Methodology

Model per country of origin

Libraries, parameters, settings	1.1 Import necessary libraries 1.2 Define all parameters and choose various settings

Load data	2.1 Read all datasets
	2.2 Select countries to model, either manually or based on clustering INFORM risk index
Generate labels and features	2.3 Loop over selected countries
	2.4 Generate features specific to the country of origin
	2.5 Generate Origin to Destination bilateral labels and features
	2.6 For each country of origin, loop over top N host countries
	2.7 Merge labels and all features - origin-specific, bilateral, and destination-specific
	2.8 Include lag features, lead labels
Process labels and features	2.9 Drop features with too many missing values
	2.10 If not imputing missing feature values, drop samples with any missing feature values
	2.11 Divide dataset into training and testing set (currently, no validation set, but include if optimizing)
	2.12 Drop destination country that is not in training set?
	2.13 Define different transformation for different features (transformed in pipeline)
	2.14 If imputing, impute missing values now
	2.15 Drop training samples with missing labels

Train/test	3.1 Loop over multiple iterations of model training/testing
	3.2 Choose a regression model or an appropriate pipeline, e.g. feature transformation + regression
Save/plot outputs	4.1 Save data and forecasts
	4.2 Plot feature coefficients (e.g. linear regression) or feature importance (e.g. random forest)
	4.3 Plot per N-Year ahead forecast

Methodology

Model per cluster of countries of origin

| *Libraries,* | 1.1 Import necessary libraries |
| *parameters, settings* | 1.2 Define all parameters and choose various settings |

Load data 2.1 Read all datasets

2.2 Learn a model per cluster of countries of origin (clustering code in github)

Generate 2.3 For each cluster, loop over selected countries

labels and features 2.4 Generate features specific to the country of origin

2.5 Generate Origin to Destination bilateral labels and features

2.6 For each country of origin, loop over top N host countries

2.7 Merge labels and all features - origin-specific, bilateral, and destination-specific

2.8 Include lag features, lead labels

Process 2.9 Drop features with too many missing values

labels and features 2.10 If not imputing missing feature values, drop samples with any missing feature values

2.11 Divide dataset into training and testing set (currently, no validation set, but include if optimizing)

~~2.12 Drop destination country that is not in training set?~~

2.13 Define different transformation for different features (transformed in pipeline)

2.14 If imputing, impute missing values now

2.15 Drop training samples with missing labels

Train/test 3.1 Loop over multiple iterations of model training/testing

3.2 Choose a regression model or an appropriate pipeline, e.g. feature transformation + regression

Save/plot outputs 4.1 Save data and forecasts

4.2 Plot feature coefficients (e.g. linear regression) or feature importance (e.g. random forest)

4.3 Plot per N-Year ahead forecast