population in the Northeast and central regions of the country.[15] It is clear from these density maps that most of the refugees are in South and big metropolitan areas like Istanbul.

Türk Telekom also made available, within the D4R Challenge, two additional datasets. Dataset 2 includes all the observations in the CDRs of a limited sample of refugee and native mobile phone users over a two-week period. Their calls can be identified at the tower level, and a new random sample is drawn for each one of the 26 two-week periods covered in the analysis. Dataset 3 includes a larger sample of Syrian refugees and Turkish natives who are followed throughout the year, but the location contained in their CDRs is coarsened at the district level to ensure that no individual user can be identified. We use these two datasets in conjunction with Dataset 1.[16] This high level of granularity allows us to calculate different segregation indices over time, across different provinces, even for different hours of the day.

4 Measures of segregation

The indices we construct to explore the extent and evolution of segregation of refugees, dissimilarity and isolation indices, are widely used in the economics, sociology and political science literature. We first define these indices and highlight their certain properties. To calculate the indices, we need to partition a geographic territory into K non-overlapping areas. This partitioning can have varying level of spatial disaggregation, ranging from the catchment areas of the 30,000 individual towers in the mobile phone network to the 957 administrative districts of the country. We label the dissimilarity index as D and the (normalized) isolation index as I^{adj}. Both indices are specific to (i) the partition of the country,

[15] A similar picture emerges if we show the distribution of the total calls made by the refugees across the country.

[16] An IT problem in the construction of the database caused the actual size of the database (as far as the time dimension is concerned) to be below its potential size. More precisely, data are missing for 82 days, mostly concentrated in March and April. We asked the scientific coordinators of the D4R Challenge, whether these gaps in the data could have been fixed. This is unfortunately not possible as the individual CDR data, which underlie this dataset, are destroyed by Türk Telekom after 12 months.

and (*ii*) the restrictions to certain areas of Turkey, but we omit these details when we define the indices below to avoid cluttering the notation. The partitioning is, however, clearly identified when we discuss specific set of results.

4.1 Dissimilarity index

The dissimilarity index D is defined by Massey and Denton (1988) as follows:

$$D \equiv \sum_{i=1}^{K} \frac{t_i |p_i - P|}{2TP(1-P)} \tag{1}$$

where t_i is total population in geographic area $i = 1, ..., K$ and p_i is the share of the minority (refugees in our case) population in area i. Based on these definitions, $m_i = t_i p_i$ is the size of the minority population in area i. The aggregate variables are denoted by capital letters: $T \equiv \sum_{i=1}^{K} t_i$ is total population, $M \equiv \sum_{i=1}^{K} m_i = TP$ is the total minority population and $P \equiv M/T$ is the share of the minority group in the total population. Based on these definitions, index D can be equivalently written as:

$$\begin{aligned} D &= \frac{1}{2(1-P)} \sum_{i=1}^{K} \left| \frac{m_i}{M} - \frac{t_i}{T} \right| \\ &= \frac{1}{2} \sum_{i=1}^{K} \left| \frac{m_i}{M} - \frac{n_i}{N} \right| \end{aligned} \tag{2}$$

where $n_i \equiv t_i - m_i$ is the native (non-minority) population in area i, and similarly $N = T - M$ is the total native population. The index $D \in [0, 1]$ has a simple interpretation as expressed above: It is the share of the minority population that needs to be relocated from high to low concentration regions to match their average distribution across the whole country.

Several limitations of the dissimilarity index have been studied in the literature. One well known problem is upward bias in the index when minority population forms a small share of the overall population or the population in certain geographic areas are small relative to the other areas (Allen et al., 2015; Mazza and Punzo, 2015). We address these issues by implementing the density-corrected dissimilarity index proposed in Allen et al. (2015). Simulations from their research shows that this method works better in correcting for bias than other commonly used approaches.

Allen et al. (2015) show that the absolute value term in D leads to upward bias as the

limiting distribution $\sqrt{n}(D - D_{pop})$ does not have mean zero. They propose the following adjusted estimator:

$$D^{\text{adj}} = \frac{1}{2}\sum_{i=1}^{K} \hat{\sigma}_i n(\hat{\theta}_i) \tag{3}$$

where

$$\hat{\sigma}_i = \frac{p_i^m(1-p_i^m)}{m_i} + \frac{p_i^n(1-p_i^n)}{n_i} \ ,$$

$\hat{\theta}_i = |p_i^m - p_i^n|/\hat{\sigma}_i$, and $n(\hat{\theta}_i)$ is the value of θ_i that maximizes

$$\phi(\hat{\theta}_i - \theta_i) + \phi(\hat{\theta}_i + \theta_i).$$

In Appendix A, we show the difference between the original and adjusted dissimilarity indices for each province in Turkey. We see that the gap is minimal (maximum is 0.4) and less than 0.01 for the 20 provinces that account for over 95 percent of the refugee stock. In other words, the original and adjusted indices lead to identical results in the empirical analysis in the rest of paper.

4.2 Isolation index

The isolation index I is defined as:

$$I \equiv \sum_{i=1}^{K}\left(\frac{m_i}{t_i}\frac{m_i}{M}\right) = \sum_{i=1}^{K}\left(p_i\frac{m_i}{M}\right) \tag{4}$$

It is a weighted average of the fraction p_i of the minority group in the population in i, where the weights are given by the share of the minority group residing in area i. While index D in (2) is *insensitive* to an identical proportional increase in the size of the minority group m_i across all areas, the isolation index $I \in [0,1]$ in (4) is not.[17] So, Massey and Denton (1988) modify the standard isolation index to limit its dependency on the share P of the minority group in the total population. This is the *adjusted* isolation index that we use in the rest of the paper:[18]

$$I^{\text{adj}} \equiv \frac{I - P}{1 - P} \tag{5}$$

[17] Simply notice that m_i/M in (4) remains unchanged, while p_i increases for all $i = 1, ..., K$, so that I unambiguously increases.

[18] We obtain similar results if we adopt the definition of the normalized isolation index by Cutler et al. (1999) and Cutler et al. (2008a), who replace in the denominator $1 - P$ with $\min\{1, M/(\min_i t_i)\} - P$.

If the individuals were matched at random, the isolation index measures the probability that members of the minority group interact with members of the majority group.

4.3 Concordance and divergence of the two indices

Our segregation indices D and I^{adj} range between 0 and 1. They attain their lowest or highest values in identical circumstances. Complete segregation corresponds to the cases when $D = I^{\text{adj}} = 1$ and all members of the minority group reside in areas where there are no members of the majority group. At the opposite extreme, $D = I^{\text{adj}} = 0$ when all areas host the same ratio of the minority group to the total population so that $p_i = P$ for all $i = 1, ..., K$.[19]

The two indices diverge, and convey different information about the segregation of the minority group for intermediate values of the indices. The dissimilarity index D in (1) is insensitive to a redistribution of the the members of the minority groups across two areas that are both either above or below their share P in the total population, while this is not the case with the adjusted isolation index I^{adj} in (5). More precisely, such a redistribution of the minority group reduces its isolation if it is directed towards area i where the share p_i of the minority group is lower.

The Index D is insensitive to an homogeneous proportional increase in the size of the minority group m_i in each area.[20] The same property does not hold for the isolation index, even in its normalized version in (5): An homogeneous change in m_i unambiguously changes p_i, P and the adjusted isolation index I^{adj} in the same direction.

4.4 Challenges of the phone data

In its most disaggregated form, the Dataset 1 reveals the time-varying level of phone activity at the level of each tower, separately for native and refugee customers. We use the level of the phone activity as a proxy for the number of the native and refugee customers at that location at that point in time. The main challenge relates to the possibility that phone usage levels might not be representative of the underlying population (Pestre et al., 2020), notably

[19] In this case, we would also have $I = P$.

[20] Formally, the value of D is unaffected if we replace each m_i with λm_i, where $\lambda > 0$. This can be readily seen from the definition of the dissimilarity index D in (2), since $\lambda m_i / \lambda M = m_i / M$ for any $\lambda > 0$.

because of (*i*) different call propensities for native and refugee callers, and (*ii*) differences in the market share of Türk Telekom within the refugee population across provinces.

We describe each of these two points in greater detail and how we address them in the Appendix B. In the case of the first issue, the dissimilarity index is unchanged if the call propensities of the two groups are time-invariant even if they are different (please see the discussion of the index above). A more serious problem arises if the individual call propensities vary across geographic locations and time. Using a second dataset provided by D4R which follows a smaller set of individual users over time and space, we argue that this problem is not present by showing there is no change in their call volumes and behaviour when these individuals change locations. In order to address the second issue, we weight the province level segregation indices by the distribution of refugees published regularly by the Ministry of Interior based on the registration numbers.

4.5 Geographic partitions

The measurement of the dissimilarity index D and of the normalized isolation index I^{adj} requires us to define the geographic partitioning of the country and the relevant spatial aggregation of the data coming from the mobile phone activity of natives and refugees. The coarser level of aggregation and larger geographic areas (such as the province level partitioning of the country) will lead to a lower value of both indices.[21] We construct four different levels of spatial partitioning: (*i*) 82 provinces, (*ii*) 957 districts, (*iii*) urban cells of 0.05×0.05 degrees, and (*iv*) catchment areas of each one of our over 30,000 phone towers. Data on call volumes referring to a specific tower is assigned to the larger geographic unit (district or province) in which the tower is located, using information on the precise GPS coordinates of the tower provided by Türk Telekom (see Salah et al., forthcoming) and the borders of these geographic units. In other words, data for a given area correspond to the combined mobile phone activities of all towers located within that area.[22]

[21] Consider two different levels of aggregation, where the second comes from a further partitioning of the spatial units used at the first level of aggregation. The calculation of the segregation indices at the province level of aggregation implicitly rests on the assumption of a uniform spatial distribution of the two populations within that level, which by definition, leads to lower levels for the indices

[22] The alternative would have been to define a Voronoi tessellation of Turkey, identifying the portion of territory that is covered by each tower. As differences are likely to be immaterial, we have not pursued this more time-consuming approach. It would have required obtaining further data from Türk Telekom, related

The two segregation indices are computed either for the entire country, or for portions of the country that are of specific interest, as mentioned in Section 4 above, such as a large province like Istanbul. We can compute the two indices for large provinces (in terms of population) that host a sizable share of the refugee population. Alternatively, we can focus on urban areas, as cities host a disproportionate share of the Syrian refugees.

4.6 Segregation indices at the national level

Figure 3 plots the weekly evolution of the dissimilarity index D and the normalized isolation index I^{adj} for the whole country when the phone call volume data are measured at the tower level. In Appendix C, we present the indices calculated at different levels of partitioning of the country (at the district level and square grids with each side measuring 0.05 degrees of latitude and longitude). Although the levels of the indices change depending on the spatial partitioning, the overall patterns are almost identical for both indices.

The dissimilarity index exhibits minimal variation throughout the year. It starts at around 0.44 in the first weeks of 2017, moves within a narrow band and declines to around 0.42 by the end of the year. We observe several short-lived increases. The first one happens around week 16, possibly corresponding to a data recording anomaly as seen in Figure 1, when several days of data were mistakenly attributed to a single day in the database. The others small spikes happen during the summer and they are likely to be driven by the internal mobility of the native population during the holidays, away from big cities to smaller towns, thus increasing the diversity in the spatial distribution of the majority relative to he minority population. Interestingly enough, relatively high values of the dissimilarity index occur during the 26^{th} and 35^{th} weeks of 2017 which correspond to the major Islamic religious holidays of *Eid-al-Fitr* (the end of Ramadan) on June 27, 2017 and *Eid-al-Adha* on September 4, 2017. The spikes are more easily observable with the district level data in Appendix C.[23]

A slightly different pattern emerges with the normalized isolation index I^{adj}. In contrast to the dissimilarity index, the normalized isolation index increases over the year when measured at the tower level. It starts at around 0.15, increases to 0.2 by June and maintains

to the orientation of the various antennas placed on each tower.

[23]Similar short-term effects of Islamic religious holidays on mobility are found by Milusheva (2020) using CDR data for Senegal.

that level until the end of the year. This pattern might reflect the increase in the number of refugee mobile phone customers of Türk Telekom over time,[24] as I^{adj} is positively related to the share of the minority group in the total population (see Section 4.3 for a discussion).

The high level of temporal distribution of the CDR data allows us to capture a very detailed picture of the segregation indices resulting from almost daily shifts in the spatial distribution of the native and refugee populations. Such level of disaggregation would not have been possible with standard cross-sectional or annual datasets, such as censuses and household surveys, that are generally used in the literature to measure segregation indices.

4.7 Segregation within provinces

In this section, we calculate both indices for each province separately, again using the tower level partitioning of space. Our objective is to see if the level of segregation differs *between* provinces and how it changes over time. Figure 4 reports the evolution of the dissimilarity index D for four provinces: two large and wealthy provinces in the West (Istanbul and Izmir) and two provinces in the Southeastern part of the country on the Syrian border (Gaziantep and Hatay). The figure reveals important variations in the extent of segregation of the refugees depending on province characteristics and location within the country. The time profile of the dissimilarity index at the province level does *not* follow the same pattern observed in Figure 3 for the whole country. These differences suggest that variations in the spatial distribution of the native population over the summer mostly occurs across rather than within Turkish provinces.

Istanbul, the largest province in Turkey with a population over 15 million, is home to over a quarter of all Syrian refugees. Their distribution is relatively stable within Istanbul as presented in top left panel of Figure 4. The dissimilarity index starts at around 0.42 and gradually dips below 0.4, following closely the levels observed at the national level in Figure 3. Izmir, the third largest province in the country and the departure point for many refugees hoping to go to Europe via crossing the Aegean Sea to Greece, starts at a slightly higher level but follows the same path with more volatility, possibly due to fewer number of observations. In the provinces near the Syrian border, refugees constitute between 10 and 20 percent of the population in 2017 as opposed to the national average of 4.5 percent. The

[24] As discussed in Section 3 above, the preferential tariff for Syrian refugees has been introduced by Türk Telekom in April 2017.

dissimilarity index as seen in the bottom panels of Figure 3 is mostly lower in these cities when compared to Istanbul or Izmir, indicating a larger share of refugees leads to higher integration. The second important observation is that the index is quite stable over time, implying refugee population has achieved some sort of an equilibrium.

When we explore the evolution of the normalized isolation index for the same four provinces in Figure 5, we see comparable patterns. Istanbul has an index that gradually increases until the last quarter and then gently declines. Izmir is quite stable throughout the year, with a significant jump during the summer months between the two religious holidays. The levels in the Southeastern border provinces are much lower at the beginning of the year, but slowly increase in relative terms. Lower index levels, again indicate more integration of the refugees, and the increase might be due to the increase in their numbers or access to cell phones.

Figures 4 and 5 also plot the confidence intervals around the province level estimates in the dissimilarity and normalized isolation indices. Confidence intervals for both indices are calculated from draws of 100 bootstrap repetitions of individual calls. Variance estimates are then calculated from these bootstrapped values of each index and confidence intervals are calculated assuming the parameters are normally distributed. Empirical call probabilities for the random draws are taken from the data for a given time and geographic disaggregation.The narrow bands around the estimated index values indicate that our estimates are quite precise. The confidence interval for large cities, like Istanbul, are especially narrow due to larger number of observations.

In general, the within province trends we see mirror those at the national level. Dissimilarity indices are highly stable over time, while isolation indices increase in the middle of the year and maintain that level. Additionally, provinces in the southeast, along the border with Syria, exhibit the lowest levels of segregation, relative to other high refugee cities. This pattern reflects earlier assimilation among refugees, as these were the first provinces to be settled in at the start of the crisis.

For the other 77 provinces in Turkey, we perform two separate exercises. First, Figure 6 plots the average value of the dissimilarity index for the whole year for each province. Starting from the left, the provinces are ordered in declining share of Syrians in the total population of the province. The value of the index for the majority of the provinces lies between 0.4 and 0.6. The national average of the index is closer to 0.4 since over 90% of

the refugees live in 20 provinces which tend to have lower index values. We also plotted the confidence bands which are calculated in the same manner as described above. Even though we see variation in the width of the band, the overall precision is again quite high. The results for the Turkish provinces is comparable to the dissimilarity indices calculated for the racial composition of the largest 300 hundred metropolitan areas in United States using the 2010 Census data. The range of the index for the US is 0.25-0.55 with a median value of 0.34.[25]

As our second exercise, Figure 7 presents the average values of the dissimilarity and normalized isolation indices in the first and last quarters of the year. Most provinces with large refugee populations and low dissimilarity indices show minimal change, with Hatay as an outlier. (The red dotted line is the 45-degree line and the size of the bubbles are proportional to the refugee population in each province.) Other smaller provinces in terms of refugee populations experience slightly bigger declines over time, indicating faster integration in areas with low initial refugee levels. We observe a similar pattern with the adjusted isolation index, presented in the lower panel of Figure 7. Provinces with larger refugee populations have lower isolation indices and, as it was the case with the four provinces discussed above, they show some increase throughout the year.

5 Segregation and mobility decisions

The previous section defined, calculated and discussed several key observations on our two segregation measures – the dissimilarity and normalized isolation indices – at the national level and within provinces over time. This section explores how segregation impacts the mobility decisions of the refugees within the country after they cross the border. More specifically, we ask if the extent of segregation at a given location – a province or a district within a province – acts as a pull or a push factor for refugees and natives as they make their internal mobility decisions.

In terms of patterns, throughout the year, the dissimilarity index exhibits relative stability within each province (using the tower-level partitioning) as well as at the national level (using tower, grid or district level partitioning). There are some fluctuations within

[25]source: https://www.censusscope.org/us/rank_dissimilarity_white_twoplus.html accessed on 3/12/2021

the year, especially during the holidays in the summer. The normalized isolation index, in contrast, exhibits some changes at the national level; it increases until the third quarter and then tapers off. At the province level, however, the normalized isolation index is more stable with more gradual changes over the year. All of these observations indicate that segregation *within* each province is relatively stable. The national level changes can be attributed to internal mobility, especially from low-isolation to high-isolation provinces.

While escaping from violence is the main reason why refugees arrive at a destination country's borders, economic and cultural factors become more important as they choose a location within that country. Academic literature identified several key pull and push factors that influence these mobility decisions (World Bank, 2018). Among the economic factors are the availability of jobs and affordable housing. Ethnic and cultural diasporas also work as pull factors as they provide financial, social and informational support mechanisms for the recently arriving refugees.

The role of segregation at a location as a determinant of mobility decisions is, however, more ambiguous. Suppose a Syrian family decides to move from Hatay on the border and the choice is between two industrial cities in the western part of the country (like Istanbul and Izmir) where there are more jobs, higher incomes and better amenities. Also, suppose the share of Syrian refugees in the total population is the same in both cities so that the diaspora effect will be the same. In one city, however, the refugees are concentrated in certain neighborhoods while they are more evenly spread out in the second one. We should emphasize once again that segregation is about the *distribution* of the minority population within a region, not about their *share* in the overall population. The first city – with more segregation – might provide stronger support mechanisms as the refugees live physically closer to each other. However, concentration might also lead to congestion externalities, lack of integration or even animosity from the locals. The focus of this section is on the relative strength of these forces.

5.1 Inter-provincial mobility decisions

We measure inter-province mobility by using Dataset 3 (from D4R) which follows and records the phone call data for the same cohort of refugees and natives for a whole year. More specifically, we construct measures of group-specific dyadic flows for both natives and refugees, labeled f_{jk}^{natives} and f_{jk}^{refugees} respectively, between province j and province k. We do this

by assigning each user a daily (provincial residence) location based on where they made or received most of their calls during that day. We then take a rolling average of that daily assignment as the most common location in the 30 days surrounding a given date to take away short trips. We keep in our sample only the users for whom we could observe at least a 90-day period. We set the origin and destination as the first and last location observed for an individual user.

We observe a higher average move rate among refugees (14.0 percent) than the sample of natives in the database (6.7 percent), indicating a relatively high degree of internal mobility.[26] Figure 8 presents a color-coded map of the number refugees who moved out of a province (out-migration) as a share of the total number of refugees who moved during the year. Darker colors represents provinces that were the origin locations for a higher share of refugees. Figure 9 presents the parallel map for the destination provinces. We see from these maps that the most important source and destination provinces are the large industrial areas (such as Istanbul) in the West as well as the border regions in the Southeast since they host the largest number of refugees. Some provinces along the Mediterranean coast where refugees can find jobs in the labor-intensive tourism related sectors other locations in the central parts of the country on the transit route to Istanbul also receive some refugees. In contrast, there is very little mobility in the Eastern and Northern parts of the country.

Using these measures, we identify the group-specific correlates of inter-provincial movements via the estimation of a standard gravity equation.[27] We match this flow data with standard bilateral gravity variables (contiguity, geodesic distance), and with destination-specific variables (population, GDP per capita, share of refugees in the population in January 2017), and with the dissimilarity D and the isolation index I^{adj} computed for destination provinces.[28] We use a Poisson Pseudo Maximum Likelihood (PPML) estimator to account for the large number of inter-provincial corridors containing zero flows.[29] Tables 1 and 2

[26] We should note once again, the natives in the database are oversampled from the provinces with a larger share of refugees which possibly leads to higher mobility rate than the national average. According to Sirkeci and Cohen (2012) the average annual internal mobility rate in Turkey during 1990-2010 was 4.2 percent.

[27] Results, which are available from the Authors upon request, are fully robust if we measure inter-provincial mobility using Dataset 2.

[28] The effect of origin-specific variables is absorbed by the inclusion of origin dummies in all specifications.

[29] The estimation sample is restricted to the 30 origin provinces with the largest refugee population, but this sample selection criterion is immaterial given that origins with larger out-going flows disproportionately

report the results from the estimation of the gravity equation, for refugees and natives, respectively.

Internal migration flows are larger between contiguous provinces, as expected, and they decline with distance. Additionally, both natives and refugees are attracted to provinces that have larger total populations and higher income per capita. These results are consistent with other studies using gravity models of migration,[30] and with the evidence provided for Syrian refugees in Turkey by Beine et al. (2021). People simply prefer to move to larger, wealthier and nearby cities.

Flows for both natives and refugees are larger towards provinces that have a higher share of refugees in the population. This variable is capturing the positive externalities generated by the presence of large diaspora networks for the refugees. For the natives, however, it is possibly picking up the effect of some unobserved determinants of the attractiveness of a province that our parsimonious specification fails to account for. However, the comparison of Tables 1 and 2 reveals that the influence exerted by this variable is significantly stronger for refugees relative to the natives.

Our main variables of interest in this gravity estimation are the segregation indices. In Table 1, we see that the refugees do not want to move to provinces characterized by a high dissimilarity index (column 4). Once we control for the overall share of refugees in the population of a province, their distribution *within* that population also matters; refugees do *not* want to live in segregated regions, at least as measured by the dissimilarity index. Normalized isolation index has a negative coefficient but it is not significant (in column 5) for the internal mobility decisions of the refugees. The stronger tendency for refugees to move towards provinces that have a higher share of refugees – combined with higher mobility rates among refugees – implies that these inter-provincial movements might explain the increase in I^{adj} that we observe in Figure 3. In the case of the natives, segregation simply has no influence on their mobility decisions as seen in columns 4-5 of Table 2.

5.2 Intra-provincial mobility decisions

A large share of internal migration flows takes place *within* provinces, between districts or even neighborhoods of a city. Most refugees live in urban areas and they might decide to

contribute to the identification of the coefficients in a PPML estimation.

[30] See, for example Beine et al. (2016), Grogger and Hanson (2011), or Özden et al. (2017)

move to another area in search of a better job, apartment or a school. They might want to be near relatives for support or, similarly, to get away from other refugees to better assimilate within the host society.

We again use Dataset 3 to see if the location choices of refugees *within* a city are also influenced by segregation levels in different part of the cities. Our analysis is based on Istanbul to address these questions. With a population over 15 million people, Istanbul is slightly smaller than Syria and larger than many countries in Europe, providing the ideal environment for our analysis. The whole province is highly urbanized and consists of 41 districts with refugees living in all of them. Most other provinces in Turkey have large rural areas and only a handful of urban districts where we would find observable number of refugees.

In Figures 10 and 11, we present the color-coded maps of the districts of Istanbul according to the share of refugees and native they receive, respectively, among those who moved within Istanbul. Refugees are moving into the areas around the Golden Horn on the European side of the city. These are historical areas dating back to the Byzantine times with older housing stock, tourist attractions and wide range of small businesses. In contrast, a larger share of natives are moving into the suburban and residential areas on Asian side, with new and relatively modern housing stock.

While the maps in Figures 10 and 11 are already giving a glimpse of different mobility patterns of the refugees and natives within the districts of Istanbul, our main goal is to identify the factors that determine these intra-provincial mobility choices. We again use a gravity model where the dependent variable is the bilateral migration flows *between* the 41 districts of Istanbul. Since the populated area covers only 1000 square miles and any pair of districts are relatively close to each other, we drop the standard gravity variables of distance and contiguity. Instead, we include the economic pull and push factors along with our segregation measures and origin fixed effects. The economic and demographic variables we consider are population, income per capita, population density, income inequality and employment/population ratio (to capture whether this is a residential or business district). We include the (natural log) of the number of refugee calls to capture the size of the refugee population.

Estimation results for the migration of refugees is presented in Table 3. The first columns reports the baseline results without the segregation indices while the second and the third

column include the dissimilarity and the normalized isolation indices, respectively. Refugees prefer to move to high income and non-residential business (high employment/population ratio) districts and avoid those with high-income inequality. In contrast, population level or density are not significant. These results indicate the importance of economic concerns for the refugees, rather than congestion or other demographic variables. The population of refugees in a district, proxied by the (natural log of the) phone call volume is highly significant, as evidence of the importance of diaspora networks. The segregation indices, our main variables of interest, have positive coefficients but are not significant and they have no impact on the other variables. Slightly different set of factors influence the mobility decisions of natives as seen in Table 4. The only significant variables are the population size and the employment/population ratio. The segregation indices have, again, large but not significant coefficients. In short, the main factors that influence the intra-province (or short range) mobility of the refugees are the economic variables and diaspora forces; the extent of segregation at the destination is not a significant factors.

5.3 Segregation by time of day

The richness of the time dimension of the D4R data allows us to explore how refugee segregation varies across different times of the day. When constructing the dissimilarity and isolation indices in the previous sections, we use weekly averages of the call volumes at the tower level. As mentioned earlier, the data are reported at the hourly intervals which allows us to calculate the distribution and segregation of refugees on an hourly basis. For this purpose, we aggregate call volumes for a specific hour of the day and tower for the entire year, separately for weekdays and weekends.

Figure 12 presents how the dissimilarity index D changes across the country during the day. We see that D exhibits a u-shaped pattern with a dip during the day. It reaches its minimum of 0.4 around 4 pm and maximum of 0.55 early in the morning and late at night. The normalized isolation index in Figure 13 shows a comparable pattern, again, with a clear dip during the middle end of the day, but an even sharper increase at night. The gaps between the minimum and the maximum values are larger than any other gap observed for the national and provincial level data for either index.

The differences in the indices during the day and at night give us clues about the extent of the residential (night time) and labor market (day time) segregation of refugees. Results

indicate segregation outside of typical work hours—which we interpret as capturing more residential segregation—is higher than during work hours in the week, when we would expect refugees and natives are more likely to be at their place of employment. The large gap between the residential and labor market segregation indices might also explain why we do not see any impact of the segregation indices on intra-provincial mobility patterns for Istanbul in the previous section. When people live and work in different areas, there might be less incentive to move.

6 Conclusion

Highly disaggregated digital data provide unique opportunities to approach important academic questions from different angles. In this paper, we use telephone usage data to identify the extent and patterns of segregation of Syrian refugees in Turkey. The data allow us to identify the volume of calls made/received by refugees and natives at the tower level for each hour in 2017. Using the phone call volume as a proxy for the population distribution, we construct a range of dissimilarity and normalized isolation indices. The richness of the data allow us to calculate the indices over time, across different provinces, even during different hours of the day.

Our first set of results regarding segregation is that the dissimilarity index is relatively stable over time both at the national and provincial levels while the isolation index shows gradual increase, possibly due to increased customer base of Turk Telekom and cell phone penetration. We see variation across provinces for both indices, with significantly lower levels of dissimilarity and isolation in provinces with higher share of refugees. In other words, if a geographic area has a small number of refugees, there seems to be more segregation.

The second set of results link segregation measures to mobility patterns, based on gravity estimations. In the case of inter-province analysis, we find that refugees choose to move to areas with high number of refugees but lower levels of segregation while natives are not influenced by it. In intra-provincal analysis based on data from Istanbul, refugee share of a district is a pull factor but segregation has no impact on refugees. Finally, we see differences in our segregation indices across the hours of the day, implying residential segregation is higher than labor market segregation and potentially explaining why segregation has no impact on intra-provincial mobility decisions.

We believe our analysis provides interesting results but only scratches the surface in terms of the opportunities provided by such data. Other interesting and more difficult questions could be answered with greater access to data, such as the market shares of the phone operators at the provincial level, as this would allow us to match CDR-based data to the underlying population and correcting for possible selection biases. The next step is represented by the analysis of the correlation between the measures of segregation and the economic outcomes of the Syrian refugees, as well as for their effects on the native population.

References

Allen, R., S. Burgess, R. Davidson, and F. Windmeijer (2015): "More reliable inference for the dissimilarity index of segregation," *The econometrics journal*, 18, 40–66.

Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, and J. Weinstein (2018): "Improving refugee integration through data-driven algorithmic assignment," *Science*, 359, 325–329.

Beine, M., L. Bertinelli, R. Cömertpay, A. Litina, J.-F. Maystadt, and B. Zou (2019): "Refugee Mobility: Evidence from Phone Data in Turkey," in *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study*, ed. by A. A. Salah, A. Pentland, B. Lepri, and E. Letouzé, Springer, 433–449.

Beine, M., L. Bertinelli, R. Cömertpay, A. Litina, and J.-F. Maystadt (2021): "A gravity analysis of refugee mobility using mobile phone data," *Journal of Development Economics*, 150, article 102618.

Beine, M., S. Bertoli, and J. Fernández-Huertas Moraga (2016): "A practitioners' guide to gravity models of international migration," *The World Economy*, 39, 496–512.

Bertoli, S. and J. Fernández-Huertas Moraga (2015): "The size of the cliff at the border," *Regional Science and Urban Economics*, 51, 1–6.

Bertoli, S. and E. Murard (2020): "Migration and co-residence choices: Evidence from Mexico," *Journal of Development Economics*, article 102330.

Blumenstock, J., G. Cadamuro, and R. On (2015): "Predicting poverty and wealth from mobile phone metadata," *Science*, 350, 1073–1076.

Blumenstock, J. and L. Fratamico (2013): "Social and spatial ethnic segregation: a framework for analyzing segregation with large-scale spatial network data," in *Proceedings of the 4th Annual Symposium on Computing for Development*, ACM, 11.

Borjas, G. J. (1992): "Ethnic capital and intergenerational mobility," *The Quarterly Journal of Economics*, 107, 123–150.

——— (1995): "Ethnicity, Neighborhoods, and Human-Capital Externalities," *The American Economic Review*, 85, 365–390.

Boy, J., D. Pastor-Escuredo, D. Macguire, and M. Moreno Jimenez, R.and Luengo-Oroz (2019): "Towards an Understanding of Refugee Segregation, Isolation, Homophily and Ultimately Integration in Turkey Using Call Detail Records," in *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study*, ed. by A. A. Salah, A. Pentland, B. Lepri, and E. Letouzé, Springer, 141–164.

Bygren, M. and R. Szulkin (2010): "Ethnic environment during childhood and the educational attainment of immigrant children in Sweden," *Social Forces*, 88, 1305–1329.

Catanzarite, L. (2000): "Brown-collar jobs: Occupational segregation and earnings of recent-immigrant Latinos," *Sociological Perspectives*, 43, 45–75.

Cavlin, A. (2020): *Syrian Refugees in Turkey: A Demographic Profile and Linked Social Challenges*, Routledge.

Cesare, N., H. Lee, T. McCormick, E. Spiro, and E. Zagheni (2018): "Promises and Pitfalls of Using Digital Traces for Demographic Research," *Demography*, 55, 1979–1999.

Clark, W. A. and S. A. Blue (2004): "Race, class, and segregation patterns in US immigrant gateway cities," *Urban Affairs Review*, 39, 667–688.

Cortese, C. F., R. F. Falk, and J. K. Cohen (1976): "Further considerations on the methodological analysis of segregation indices," *American Sociological Review*, 630–637.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107, 455–506.

——— (2008a): "Is the melting pot still hot? Explaining the resurgence of immigrant segregation," *The Review of Economics and Statistics*, 90, 478–497.

——— (2008b): "When are ghettos bad? Lessons from immigrant segregation in the United States," *Journal of Urban Economics*, 63, 759–774.

Duncan, O. D. and B. Duncan (1955): "A methodological analysis of segregation indexes," *American Sociological Review*, 20, 210–217.

Goldsmith, P. R. (2009): "Schools or neighborhoods or both? Race and ethnic segregation and educational attainment," *Social Forces*, 87, 1913–1941.

Grogger, J. and G. H. Hanson (2011): "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 95, 42–57.

Guimaraes, P., O. Figueirdo, and D. Woodward (2003): "A tractable approach to the firm location decision problem," *Review of Economics and Statistics*, 85, 201–204.

Hall, M. (2013): "Residential integration on the new frontier: Immigrant segregation in established and new destinations," *Demography*, 50, 1873–1896.

Hamilton, E. R. and R. Savinar (2015): "Two sources of error in data on migration from Mexico to the United States in Mexican household-based surveys," *Demography*, 52, 1345–1355.

Hanson, G. H. (2006): "Illegal migration from Mexico to the United States," *Journal of Economic Literature*, 44, 869–924.

Hu, W., R. He, J. Cao, L. Zhang, H. Uzunalioglu, A. Akyamac, and C. Phadke (2019): "Quantified Understanding of Syrian Refugee Integration in Turkey," in *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study*, ed. by A. A. Salah, A. Pentland, B. Lepri, and E. Letouzé, Springer, 201–223.

Ibarraran, P. and D. Lubotsky (2007): "Mexican immigration and self-selection: New evidence from the 2000 Mexican census," in *Mexican immigration to the United States*, ed. by G. J. Borjas, University of Chicago Press, 159–192.

Iceland, J. (2004): "Beyond black and white: metropolitan residential segregation in multi-ethnic America," *Social Science Research*, 33, 248–271.

Iceland, J. and K. A. Nelson (2008): "Hispanic segregation in metropolitan America: Exploring the multiple forms of spatial assimilation," *American Sociological Review*, 73, 741–765.

Iceland, J. and M. Scopilliti (2008): "Immigrant residential segregation in US metropolitan areas, 1990–2000," *Demography*, 45, 79–94.

Jahn, J., C. F. Schmid, and C. Schrag (1947): "The measurement of ecological segregation," *American Sociological Review*, 12, 293–303.

Järv, O., K. Müürisepp, R. Ahas, B. Derudder, and F. Witlox (2015): "Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia," *Urban Studies*, 52, 2680–2698.

Leonard, J. S. (1987): "The interaction of residential segregation and employment discrimination," *Journal of Urban Economics*, 21, 323–346.

Logan, J. R., W. Zhang, and R. D. Alba (2002): "Immigrant enclaves and ethnic communities in New York and Los Angeles," *American Sociological Review*, 299–322.

Marquez, N., K. Garimella, O. Toomet, I. Weber, and E. Zagheni (2019): "Segregation and Sentiment: Estimating Refugee Segregation and Its Effects Using Digital Trace Data," in *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study*, ed. by A. A. Salah, A. Pentland, B. Lepri, and E. Letouzé, Springer, 265–282.

Massey, D. S. and N. A. Denton (1988): "The dimensions of residential segregation," *Social Forces*, 67, 281–315.

——— (1993): *American apartheid: Segregation and the making of the underclass*, Harvard University Press.

Mayer, S. E. (2002): "How economic segregation affects children's educational attainment," *Social Forces*, 81, 153–176.

Mazza, A. and A. Punzo (2015): "On the upward bias of the dissimilarity index and its corrections," *Sociological Methods & Research*, 44, 80–107.

Milusheva, S. (2020): "Predicting Dynamic Patterns of Short-Term Movement," *The World Bank Economic Review*, 34, S26–S34.

Özden, Ç., M. Packard, and M. Wagner (2017): "International migration and wages," *Revue d'economie du developpement*, 25, 93–133.

Parsons, C. and P.-L. Vézina (2018): "Migrant networks and trade: The Vietnamese boat people as a natural experiment," *The Economic Journal*, 128, F210–F234.

Patel, K. and F. Vella (2013): "Immigrant networks and their implications for occupational choice and wages," *Review of Economics and Statistics*, 95, 1249–1277.

Pestre, G., E. Letouzé, and E. Zagheni (2020): "The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates," *The World Bank Economic Review*, 34, S89–S97.

Salah, A. A., A. Pentland, B. Lepri, E. Letouzé, Y.-A. de Montjoye, X. Dong, Ö. Dagdelen, and P. Vinck (2019): "Introduction to the Data for Refugees Challenge on Mobility of Syrian Refugees in Turkey," in *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study*, ed. by A. A. Salah, A. Pentland, B. Lepri, and E. Letouzé, Springer, 3–27.

Silm, S. and R. Ahas (2014): "The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset," *Social Science Research*, 47, 30–43.

Sirkeci, I. and J. H. Cohen (2012): "Internal mobility of the foreign born in Turkey," in *Minority Internal Migration in Europe*, ed. by N. Finney and G. Catney, Routledge, 132–152.

Sjaastad, L. A. (1962): "The costs and returns of human migration," *Journal of Political Economy*, 70, 80–93.

Sterly, H., B. Etzold, L. Wirkus, P. Sakdapolrak, J. Schewe, C.-F. Schleussner, and B. Hennig (2019): "Assessing Refugees' Onward Mobility with Mobile Phone Data—A Case Study of (Syrian) Refugees in Turkey," in *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study*, ed. by A. A. Salah, A. Pentland, B. Lepri, and E. Letouzé, Springer, 201–223.

Taeuber, K. E. and A. F. Taeuber (1976): "A practitioner's perspective on the index of dissimilarity," *American Sociological Review*, 41, 884–889.

Trounstine, J. (2016): "Segregation and inequality in public goods," *American Journal of Political Science*, 60, 709–725.

Turner, M. A. (2008): "Residential Segregation and Employment Inequality," in *Segregation: The rising costs for America*, ed. by J. H. Carr and N. K. Kutty, Routledge, 151–196.

Williams, D. R. and C. Collins (2001): "Racial residential segregation: a fundamental cause of racial disparities in health," *Public Health Reports*, 116, 404.

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

Tables and Figures

Table 1: Gravity estimation within Turkey for refugees

	Dependent variable: f_{jk}^{refugees}				
	(1)	(2)	(3)	(4)	(5)
Contiguity$_{jk}$	0.428	0.447	0.422	0.381	0.417
	(0.154)***	(0.157)***	(0.152)***	(0.151)**	(0.156)***
Ln(Distance$_{jk}$)	-0.295	-0.401	-0.275	-0.266	-0.275
	(0.064)***	(0.092)***	(0.069)***	(0.067)***	(0.069)***
Ln(Population$_k$)		1.164	1.065	1.042	1.062
		(0.054)***	(0.044)***	(0.045)***	(0.043)***
Ln(GDP/capita$_k$)		0.066	0.644	0.582	0.639
		(0.115)	(0.112)***	(0.110)***	(0.115)***
Refugee share$_k$			7.905	7.719	7.949
			(0.365)***	(0.373)***	(0.409)***
Dissimilarity index$_k$				-2.187	
				(0.536)***	
Isolation index$_k$					-0.214
					(0.701)
Origin FE	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No
Pseudo-R^2	0.66	0.58	0.63	0.64	0.63
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400
Aggregate sample	15,497	15,497	15,497	15,497	15,497

Notes: results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province refugee migration flows calculated from the D4R Dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as of Jan. 12 2017. Segregation indices are calculated using call records from the first quarter of 2017 using D4R Dataset 1. Distance variables are calculated as the geodesic distance (in Km) from the geographic centers of each province.

Table 2: Gravity estimation within Turkey for natives

	Dependent variable: f_{jk}^{natives}				
	(1)	(2)	(3)	(4)	(5)
Contiguity$_{jk}$	0.672	0.635	0.623	0.624	0.615
	(0.144)***	(0.138)***	(0.138)***	(0.139)***	(0.141)***
Ln(Distance$_{jk}$)	-0.368	-0.307	-0.292	-0.289	-0.290
	(0.073)***	(0.077)***	(0.076)***	(0.076)***	(0.078)***
Ln(Population$_k$)		0.956	0.920	0.921	0.916
		(0.040)***	(0.040)***	(0.040)***	(0.041)***
Ln(GDP/capita$_k$)		0.418	0.544	0.527	0.513
		(0.095)***	(0.099)***	(0.106)***	(0.104)***
Refugee share$_k$			2.083	1.963	2.032
			(0.731)***	(0.828)**	(0.787)***
Dissimilarity index$_k$				-0.238	
				(0.647)	
Isolation index$_k$					-0.839
					(0.626)
Origin FE	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No
Pseudo-R^2	0.59	0.58	0.58	0.58	0.58
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400
Aggregate sample	13,142	13,142	13,142	13,142	13,142

Notes: results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province native migration flows calculated from the D4R Dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as of Jan. 12 2017. Segregation indices are calculated using call records from the first quarter of 2017 using D4R Dataset 1. Distance variables are calculated as the geodesic distance (in Km) from the geographic centers of each province.

Table 3: Gravity estimation within Istanbul for refugees

	(1)	(2)	(3)
Ln(Population$_k$)	0.156	0.256	0.250
	(0.320)	(0.277)	(0.305)
Ln(Income$_k$)	0.331	0.308	0.322
	(0.162)**	(0.165)**	(0.163)**
Employment/resident ratio$_k$	0.148	0.170	0.161
	(0.078)*	(0.086)**	(0.084)**
Population Density$_k$	0.068	0.077	0.071
	(0.150)	(0.153)	(0.153)
Income Inequality$_k$	-0.318	-0.270	-0.296
	(0.143)**	(0.139)*	(0.142)**
Ln(refugee calls$_k$)	0.641	0.595	0.568
	(0.107)***	(0.120)***	(0.154)***
Dissimilarity index$_k$		1.643	
		(1.460)	
Isolation index$_k$			1.792
			(2.525)
Origin FE	Yes	Yes	Yes
Dest FE	Yes	No	No
Origin-destination pairs	1,444	1,444	1,444

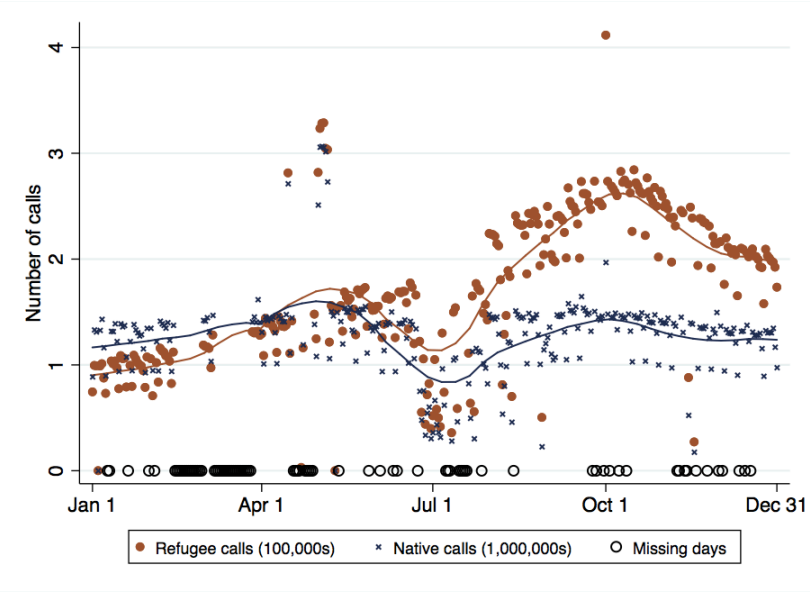
Notes: results from Poisson pseudo-maximum likelihood (PPML) regressions of refugee mobility across districts of Istanbul; refugee migration flows calculated from the D4R Dataset 3; Segregation indices are calculated using call records from the first quarter of 2017 using D4R Dataset 1.

Table 4: Gravity estimation within Istanbul for natives

	(1)	(2)	(3)
Ln(Population$_k$)	1.178	1.247	1.382
	(0.284)***	(0.308)***	(0.329)***
Ln(Income$_k$)	0.013	-0.039	-0.017
	(0.191)	(0.190)	(0.195)
Worker/resident ratio$_k$	0.219	0.244	0.253
	(0.089)**	(0.098)**	(0.098)***
Population Density$_k$	-0.052	-0.028	-0.039
	(0.091)	(0.093)	(0.093)
Income Inequality$_k$	0.097	0.147	0.134
	(0.163)	(0.162)	(0.162)
Ln(refugee calls$_k$)	-0.180	-0.188	-0.307
	(0.199)	(0.195)	(0.221)
Dissimilarity index$_k$		1.859	
		(1.327)	
Isolation index$_k$			4.158
			(2.286)
Origin FE	Yes	Yes	Yes
Dest FE	Yes	No	No
Origin-destination pairs	1,444	1,444	1,444

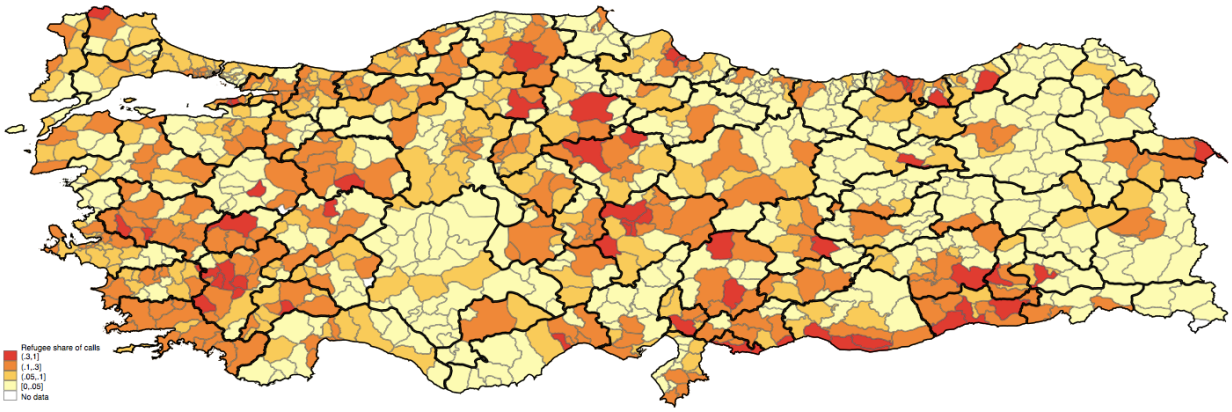
Notes: results from Poisson pseudo-maximum likelihood (PPML) regressions of native flows across districts of Istanbul; native migration flows calculated from the D4R Dataset 3; segregation indices are calculated using call records from the first quarter of 2017 using D4R Dataset 1.

Figure 1: Summary of call volume and missing data



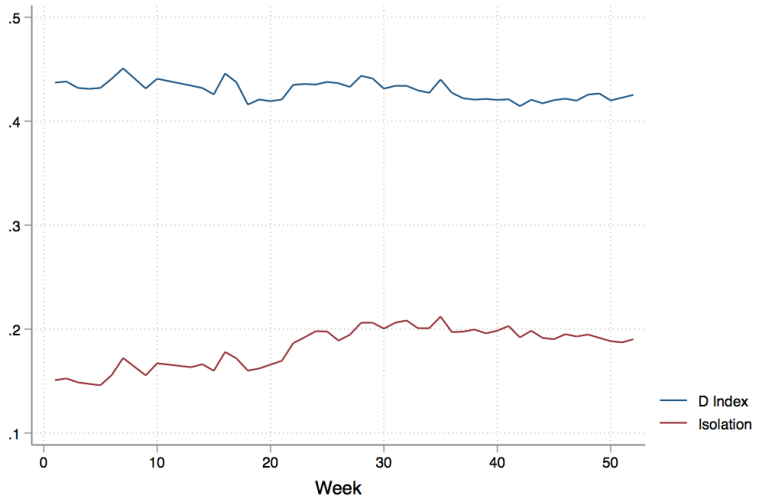
Notes: total call volume by day and refugee status. Black circles indicate days in which no data is available.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom.

Figure 2: Refugee share of calls by district



Notes: Refugee share of total calls by district in all of 2017; thick black lines indicate province borders.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom.

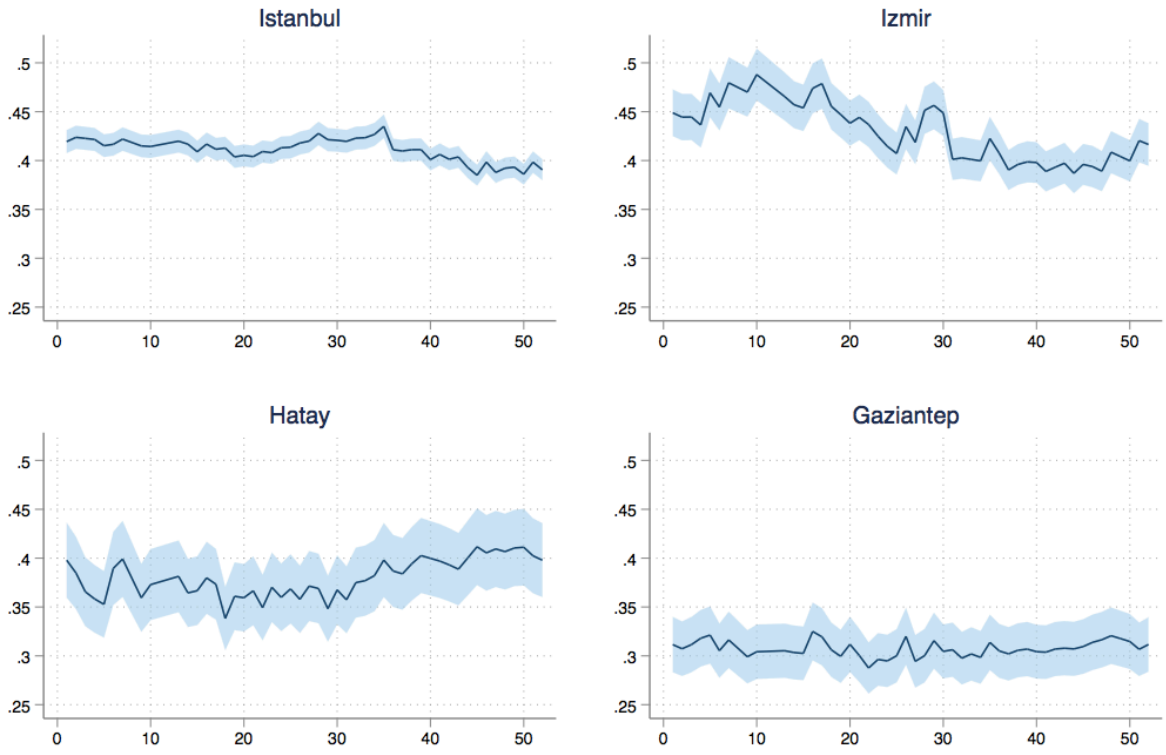
Figure 3: Evolution of the dissimilarity and normalized isolation indices



Notes: Evolution of the indices D and I^{adj} measured at the weekly level from January to December 2017 using information on call volumes by natives and refugees at the antenna level of aggregation. We average weekly province-level indices using as weights the share of the total refugee population in each province in January 2017.

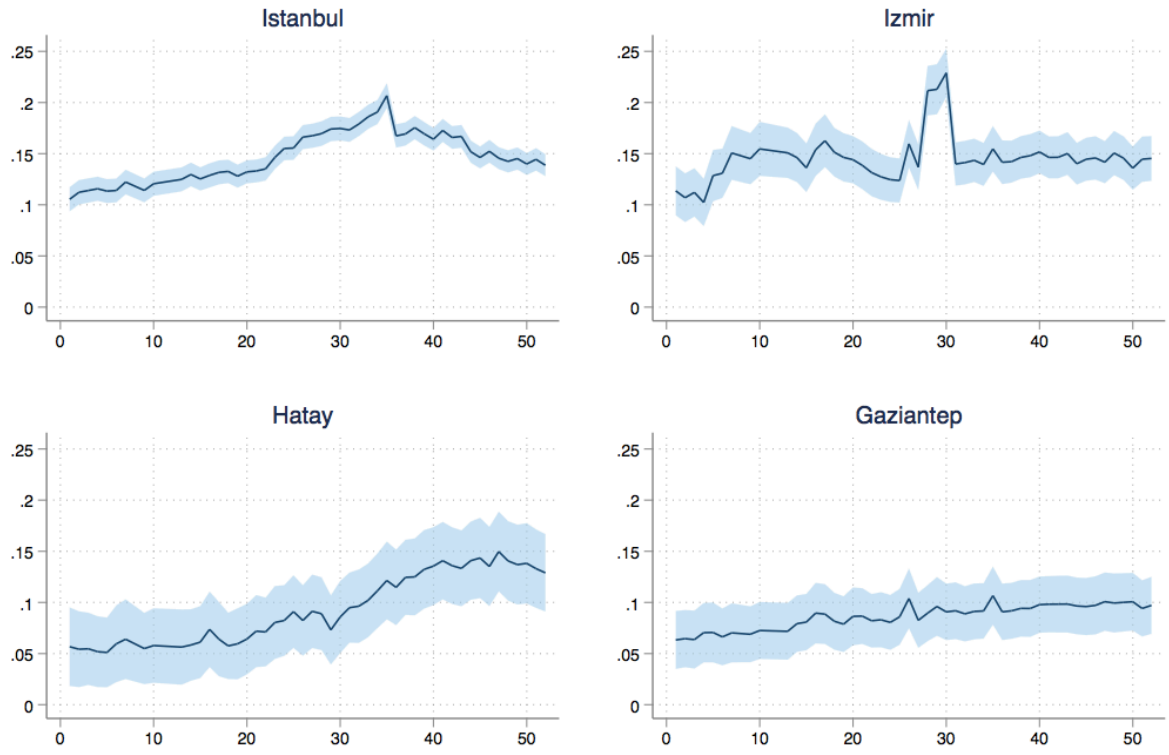
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interior.

Figure 4: Evolution of the dissimilarity index by province



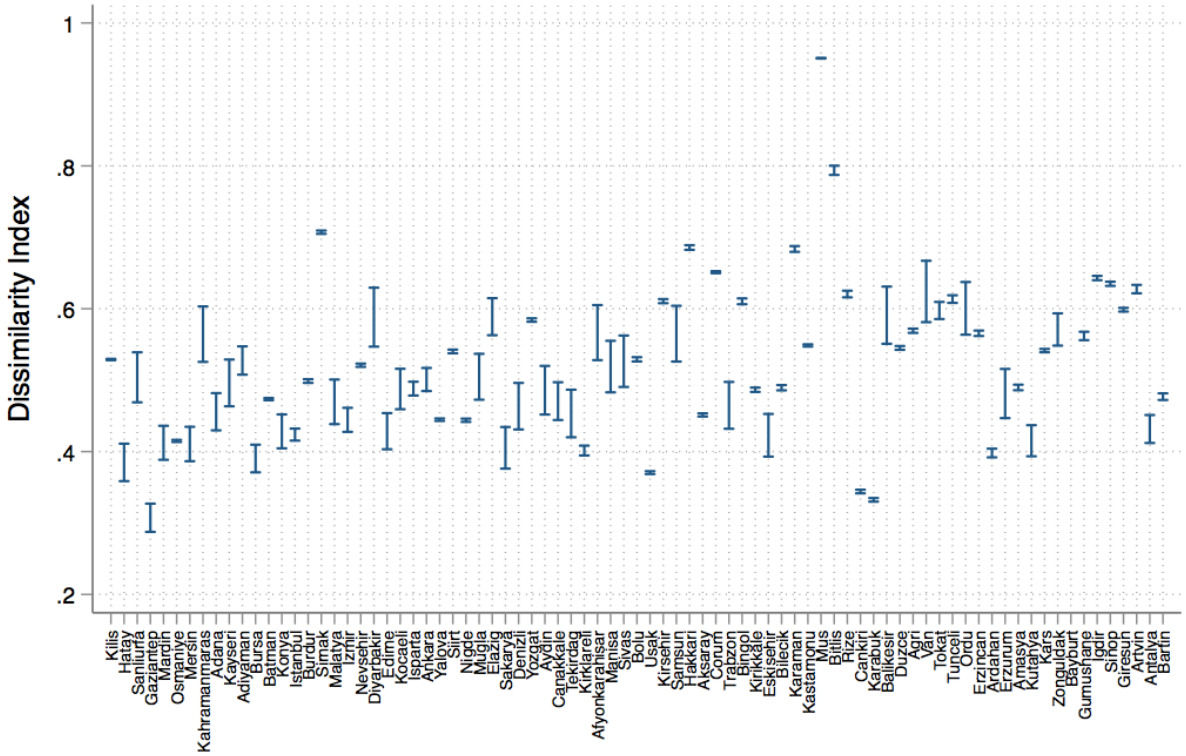
Notes: evolution of the dissimilarity isolation index D (with 95 percent confidence intervals) measured at the weekly level from January to December 2017 using information on call volumes by natives and refugees at the level of towers in four provinces; we average weekly province-level dissimilarity indices using as weights the share of the total refugee population in each province in January 2017.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Figure 5: Evolution of the isolation index by province



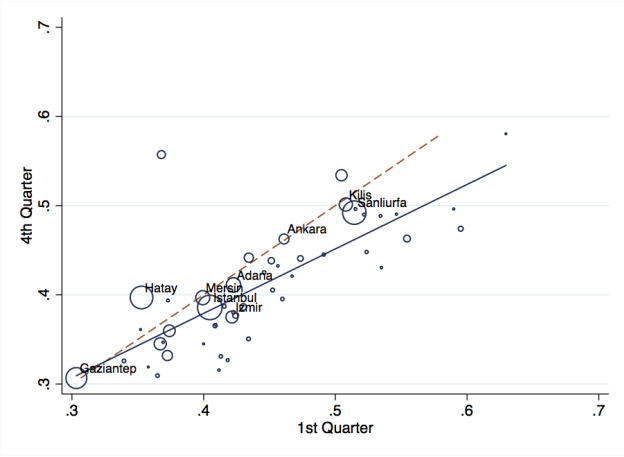
Notes: evolution of the normalized isolation index I^{adj} (with 95 percent confidence intervals) measured at the weekly level from January to December 2017 using information on call volumes by natives and refugees at the level of towers in four provinces; we average weekly province-level dissimilarity indices using as weights the share of the total refugee population in each province in January 2017.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Figure 6: Dissimilarity index by province with confidence intervals

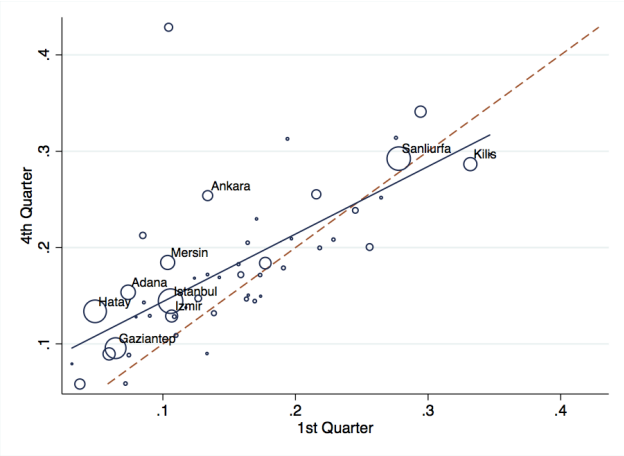


Notes: The dissimilarity index D (with 95 percent confidence intervals) measured for the whole year using information on call volumes by natives and refugees at the level of towers in all provinces separately.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Figure 7: Dissimilarity and isolation indices in 2017Q1 and 2017Q4



(a) Dissimilarity index D

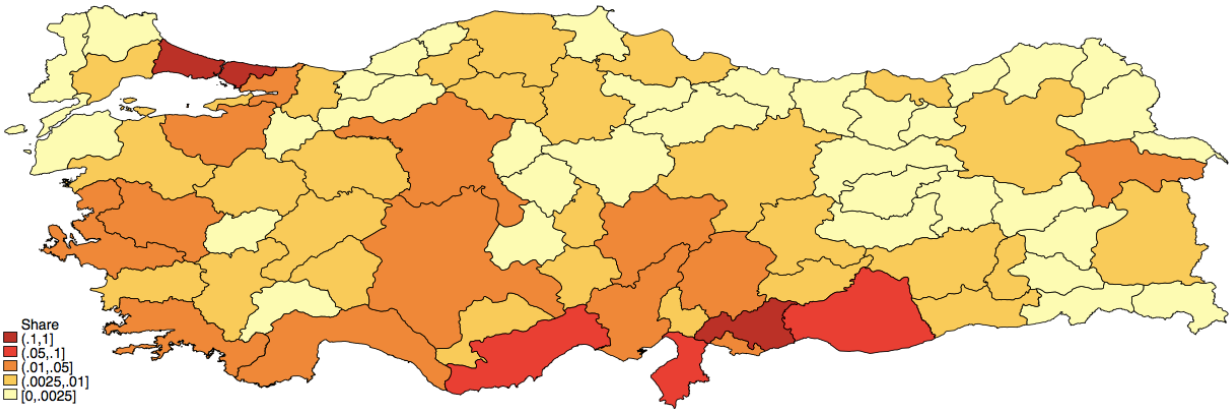


(b) Normalized isolation index I^{adj}

Notes: comparison of the province-level dissimilarity and isolation index in the first (2017Q1) and in the fourth (2017Q4) quarter of 2017 using information on call volumes by natives and refugees at the level of towers; the size of each bubble is proportional to the number of refugees in January 2017 in each province; the red dashed line is the 45 degrees line, while the blue line reflects a (weighted) regression of either of the two indices in 2017Q4 on its value in 2017Q1.

Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

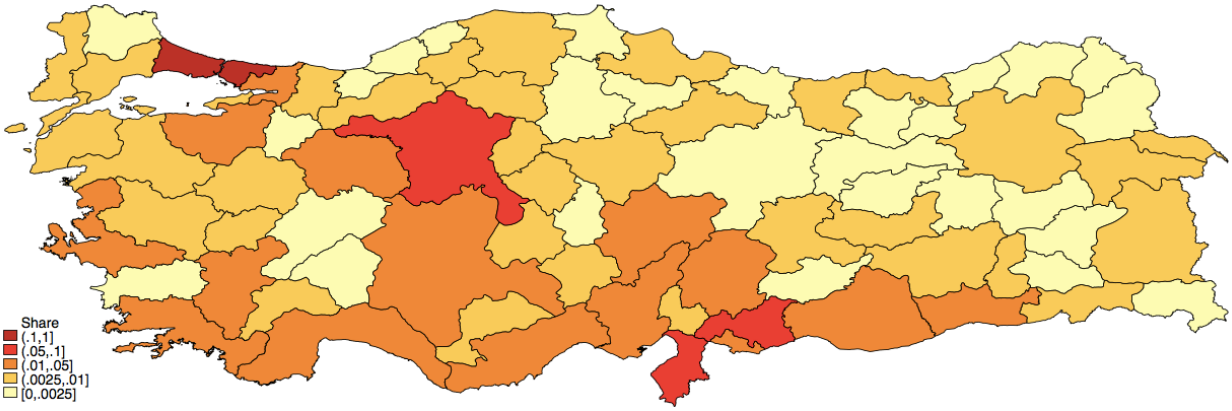
Figure 8: Origin provinces of refugees who moved during the year



Notes: Share of each province for outgoing flows of refugees.

Source: Authors' elaboration on D4R Dataset 3 by Türk Telekom.

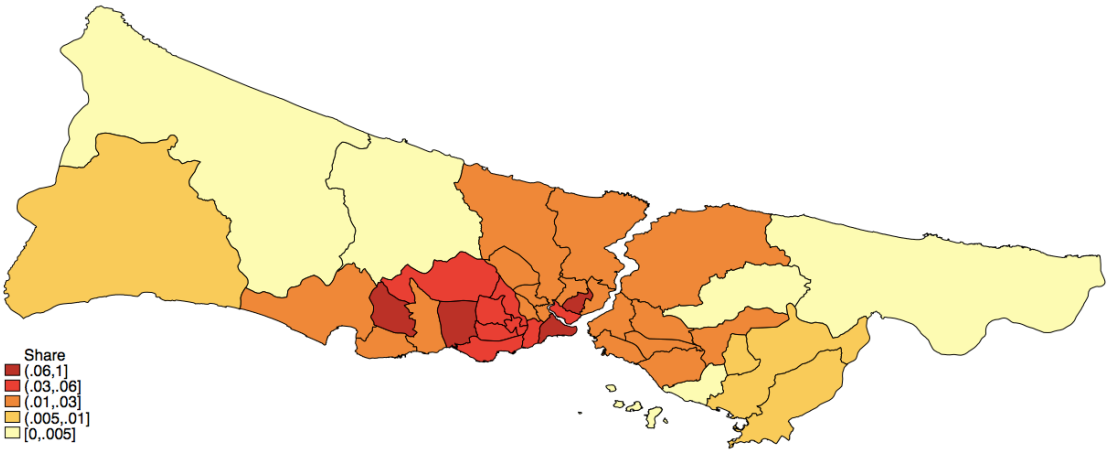
Figure 9: Destination provinces of refugees who moved during the year



Notes: Share of each province for incoming refugee flows.

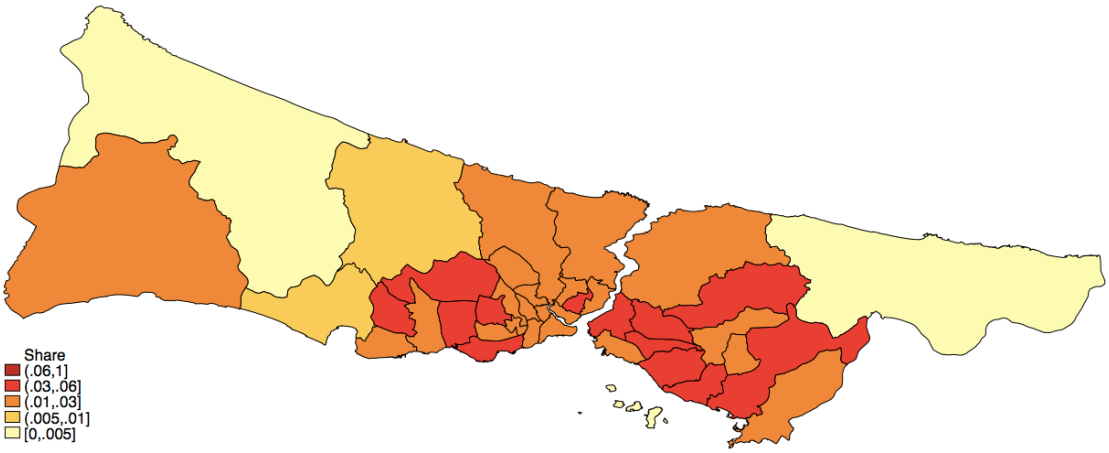
Source: Authors' elaboration on D4R Dataset 3 by Türk Telekom.

Figure 10: Destination districts of refugees who moved within Istanbul



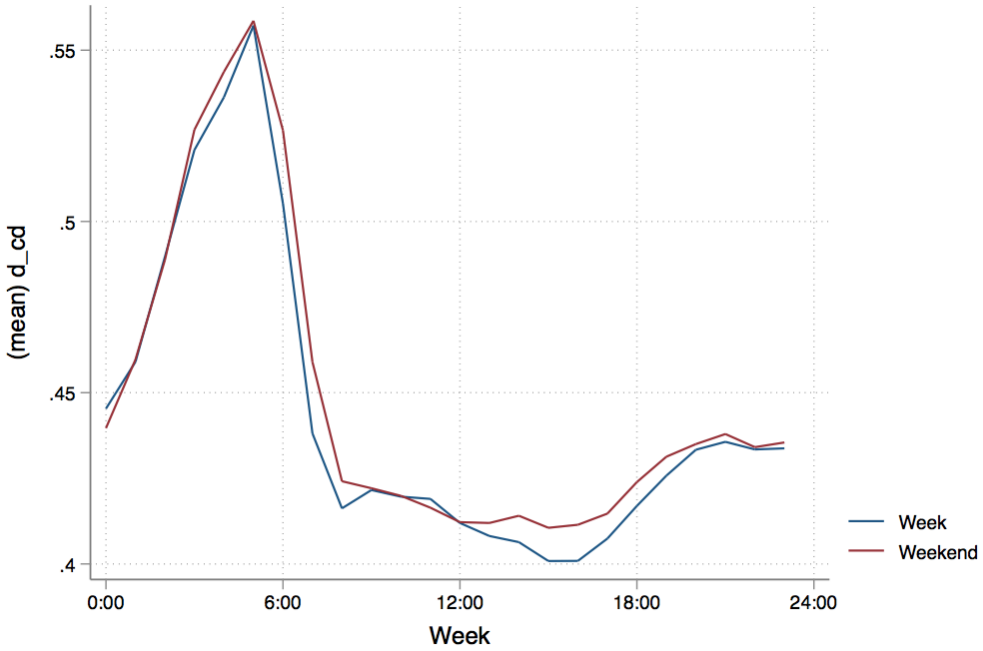
Notes: Share of each destination district for refugees who moved between districts in Istanbul.
Source: Authors' elaboration on D4R Dataset 3 by Türk Telekom.

Figure 11: Destination districts of natives who moved within Istanbul



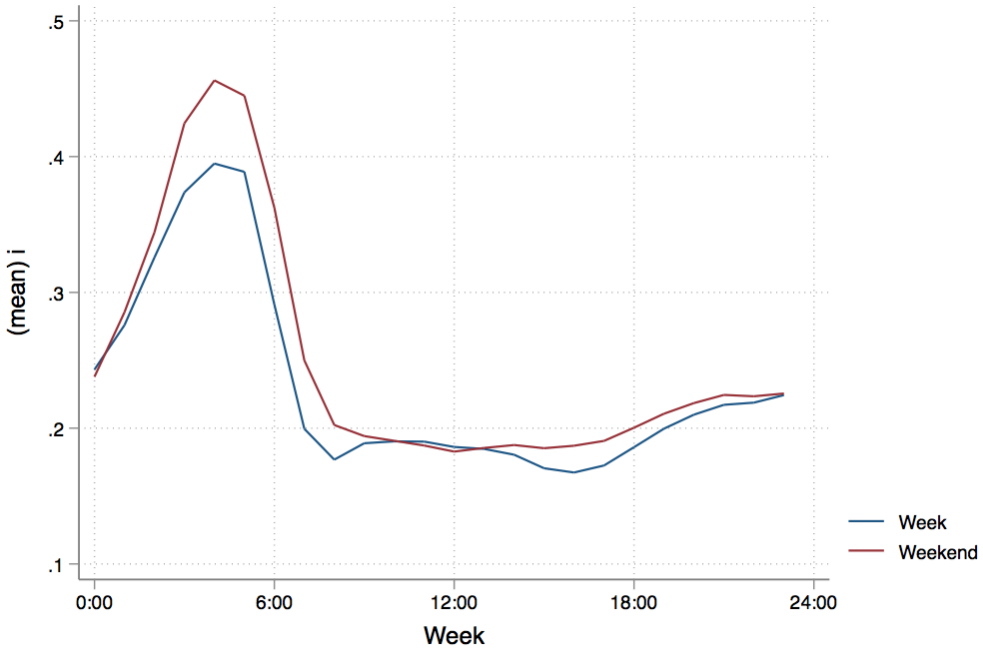
Notes: Share of each destination district for natives who moved between districts in Istanbul.
Source: Authors' elaboration on D4R Dataset 3 by Türk Telekom.

Figure 12: Dissimilarity index by hour of day



Notes: evolution of the dissimilarity index D measured at the hourly level aggregated across all of 2017; we average hourly province-level dissimilarity indices using as weights the share of the total refugee population in each province in January 2017.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Figure 13: Isolation index by hour of day



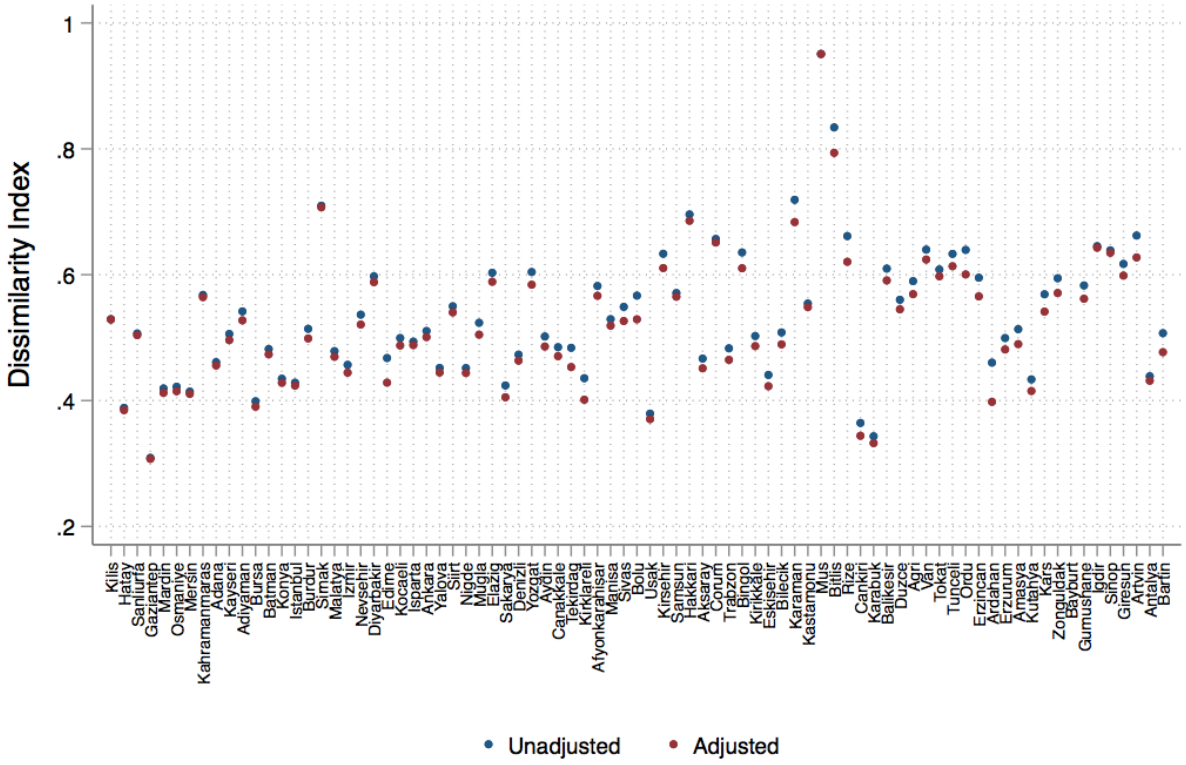
Notes: evolution of the normalized isolation index I^{adj} measured at the hourly level aggregated across all of 2017; we average hourly province-level isolation indices using as weights the share of the total refugee population in each province in January 2017.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Appendix

A Adjusted dissimilarity index

As discussed in Section 4.1, the standard dissimilarity index might have upward bias if the population of minority is small in certain sub-regions or the populations of some sub-regions are small relative to the overall population. We implement the correction proposed by Allen et al. (2015). Figure 14 below presents these adjusted dissimilarity indices for all of the provinces. We see that the gap between the original and adjusted values of the index increase as the share of the Syrians in the population decline. The largest gap is, however, only 0.04, confirming that our estimates of the indices are relatively precise and giving us confidence for the next stage of our analysis.

Figure 14: Adjusted Dissimilarity index for all provinces



Notes: The dissimilarity index D measured for the whole year using information on call volumes by natives and refugees at the level of towers in all provinces separately.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

B Data challenges

The dataset provides information on the call volume during each hour at each tower for each type of customer (hour-tower-customer type triplet), and not directly on the number of each of the two types of customers who are present on an hourly basis in the catchment area covered by each tower.[31] This data will give us an unbiased measure of the share p_i of

[31] Analyzing call volumes rather than the number of (native and refugee) phone lines connected to each tower directly avoids the problems arising from a single customer having multiple lines (see Salah et al., forthcoming).

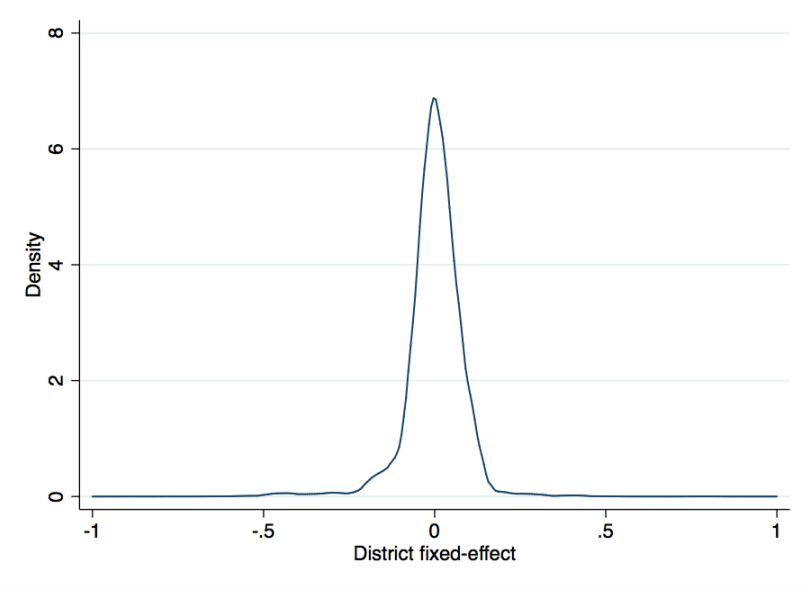
the minority (refugee) group in each area i only if the propensity of the two groups to make and receive calls coincide. Time-invariant differences on call propensities between the two groups do not pose problems for the computation of the dissimilarity index D, because (as discussed in Section 4.3), the index is invariant to a proportional increase in the size of the minority group m_i in each area. Differences in propensities to make phone calls are identical to a proportional change (increase or decrease) in the size of the refugee population across regions. However, a higher propensity to make and receive phone calls by the minority group results in a higher measure of the isolation index I^{adj}, as it would inflate p_i.[32]

A problem may arise, though, if individual call propensities vary geographically, for example, across provinces. That is, if mobile-phone users are more likely to make more calls in Istanbul than Ankara, in which case, differences in call volume may not necessarily indicate differences in population. As a check, we use another dataset provided by Türk Telekom to see if individual level call propensities vary across locations. Dataset 2 follows a small subset of individual Türk Telekom users over time, identifying the time and location (at the district level) of each call. Using this data, we calculate daily call volume of each individual and assign users to a district based on where they make the most number of calls in a given day. We then regress (the natural logarithm of) this call volume on date and district fixed-effects. Next, we calculate the residuals after taking out any province-level effects from the estimated district-level fixed-effects. The distribution of these district fixed-effects, weighted by individual-day observations, is centered around zero with a small standard deviation as present in Figure 15 below. 90 percent of our individual observations are located in districts within 10 percent (or 10 log points) of the average district (which has a fixed effect of zero). Additionally, districts further away from this average district tend to be those with much smaller samples, indicating that at least some of the observed variation is driven by small sample sizes. This analysis shows that the call volume is a convincing proxy for population distribution over geographic areas.[33]

[32] We should note the resulting dependency is lower with I^{adj} than with the non-normalized version I of the index.

[33] The results of this exercise are available upon request.

Figure 15: District level differences in log call propensities



Notes: Figure shows a weighted kernel density plot of district-level fixed-effects from a regression of individual by day log call rates, residualizing out any province-constant effects. Plot is weighted by the number of individual caller-by-day observations at the district level.
Source: Authors' elaboration on D4R Dataset 2 by Türk Telekom.

The second problem arises if the overall market share of Türk Telekom and its penetration among the refugee population show large variation across the 82 Turkish provinces. An analysis of the Datasets 1, 2 and 3 shows that the share of refugees across provinces does not perfectly align with the distribution of refugees based on official statistics. This is because uptake of Türk Telekom service is not identical across provinces. In order to explore this issue further, we combine our dataset with the official statistics on the number of Syrian refugees at the province level published periodically by the Turkish Ministry of Interior. When calculating the segregation indices at the national level, we weight the province-level segregation indices by the distribution of refugees across provinces, as reported by the government.

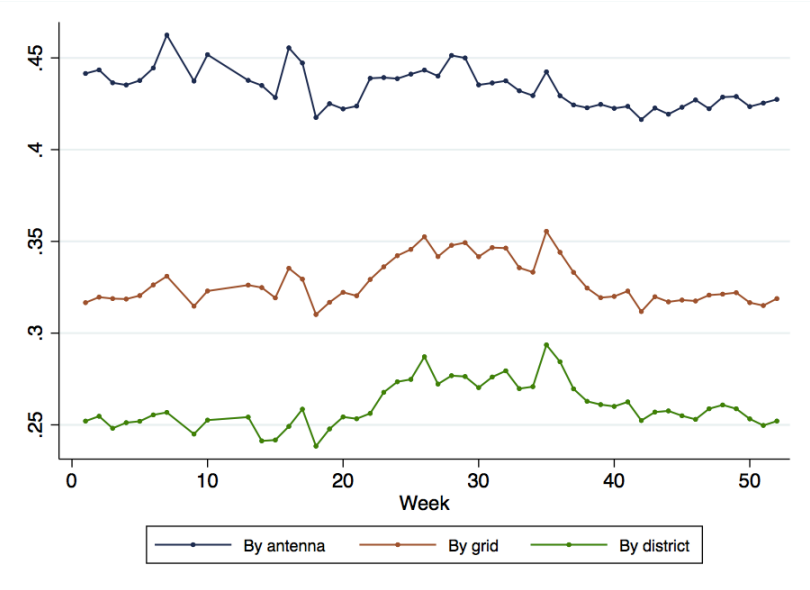
C Different levels of geographic partitioning

In Section 4.5 we discuss how different partitioning of the country leads to different values for the segregation indices. The first two figures below present the adjusted dissimilarity and normalized isolation indices calculated at different levels of partitioning - cell tower, district and grid levels. Turkey is split into 976 (NUTS3 level) districts with varying areas and populations. The square grid is formed by partitioning the country into 6,820 equal-sized geographic squares with each side measuring 0.05 degrees of latitude and longitude. Moving from districts to grids and, then, to the (basic) tower level results in an increase in the value of D from an average of around 0.25 to 0.32, and, then, to 0.43. Similarly, the value of normalized isolation index I^{adj} increases by around 0.05 as we move from districts to grid to the tower level partitioning. For both indices, however, the overall patterns are maintained.

The last Figure reports the evolution of the dissimilarity and the normalized isolation indices limiting only to towers located in urban areas. The trends of the two indices are similar to those observed for the entire country. We notice a slight reduction in dissimilarity and a similarly small increase in isolation indices. The main difference with respect to Figure 3 is that the seasonal effects or variations due to religious holidays are much weaker when we focus only on urban areas,[34] suggesting that internal mobility of the native population predominantly occurs out of urban areas, rather than across them.

[34] There is a minor isolated peak in I^{adj} in urban areas in correspondence to the *Eid-al-Adha* in the week of September 4, 2017.

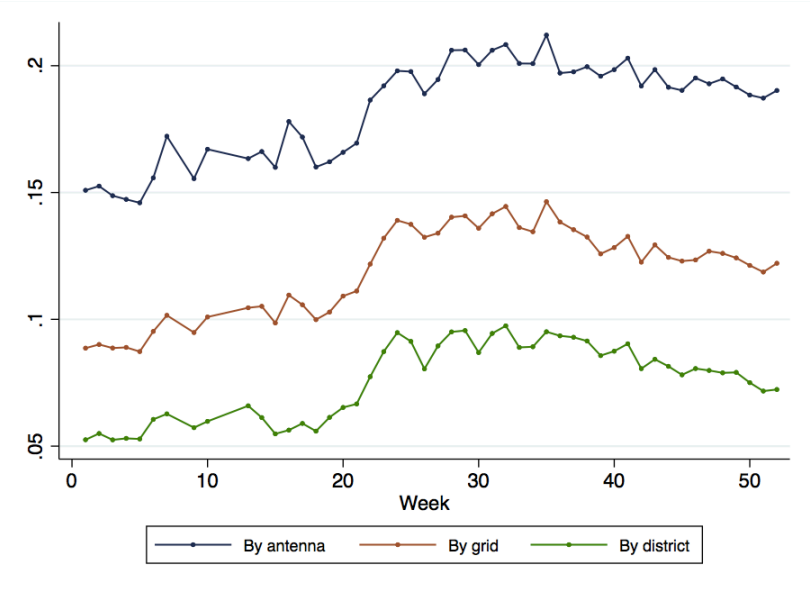
Figure 16: Evolution of the dissimilarity index



Notes: evolution of the dissimilarity index D measured at the weekly level from January to December 2017 using information on call volumes by natives and refugees at different levels of aggregation (antenna, grid and district); we average weekly province-level dissimilarity indices using as weights the share of the total refugee population in each province in January 2017.

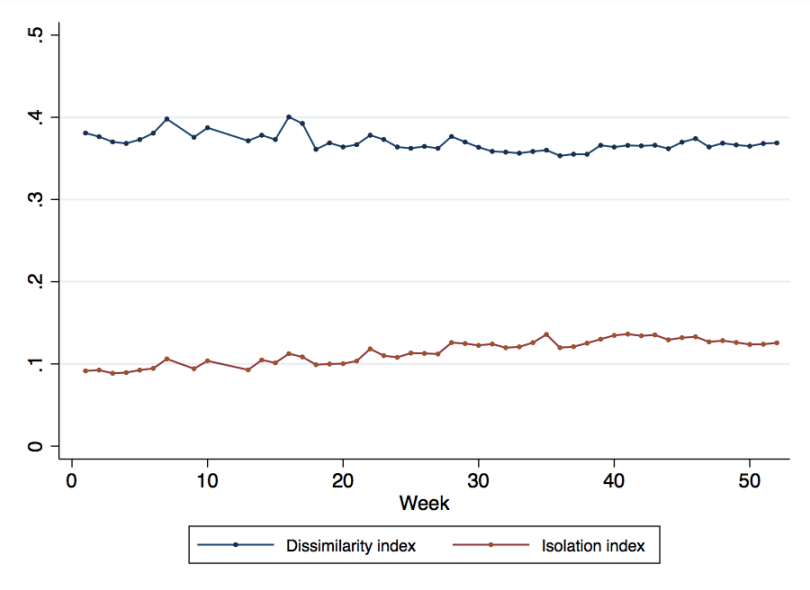
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Figure 17: Evolution of the isolation index



Notes: evolution of the normalized isolation index I^{adj} measured at the weekly level from January to December 2017 using information on call volumes by natives and refugees at different levels of aggregation (antenna, grid and district); we average weekly province-level normalized isolation index using as weights the share of the total refugee population in each province in January 2017.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.

Figure 18: Evolution of the dissimilarity and isolation indices in urban areas



Notes: evolution of the dissimilarity and of the normalized isolation index in urban areas measured at the weekly level from January to December 2017 using information on call volumes by natives and refugees in urban areas; a tower is assigned to urban areas if there are at least 15 other towers within a 5 km radius; weights are calculated as province share of total refugee population in January, 2017.
Source: Authors' elaboration on D4R Dataset 1 by Türk Telekom and data on refugees registered in each Turkish province in January 2017 from the Ministry of Interiors.