



**United Nations Economic Commission for Europe
Statistical Division**

Other Methods to Estimate Emigration

Jason Schachter, Statistician

United Nations Economic Commission for Europe

Workshop on Migration Statistics

Chisinau, Republic of Moldova, 8-9 September, 2014

Measurement

- Emigration is extremely difficult to measure (no country does it well)
- In addition to complete registration data, there are several regularly used methods of estimation
 - survey questions in countries of origin about persons living abroad
 - consular (or similar) data in countries of destination
 - migration data from destination countries

Some alternative methods

- ❖ Residual Methods
 - Population in country of origin
 - Population in countries of destination
- ❖ Panel Attrition Methods
- ❖ Indirect Methods
- ❖ Statistical Modeling

Residual Methodology, aka Intercensal Cohort-Component Method



- “Simplest” way to measure **net** emigration
- Resident population counted at two points in time (using the same data source, usually a census)
- Any difference in population size not attributable to births or deaths is due to net migration.
 - Accurate birth and death information (birth and death rates) needed
 - Each data source (e.g. census) needs to have extremely similar coverage.

Formula

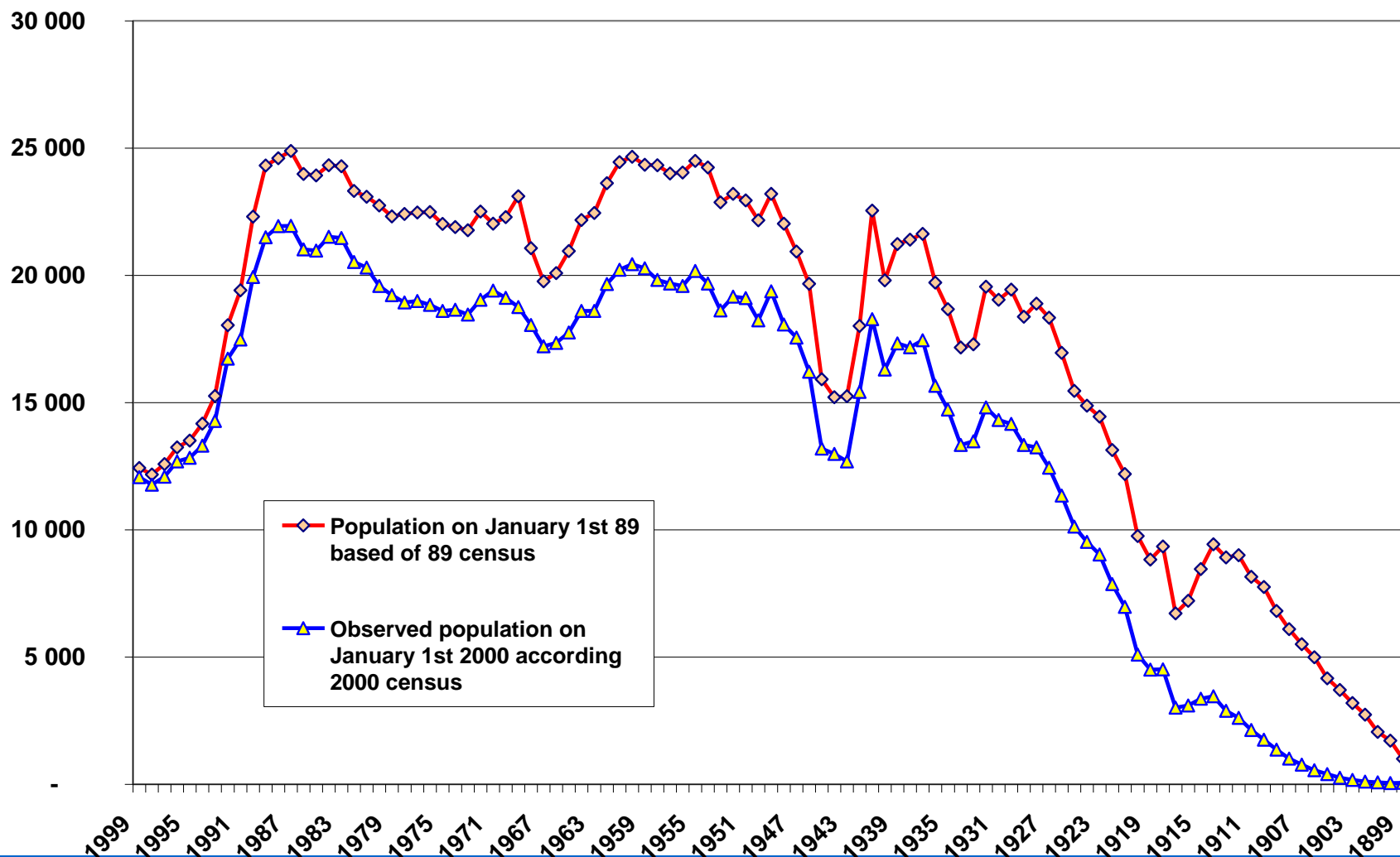
- ❖ Emigration is the difference or residual between the *expected* population and the *actual* population enumerated in the later census

- ❖
$$E_{t1-t2} = P_{t1} + B_{t1-t2} - D_{t1-t2} - P_{t2}$$

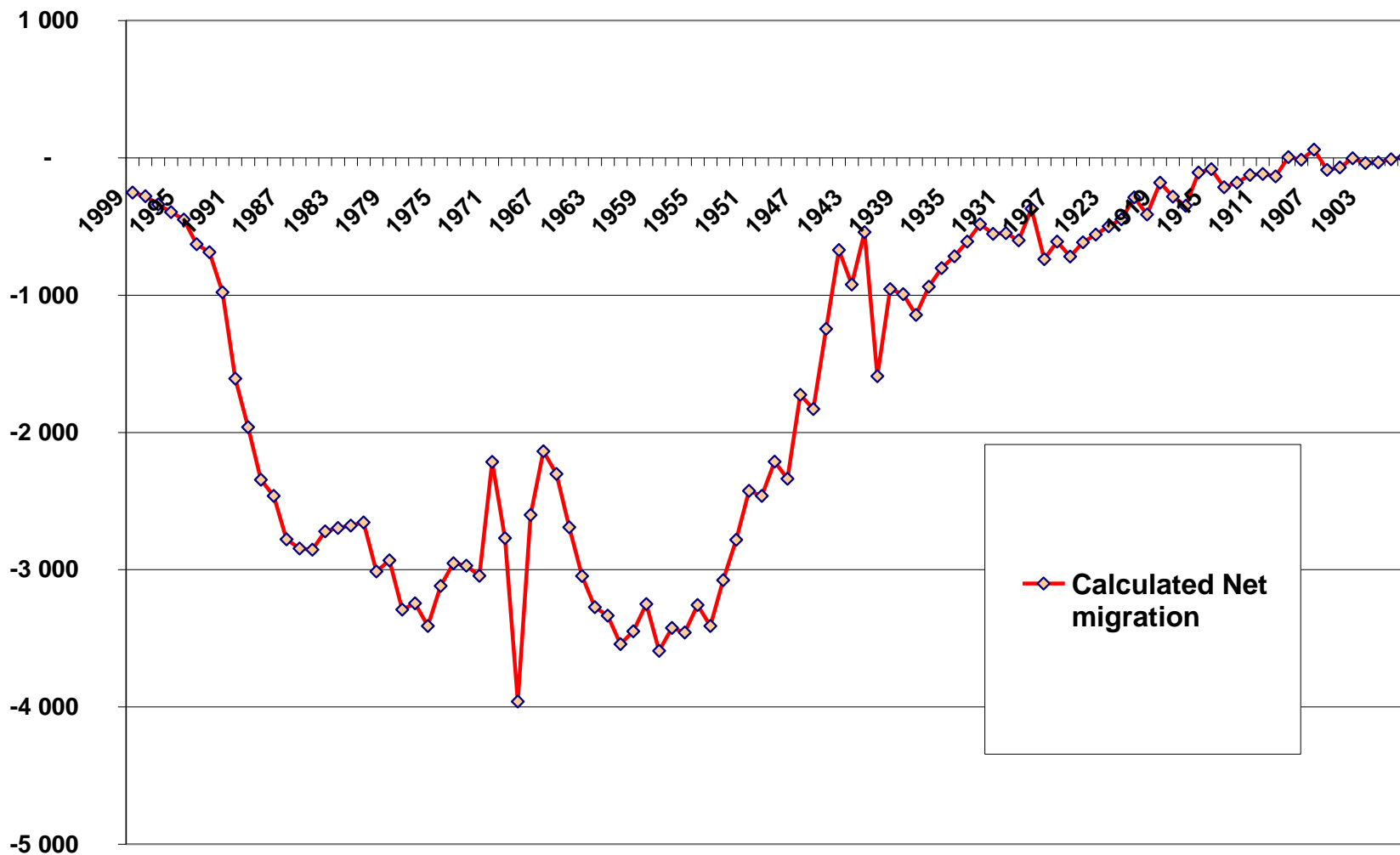
Conceptual Example

- ❖ Population change = $B(\text{irths}) - D(\text{eaths}) + / - N(\text{et})M(\text{igration})(I - O)$
- ❖ T_1 : population = 100,000
- ❖ T_2 : population = 105,000
- ❖ Between T_1 and T_2 : 15,000 births, 5,000 deaths
- ❖ Residual component = Net migration (-5,000) (Net emigration = 5,000)

Estonia Example 1989 & 2000 (Poulain 2006): Age structure of population



Age-specific net migration for Estonia between two censuses



Limitations

- Coverage error (under or over)
 - For a high emigration country, improved coverage from one Census to the next might underestimate true levels of emigration, as population counts might be larger due to improved coverage, not lower emigration.
- Provides little information about other emigrant characteristics (e.g. education) or the size of the in- and outflows.

Emigration to countries of destination

- ❖ Method (modified) can also be used to estimate native/citizen emigration using data from destination countries.
- ❖ Method repeated for all destination countries of interest (major destinations)

Methodology

- ❖ Observe stock of persons from a specific country of origin at two points in time (T_1 and T_2) from similar sources (e.g. Census)
- ❖ Survive T_1 population of interest to T_2 (using age and sex specific life table survival rates)
 - For foreign-born estimates birth rates not needed since children born abroad considered native
- ❖ Difference between survived and observed population is “net migration” to a specific country
 - Divide by time period for “rough” estimate of annual net migration
 - ◆ Combine with annual inflow data to country of origin to estimate outflows



Project for US Census Bureau: Using international stock data to estimate “net migration”

- ❖ Investigated feasibility of using international data sources to estimate the number of US citizens moving abroad (2005)

- ❖ Requested data on US citizens and/or US born from two most recent censuses: Canada, France, Italy, Poland, and the United Kingdom
 - UNECE/EUROSTAT data exchange initiative

Project expanded (2008) to include data from all countries with available data and estimates used for US Census Bureau’s estimates of native emigration

Specific Methodology

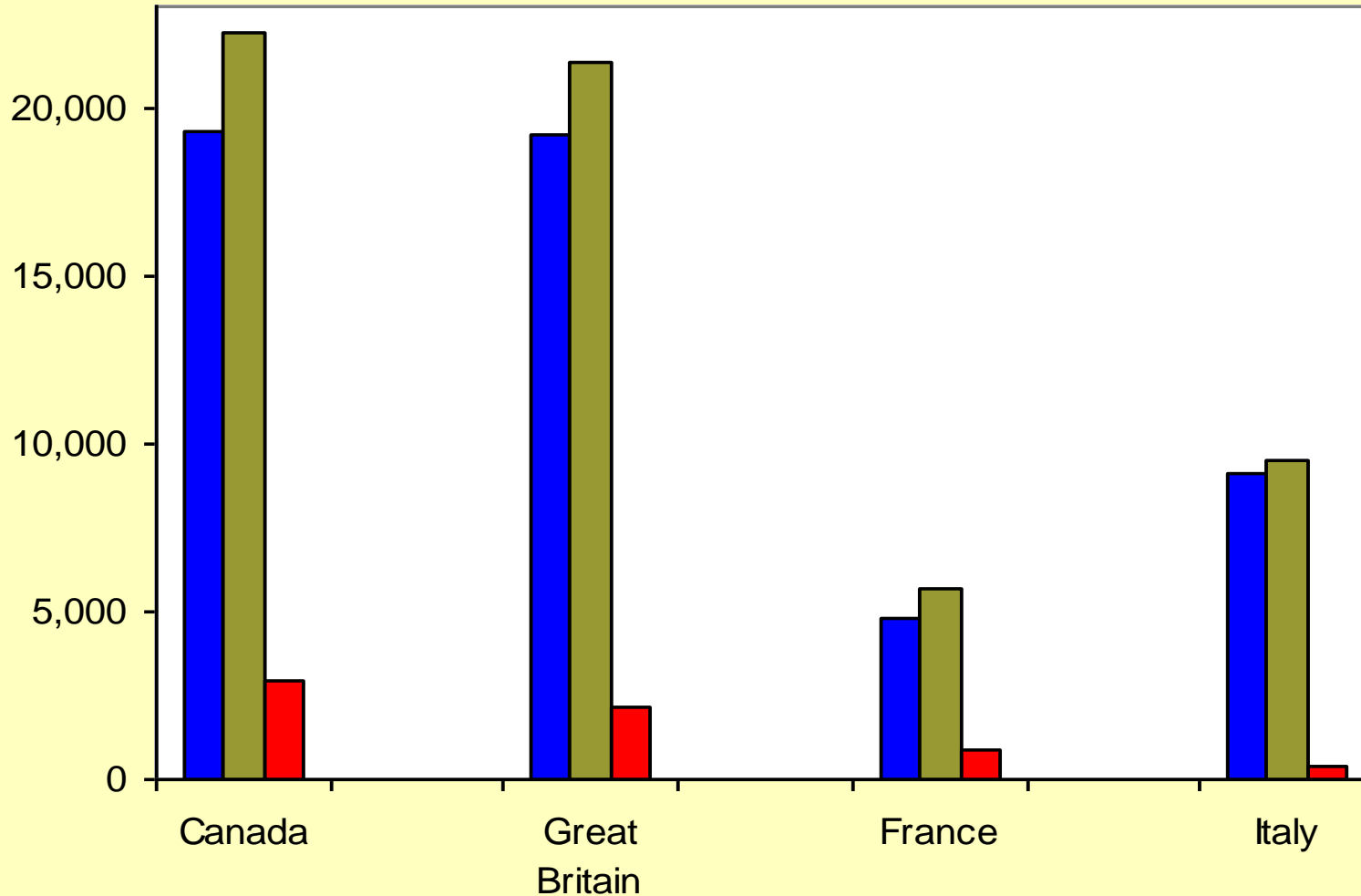


- ❖ From two most recent Censuses, establish a T_1 (circa 1990 Census) population of US born and/or citizens, and a similar T_2 (circa 2000 Census) observed population.
- ❖ **US born:** T_1 population survived (using age and sex specific death rates based on the resident US population at T_1) a number of years equal to the difference between T_1 and T_2 .
- ❖ Difference between survived T_1 and observed T_2 population is assumed to measure “net international migration” between the US and given country over the T_2-T_1 time period.
 - Figure then divided by the time period to yield an average annual net migration figure

Methodology (cont.)

- ❖ **US citizens:** same methodology with two additional components:
 - Birth of children to US female citizens (added to the survived T_1 total)
 - US citizens who naturalized to another country –and revoked US citizenship- (subtracted from the survived T_1 total).
- ❖ Proved difficult to incorporate additional components
 - Number of births to US females is not necessarily equal to number of new US citizens (also misses potential births to non-US nationals married to US males).
 - Not able to find public records on number of US citizens who naturalized in these countries.

Estimated Annual Inflow, Outflow, and Net Migration of US Born to/from the United States



■ Estimated Inflow ■ Estimated Outflow ■ Average annual net migration gain from US

Method had a number of assumptions/limitations

- ❖ US population living abroad has the same age and sex-specific death and birth rates as the US resident population
- ❖ Made assumptions about the age distribution of Americans living abroad (people were equally distributed within each age cohort)
- ❖ Improved coverage (measurement) of foreigners/foreign-born from one Census to next might be the actual reason for apparent “net migration gain”
- ❖ Measurement error around figures, from sample based data, could result in “net migration”
- ❖ Limited information on dual-citizens

Panel Attrition Methods

- ❖ Attrition from administrative data or household surveys
 - e.g. foreign born (to estimate return migration)
- ❖ Longitudinal data
 - Decompose missing data into its components, including emigration
- ❖ Provides short-term estimates

Limitations

- ❖ Administrative data systems and panel surveys are typically not designed for measuring migration, therefore, data sources might not contain all data necessary to estimate emigration without making huge data assumptions.

Example: CPS matching method (Van Hook 2006)

- ❖ Used to estimate emigration of foreign born from USA, but also applicable to general population
- ❖ CPS Labour Force Survey
 - Quasi-longitudinal design (respondents in (4 months), out of (8 months), and in (4 months) sample for period of time)
 - Sampling frame made up of addresses, not individuals, therefore, if respondents move to a new address they drop out of sample.

Estimation Methodology

- ❖ Calculate probability that a respondent will not be included in subsequent waves of the survey to estimate emigration.
- ❖ Estimate probabilities of internal migration, death & non-response: rest is emigration
- ❖ Emigration probability then applied to total (foreign born) population to get estimates of those who left country

Indirect Estimation Methods

- ❖ Estimate emigration using responses to questions on household surveys about the residence of household members or relatives (e.g. Zaba 1985, Zlotnik 1987)
 - Method 1: Mothers asked current residence of all **children** ever born (adjust for orphans and emigrated mothers)
 - Method 2: All respondents asked current residence of **siblings** (adjust for multiple reporting or emigration of emigrant's siblings)

Child Method

- ❖ Number of emigrants calculated by dividing number of emigrant children with a surviving mother by the proportion not orphaned in each age group.
 - Assumes likelihood of emigration and being an orphan is independent.
- ❖ Emigration of children whose mother has also emigrated can be estimated if the survey collects information on the residence of the mother.
 - Assumes likelihood of child emigrating is independent of the likelihood of the mother emigrating. Problematic for young emigrants that are likely to migrate with their family (underestimates emigration)

Statistical Modeling

- ❖ Used to produce migration estimates, especially when data are missing or incomplete
 - Individual level: likelihood an individual will migrate (esp. age)
 - Aggregate level (geographers): probability that a migration stream originated in a particular region or country (migration flow matrix: models estimate missing cells)
 - ◆ Demographic fixed-rate, gravity, entropy maximization, and log-linear models

Advantages

- ❖ Overcome issues of sparse or missing data
- ❖ Can highlight migration patterns that can be generalized to other populations
- ❖ Help assess the validity of estimates

Disadvantages

- ❖ Individual: data needs to be complete (migration history)
- ❖ Aggregate: data needs to be harmonized between countries
- ❖ A number of assumptions are used (e.g. “expert opinion”), which can yield different results
- ❖ Technically difficult
- ❖ Place in official statistics?

Conclusions

- ❖ A number of alternative methods to estimating emigration
- ❖ All have their own limitations
- ❖ Some are rarely used and difficult to compute
- ❖ A combination (triangulation) of different sources and methods could yield best estimates of emigration