Analysis of Migration Flows

Griffith Feeney

The Population Census

- Is by far the best source for data on migration flows
- Even in countries with high mobility, relatively small proportions of the total population migrate
- Even small numbers of units give hundreds or thousands of migration streams; few surveys are large enough to deliver this data

Standard Format for Migration Flow Tables: 1

- Some variation on *place of previous residence* by *place of current residence, for example* ...
- **Total population** *by* **place of birth** *and* **place of current residence**
- **Population five years old and over** *by* **place of residence five years ago** *and* **place of current residence**

Standard Format for Migration Flow Tables: 2

- **Total population** *by* **place of previous residence** *and* **place of current residence**
- **Population five years of age and under** *by* **place of birth** *and* **place of current residence** (*why this one?*)
- List of places of previous residence may or may not match place of current residence; any two “places” define two “flows”

Availability and Analysis

- Tables of this kind have been produced for many decades
- One would suppose methods for analyzing them would be well developed, well known, and widely applied
- This seems not to be the case, however; do you know of methods, studies?

Caveats

- In all references to the word “migrant” pay very close attention to *precisely* what is referred to
- Pay close attention to how finely “place of previous residence” is specified on the census questionnaire
- Beware of large numbers of “not stated” responses for place of previous residence

The Essential Challenge

- Too many flows; with only 20 regions, we have nearly 400 flows
- For many purposes we would like much more detail than 20 subnational units
- How to avoid being overwhelmed by this level of detail? How to use the data to tell an intelligible story?

First Cut

- The first solution is evasive; forget flows as such and ask *how much migration is there?*
- Define a migrant as any person counted in an off-diagonal cell and compute migrants as a percentage of total population in all cells
- Expect this figure to be less than ten percent, perhaps less than one percent

Second Cut

- Now look at flows; list them all, sort them in descending order of magnitude, and calculate the cumulative percentage of migrants accounted for by the k largest flows for each $k = 1, 2, \ldots$
- Plot the cumulative percentages against k; focus on the largest flows; see if the ‘80-20’ rule applies

Third Cut: 1

- Consider how many migrants we *expect* in each flow
- One major influence is the population of the sending and receiving areas
- Other things being equal, an area with greater population will tend both to send and to receive larger numbers of migrants

Third Cut: 2

- Suppose that relative sizes of the populations of the different regions were the *only* influence on the size of streams
- What flows would we observe in this case?
- Every place would receive migrants from and send migrants to other places in proportion to the population of those places

Third Cut: 3

- How to work out the numbers of migrants in each flow this implies?
- Construct a table with *origins for rows* and *destinations for columns*
- Assume origin and destination units are the same
- Let p_i denote the population of the i-th unit as a proportion of total population

Third Cut (*continued*)

- Compute the product $p_i p_j$ and enter in the ij-th cell of the table
- Put the diagonal entries to zero and 'normalize' the remaining entries
- The resulting table has the desired property; in-migrants in proportion to population of place of origin, out-migrants in proportion to place of destination

Third Cut (*continued*)

- Apply these "model" proportions to the total number of migrants to obtain model *numbers* of migrants
- Express actual numbers of migrants as the sum of the model numbers and a residual term: Actual = Model + Residual, or …
- **Residual = Actual - Model**

What Have We Accomplished?

- If relative size of populations of units were the *only* influences, the residuals would all be zero
- Deviations of residuals from zero thus indicate the operation of *other* influences
- In looking at the residuals, in other words, we have “controlled for” size of population as an influence on size of migration flow

The Question for Data Analysis

- Is therefore: which residuals are **BIG!**
- Big residuals tell us that something other than relative population size is operating; we can then explore *what*
- If we find a small number of relatively large residuals, we can zero out these streams and repeat, aiming for a ‘well behaved’ set of residuals

How to Look at Residuals?

- For small sets of numbers, under 50 or so, use a (hand or computer constructed) **stem and leaf plot**
- For larger sets of numbers the best display is probably the **quantile plot**; see William S. Cleveland's ***Visualizing Data***, pages 17-20; unfortunately, your spreadsheet probably doesn't provide this feature

For An Example

- Of an analysis of this kind see the notes on Myanmar

Review of Key Points

- Migration flows reflect three kinds of influences: the **overall level** of migration
- the **relative size** of the units between which flows are defined
- **Other influences** on migration
- We are most interested in the last; we need to *control for* the first two influences to see the effect of other influences

Questions?
Comments?
Discussion?