Bansak et al (2018), "Improving refugee integration through data-driven algorithmic assignment."

Sociology Statistics Reading Group

Hannah Postel

Princeton University

29 March 2018

Outline

1. Context
2. Data & methodology
3. Machine learning for public policy

Institutional/political context

- 65 million forcibly displaced, 21 million refugees

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"
- Three 'safe' outcomes: voluntary repatriation, local integration, or third country resettlement.

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"
- Three 'safe' outcomes: voluntary repatriation, local integration, or third country resettlement.
  - Achieved by less than 2% of refugees.

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"
- Three 'safe' outcomes: voluntary repatriation, local integration, or third country resettlement.
  - Achieved by less than 2% of refugees.
- Only 1 in every 200 refugees is resettled.

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"
- Three 'safe' outcomes: voluntary repatriation, local integration, or third country resettlement.
  - Achieved by less than 2% of refugees.
- Only 1 in every 200 refugees is resettled.
  - Only ~9% qualify (most vulnerable).

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"
- Three 'safe' outcomes: voluntary repatriation, local integration, or third country resettlement.
  - Achieved by less than 2% of refugees.
- Only 1 in every 200 refugees is resettled.
  - Only ~9% qualify (most vulnerable).
  - Of these, only ~14% are put forward (annually).

Institutional/political context

- 65 million forcibly displaced, 21 million refugees
  - Refugee: someone "forced to flee his or her country because of persecution, war, or violence"
- Three 'safe' outcomes: voluntary repatriation, local integration, or third country resettlement.
  - Achieved by less than 2% of refugees.
- Only 1 in every 200 refugees is resettled.
  - Only ~9% qualify (most vulnerable).
  - Of these, only ~14% are put forward (annually).
- With limited options, important to achieve best possible outcomes for those resettled.

Growing body of research

Transfer systems

- Countries buy and sell refugee resettlement quotas (Hathaway and Neve 1997, Schuk 1997, Bubb, Kremer, and Levine 2011, Moraga and Rapoport 2014).
- Explictly focused on political tractability over integration outcomes.

Growing body of research

Transfer systems

- Countries buy and sell refugee resettlement quotas (Hathaway and Neve 1997, Schuk 1997, Bubb, Kremer, and Levine 2011, Moraga and Rapoport 2014).
- Explictly focused on political tractability over integration outcomes.

Matching

- Building on Roth: kidney exchange, med school matching.
- Jones & Teytelboym 2016: two-sided match; local and/or international.

Growing body of research

Transfer systems

- Countries buy and sell refugee resettlement quotas (Hathaway and Neve 1997, Schuk 1997, Bubb, Kremer, and Levine 2011, Moraga and Rapoport 2014).
- Explictly focused on political tractability over integration outcomes.

Matching

- Building on Roth: kidney exchange, med school matching.
- Jones & Teytelboym 2016: two-sided match; local and/or international.

Bansak et al identify "practical barriers," noting: "in contrast, our data-driven approach can be immediately implemented by using existing data to optimize integration outcomes" (326).

Goals

Main factors affecting integration success: geographic context, individual characteristics, synergies between the two.

Goals

Main factors affecting integration success: geographic context, individual characteristics, synergies between the two.

Match refugees and locations to yield optimal outcomes (here, probability of employment).

Goals

Main factors affecting integration success: geographic context, individual characteristics, synergies between the two.

Match refugees and locations to yield optimal outcomes (here, probability of employment).

Compare two systems: The US, where resettlement agencies consider individual refugee characteristics in destination choice, and Switzerland, where asylum-seekers face proportional random assignment to cantons.

Goals

Main factors affecting integration success: geographic context, individual characteristics, synergies between the two.

Match refugees and locations to yield optimal outcomes (here, probability of employment).

Compare two systems: The US, where resettlement agencies consider individual refugee characteristics in destination choice, and Switzerland, where asylum-seekers face proportional random assignment to cantons.

Method:

Goals

Main factors affecting integration success: geographic context, individual characteristics, synergies between the two.

Match refugees and locations to yield optimal outcomes (here, probability of employment).

Compare two systems: The US, where resettlement agencies consider individual refugee characteristics in destination choice, and Switzerland, where asylum-seekers face proportional random assignment to cantons.

Method:

- Model
- Map
- Match

Data

United States

- Information on 33,728 working-age refugees resettled 2011-2016.
  - 43 locations: those receiving 200+ over entire period.
  - Covariates: sex, English-speaking, age, education, origin country, year & month of arrival, employment status.
  - Training through 2016Q2, 'free refugees' from 2016Q3 (n=919) as test.

Data

United States

- Information on 33,728 working-age refugees resettled 2011-2016.
  - 43 locations: those receiving 200+ over entire period.
  - Covariates: sex, English-speaking, age, education, origin country, year & month of arrival, employment status.
  - Training through 2016Q2, 'free refugees' from 2016Q3 (n=919) as test.

Switzerland

- Asylum seekers receiving subsidiary protection 1993-2013 (n=22,159).
  - 25 of 26 cantons.
  - Same covariates as above plus marital status, Christian/Muslim, French-speaking.
  - Training through 2012, 'free refugees' in 2013 (n=888) as test.

Data

United States

- Information on 33,728 working-age refugees resettled 2011-2016.
  - 43 locations: those receiving 200+ over entire period.
  - Covariates: sex, English-speaking, age, education, origin country, year & month of arrival, employment status.
  - Training through 2016Q2, 'free refugees' from 2016Q3 (n=919) as test.

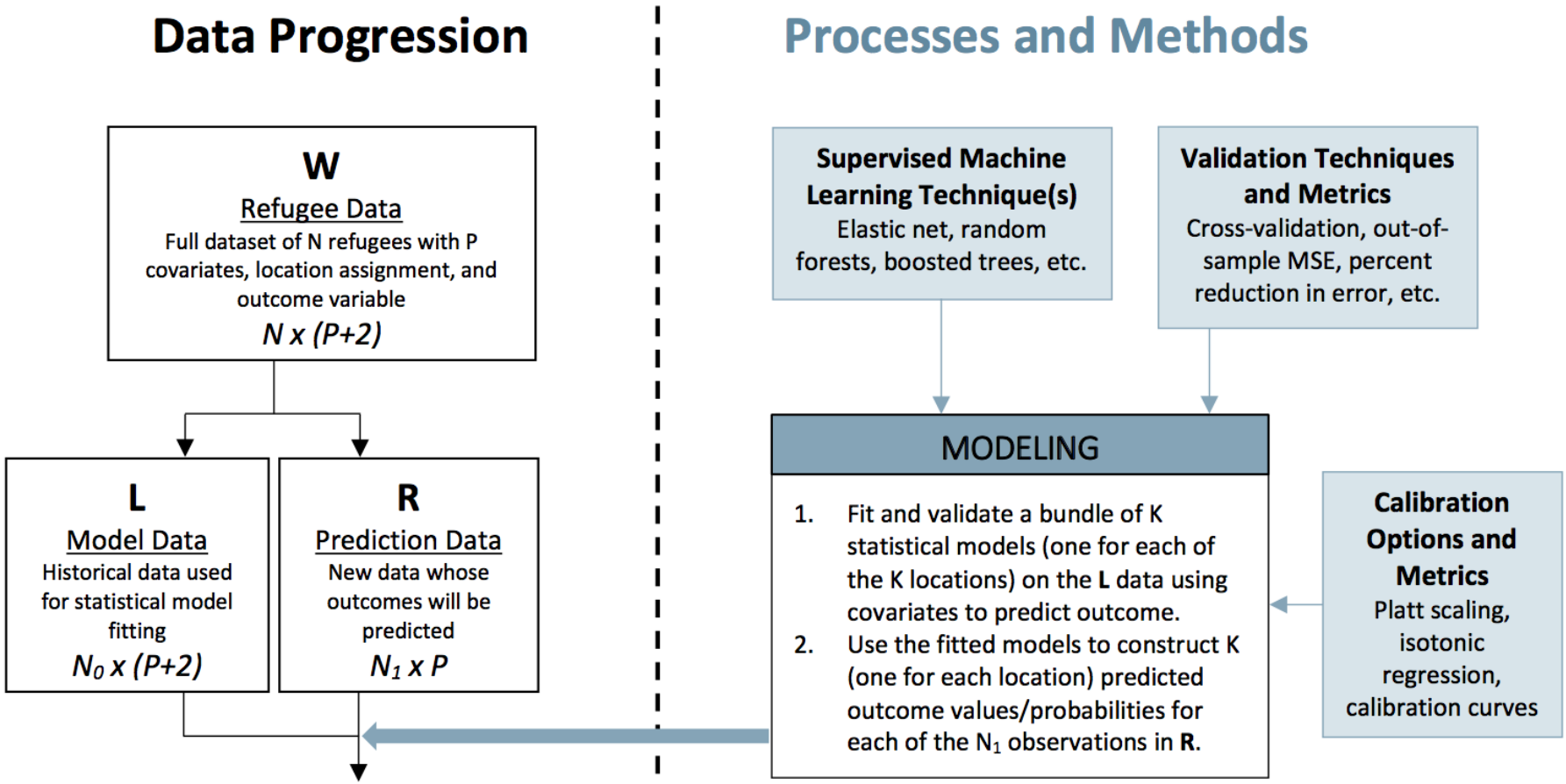
Switzerland

- Asylum seekers receiving subsidiary protection 1993-2013 (n=22,159).
  - 25 of 26 cantons.
  - Same covariates as above plus marital status, Christian/Muslim, French-speaking.
  - Training through 2012, 'free refugees' in 2013 (n=888) as test.

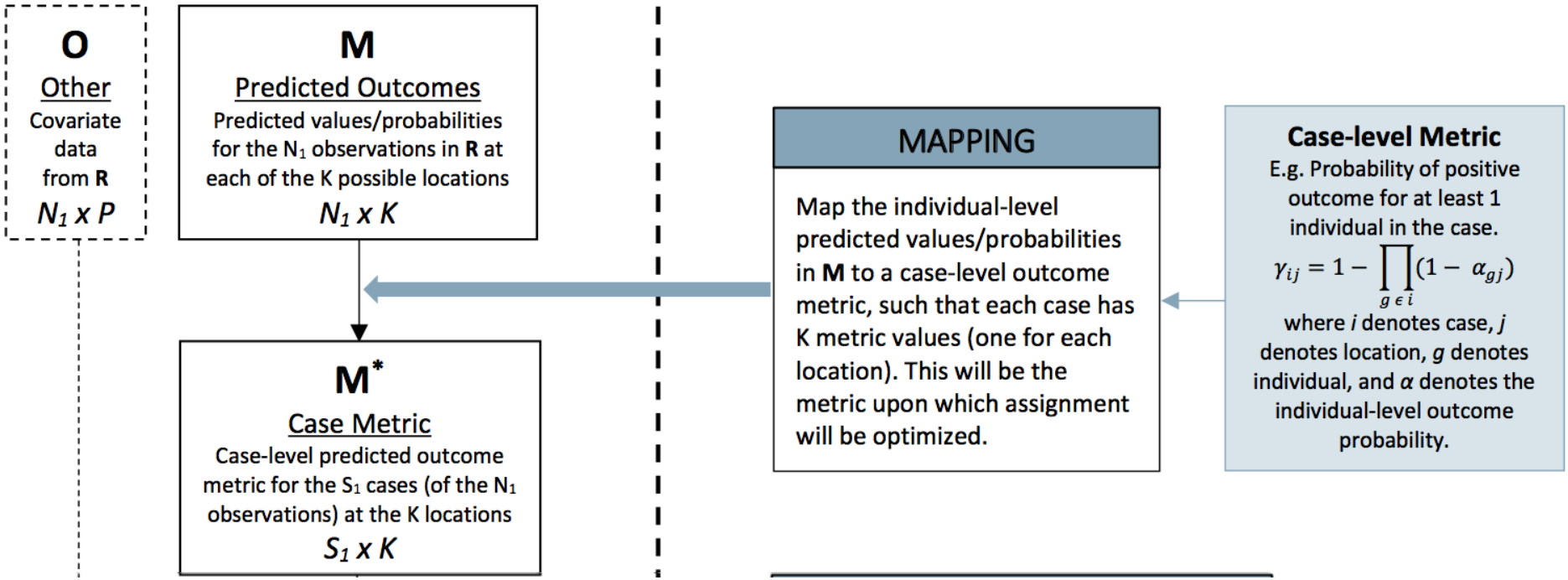
Training/test split is the key to this whole endeavor!

Step One: Model

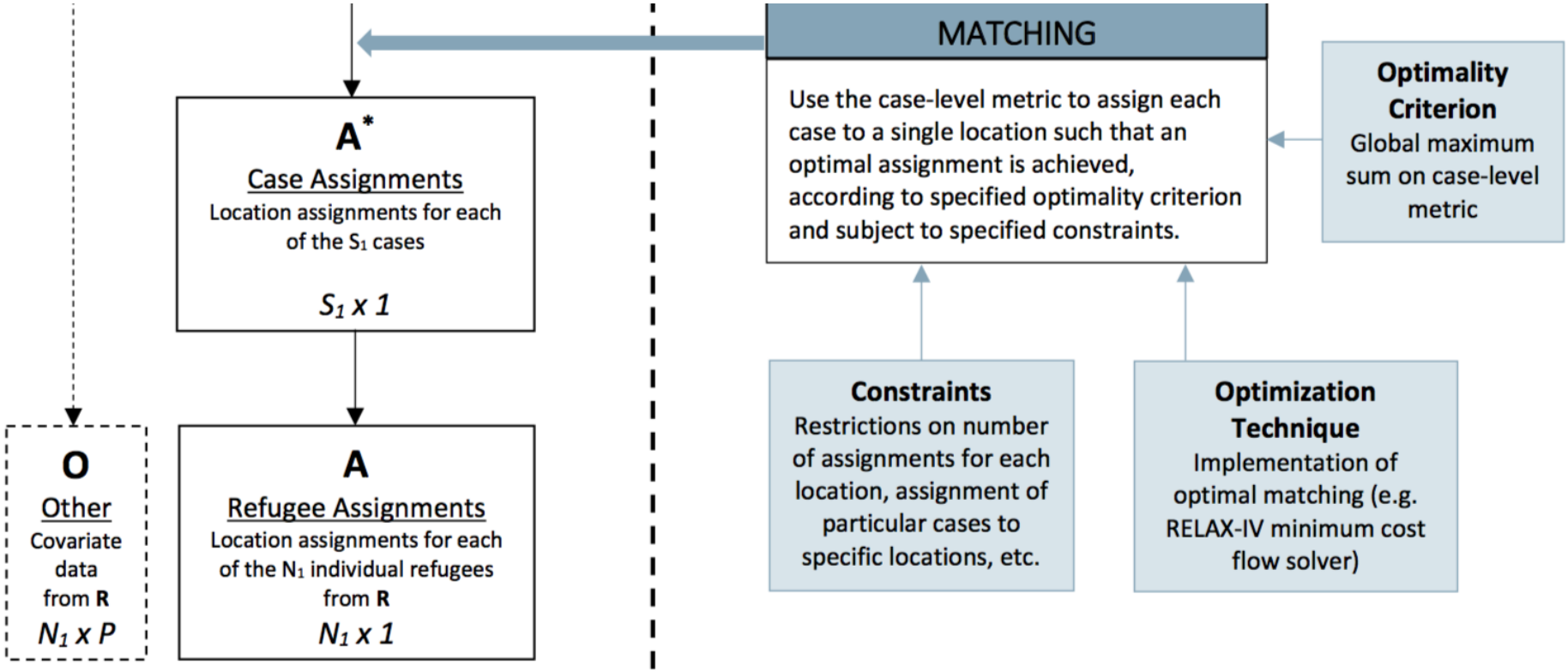
Fig. S3: **Data-driven algorithm for refugee assignment.** Figure shows a schematic of the main stages of the algorithm.



Step Two: Map


O
Other
Covariate data from R
$N_1 \times P$
M
Predicted Outcomes
Predicted values/probabilities for the N_1 observations in R at each of the K possible locations
$N_1 \times K$
M^*
Case Metric
Case-level predicted outcome metric for the S_1 cases (of the N_1 observations) at the K locations
$S_1 \times K$
MAPPING
Map the individual-level predicted values/probabilities in M to a case-level outcome metric, such that each case has K metric values (one for each location). This will be the metric upon which assignment will be optimized.
Case-level Metric
E.g. Probability of positive outcome for at least 1 individual in the case.
$\gamma_{ij} = 1 - \prod_{g \in i}(1 - \alpha_{gj})$
where i denotes case, j denotes location, g denotes individual, and α denotes the individual-level outcome probability.

Step Three: Match



Results

- Predicted probability of employment for median refugee approximately doubled in both countries.
- By US location, optimized assignment increased employment by 41% over baseline on average.

Results

- Predicted probability of employment for median refugee approximately doubled in both countries.
- By US location, optimized assignment increased employment by 41% over baseline on average.

This is huge, but. . .

Results

- Predicted probability of employment for median refugee approximately doubled in both countries.
- By US location, optimized assignment increased employment by 41% over baseline on average.

This is huge, but...

- Simulating outcome based on old system.

Results

- Predicted probability of employment for median refugee approximately doubled in both countries.
- By US location, optimized assignment increased employment by 41% over baseline on average.

This is huge, but...

- Simulating outcome based on old system.
- No one will ever agree to do an RCT of this.

Personalized policy

- Often we target policies to the average person, or a group based mainly on location.
- These methods enable us to (potentially) choose the best treatments for individual groups or people.
- Other applications you can think of?

Ethics

- Main pushback has been lack of refugee choice.
- Is this is still an improvement over the status quo?
- People get weird about computers making human-impact decisions.

Broader questions on social science for policy

- Impressive attempt to tailor to current context, but does not fully reflect reality in some key ways.
  - E.g. sociopolitical context at destination as major integration factor.
- Think-tank industry built to narrow distance between research and policy, but gaps remain.
- Should we still design research explicitly targeting policy even with missing/inaccurate information and/or low chance of implementation?