

**A framework for measuring international labour migration
based on a combination of data sources with numerical illustration in the case of Turkey**

Farhad Mehran, ILO Consultant
mehranxfarhad@yahoo.com

25 May 2020 [Revised 26 May 2020, 19 October 2020]

- I. International migrant workers
 - a. Usual residents of country of measurement
[Draft completed]
 - b. Not usual residents, or non-residents foreign workers
[To be drafted]
- II. For-work international migrants
[To be drafted]
- III. Return international migrant workers
[To be drafted]
- IV. Citizens or Native-born working abroad
[To be drafted]

1. Introduction

Most countries dispose a variety of data sources on international labour migration. The data sources are often disparate and no single source provide a comprehensive coverage of the country's international labour migration. The recent international guidelines concerning statistics of international labour migration¹ distinguishes the sources of stock and flow statistics as follows:

- *sources generating stock statistics*: population censuses; household surveys, in particular labour force surveys (especially relevant for certain groups, e.g. category 14(a) (usual residents) of international migrant workers); specialized migration and demographic surveys; surveys limited to, or focused on, particular populations or domains (such as surveys near international borders, surveys of refugee camps); and establishment census and sample surveys;
- *sources generating flow statistics*: border registrations; statistics of residence permits issued; statistics of work permits issued, statistics of visas issued; departure registrations; and household surveys;
- *sources that can generate both flow and stock statistics*: population registers; household surveys, registers of foreigners; tax and social security registrations; and registrations for use of utilities (e.g. phone, electricity);
- *other sources*: certain groups of international migrant workers who are in need of international protection may require special surveys with targeted samples due to their particular characteristics and circumstances. Appropriate administrative sources of data, where available, may supplement targeted sample surveys or may even serve as an alternative to such surveys.

The international guidelines emphasize that: “These different statistical sources should be treated as complementary, to be used in combination in order to derive comprehensive sets of statistics to the extent possible.” The purpose of the present note is to lay-out a framework for carrying out this task in a systematic and orderly fashion. The main elements of the proposed framework are described in the next section and illustrated in the case of Turkey in the subsequent sections.

A companion framework is being developed for combining data sources on international labour migration at the individual record level. The framework is illustrated with a numerical example on the measurement of “frontier workers” in Switzerland, a sub-category of international migrant workers, category (b), i.e. non-resident foreign workers.

2. Framework for combining data from different sources

Consider the estimation of the total stock of international migrant workers (category 14a) of the international guidelines concerning statistics of international labour migration. It refers to “international migrants who, during a specified reference period, were in the labour force of the country of their usual residence, either in employment or in unemployment.” The

¹ ILO *Guidelines concerning statistics of international labour migration*, 20th International Conference of Labour Statisticians, Geneva, 2018, para. 14.

problem at hand is to estimate the total stock for a specified reference period based on existing data sources in a given country.

The proposed framework for carrying out this task is based on the construction of a hierarchical set of estimates, starting with “international migrants”, among them “international migrants of working age” and finally “international migrant workers (category 14a),

- International migrants
- International migrants of working age
- International migrant workers (category 14a)

The construction of the estimates of a universe U , say, international migrants, using existing data sources involves four steps as follows:

- 1) Compilation
- 2) Comparison
- 3) Reconciliation
- 4) Balancing

Compilation means listing, gathering and documenting the existing data sources, say, $X_1, X_2, X_3, \dots, X_n$. The end product of this step is a set of independent tables corresponding to each data source. Documenting means describing the data source in terms of its scope and coverage, listing its contents and providing definitions of the items, specifying the number of records, and explaining how records are added or deleted.

Comparison means reviewing the scope and coverage and the underlying concepts and definitions, and determining the relative strengths and limitations of each data source. The end product of this step is a set of comparative tables where each column of a table corresponds to data from one data source.

Reconciliation means taking two data sources at the time and trying to quantify, with double-entry book-keeping, the differences in coverage and definitions as to arrive as close as possible to the target universe U with a minimum outstanding balance between the two sources.

Let X_1 and X_2 be the two data sets selected on the basis of considerations in the preceding step. The reconciliation starts with the construction of a double-entry table, where X_1 is set on the one side of the table, say the left-hand side, and X_2 on the other side, say the right-hand side. Reconciliation then means finding elements Y_1 and Z_1 to add and subtract in X_1 , and similarly, elements Y_2 and Z_2 , to add and subtract in X_2 , so that the resulting reconciled table can be expressed as, $U_1 = X_1 + Y_1 - Z_1$ and $U_2 = X_2 + Y_2 - Z_2$, where the coverages, U_1 and U_2 , have been made as close as possible to U , and the balancing item, defined as the difference between U_1 and U_2 , has been made as small as possible. It should be mentioned

that in practice, the different adjustments in the reconciliation process may sometimes take the form of multiplicative factors, rather than the form of additive or subtractive elements.

Where more than two data sources are involved, the reconciliation process continues until all data sources have been reconciled. At each stage after the first, the new data source is reconciled with the reconciled figure obtained from the previous stage. Thus, if X_3 is the new data source, and $\widehat{X_1 X_2}$ is the reconciled figure obtained from the reconciliation of the earlier data sources, X_1 and X_2 , then X_3 is set in the left-hand side of the new table and $\widehat{X_1 X_2}$ is moved to the right-hand side of the new table. The reconciliation process is now carried out on the basis of the new data source X_3 and the hybrid data source formed by the combination of X_1 and X_2 , and so on.

Balancing is the final step of the estimation process. It is a judgmental evaluation of the last reconciliation table. Given that the value of the target universe U is not known, a choice must be made between the two end figures in the last reconciliation table. This choice may be made on the basis subjective considerations or on the basis of the magnitude of the effect of reconciliation. A figure obtained with the need for a large amount of reconciliation may be regarded as less reliable than one obtained with less reconciliation.

In the subsequent sections of the paper, a numerical illustration is presented with preliminary data from Turkey.

3. Estimation of the stock of international migrants in Turkey

In estimating the stock of international migrants in Turkey, it is convenient to separate Syrian refugees from others, and divide the total universe into three groups as follows:

1. International migrants (Non-Syrians)
2. Syrian refugees
 - a. Syrian refugees with resident permits
 - b. Syrian refugees without resident permits, under temporary protection

Data sources on groups 1 and 2a are essentially the same and can be treated in a similar fashion. Data sources on 2b, however, are different in nature, based mostly on national and international aid agencies and non-governmental organizations. The numerical illustration here focuses on group 1 only.

Compilation of data sources

As part of a comprehensive study on the measurement of international labour migration in Turkey, a recent ILO report examined more than 10 data sources on international migration in Turkey² including:

Administrative data sources

- Migration statistics – Directorate General of Migration Management (DGMM)

² ILO, "Measuring International Labor Migration in Turkey," ILO consultant, 2020.

- Social security registration data – Social Security Institute (SSI)
- Border statistics – Ministry of Culture and Tourism
- Establishment data – Ministry of Industry and Technology
- Address-based population register system (ADNKS)

Household-based surveys

- Labour force survey (LFS)
- Survey of income and living conditions (SILC)
- Household budget survey

Establishment-based surveys

- Establishment census
- Establishment sample surveys

This list essentially fulfils the first step of the present measurement framework, that is “compilation of the data sources”. It corresponds to the list of data sources X_1, X_2, \dots, X_n , mentioned earlier with $n=10$.

Comparison of data sources

The next step is the comparison of the data sources, that is the review of their scope and coverage and their underlying concepts and definitions, with the aim of determining their relative strengths and limitations. This step has also been accomplished in the ILO report where the information available in each source is described and the contents analyzed in the light of the ILO guidelines on statistics of international labour migration.

Three data sources emerge as most comprehensive and relevant for the measurement of international migrant workers in Turkey. These are labelled here as:

X_1 = Migration statistics – Directorate General of Migration Management (DGMM)

DGMM is part of the Ministry of Interior of Turkey. It is responsible “to implement policies and strategies related to migration,” and in particular to “carry out tasks and procedures related to foreigners’ entry into, stay in, exit and removal from Turkey.”³ It maintains statistics on entrance and exit, residence permits, irregular migration, international protection, temporary protection, victims of human trafficking, return statistics and removal centres.

X_2 = Address-based population register system (ADNKS)

ADKNS is an “Address Based Population Registration System” maintained by TurkStat, the national statistical institute of Turkey. The system is based on the integration of current administrative registers to “obtain updated information on [the] size and characteristics of [the] population of localities and to follow up the population movements regularly.”⁴

X_3 = Labour force survey (LFS)

³ <http://en.goc.gov.tr>

⁴ <http://turkstat.gov.tr>

The labour force survey is conducted on a quarterly basis by TurkStat.⁵ It covers the population resident in Turkey including all persons with resident permits who have been in Turkey for at least 12 months or with an intention of stay of at least 12 months. The sample is based on ADNKS. The survey provides estimates of foreign nationals within the scope of the survey. In principle, estimates of the total foreign population within the scope of the survey data can be obtained for the working age population as well as the population below working age.

Reconciliation of data sources

In line with the third step of the measurement framework, the three main data sources X_1 , X_2 , and X_3 are reconciled. The reconciliation process may start, in principle, with any combination of two data sources (X_1 and X_2 , or X_1 and X_3 , or X_2 and X_3). In the present context, however, it was found more convenient to start with the first two data sources (X_1 =DGMM and X_2 =ADKNS). Both are administrative sources and related to each other as one (X_1) feeds the other (X_2). Also, this ordering leaves the labour force survey (X_3) as the last data source for reconciliation. The LFS is the main source of subsequent estimation of the international migrants of working age and the stock of international migrant workers.

- Reconciliation of data sources X_1 (DGMM) and X_2 (ADKNS)

Table A attempts to lay-out the main elements of reconciliation of the stock of international migrants in 2019 based on the two main administrative sources in Turkey. Data source X_1 (DGMM) is set in the left-hand side of the table and data source X_2 (ADKNS) in the right-hand side.

Table A. Calculation of the stock of international migrants, Turkey 2019
Reconciliation of data from DGMM and ADNKS

	DGMM ¹	Number		ADNKS ²	Number
1	Total	5'456'223	1	Total	1'513'180
2	Foreign nationals with resident /work permits	1'101'030	2	Foreign nationals who reported stay not exceeding 3 months	Y
3	Apprehended undocumented foreign nationals	454'662	3	Other foreign nationals who could not supply valid address	Z
4	Syrian refugees under temporary protection (excludes Syrians with resident permits already accounted in 2) ³	3'576'370	4	Former Turkish citizens resident in Turkey (blue card)	-t
5	Foreign nationals under international protection (56'417) + others awaiting to be resettled or applications processed	324'161	5		
6	Total = (2)+(3)+(5)	1'879'853	6	Total = (1)+(2)+(3)	1'513'180 + y+z-t

⁵<http://www.turkstat.gov.tr/PreHaberBultenleri.do;jsessionid=5JyBYvTTDnw85nXyBLHGwqTk1nFnTS9LtgX1S7XC7TVsjTvyJ4d!19786689?id=21579>

7	Undocumented foreign nationals not apprehended	X	7		
8	Total = (6)+(7)	1'879'853 +x	8	Total = (1)+(2)+(3)	1'513'180 + y+z-t
9	Balancing item	366'673+x-y-z+t	9		

Sources: ¹Directorate General of Migration Management (DGMM), "Measuring International Labor Migration in Turkey," ILO consultant, Table 4.1, ²Address Based Population Registration System (ADNKS), TurkStat, ILO Consultant paper, Table 4.2a.

Note:³ The case of Syrian refugees under temporary protection is treated separately outside the scope of the present paper.

In Table A, the unknown statistics are labelled as "x", "y", "z" and "t". In principle, data on y, z and t can be obtained from the agency providing the underlying population data to TurkStat, as the information on "Foreign nationals who reported stay not exceeding 3 months" and "Foreign nationals who could not supply valid address" must be known to them. An initial estimate of x, "undocumented foreign nationals not apprehended" can then be derived from the balancing item of the table:

$$x_0 = y + z - t - 366'673$$

In principle, the value of the estimate must be positive. If it turns out to be negative, more careful scrutiny of the underlying data should be carried out. To fix ideas, suppose the number foreign nationals who reported stay not exceeding 3 months to be $y=800'000$ and the number of other foreign nationals who could not supply a valid address to be $z=50'000$, and the number blue card holders $t=100'000$, then x_0 would be $383'327$, a value almost equal to the number of apprehended undocumented foreign nationals, $454'662$. This ratio of about 1 to 1 of not apprehended to apprehended undocumented foreign nationals would then be reviewed by DGMM, and a new estimate of x would be produced,

$$x_1 = \max(x_0, \delta \times 454'662)$$

where δ is the DDGM's assessment of the not apprehended to apprehended ratio.

If this procedure cannot be conducted or the result is not considered sufficiently reliable, an alternative more elaborate estimation of the undocumented foreign nationals may be envisaged. An example is given by confronting data on residence permit by year of arrival and household survey data on international migrants by year of arrival.⁶ The calculation requires a sufficiently long, time series and the availability of a question on "year arrival" in the household survey questionnaire.

- Reconciliation with data source X_3 (LFS)

⁶ George Lemaitre, "A general method for estimating the number of unauthorized immigrants using standard data sources" (version 2), IFMS 2020, Cairo, Egypt, 20 January 2020. A reformulated methodology based on linear models is described in Farhad Mehran, "Estimation of undocumented international migrants (and workers)" preliminary draft, 15 May 2020.

The result of the reconciliation process of data sources X_1 (DGMM) and X_2 (ADKNS) is on line 8 of Table A (left panel). The balancing item which expresses the outstanding part of the reconciliation is on line 9. The reconciled value now moves to the right-hand side of the new Table B for reconciliation with the third data source, the labour force survey, X_3 (LFS).

Table B. Calculation of the stock of international migrants, Turkey 2019
Reconciliation of household labour force survey and administrative data

	LFS ¹	Number		Combined DGMM & ADNKS ²	Number
1	Total	LFS estimate 2019	1	Total	1'879'853 +x
2	Foreign nationals who stay or intend to stay more than 90 days but less than 12 months in Turkey.	u	2		
3	Foreign nationals not member of household within scope of survey (e.g., living at worksite, garage, hospital, ...)	v	3		
4	Foreign nationals (Syrian refugees) under temporary protection living in household within scope of survey	-w	4		
5	Former Turkish citizens resident in Turkey (blue card)	-s	5		
6	Total = (1)+(2)+(3)-(4)-(5)	LFS estimate +u+v-w-s	6	Total	1'879'853 +x
7	Balancing item	LFS estimate - 1'879'853 +u+v-w-s-x	7		

Sources: ¹ TurkStat, Household Labour Force Survey. ² Reconciled DGMM and ADNKS total (Table A).

The new unknown data in Table B are labelled as “u”, “v”, “w” and “s”. Data on u, w and s, can, in principle, be obtained from counts during field operations of the survey. The estimates can be constructed using the design weights of locations where the units were found. [By design weight, it is meant the sampling weights before calibration to the aggregate total.]

An initial estimate of v, “Foreign nationals not member of household within scope of survey (e.g., living at worksite, garage, hospital, ...)” can then be derived from the balancing item of the table:

$$v_0 = (1'879'853 + x) - (LFS\ estimate + u - w - s)$$

The resulting value may be positive or negative depending on the under or over coverage of the survey in relation to the combined DGMM & ADNKS total. If $v_0 < 0$, it means that DGMM has under coverage even after the adjustment for not apprehended undocumented foreign nationals. If $v_0 > 0$, it means that DGMM has over coverage beyond the adjustment for not apprehended undocumented foreign nationals. If necessary, a new estimate may be developed, perhaps, by field operations in a sample of locations where households are listed and the proportion of foreign nationals not member of household within scope of survey (e.g., living at worksite, garage, hospital, ...) is obtained. The resulting new estimate, $v=v_1$, would then be used to calculate the reconciled estimate of the total number of international migrants,

$$\text{Reconciled (LFS, DGMM, ADNKS)} = \text{LFS estimate (2019)} + u + v - w - s$$

Final balance

At the end of any reconciliation process, a balancing item is expected to remain. It reflects the part of the differences among the data sources that could not be reconciled, either because the data for reconciliation were not available or because no further explanation for the differences could be identified.

In the present example, the last balancing item is in line 7 of Table B. If its numeric value cannot be determined, then the left-hand total in line 6 would be different than the corresponding right-hand total in line 6. In this case, a choice has to be made between one of the two totals. This choice is called the final balance. As mentioned earlier, this choice may be made on the basis subjective considerations or on the basis of the magnitude of the effect of reconciliation following the principle that a figure obtained with the need for a large amount of reconciliation may be regarded as less reliable than one obtained with less reconciliation.

A fuller description of this process may be provided when the proposed framework is in reality implemented in Turkey and the various components (x, y, z, t, s, u, v, and w) are actually quantified.

4. Estimation of the stock of International migrants of working age in Turkey

The estimate of international migrants derived at the end of the reconciliation of the data of LFS, DDGM and ADNKS in Table B will now become the base for the calculation of the international migrants of working age (15 years and over). A simple method shown in Table C below is to calculate the proportion of the working age population in the total foreign-born (or non-citizen) from the ADNKS and then apply the proportion to the reconciled base population.

**Table C. Calculation of the stock of international migrants of working age, Turkey 2019
Based on ADNKS age data**

	ADNKS ¹	Number		Reconciled (LFS,DGMM,ADNKS) ²	Number
1	Total foreign nationals	1'531'180	1	Total	Reconciled (LFS,DGMM,ADNKS)
2	Foreign nationals in age group 0-14 years	m	2		
3	Foreign nationals 15 years old and over	n	3		
4	Share of population 15+ years %	$p=m/n$	4		
5			5	Total foreign population 15+ years	$p \times$ Reconciled (LFS,DGMM,ADNKS)

Sources: ¹ TurkStat, ADNKS. ² Reconciled (LFS,DGMM,ADNKS) total (Table B).

Note: ³ Depending on the availability and quality of the data on place of birth in ADNKS, the calculations may also be carried out on the basis of foreign born population.

The choice of ADNKS in Table C is based on the consideration that the coverage of the register is broader than that of the LFS and would provide a more accurate estimate of p for foreign nationals. In fact, age data are also available on the components y and z of Table A, their use in the calculation of p may improve the accuracy of the estimate. If this choice is considered inadequate, the calculations in Table C may be based on the age data of LFS. In any case, it should be mentioned that it would be more appropriate to calculate p for men and women separately, in practice.

5. Estimation of the stock of international migrant workers in Turkey

The final stage is the estimation of the stock of the international migrant workers (category a) based on the reconciled estimate of international migrants obtained in Table C. The calculation is shown in the following Table D. Essentially, it amounts to adding to the estimate of the international migrant workers within the scope of LFS, an estimate of international migrant workers outside the scope of LFS.

Table D. Calculation of the stock of international migrant workers (category a) Turkey 2018 [To be updated for 2019]

	LFS ¹	Number		Reconciled (LFS,DGMM,ADNKS) ²	Number
1	LFS international migrants of working age	1,698,906-w	1	International migrants of working age	pxReconciled (LFS,DGMM,ADNKS)
2	In labour force	807,949-win	2		
3	Outside labour force	890,947-wout	3		
4	Other international migrants of working age	Reconciled – 1,698,906+w	4		
5	Labour force participation rate	r	5		
6	In labour force	r x (4)	6		
7	Outside labour force	(4)-r*(4)	7		
8	Total international migrant workers (category a)	(2)+(6)	8		

Sources: ¹ TurkStat, Household Labour Force Survey. Reported in ILO Consultant report, Table 4.5a, ²pxReconciled (LFS,DGMM,ADNKS) Table C.

The latter estimate may be obtained by simply applying the labour force participation rate of international migrants within the scope of LFS (r_o) to the estimate of the other international migrants of working age, where

$$r_o = \frac{807,949}{1,698,906} = 0.476$$

or, preferably,

$$r_o = \frac{807,949 - win}{1,698,906 - w} =$$

where w is the LFS estimate of foreign nationals (Syrian refugees) of working age under temporary protection living in household within scope of survey and w_{in} is the LFS estimate of those in the labour force.

This initial value of r may be revised based on quantitative or qualitative information on the population of international migrants outside the scope of LFS. For example, if it is known that the educational attainment of this population is predominately below secondary level of education, then the LFS estimate of labour force participation rate of that category persons may be applied. Or, if it is known that the international migrants outside the scope of LFS are predominately in large cities (Istanbul, Ankara, Izmir), then the corresponding LFS labour force participation rate(s) may be used.

6. Overall data requirement for Turkey

The various data items of Tables A, B, C and D for the calculation of the stock of international migrant workers (category a) are assembled together by data source in Table E below. The 'OK' sign in the last column of the table indicates that the data item is available and presented in the report. The sign "OK?" means that the data item should, in principle, be available from the indicated source, but not reflected in the present report. The extraction of the data may in practice require meetings with relevant officials to provide clarifications on the data item and explanations on the context of the exercise. The data items marked with "see text" concern the two variables for which special treatments are needed as described in the text.

Table E. Overall data requirements

DDGM	Data item		
Table A (1)	Total registered foreign nationals	-	OK
Table A (2)	Foreign nationals with resident/work permits	-	OK
Table A (3)	Apprehended undocumented foreign nationals	-	OK
Table A (4)	Syrian refugees under temporary protection	-	OK
Table A (5)	Applicants for international protection	-	OK
Table A (7)	Undocumented foreign nationals not apprehended	x	see text
ADNKS	Data item		
Table A (1)	Total foreign nationals	-	OK
Table A (2)	Foreign nationals who reported stay not exceeding 3 months	y	OK?
Table A (3)	Other foreign nationals who could not supply valid address	z	OK?
Table A (4)	Former Turkish citizens resident in Turkey (blue card)	t	OK?
Table C (2)	Foreign nationals in age group 0-14 years	m	OK?
Table C (3)	Foreign nationals 15 years old and over	n	OK?
HLFS	Data item		
Table B (1)	Total foreign nationals in households within scope of survey	-	OK
Table B (2)	Foreign nationals in Turkey 9-12 months or with intention to stay 9-12 months	u	OK?
Table B (3)	Foreign nationals not members of households	v	see text
Table B (4)	Foreign nationals (Syrian refugees) under temporary protection in households within scope of survey	w	OK?
Table B (5)	Former Turkish citizens resident in Turkey (blue card)	s	OK?
Table D (1)	Total foreign nationals of working age in households within scope of survey	-	OK
Table D (2)	Total foreign nationals of working age in labour force, in households within scope of survey	w_{in}	OK?
Table D (3)	Total foreign nationals of working age outside labour force, in households within scope of survey	w_{out}	OK?

In practice, the calculations of the reconciliation Tables A, B, C, D, and all data items listed in Table D should be obtained, to the extent feasible and possible, for men and women, separately.
