Nowcasting refugee population stock figures

Andrea Pellandra, UNHCR

February, 2020

UNHCR
The UN Refugee Agency

UNHCR Statistics

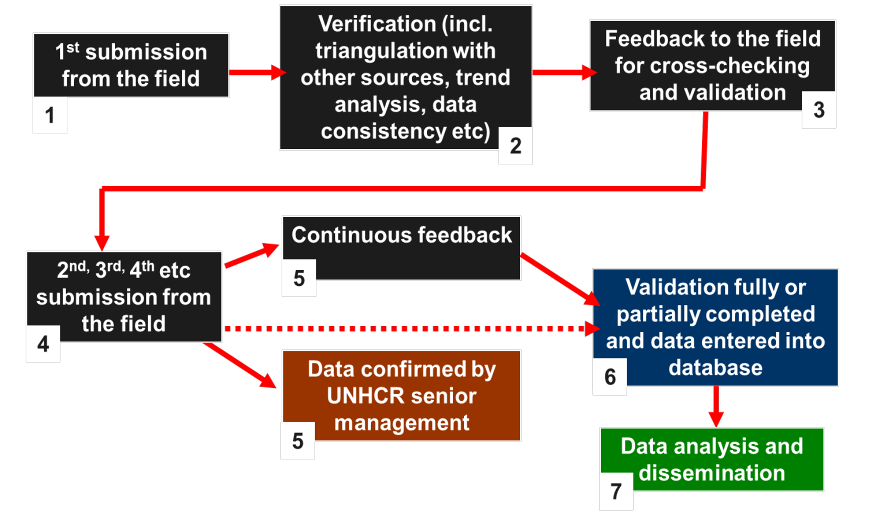
- UNHCR is the official source of statistics for refugees and forcibly displaced
- Official Statistical reviews conducted twice a year
- Population estimates released following global reporting, compilation and validation process
- All data submitted by countries to country operations and entered in global database
- Verified for consistency by the UNHCR Statistical Unit

UNHCR Populations of concern

- 74.8 million people:
  - Refugees (20.4 million under UNHCR mandate)
  - Asylum-seekers (3.5 million)
  - Internally-displaced (41.4 million)
  - Stateless (3.9 million)
  - Returned refugees (0.6 million)
  - Returned IDPs (2.3 million)
  - Others of concern and displaced Venezuelans outside of asylum system (3.8 million)

Annual Statistics Report: from the field to the HC

Biannual Statistics Review Exercise from 192 operations



Statistical analysis and products

- High visibility
- Many media requests: ensuring accurate and relevant data is available
- Most tweeted images and high media impact
- Increased evidence-based advocacy and policies

UNHCR The UN Refugee Agency

Main data sources

- Direct refugee registration databases
  - UNHCR proGres database
  - National registration systems
- Administrative asylum systems
  - Different national institutions may be responsible for data production in the area of asylum and refugee matters (National Statistical Offices, Ministry of Interior, Ministry of Justice, Administrative Tribunals).

Motivation and Challenges

- Usually released with lag of several months
- Need for more frequent updates with shorter lag
- Different data sources in different operations
- Lack of registration data
- Government owned data
- Sometimes different data sources for different populations in same operation

Methodology 1: registration data (stock)

- Refugee pop estimated through ProGres data
  - If V4 then drawn from database/DataPort
  - If V3 then data requested from ProGres instances each month from operation.
  - Where required, follow up with the operation and complement with operational data (unregistered populations, ie. SUD, AFG, YEM, ETH, BGH, DRC, CAM, etc.)
  - Outliers checked and followed up on.
  - Non proGres registration (eg Turkey) relies on published gvt figures

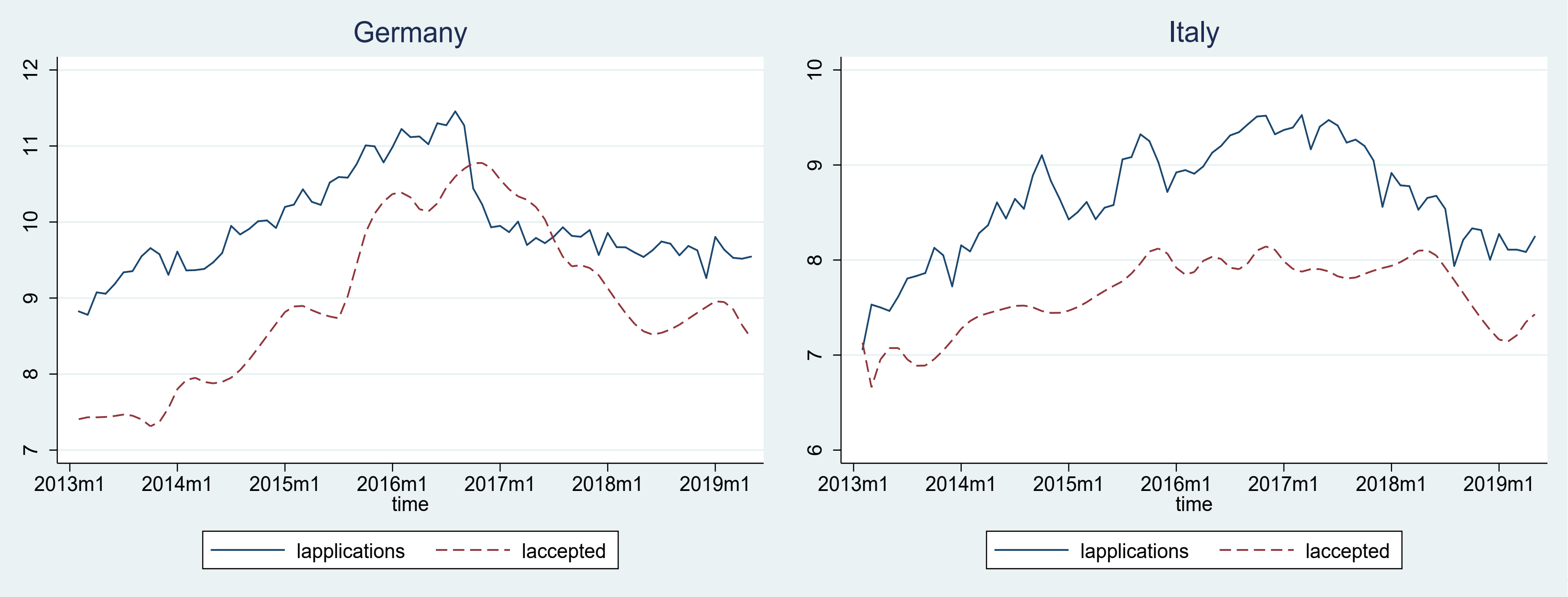
Methodology 2: RSD data (flows)

- Asylum decisions reported on monthly basis to UNHCR
  - No refugee pop reported (UNHCR already estimating)
  - Positive decisions added to stock
  - Deductions made based on assumed length of time in refugee status (usually 10 years, monthly deduction assumed to be linear)

Methodology 3: Time series estimation

- Number of asylum seekers in the current year is highly related to the numbers in last year
  - The application of asylum seekers in current period is positively correlated with the applications in previous period
  - The recognition of asylum seekers in last year is positively correlated with the applications in previous period
- Time series analysis

Methodology 3: Time series estimation



UNHCR
The UN Refugee Agency

Methodology 3: time series projections (monthly)

- Asylum applications and recognitions entangled in a dynamic relationship
- Time series analysis provides a convenient framework to analysis the dynamic relationship between recognitions and applications
- Granger causality test shows applications Granger-cause recognitions in the majority of countries
- Estimate multiple VARs of varying lag lengths and compute multiple test statistics

Methodology 3: time series projections

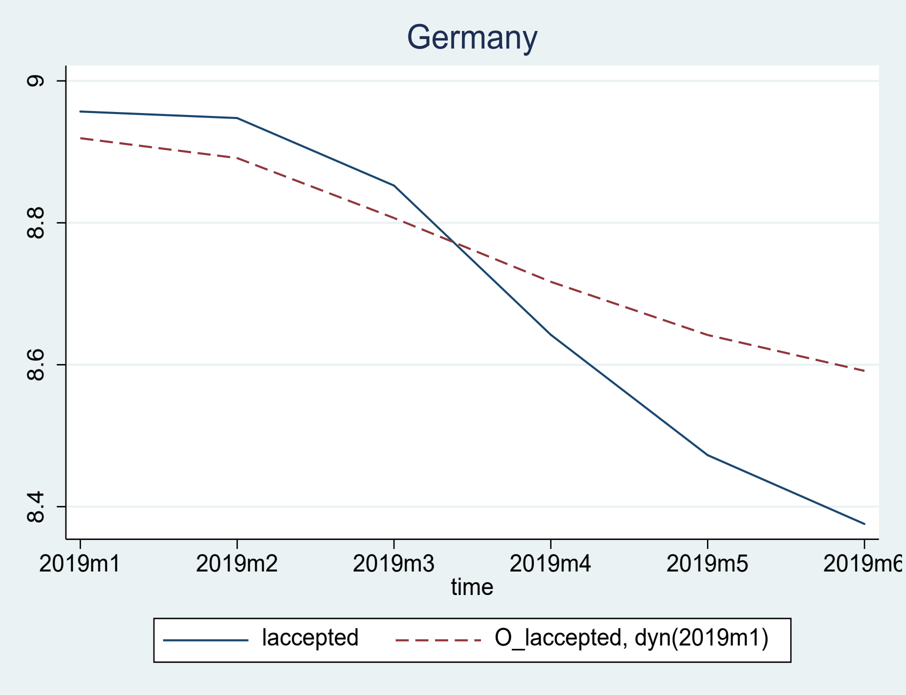
- Decisions reported quarterly to Eurostat
  - apportioned to months through least squares distribution
  - forecast based on reduced-form Vector Autoregression multivariate time series model

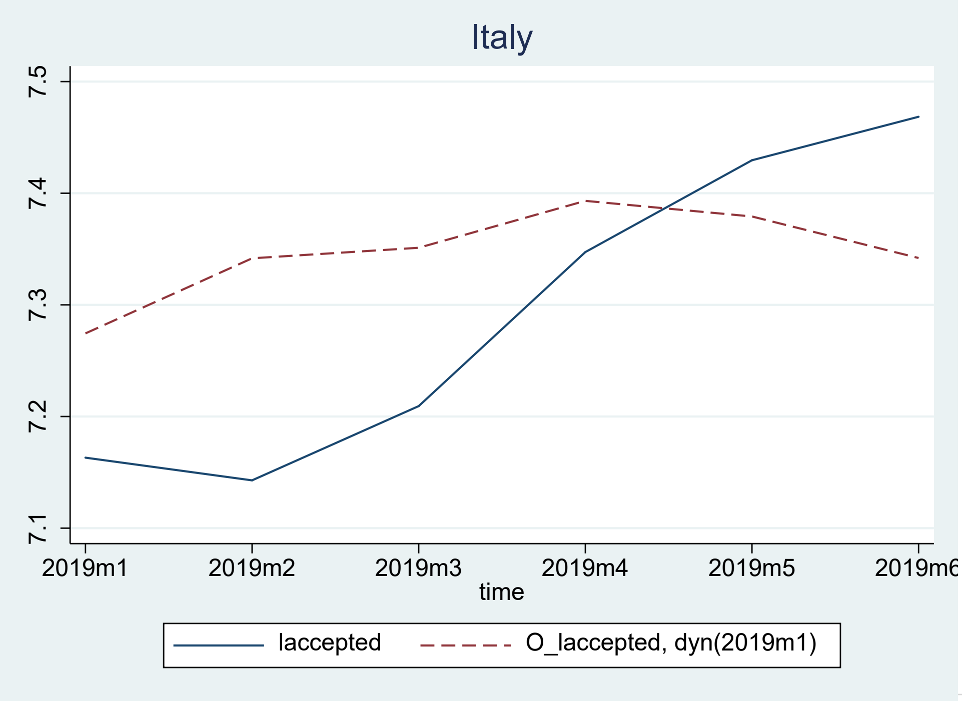
$$\mathbf{Z}_t = \delta + \Theta_1 \mathbf{Z}_{t-1} + \Theta_2 \mathbf{Z}_{t-2} + .. + \Theta_p \mathbf{Z}_{t-p} + \boldsymbol{\epsilon}_t.$$

$$\boldsymbol{Z}_t = \begin{pmatrix} A_t \\ D_t \end{pmatrix}$$

  - positive decisions added to previous period stock

Out of sample forecasts





Methodology 4: gravity model (yearly)

- Different conditions caused ‘Gravity’ between the countries, which will pull or push the asylum seekers.
- Gravity Model: an intuitive framework to understand the determinants of flows between countries
  - Used in study international trade for analyzing economic circumstance relevant to movement of goods, services and people across the border
  - Growing extensively in the study of international migration and labor movements
  - Can be used to predict the number of refugee flows

Methodology 4: gravity model (yearly)

- Decisions reported to UNHCR in PSR (biannual)
  - projected based on panel data gravity model
  - apportioned to months through least squares distribution method
  - added to previous year stock

Gravity model

- Specification of the Gravity Model:
  - Primary variables
    - Economy of the origin and destination countries, often measured by the GDP
    - Population of the origin and destination countries
  - Secondary variables (commonly included)
    - Distance between the origin and destination countries
    - If the they share a contiguous border
    - If the they had a colonial relationship in history
    - If they share a common language
  - Variables to reflect conflicts or instability in origin countries
    - Number of Deaths of civilians in conflicts
    - Political stability
    - INFORM: risk assessment for humanitarian crises and disasters

UNHCR
The UN Refugee Agency

Gravity model of international migration

$$\mathbf{Ln}(\mathbf{X_{ij}}) = \mathbf{G} + \boldsymbol{\beta}_0\, \mathbf{Ln}(A_{\mathbf{ij(t-1)}}) + \boldsymbol{\beta}_1\, \mathbf{Ln}(\mathbf{Y_i}) + \boldsymbol{\beta}_2\, \mathbf{Ln}(\mathbf{Y_j}) + \boldsymbol{\beta}_3\, \mathbf{Ln}(Z_{\mathbf{i}}) + \boldsymbol{\beta}_4\, \mathbf{Ln}(\mathbf{Z_j}) + \boldsymbol{\beta}_5\, \mathbf{Ln}(\mathbf{D_i}) + \boldsymbol{\lambda}_{\mathbf{j}} + \boldsymbol{\chi}_{\mathbf{i}} + \boldsymbol{\gamma}_{\mathbf{ij}} * \mathbf{ln}(\boldsymbol{\phi}_{\mathbf{ij}}) + \mathbf{e_{ij}}$$

where:

X_{ij} represents flows (positive decisions) from country of origin i to country of asylum j;

A_{ij} represent previous year applications

Y_i represents GDP of country i;

Y_j represents GDP of country j;

Z_i represents population of country i;

Z_j represents population of country j;

D_i represents death in combat in origin country

Variables χ_i, γ_{ij} and λ_j are fixed effects associated to each country (origin and asylum), and year.

The ϕ_{ij} represents a vector of control variables associated with migration costs:

- distance, colonial ties, common language, common border

e_{ij} is the error associated to unobservables

Variables and data sources

Predictor Variables	Data source	links
Refugee population	UNHCR	http://popstats.unhcr.org/en/overview
GDP data	World Economic Outlook (IMF)	https://www.imf.org/external/pubs/ft/weo/2019/01/weodata/index.aspx
Population data	World Population Prospects	https://population.un.org/wpp/
Distance / If share contiguous border / Has ever Colonial / If use the same language	CEPII (Centre d'Études Prospectives et d'Informations Internationales)	http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8
Deaths of civilians in conflicts in country of origin	Uppsala Conflict Data	https://www.pcr.uu.se/research/ucdp/
ACLED	Armed Conflict Location & Event Data Project	https://www.acleddata.com/
……..		

UNHCR
The UN Refugee Agency

Prediction

Two datasets:

- Train Dataset:
  - Data from 2000-2017
    - - Modeling
    - - Predicting 2018 application and recognition
- Test Dataset:
  - Data for only 2018
    - - Comparing the prediction with actual application and recognition number

Prediction (contd.):

Heckman Correction

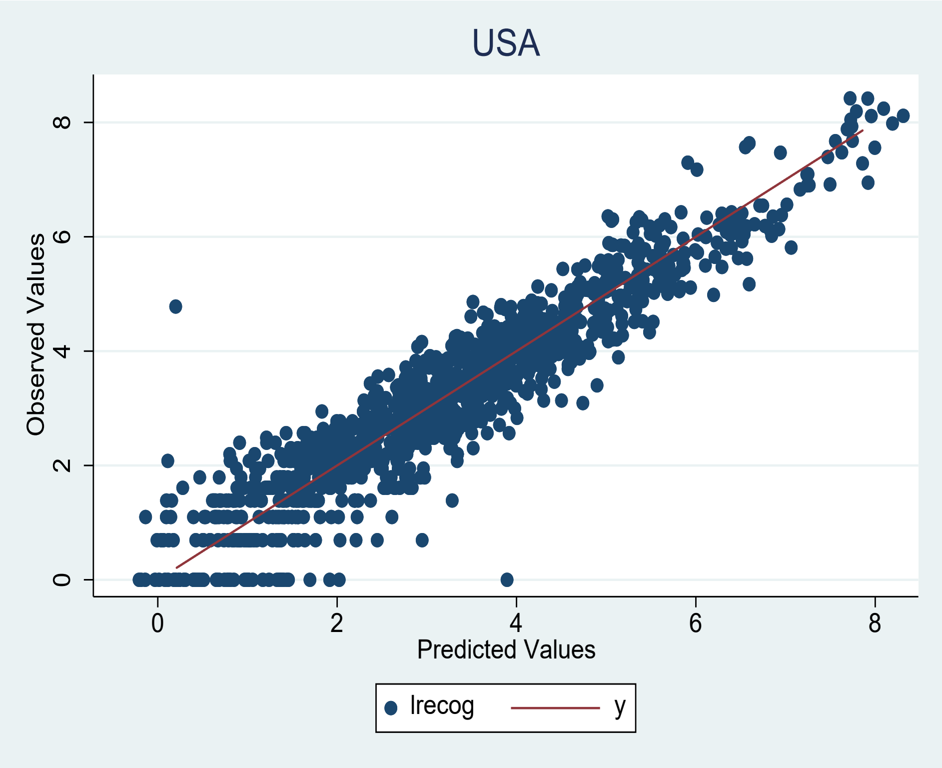
- To deal with the zero and missing values
- To explain the asymmetric flows of refugees within countries.

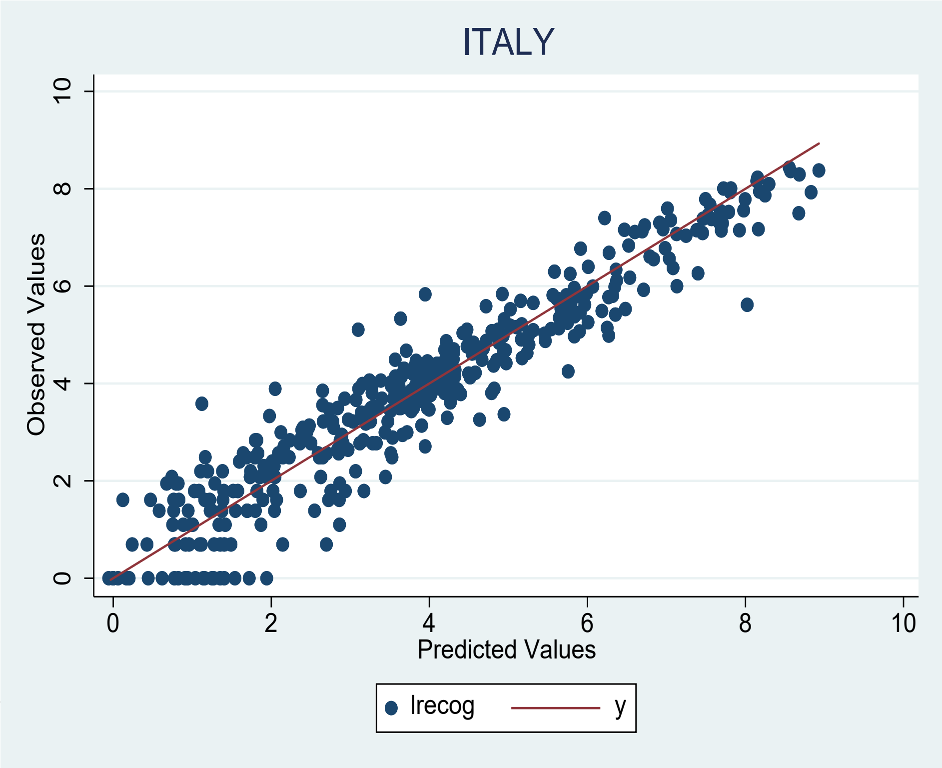
How it corrects the estimation bias of the other models we have used?

- 1st Step: Probit Regression
- 2nd Step: The result of the Probit Regression are incorporated in the model to correct the effect of zeros and NAs.
- We use colonial ties as exclusion restriction

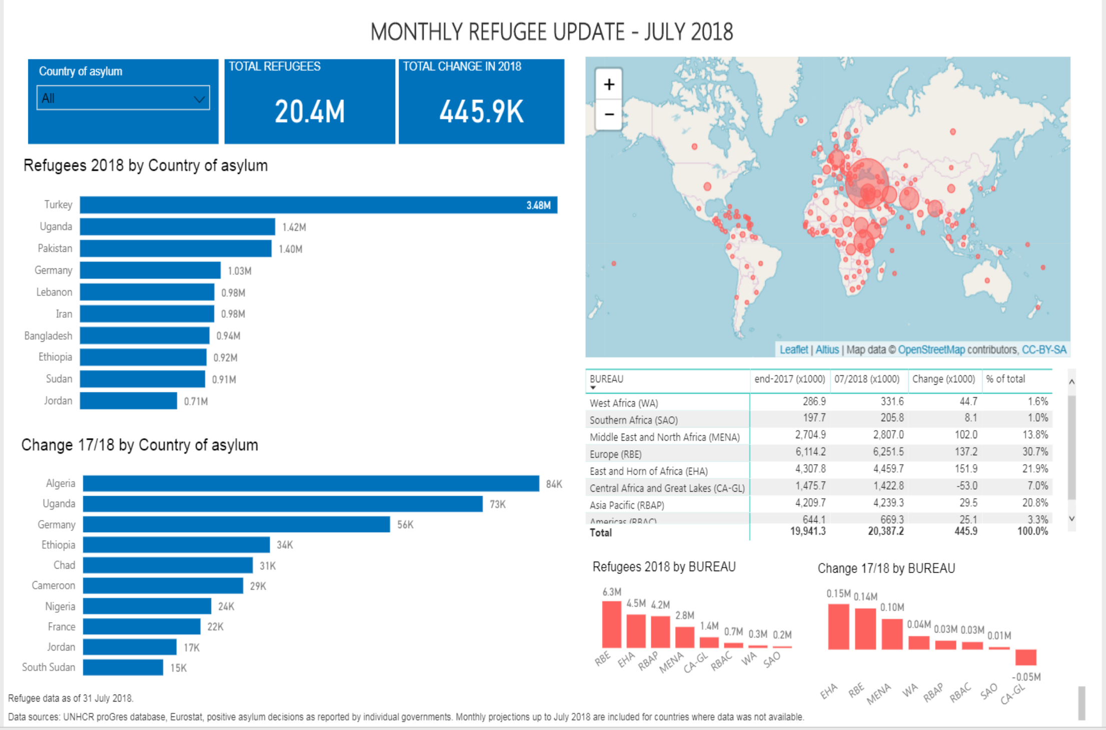
UNHCR The UN Refugee Agency

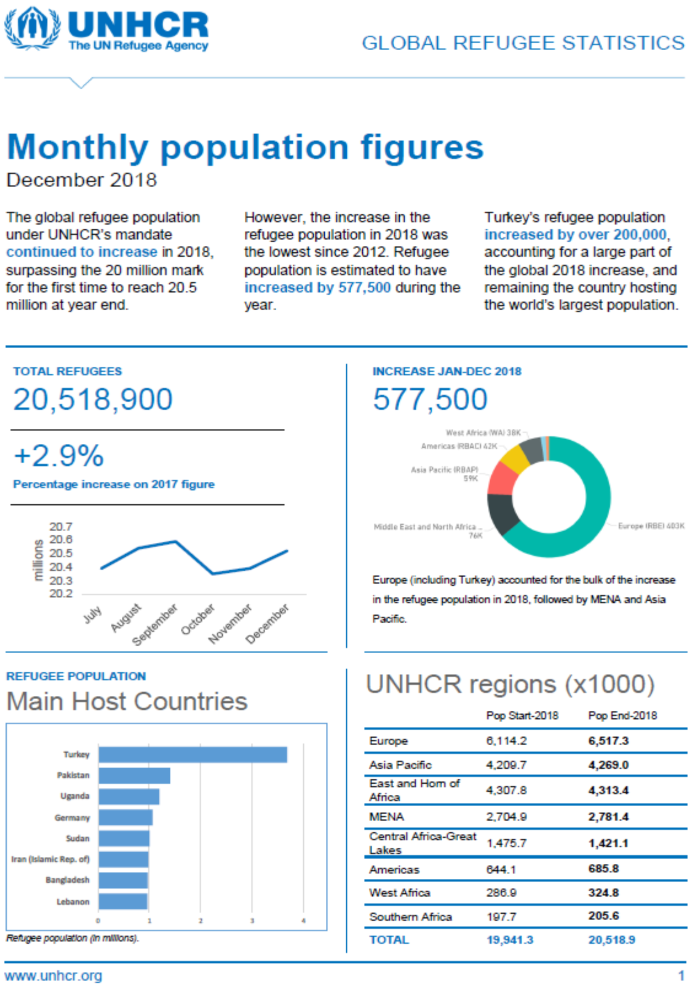
Prediction (contd.):
Heckman Correction





Dissemination





UNHCR The UN Refugee Agency

GLOBAL REFUGEE STATISTICS

Monthly population figures

December 2018

The global refugee population under UNHCR's mandate continued to increase in 2018, surpassing the 20 million mark for the first time to reach 20.5 million at year end.

However, the increase in the refugee population in 2018 was the lowest since 2012. Refugee population is estimated to have increased by 577,500 during the year.

Turkey's refugee population increased by over 200,000, accounting for a large part of the global 2018 increase, and remaining the country hosting the world's largest population.

TOTAL REFUGEES

20,518,900

+2.9%

Percentage increase on 2017 figure

INCREASE JAN-DEC 2018

577,500

Europe (including Turkey) accounted for the bulk of the increase in the refugee population in 2018, followed by MENA and Asia Pacific.

REFUGEE POPULATION

Main Host Countries

Refugee population (in millions).

UNHCR regions (x1000)

	Pop Start-2018	Pop End-2018
Europe	6,114.2	**6,517.3**
Asia Pacific	4,209.7	**4,269.0**
East and Horn of Africa	4,307.8	**4,313.4**
MENA	2,704.9	**2,781.4**
Central Africa-Great Lakes	1,475.7	**1,421.1**
Americas	644.1	**685.8**
West Africa	286.9	**324.8**
Southern Africa	197.7	**205.6**
TOTAL	**19,941.3**	**20,518.9**

www.unhcr.org

1

Current challenges

- Data not validated by operations, how to disseminate
- SDG indicators
  - Indicator 10.7.4 refugees as % of origin country population
  - Priority indicators disaggregated by forced displacement
- Data disaggregation by age, sex, geogr. location
- Estimation beyond refugees & asylum seekers (ie. IDPs)
- Exploring machine learning techniques

THANK YOU!