# Models of migration

Frans Willekens

**Colorado Conference on the Estimation of Migration**

24 – 26 September 2004

# Introduction

- Migration : change of residence (relocation)
- Migration is situated in time and space
  - Conceptual issues
    - Space: administrative boundaries
    - Time: duration of residence or intention to stay
  - Measurement of migration
    - Event: ‘migration’
    - Person: ‘migrant’

# Data types applied to life-history data: the case of migration

# Data types applied to migration

- Micro-data: data on individuals or households
  - **Status data**:
    - Current status (*state occupied*)
      - migrant status (e.g. ever migrated / never migrated in given period)
      - current place (region) of residence
    - Place of residence at two points in time: transition data (***migrant data***)
      - Time interval of fixed length: e.g. census and 5 years prior
        $\Rightarrow$ "Where did you live 5 years ago?"
      - Time interval variable: e.g. census and place of birth
        $\Rightarrow$ "Place of birth"
    - Place of residence at 3 or more points in time

# Data types applied to migration

- Micro-data:
  - **Event data** : ***migration data (movement data)***
    - Migration during given period (yes/no): migrant status
    - Ever migrated?
    - Number of migrations (quantum)
    - Timing of migration (tempo)
      - Time scale: calendar time, age, process time (time since event-origin)
      - Measurement of time: exact time, time interval (discrete time, e.g. month, year)
      - Timing of all migrations vs timing of last migration

# Data types applied to migration

- Grouped data: data on groups of individuals or households (actors)
  - **Status data**:
    - Current status: number of actors (subjects) in given status
    - Number of actors by place of residence at two points in time: transition data (migrant data) ***CENSUS***
    - Number of actors by place of residence at 3 or more points in time
  - **Event data**:
    - Number of events during given period ***POP. REGISTER***

# The Lexis diagram

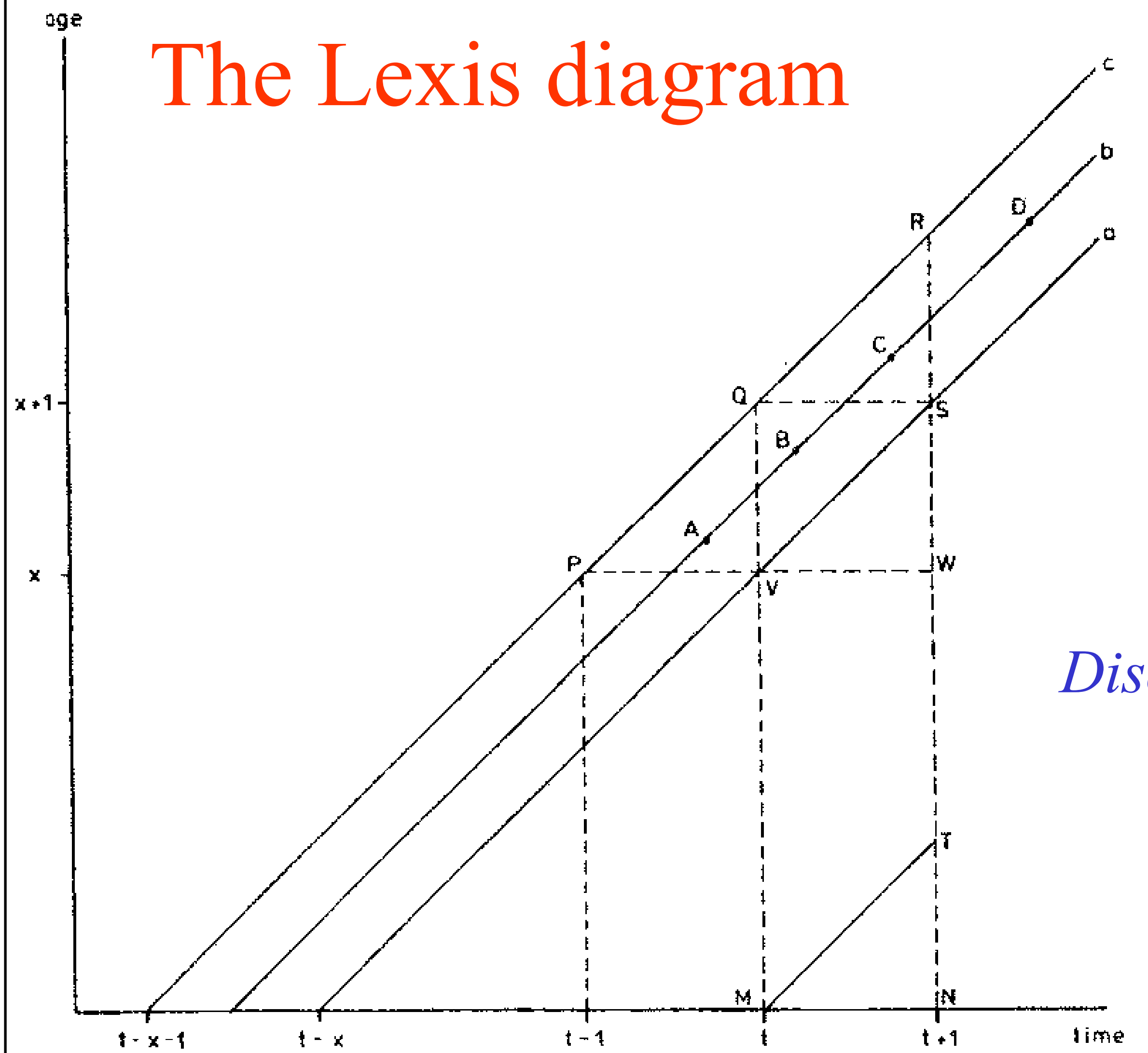


Figure 2 Lexis diagram

*Discrete age/time*

# Individual as ‘carrier’ of attributes
# Life course as a sequence of attributes

- Personal attribute at age x or time t
- Attributes change in the course of life: events
- Describe life course
  - By attributes at each age: **status-based approach**
  - By initial attributes and changes in attributes: **event-based approach**
- Changes (and events) occur in *continuous time* but they are often measured in *discrete time*

# State-space approach to life histories

- Attribute = state
- Set of possible attributes = state space
- Attribute at age x = state occupied at age x
  - State occupancy
- Change in attribute = state transition
  - Direct transition
  - Discrete-time transition
  - Attributes of transitions: state of origin, state of destination, reason
- Timing of state transitions
  - *Age structure of migration* (age/duration dependence)
- Dependence of destination on origin
  - *Spatial structure of migration*

# Probability models of migration

- Risk indicators: risk of a transition
  - Number of transitions during unit interval: counts
  - How likely is a transition in unit interval: probability
  - Timing of transition: transition rate
- Underlying random mechanism
  - Count data: Poisson models
  - Probability (or proportion): logit model and logistic regression
  - Rate: transition rate model
    - Rate = occurrences / exposure

# Model 1: state occupancy

- $Y_k(x)$ State occupied at x
- ${}_k\pi_i(x) = \Pr\{Y_k(x)=i\}$ State probability
  - Identical individuals: ${}_k\pi_i(x) = \pi_i(x)$ for all k
  - Individuals differ in some attributes:

    ${}_k\pi_i(x) = \pi_i(x,Z)$, Z = covariates
    - Prob. of residing in i region by region of birth
- Statistical inference: MLE of $\pi_i(x)$
  - Multinomial distribution

$$\Pr\{N_1 = n_1, N_2 = n_2, \ldots\} = \frac{m!}{\prod_{i=1}^{I} n_i!} \prod_{i=1}^{I} \pi_i^{n_i}$$

# Model 1: state occupancy

- Statistical inference: MLE of state probability $\pi_i$
  - Multinomial distribution

$$\Pr\{N_1 = n_1, N_2 = n_2, \ldots\} = \frac{m!}{\prod_{i=1}^{I} n_i!} \prod_{i=1}^{I} \pi_i^{n_i}$$

  - Likelihood function $L = \prod_{i=1}^{I} \pi_i^{n_i}$
  - Log-likelihood function $l = \ln(L) = \sum_{i=1}^{I} n_i \ln(\pi_i)$
  - MLE $\hat{\pi}_i = \frac{n_i}{m}$
  - Expected number of individuals in i: $E[N_i]=\pi_i m$

# Model 1: State occupancy with covariates

$$\log it[\pi_i(Z)] = \ln \frac{\pi_i(Z)}{1-\pi_i(Z)} = \eta_i = \beta_{i0} + \beta_{i1} Z_1 + \beta_{i2} Z_2 + \beta_{i3} Z_3 + \ldots$$

$$\pi_i = \frac{\exp(\eta_i)}{\exp(\eta_1) + \exp(\eta_2) + \ldots + 1 + \ldots} = \frac{\exp(\eta_i)}{\sum_{j=1}^{I} \exp(\eta_j)}$$

multinomial logistic regression model

# Model 2: Transition probabilities

- State probability ${}_k\pi_i(x,Z) = \Pr\{Y_k(x,Z)=i \mid Z\}$

- Transition probability

$$Pr\{Y(x+1) = j \mid Y(x), Y(x-1), \ldots; Z\} = Pr\{Y(x+1) = j \mid Y(x); Z\}$$

$$\Pr\{Y(x+1) = j \mid Y(x) = i\} = p_{ij}(x)$$

*discrete-time transition probability*

Migrant data; Option 2

- Transition probability as a logit model

$$\operatorname{logit}[\pi_j(x+1)] = \beta_{j0} + \beta_{j1} Y_i(x) \qquad p_{ij}(x) = \frac{\exp\left[\beta_{j0} + \beta_{j1} Y_i(x)\right]}{\sum_{r=1}^{I} \exp[\beta_{j0} + \beta_{j1} Y_r(x)]}$$

# Model 2: Transition probabilities

- Transition probability as a logit model

$$\operatorname{logit}[\pi_j(x+1)] = \beta_{j0}(x) + \beta_{j1}(x) Y_i(x)$$

$$p_{ij}(x) = \frac{\exp\left[\beta_{j0}(x) + \beta_{j1}(x) Y_i(x)\right]}{\sum_{r=1}^{I} \exp[\beta_{j0}(x) + \beta_{j1}(x) Y_r(x)]}$$

with $\beta_{jo}(x)$ = logit of residing in j at x+1 for reference category (not residing in i at x) and $\beta_{j0}(x) + \beta_{j1}(x)$ = logit of residing in j at x+1 for resident of i at x.

# Model 2: Transition probabilities with covariates

Illustration 1 - Micro-data

- Covariate: region of birth

$$p_{ij}(x) = \frac{\exp\left[\eta_{ij}(x)\right]}{\sum_{r=1}^{I} \exp[\eta_{ij}(x)]}$$

with $$\eta_{ij}(x) = \beta_{ij0}(x) + \beta_{ij1}(x) Z_1 + \beta_{ij2}(x) Z_2 + \beta_{ij3}(x) Z_3 + \ldots$$

e.g. $Z_k = 1$ if k is region of birth ($k \neq i$); 0 otherwise.

$\beta_{ij0}$ (x) is logit of residing in j at x+1 for someone who resides in i at x and was born in i.

multinomial logistic regression model

# Model 2: Transition probabilities with covariates

Illustration 1 - Macro-data
- Covariate: underfive (or infant) migration probability

$$p_{ij}(x) = \frac{\exp[\eta_{ij}(x)]}{\sum_{r=1}^{I} \exp[\eta_{ij}(x)]}$$

with $$\eta_{ij}(x) = \beta_0(x) + \beta_1(x)\, p_{ij}(-5)$$

Rogers, Muhidin, Jordan, Lea (2004, p. 8):

linear model with regression coefficients independent of x

$$p_{ij}(x) = a + b\, p_{ij}(-5)$$

# Transition rates

$$\mu_{ij}(x) = \lim_{(y-x)\to 0} \frac{p_{ij}(x,y)}{y-x} \qquad \text{for i} \neq \text{j}$$

$\mu_{ii}(x)$ is defined such that $\sum_j \mu_{ij}(x) = 0$

Hence $\mu_{ii}(x) = \sum_{j \neq i} \mu_{ij}(x) = \lim_{(y-x)\to 0} \frac{1 - p_{ij}(x)}{y-x}$

*Force of retention*

# Transition rates: matrix of intensities

$$\boldsymbol{\mu}(\mathrm{x}) = \begin{bmatrix} \mu_{11}(\mathrm{x}) & -\mu_{21}(\mathrm{x}) & . & . & -\mu_{\mathrm{I}1}(\mathrm{x}) \\ -\mu_{12}(\mathrm{x}) & \mu_{22}(\mathrm{x}) & . & . & -\mu_{\mathrm{I}2}(\mathrm{x}) \\ . & . & . & . & . \\ . & . & . & . & . \\ -\mu_{1\mathrm{I}}(\mathrm{x}) & -\mu_{2\mathrm{I}}(\mathrm{x}) & . & . & \mu_{\mathrm{II}}(\mathrm{x}) \end{bmatrix}$$

Discrete-time transition probabilities:

$$\mathbf{P}(\mathrm{x},\mathrm{y}) = \begin{bmatrix} p_{11}(x,y) & p_{21}(x,y) & . & . & p_{N1}(x,y) \\ p_{12}(x,y) & p_{22}(x,y) & . & . & p_{N2}(x,y) \\ . & . & . & . & . \\ . & . & . & . & . \\ p_{1N}(x,y) & p_{2N}(x,y) & . & . & p_{NN}(x,y) \end{bmatrix}$$

$$\frac{d\mathbf{P}(x)}{dx} = -\boldsymbol{\mu}(x)\mathbf{P}(x)$$

# Transition rates: piecewise constant transition intensities (rates)

$$\mathbf{P}(x,y) = \exp\left[-(y-x)\mathbf{M}(x,y)\right]$$

$$\exp(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \ldots$$

$$\exp[-(y-x)\mathbf{M}(x,y) = \mathbf{I} - (\mathrm{y}-\mathrm{x})\mathbf{M}(x,y) + \frac{(\mathrm{y}-\mathrm{x})^2}{2!}\left[\mathbf{M}(x,y)\right]^2 - \frac{(\mathrm{y}-\mathrm{x})^3}{3!}\left[\mathbf{M}(x,y)\right]^3 + \ldots$$

$$\mathbf{P}(\mathrm{x},\mathrm{y}) = \left[\mathbf{I} + \tfrac{1}{2}\mathbf{M}(\mathrm{x},\mathrm{y})\right]^{-1}\left[\mathbf{I} - \tfrac{1}{2}\mathbf{M}(\mathrm{x},\mathrm{y})\right]$$

# Transition rates: generation and distribution

$$\mu_{ij}(x) = \mu_{i+}(x)\,\xi_{ij}(x)$$

where $\xi_{ij}(x)$ is the probability that an individual who leaves i selects j as the destination. It is the conditional probability of a *direct transition* from i to j.

Competing risk model

$$\begin{bmatrix} \mu_{11}(x) & -\mu_{21}(x) & . & . & -\mu_{I1}(x) \\ -\mu_{12}(x) & \mu_{22}(x) & . & . & -\mu_{I2}(x) \\ . & . & . & . & . \\ . & . & . & . & . \\ -\mu_{1I}(x) & -\mu_{2I}(x) & . & . & \mu_{II}(x) \end{bmatrix} = \begin{bmatrix} \xi_{11}(x) & -\xi_{21}(x) & . & . & -\xi_{I1}(x) \\ -\xi_{12}(x) & \xi_{22}(x) & . & . & -\xi_{I2}(x) \\ . & . & . & . & . \\ . & . & . & . & . \\ -\xi_{1I}(x) & -\xi_{2I}(x) & . & . & \xi_{II}(x) \end{bmatrix} \begin{bmatrix} \mu_{1+}(x) & 0 & . & . & 0 \\ 0 & \mu_{2+}(x) & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \mu_{I+}(x) \end{bmatrix}$$

# Transition rates: generation and distribution with covariates

Log-linear model
$$m_i = \exp\left[\beta_{i0} + \beta_{i1} Z_1 + \beta_{i2} Z_2 + \ldots\right]$$

$$\ln m_i = \beta_{i0} + \beta_{i1} Z_1 + \beta_{i2} Z_2 + \ldots$$

Cox model
$$m_i(x) = m_{io}(x) \exp\left[\beta_{i0} + \beta_{i1} Z_1 + \beta_{i2} Z_2 + \ldots\right]$$

# From transition probabilities to transition rates
# The inverse method (Singer and Spilerman)

$$\mathbf{P}(\mathrm{x,y}) = \left[\mathbf{I} + \tfrac{1}{2}\mathbf{M}(\mathrm{x,y})\right]^{-1}\left[\mathbf{I} - \tfrac{1}{2}\mathbf{M}(\mathrm{x,y})\right]$$

$$\mathbf{M}(x,y) = \tfrac{y-x}{2}\left[\mathbf{I} - \mathbf{P}(x,y)\right]\left[\mathbf{I} + \mathbf{P}(x,y)\right]^{-1}$$

From 5-year probability to 1-year probability:

$$\mathbf{P}(x,x+1) = \exp\left[-\mathbf{M}(x,x+1)\right]$$

# Count data

Poisson model: $$\Pr\{N_\mathrm{i} = \mathrm{n_i}\} = \frac{\lambda_i^{n_i}}{n_i!}\exp[-\lambda_\mathrm{i}]$$

Covariates: $$E[N_i] = \lambda_i = \exp[\beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \ldots]$$

$$\ln\lambda_i = \beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \ldots$$

The log-rate model is a log-linear model with an offset:

$$E\left[\frac{N_i}{PY_i}\right] = \frac{\lambda_i}{PY_i} = \exp[\beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \ldots]$$

$$E[N_i] = \lambda_i = PY_i\exp[\beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \ldots]$$

# Incomplete data

Poisson model: $$\Pr\{N_{ij} = n_{ij}\} = \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!}\exp[-\lambda_{ij}]$$

Data availability: $$E\lfloor N_{ij} \rfloor = \lambda_{ij} = \alpha_i \, \beta_j$$

The ***maximization*** of the probability is equivalent to maximizing the log-likelihood $l = \sum_{ij}\left[n_{ij}\ln[\alpha_i\beta_j] - \alpha_i\beta_j\right]$

$$\hat{\alpha}_i = \frac{n_{i+}}{\sum_j \hat{\beta}_j} \qquad \hat{\beta}_j = \frac{n_{+j}}{\sum_i \hat{\alpha}_i}$$

The EM algorithm results in the well-known expression

$$\lambda_{ij} = \frac{n_{i+}}{n_{++}} n_{+j}$$

# Conclusion

- Unified perspective on modeling of migration: probability models of counts, probabilities (proportions) or rates (*risk indicators*)
- State occupancies and state transitions
  - Transition rate = exit rate * destination probabilities

Timing of event — Direction of change