

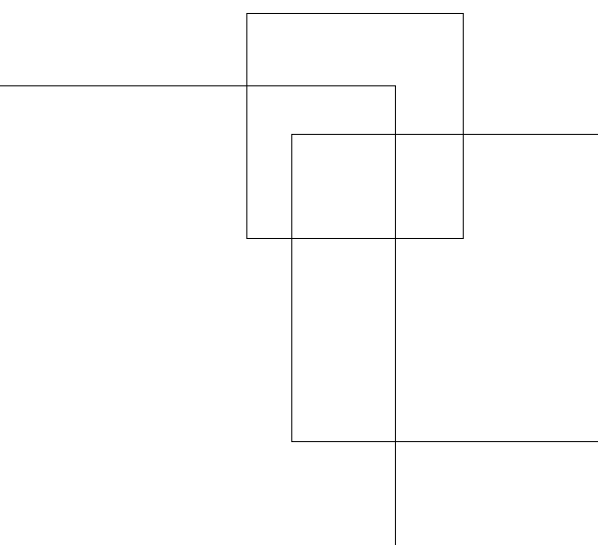


International  
Labour  
Office

RESEARCH DEPARTMENT ||||| WORKING PAPER NO. 46 ||||||

# Identifying Syrian Refugees in Turkish Microdata

LUIS PINEDO ||||||



|||||

SEPTEMBER 2019

## Table of contents

Acknowledgements

Abstract

1	Introduction .....	1
2	An identification strategy for Syrian refugees in the HLFS .....	3
3	Survey weights' adjustments .....	8
4	Conclusion .....	10
	References .....	11
	Appendix .....	12
	Appendix .....	14
	Appendix .....	15

## Acknowledgements

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

## 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 normal flow of migration that arrived to Turkey during the Syrian war. I propose to eliminate the regular 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, non-regular migration, i.e. Syrian refugees. The results show that the age distribution of the non-regular 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.

Keywords: Syrian refugees, Turkish microdata, Identification strategy

JEL classification: C-81, F-22, J-11



## 1 Introduction

The succession of revolts that followed the Arab Spring was typically characterized by short-termed demonstrations and/or outbursts of violence in most of the affected countries; all but one: Syria. Since March 2011 until now 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<sup>1</sup> of registered refugees of which 3.6 million<sup>2</sup> were welcomed by Turkey under the temporary protection regime.<sup>3</sup> For Turkey in particular, this unprecedented situation has not only produced an humanitarian emergency but also has likely affected the lives of millions of Turkish people. In this context, the demand for policy responses is pressing and so does the demand for relevant information. This report aims, precisely, at filling an information gap by means of a strategy that would allow using the primary source for labour market statistics of Turkey, the Household Labour Force Survey, for the creation of statistics on Syrian refugees.

The use of microdata when informing about the Syrian refugees 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) study the effect of Syrian refugees' arrival on the price level of goods, 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 labour costs onto the goods' prices. In addition, Tümen (2016) also finds that natives not only have lower chances of finding an informal job, but also higher chances of finding a formal one 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 found in 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 of finding a large displacement of Turkish natives from the informal sector due to the arrival of the refugee population, also find a net displacement of women and the low educated away from the labour 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 the Syrian population, it lacks, beyond a few basic questions, deep information with regards Syrians' labour market performance.

Other ad-hoc surveys on Syrian refugees' socio-economic 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, 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, currently active and had at least an employee. On the negative side, this database is not meant to be nationally representative, as it is also confirmed by its authors. 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 per cent of Syrian refugees have

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

<sup>2</sup> Data from the DGMM, updated as of 4<sup>th</sup> April, 2019.

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

university degrees. The extent to which Syrian refugees have such high qualification might be have been exaggerated due to the survey mode but it brings to the spotlight the importance of developing a model for recognition of prior learning of refugees.

Lack of data affects the depth of any research on Syrian refugees; an example can be found in Yavçan (2017), where the author tries 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), who, in 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 with the availability found for Lebanon, where at least 2 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.<sup>4</sup>

**Household Labour Force Survey.** The fact that the refugee population in Turkey represents 4.4 per cent of the population living in Turkey<sup>5</sup> creates a growing need for nationally representative data on Syrian refugees in Turkey that is not currently being fulfilled. Fortunately, the relatively high proportion of Syrian refugees in the Turkish economy might have as well opened the door to the use of nationally representative microdata from the Turkish Statistical Institute. Using Turkstat databases for analysing Syrian refugees is not straightforward, though. 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.<sup>6</sup> An exception to this survey methodology is given by the Household Labour Force Survey<sup>7</sup> which instead of families targets addresses, thus allowing interviewers to find Syrian families under certain conditions.

Even though Syrians refugees now 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 normal flow of migrants coming to Turkey (see Appendix A for a quick visual inspection of how this flow looks like). In practice, we propose an indirect identification whereby we remove non-Syrian migrants during the 2011-2017 period, thus finding Syrian refugees as a leftover. Full details of this methodology can be found in Section 2 of this article, including a comparison between the Syrians' age distribution found in the HLFS and the one recorded by the DGMM. The rest of the report is mostly based on data from the 2017 HLFS, which we exploit to gain some understanding about Syrian refugees living and working conditions. In addition, wherever appropriate, we compare Syrians' statistics with those of Turkish natives with the intention of having a reference point.

The indirect identification method proposed in this article should capture all irregular migration that came into Turkey between 2011 and 2017. In practice, this means all Syrians who migrated during that period have been identified, including those covered by the temporary protection regime, those with short-term residence permits and Syrian who acquired the Turkish nationality. It should be noted that some other migrants (particularly those coming from Iraq or Afghanistan in recent times) might have also been included in the group. Still, throughout the report we refer to the group as a whole as 'Syrian refugees', first, because of the likely similar characteristics in terms of labour market indicators and, second, because of the small number other migrants would represent in comparison with Syrians refugees.

---

<sup>4</sup> Syrian refugees are under-represented in the Labour Force Survey, however their survey weights have been adjusted to add-up to their total population.

<sup>5</sup> As reported by the DGMM at [http://www.goc.gov.tr/icerik6/temporary-protection\\_915\\_1024\\_4748\\_icerik](http://www.goc.gov.tr/icerik6/temporary-protection_915_1024_4748_icerik).

<sup>6</sup> Their information is kept separately by the DGMM.

<sup>7</sup> See [http://tuik.gov.tr/MicroVeri/Hia\\_2017/turkce/index.html](http://tuik.gov.tr/MicroVeri/Hia_2017/turkce/index.html) for more information.

In what follows, Section 2 explains the matching strategy utilized to isolate the refugees. Section 3 presents a post-stratification adjustment to take into account the fact that the HLFS may be representing 3 million individuals more. At last, Section 4 concludes.

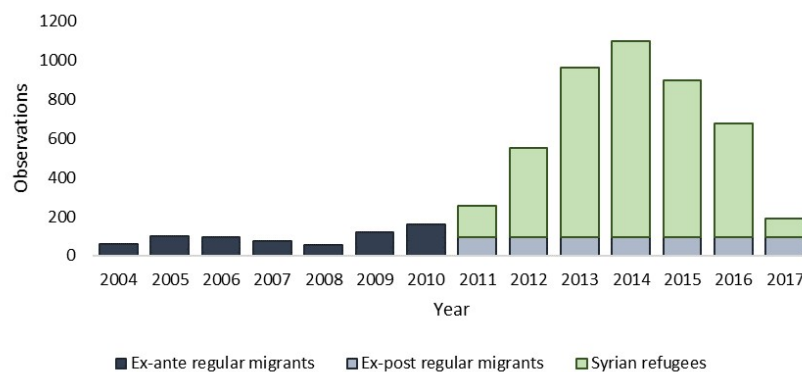
## 2 An identification strategy for Syrian refugees in the HLFS

The number of foreign-born individuals present in the 2017 Household Labour Force Survey (HLFS) is six times larger among those who arrived between 2011 and 2017 than among those that did so between 2004 and 2010, suggesting the survey may have captured Syrian refugees. Unfortunately, the publicly available microdata of the HLFS does not contain information on the country of origin and, even though we 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 that preceded the Syrian war (2004-2010). As a result, Syrian refugees are probably mixed up in the data with the hereinafter called ‘regular’ migrants, thus preventing a direct identification of Syrian refugees.

### Assumptions

In order to identify the Syrian refugees present in the sample we pursue an indirect identification strategy. Instead of finding Syrians among the 2011-2017 migrants we find those who are not, to then remove them from the sample (see Figure 1 for a visual explanation of the idea) with Syrian refugees being a ‘left-over’ of the procedure. For this strategy to work we assume that there is a relatively constant flow of what we call ‘regular’ migrants. This assumption is critical in that it provides the number of observations that need to be removed from the ex-post migrants’ group -migrants arrived between 2011 and 2017.

**Figure 1: Number of foreign-born individuals in Turkey by year of arrival, 2004-2017**



Source: 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 regular migrants’) from the total number of observations in each of the years between 2011 and 2017 (ex-post migrants).

In addition, ex-ante and ex-post ‘regular’ migrants, some of them thought to be Turkish-German by Bel-Air (2016), are assumed to share similar socio-economic characteristics which are (1) observable in the microdata and (2) significantly different from those of Syrian refugees. This allows for the separation of ‘regular’ 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. we 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, we would expect Syrian refugees and ‘regular’ migrants



to differ. It should be noted that for the sake of relevance, the comparison is done at the family level. This is because we 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; that is to minimize the noise due to mixing<sup>8</sup> with the local population and/or other migrants.

It stems from Table 1 that sharp differences exist between the families who arrived to Turkey between 2004 and 2010 and those who did so 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 existence of a much younger population. The relatively high number of households with a widow is a hint that points at the possibility of having some ex-post migrant families escaping from a civil war. In addition, we observe significantly less individuals with tertiary education, less female workers and more NEETs among ex-post migrants, which strongly suggests the existence of strong differences at the cultural and at the socio-economic level between the two groups under analysis. At last, the high prevalence of informal workers among ex-post migrants, 43 per cent of their households have at least one such worker, fits well with these families being Syrian as not many of their members have managed to obtain a working permit. In sum, based on the observed differences, it is reasonable to argue that ex-post migrants constitute a different group which, in turn, supports the use of matching.

**Table 1: Summary statistics at the family level: Before matching**

<i>Variable</i>	<i>Migrant families 04-10</i>	<i>Migrants Families 11-17</i>	<i>Ratio (ex-post/ex-ante)</i>
<i>Family size</i>	1.68	3.34	2.00
<i>Proportion of 0-14</i>	0.09	0.20	2.23
<i>Proportion of 15-24</i>	0.16	0.24	1.50
<i>Proportion of 15+ women</i>	0.75	0.63	0.83
<i>Existence of a widow</i>	0.03	0.08	2.37
<i>Existence tertiary educ.</i>	0.41	0.23	0.55
<i>Proportion of 15-24 students</i>	0.67	0.20	0.29
<i>Proportion of 15+ female workers</i>	0.31	0.14	0.44
<i>Proportion of 15+ NEETs</i>	0.50	0.59	1.18
<i>Number of workers</i>	0.51	0.78	1.53
<i>Number of informal workers</i>	0.18	0.63	3.50
<i>Existence of male garment workers</i>	0.02	0.11	6.61

Source: 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. The proportions of 15-24 students and 15+ female workers are calculated for families who, respectively, have 15-24 aged old individuals and 15+ women. In all cases it can be rejected that the difference in means is equal to zero at the 95 per cent confidence level.

## Matching

The matching of ex-post 'regular' migrant families with ex-ante 'regular' 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 'regular' migrant family for ex-ante 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 the lack of replacement.

<sup>8</sup> This noise is particularly acute among ex-ante migrants, with a high tendency to live in mixed households.

**Model estimation.** Propensity scores are built with the help of a Logit model so as to maintain the probabilities of being a ‘regular’ 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:

$$Pr(n_i = 1|X_i) = G(\beta X_i), \quad (1)$$

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

$$\begin{aligned} \beta X_i = & \beta_0 + \beta_1 kids_i + \beta_2 young_i + \beta_3 women_i + \beta_4 widow_i + \beta_5 university_i + \beta_6 student_i \\ & + \beta_7 fem\_work_i + \beta_8 NEET_i + \beta_9 workers_i + \beta_{10} informal_i + \beta_{11} garment_i + \vec{\delta} R_i + \epsilon_i, \end{aligned} \quad (2)$$

where *kids*, *young* and *women* denote the proportion in the household of, respectively, individuals aged 0-14, individuals aged 15-24 and women. The variable *widow* stands for the existence of at least 1 widow in the family. Likewise, *university* denotes the existence of at least 1 person with tertiary studies in the family. The variables *fem\_work*, *NEET* and *student* are proportions of (1) women working among +15 women, (2) NEETs among the +15 population and (3) students among the 15-24 population; in all cases a zero is assigned when the proportion is not applicable (i.e. there are no women in the family).<sup>9</sup> At last, *workers* and *informal* measure, respectively, the number of workers and informal<sup>10</sup> workers in the household while the variable *garment* controls for the existence of at least one male worker in the textile, garment, footwear and leather industry (TCLF).<sup>11</sup> On top of the above-mentioned variables, regional dummies are included in the matrix denoted by the letter R.

**Table 2: Estimates of the probability of being a ‘regular’ migrant family**

Variable	Probability	Variable	Probability
<i>Prop. 0-14</i>	-0.26***	<i>Prop. +15 female workers</i>	0.13*
<i>Prop. 15-24</i>	-0.28***	<i>Prop. +15 NEETs</i>	-0.16**
<i>Prop. +15 women</i>	0.07	<i>Number of workers</i>	0.03
<i>Exist widows</i>	-0.13*	<i>Number of informal workers</i>	-0.15*
<i>Exist +15 tertiary educ.</i>	0.06*	<i>Prop. informal worker</i>	-0.14
<i>Prop. 15-24 students</i>	0.23***	<i>Exist male garment worker</i>	-0.19*

Significance: \*\*\* at 99%, \*\* at 95%, \* at 90%. Pseudo  $R^2$  : 0.1680.

Table 2 contains the marginal probability of being a ‘regular’ 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 the socio-economic and cultural differences shown in Table 1 are significantly different from zero; for example, it can be seen that the existence of at least a widow in the household raises the probability of having found a Syrian family by 13 per cent, while the existence of at least one informal worker increases the mentioned probability by 30 per cent. With respect to the proportions, the results show that an increase of 0.1 in the proportion of 0-14 kids in the family lowers the probability of being a ‘regular’ migrant family by 2.8 per cent. In addition, it is found that ‘regular’ migrant families have a much higher propensity to live in the subregions of Antalya and Van (not shown in Table 2 for space reasons) than Syrian refugees.

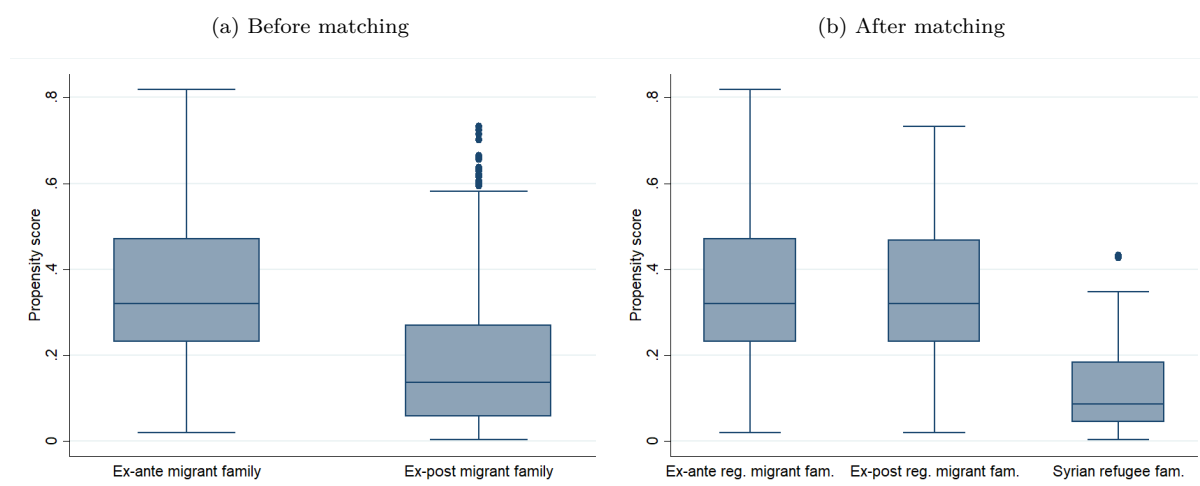
<sup>9</sup> The population proportions (*young*, *women*) provide the marginal probability for those families containing the applicable population but lacking the characteristic. For instance, if there are 15-24 individuals but no one is studying.

<sup>10</sup> Informality occurs whenever contributions to social security are not provided on behalf of the worker.

<sup>11</sup> ISIC rev.4 codes 13, 14 and 15.

**Identification.** Given the marginal probabilities shown in Table 2 we build propensity scores for each of the 1,756 families in the sample. Then, every ex-ante migrant family is matched with an ex-post migrant family and the 1,002 leftover families are labelled ‘Syrian refugees’. The propensity scores of ex-ante and ex-post migrant families are shown in Figure 2a confirming these two groups of migrant families are very different from each other. Figure 2b shows the propensity scores *after* the matching is done, with probabilities assigned separately for ex-ante ‘regular’ migrant families (those arrived between 2004 and 2010), ex-post ‘regular’ migrant families (matched families from 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 similar distribution of propensity scores. This is reassuring with respect to the first identification assumption, the existence of a constant flow of regular migrants.

**Figure 2: Propensity scores, before and after matching**



Source: 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 from top to bottom, the third quantile, the median and the first quantile. Part (a) shows the distribution before matching for foreign-born families who arrived, respectively, between 2004-2010 and between 2011 and 2017. Part (b) splits the scores of 2011-2017 migrant families between those families matched (ex-post ‘regular’ migrant families) and those unmatched (Syrian families).

The resulting matching can also be tested with the help of the same variables shown in Table 1; in this regard, Table 3 provides averages for 12 family-level variables for all three groups identified, ex-ante and ex-post ‘regular’ 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 ‘regular’ migrants. These differences are even more dramatic than in Table 1; for example the average Syrian refugee family has 3.85 members compared to 2.00 members living in the ex-post ‘regular’ migrant families.<sup>12</sup> Other revealing examples include the number of informal workers (0.81 vs. 0.16), the existence of female workers (0.08 vs. 0.29) and the proportion of 15-24 students (0.11 vs. 0.63).

Certain dissimilarities can still be found between the two groups of ‘regular’ migrant families, the ex-ante and the ex-post. These differences are not necessarily a signal of a lack of comparability between the two

<sup>12</sup> Ex-post migrants families’ (i.e. Syrian and ex-post ‘regular’ families together) size is 3.31 before the separation, refer to Table 1.

groups since they might be due to the time spanned between the arrival of ex-ante migrants and the present time. For example, the fact that ex-ante migrants are 7 years older than ex-post migrants might explain the imperfect match in, say, the lower percentage of ex-post ‘regular’ migrant families’ households where at least one individual holds a tertiary degree.

**Table 3: Summary statistics at the family level: After matching**

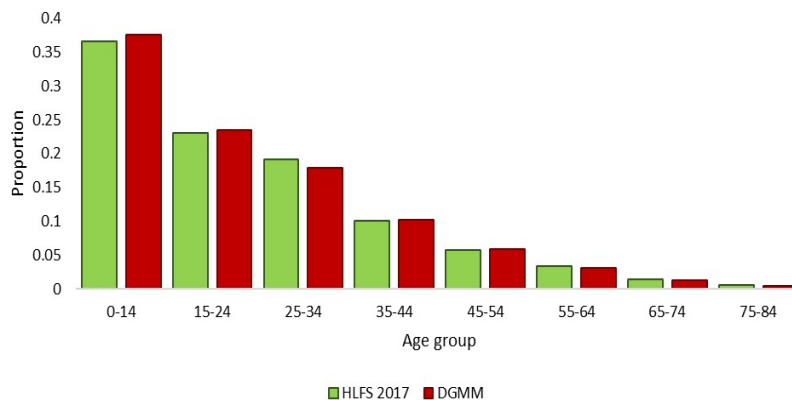
<i>Variable</i>	<i>Regular migrant families '04-'10</i>	<i>Regular migrant families '11-'17</i>	<i>Syrian families</i>
<i>Family size</i>	1.68	2.00	3.85
<i>Proportion of 0-14</i>	0.09	0.09	0.24
<i>Proportion of 15-24</i>	0.16	0.16	0.26
<i>Proportion of 15+ women</i>	0.75	0.76	0.58
<i>Existence of a widow</i>	0.03	0.03	0.09
<i>Existence tertiary educ.</i>	0.41	0.38	0.17
<i>Proportion of 15-24 students</i>	0.67	0.63	0.11
<i>Proportion of 15+ female workers</i>	0.31	0.29	0.08
<i>Proportion of 15+ NEETs</i>	0.50	0.50	0.62
<i>Number of workers</i>	0.51	0.47	0.90
<i>Number of informal workers</i>	0.18	0.16	0.81
<i>Existence of male garment workers</i>	0.01	0.01	0.14

The table shows averages of the mentioned statistics at the family level. The ‘existence’ refers to the existence of at least one person with the mentioned characteristic. The proportions of 15-24 students and 15+ female workers are calculated for families who, respectively, have 15-24 aged old individuals and 15+ women. In all cases it can be rejected that the difference in means between Syrian families and ex-post ‘regular’ migrant families is equal to zero at the 95 per cent confidence level.

## Quality check

The sharp increase in foreign-born individuals captured by the HLFS since the onset of the Syrian civil war and the marked differences in the socio-economic indicators shown by those identified as Syrians leave little doubt about them belonging to refugees. However, questions might still arise about the specific sub-population represented by those captured by the matching methodology.

**Figure 3: Syrian refugees’ population pyramid, HLFS vs. DGMM**



Source: Household Labour Force Survey 2017, Directorate General of Migration Management and author’s own calculations. Notes: The figures shows the proportion of Syrian refugees in a number of age groups from two sources, the HLFS 2017 and the DGMM. A chi-squared test cannot reject that the HLFS and DGMM data are drawn from the same distribution.

As a quality control check, Figure 3 compares the population pyramid of the 3,858 Syrian refugees identified as such in the HLFS with the population pyramid of the Syrian refugees under temporary protection registered by the Turkish Directorate General of Migration Management.<sup>13</sup> The figure shows the proportion held by each of the 8 age groups in which the population has been split. The conclusion is that the age distribution of those identified as Syrian refugees in the HLFS have a very similar age distribution to the one of those they are supposed to be representing in all age groups under consideration.<sup>14</sup>

### 3 Survey weights' adjustments

**Background.** The Household Labour Force Survey (HLFS) covers<sup>15</sup> all settlements in Turkey, thus, providing nationally representative figures on all residents with the exception of the non-institutional population. In practice, though, the coverage is further restricted to Turkish natives residing in Turkey and foreigners with long-term residence permits, see pages 8-9 from İçduygu (2013). This restriction, which can be perceived as a minor issue turns out to exclude several millions of Syrian refugees that currently populate Turkey.

The source of the exclusion revolves around the Address Based Population Registration System (ABPRS), a registry set up by the Law 5490 of 2006 on Population Services which is used by the Turkish Statistical Institute to sample addresses. This system<sup>16</sup> matches, for foreigners with residence permits of at least 6 months,<sup>17</sup> 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 foreigners 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 excluding Syrian refugees from (1) the need of having a residence permit and (2) being registered in the ABPRS as their addresses are kept in a separated registry by the Directorate General of Migration Management.

**Syrian refugees in the HLFS.** In spite of the initial inability of covering individuals under the temporary protection regime, some of the interviewed households in the HLFS (approximately 1,000 households) are occupied by foreigners whom, 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; on the one hand around 3,858 Syrian refugees are now representing more than a million<sup>18</sup> Turkish citizens (including foreigners with long-term residence permits) even though their socio-economic characteristics are far from comparable to the ones of the people they are supposed to be representing. On the other hand, since the sample now 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 Turkish and long-term foreign residents plus 3.19 million of 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).

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

<sup>14</sup> It should be noted that we are 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 who acquired the Turkish nationality could not be retrieved from Turkstat. This could explain why according to the HLFS estimates Syrian refugees are slightly older.

<sup>15</sup> See [http://tuik.gov.tr/MicroVeri/Hia\\_2017/english/meta-data/index.html](http://tuik.gov.tr/MicroVeri/Hia_2017/english/meta-data/index.html) for more details.

<sup>16</sup> See Taştı (2009) for more information on how the system works.

<sup>17</sup> As mentioned in Bel-Air (2016).

<sup>18</sup> Expanded number of Syrian refugees using the original survey weights of the HLFS 2017.

**Non-response adjustment for Turkish residents.** We propose to solve the former problem by treating the existence of Syrian refugees as a non-response 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 upwards adjustment of the survey weights. In this research we perform the adjustment by multiplying each non-Syrian refugee observation's survey weight,  $w$ , by a sub-region specific adjustment factor. These adjustment factors,  $f_j^{adj}$ , are defined at the NUTS-2<sup>19</sup> subregion level,  $j \in (1, J)$ , as follows

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

where  $N$  is the total number of observations in the sample,  $T$  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<sup>20</sup> by multiplying the original survey weights by the region-specific adjustment factor,

$$w_{i,j}^{adj} = w_{i,j} f_j^{adj} \quad \text{for all non-Syrian refugees.} \quad (4)$$

**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 performed provided that something close to a census informing us of the total count of Syrian refugees in the country exists and provided 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.<sup>21</sup> The latter assumption can be justified 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, we 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 Syrians 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,  $p_j^{-1}$ , where the proportion is defined as

$$p_j = \frac{\sum_{i=1}^N i_j}{\sum_{i=1}^N w_{i,j}}, \quad (5)$$

and  $N$  represents the total number of observations in the sample. In addition, because 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 we add a correction factor that makes the sum of the weighted sample add up to the official figure of Syrian refugees as of July 2017,  $\bar{S}$ .<sup>22</sup> Survey weights are defined by

<sup>19</sup> See [https://ec.europa.eu/regional\\_policy/en/policy/what/glossary/n/nuts/](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.

<sup>20</sup> It should be noted that standard errors will increase as a result of the non-response 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.

<sup>21</sup> Even though these figures are published at the NUTS-3 level (provinces), we disregard the provincial distribution because we suspect Syrian refugees have incentives to re-distribute themselves within Turkey to areas with a higher number of job opportunities, for instance, areas like Bursa or İstanbul.

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

multiplying the inverse proportion of being selected in a particular subregion by the adjustment factor as follows,

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

where  $S$  is the number of Syrian refugees in the sample. It should be noted that, in practice, the correction factors divide the actual population of Syrians by the expanded population of Syrians that 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 B.1.

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 be found in Table C.1 (Appendix C) at the subregional level (NUTS-2), the lowest level of geographical disaggregation allowed 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 İstanbul, Bursa and Konya. This pattern, which could be the natural result of refugees' job search efforts, can be visualized with the help of maps in Figure C.1 (official distribution), C.2 (HLFS distribution) and C.3 which shows the difference between the official and the HLFS-estimated refugees' geographical distribution.

## 4 Conclusion

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 these groups down. This article proposes the use of the Turkey Household Labour Force Survey to overcome the information deficit with regards Syrians in Turkey. In particular, I propose an indirect identification strategy to isolate Syrian refugees from other 'regular' migrants, since both are grouped together in the publicly available microdata.

The identification strategy produces a population pyramid for HLFS refugees that is comparable to the age profile recorded by the Turkish Directorate General for Migration Management. This should allow researches and authorities alike to produce more accurate investigations and to build better-informed policies aimed at the Syrian population. In addition, I show that Syrian refugees might have internally migrated from south-eastern provinces bordering with Syrian to more industrialized areas of Turkey like Bursa, Konya or İstanbul. This pattern of internal migration would need to be confirmed by other instruments yet it suggests that a re-allocation of funds and humanitarian efforts might be due.



## References

- Alsharabati, C.; Nammour, J. 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.
- Del Carpio, X.V.; Wagner, M. 2015. *The impact of Syrian refugees on the Turkish labor market*, Policy Research Working Paper 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, November 27-29<sup>th</sup>.
- Kaymaz, T.; Kadboy, O. 2016. *Syrians in turkey - the economics of integration*, Alsharq Forum, Expert Brief.
- Konun, B.B.; Tümen, S. 2016. *Immigration and prices: Quasi-experimental evidence from Syrian refugees in Turkey*, Central Bank of Turkey Working Paper No 16/01.
- 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", in *American Economic Review*, Vol. 106, No. 5, pp. 456-460.
- Uçak, S.; Raman, K. 2017. *Another side of the story: A market assessment of syrian smes in turkey*, Building Markets.
- Yavçan, B. 2017. "On governing the syrian refugee crisis collectively: The view from turkey.", in *Near Futures Online*, 1, 'Europe at a crossroads' March 2016.



## Appendix A: Derived variables from the HLFS

**Number of children aged 0-14.** The data provided in the HLFS by the Turkish Statistical Institute only refers 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.

**Year of arrival to Turkey of persons born abroad.** One the questions available in the HLFS asks respondents whether they 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 the first paragraph of Appendix A. For most<sup>23</sup> 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.

**Table A1: Construction of the variable ‘year of arrival to Turkey’**

	<i>Variable</i>	<i>Sub-population</i>	<i>Obs. HLFS 2017</i>
Year of arrival	<code>buil_yil</code>	<code>dogum_yer=Abroad</code>	8,310
		<code>buil_yasama=No</code> <sup>24</sup>	
		<code>onceki_ikamet=Abroad</code>	
	<code>tr_yil</code>	<code>dogum_yer=Abroad</code>	1,377
		<code>buil_yasama=No</code>	
		<code>onceki_ikamet=Turkey</code>	
		<code>yurtdisi_durum=Yes</code>	

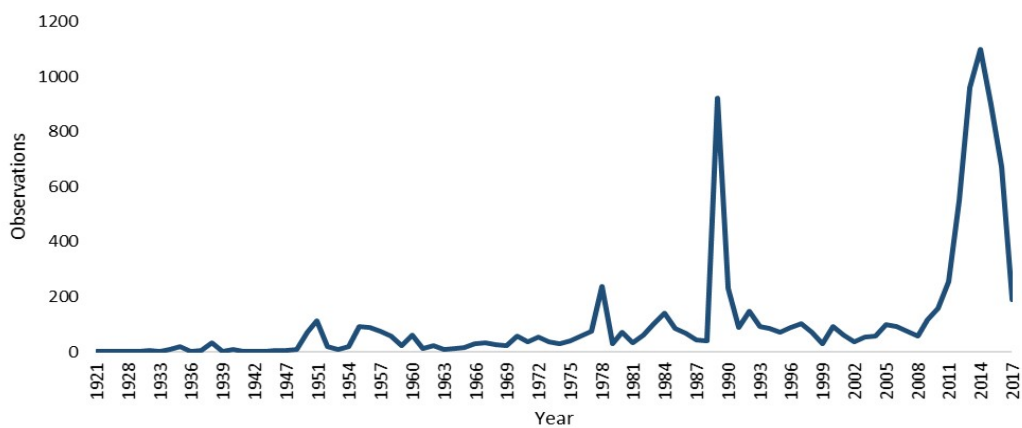
The full list of logical skips used to build the year of arrival is shown in Table A1, where the column ‘sub-population’ presents the conditions that respondents need to fulfill for their year of arrival to come from either of the two options. It should be noted that the condition referring to the variable `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 a second one after 2011, right after the Syrian civil war.

<sup>23</sup> 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 that conform the group of foreign-born who have changed provinces within Turkey at least once in their lives.

<sup>24</sup> 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



## Appendix B: Post-stratification weights

**Table B.1: Re-weighting by subregion (NUTS-2), 2017 HLFS**

<i>Subregion</i>	<i>Provinces</i>	<i>Turkish</i>	<i>Syrian refugees</i>	
		<i>Adj. factor</i>	<i>Inv. prob.</i>	<i>Adj. weight</i>
İstanbul	İstanbul	1.0203	338	1,511
Tekirdağ	Tekirdağ, Edirne, Kırklareli	1.0024	109	485
Balıkesir	Balıkesir, Çanakkale	1.0016	96	428
İzmir	İzmir	1.0078	196	876
Aydın	Aydın, Denizli, Muğla	1.0086	157	700
Manisa	Manisa, Afyon, Kütahya, Uşak	1.0039	171	765
Bursa	Bursa, Eskişehir, Bilecik	1.0212	195	871
Kocaeli	Kocaeli, Sakarya, Düzce, Bolu, Yalova	1.0044	188	838
Ankara	Ankara	1.0071	192	856
Konya	Konya, Karaman	1.0442	83	370
Antalya	Antalya, Isparta, Burdur	1.0082	166	743
Adana	Adana, Mersin	1.0293	183	817
Hatay	Hatay, Kahramanmaraş, Osmaniye	1.0141	187	837
Kırıkkale	Kırıkkale, Aksaray	1.0239	85	378
	Niğde, Nevşehir, Kırşehir			
Kayseri	Kayseri, Sivas, Yozgat	1.0085	160	716
Zonguldak	Zonguldak, Karabük, Bartın	1.0007	110	490
Kastamonu	Kastamonu, Çankırı, Sinop	1.0032	55	246
Samsun	Samsun, Tokat, Çorum, Amasya	1.0020	139	620
Trabzon	Trabzon, Ordu, Giresun	1.0003	120	552
	Rize, Artvin, Gümüşhane			
Erzurum	Erzurum, Erzincan, Bayburt	1.0013	75	335
Ağrı	Ağrı, Kars, Iğdır, Ardahan	1.0010	77	343
Malatya	Malatya, Elazığ, Bingöl, Tunceli	1.0015	120	537
Van	Van, Muş, Bitlis, Hakkâri	1.0000	92	411
Gaziantep	Gaziantep, Adıyaman, Kilis	1.0264	144	642
Şanlıurfa	Şanlıurfa, Diyarbakır	1.0126	177	791
Mardin	Mardin, Batman, Şırnak, Siirt	1.0020	159	711

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 add up in the table due to rounding.

## Appendix C: Geographical distribution of Syrian refugees

Table C.1: Syrian refugees under temporary protection in Turkey by subregion (NUTS-2), 2017

<i>Subregion</i>	<i>Provinces</i>	<i>Syrian refugees in Turkey</i>		
		<i>Official</i>	<i>HLFS estimated</i>	<i>Difference</i>
İstanbul	İstanbul	497,135	1,093,701	596,565
Bursa	Bursa, Eskişehir, Bilecik	113,989	270,691	156,702
Konya	Konya, Karaman	76,744	201,085	124,341
Adana	Adana, Mersin	308,641	378,422	69,782
Kırıkkale	Kırıkkale, Aksaray	15,055	68,110	53,054
	Niğde, Nevşehir, Kırşehir			
Aydın	Aydın, Denizli, Muğla	26,483	74,111	47,628
Antalya	Antalya, Isparta, Burdur	15,438	51,517	36,079
Manisa	Manisa, Afyon, Kütahya, Uşak	13,554	49,325	35,771
Ankara	Ankara	75,881	96,429	20,547
Kocaeli	Kocaeli, Sakarya, Düzce, Bolu, Yalova	46,533	63,759	17,226
Samsun	Samsun, Tokat, Çorum, Amasya	7,579	22,671	15,092
Kastamonu	Kastamonu, Çankırı, Sinop	1,449	9,719	8,270
Erzurum	Erzurum, Erzincan, Bayburt	950	8,052	7,103
Ağrı	Ağrı, Kars, Iğdır, Ardahan	1,333	4,287	2,954
Zonguldak	Zonguldak, Karabük, Bartın	883	3,300	2,417
Balıkesir	Balıkesir, Çanakkale	6,222	8,235	2,013
Tekirdağ	Tekirdağ, Edirne, Kırklareli	15,719	14,525	-1,194
	Trabzon, Ordu, Giresun			
Trabzon	Rize, Artvin, Gümüşhane	4,027	3,191	-836
	Van, Muş, Bitlis, Hakkâri			
Van	Van, Muş, Bitlis, Hakkâri	4,662	0	-4,662
İzmir	İzmir	112,881	103,718	-9,163
Malatya	Malatya, Elazığ, Bingöl, Tunceli	29,689	12,290	-16,769
Kayseri	Kayseri, Sivas, Yozgat	67,207	48,966	-18,241
Mardin	Mardin, Batman, Şırnak, Siirt	136,673	12,322	-124,352
Gaziantep	Gaziantep, Adıyaman, Kilis	497,371	210,195	-287,176
Şanlıurfa	Şanlıurfa, Diyarbakır	466,811	134,796	-332,015
Hatay	Hatay, Kahramanmaraş, Osmaniye	536,986	134,500	-402,486

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 C.1: Number of Syrian refugees under temporary protection by subregion, DGMM July 2017

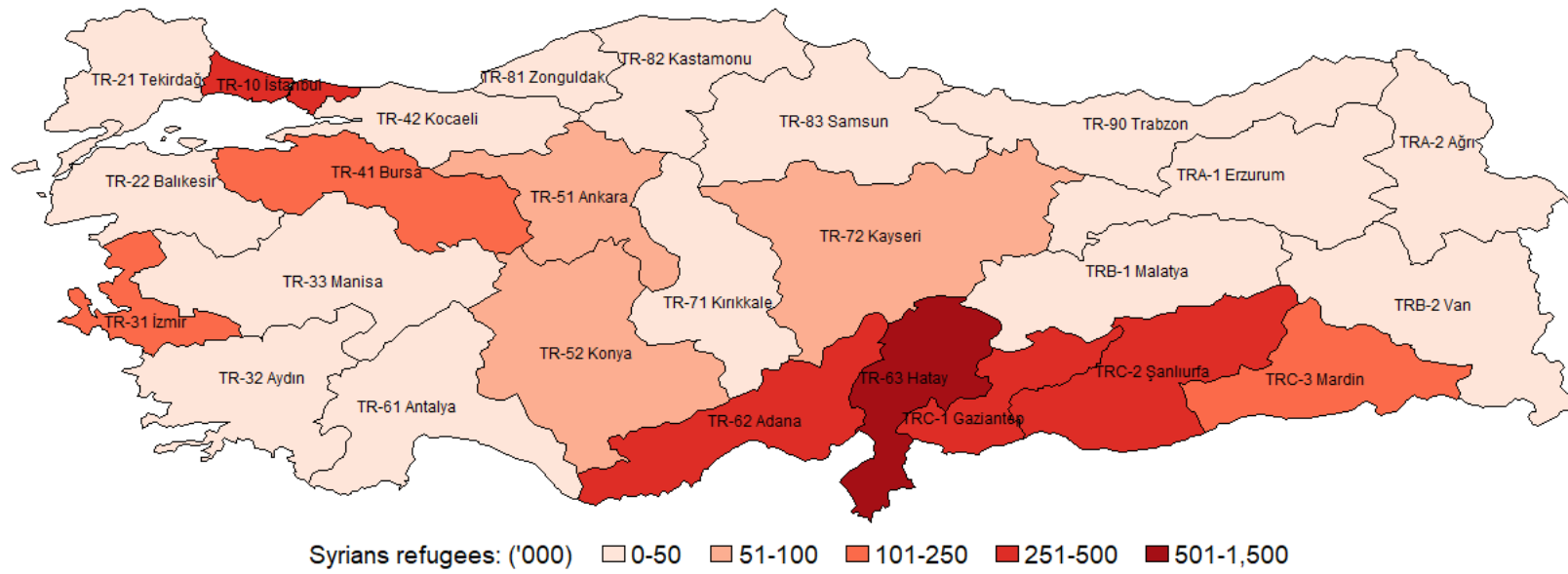


Figure C.2: Number of Syrian refugees under temporary protection by subregion, HLFS 2017

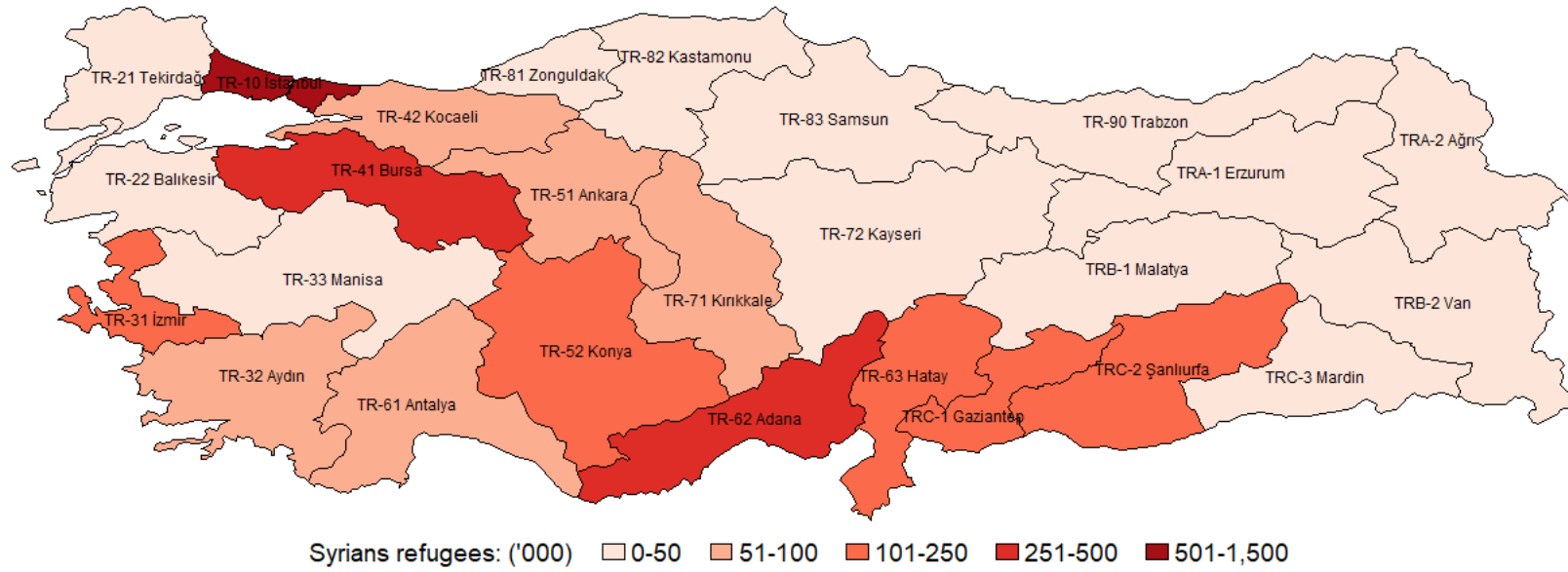


Figure C.3: Difference in Syrian refugees under temporary protection by subregion, HLFS vs. DGMM

