A Practitioner's Guide to Handling Irregularities Resulting from the 2014 Revisions to the Turkish Household Labor Force Survey

Murat Demirci[*]
Koç University

Meltem Poyraz[**]
Boğaziçi University

Abstract

We document the implications of the 2014 revisions to the Turkish Household Labor Force Survey and offer guidance on how to handle the irregularities in population and unemployment statistics that resulted from two particular revisions. First, new population projections were adopted to assign survey weights. Second, a narrower definition of unemployment was adopted. We propose methods to adjust the survey weights for the pre-2014 period in order to discern changes in population statistics by age groups and regions without interruption over time and to calculate the unemployment rates according to both broader and narrower definitions since 2004.

Keywords: Household Labor Force Survey, Survey Methods, Unemployment Statistics, Population Statistics
JEL Classification: C83, J10, J64

Türkiye Hanehalkı İşgücü Anketi 2014 Revizyonlarından Kaynaklanan Düzensizliklerin Ele Alınmasına Yönelik Bir Uygulayıcı Rehberi

Özet

Bu çalışmada TÜİK Hanehalkı İşgücü Anketi'ninde 2014 yılında yapılan düzenlemelerden kaynaklanan bazı sonuçları bulguluyor, nüfus ve işsizlik istatistiklerinde yapılan revizyonların özellikle ikisinden kaynaklanan sorunların nasıl ele alınması gerektiği hakkında yol gösteriyoruz. İlk revizyon, anket ağırlıklarının dayandırıldığı yeni nüfus projeksiyonlarının kabul edilmesidir. İkincisi ise, dar tanımlı bir işsizlik tanımına geçilmesidir. Bu çalışmada 2014 öncesindeki anket ağırlıklarını nüfus istatistiklerinde yaş bazında ve bölgesel olarak kopma olmayacak şekilde yeniden ayarlamak ve işsizlik oranlarını geniş ve dar tanımlı olarak 2004 yılından itibaren hesaplamak için yöntemler öneriyoruz.

Anahtar Kelimeler: Hanehalkı İş Gücü Anketi, Anket Yöntemleri, Nüfus İstatistikleri, İşsizlik İstatistikleri
JEL Sınıflandırması: C83, J10, J64

[*] Murat Demirci. Department of Economics, Koç University, Sarıyer, Istanbul 34450 Turkey. E-mail: mudemirci@ku.edu.tr. ORCID: 0000-0002-3255-2809

[**] Meltem Poyraz. Department of Economics, Boğaziçi University, Bebek, Istanbul 34342 Turkey. E-mail: meltem.poyraz@boun.edu.tr. ORCID: 0000-0001-5779-0259

L abor force surveys have been widely used by researchers and policy makers to analyze various dimensions of labor markets.[1] Institutions responsible for conducting such surveys periodically revise survey items or sampling methodologies with the purpose of improving accuracy.[2] However, such revisions raise the risk of generating inconsistencies across different waves of a survey. In this paper, we explore potential irregularities that stem from the 2014 revisions to the Turkish Household Labor Force Survey (THLFS) and propose methods to deal with them.

The first major revision in 2014 was the change in the underlying population projections that have been used to assign survey weights to observations. Tabulations based on the sum of survey weights show that the population of older people and the populations of NUTS-2 regions display abnormal changes around 2014, the first year the new population projections were used. Incorrect assumptions about death rates or inter-regional migration patterns in the previous population projections are likely to be the main reasons for the observed anomalies in population statistics. We propose an alternative method of assigning survey weights for the pre-2014 period with the purpose of achieving a reasonable reflection of annual changes in the population statistics by age groups and by regions for each gender. To achieve this, we make two assumptions: i) the death rates utilized for the elderly in the new population projections can be applied to the elderly population of the earlier period, and ii) the inter-regional migration rates follow a linear trend within short periods of time.

Another major revision in the implementation of the THLFS in 2014 was the change in the criterion for the classification of unemployment. As of 2014, individuals who searched for a job in the previous three months but not in the previous four weeks are no longer classified as unemployed.[3] The adoption of the narrower definition of unemployment has led to an underestimation of the number of unemployed people in the post-2014 period compared to earlier years. In this study, we propose a method to eliminate the break that occurred in unemployment statistics in 2014 due to this change in the definition of unemployment. For that purpose, we employ the 2009-2013 waves of the THLFS that report the information on job search behavior within the previous four weeks and also within the previous three months. Using this information, we estimate the parameters of a model that explains job search behavior in the previous three months but not in the previous four weeks. We then use these estimates to infer the population of unemployed people and the unemployment rate based on the job search criterion in the previous three months for the post-2014 period. We also adopt a similar strategy to predict the unemployment statistics based on the criterion of searching for a job in the previous four weeks for the pre-2009 period.

---

[1] For instance, researchers have used such surveys in Turkey to analyze female labor force participation (Dayioğlu and Kirdar 2010, Karaoglan and Okten 2015, Tunali et al. 2019), the impact of immigration (Aydemir and Kirdar 2017a, Ceritoglu et al. 2017, Akgündüz and Torun 2020, Araci et al. 2021), the effect of education on earnings (Aydemir and Kirdar 2017b, Ozturk and Tumen 2018, and Patrinos et al. 2020), determinants of educational attainment (Polat 2017, Caner et al. 2019, Demirci and Poyraz 2021), and wage inequality (Bakis and Polat 2015).

[2] There are two prominent examples of such revisions and how researchers handle each of them in the U.S. context. First, the number of computer-related occupations in labor force surveys has increased noticeably over time. Meyer and Osborne (2005) developed a consistent classification of occupation codes that appear in different waves of the labor force surveys. Second, the coding of educational attainment has changed over time. In particular, the number of completed schooling years was used until 1992 for this measurement, whereas the type of degrees attained has been used since then. Jaeger (1997) has proposed a method to convert the years of schooling to degree categories. To do so, he uses the fact that the survey collects information on educational attainment both in terms of schooling years and degree categories. In this paper, we adopt a similar strategy by using the information related to job searches in the previous three months and in the previous four weeks in the 2009-2013 THLFS.

[3] Researchers had to conduct their analysis with shorter time series because of this break in unemployment statistics. For instance, only the data from the pre-2014 period were used by Gil-Alana et al. (2019) to analyze the persistence of unemployment and by Demirci and Poyraz (2021) to analyze the impact of high unemployment rates on school enrollment of youth in Turkey.

We also provide researchers with practical guidance on how to use our suggested weights and the extended unemployment series to mitigate the irregularities that stem from the 2014 revisions to the THLFS. First, the suggested weights in this paper highlight the noticeable underestimation of the elderly population in the earlier waves of the THLFS. In particular, tabulations that are based on the suggested weights imply that the population of people older than 65 was underestimated by 5.7% in the 2013 THLFS. We therefore recommend that researchers use the suggested weights of this paper in their studies analyzing the statistics of the elderly. Second, we highlight that when the original weights are used there are abnormalities in the evolution of regional population statistics of the individuals aged 15 or older around 2014. For instance, the population of the Istanbul region increased by 9.3% from 2013 to 2014 according to the original weights provided in the THLFS, whereas it increased by 2.5% according to the weights suggested in this paper. The latter figure is more in line with the annual growth rate of the population of the Istanbul region over the 2004-2019 period, which is 1.9%. We consequently advise researchers to use our suggested weights in their robustness checks, particularly if their studies rely on regional variation, which is highly correlated with the abnormalities in the regional statistics that our methodology points out. Third, our results show that the normalized versions of the statistics appear to be less affected by the revisions in 2014. For example, our method shows that the number of employed people in the Istanbul region was underestimated by 4.2% in 2013 if the original weights are used, whereas the normalized version of this statistic (i.e., the employment rate) has differed by less than 1 percentage point across the two weighting schemes considered. Thus, researchers who prefer to use the original weights should work with the normalized versions of the statistics. Fourth, the change in the definition of unemployment created a break in the time series of unemployment statistics in 2014. Our analysis of the unemployment statistics indicate that the number of unemployed people (the unemployment rate) would be 10.2% (1.2 percentage points) larger in 2014 if the criterion of a job search in the previous three months had been used to define unemployment. We recommend our predicted unemployment series to researchers who need to use a longer time series of unemployment statistics.

This paper proceeds as follows: The second section discusses the revisions adopted in 2014 in detail. The third section describes the methodology we use to obtain the suggested survey weights, and the fourth reports the impact of using the suggested weights on population and employment statistics. The fifth describes the model we use to predict unemployment statistics based on the criteria of a job search either in the previous three months or in the previous four weeks and discusses the resulting unemployment series. We conclude in the sixth section.

The 2014 Revisions to the Turkish Household Labor Force Survey

The Turkish Statistical Institute (TURKSTAT) conducts the THLFS as repeated cross-sectional surveys to collect detailed information on the economic activities of individuals aged 15 and over. The sampling design of the THLFS is rotating. The same address is visited once in each of two consecutive quarters, left out in the following two quarters, and visited again in the fifth and the sixth quarters. A survey weight is assigned to each individual in the THLFS such that the sample would be representative of each NUTS-2 region since the 2004 wave. Since we analyze regional statistics in this study, we focus on the post-2004 period.

In 2014, TURKSAT implemented five major revisions to the THLFS in accordance with the framework of the European Union criteria (TURKSTAT, 2014). These changes can be summarized as follows:

i)   The reference period of the survey was distributed among all 52 weeks of the year, while it had only been the first week of each month before 2014.

ii)   The sampling design was adjusted to ensure that 50% of visited addresses would be the same across the two consecutive quarters of the survey in a given year and across the same quarters of the two subsequent years.

iii)   The sampling frame was adjusted to account for a new administrative classification adopted by the Turkish government, which expanded the number of greater municipalities.

iv)   Population projections that have been used to assign survey weights were revised.

v)   The reference period of a job search for an individual to be classified as unemployed was changed from the last three months to the last four weeks.

The first two revisions were adopted to capture the seasonality in labor market statistics more accurately. These revisions would generate inconsistencies in labor market statistics only if employment or job search activities in the first week of the month were different than those during the rest of the month. We believe that there is no reason for such a systematic difference to exist over the weeks of a month. Also, since neither the month nor the week of the survey is provided in the THLFS micro data, we cannot explore the implications of these revisions in this study.

The Turkish government changed the administrative status of thirteen provinces to that of "greater municipality" by law in 2012 (Official Gazette, 2012). The third revision makes the sampling frame of the THLFS consistent with this new administrative classification. Most villages that were classified as rural areas before 2014 were considered urban areas within greater municipalities after the revision. Thus, the percentage of people residing in urban areas increased from 72.3% in 2013 to 86.5% in 2014 (TURKSTAT, 2014). The analysis of employment statistics by rural or urban areas over time became unfeasible due to this compositional change. Indeed, TURKSTAT stopped reporting the variable indicating the urban residency status of individuals in 2014.

TURKSTAT employs population projections to assign a survey weight to each observation in the data. These population projections take the most current knowledge regarding the civilian population in Turkey for a particular year. They then rely on certain assumptions about death rates and inter-regional migration patterns to infer the population statistics by age and regions for the following years. As the fourth revision above states, TURKSTAT started to use updated population projections in 2014 and acknowledged the underestimation of the population of older people in the earlier waves of the THLFS (TURKSTAT, 2014). As an illustration of this underestimation, in the left panel of Figure 1 we plot the population of people older than 65 during the 2004-2019 period based on the original weights provided in the THLFS data. There is a jump in the population of older people of each gender in 2014, which suggests an underestimation of their population before 2014. This underestimation in the earlier waves of the THLFS is likely to stem from the overestimation of death rates in the previous population projections that were used to assign survey weights.[4]

There have been some earlier attempts by TURKSTAT and other researchers to deal with the implications of the fourth revision by providing revised time series of population and labor market statistics for the pre-2014 period that are consistent with the post-2014 period. However,

---

[4] After the establishment of the Address Based Population Registration System (ABPRS) in 2007, TURKSTAT started using new population projections. Population statistics by age, gender, and regions based on ABPRS differed from those based on the earlier projections. Yet we do not observe a break in the population statistics around 2008 in the THLFS data because TURKSTAT revised micro data sets between 2004 and 2008 to obtain comparability over time (ILO, 2016).

these studies have focused on the statistics at the national level but not for smaller demographic groups. In particular, TURKSTAT (2015) released a revised population series of all residents aged 15 or above for the 2005-2013 period; however, they did not provide the revised population statistics by age group. According to the revised population series, the population of civilians aged 15 or above were underestimated by 376 thousand in the 2013 THLFS. This problem was less evident in the 2005 THLFS as the extent of underestimation decreases to 3 thousand individuals. TURKSTAT (2015) also revised the series of employment, unemployment and labor force participation for the 2005-2013 period to make them consistent with the post-2014 data. Another study, Uysal and Kavuncu (2018), inferred employment statistics by gender for the pre-2014 period by applying the growth rates of the associated series realized during the post-2014 period to the earlier periods. We contribute to this discussion by generating the suggested survey weights in the individual-level data, which can be used by researchers to calculate population or employment statistics for any group of interest. We also contribute to this discussion by highlighting the anomalies in regional statistics, which our suggested weights also help to correct. As an example of the anomalies at the regional level, in the right panel of Figure 1 we display the population of people who live in the Istanbul region, the most developed and the most populated region of Turkey. As shown in the figure, the population of Istanbul displayed a noticeable increase in 2014. A potential reason for this abnormal jump is the underestimation of migration to Istanbul in the previous population projections that were used to assign survey weights in the earlier waves of the THLFS.

Figure 1   Population of Older People and Population in Istanbul
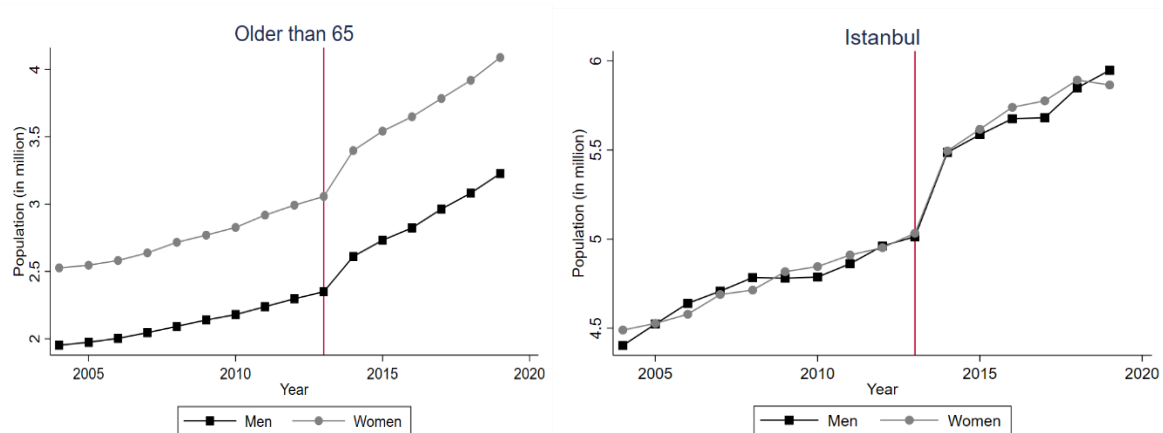
Notes: The population statistics in these figures are calculated as the sum of the original survey weights provided in the THLFS data across all individuals in the specified group of interest. The vertical line indicates the year 2013, the last year before the revisions were adopted in the THLFS.

The fifth revision in 2014 was the change in the criterion of classifying individuals as unemployed. Individuals who searched for a job in the previous three months were counted as unemployed before the 2014 THLFS, whereas only individuals who searched for a job in the previous four weeks were counted as unemployed in the 2014 and later waves of the THLFS. Thus, a narrower sample of individuals has been considered as unemployed since 2014, which makes the comparison of unemployment statistics over time unfeasible. To understand the potential effects of this revision, we compare the population of individuals who searched for a job in the previous four weeks to those who searched for a job at any time in the previous three months,

which were made available separately in the 2009-2013 THLFS. The population of the former was 2 million 520 thousand in 2013, whereas the population of the latter was 2 million and 849 thousand. These tabulations suggest that there would be about a 12% difference in the unemployment rate between the narrower and broader definitions of unemployment. In this study, we provide a method to obtain the unemployment rates based on each unemployment criterion without a break over the whole 2004-2019 period.

Methodology for Calculating Suggested Survey Weights

As highlighted in the previous section, there are discontinuities in the population statistics by age and by regions in the THLFS data in 2014. In this section, we propose a method to adjust survey weights for the pre-2014 period such that the resulting population statistics will change continuously around the revision year 2014. Let cohort gc define people of gender g who belong to birth cohort c and P_{gcrt} denote the population of cohort gc who reside in NUTS-2 region r at time t. P_{gcrt} is calculated as the sum of the original survey weights provided in the THLFS for the people of cohort gc who live in region r at time t. Our aim is to predict an alternative to P_{gcrt}, namely $\hat{P}_{gcrt}$, that does not show irregularities around the year 2014.

The outline of our strategy to obtain $\hat{P}_{gcrt}$ is as follows: we take the population of civilians aged 15 or above as given by the revised population statistics reported by TURKSTAT (2015), then we apply the following steps separately for each gender: (i) determine cohorts who are older than age 65 in the pre-2014 period, and then apply the death rates observed in the new population projections to infer the population of these older cohorts in each region in the pre-2014 period, (ii) calculate the total population of all young cohorts in the pre-2014 period by subtracting the total population of all older cohorts from the revised population statistics of TURKSTAT (2015), (iii) calculate the population of each younger cohort at the national level in the pre-2014 period according to their distribution observed in the original THLFS data, and (iv) allocate the population of these younger cohorts to each region by assuming a linear time trend in inter-regional migration.

We apply this method first for the 2008-2013 period and then for the 2004-2007 period to maintain consistency of population statistics over time. Finally, we calculate the adjustment factors associated with survey weights for the people of each cohort gc who live in region r at time t as the ratio of $\hat{P}_{gcrt}$ to P_{gcrt} for all years in the 2004-2013 period. The details of our methodology are discussed below.

Older cohorts in the 2008-2013 period

We define birth cohorts at 5-year intervals. In particular, the birth cohorts correspond to groups of people who are aged 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84 or above 85 in 2013. Let index o denote the cohorts of people who were aged 65 and above in 2008 (i.e., those aged 70 and above in 2013) and let Y_{gort} be the natural logarithm of the population of people of gender g in the older cohort o (i.e., the cohort go) who live in region r at time t.

To infer the death rates of older cohorts, we first estimate the following equation with the 2014-2019 THLFS data separately for each older cohort and gender by pooling data from all regions

$$(1) \qquad\qquad Y_{gort} = \delta_{gor} + \alpha_{go} T_t + \varepsilon_{gort}$$

where δ_{gor} denotes the fixed effects associated with each gender, birth cohort, and region, T_t is the linear time trend (calculated as t-2013), and ε_{gort} denotes the error term. The key coefficient of interest in this regression is α_{go}, which shows the nation-wide death rate of the cohort go.

Next, we estimate the following equation with the 2014-2019 THLFS data separately for each older cohort, gender, and region

$$(2) \qquad\qquad Y_{gort} = \gamma_{gor} + \beta_{gor} T_t + e_{gort}$$

where γ_{gor} denotes the constant associated with the cohort go in region r, T_t is the linear time trend (calculated as t-2013), and e_{gort} denotes the error term. Using the estimates of this equation, we predict the logarithm of the population of the cohort go in region r in 2014 as $\hat{Y}_{gor,2014} = \hat{\gamma}_{gor} + \hat{\beta}_{gor}$.[5] Then, we infer the population of the cohort for each year t in the 2008-2013 period as

$$(3) \qquad\qquad \hat{P}_{gort} = \exp[\hat{Y}_{gor,2014} + \hat{\alpha}_{g\tilde{o}}(t - 2014)]$$

where $\tilde{o}$ refers to the birth cohort which is 5-years younger than the birth cohort o.

Thus, the main assumption made with this inference is that the death rate of a particular age group in the 2008-2013 period is not different than the death rate of the same age group observed in the 2014-2019 period. This assumption may not hold especially in the long-term because death rates faced by a particular age group might decline over time. However, since we analyze a short period of time in this study, we do not expect significant changes in death rates. We also do not model the inter-regional migration of the elderly people. Since the migration rate of older people is presumably negligible, we believe that this modelling choice is innocuous.

As an illustration of our methodology, in Figure 2, we plot the observed and predicted population of men who were born between 1934 and 1938 (i.e., those aged between 70 and 74 in 2008) and who currently reside in the Izmir region. We first obtain the predicted value of the population of this birth cohort for the year 2014 from the linear fit obtained from the post-2014 data. Then, starting from this predicted value, we extend the population series until 2008 by applying the observed death rate for this age group in the post-2014 period.

Finally, we calculate the total population of all older cohorts in each year t between 2008 and 2013 as

$$(4) \qquad\qquad P_t^{old} = \sum_g \sum_o \sum_r \hat{P}_{gort}.$$

---

[5] Although we can observe the population of people of each cohort, gender, and region directly from the 2014 THLFS, we prefer to use the predicted populations that rely on the trend observed for the whole 2014-2019 data because a population figure from only one year might be subject to large measurement errors.

Figure 2   An Illustration of Our Method of Predicting the Population of the Elderly
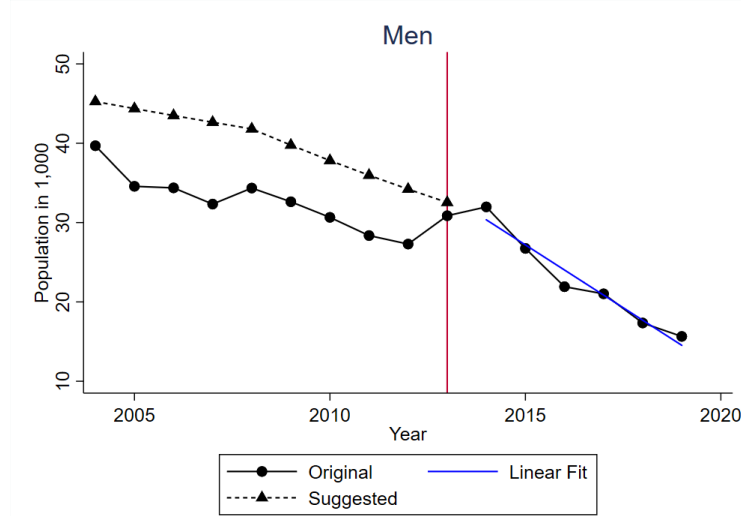
Notes: The 2004-2019 THLFS data. The population series labeled as "original" is calculated as the sum of the original survey weights across all men who were born between 1934 and 1938 and resided in the Izmir region at the time survey was conducted. The "linear fit" plots the predicted population of the corresponding group over the 2014-2019 period. To plot the dotted line, we take the predicted value for the year 2014 from the linear fit and extend it until 2008 by applying the national-level death rate associated with the age group that the 1934-38 birth cohort belonged to during the 2008-2013 period. Then, we take the predicted value for the year 2008 from the dotted line and extend it until 2004 by applying the national-level death rate associated with the age group that the 1934-38 birth cohort belonged to during the 2003-2008 period. The vertical line indicates the year 2013.

Younger cohorts in the 2008-2013 period

Let P_t^{young} be the total population of all younger cohorts (i.e., younger than 65 in 2008) at time t. We calculate P_t^{young} by subtracting P_t^{old} from the revised statistics reported by TURKSTAT (2015) for the civilian population who are aged 15 and older. Once we obtain P_t^{young}, we then infer the population of each young cohort in region r at time t as described below.

Let index y denote cohorts of people who were younger than 65 in 2008. Let $f_t(gy)$ denote the fraction of individuals of gender g in birth cohort y (i.e., cohort gy) among all individuals of younger cohorts at time t. We calculate the values of $f_t(gy)$ based on the original weights provided in the 2008-2013 THLFS data. Since the death rates are almost zero for these younger cohorts, we assume that this way of calculating $f_t(gy)$ does not suffer from the miscalculation of death rates as in the previous population projections and is therefore accurate.

Next, we need to determine how individuals from each young cohort are distributed over NUTS-2 regions. Let $s_{gyt}(r)$ denote the fraction of individuals who live in region r at time t among all individuals of cohort gy. We calculate $s_{gyt}(r)$ for the years 2008 and 2014 based on the original weights in the THLFS. Then, we linearly interpolate the values of $s_{gyt}(r)$ for the years between 2008 and 2014 for each cohort gy and region r. The underlying assumption behind this method is that the inter-regional migration rates followed a linear trend over the 2008-2014 period. Although the migration rate between regions might be affected by local labor market conditions, capturing this dimension requires a more sophisticated model. We believe that our assumption of

a linear trend is innocuous given the short period of time being studied. Thus, we simply employ a linear interpolation instead of a more sophisticated model.

Lastly, we infer the population of the cohort gy in region r for each year t in the 2008-2013 period as

$$(5) \qquad\qquad \hat{P}_{gyrt} = s_{gyt}(r) \; f_t(gy) \; P_t^{young}.$$

The 2004-2007 period

After obtaining the population of each cohort in the 2008-2013 period, we proceed to infer the population of each cohort in the 2004-2007 period by applying the same method. In practice, we redefine birth cohorts as groups of people who are aged 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84 or above 85 in 2008. For the cohorts who are aged 65 and above in 2003 (i.e., those aged 70 and above in 2008), we then infer the population of each cohort by using the data from the 2008-2013 period. In particular, we estimate the parameters of the models in Equation (1) and (2) with the data from the 2008-2013 period and use these estimates in Equation (3) to infer the population statistics of older cohorts in each year of the 2004-2007 period. It is important to highlight that the population statistics that we use at this stage for the older cohorts of the 2008-2013 are predicted by the method described in the *Older cohorts in the 2008-2013 period* subsection to enhance the consistency of statistics over time. As the example in Figure 2 shows, this method generates reasonable population series of the older cohorts without having a break around the year 2008. Next, we calculate the regional population of younger cohorts in the 2004-2007 period by first taking the national-level distribution of each cohort as observed in the original THLFS data and then by assuming linear inter-regional migration patterns between 2004 and 2008 as described in the *Younger cohorts in the 2008-2013 period* subsection.

Calculation of Suggested Survey Weights

Knowing P_{gcrt} and $\hat{P}_{gcrt}$ for individuals of each cohort gc in region r at time t allows us to calculate the associated adjustment factors as follows:

$$(6) \qquad\qquad adj\_factor_{gcrt} = \frac{\hat{P}_{gcrt}}{P_{gcrt}}.$$

We then adjust the survey weight for individual i who is in cohort gc in region r at time t as

$$(7) \qquad\qquad \hat{w}_i = w_i * adj\_factor_{gcrt}$$

where w_i denotes the original weights provided in the THLFS data.

Population and Employment Statistics with the Suggested Survey Weights

In this section, we provide summary statistics of the adjustment factors and analyze how population and employment statistics differ when the suggested weights, instead of the original weights, are used. In Table 1, we provide the average adjustment factors associated with each birth cohort and gender. The adjustment factors are noticeably larger for older cohorts and for men. For

instance, the results suggest that the original weights for individuals born before 1929 need to be multiplied by 1.378 for men and 1.189 for women, on average. On the other hand, the average adjustment factors are close to 1 for the cohorts of both genders born after 1939.

Table 1   Adjustment Factors for Survey Weights by Birth Cohorts and Gender

	Men	Women
Those born before 1929	1.378	1.189
	(0.821)	(0.507)
Those born between 1929-1933	1.247	1.138
	(0.345)	(0.260)
Those born between 1934-1938	1.084	0.971
	(0.196)	(0.170)
Those born between 1939-1943	1.014	1.033
	(0.141)	(0.124)
Those born between 1944-1948	1.010	1.005
	(0.123)	(0.108)
Those born between 1949-1953	1.005	1.006
	(0.107)	(0.101)
Those born between 1954-1958	0.999	1.006
	(0.090)	(0.074)
Those born between 1959-1963	1.007	0.998
	(0.076)	(0.080)
Those born between 1964-1968	0.993	0.999
	(0.056)	(0.061)
Those born between 1969-1973	0.991	0.996
	(0.064)	(0.061)
Those born between 1974-1978	0.996	0.990
	(0.061)	(0.065)
Those born between 1979-1983	1.001	0.996
	(0.073)	(0.056)
Those born between 1984-1988	1.010	1.007
	(0.086)	(0.071)
Those born between 1989-1993	0.997	1.002
	(0.072)	(0.074)
Those born between 1994-1998	0.989	0.993
	(0.086)	(0.091)

Notes: Each cell shows the average of the adjustment factors associated with the specified birth cohort and gender. The standard deviations across NUTS-2 regions and years within each group are provided in parentheses.

In Figure 3, we plot two population series for cohorts who were born before 1943: one calculated with the original weights (solid line) and the other with the suggested weights (dotted line). The population series with the suggested weights do not display any unreasonable change around the revision year of 2014. This shows that our methodology addresses one of the major problems that stem from the 2014 revisions. Moreover, the population of older men based on the suggested weights is considerably larger than the one based on the original weights for the years before 2014. This underestimation is observed over the whole 2004-2013 period. Our tabulations suggest that the population of older female cohorts is also underestimated for the years before 2014 but at a magnitude that is smaller and also diminishes towards 2004.

Figure 3   Population of the Elderly with the Original and the Suggested Weights
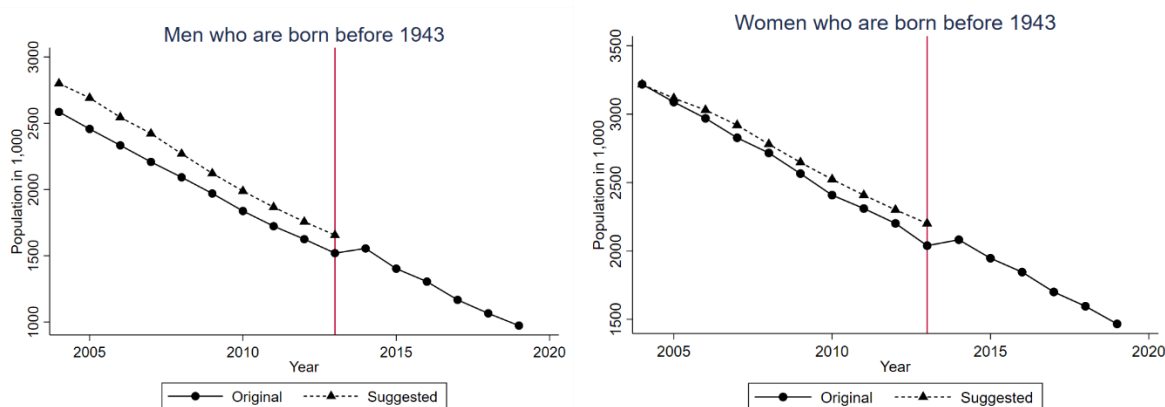
Notes: The 2004-2019 THFLS data. The solid lines are obtained as the sum of the original survey weights across all individuals in the specified group of interest, whereas the dotted lines are the sum of the survey weights suggested in this paper across the same group of individuals. The vertical line indicates the year 2013.

In Table 2, we provide the national-level aggregate employment statistics that are calculated separately with the original and the suggested weights. Employment statistics at the national level do not change to a great extent for both genders once the suggested weights are used. This result stems from the fact that the suggested weights associated with working-age people are only slightly different than the original weights. TURKSTAT (2014) also states that the revision of the population projections mostly affected the population of older people.

Table 2   National-Level Employment Statistics (in thousands)

	Employed Men			Employed Women		
Year	Original	Suggested	% Deviation	Original	Suggested	% Deviation
2004	14585	14547	0.26%	5047	5013	0.66%
2005	14959	14907	0.34%	5108	5075	0.65%
2006	15165	15144	0.14%	5258	5239	0.37%
2007	15382	15376	0.04%	5356	5347	0.17%
2008	15598	15616	-0.11%	5595	5595	0.01%
2009	15406	15414	-0.05%	5871	5857	0.24%
2010	16170	16225	-0.34%	6425	6435	-0.16%
2011	17137	17209	-0.42%	6973	7000	-0.38%
2012	17512	17523	-0.06%	7309	7307	0.02%
2013	17883	17915	-0.18%	7641	7657	-0.21%
2014	18244	x	x	7689	x	x

Notes: The table shows the population of employed individuals of each gender over time in thousands. The populations are calculated separately using the original survey weights provided in the THLFS data and the survey weights suggested in this paper. The last column of each panel shows the percentage deviation of the original statistics of interest from the corresponding statistics, which are calculated with the suggested survey weights.

Next, we calculate the population of individuals aged 15 and older in each NUTS-2 region separately with the original and the suggested weights. Table 3 shows the percentage deviations of the statistics using the original weights from the corresponding statistics using the suggested weights in the year 2013. Our results suggest a considerable degree of underestimation of the population in the Istanbul, Manisa, and Erzurum regions and overestimation of the population in the Edirne, Hatay, and Diyarbakır regions. In Figure 4, we plot the evolution of the populations of these NUTS-2 regions over time. The figure shows that the issues of overestimation and underestimation are apparent only in 2013 and in a few earlier waves of the THLFS.

Table 3   Deviations of the Original Statistics from the Suggested Statistics by NUTS-2 Regions

	Population		Employment		Employment Rate	
	Men	Women	Men	Women	Men	Women
Istanbul	-6.42	-6.06	-4.71	-4.03	1.82	2.17
Edirne	9.36	5.01	9.46	5.85	0.09	0.80
Balıkesir	-1.65	0.55	-1.45	0.43	0.20	-0.12
İzmir	0.34	-0.96	1.05	-0.17	0.70	0.80
Denizli	2.96	5.24	3.64	5.29	0.66	0.04
Manisa	-8.79	-5.76	-8.07	-5.81	0.79	-0.06
Bursa	3.27	4.73	1.91	3.91	-1.32	-0.78
Kocaeli	3.85	4.44	4.54	5.49	0.67	1.00
Ankara	-1.07	-2.11	-1.65	-2.76	-0.59	-0.67
Konya	0.23	-1.17	-0.62	-2.51	-0.84	-1.36
Antalya	-0.59	0.00	0.54	0.24	1.14	0.24
Adana	-1.09	-1.65	-0.59	-1.95	0.51	-0.31
Hatay	5.64	6.67	7.36	5.83	1.62	-0.79
Nevşehir	3.02	2.68	2.98	2.28	-0.04	-0.39
Kayseri	-3.55	-2.69	-1.90	-1.42	1.71	1.31
Zonguldak	0.62	2.46	2.04	1.58	1.41	-0.86
Kastamonu	2.69	3.89	-0.10	-0.36	-2.72	-4.09
Samsun	-3.80	-2.36	-2.71	-2.43	1.13	-0.07
Trabzon	5.48	-0.56	3.31	-1.09	-2.06	-0.53
Erzurum	-11.44	-8.25	-9.17	-3.50	2.56	5.17
Kars	-2.68	-1.32	-0.79	-0.05	1.94	1.29
Malatya	2.74	0.02	2.26	-1.53	-0.46	-1.55
Van	-2.58	1.25	-0.51	2.56	2.13	1.30
Gaziantep	2.23	-1.37	2.01	-0.14	-0.22	1.24
Diyarbakır	5.18	4.98	5.71	6.12	0.51	1.08
Siirt	1.77	3.54	4.41	7.21	2.59	3.54

Notes: The 2013 THFLS. Each cell shows the percentage deviations of the original statistics of interest from the corresponding statistics, which are calculated with the suggested survey weights.

Figure 4 Population Statistics by Regions with the Original and the Suggested Weights
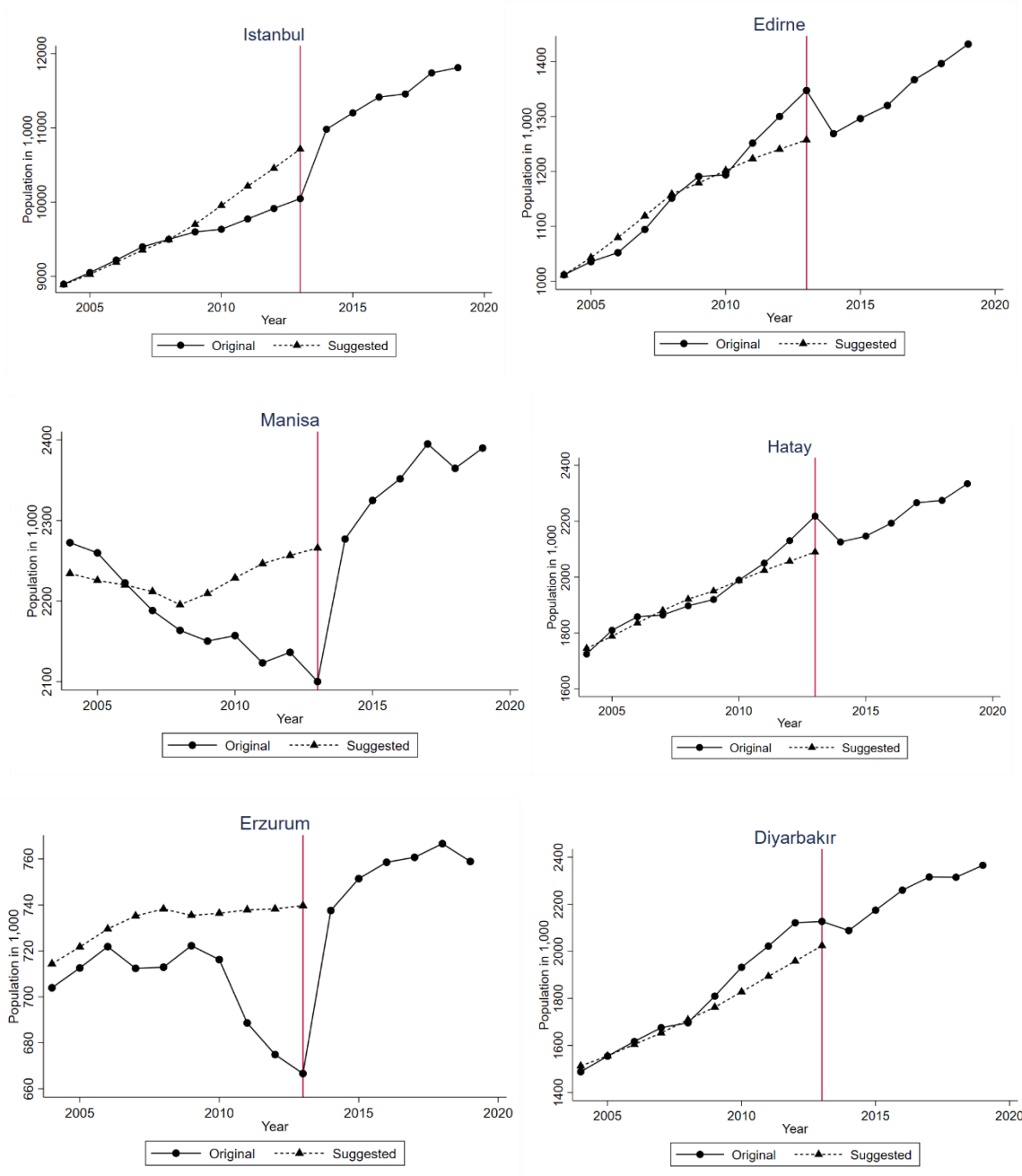
Notes: The 2004-2019 THFLS data. The solid lines are obtained as the sum of the original survey weights across all individuals who were aged 15 or older and resided in the specified region. The dotted lines are obtained as the sum of the survey weights suggested in this paper. The vertical line indicates the year 2013.

Lastly, in Table 3, we report the percentage deviation of the number of employed people in the NUTS-2 regions based on the original weights from the corresponding statistics based on the suggested weights. Although employment statistics do not change at the national-level, they differ noticeably at the NUTS-2 level after applying the suggested weights. The type of weights used matters particularly for the employment statistics of some regions, including Istanbul, as it

does for population statistics. Since both the population and employment statistics of regions change with the suggested weights, the employment rate (i.e., the ratio of employment to population) stays relatively stable. For instance, the average of the absolute difference between original and suggested population (employment) statistics in 2013 is 3.61 (3.21) percent for men and 3.06 (2.87) percent for women across all regions. On the other hand, the average of the absolute difference between the original and suggested employment rates is 1.17 percent for men and 1.21 percent for women.

Unemployment Statistics

Methodology

The 2009-2013 waves of THLFS data collected information on job searches both in the previous four weeks and in the previous three months. Thus, for the 2009-2013 period, it is possible to assign non-employed people into three groups: (i) those who searched for a job in the previous four weeks, (ii) those who searched for a job in the previous three months but not in the previous four weeks, and (iii) those who were not in the labor force (i.e., those who were neither employed nor searched for a job in the previous three months). With this information, we are able to generate the unemployment measure that is based on the broader definition (i.e., non-employed people who searched for a job in the previous three months) for the 2014-2019 period and also to generate the unemployment measure that is based on the narrower definition (i.e., non-employed people who searched for a job in the previous four weeks) for the 2004-2008 period.

In order to generate unemployment statistics that are based on the broader definition for the 2014-2019 period, we first estimate a probit model (Model 1) with a sample that consists of the last two of the aforementioned groups, namely those who searched for a job in the previous three months but not in the previous four weeks, as well as those who were not in the labor force. We construct the sample this way because individuals who belong to these groups in the 2014-2019 period are considered to have been out of labor force according to the narrower definition of unemployment, and we are interested in what fraction of these individuals would be classified as unemployed according to the broader definition of unemployment. The dependent variable in this probit model takes the value of one for those who searched for a job in the previous three months but not in the previous four weeks, and zero for the rest of this sample. The coefficient estimates are then used to predict the likelihood of a job search in the previous three months but not in the previous four weeks for each individual considered to have been out of the labor force according to the narrower definition in the 2014-2019 period. By adding the sum of these predicted probabilities to the population of individuals who searched for a job in the previous four weeks, we obtain the unemployment measure that is based on the broader definition for the 2014-2019 period.

Similarly, to generate the unemployment statistics that are based on the narrower definition for the 2004-2008 period, we estimate a probit model (Model 2) with a sample that consists of the first two of the aforementioned groups, namely those who searched for a job in the previous four weeks and those who searched for a job in the last three months but not in the previous four weeks. We construct the sample in this manner because individuals who belong to these groups in the 2004-2008 period are considered to have been unemployed according to the broader definition, and we are interested in what fraction of these individuals would no longer have been classified as unemployed according to the narrower definition. The dependent variable takes the value of one for those who searched for a job in the previous three months but not in the previous four weeks,

and zero for the rest of this sample. The coefficient estimates are then used to predict the likelihood of a job search in the previous three months but not in the previous four weeks for each individual who searched for a job in the last three months in the 2004-2008 period. By subtracting the sum of these predicted probabilities from the population of individuals who searched for a job in the previous three months, we get the unemployment measure that is based on the narrower definition for the 2004-2008 period.

The formulation of the probit model in our context can be described as follows. Let Y_i be a dummy variable that takes the value of 1 for individuals who searched for a job in the previous three months but not in the previous four weeks and 0 for all other individuals in the sample of interest. According to the model, the probability of searching for a job in the previous three months but not in the previous four weeks for individual i is equal to $\Phi(\beta X_i)$ where Φ denotes the cumulative distribution function for the normal distribution, X_i shows the vector of explanatory variables for individual i, and β denotes the vector of parameters.

We use various explanatory variables in each probit model. In particular, we control for individual-specific demographic characteristics, including age, educational attainment (with the dummy variables for middle school, high school, and college graduates), current status of enrollment, and marital status. We also include the following characteristics of each household as explanatory variables: the number of adult individuals (i.e., those aged 15 or over), the number of employed adults, the number of formally employed adults, and the total wage earned by adult wage workers. Moreover, we control for the employment history of individuals with a dummy variable for having had a previous job and with the length of the non-employment spell in years.[6] We also control for the effect of the most recent labor market activity with dummy variables for searching for a job, being retired, being enrolled, doing housework (only for women) and being a soldier (only for men) in the last year.[7] Lastly, we include the unemployment rate that is observed in the survey year in the NUTS-2 region of residency in addition to dummy variables for NUTS-2 regions indicating the residency.

We estimate each model separately for each gender by using both the original and suggested weights. We use the same set of explanatory variables in both models, but the regional unemployment rate is calculated differently in each. In particular, it is calculated as the ratio of the number of people who have searched for a job in the previous four weeks to the size of the labor force in Model 1, and as the ratio of the number of people who have searched for a job anytime in the previous three months to the size of the labor force in Model 2.

Probit Estimates

Table 4 provides the coefficient estimates of each probit model that employs the survey weights suggested in the *Methodology* section. The estimates based on the original weights are presented in Appendix Table A1. In this section, we interpret only the estimates obtained by using the suggested weights, as the estimates obtained by using each type of weights are very similar.

---

[6] A non-employment spell is calculated as the year the survey was conducted minus the year of the last job. For individuals who never worked, the non-employment spell is calculated by deducting 15 for primary or middle school graduates, 18 for high school graduates, and 22 for college graduates from the year the survey was conducted.

[7] In our specification, we include the dummy variable as indicator of having conducted housework in the previous year only for women, because no man in our sample is coded as having conducted this behavior as his main activity. Similarly, we use the dummy variable for having been a soldier only for men, because no woman is coded as having been a soldier in our sample.

Table 4   Coefficient Estimates of the Probit Models with the Suggested Weights

	Model 1		Model 2	
	Men	Women	Men	Women
Age	-0.0221***	-0.0130***	0.0024**	0.0021
	(0.0007)	(0.0007)	(0.0009)	(0.0014)
Middle school graduates	-0.1478***	-0.0493**	-0.0508**	-0.0173
	(0.0231)	(0.0251)	(0.0211)	(0.0334)
High school graduates	0.0051	0.3024***	-0.0303	-0.0262
	(0.0220)	(0.0202)	(0.0213)	(0.0278)
College graduates	0.2788***	0.6527***	-0.1073***	-0.0945***
	(0.0277)	(0.0252)	(0.0285)	(0.0317)
Married	0.2877***	-0.2046***	-0.0331	0.0443*
	(0.0233)	(0.0184)	(0.0212)	(0.0255)
Enrolled in a school	0.4556***	0.1334***	-0.0293	0.0042
	(0.0325)	(0.0330)	(0.0297)	(0.0289)
Number of adults in household	-0.0143**	0.0050	0.0024	-0.0167*
	(0.0059)	(0.0060)	(0.0058)	(0.0093)
Number of employed in household	-0.0819***	-0.0646***	0.0082	0.0458**
	(0.0125)	(0.0124)	(0.0124)	(0.0182)
Number of formal employed in household	0.014	0.0078	-0.0249	-0.0379*
	(0.0170)	(0.0160)	(0.0192)	(0.0227)
Total earnings in household (in 1,000 TL)	-0.0008	-0.0005	-0.0051	0.008
	(0.0066)	(0.0075)	(0.0161)	(0.0131)
Have previously worked	0.6216***	0.3530***	0.0234	0.0336
	(0.0241)	(0.0197)	(0.0335)	(0.0299)
Non-employment spell (in years)	-0.0386***	-0.0128***	0.0126***	0.0054***
	(0.0027)	(0.0011)	(0.0033)	(0.0020)
Searching for a job last year	0.5026***	1.1573***	0.0387**	0.0126
	(0.0194)	(0.0303)	(0.0172)	(0.0288)
Retired last year	-1.0569***	-0.9626***	0.2482***	0.3764***
	(0.0275)	(0.0631)	(0.0453)	(0.1222)
Enrolled in a school last year	-0.7083***	-0.5252***	0.1577***	0.1538***
	(0.0336)	(0.0384)	(0.0362)	(0.0408)
Housework last year	x	-0.3921***	x	0.2297***
		(0.0219)		(0.0305)
Military duty last year	0.4549***	x	-0.0421	x
	(0.0444)		(0.0373)	
Unable to work last year	-1.0111***	-0.8868***	0.2526***	0.3264**
	(0.0463)	(0.0681)	(0.0769)	(0.1417)
Regional unemployment rate	6.5112***	2.7212***	-3.3453***	-1.8095**
	(0.6541)	(0.6394)	(0.6099)	(0.8392)
Constant	-2.3362***	-2.3970***	-1.3995***	-1.5672***
	(0.0875)	(0.0908)	(0.0886)	(0.1066)
NUTS-2 Fixed Effects	Yes	Yes	Yes	Yes
Pseudo R-squared	0.3521	0.2752	0.0245	0.0445
Number of Observations	293845	722596	63440	32514

Notes: The table shows the coefficient estimates of the probit models where the dependent variable takes the value of one for individuals who searched for a job in the previous three months but not in the previous four weeks. The models are estimated using the survey weights suggested in this paper. The sample consists of individuals who are non-employed and searched for a job in the previous four weeks in Model 1 and those who searched for a job in the previous three months in Model 2.

Since Model 1 and Model 2 are estimated by using different samples of individuals, the interpretations of the coefficient estimates differ between models. In Model 1, searching for a job in the previous three months is an indicator of attachment to the labor force since the sample consists of individuals who are not in the labor force based on the narrower definition of unemployment (i.e., those who are not employed and have not searched for a job in the previous four weeks). On the other hand, in Model 2, the sample consists only of individuals who have searched for a job anytime in the previous three months. Thus, searching for a job in the previous three months but not in the previous four weeks is an indicator of being detached from the labor force in this model. As an example of this difference in the interpretation of estimates across the models, we compare the estimated coefficients of the dummy variable for being a college graduate. The coefficient of the dummy variable for being a college graduate has a positive sign in Model 1, which means that college graduates are more likely to have searched for a job in the previous three months among the sample of people who are not in the labor force. Meanwhile, the sign of the coefficient of the same dummy variable in Model 2 is negative, which means that college graduates are less likely to stop looking for a job in the previous four weeks among the sample of people who have searched for a job anytime in the previous three months.

Estimates of Model 1 show that individuals of both genders who are younger, more educated, and currently enrolled in a school are more likely to have searched for a job in the previous three months among the sample of people who are not in the labor force. The effect of being married differs between genders such that the likelihood of a job search in the previous three months decreases with marriage among women and increases among men. Only a few of the household-specific variables, such as the number of adults or the number of employed individuals in a household, are statistically significant determinants of having searched for a job in the previous three months. On the other hand, most of the variables describing labor market activities prior to the survey year are significant determinants in Model 1. In particular, the estimates show that individuals who were retired, enrolled in a school, conducted housework (only for women), or who were unable to work due to disability or illness in the preceding year, in addition to those who had a previous job or a longer non-employment spell, were less likely to have searched for a job in the three months preceding the survey.

In Model 2, compared to Model 1, a smaller number of variables significantly affect the likelihood of having searched for a job in the previous three months but not in the previous four weeks among the sample of unemployed people according to the broader definition. Specifically, college education and the number of household members for both genders; marriage status, the number of employed and formally employed individuals in a household for women; and age for men are statistically significant determinants. Similar to Model 1, most of the variables about labor market activities prior to the survey year significantly affect the likelihood of a job search in the previous three months but not in the previous four weeks. In particular, individuals of both genders who had longer non-employment spells and those who were retired, enrolled in a school, did housework (only for women) or were unable to work due to disability or illness in the same period of the previous year were more likely to have searched for a job in the previous three months but not in the previous four weeks.

We also find a great deal of heterogeneity among people who live in different regions. We provide the coefficient estimates associated with each NUTS-2 region for both models. The estimates with the suggested weights are given in Appendix Table A2, and the ones with the original weights are given in Appendix Table A3. In both models, we also find that the regional

unemployment rate is a statistically significant determinant of the likelihood of job search in the last three months but not in the last four weeks.

Predicted Unemployment Rates

Table 5 displays the observed and predicted populations of individuals who searched for a job in the previous three months but stopped searching in the previous four weeks over the 2004-2019 period. We report these numbers based on the survey weights suggested in this paper (the left panel of Table 5) and also based on the original weights (the right panel of Table 5). In both models with each type of weights, the predicted population is close to the observed population, which shows the goodness of fit for both the probit models that we estimate.

Table 5   Population of individuals who searched for a job in the previous three months but not in the previous four weeks (in thousands)

|	Suggested Weights			Original Weights		
	Observed	Model 1	Model 2	Observed	Model 1	Model 2
2004	--	--	272.9	--	--	274.8
2005	--	--	278.7	--	--	279.9
2006	--	--	277.9	--	--	278.1
2007	--	--	287.4	--	--	287.6
2008	--	--	307.5	--	--	308.1
2009	398.5	392.5	375.4	400.4	395.0	377.0
2010	332.1	342.6	352.4	332.8	343.3	353.3
2011	302.3	307.7	317.1	304.1	308.8	318.7
2012	327.2	319.7	306.6	329.2	322.2	308.4
2013	326.8	321.6	335.3	328.6	323.2	337.8
2014	--	324.1	--	--	324.7	--
2015	--	337.0	--	--	337.5	--
2016	--	363.6	--	--	364.1	--
2017	--	375.0	--	--	375.7	--
2018	--	386.7	--	--	387.6	--
2019	--	469.5	--	--	470.6	--

Notes: The table shows the observed and predicted population of individuals who searched for a job in the previous three months but not in the previous four weeks (in thousands). The left (right) panel displays the predicted population based on the coefficient estimates of the probit models estimated using the suggested (original) weights.

In the year 2014, the predicted population of people who searched for a job in the previous three months but stopped searching in the previous four weeks was 324.1 thousand based on the coefficient estimates obtained with the suggested weights.[8] This finding suggests that the population of unemployed people would have been 324.1 thousand larger if the 2014 THLFS used the broader definition of unemployment (i.e., a job search in the previous three months). Moreover,

---

[8] We use the STATA's intrinsic function "predict" to estimate these probabilities. If the probabilities are calculated based on the estimates reported in Table 4, they might slightly differ because of rounding errors. In fact, the corresponding number is 324.6 (instead of 324.1) when the probabilities are calculated based on the estimates in Table 4. Since the function "predict" uses more than four digits in the values of coefficients, it provides more accurate estimates.

in the year 2008, the predicted population of people who searched for a job in the previous three months but stopped searching in the previous four weeks was 307.5 thousand. This result suggests that the population of unemployed people would have been 307.5 thousand smaller if the 2008 THLFS had used the narrower definition of unemployment (i.e., a job search in the previous four weeks).

Figure 5   Unemployment Rates According to the Narrower and the Broader Definition**s**
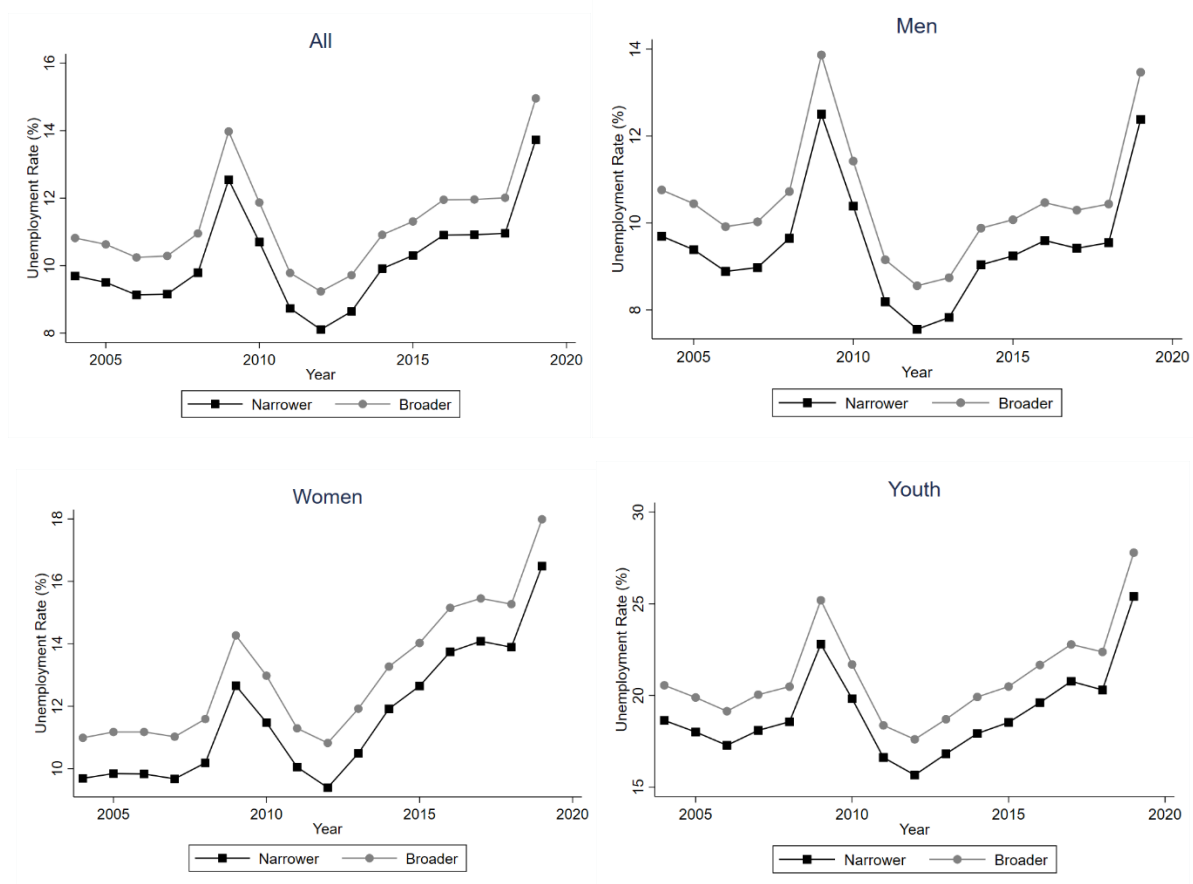
Notes: The 2004-2019 THFLS data. The black lines show the unemployment rates based on the narrower definition (i.e., a job search in the previous four weeks) among the specified group of people. The unemployment rates based on the narrower definition for the 2004-2008 period are obtained from the predicted probabilities according to the Model 2 estimated using the suggested weights. The gray lines show the unemployment rates based on the broader definition (i.e., a job search in the previous three months) among the specified group of people. The unemployment rates based on the broader definition for the 2014-2019 period are obtained from the predicted probabilities according to the Model 1 estimated using the suggested weights.

Next, we use the coefficient estimates of the probit models with the suggested weights to predict the populations of unemployed people based on the narrower definition for the 2004-2008 period and based on the broader definition for the 2014-2019 period. Then, we calculate unemployment rates based on each definition as the ratio of the corresponding population of unemployed people to that of people in the labor force (i.e., the sum of unemployed and employed people). In Figure 5, we plot both unemployment rates for the whole 2004-2019 period. By design, the unemployment rate based on the broader definition is higher than the one based on the narrower

definition. However, the ratio of these two unemployment measures changes over time. Specifically, according to our predictions, the ratio of the broader unemployment rate to the narrower one was 1.114 in 2009, 1.124 in 2013, and 1.089 in 2019. In Figure 5, we also provide the unemployment rates for men, women, and youth (i.e., those who are aged between 15 and 24) based on both definitions. Although the unemployment rates in each year and their evolution over time differ across these demographic groups, the ratio of the broader unemployment rate to the narrower one is similar in a given year. For instance, our predictions suggest that this ratio was 1.087 for men, 1.091 for women, and 1.093 for youth in 2014.

Conclusion

In this study, we highlight some irregularities in population and labor market statistics that result from the 2014 revisions to the THLFS and provide methods to correct them. The suggested survey weighting scheme in our study eliminates the issue of abnormal changes around the year 2014 in the population statistics of older people and in the population statistics by regions. We also propose a method to calculate the unemployment statistics according to both the narrower (job search in the previous four weeks) and the broader (job search in the previous three months) definitions of unemployment over the whole 2004-2019 period.

Our results provide several practical suggestions for the users of the THLFS data. We recommend that researchers who work on issues regarding the elderly population use our weights in their studies for robustness purposes. Also, studies that rely on the regional variation need to be conducted carefully. In particular, we suggest researchers initially check the correlation between the regional variation measures they are using and the regional variation in abnormalities that we report in Table 3 of this paper. In the case of high correlation, we strongly recommend using the suggested weights in this paper for robustness purposes. Lastly, we advise using the normalized versions of labor force statistics, such as the employment rate rather than the level of employment, to eliminate the confounding impact of the 2014 revisions on the outcomes in studies using the original survey weights.

To facilitate the application of our methods, we make the adjustment factors, which can be easily merged with the micro-level data of the THLFS to calculate the suggested weights, available upon request. In addition, we also provide the national and regional unemployment statistics for various demographic groups, including men, women, and youth, upon request. Researchers can also employ the parameter estimates of our probit models to generate the unemployment rates for other demographic groups of interest.

References

Akgündüz, Y.E. and Torun, H. (2020). "Two and a Half Million Syrian Refugees, Tasks and Capital Intensity," *Journal of Development Economics,* 145: 102470.

Araci, D.T., Demirci, M., and Kirdar M.G. (2021). "Development Level of Hosting Areas and the Impact of Refugees on Natives' Labor Market Outcomes." IZA Discussion Papers No: 14267.

Aydemir, A. and Kirdar, M.G. (2017a). "Quasi-Experimental Impact Estimates of Immigrant Labor Supply Shocks: The Role of Treatment and Comparison Group Matching and Relative Skill Composition," *European Economic Review,* 98: 282–315.

Aydemir, A. and Kirdar, M.G. (2017b). "Low Wage Returns to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey," *Oxford Bulletin of Economics and Statistics,* 79(6): 1046-1086.

Bakis, O. and Polat, S. (2015). "Wage Inequality in Turkey, 2002–10," *Economics of Transition,* 23(1): 169-212.

Caner, A., Demirel, M. and Okten, C. (2019). "Attainment and Gender Equality in Higher Education: Evidence from a Large Scale Expansion." IZA Discussion Papers No. 12711.

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

Dayıoğlu, M. and Kırdar, M.G. (2010). *Determinants of and Trends in Labor Force Participation of Women in Turkey*. World Bank.

Demirci, M. and Poyraz, M. (2021). "Post-Compulsory Schooling of Youth in Turkey during the Great Recession: A Case of Pro-cyclical Enrollment," Koç University-TUSIAD Economic Research Forum Working Papers 2117.

Gil-Alana, L.A., Ozdemir, Z.A., and Tansel, A. (2019). "Long Memory in Turkish Unemployment Rates," *Emerging Markets Finance and Trade,* 55(1): 201-217.

ILO (International Labor Organization) (2016). "Turkey-Household Labour Force Survey 2015" retrieved from https://www.ilo.org/surveyLib/index.php/catalog/1230/pdf-documentation

Jaeger, D.A. (1997). "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers," *Journal of Business & Economic Statistics,* 15(3): 300-309.

Karaoglan, D. and Okten, C. (2015) "Labor-Force Participation of Married Women in Turkey: A Study of the Added-Worker Effect and the Discouraged-Worker Effect," *Emerging Markets Finance and Trade,* 51(1): 274-290

Meyer, P.B. and Osborne, A.M. (2005). *Proposed Category System for 1960-2000 Census Occupations*. US Department of Labor, Bureau of Labor Statistics, Office of Productivity and Technology.

Official Gazette. (2012). "On Üç İlde Büyükşehir Belediyesi ve Yirmi Altı İlçe Kurulması ile Bazı Kanun ve Kanun Hükmünde Kararnamelerde Değişiklik Yapılmasına dair Kanun" retrieved from https://www.resmigazete.gov.tr/eskiler/2012/12/20121206-1.htm

Ozturk, A. and Tumen, S. (2018). "Education and Labor Market Consequences of Student Protests in Late 1970s and the Subsequent Military Coup in Turkey." IZA Discussion Papers No. 11733.

Patrinos, H.A., Psacharopoulos, G., and Tansel, A. (2020) "Private and Social Returns to Investment in Education: the Case of Turkey with Alternative Methods," *Applied Economics*, 53(14): 1-21.

Polat, S. (2014). "The Expansion of Higher Education in Turkey: Access, Equality and Regional Returns to Education," *Structural Change and Economic Dynamics,* 43: 1-14.

Tunali, I., Kirdar, M.G., and Dayioglu, M. (2019). "Female Labor Force Participation in Turkey: A Synthetic Control (Panel) Analysis, 1988-2013." IZA Discussion Papers No. 12844.

TURKSTAT. (2014). "The Explanations of New Regulations in Household Labor Force Surveys", retrieved from https://turkstatweb.tuik.gov.tr/MicroVeri/Hia_2014/english/downloads/explanations.pdf

TURKSTAT. (2015). "Times Series- Basic Labour Force Indicators, 2005-2014 January" retrieved from https://turkstatweb.tuik.gov.tr/PreHaberBultenleri.do?id=16007

Uysal, G. and Kavuncu, F. (2019) "HİA İstatistiklerinin Toplumsal Cinsiyet Ayrımında Revizyon," BETAM Bilgi Notu 18/17.

Online Appendix

Table A1: Coefficient Estimates of the Probit Models with the Original Weights

	Model 1		Model 2	
	Men	Women	Men	Women
Age	-0.0218***	-0.0129***	0.0024***	0.0020
	(0.0008)	(0.0007)	(0.0009)	(0.0014)
Middle school graduates	-0.1463***	-0.0471*	-0.0505**	-0.0137
	(0.0230)	(0.0251)	(0.0210)	(0.0333)
High school graduates	0.0071	0.3007***	-0.0289	-0.0281
	(0.0219)	(0.0202)	(0.0212)	(0.0277)
College graduates	0.2817***	0.6503***	-0.1060***	-0.0967***
	(0.0276)	(0.0251)	(0.0284)	(0.0317)
Married	0.2851***	-0.2062***	-0.0340	0.0431*
	(0.0232)	(0.0183)	(0.0211)	(0.0254)
Enrolled in a school	0.4563***	0.1331***	-0.0254	0.0055
	(0.0323)	(0.0330)	(0.0294)	(0.0289)
Number of adults in household	-0.0146**	0.0047	0.0017	-0.0167*
	(0.0058)	(0.0060)	(0.0058)	(0.0093)
Number of employed in household	-0.0834***	-0.0651***	0.0070	0.0444**
	(0.0123)	(0.0123)	(0.0123)	(0.0182)
Number of formal employed in household	0.0165	0.0103	-0.0230	-0.0353
	(0.0166)	(0.0159)	(0.0191)	(0.0227)
Total earnings in household (in 1,000 TL)	-0.0001	-0.0005	-0.0045	0.0077
	(0.0043)	(0.0074)	(0.0160)	(0.0131)
Have previously worked	0.6209***	0.3503***	0.0233	0.0332
	(0.0239)	(0.0196)	(0.0334)	(0.0298)
Non-employment spell (in years)	-0.0385***	-0.0128***	0.0125***	0.0055***
	(0.0027)	(0.0011)	(0.0033)	(0.0020)
Searching for a job last year	0.5001***	1.1592***	0.0405**	0.0142
	(0.0193)	(0.0301)	(0.0172)	(0.0287)
Retired last year	-1.0661***	-0.9725***	0.2444***	0.3602***
	(0.0275)	(0.0621)	(0.0451)	(0.1206)
Enrolled in a school last year	-0.7081***	-0.5333***	0.1564***	0.1507***
	(0.0334)	(0.0382)	(0.0359)	(0.0406)
Housework last year	x	-0.3955***	x	0.2267***
		(0.0218)		(0.0304)
Military duty last year	0.4563***	x	-0.0379	x
	(0.0443)		(0.0372)	
Unable to work last year	-1.0096***	-0.8942***	0.2591***	0.3281**
	(0.0461)	(0.0677)	(0.0766)	(0.1413)
Regional unemployment rate	6.5319***	2.6845***	-3.4492***	-1.9695**
	(0.6510)	(0.6369)	(0.6075)	(0.8378)
Constant	-2.3443***	-2.3943***	-1.4014***	-1.5561***
	(0.0871)	(0.0907)	(0.0881)	(0.1065)
NUTS-2 Fixed Effects	Yes	Yes	Yes	Yes
Pseudo R-squared	0.3494	0.2746	0.0243	0.0444
Number of Observations	293845	722596	63440	32514

Notes: The table shows the coefficient estimates of the probit models where the dependent variable takes the value of one for individuals who searched for a job in the previous three months but not in the previous four weeks. The models are estimated with the original survey weights provided in the THLFS data. The sample consists of individuals who are non-employed and have not searched for a job in the previous four weeks in Model 1 and those who searched for a job in the previous three months in Model 2.

Table A2: Estimated NUTS-2 Coefficients of the Probit Models with the Suggested Weights

| | Model 1 | | Model 2 | |
	Men	Women	Men	Women
İstanbul	*omitted*		*omitted*	
Edirne	0.0643	0.4461***	0.2744***	0.4583***
Balıkesir	-0.0660	0.2363***	0.3675***	0.4434***
İzmir	0.0421	0.3147***	0.1470***	0.2220***
Denizli	0.0380	0.5049***	0.3015***	0.5125***
Manisa	-0.0365	-0.0045	0.1348***	0.0966
Bursa	0.1219***	0.1259***	0.3606***	0.2391***
Kocaeli	0.3181***	0.3759***	0.4213***	0.5161***
Ankara	0.2047***	0.1983***	0.3275***	0.3663***
Konya	0.1165**	0.3526***	0.4282***	0.6108***
Antalya	-0.0182	0.2977***	0.2060***	0.2287***
Adana	0.1742***	0.4560***	0.3931***	0.5715***
Hatay	0.1449***	0.3756***	0.4350***	0.6633***
Nevşehir	-0.0716	0.1779***	0.2498***	0.3844***
Kayseri	0.2646***	0.2225***	0.4205***	0.2531***
Zonguldak	0.0005	0.4249***	0.0768	0.3618***
Kastamonu	0.1017	0.3885***	0.3060***	0.2858***
Samsun	-0.0265	0.3170***	0.3169***	0.4897***
Trabzon	0.3348***	0.4037***	0.5849***	0.7059***
Erzurum	0.2128***	0.1135*	0.2566***	0.5869***
Kars	0.2579***	0.1590**	0.5374***	0.7341***
Malatya	0.1885***	0.5064***	0.6262***	0.9663***
Van	-0.1002**	0.0364	0.3973***	0.5361***
Gaziantep	0.0719*	-0.0586	0.4346***	0.3784***
Diyarbakır	-0.3987***	-0.2191***	0.4833***	0.7910***
Siirt	-0.0819*	-0.0186	0.4381***	0.6723***

Notes: The table shows the estimated NUTS-2 coefficients associated with the probit models displayed in Table 4.

Table A3: Estimated NUTS-2 Coefficients of the Probit Models with the Original Weights

| | Model 1 | | Model 2 | |
	Men	Women	Men	Women
İstanbul	omitted		omitted	
Edirne	0.0600	0.4542***	0.2784***	0.4570***
Balıkesir	-0.0760	0.2347***	0.3595***	0.4370***
İzmir	0.0383	0.3125***	0.1521***	0.2232***
Denizli	0.0312	0.5098***	0.3039***	0.5155***
Manisa	-0.0396	-0.0025	0.1360***	0.0938
Bursa	0.1152***	0.1311***	0.3598***	0.2419***
Kocaeli	0.3179***	0.3771***	0.4225***	0.5191***
Ankara	0.1959***	0.1991***	0.3271***	0.3631***
Konya	0.1115**	0.3513***	0.4294***	0.6137***
Antalya	-0.0260	0.3014***	0.2006***	0.2295***
Adana	0.1668***	0.4608***	0.3982***	0.5776***
Hatay	0.1395***	0.3779***	0.4382***	0.6565***
Nevşehir	-0.0696	0.1769***	0.2513***	0.3857***
Kayseri	0.2585***	0.2242***	0.4223***	0.2485***
Zonguldak	-0.0056	0.4260***	0.0808	0.3653***
Kastamonu	0.1029	0.3862***	0.3070***	0.2850***
Samsun	-0.0298	0.3147***	0.3172***	0.4850***
Trabzon	0.3282***	0.3930***	0.5866***	0.6969***
Erzurum	0.2217***	0.1172*	0.2512***	0.5808***
Kars	0.2658***	0.1713***	0.5284***	0.7348***
Malatya	0.1871***	0.5112***	0.6336***	0.9679***
Van	-0.0997**	0.0461	0.4029***	0.5313***
Gaziantep	0.0741*	-0.0667	0.4371***	0.3628***
Diyarbakır	-0.3983***	-0.2221***	0.4849***	0.7824***
Siirt	-0.0766*	-0.011	0.4422***	0.6729***

Notes: The table shows the estimated NUTS-2 coefficients associated with the probit models displayed in Table A1.