Pairing Administrative Datasets with Google Trends to Infer Migrant Flows and Sending Country Impacts

Susan Godlonton
Williams College

Caroline Theoharides
Amherst College

January, 2018
OECD IFMS Presentation
Research Overview

• Are Google Trends data a valid proxy for:
  • migration intensity, and
  • destination specific contract migrant worker flows?

• How do we address this question?
  • Link rich administrative data from the Philippines on contract migration to Google Trends search data

• Why important?
  • Individual micro-level data on migration akin to that in the Philippines is rare; migration data collection can be costly
  • Potentially open up low-cost opportunities for quantitative analysis
  • Enable research on migration questions in under-studied contexts
  • Understand opportunities and limits of google trends as a measurement tool for migration
Philippines Context and Data

• Philippines Migration:
  • Philippines is one of world’s largest senders of migrants
  • Wide variety of destinations and occupations

• Rich administrative data available:
  • Individual micro-level data on contract migrants from the Philippines from 1992-2016
  • Data include country of destination and municipality of origin
  • Enables measurement of destination-specific historical migration density by municipality
Google Trends Data

• Extractions
  • Spans time period 2004 – present
  • Available at monthly level in Philippines (aggregate to year)
  • Maximum # searches indexed to 100 ➞ i.e. doesn’t provide actual number of searchers but unit of observation with highest search count set to 100; others indexed to the max

• Two types of extractions:
  • Search for “OFW”, “Work abroad” “Jobs abroad” disaggregated by region in the Philippines (i.e. unit = region-month)
  • Search for work abroad by destination country “X” for whole country (i.e. unit = country-month)
Approach

• Compare search index with administrative data to see if searches for migration related terms are a good proxy for
  • Migration intensity by region; and
  • Migration destination flows

• Use non-parametric analysis to compare data sources
  • Migration intensity by region-year: Compare three search terms “OFW,” “jobs abroad,” “work abroad” to actual flows
  • Migrant destination flow by year: Compare “Work in X”
Migration intensity comparison (region-year)

- Jobs abroad
- Work abroad
- OFW
Migrant destination (region-year)
What we learn?

• **Opportunities: What we can do with google trends data?**
  • Variation in search intensity mimics variation in actual migrant flows
    → research utilizing variation in migrant flows seems plausible
    → mindful of local terminology (variation between OFW vs work abroad)
  • Variation in destination country mimics variation in low to mid-range destinations
    → High destination countries: other sources of information more important (peers, recruiters)

• **Limitations: What we cannot do with google trends data?**
  • Levels are indexed → can’t extract levels of migration
  • Not ideal for capturing high destination country variation; and high migrant intensity locales