Caro, Luis Pinedo

Article

Identifying Syrian refugees in Turkish microdata

IZA Journal of Development and Migration

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Caro, Luis Pinedo (2020) : Identifying Syrian refugees in Turkish microdata, IZA Journal of Development and Migration, ISSN 2520-1786, Sciendo, Warsaw, Vol. 11, Iss. 1, pp. 1-21, https://doi.org/10.2478/izajodm-2020-0017

This Version is available at:
https://hdl.handle.net/10419/298688

*Luis Pinedo-Caro**

Identifying Syrian refugees in Turkish microdata

Abstract

This article proposes a strategy to identify Syrian refugees in Turkey's Household Labour Force Survey (HLFS). Even though Turkey's HLFS contains information on the migrants' year of arrival to Turkey, it does not provide details on their nationalities. This unfortunate feature mixes Syrian refugees with the standard flow of migration who arrived to Turkey during the Syrian war. I propose to eliminate the standard flow of migrants arrived between 2011 and 2017 by matching them (based on their characteristics) with the migrants arrived in the 2004–2010 period. This method obtains, indirectly, nonstandard migration, i.e., Syrian refugees. The results show that the age distribution of the nonstandard migrants identified matches the age distribution of Syrian refugees as officially released by the Turkish government. At last, I propose a post-stratification adjustment of the survey weights to find the actual geographical distribution of Syrian refugees in Turkey.

Current version: July 20, 2020
Keywords: Syrian refugees, Turkish microdata, identification strategy
JEL codes: C-81, F-22, J-11
Corresponding author: Luis Pinedo-Caro
pinedo@ilo.org

ILO Office for Turkey, Ferit Recai Ertuğrul Caddesi No.4, 06450 Oran, Ankara. E-mail: pinedo@ilo.org

Cite as: Pinedo-Caro. *IZA Journal of Development and Migration* (2020) 11:17.
https://doi.org/10.2478/izajodm-2020-0017

1 Introduction

The succession of revolts that followed the Arab Spring was typically characterized by short-term demonstrations and/or outbursts of violence in most of the affected countries; all but one, Syria. Since March 2011 until present, none of the multiple belligerents fighting in Syria have been able to regain full control of the country, causing, according to UNHCR, more than 5.68 million[1] of registered refugees out of which 3.6 million[2] were welcomed by Turkey under the temporary protection regime.[3] For Turkey, in particular, this unprecedented situation has not only produced an humanitarian emergency but also likely affected the lives of millions of Turkish people. In this context, the demand for policy responses is pressing and so does the demand of relevant information. This report aims precisely at filling an information gap by means of a strategy that would allow using the primary source for labor market statistics of Turkey, the Household Labour Force Survey (HFLS), for the creation of statistics on Syrian refugees.

The use of microdata when informing about the Syrian refugee crisis has been scarce. Some research has been conducted using macroeconomic data with regards to the Syrians' regional presence. For example, Konun and Tümen (2016) and Tümen (2016) studied the effect of Syrian refugees' arrival on the price level of goods, a finding that the goods whose production process intensely employs informal workers showcased a decline in their prices. This would be explained by Syrian workers replacing Turkish natives in informal jobs at a cheaper rate, passing the lower labor costs onto the goods' prices. In addition, Tümen (2016) also found that natives have both lower chances of finding an informal job and higher chances of finding a formal one. The latter might be due to the increase in the provision of public services caused by the arrival of the refugees. Another article analyzing the impact of Syrian refugees is Del Carpio and Wagner (2015), this time by combining microdata from the Turkish Labour Force Survey with macro data on the number of refugees by region. These authors, in addition to finding a large displacement of Turkish natives from the informal sector due to the arrival of the refugee population, also found a net displacement of women and the low educated away from the labor market.

Despite some successful attempts at producing studies on the impact of Syrian refugees at the macroeconomic level, little is known about their personal circumstances. One of the most remarkable attempts from a sociological point of view is the Syrian Barometer (see Erdoğan 2017), a national level survey covering 11 provinces and interviewing 1,235 Syrian families, reaching out, in total, 7,591 Syrians. Even though attractive in terms of understanding Turkish nationals' sentiment with regards to the Syrian population, it lacks, beyond a few basic questions, relevant information with regards to Syrians' labor market performance.

Other ad-hoc surveys focused on Syrian refugees' socioeconomic conditions are not as ambitious and the few existing sources lack national representativeness. Still, a remarkable effort in gathering data at the microeconomic level can be found in Uçak and Raman (2017). This research uses a survey on Syrian-owned SMEs to provide a snapshot of this type of companies, including the value of having them for the Turkish economy. With regards to the dataset,

1 According to https://data2.unhcr.org/en/situations/syria accessed as of April 4, 2019.

2 Data from the Directorate General of Migration Management (DGMM), updated as of April 4, 2019.

3 See https://help.unhcr.org/turkey/information-for-syrians/temporary-protection-in-turkey/ for more information on this regime.

which can be taken as a small-scale enterprise survey, it included visits to 230 businesses equally split between Istanbul and Gaziantep on the condition that they were legally established, are currently active, and had at least one employee. On the negative side, this database is, as confirmed by its authors, not meant to be nationally representative. Data collection efforts can also show glimpses of creativity, as in Kaymaz and Kadboy (2016), where the authors make use of a survey carried out on migration routes to find that around 30% of Syrian refugees have university degrees. Even though the extent to which Syrian refugees have such high qualifications may have been exaggerated due to the survey mode, it brings to the spotlight the importance of developing a model for the recognition of refugees' prior learning.

Lack of data may affect the depth and relevance of research on Syrian refugees; for example, Yavçan (2017) tried to illustrate the challenges faced by Turkey regarding Syrian refugees resorting to a small survey done by UNCHR in some Greek islands. Another example is Cagaptay (2014), in which an attempt to gauge the impact of Syrian refugees on the ethnic and sectarian balance of south-eastern provinces has to rely on data from the 1960 Census because it was the last one that collected data on ethnicity. The lack of nationally representative data on Syrian refugees in Turkey is in contrast to the availability found in Lebanon, where at least two such surveys have been carried out (see Alsharabati and Nammour 2016 or BRIC 2013), or in Jordan, where Syrian refugees can be identified within the Labour Force Survey.[4]

1.1 Household Labour Force Survey

The fact that the refugee population in Turkey represents 4.4% of the population living in Turkey[5] creates a growing need for nationally representative data on Syrian refugees in Turkey which is not currently fulfilled. Fortunately, the relatively high proportion of Syrian refugees in the Turkish economy might have opened the door as well to the use of nationally representative microdata from the Turkish Statistical Institute. However, using Turkstat databases for analyzing Syrian refugees is not straightforward. Household surveys in Turkey usually target families that are inscribed in the Address Based Population Registration System (ABPRS) and Syrian refugees under temporary protection are not included in that registry.[6] An exception to this survey methodology is given by the HFLS[7] that instead of families targets addresses, thus allowing interviewers to find Syrian families under certain conditions.

At present, even though Syrian refugees take part of the HLFS, their identification is not direct; the HLFS publicly available microdata does not provide the nationality of those classified as foreign-born, thus mixing up Syrians with the standard flow of migrants coming to Turkey (see Appendix A for a quick visual inspection of how this flow looks like). In this article, I propose an indirect identification method, whereby removing the standard migrants of the 2011–2017 period allows me to find nonstandard migration as a leftover. This method is meant to identify all nonstandard migrants who came into Turkey between 2011 and 2017. In practice, this group contains all Syrian refugees who migrated during that period, including those covered by the temporary protection regime, those with short-term residence permits, and

4 Syrian refugees are underrepresented in the Labour Force Survey; however, their survey weights have been adjusted to add-up to their total population.

5 As reported by the DGMM at http://www.goc.gov.tr/icerik6/temporary-protection 915 1024 4748 icerik

6 Their information is kept separately by the DGMM.

7 See http://tuik.gov.tr/MicroVeri/Hia 2017/turkce/index.html for more information.

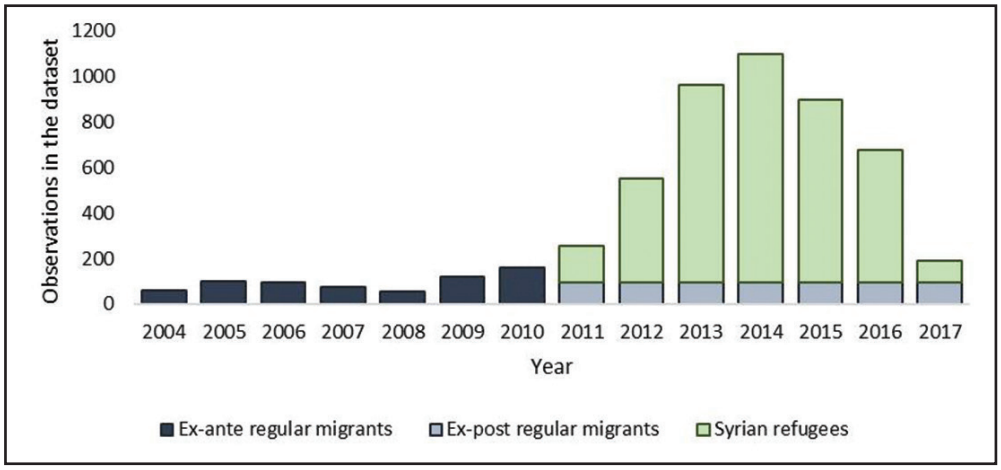
those who acquired the Turkish nationality. It should be noted that some other migrants (particularly those coming from Iraq or Afghanistan in recent times) may have also been included in the group. Still, throughout the report I refer to the group as a whole as "Syrian refugees" because Syrian refugees constituted 89.7% of nonstandard migration in 2017.[8]

In what follows, Section 2 explains the matching strategy I used to isolate Syrian refugees. Section 3 presents a post-stratification adjustment that calculates new survey weights for Syrian refugees. This section also uses these newly created weights to estimate the geographical distribution of Syrian refugees in Turkey including a comparison with the official distribution. Finally, Section 4 concludes.

2 An identification strategy for Syrian refugees in the HLFS

In the 2017 Household Labour Force Survey (HLFS) the number of foreign-born individuals who arrived between 2011 and 2017 is six times larger than the number of migrants who did so between 2004 and 2010. Unfortunately, the publicly available microdata of the HLFS do not contain information on the country of origin, and even though I suspect that Syrian refugees make up for the majority of observations among those who migrated between 2011 and 2017, they are unlikely to be the only foreigners who entered Turkey since the onset of the Syrian crisis. This hypothesis is supported by Figure 1, which shows the existence of a relatively constant number of foreign-born individuals arriving to Turkey during the years preceding the Syrian war (2004–2010). As a result, Syrian refugees are probably mixed up in the data with the hereinafter called "standard" migrants, thus preventing a direct identification of Syrian refugees.

Figure 1 Number of foreign-born individuals in Turkey by the year of arrival, 2004–2017.



Sources: Household Labour Force Survey 2017 and author's own calculations.

Notes: The figure shows the number of foreign-born observations living in Turkey. "Syrian refugees" are obtained by subtracting the average number of foreign-born individuals during the 2004–2010 period (the so-called ex-ante standard migrants) from the total number of observations in each of the years between 2011 and 2017 (ex-post migrants).

8 Figures obtained from publicly available data in the website of the Directorate General for Migration Management, Ministry of Interior of Turkey.

2.1 Assumptions

To identify the Syrian refugees present in the sample, I pursue an indirect identification strategy. Instead of finding Syrians among the 2011–2017 migrants, I find those who are not and then remove them from the sample (see Figure 1 for a visual explanation of the idea) with Syrian refugees who are identified as a "leftover" of the procedure. For this strategy to work, I assume that there is a relatively constant flow of what I call "standard" migrants. In particular, I assume that during the 2011–2017 period there was, on top of Syrian refugees, as many migrants as there were during the 2004–2010 period. This assumption is based on the findings of Korfalı and Acar (2018); their chapter shows how the flow of migrants from Central and Eastern Europe (which constitutes 40% of the total migration) to Turkey remained unaffected after the Syrian refugees started entering into Turkey. In practice, this assumption provides the number of observations that need to be removed from the ex-post migrants' group, i.e., those arrived between 2011 and 2017.

In addition, ex-ante and ex-post "standard" migrants, some of them thought to be Turkish-German by Bel-Air (2016), are assumed to share similar socioeconomic characteristics that are (1) observable in the microdata and (2) significantly different from those of Syrian refugees. This allows for the separation of "standard" migrants from Syrian refugees in the ex-post migrants' group. If, for instance, the ex-ante and the ex-post migrants' groups were identical, the matching would be trivial and refugees would not be identified, i.e., I would be removing ex-post migrants at random which does not help more than no matching at all. The comparability of ex-ante and ex-post migrants is tested (see Table 1) by comparing mean values of variables where, in principle, I would expect Syrian refugees and "standard" migrants to differ. It should be noted that for the sake of relevance, the comparison is done at the family level. This is because

Table 1 Summary statistics at the family level: before matching

Variable	Migrant families (2004–2010)	Migrants families (2011–2017)	Ratio (ex-post/ex-ante)
Family size	1.68	3.34	2.00
Proportion of 0–14	0.09	0.20	2.23
Proportion of 15–24	0.16	0.24	1.50
Proportion of 15+ women	0.75	0.63	0.83
Existence of a widow	0.03	0.08	2.37
Existence tertiary educ.	0.41	0.23	0.55
Existence of 15–24 students	0.15	0.10	0.65
Existence of 15+ female workers	0.27	0.13	0.50
Proportion of 15+ NEETs	0.50	0.59	1.18
Existence of workers	0.42	0.55	1.31
Number of informal workers	0.18	0.63	3.50
Existence of male garment workers	0.02	0.11	6.61
Existence of non-migrants	0.57	0.27	0.47

Sources: Household Labour Force Survey 2017 and author's own calculations.

Notes: The table shows averages at the family level for a number of variables. The wording "existence" refers to the existence of at least one person with the mentioned characteristic in the household. In all cases, it can be rejected that the difference in means is equal to zero at the 95% confidence level.

I match families—as opposed to individuals—so as to keep within-household coherence. Moreover, only individuals arrived during the prescribed period are included as part of the family, i.e., to minimize the noise due to mixing[9] with the local population and/or other migrants. It stems from Table 1 that sizable differences exist between the families who arrived to Turkey between 2004 and 2010 and between 2011 and 2017. For example, family size doubles among ex-post migrants, as it roughly does the proportion of children aged 0–14, proving the arrival of a much younger population in recent times. In turn, the higher number of households with a widow suggests the existence of families that may have escaped from a war. Moreover, I observe significantly less individuals with tertiary education, less female workers, and more NEETs among ex-post migrants, which hints the existence of strong differences at the cultural and at the socioeconomic level between the two groups under analysis. I also find that 65% of the workers (the informality rate in Turkey is 35.0%) coming from ex-post migrant families are not registered with the social security institute. This fact fits well with the existence of an underlying population of Syrian refugees because it is known that they had received very few work permits at the time of the survey.[10] In sum, based on the observed differences, it is reasonable to argue that ex-post migrants constitute a different group that in turn supports the use of matching.

2.2 **Matching**

The matching of ex-post "standard" migrant families with ex-ante "standard" migrant families uses nearest neighbor propensity score without replacement. In practice, this translates into the calculation of a probability (propensity score) of being an ex-ante "standard" migrant family for ex-post migrant families based on observable characteristics like the ones shown in Table 1. Then, based on the scores every ex-ante migrant family is matched with the ex-post migrant family who has the closest score—the nearest neighbor—and is not considered again for matching, hence there is the lack of replacement.

2.3 **Model estimation**

The variables chosen to be part of the propensity score calculations are selected based on the expectation of a different prevalence in "standard" migrant families and refugee families. No other consideration was taken since, according to Caliendo and Kopeinig (2008), matching is not intended to estimate structural parameters. In addition, I follow Rubin and Thomas (1996)'s recommendation against "trimming" models for the sake of parsimony. As a result, I do not remove variables based on their parameters' statistical significance provided there are reasonable doubts with respect to their relationship with being a "standard" migrant family.

The variables and their definitions are summarized in Table 2 for convenience. It should be noted that all variables are defined for the whole population of households. This also applies to variables defined for 15+ individuals because there is no household without at least one individual from said age group. In total, I use 12 variables that cover demographics (kids, young, women, widow), educational attainment (university, student), labor market indicators (fem work, NEET,

9 This noise is particularly acute among ex-ante migrants, with a higher tendency to live in mixed households.

10 See https://www.asylumineurope.org/reports/country/turkey/access-labour-market-0 for the exact figure and a provincial distribution.

Table 2 Variable description

Mnemonic	Short description	Full description
kids	Prop. 0–14	Proportion of people aged 0–14 in the family.
young	Prop. 15–24	Proportion of people aged 15–24 in the family.
women	Prop. 15+ women	Proportion of women among 15+ family members.
widow	Exist widows	Existence of at least one widow in the family.
university	Exist 15+ tertiary educ.	Existence of at least one 15+ university graduate.
student	Exist 15–24 students	Existence of at least one student aged 15–24.
fem work	Exist 15+ female workers	Existence of at least one working women.
NEET	Prop. 15+ NEETs	Proportion of NEETs among 15+ family members.
workers	Exist workers	Existence of at least one 15+ worker in the family.
informal	Number informal workers	Number of 15+ informal workers in the family.
garment	Exist male garment worker	Existence of at least one 15+ male garment worker.
turkish	Exist nonmigrants	Existence of at least one nonmigrant in the family.

Sources: Household Labour Force Survey 2017 and author's own calculations.
Note: The table describes the variables used in Equation (1).

workers, informal, garment), and the existence of mixed families (turkish). It is also worth noting that informality is defined using the existence of contributions to the social security institute. Moreover, I define the garment sector using ISIC[11] rev.4 codes 13, 14, and 15. On top of the previously described variables, regional dummies are represented by the matrix R in Equation (1).

Propensity scores are built with the help of a Logit model so as to maintain the probabilities of being a "standard" migrant family bounded between 0 and 1. The model is defined for the *i*th family using the logistic cumulative distribution function (CDF) as follows:

$$Pr(n_i = 1 \mid X_i) = G(\beta X_i), \tag{1}$$

where the probability of being a "standard" ($n = 1$) migrant family given some characteristics (X) is given by the logistic function $G(\cdot)$. The arguments inside this function are given by:

$$\begin{aligned}\beta X_i = \beta_0 + \beta_1 \text{kids}_i + \beta_2 \text{young}_i + \beta_3 \text{women}_i + \beta_4 \text{widow}_i + \beta_5 \text{university}_i + \beta_6 \text{student}_i + \beta_7 \text{femwork}_i \\ + \beta_8 \text{NEET}_i + \beta_9 \text{workers}_i + \beta_{10} \text{informal}_i + \beta_{11} \text{garment}_i + \beta_{12} \text{turkish}_i + \delta R_i + E_i\end{aligned} \tag{2}$$

Table 3 contains the marginal probability of being a "standard" migrant family after estimating the Logit model for 1,756 families, of which 377 are ex-ante migrant families and 1,379 are ex-post migrant families. The estimates confirm that most of the socioeconomic and employment-related variables shown in Table 1 are differential factors between migrant groups even after controlling for all of them at the same time. For example, it can be seen that living in a mixed household with a Turkish native decreases the probability of having found a Syrian family by 12%, while the same probability increases by 10% for every informally employed migrant found in the household. With respect to the proportions, the results show that an increase of 0.1 in the proportion of 15–24 kids in the family lowers the probability of being a "standard" migrant family by 1.9%. In addition, it is found that "standard" migrant families have a much higher propensity to live in the regions of Antalya and Van (data not shown in Table 2 due to space reasons). Still, geographical differences are much smaller than expected;

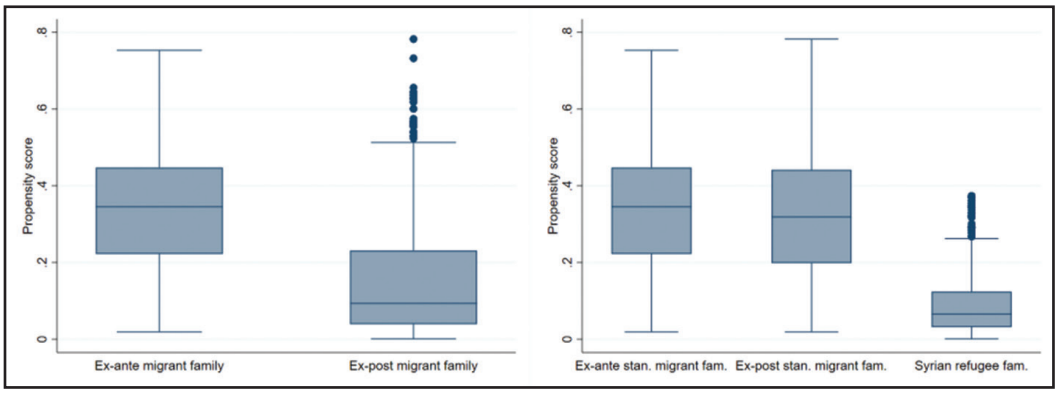
11 International Standard Industrial Classification.

Table 3 Estimates of the probability of being a "standard" migrant family, marginal effects

Variable	Marginal effect	Variable	Marginal effect
Prop. 0–14	−0.05	Exist 15+ female workers	0.10***
Prop. 15–24	−0.19***	Prop. 15+ NEETs	−0.15***
Prop. 15+ women	0.02	Exist 15+ workers	−0.09**
Exist widows	−0.04	Number of informal workers	−0.10***
Exist 15+ tertiary education	0.05**	Exist male garment worker	−0.07
Exist 15–24 students	0.12**	Exist nonmigrants	0.12***

Notes: Significance: ** at 5%, *** at 1%. Pseudo R^2: 0.1715. Estimated with 1,756 families.

Figure 2 Propensity scores, before and after matching: (a) before matching and (b) after matching.



Sources: Household Labour Force Survey 2017 and author's own calculations.

Notes: The box plots show propensity scores distributions. The three horizontal lines of the blue boxes denote the third quantile, the median, and the first quantile, from top to bottom. (a) shows the distribution before matching for foreign-born families who arrived, respectively, between 2004 and 2010 and between 2011 and 2017. (b) splits the scores of 2011–2017 migrant families between those families matched (ex-post "standard" migrant families) and those unmatched (Syrian families).

most Syrian refugees were initially registered in south eastern provinces of Turkey and hints refugees may have migrated to other regions.

2.4 Identification

Given the marginal probabilities shown in Table 2, I build propensity scores for each of the 1,756 families of the sample. Then, every ex-ante migrant family is matched with an ex-post migrant family and the 1,002 leftover families are labeled "Syrian refugees." The propensity scores of ex-ante and ex-post migrant families are shown in Figure 2(a), at it can also be seen in Table 1, that these two groups of migrant families are remarkably different from each other. Figure 2 shows propensity scores *after* matching is done for ex-ante "standard" migrant families (those arrived between 2004 and 2010), ex-post "standard" migrant families (matched families among those arrived between 2011 and 2017), and Syrian refugee families (unmatched families arrived between 2011 and 2017). In addition to isolating a group of Syrian families which is markedly distinct from earlier migrants, the matching has been able to create a control group with an almost identical distribution of propensity scores. This can be interpreted

Table 4 Summary statistics at the family level: after matching

Variable	"Standard" migrant families (2004–2010)	"Standard" migrant families (2011–2017)	Syrian families
Family size	1.67	1.91	3.88
Proportion of 0–14	0.09	0.09	0.25
Proportion of 15–24	0.16	0.15	0.27
Proportion of 15+ women	0.75	0.77	0.57
Existence of a widow	0.03	0.04	0.09
Existence tertiary educ.	0.41	0.38	0.17
Existence of 15–24 students	0.14	0.14	0.07
Existence of 15+ female workers	0.26	0.25	0.07
Proportion of 15+ NEETs	0.50	0.51	0.62
Existence of workers	0.42	0.39	0.61
Number of informal workers	0.18	0.16	0.81
Existence of male garment workers	0.02	0.02	0.14
Existence of non-migrants	0.57	0.55	0.16

Notes: The table shows averages for a number of variables at the family level. The term "existence" refers to the existence of at least one person with the mentioned characteristic. In all cases, it can be rejected that the difference in means between Syrian families and ex-post "standard" migrant families is equal to zero at the 95% confidence level.

in positive terms with regards to the second identification assumption as there seems to exist a group of migrants in the 2011–2017 period who share, on average, similar characteristics with those who arrived between 2004 and 2010.

The resulting matching can also be tested with the help of the same variables shown in Table 1. In this regard, Table 4 provides averages for 12 family-level variables for all of the three groups identified: ex-ante "standard" migrant families, ex-post "standard" migrant families, and Syrian refugee families. Overall, the matching provides a cleansing effect over all the statistics under analysis by increasing the differences between the averages held by Syrian refugee families and "standard" migrants. For example, the average Syrian refugee family has 3.85 members compared with 1.91 members living in the ex-post "standard" migrant families.[12] Other revealing examples include the number of informal workers (0.81 vs. 0.16), the existence of female workers (0.07 vs. 0.25), and the existence of a person with an university degree (0.38 vs. 0.17).

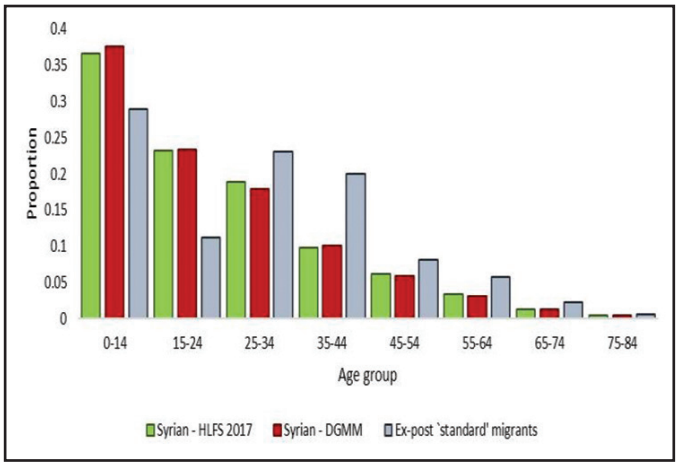
Certain dissimilarities can still be found between ex-ante and ex-post "standard" migrant families. These differences do not necessarily signal a lack of comparability between the two groups since they might be due to the time spanned between the arrival of ex-ante migrants and the time of the survey, 2017. For example, the fact that ex-ante migrants are 7 years older than ex-post migrants might explain the lower percentage of ex-post "standard" migrant families where at least one individual holds a tertiary degree.

2.5 **Quality check**

The dramatic increase in foreign-born individuals captured by the HLFS since the onset of the Syrian civil war and the marked differences in the socioeconomic indicators shown by those

12 Ex-post migrants families' (i.e., Syrian and ex-post "standard" families together) size is 3.34 before the separation,

Figure 3 Syrian refugees' population pyramid, HLFS vs. DGMM.



Sources: Household Labour Force Survey 2017, Directorate General of Migration Management, and author's own calculations.

Notes: The figure shows the proportion of Syrian refugees in a number of age groups from two sources: the HLFS 2017 and the DGMM. A chi-square test cannot reject that the HLFS and DGMM data are drawn from the same distribution. It also shows the proportion of ex-post "standard" migrants for comparison.

identified as Syrians leaves little doubt about them belonging to refugees. However, questions might still arise about the specific subpopulation represented by those captured by the matching methodology (refer Table 1).

As a quality control check, Figure 3 compares the population pyramid of the 3,858 Syrian refugees identified as such by the matching methodology with the pyramid of (1) the Syrian refugees under temporary protection registered by the Turkish Directorate General of Migration Management and (2) the "standard" ex-post migrants.[13] The figure shows the proportion of the group-specific population held by each of the eight age groups in which the population has been split. The age distribution of those identified as Syrian refugees in the HLFS is very similar to those who are supposed to be representing in all of the age groups under consideration.[14] In addition, the comparison between the age distribution of Syrian refugees and ex-post "standard" migrants shows that the matching is not trivial and the procedure has actually removed the noise that was surrounding the Syrian refugee population in the data.

3 Survey weights' adjustments

3.1 Background

The HLFS covers[15] all settlements of Turkey at the sampling stage, thus providing nationally representative figures on all residents with the exception of the noninstitutional population. In practice, though, the coverage is further restricted to Turkish natives residing in Turkey and foreigners with long-term residence permits (see İçduygu 2013, pp. 8–9). This restriction, which

13 Data retrieved from http://www.goc.gov.tr/icerik6/temporary-protection 915 1024 4748 icerik

14 It should be noted that I am comparing figures on Syrian refugees under temporary protection with estimates from the HLFS that represent *all* Syrian refugees. This is because the age breakdown of Syrian refugees with short-term residence permits and those with Turkish nationality could not be retrieved from Turkstat.

15 For more details, see http://tuik.gov.tr/MicroVeri/Hia 2017/english/meta-data/index.html

can be perceived as a minor issue, turns out to exclude several millions of Syrian refugees which currently populate Turkey.

The source of the exclusion revolves around the ABPRS, a registry setup by the Law 5490 of 2006 on Population Services which is used by the Turkish Statistical Institute to sample addresses. This system[16] matches, for foreigners with residence permits of at least 6 months,[17] addresses from the National Address Database (NAD) with passport numbers before storing the information in the ABPRS. The problem is that most Syrian refugees have not received neither a residence permit nor the Turkish nationality; according to Article 20, point (g) of the Law 6458 of 2013 on Foreigners and International Protection "a residence permit shall not be required from those foreigner holders of the documents listed in paragraph 7 of Article 69 as well as the first paragraphs of Articles 76 and 83." The mentioned paragraphs make reference to those applying for international protection in the different phases of the application process, in practice exempting Syrian refugees from (1) needing residence permits and (2) being registered in the ABPRS as their addresses are kept in a separate registry by the Directorate General of Migration Management.

3.2 **Syrian refugees in the HLFS**

In spite of the initial inability to covering individuals under the temporary protection regime, some of the interviewed households in the HLFS (approximately 1,000 households) are occupied by foreigners who, given the year of arrival to Turkey (among other characteristics), are likely to be Syrian refugees.

Two problems arise from the appearance of Syrian refugees in the HLFS sample. First, around 3,858 Syrian refugees are currently representing more than 1 million[18] Turkish citizens (including foreigners with long-term residence permits) even though their socioeconomic characteristics are far from comparison with the ones who are supposed to be representing. Second, since the sample currently includes Syrian refugees, the total population represented by the sample should be increased to 81.6 million as of July 2017, i.e., 78.6 million Turkish citizens and long-term foreign residents plus 3.19 million Syrian refugees as estimated by the DGMM (including Syrian who acquired the Turkish nationality, those on short-term residence permits, and those covered by the temporary protection regime).

3.3 **Nonresponse adjustment for Turkish residents**

I propose to solve the former problem by treating the existence of Syrian refugees as a nonresponse problem, i.e., as if the Turkish family that should have been interviewed was not present at home at the time of the visit. By following this assumption, the expanded number of Turkish people is down to 77.6 million, thus requiring an upward adjustment of the survey weights. This adjustment is performed by multiplying each non-Syrian refugee observation's survey

16 See Taştı (2009) for more information on how the system works.

17 As mentioned in Bel-Air (2016).

18 Expanded number of Syrian refugees using the original survey weights of the HLFS 2017.

weight, w, by a subregion-specific adjustment factor. These adjustment factors, f^{adj}, are defined at the NUTS-2[19] subregion level, $j \in (1, J)$, as follows:

$$f_j^{adj} = \frac{\sum_{i=1}^{N} w_{i,j}}{\sum_{i=1}^{T} w_{i,j}} \tag{3}$$

where N is the total number of observations in the sample and T is the number of Turkish natives plus foreigners with long-term residence permits, and a result of adding up survey weights the numerator and the denominator are equal to the respective expanded populations in a given subregion. Adjusted survey weights, w^{adj}, are then created[20] by multiplying the original survey weights by the region-specific adjustment factor,

$$w_{i,j}^{adj} = w_i f_j^{adj} \text{ for all non-Syrian refugees} \tag{4}$$

3.4 **Post-stratification adjustment for Syrian refugees**

The problem related to the representativeness of the Syrian refugees' sample is more contentious. To start with, the survey weights initially assigned to them in the HLFS have little value because they were meant for other people; they are consequently dropped altogether. In this case, a post-stratification adjustment can be used provided that something close to a census informing us of the total count of Syrian refugees in the country exists and provided that the sample of Syrian refugees is randomly drawn. The former is fulfilled by figures on the total population of Syrian refugees in Turkey regularly published by the Directorate General of Migration Management.[21] The latter assumption can be fulfilled by arguing that Syrian refugees were not expected to appear in the sample and, since the original sampling was meant to be representative of all regions of Turkey so are the households with Syrian refugees found by mistake. In other words, I do not expect the appearance of households with Syrians to happen more often in Adana than in Samsun other than by the fact that there are more Syrian refugees living in Adana than in Samsun.

The survey weights for Syrian refugees are assumed to be a function of the inverse proportion a person has of being selected in a specific subregion, p^{-1}, where the proportion is defined as

$$p_j = \frac{\sum_{i=1}^{N} ij}{\sum_{i=1}^{N} w_{i,j}} \tag{5}$$

and N represents the total number of observations in the sample. In addition, since the number of visits to mistaken households is not meant to have the necessary proportion for the weights to add up to the actual population of Syrian refugees, I add a correction factor $\overline{S}$ that makes the

19 See https://ec.europa.eu/regional policy/en/policy/what/glossary/n/nuts/ for an explanation on the statistical regional units classification. In Turkey, there are 26 subregions at the NUTS-2 level.

20 It should be noted that standard errors will increase as a result of the nonresponse adjustment. Users may want to consider the use of replication methods, including bootstrap when carrying out analysis with the proposed methodology to take into account the added uncertainty.

21 Even though these figures are published at the NUTS-3 level (provinces), I disregard the geographical distribution because I suspect that Syrian refugees have incentives to redistribute themselves within Turkey to areas with a higher number of job opportunities, for example, Bursa or Istanbul.

sum of the weighted sample add up to the official figure of Syrian refugees as of July 2017.[22] Survey weights are defined by multiplying the inverse proportion of being selected in a particular subregion by the adjustment factor as follows:

$$w_{i,j}^{adj} = p_j^{-1} \frac{\overline{S}}{\sum_{j=1}^{J} p_j^{-1} \sum_{i=1}^{S} i_j} \text{ for Syrian refugees} \tag{6}$$

where S is the number of Syrian refugees in the sample. It should be noted that, in practice, the correction factor divides the actual population of Syrians by the expanded population of Syrians which arises from the probability of choosing a person in a particular subregion. Both the adjustment factors for Turkish residents and the adjusted weights for Syrian refugees can be found in Table C1 in Appendix.

The application of this post-stratification adjustment allows me to estimate the actual geographical distribution of Syrian refugees. This distribution (HLFS) together with the official distribution as published by the government of Turkey can also be found in Table C1 in Appendix at the subregional level (NUTS-2), the lowest level of geographical disaggregation provided in the microdata. The comparison shows the existence of an internal migration pattern from Syrian-bordering subregions (notably Hatay, Şanlıurfa, and Gaziantep) to more industrialized areas such as Istanbul, Bursa, or Konya. This pattern, which could be the natural result of refugees' job search efforts, can be visualized with the help of maps in Figures C1 (official distribution), C2 (HLFS distribution), and C3 in Appendix which show the difference between the official and the HLFS-estimated refugees' geographical distributions.

4 Conclusions

The Syrian refugees hosted by Turkey have a higher risk of facing poverty and working conditions' deficits. As it is often the case with migrant populations, those more in need of help are also the ones for whom less information can be found due to the difficulties in tracking down these groups. This article proposes the use of the Turkey HLFS to overcome the information deficit with regards to Syrians in Turkey. In particular, I propose an indirect identification strategy to isolate Syrian refugees from other "standard" migrants, since both are grouped together in the publicly available microdata.

The identification strategy produces a population pyramid for HLFS refugees, that is comparable with the age profile recorded by the Turkish Directorate General for Migration Management. In addition, I show that Syrian refugees might have internally migrated from south-eastern provinces bordering Syria to more industrialized areas of Turkey like Bursa, Konya, or Istanbul. This pattern of internal migration would need to be confirmed by other instruments, yet it suggests that a reallocation of funds and humanitarian efforts might be due.

In addition, this methodology should allow researchers to use the full depth of Turkey's labor force survey for the study of the Syrian refugee population. This includes the creation of labor market indicators for this group such as those based on formality rates, average earnings, and details about the employment structure or the educational background. These and

22 This includes DGMM estimates on Syrian refugees under temporary protection and short-term residence permits as well as Syrians who acquired the Turkish nationality.

other results should allow policy makers and authorities alike to build better informed policies, including active labor market policies aimed at the Syrian population.

Acknowledgments

The author would like to thank Hakkı Özel, Numan Özcan, Burçe Ünver, and the Program of Support for Refugees' team of the ILO office for Turkey for useful discussions as well as two anonymous referees for valuable comments. The provision of the 2017 Household Labour Force Survey by the Turkish Statistical Institute under the license 27964695-622.03-E.26626 is appreciated.

References

Alsharabati, C.; J. Nammour (2016): Survey on Perceptions of Syrian Refugees in Lebanon. Institut des Sciences Politiques (Université Saint-Joseph de Beyrouth).

Bel-Air, F. D. (2016): Migration profile: Turkey. Policy Brief, Issue 2016/09, December.

BRIC (2013): Survey on the livelihoods of Syrian Refugees in Lebanon. Beirut Research and Innovation Center Research Report, November 2013.

Cagaptay, S. (2014): The Impact of Syria's Refugees on Southern Turkey. Policy Focus 130, Washington Institute for Near East Policy.

Caliendo, M.; S. Kopeinig (2008): Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys 22(1), 31-72.

Del Carpio, X. V.; M. Wagner (2015): The Impact of Syrian Refugees on the Turkish Labor Market. Policy Research Working Paper No. 7402, World Bank Group, August 2015.

Erdoğan, M. (2017): Syrians Barometer – 2017: A Framework for Achieving Social Cohesion with Syrians in Turkey. Executive Summary.

İçduygu, A. (2013): Turkey and International Migration 2012-2013. Report prepared for The Annual Meeting of the OECD Expert Group on Migration, Paris, 27-29November.

Kaymaz, T.; O. Kadboy (2016): Syrians in Turkey - The Economics of Integration. Alsharq Forum, Expert Brief.

Konun, B. B.; S. Tümen (2016): Immigration and Prices: Quasi-Experimental Evidence from Syrian Refugees in Turkey. Central Bank of Turkey Working Paper No 16/01.

Korfalı, D. K.; T. Acar (2018): Migration from Central and Eastern Europe to Turkey, in: Peter, S.; M. van Ostaijen (eds.), Between Mobility and Migration. IMISCOE Research Series. Cham: Springer, 227-248.

Rubin, D.; N. Thomas (1996): Matching Using Estimated Propensity Scores: Relating Theory to Practice. Biometrics 52(1), 249-264.

Taştı, E. (2009): Address Based Population Registration System, 2007 Population Census in Turkey. Conference Paper.

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

Uçak, S.; K. Raman (2017): Another Side of the Story: A Market Assessment of Syrian SMEs in Turkey. Syria: Building Markets.

Yavçan, B. (2017): On Governing the Syrian Refugee Crisis Collectively: The View from Turkey. Near Futures Online, 1, 'Europe at a crossroads' March 2016.

Appendix

A Derived variables from the HLFS

A.1 Number of children aged 0–14

The data provided in the HLFS by the Turkish Statistical Institute refer only to those aged 15 or older; however, information regarding the number of people aged 0–14 living in the household can be retrieved. This variable is calculated by subtracting the variable hh buyukluk, which contains the number of people in the household (including children), with a variable of our own creation that contains the number of household members aged 15 or more. Some pieces of household-level information are assigned to the children, for instance, the region (NUTS-1), the subregion (NUTS-2), and the survey weight; in addition, children living in a household whose head is foreign-born are given the year of arrival of the head of the household provided he/she arrived not before 14 years since the time of the survey. In all other cases, it is assumed that the children were born in Turkey.

A.2 Year of arrival to Turkey of persons born abroad

One of the questions available in the HLFS is whether the respondents were born abroad or in Turkey. As an example, this group contains 10,032 observations in the 2017 HLFS, including children aged 0–14 (see first paragraph of this appendix). For most[23] of the foreign-born population, the dataset also provides information on their year of arrival to Turkey. The variable "year of arrival" is built in two steps as its information comes from two sources: the variable buil **yil** for those who live in the same province since their arrival to Turkey and the variable **tr yil** for those who changed provinces within Turkey at least once since their arrival *and* have lived abroad for at least 12 months.

The full list of logical skips used to build the year of arrival is shown in Table A1, where the column "subpopulation" presents the conditions that respondents need to fulfill for their year of arrival to come from either of the two options. It should be noted that the condition referring to the variable buil yasama (which asks whether the person have permanently lived in the current province) is shown for completeness but it does not make any difference as it is logically impossible to have lived the whole life in the same Turkish province while being born abroad.

The resulting variable is plotted in Figure A1 which shows the number of observations by year of arrival. It can be observed the existence of two peaks, one in 1989, which coincides with the migration/expulsion of Turks from Bulgaria, and the second after 2011, right after the Syrian civil war.

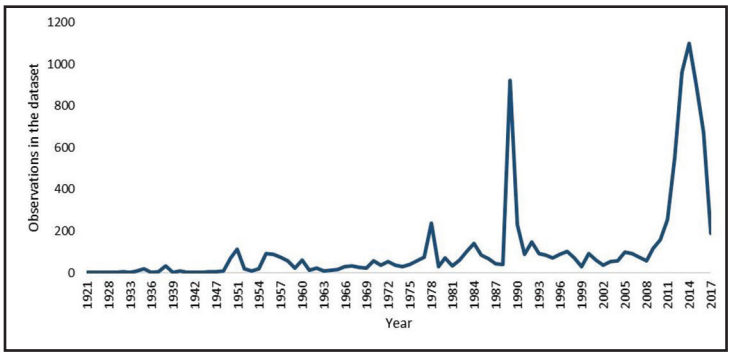
23 The survey does not ask the arrival year to foreign-born residents who have changed their province of residence if they have lived less than 12 months outside Turkey. For instance, in the 2017 HLFS, this group totals 344 observations out of the 1,721 which conform the group of foreign-born who have changed provinces within Turkey at least once in their lives.

Table A1 Construction of the variable 'year of arrival to Turkey'.

Year of arrival	Variable	Subpopulation	Observed HLFS 2017
	buil yil	dogum yer=Abroad	8,310
		buil yasama=No*	
		onceki ikamet=Abroad	
	tr yil	dogum yer=Abroad	1,377
		buil yasama=No	
		onceki ikamet=Turkey	
		yurtdisi durum=Yes	

Notes: *In practice, it is not possible for a person born abroad to have lived permanently in the same Turkish province his/her whole life. There is only one observation (out of 10,032) in the 2017 HLFS for which a Yes is recorded in 'buil yasama' despite being born abroad and it is disregarded.

Figure A1 Born abroad by year of arrival to Turkey.



B **Post-stratification weights**

Table B1 Re-weighting by subregion (NUTS-2), 2017 HLFS

	Turkish		Syrian refugees	
Subregion	**Provinces**	**Adj. factor**	**Inv. prob.**	**Adj. weight**
İstanbul	İstanbul	1.0203	338	1,511
Tekirdağ	Tekirdağ, Edirne, Kırklareli	1.0024	109	485
Balıkesir	Balıkesir, Ç anakkale	1.0020	96	428
İzmir	İzmir	1.0076	196	876
Aydın	Aydın, Denizli, Mŭgla	1.0079	157	700
Manisa	Manisa, Afyon, Kütahya, Uşak	1.0041	171	765
Bursa	Bursa, Eskişehir, Bilecik	1.0207	195	871
Kocaeli	Kocaeli, Sakarya, Düzce, Bolu, Yalova	1.0032	188	838
Ankara	Ankara	1.0077	192	856
Konya	Konya, Karaman	1.0429	83	370
Antalya	Antalya, Isparta, Burdur	1.0081	166	743
Adana	Adana, Mersin	1.0268	183	817
Hatay	Hatay, Kahramanmaraş, Osmaniye	1.0143	187	837
Kırıkkale	Kırıkkale, Aksaray	1.0263	85	378
	Niğde, Nevşehir, Kırşehir			
Kayseri	Kayseri, Sivas, Yozgat	1.0075	160	716
Zonguldak	Zonguldak, Karabük, Bartın	1.0008	110	490
Kastamonu	Kastamonu, Ç ankırı, Sinop	1.0029	55	246
Samsun	Samsun, Tokat, Ç orum, Amasya	1.0026	139	620
Trabzon	Trabzon, Ordu, Giresun	1.0003	124	552
	Rize, Artvin, Gümüşhane			
Erzurum	Erzurum, Erzincan, Bayburt	1.0016	75	335
Ağrı	Ağrı, Kars, Iğdır, Ardahan	1.0009	77	343
Malatya	Malatya, Elazığ, Bingöl, Tunceli	1.0008	120	537
Van	Van, Muş, Bitlis, Hakkâri	1.0004	92	411
Gaziantep	Gaziantep, Adıyaman, Kilis	1.0255	144	642
Şanlıurfa	Şanlıurfa, Diyarbakır	1.0161	177	791
Mardin	Mardin, Batman, Şırnak, Siirt	1.0019	159	711

Notes: The numbers from the column “Adj. weight” multiply the numbers of the column “Inv. prob.” by a factor of 4.4673; the multiplications might not be added up in the table due to rounding.

Source: Household Labour Force Survey 2017 and author’s own calculations.

C Geographical distribution of Syrian refugees

Table C1 Syrian refugees under temporary protection in Turkey by subregion (NUTS-2), 2017

		Syrian refugees in Turkey		
Subregion	**Provinces**	**Official**	**HLFS estimated**	**Difference**
İstanbul	İstanbul	497,135	1,148,199	651,064
Bursa	Bursa, Eskişehir, Bilecik	113,989	260,277	146,288
Konya	Konya, Karaman	76,744	198,198	121,454
Kırıkkale	Kırıkkale, Aksaray	15,055	80,724	65,669
	Niğde, Nevşehir, Kırşehir			
Adana	Adana, Mersin	308,641	358,830	50,189
Antalya	Antalya, Isparta, Burdur	15,438	60,270	44,832
Aydın	Aydın, Denizli, Mŭgla	26,483	69,100	42,617
Manisa	Manisa, Afyon, Kütahya, Uşak	13,554	54,504	40,950
Ankara	Ankara	75,881	111,672	35,791
Samsun	Samsun, Tokat, Ç orum, Amasya	7,579	29,328	21,749
Erzurum	Erzurum, Erzincan, Bayburt	950	9,483	8,533
Kastamonu	Kastamonu, Ç ankırı, Sinop	1,449	8,748	7,299
Balıkesir	Balıkesir, Ç anakkale	6,222	11,760	5,538
Kocaeli	Kocaeli, Sakarya, Düzce, Bolu, Yalova	46,533	51,336	4,803
Ağrı	Ağrı, Kars, Iğdır, Ardahan	1,333	3,696	2,363
Zonguldak	Zonguldak, Karabük, Bartın	883	2,892	2,009
Tekirdağ	Tekirdağ, Edirne, Kırklareli	15,719	16,730	1,011
Van	Van, Muş, Bitlis, Hakkâri	4,662	5,642	980
Trabzon	Trabzon, Ordu, Giresun	4,027	4,352	325
	Rize, Artvin, Gümüşhane			
İzmir	İzmir	112,881	106,764	−6,117
Malatya	Malatya, Elazığ, Bingöl, Tunceli	29,689	7,965	−21,724
Kayseri	Kayseri, Sivas, Yozgat	67,207	42,480	−24,727
Mardin	Mardin, Batman, Şırnak, Siirt	136,673	12,672	−124,001
Şanlıurfa	Şanlıurfa, Diyarbakır	466,811	180,728	−286,083
Gaziantep	Gaziantep, Adıyaman, Kilis	497,371	208,257	−289,114
Hatay	Hatay, Kahramanmaraş, Osmaniye	536,986	143,244	−393,742

Notes: Official figures on Syrian refugees under temporary protection from the DGMM as of July 2017. Estimated figures are calculated with the help of adjusted survey weights that take into account the proportion of the population sampled in each subregion as well as a correction factor. Differences of less than 20,000 should be disregarded due to the small sample size.

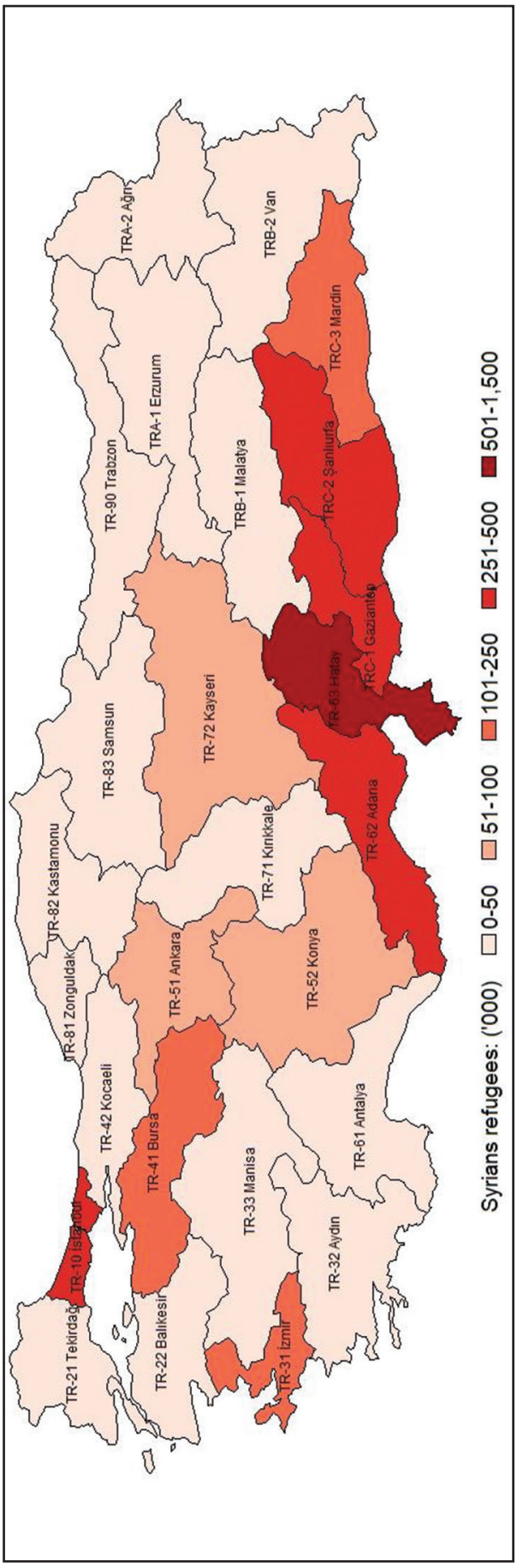


Figure C1 Number of Syrian refugees under temporary protection by subregion, DGMM July 2017.

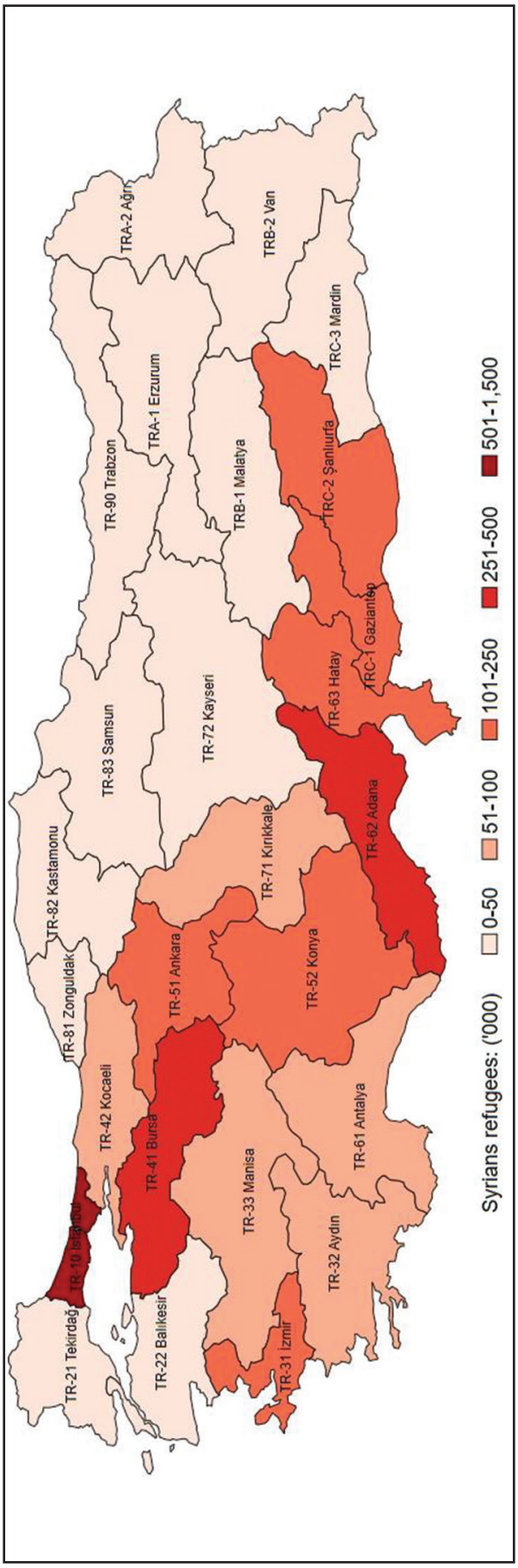


Figure C2 Number of Syrian refugees under temporary protection by subregion, HLFS 2017.

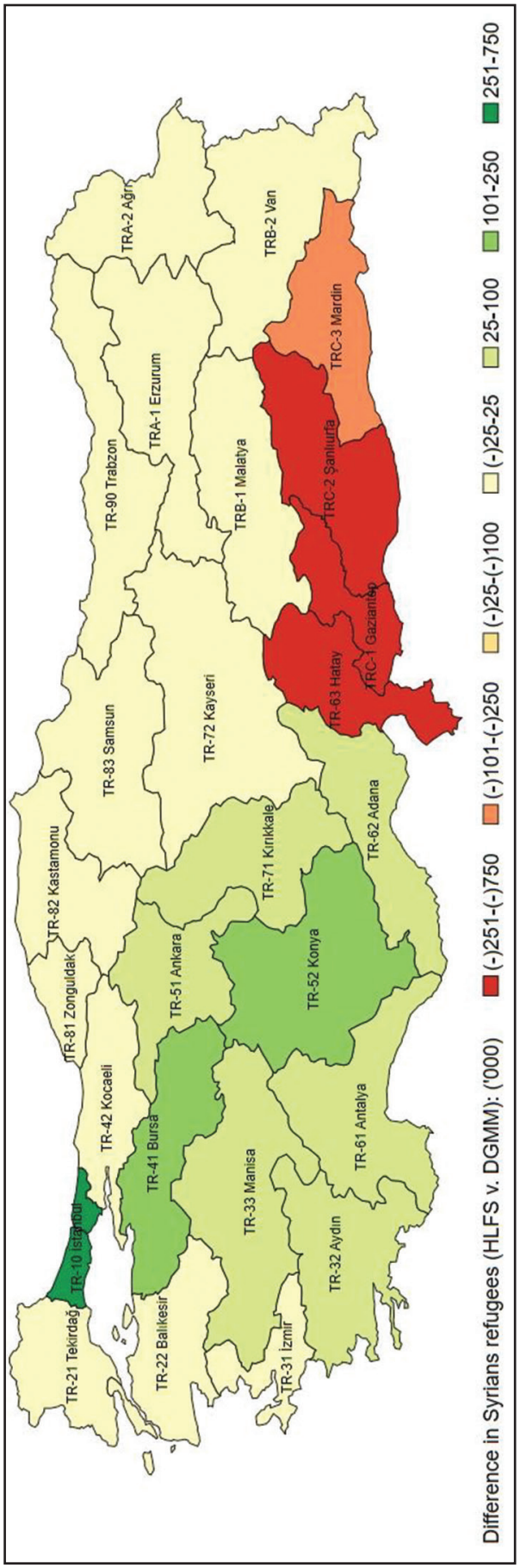


Figure C3 Difference in Syrian refugees under temporary protection by subregion, HLFS vs. DGMM.