6

Models of migration: observations and judgements

Frans Willekens
Netherlands Interdisciplinary Demographic Institute, The Hague

6.1 Introduction

The monitoring of international migration in Europe calls for an adequate representation of migration processes and adequate data on the processes.[1] Models are abstract representations of some portion of the real world. They are important for our understanding of the world. When the founding father of econometrics, Jan Tinbergen, received the Nobel Prize in 1969, he delivered the Nobel Prize Lecture. In it, he asserted that models force us to present an internally consistent theory and to confront that theory with reality (Tinbergen 1981:17). Around the same time, Britton Harris, a founding father of quantitative urban planning research, viewed properly constructed models as theories (Harris 1983). More recently, Tom Burch has expressed the view that models are the central element of scientific knowledge (Burch 1999, 2003). In his view, a model represents theory when it starts from empirical observations and arrives at an abstract and therefore general description of the real world (Burch 2003:269). Burch argues that models do not have to be true to be useful. Sufficient realism to the purpose at handsupports understanding,

[1] For a discussion, see e.g. Willekens (1994) and Chapters 2 to 5 of this book.

International Migration in Europe: Data, Models and Estimates Edited by J. Raymer and F. Willekens

explanation, prediction and policy guidance. The models that are reviewed in this chapter are abstract but they acknowledge that some people are more likely to migrate than other people and that the likelihood of migration and the directions of migration are determined by characteristics of potential migrants, by contextual factors at the place of origin (e.g. push factors) and at the place of destination (e.g. pull factors) and intervening factors. The models also acknowledge the fact that the occurrence and timing of migration can never be predicted with certainty, which points to the role of chance. Finally, the models provide a vehicle for integrating systematic factors and random factors in the prediction of migration.

The aim of this chapter is to provide a perspective on the modelling of migration that distinguishes between systematic and random factors and that accommodates several migration measures and several data types, including statistical and judgemental data. Migration is a renewable or repeatable event. Renewable events are characterised by quantum and tempo. The *quantum* or level of migration is measured by the number of migrants during a period of time, the share of migrants in a population, or the rate of migration. These measures are related in some way. The *tempo* refers to the timing of migration, more particularly the ages at which migration occurs and the interval between successive migrations. The migration models are rooted in survival models that have been developed for time-to-event data, where the event is a transition from one state of existence to another. Migration models are transition models. The dependent variable is a count, a rate or a probability.

This chapter consists of six sections. Section 6.2 is a non-technical discussion of issues in migration modelling. The issues include the distinction between migrations and migrants, the spatial pattern of migration and the age structure of migrants. The section also introduces the use of regression models for predicting migration flows from incomplete data. Section 6.3 reviews the essentials of probability theory. Section 6.4 is the main part of the chapter. It is a review of migration models from the perspective of probability theory. Different models are associated with different data types. Harmonisation of migration data and comparison of migration levels in space and time require that one data type can be converted into another type. Probability models serve that function. Some models predict the number of migrants during a given interval, others predict the probability that a randomly observed individual is a migrant, while still other models predict the rate of migration. The parameters of the probability models are generally estimated from observations on migration. Empirical evidence on migration is often lacking, incomplete or of questionable quality. Parameters estimated from inadequate data may be misleading. To obtain accurate parameter estimates, hard evidence may be augmented by soft evidence such as expert opinions and judgements. Section 6.5 discusses migration modelling in the absence of adequate data. The method, which is widely used in migration analysis, is presented as a special case of the EM (expectation–maximisation) algorithm. The EM algorithm is the most widespread statistical method for model estimation when data are incomplete. The chapter demonstrates that the modelling of migration can benefit from recent developments in modelling of life events and life histories. Section 6.6 concludes the chapter.

6.2 Data types and data structure

Migration is an event and the person who migrates is a migrant. The event must be properly defined. Any migration involves relocation but not all relocations are migrations. Migration is generally defined as a change of usual residence (address) beyond administrative boundaries. Relocations that do not involve a change of address do not qualify as migrations. They may be travel, commuting or temporary change of residence. These spatial movements may need to be considered because they may lead to a migration. The administrative unit considered in the definition of migration may be a village, a town, a district or a country. The general definition of the event does not change with the spatial unit. The definition of migration often involves a temporal dimension – an intended duration of stay or an actual duration of residence. Sometimes relocation beyond an administrative boundary involves a change of address but does not meet the duration criterion and is therefore not viewed as a migration. The migration models reviewed in this chapter focus on the spatial dimension of migration and disregard the time dimension. Any relocation across administrative boundaries involving a change of address is a migration irrespective of the intended or actual duration of stay at the destination. If the address is changed for a very small duration, the event is seen as a migration. Migration is said to be an event in continuous time. The person relocating beyond administrative boundaries is a migrant.

A major problem in migration studies is the confusion that exists between concept and measurement. Although the concept of migration is unambiguous, for practical reasons migration is often defined in terms of its measurement. For instance, relocation (a change of address) is sometimes measured by comparing the addresses at two points in time, a fixed or variable number of years apart (one year, five years, lifetime). If the addresses are in different administrative units, a migration is said to have occurred although the event of migration has not been recorded. Several authors therefore make a distinction between migrant data that result from a comparison of addresses at two points in time and migration data. Different lengths of the time interval have occupied researchers for years (see e.g. Long and Boertlein 1981; Kitsul and Philipov 1981; Courgeau 1982; Rogers *et al.* 2003b). As early as 1973, Courgeau (1973) distinguished between migration and migrant. A migration is an event and a migrant is a person.

Events occur in continuous time, i.e. an event may occur at any time. Although an event occurs in continuous time it is generally not feasible to record the exact time or date of the event. The month or year of occurrence is recorded instead, i.e. the date of the event is recorded in discrete time. The measurement approach that focuses on events and records the timing of events in continuous or discrete time is referred to as the event-based approach (Willekens 2001). A different approach is to measure events indirectly by comparing the place of residences at two points in time. For instance, in the United States census, migrations are measured by comparing the place of residence at census night and the place of residence five years prior to the census. This approach is the status-based approach. The status-based measurement of migration distinguishes time intervals and therefore takes place in discrete time. The distinction between continuous time and discrete time and the distinction between the event-based approach and the status-based approach are essential. In the literature on migration, the measurement of events in continuous or discrete time has been referred to as the movement approach and the measurement of a migration by comparing places of residence at two points in time as the transition approach (Ledent 1980; Rees and Willekens 1986). Following Courgeau (1973), the event-based type of data is often referred to as migration data and the status-based type of data as migrant data. Other authors distinguish between direct transitions (event-based) and discrete-time transitions (status-based).

The different ways of measuring migration lead to different data types. The distinction between the event-based approach and the status-based approach has been accepted in the literature as a basis for a typology of data types associated with the measurement of migration (see e.g. Rees and Willekens 1986; Bell *et al.* 2002:437; Duke-Williams and Blake 1999; Kupiszewski and Kupiszewska 2003). The authors also discuss the sources of the differences. Typically, event data are associated with population registers and (discrete-time) transition data with population censuses.

Events have characteristics and persons have characteristics and these should be kept separate. Characteristics of a migration include the origin and destination of migration, and the reason for migration. Examples include rural–urban migration, international migration by country of origin and country of destination, marriage migration, family reunion, job-related migration and forced migration. Note that the destination is the current place of residence, which is also a characteristic of the person who migrates, i.e. the migrant. Age, sex, level of education, marital status, employment status, country of birth and country of residence at a given time are characteristics of migrants. In the present chapter and in life history analysis in general, age is treated differently from the other characteristics. Age is a duration variable, which measures the time elapsed since a reference event or event origin. Any event can be selected as the event origin (e.g. birth, marriage, last migration). If the reference event is birth, the duration variable is age. If, on the other hand, the reference event is the last migration, the duration variable is the duration of current residence. The personal characteristics except age are referred to as covariates.

The modelling framework presented in this chapter is rooted in the multivariate analysis of time-to-event data, also known as multistate survival analysis. In multistate analyses a personal attribute such as the place of residence is denoted as a state and the variable that identifies the state is the state variable. The collection of all possible states is the state space. In mathematics and engineering, multistate models are known as state-space models. A change of attribute, i.e. a change of state, is a transition. In this chapter two key concepts are distinguished. The first is *state occupancy*. It is the state occupied at a given point in time (e.g. at a given exact age). The second is the *state transition*. It refers to a change in state occupied. Transitions may be expressed in continuous time or discrete time (Andersen *et al.* 1993:93). Transitions in continuous time are referred to as *direct transitions* or events (Rajulton 1999:5). *Discrete-time transitions* are identified by comparing states occupied at two consecutive points in time (t), e.g. the region of residence at t_1 and the region of residence at t_2.

Characteristics of migrations and migrants are specified at the individual level. An individual occupies an address and has a set of attributes. At the population level, the distribution of addresses and attributes is the subject of study. With a distribution is associated a *data structure*: age structure, covariate structure, motivational structure and spatial structure. The covariate structure relates to the attributes of the migrants, e.g. country or region of birth and sex, but also employment status and marital status. The motivational structure relates to the reasons for migration. The spatial structure relates to the origins and destinations of migrations. In the study of migration, the spatial structure is of particular relevance. The origins and destinations of migration flows define a spatial pattern. The spatial distribution of potential destination locations, their attributes and the interlinkages that connect them shape migration flows in ways which accord them spatial structure that is likely to affect the directions of subsequent migrations (Rogers *et al.* 2001). When residents of a given area are more likely to move to a particular destination rather than to another destination, a spatial dependence exists. Spatial dependences generate spatial structure. The modelling framework proposed in this chapter captures different spatial dependences and spatial structures. Based on a review of the recent literature on spatial structure, Bell *et al.* (2002:436) argued that four broad dimensions of spatial structure could be recognised, relating to (6.1) migration intensities, (6.2) migration distance, (6.3) migration connectivity and (6.4) the effect of migration on the redistribution of the population. Recently, Rogers *et al.* (2001, 2002a,b, 2003a) proposed statistical models to capture the spatial structure of migration flows or a migration system. The approach is adopted in this chapter. The proposed framework encompasses different types of spatial dependence. Consider the following examples. Spatial dependence is absent if the destination is not affected by the origin of the migrant. The spatial focus, which is the concentration of migration in a few flows, is another example. The spatial focus is measured using a variety of indices to capture the extent to which migration flows between regions are concentrated or dispersed (Plane and Mulligan 1997; Rogers and Sweeney 1998; Rogers and Raymer 1998). The study of spatial focusing also includes the spatial dominance exerted in varying degrees by destinations on origins (Pooler 1992).

Each structure calls for a different modelling approach that may be logically integrated in the comprehensive framework. The age structure is modelled by models of age (duration) dependence. The Rogers–Castro model migration schedule is an example (see Chapter 8). The covariate structure is described by transition rate models and logistic regression models. The spatial structure is captured by spatial interaction models. Conventional spatial interaction models capture the effects of distance on the level and direction of migration flows (see e.g. Mueser 1989; Sen and Smith 1995). Most spatial interaction models today include other variables. Underlying the various spatial interaction models of migration has been a recognition that the decision to move both shapes and is shaped by the population geography within which the movement takes place. The spatial structure changes over time and, although the spatial structure is often remarkably stable even in periods of socio-economic change, models of spatial structure should be able to capture continuity and change. In the unlikely case that the four structures are independent, the structures can be modelled separately. Different dependences or interaction effects are identified and integrated into the framework. For instance, the age structure of migration may differ by origin and/or destination and/or covariate and/or reason for migration.

A unified perspective on the modelling of migration flows has four significant advantages. First, it provides a single, comprehensive framework for the *analysis of data* on migration. Data may be of different types. Second, it provides a framework for the *harmonisation of migration statistics*. Migration data are obtained in many different ways. The estimation of comparable indicators of levels and trends of migration (and direction of migration) and the comparative analysis of patterns of migration require comparable data or techniques for converting one data type into another, e.g. migrant data into migration data.

Third, it provides a framework for the*prediction (estimation) of missing data on migration*. The prediction of a missing value or a set of missing values on the basis of available data is similar to the imputation of missing values. Imputation is receiving much attention in the literature. Methods for statistical data imputation may be divided into two broad groups: model imputation and donor imputation. In model imputation, the imputed values are directly derived from a data model, i.e. a statistical or demographic model of the data. Common data models take the format of regression models. The regression model is estimated from the available data that may be augmented with qualitative information on migration (expert opinion, judgemental data). In donor imputation, the imputed values are derived from a set of observed values (donors). The imputed value is based on information in the closest valid record (nearest neighbour matching characteristics that are not missing). Post-imputation edits (post-editing) make sure that the nearest neighbour is close enough to be used as a donor. All available auxiliary information is used to assure a best estimate or imputation. The estimation of missing migration data may benefit from the literature on imputation. Donor imputation is not considered in this chapter.

A fourth advantage of a unified perspective is that seemingly different migration models may be grouped into classes of models. For instance, the gravity model, the entropy model and the log-linear model of interregional migration have a common structure that links them to the family of generalised linear models (Willekens 1983). The migration pool model, which is much used in regional population projections, is a special case of the origin–destination migration flow model. The migrant pool model[2] is a migration flow model with the origin–destination interaction removed.

[2] A survey of European official subnational population projection practices in the 1990s by van Imhoff *et al.* (1994) revealed that one variant in common use is the migration pool model. In the migration pool model, migration is projected in two stages. The first stage is the projection of the number of outmigrations from each region. The migrants are placed in a common pool. In the second stage, the migrations in the pool are distributed over the possible destinations. For a recent illustration of the two-stage modelling of migration, see van Wissen *et al.* (2005) and Stillwell (2005).

Several authors have assessed the impact of choice of migration model in the context of regional population projections (Kupiszewski and Kupiszewska 2003; Wilson and Bell 2004; Stillwell 2005; van Wissen *et al.* 2005).

The framework for the modelling of migration is still incomplete. Four limitations are singled out. First, it does not include a measure of the reliability of the estimates or the degree of confidence that one may attach to the estimates. Second, it does not cover the indirect estimation of migration from data on *populations* at two points in time and natural increase during the period. Third, it is of no use to estimate undocumented migration if undocumented migrants are not included in the aggregate data or if qualitative information (e.g. expert opinion, educated guess) is lacking. Fourth, the framework does not yet include a unified approach to model temporal changes in migration flows. In the 1980s, Plane and Rogerson (1986) and Jackson *et al.* (1990) proposed causative matrices to link matrices of migration rates from one time period to another and to extrapolate migration rates using a geometric regression based on two data points. The causative method was recently discussed by De Mesnard (2004) in the context of structural changes in input–output analysis. The method is not considered in this chapter.

6.3 Probability models: generalities

Observations on migration are manifestations of a migration process. People migrate at different times in different directions for different reasons. The migration process that is revealed to the observer depends on the observation, i.e. the measurement of the migration process. The true migration process is often referred to as the underlying process. *Data models* describe the observations while *process models* describe the underlying process. Our aim is to model processes rather than observations on the processes.

Migration is the result of several factors. They include personal factors and contextual or situational factors. The latter include push, pull factors and intervening factors. Migration is however also an outcome of chance, i.e. random factors. As a consequence, an observation on migration is in part a result of chance. To separate the effects of systematic factors from the effects of random factors, the chance process is made explicit, which is done by specifying a probability model rather than a deterministic model. A probability model is a mathematical representation of a random phenomenon (see e.g. Taylor and Karlin 1998). The variable that describes an observation or a set of observations is a *random variable* and the distribution of the random variable reflects the random distribution of the observations. Many authors do not use random variables but work in terms of variates. A variate is a generalisation of the idea of a random variable. It has similar probabilistic properties but is defined without reference to a particular probabilistic experiment (Evans *et al.* 2000:4). A *variate* is the set of all random variables that obey a given probabilistic law. A multivariate is a vector or set of elements, each of which is a variate. Different types of observations (data types) are represented by different variates. Counts are represented by Poisson variates. Like counts, they can take on only nonnegative integer values. The Poisson model is the probability model for counts. A response variable that has only two categories (yes, no; success, failure) is a binary response variable. A single observation may be viewed as the outcome of a Bernoulli trial, and a sequence of observations as a sequence of Bernoulli trials. The number of successes in a number of trials is a binomial variate. The number of failures before the first success is a geometric variate, and the number of failures before the *k*th success is a negative binomial variate. A variate may take on several values (a finite number in the case of a discrete variate and an infinite number in the case of a continuous variate). In this chapter we consider discrete variates. With each value of the discrete variable may be associated a probability, and the distribution of the values is a probability distribution. The distribution of observations (empirical distribution) is approximated by a theoretical distribution, such as the Poisson distribution and the binomial distribution.

The probability distributions describe the random phenomenon. Theoretical distributions are fully characterised by one parameter or a few parameters. For instance, the normal distribution has two parameters (mean and variance). The Poisson distribution has a single parameter (mean and variance are equal). The values of the parameters of the distributions depend on the systematic factors such as personal characteristics and situational variables. In order to predict the parameter value from the factors (predictors), the parameter may need to be transformed to assure that the predicted value is within the range of acceptable values. The Poisson distribution has a single parameter, which must be nonnegative. To assure that the value is nonnegative irrespective of the values of the systematic factors or predictors, the log (logarithmic) transformation is used. The binomial distribution has a single parameter too – the probability of success – generally denoted by p for probability. Its value must be between 0 and 1, and to assure that the requirement is met, a logit transformation is often used to relate the parameter to the systematic factors. The logit transformation imposes a logistic distribution onto p. That means that, as the predictors vary from $-\infty$ to $+\infty$, the parameter p varies between 0 and 1 following a cumulative logistic distribution. Alternative transformations include the probit (cumulative normal distribution) and the log-log transformation (cumulative extreme value distribution).

Now we consider the Poisson model and the logit model in more detail. The Poisson variate describes the number of migrations during an interval of unit length (e.g. year, month). The number of migrations is not restricted in any way. Subjects in a (sample) population may experience an event more than once during the unit interval. The probability of observing n migrations during the interval is given by the Poisson model

$$\Pr\{N = n\} = \frac{\lambda^n}{n!}\exp(-\lambda), \tag{6.1}$$

where N is the variate denoting the number of migrations, n is the observed number of migrations during a unit interval, and λ is the expected number of migrations during the unit interval. The latter is the parameter of the Poisson model. Note that λ is not the same as the migration rate. The migration rate is the expected number of migrations during the unit interval per person, i.e. the number of migrations per person-year or person-month. It is obtained by dividing the number of migrations during the interval by the total time spent during the interval by all subjects combined. The migration rate will be denoted byμ.

The parameter of the Poisson model may be made dependent on covariates:

$$E[N] = \lambda = \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots). \tag{6.2}$$

The model may be written as a log-linear model:

$$\ln \lambda = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots. \tag{6.3}$$

In principle, Z_p can be any covariate. In conventional log-linear analysis, all covariates are discrete or categorical. The observations on event occurrences may therefore be arranged in a contingency table. The covariates refer to rows, columns, layers and combinations of these (to represent interaction effects). Log-linear models of age and spatial structures of migration flows are studied by Rogers *et al.* (2003a) among others.

The migration rate associated with the parameter of the Poisson model is the expected number of migrations divided by the person-years lived during the interval (assuming that the unit of time is a year). Let the person-years be denoted by PY. The migration rate is $\mu = \lambda/PY$. The person-years are assumed to be known and independent of the number of migrations. Hence, PY is not a variate.

The Poisson variate describes the number of migrations during an interval. It is applicable when event data are available (event-based approach). Migration is often measured by comparing the places of residence at two consecutive points in time. In that status-based approach, a migrant is an individual whose residence at time t is different from the residence at $t-1$. Multiple migrations during the interval are not recorded and the number of migrants can never be larger than the number of subjects in the (sample) population. If an individual whose residence at time t differs from that at $t-1$ is coded 1 and an individual whose residence is the same (although he may have changed residence several times during the interval) is coded 0, then the number of migrants in a (sample) population is a binomial variate. The probability of observing n migrants among a (sample) population of m individuals is the binomial distribution with parameter p and index m:

$$\Pr\{N_m = n\} = \frac{m!}{n!(m-n)!} p^n (1-p)^{m-n}. \tag{6.4}$$

The parameter p is the probability of being a migrant. The logit model relates p to predictors while assuring that the predicted value of p is between 0 and 1. The logit model is

$$\mathrm{logit}(p) = ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots, \tag{6.5}$$

where Z_p can be any covariate and $p/(1-p)$is the odds of being a migrant.

In many applications, m is the total number of migrations out of a given region (of origin) and n is the number of migrations that end in a given destination. In that case, $m - n$ is the number of outmigrations that end in another destination. The expected number of migrations during an interval that end in a given destination rather than in any of the other destinations is λp. The probability of observing n migrations to the given destination during a unit interval is given by the Poisson distribution with parameter λp with $\lambda = \mu PY$. If the destination choice is independent of the decision to outmigrate, then μ (and λ) and p can be studied separately (see e.g. Hachen 1988). If the decision to outmigrate is not independent of the destination μ and p must be studied simultaneously.

If the possible number of destinations exceeds two, then the distribution of the outmigrations over the destinations say three is given by the multinomial distribution

$$\Pr\{N_1 = n_1, N_2 = n_2, N_3 = n_3\} = \frac{m!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}, \tag{6.6}$$

which may be written in general as

$$\Pr\{N_1 = n_1, N_2 = n_2, ...\} = \frac{m!}{\prod_{i=1}^{I} n_i!} \prod_{i=1}^{I} p_i^{n_i}, \tag{6.7}$$

where I is the number of possible destinations, n_i is the number of individuals ending in i and with $\sum p_i = 1$and $\sum N_i = \sum n_i = m$.

6.4 Probability models of migration

This section reviews models of state occupancies and state transitions. State occupancy is expressed in terms of the probability that an individual selected at random from a (sample) population occupies a given state. It is approximated by the proportion of the (sample) population in a given state. State transitions during an interval (in continuous time or discrete time) are expressed in terms of three risk indicators: counts, probabilities (proportions) and rates. Probabilities relate transitions to the *risk set* (population at risk) at the beginning of an interval. The risk set accounts for attrition during the interval for reasons unrelated to the transition being studied (censoring), which is the migration from the current place of residence to a given destination. Rates relate transitions to *exposure*. Probabilities are obtained as the ratio of the number of events (in continuous time or discrete time) to the risk set; rates are obtained as the ratio of the number of events (in continuous time) to the exposure time. Counts refer to the number of events that occur during a given interval. It is the numerator in probability and rate measures.

In developing probability models of migration and migrants, the occurrence of an event (migration) is assumed to be the result of an underlying random mechanism. The occurrence of a migration depends on both personal attributes (systematic factors) and chance. Our approach is to model the random mechanism by specifying a probability model. A problem is that a migration is not necessarily associated with a single random mechanism. Different mechanisms may result in the same event of migration. Hence different probability models may describe the level and direction of migration. If an event is observed, it should be possible to identify the set of plausible mechanisms and to identify the mechanism that *most likely* produced the event. To determine the most likely mechanisms and the model that describes that mechanism, the maximum likelihood method is applied. The method identifies the 'best model', i.e. the model that has the greatest probability of predicting the observations on event occurrence. That model describes the random mechanism that most likely underlies the event.

6.4.1 State probabilities

Let S denote the state space: $S = \{1, 2, 3, \ldots, i, \ldots, I\}$. The state space contains I possible geographical areas or regions. At a given age, an individual resides in one area and one area only. In other words, the states are mutually exclusive. Let $Y_k(x)$ be a polytomous random variable denoting the state occupied by individual k at exact age x. The probability that an individual resides in state i is the *state probability*. The probability that individual k resides in state i at exact age x is ${}_k\pi_i(x) = Pr\{Y_k(x) = i\}$. If all individuals are independent and identical, then ${}_k\pi_i(x) = \pi_i(x)$ for all k. If individuals differ in a few characteristics only or if a few characteristics suffice to predict the state occupied at age x, then ${}_k\pi_i(x) = \pi_i(x, Z)$, where Z represents a specific combination of characteristics or covariates. The probability that individual k occupies state i at exact age x depends on the covariates only, and individuals with the same covariates have the same state probability.

The state occupied at x may be denoted differently. Let $Y_{ki}(x)$ be an indicator variable (binary) which is 1 if individual k occupies state i at x and 0 otherwise. The state probability is the probability that the random variable takes on the value of 1.

Consider a sample of m individuals. We do not consider covariates, implying that all individuals are identical. Covariates are introduced below. In addition, age is omitted for convenience. The number of individuals observed in state i is

$$N_i = \sum_{k=1}^{m} Y_{ki}. \tag{6.8}$$

The probability of observing N_1 individuals in state 1, N_2 in state 2, N_3 in state 3, etc., is given by the multinomial distribution

$$\Pr\{N_1 = n_1, N_2 = n_2, \ldots\} = \frac{m!}{\prod_{i=1}^{I} n_i!} \prod_{i=1}^{I} \pi_i^{n_i}, \tag{6.9}$$

where n_i is the observed number of individuals in i and with $\sum \pi_i = 1$ and $\sum N_i = \sum n_i = m$. The most likely values of the parameters π_i, given the data, are obtained by maximising the likelihood that the model predicts the data, which is the maximum likelihood method. The values of π_i $(i = 1, 2, \ldots, I)$ that maximize the above multinomial distribution is $\hat{\pi}_i = n_i/m$.

The expected number of individuals occupying state i is $E[N_i] = \pi_i m$ and the variance is $\mathrm{Var}[N_i] = \pi_i(1 - \pi_i)m$. The probability that an individual is found in state i is the expected value of $Y_i : \pi_i = E[Y_i]$. The variance of Y_i is

$$\mathrm{Var}[Y_i] = \mathrm{Var}[N_i/m] = \mathrm{Var}[N_i]/m^2 = [\pi_i(1 - \pi_i)]/m.$$

The variance declines with increasing sample size.

Now we introduce covariates. They are denoted by $Z\,(Z = \{Z_1, Z_2, Z_3, \ldots\})$, and Z_p may represent a single attribute or a combination of attributes (to denote interaction effects). The state probability $\pi_i(Z)$ that an individual with covariates Z occupies state i is given by the logit equation

$$\mathrm{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \eta_i = \beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \beta_{i3}Z_3 + \cdots, \qquad (6.10)$$

where $\pi_i/(1 - \pi_i)$ is the odds of occupying state i. The logit transformation assures that the state probabilities lie between 0 and 1, and that their sum is equal to unity. The value of η may range from $-\infty$ to $+\infty$, but the value of π_i stays within 0 and 1. To obtain the probabilities, the logit scale is converted into the probability scale

$$\pi_i = \frac{\exp(\eta_i)}{\exp(\eta_1) + \exp(\eta_2) + \cdots + 1 + \cdots} = \frac{\exp(\eta_i)}{\sum_{j=1}^{I} \exp(\eta_j)}, \qquad (6.11)$$

where the 1 is associated with the reference category. The model is the multinomial logistic regression model.

6.4.2 Transition probabilities

The state occupied at a given age generally depends on the states occupied at previous ages, in addition to personal attributes at the given age. Hence the probability of being in state j at x+1 (or more generally y) depends on the states occupied at previous ages. It is often assumed that only the most recent state occupancy is relevant:

$$\Pr\{Y(x+1) = j \mid Y(x), Y(x-1), \cdots; Z\} = \Pr\{Y(x+1) = j \mid Y(x); Z\}. \qquad (6.12)$$

If the state occupied at x is i, then

$$\Pr\{Y(x+1) = j \mid Y(x) = i\} = p_{ij}(x). \qquad (6.13)$$

Here $p_{ij}(x)$ is the probability that an individual who resides in state i at x resides in state j at x+1. It is the discrete-time transition probability. The interval can be of any length but is generally one or five years. This model is suited for describing migrant data, i.e. data that infer migration by recording the places of residence at two consecutive points in time.

The status dependence may also be written as

$$\text{logit}[\pi_j(x+1)] = \beta_{j0}(x) + \beta_{j1}(x)Y_i(x), \tag{6.14}$$

where $Y_i = 1$ if state i is occupied at x and 0 otherwise. Hence the transition probability may be written as

$$p_{ij}(x) = \frac{\exp[\beta_{j0}(x) + \beta_{j1}(x)Y_i(x)]}{\sum_{r=1}^{I}\exp[\beta_{j0}(x) + \beta_{j1}(x)Y_r(x)]}. \tag{6.15}$$

The transition probabilities may depend on covariates in a way that is similar to that of state probabilities. Transition probabilities out of a given state i that depend on covariates may be estimated using multinomial logistic regression software.

6.4.3 Transition rates

The transition probabilities discussed in Section 6.4.2 are defined for discrete time intervals. They depend on the number of persons at risk at the beginning of the interval, which is generally known as the risk set. The probabilities are not directly related to the *duration* that individuals in i are at risk of migrating to j. Since the event of migration (direct transition) may occur at any time during the interval from x to y (with $y = x+1$, for instance), the transition probability is defined for very small intervals. The probability that an individual in i transfers to j during an infinitesimally small interval following x is the instantaneous rate of transition:

$$\mu_{ij}(x) = \lim_{(y-x)\to 0}\left(\frac{p_{ij}(x, y)}{y-x}\right) \quad \text{for } i \text{ not equal to } j. \tag{6.16}$$

The instantaneous rate of transition is also known as the transition intensity and the force of transition. The term $\mu_{ii}(x)$ is defined such that $\sum_j \mu_{ij}(x) = 0$. The quantity $\mu_{ii}(x)$ is nonnegative. It is sometimes referred to as the intensity of passage because it relates to the transition from i to any other state different from i. Schoen (1988:65) refers to it as the 'force of retention'.

The intensities are the basic parameters of a continuous-time multistate process. Under the restrictive Markov assumption, the probability that an individual leaves a state depends only on the state occupied and the individual's age. It is independent of other characteristics.

The matrix of instantaneous rates with off-diagonal elements $-\mu_{ij}(x)$ and with $\mu_{ii}(x)$ on the diagonal is known as the *generator* of the stochastic process $\{Y_k(x); x \geqslant 0\}$ (Çinlar 1975:256). The matrix is denoted by $\boldsymbol{\mu}(x)$, and has the following configuration:

$$\boldsymbol{\mu}(x) = \begin{bmatrix} \mu_{11}(x) & -\mu_{21}(x) & \cdots & -\mu_{I1}(x) \\ -\mu_{12}(x) & \mu_{22}(x) & \cdots & -\mu_{I2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ -\mu_{1I}(x) & -\mu_{2I}(x) & \cdots & \mu_{II}(x) \end{bmatrix}. \tag{6.17}$$

Note that

$$\lim_{(y-x)\to 0}\left(\frac{\mathbf{P}(x,y)-\mathbf{I}}{y-x}\right)=-\boldsymbol{\mu}(x). \tag{6.18}$$

The matrix of discrete-time transition probabilities is:

$$\mathbf{P}(x,y)=\begin{bmatrix} p_{11}(x,y) & p_{21}(x,y) & \cdots & p_{N1}(x,y) \\ p_{12}(x,y) & p_{22}(x,y) & \cdots & p_{N2}(x,y) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N}(x,y) & p_{2N}(x,y) & \cdots & p_{NN}(x,y) \end{bmatrix}. \tag{6.19}$$

An element of $\mathbf{P}(x,y)$, namely $p_{ij}(x,y)$, denotes the probability that an individual who is in state i at exact age x is in state j at exact age y. The Markovian assumption implies the following relationship between $\mathbf{P}(x,\ x+v)$ and $\mathbf{P}(x+v,\ y)$:

$$\mathbf{P}(x,\ y)=\mathbf{P}(x,x+v)\cdot\mathbf{P}(x+v,\ y). \tag{6.20}$$

Subtraction of $\mathbf{P}(x+v,\ y)$ from both sides of the equation yields

$$\frac{\mathbf{P}(x,y)-\mathbf{P}(x+v,y)}{v}=\frac{[\mathbf{P}(x,x+v)-\mathbf{I}]\mathbf{P}(x+v,y)}{v} \tag{6.21}$$

and

$$\lim_{v\to 0}\frac{\mathbf{P}(x,y)-\mathbf{P}(x+v,y)}{v}=\lim_{v\to 0}\frac{[\mathbf{P}(x,x+v)-\mathbf{I}]\,\mathbf{P}(x+v,y)}{v} \tag{6.22}$$

or

$$\frac{\mathrm{d}\mathbf{P}(x)}{\mathrm{d}x}=-\boldsymbol{\mu}(x)\mathrm{P}(x). \tag{6.23}$$

Recall that

$$\lim_{(y-x)\to 0}\frac{\mathbf{P}(x,y)-\mathbf{I}}{y-x}=-\boldsymbol{\mu}(x). \tag{6.24}$$

Multiplying both sides with the vector of state probabilities at age x, $\mathbf{P}(x)$, leads to

$$\frac{\mathrm{d}\mathbf{P}(x)}{\mathrm{d}x}=-\boldsymbol{\mu}(x)\mathbf{P}(x), \tag{6.25}$$

where $\mathbf{P}(x)$ is a vector of state probabilities.

The model is a system of differential equations. In multistate demography, two avenues are followed to solve the system. Both introduce age intervals (Rogers and Willekens 1986:370ff.). The first avenue postulates a piecewise constant intensity function, $\mu(t)=\mu(x)$ in the interval from x to $y\,(x\leq t<y)$. This implies an

exponential distribution of demographic events within each age interval. The model that results is referred to as the *exponential model*. The second avenue postulates a piecewise linear survival function. A piecewise linear survival function is obtained when demographic events are uniformly distributed within the age intervals. The model that results is referred to as the *linear model*. The first avenue is followed by van Imhoff (1990) and van Imhoff and Keilman (1991) among others; the second by Willekens and Drewe (1984) among others. The state occupancies and the sojourn times must be estimated simultaneously from the population at the beginning of the interval and the events during the interval.

To solve the system of differential equations, it may be replaced by a system of integral equations:

$$\mathbf{P}(x, y) = \mathbf{I} - \int_0^{y-x} \boldsymbol{\mu}(x+t)\,\mathbf{P}(x, x+t)\,\mathrm{d}t. \tag{6.26}$$

To derive an expression involving transition rates during the interval from x to y, we write

$$\mathbf{P}(x, y) = \mathbf{I} - \left[\int_0^{y-x} \boldsymbol{\mu}(x+t)\,\mathbf{P}(x, x+t)\,\mathrm{d}t\right]\left[\int_0^{y-x} \mathbf{P}(x, x+t)\,\mathrm{d}t\right]^{-1} \times \left[\int_0^{y-x} \mathbf{P}(x, x+t)\,\mathrm{d}t\right] \tag{6.27}$$

and

$$\mathbf{P}(x, y) = \mathbf{I} - \mathbf{M}(x, y)\,\mathbf{L}(x, y), \tag{6.28}$$

where $\mathbf{M}(x, y)$ is the matrix, with elements $m_{ij}(x, y)$, of average transition rates during the interval from x to y and

$$\mathbf{L}(x, y) = \int_0^{y-x} \mathbf{P}(x, x+t)\,\mathrm{d}t$$

is the sojourn time spent in the different states between ages x and y per person in a state occupied at age x.

6.4.3.1 Exponential model

The transition intensities $\mu(x)$ are assumed to remain constant during the age interval from x to y and to be equal to the model transition rates $\mathbf{M}(x, y)$. It is furthermore assumed that they can be estimated by empirical occurrence–exposure rates for that age interval. This assumption is consistent with the general assumption in demography that life-table rates are equal to empirical rates. In this chapter no separate notation is used for model rates and empirical rates. The matrix of transition probabilities between x and y is

$$\mathbf{P}(y, x) = \exp\left[-(y-x)\mathbf{M}(y, x)\right], \tag{6.29}$$

where $\mathbf{M}(x, y)$ is the matrix of empirical occurrence–exposure rates or transition rates for the age interval from x to y and $\mu_{ij}(t) = m_{ij}(x, y)$ for $x \leq t < y$ and $\boldsymbol{\mu}(t) = \mathbf{M}(x, y)$ for $x \leq t < y$.

A number of methods exist to determine the value of $\exp[-\mathbf{M}]$ (see e.g. Director and Rohrer 1972:431ff.; Aoki 1976:387; Strang 1980:206). For example, the Taylor series expansion may be used for $\exp(\mathbf{A})$,

$$\exp(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots . \tag{6.30}$$

Hence one obtains

$$\begin{aligned} \exp[-(y-x)\mathbf{M}(x, y)] = {} & \mathbf{I} - (y-x)\mathbf{M}(x, y) + \frac{(y-x)^2}{2!}[\mathbf{M}(x, y)]^2 \\ & - \frac{(y-x)^3}{3!}[\mathbf{M}(x, y)]^3 + \cdots \end{aligned} \tag{6.31}$$

(see also Schoen 1988:72).

The transition rates $\mathbf{M}(x, y)$ are estimated from the data. The transition rate $m_{ij}(x, y)$ is equal to the number of moves (or direct transitions) from i to j during the interval from x to y, divided by the exposure in the state i:

$$m_{ij}(x, y) = \frac{n_{ij}(x, y)}{L_i(x, y)}, \tag{6.32}$$

where $n_{ij}(x, y)$ is the observed number of moves from i to j during the interval and $L_i(x, y)$ is the duration in i exposed to the risk of moving to j. It is the sojourn time in i during the (x, y) interval. Exposure is measured in person-months or person-years. In the case of two states, the rate equation may be written as

$$\begin{bmatrix} m_{11}(x, y) & -m_{21}(x, y) \\ -m_{12}(x, y) & m_{22}(x, y) \end{bmatrix} = \begin{bmatrix} n_{11}(x, y) & -n_{21}(x, y) \\ -n_{12}(x, y) & n_{22}(x, y) \end{bmatrix} = \begin{bmatrix} L_1(x, y) & 0 \\ 0 & L_2(x, y) \end{bmatrix}^{-1}, \tag{6.33}$$

where $m_{11}(x, y) = m_{12}(x, y)$ and $m_{22}(x, y) = m_{21}(x, y)$. In matrix notation:

$$\mathbf{M}(x, y) = \mathbf{n}(x, y)\,[\mathbf{L}(x, y)]^{-1}. \tag{6.34}$$

Let $\overline{\mathbf{L}}(x, y)$ be the vector of sojourn times containing the diagonal elements of $\mathbf{L}(x, y)$ and let $\mathbf{K}(x)$ be a vector with the state occupancies at age x by surviving cohort members as its elements:

$$\mathbf{K}(x) = \begin{bmatrix} K_1(x) \\ K_2(x) \end{bmatrix}, \tag{6.35}$$

with $K_i(x)$ the number of cohort members in state i at exact age x. The vector of sojourn times by all cohort members in the various states is obtained by the following equation:

$$\overline{\mathbf{L}}(x,y) = \left[\int_0^{y-x} \mathbf{P}(x, x+t)\,\mathrm{d}t\right] \mathbf{K}(x). \tag{6.36}$$

Since the transition intensities are constant in the interval from x to y, the equation may be written as follows:

$$\overline{\mathbf{L}}(x,y) = \left[\int_0^{y-x} \exp\left[-t\,\mathbf{M}(x,y)\right]\,\mathrm{d}t\right] \mathbf{K}(x). \tag{6.37}$$

Integration yields

$$-\left[\mathbf{M}(x,y)\right]^{-1} \left|\exp\left[-t\,\mathbf{M}(x,y)\right]\right|_0^{y-x}, \tag{6.38}$$

which is equal to

$$\left[\mathbf{M}(x,y)\right]^{-1} \left\{\mathbf{I} - \exp\left[-(y-x)\,\mathbf{M}(x,y)\right]\right\}. \tag{6.39}$$

Hence the sojourn times in the various states during the (x, y) interval are given by

$$\overline{\mathbf{L}}(x,y) = \left[\mathbf{M}(x,y)\right]^{-1} \left\{\mathbf{I} - \exp\left[-(y-x)\,\mathbf{M}(x,y)\right]\right\} \mathbf{K}(x). \tag{6.40}$$

6.4.3.2 Linear model

To solve $\mathbf{P}(x,y) = \mathbf{I} - \mathbf{M}(x,y)\,\mathbf{L}(x,y)$, one may introduce an approximation of $\mathbf{L}(x,y)$. A simple approximation is that $\mathbf{P}(x, x+t)$ is linear in the interval $x \le x+t < y$. Hence $\mathbf{L}(x,y)$ may be approximated by a linear integration:

$$\mathbf{L}(x,y) = \int_0^{y-x} \mathbf{P}(x, x+t)\,\mathrm{d}t \approx \frac{y-x}{2}\left[\mathbf{I} + \mathbf{P}(x,y)\right]. \tag{6.41}$$

Introducing this expression in the equation to be solved gives

$$\mathbf{P}(x,y) = \mathbf{I} - \frac{y-x}{2}\mathbf{M}(x,y)\left[\mathbf{I} + \mathbf{P}(x,y)\right], \tag{6.42}$$

$$\mathbf{P}(x,y) = \mathbf{I} - \frac{y-x}{2}\mathbf{M}(x,y)\,\mathbf{I} - \frac{y-x}{2}\mathbf{M}(x,y)\,\mathbf{P}(x,y), \tag{6.43}$$

$$\mathbf{P}(x,y) + \frac{y-x}{2}\mathbf{M}(x,y)\,\mathbf{P}(x,y) = \mathbf{I} - \frac{y-x}{2}\mathbf{M}(x,y)\,\mathbf{P}(x,y) \tag{6.44}$$

$$\mathbf{P}(x,y)=\left[\mathbf{I}+\frac{y-x}{2}\mathbf{M}(x,y)\right]^{-1}\left[\mathbf{I}-\frac{y-x}{2}\mathbf{M}(x,y)\right]. \tag{6.45}$$

The linear approximation implies the assumption that the events are uniformly distributed over the interval. The assumption is adequate when the transition rates are small or the interval is short. It can be shown that the linear model is an approximation to the exponential model that retains the first three terms of the Taylor series expansion.

The instantaneous rates of transition $\mu_{ij}(x)$ may be written as the product of two terms, a rate and a probability. The first is the instantaneous rate of leaving state of origin i irrespective of destination, and the second is the conditional probability of selecting j as the destination provided the state of origin is left (i.e. upon leaving i). The first term, the exit rate, determines the timing of the transition, while the second, the destination probability, determines the destination (new attribute). The exit rate is

$$\mu_{i+}(x)=\mu_i(x)=\sum_{j\neq i}\mu_{ij}(x).$$

Note that $\mu_i(x)=\mu_{ii}(x)$. The transition rate is

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

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. Note that the above expression is that of a competing risk model or a transition rate model with multiple destinations (Blossfeld and Rohwer 2002). In the terminology of competing risks, the first term is the rate of the event and the second term (destination) indicates the type of the event. If the occurrence of the event and the type of event are unrelated, the two terms may be estimated and studied separately (Hachen 1988:29; Sen and Smith 1995:372). The first term is studied using a transition rate (hazard rate) model; the second using a logit model or a logistic regression model.

In the migration literature, the first term $\mu_i(x)$ is known as the generation component and the second $\xi_{ij}(x)$ as the distribution component (Rogers *et al.* 2002b). It is often assumed that the factors that influence the level of outmigration differ from those that influence the choice of destination, and consequently the generation and the distribution components are modelled separately. If, in addition, the choice of destination is independent of the origin, the model that results is the migrant pool model.

The discrete-time transition probabilities are related to the transition intensities and the transition intensities are related to the probabilities of direct transition in an interesting way. The off-diagonal elements of $\mathbf{M}(x,y)$ may be replaced by $-m_{i+}(x,y)\xi_{ij}(x)$, where $m_{i+}(x,y)$ is the rate of leaving i, which is assumed to be constant in the interval from x to y. The diagonal elements are $m_{i+}(x, y)$. The $\boldsymbol{\mu}$ matrix may be written as

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

$$\times \begin{bmatrix} \mu_{1+}(x) & 0 & \cdots & 0 \\ 0 & \mu_{2+}(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{I+}(x) \end{bmatrix} \tag{6.46}$$

Now we introduce covariates. As above, they are denoted by $Z(Z = \{Z_1, Z_2, Z_3, \ldots\})$. The exit rate is modelled using a transition rate model for a single event (leaving the state of origin). The elementary transition rate model is the basic exponential model, with the rate being independent of age (Blossfeld and Rohwer 2002):

$$m_i = \exp\left(\beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \cdots\right). \tag{6.47}$$

The model may be written as a log-linear model

$$\ln m_i = \beta_{i0} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + \cdots \tag{6.48}$$

The model is also known as the log-rate model (see e.g. Yamaguchi 1991: Chapter 4).

The age dependence may be introduced in two ways: nonparametric and parametric. In the first approach, the population is stratified by age and a transition rate is estimated for each age separately. In the parametric approach, age dependence is represented by a model. A common model is the Gompertz model, which imposes onto the transition rate an exponential change with duration. The Gompertz model has two parameters and each may be made dependent on covariates – for detailed treatment, see Blossfeld and Rohwer (2002). Other parametric models of duration dependence may be used. In migration studies, the model migration schedule is a common representation of the age dependence of the migration rate. Each parameter of the model may be related to covariates. In practice, only one or a selection of parameters is assumed to depend on covariates. The computer package for transition data analysis (TDA) developed by Rohwer and used by Blossfeld and Rohwer (2002) has a facility for user-defined rate models (Rohwer and Pötter 1999: Section 6.17.5). The program may be downloaded from Professor G. Rohwer's homepage at http://www.stat.ruhr-uni-bochum.de/ . The manual can be downloaded from the same site.

In some cases, the researcher is not interested in the age dependence of migration rates, but in the effect of covariates on the migration level. Rather than omitting age altogether, as in the basic exponential model, the migration rate is allowed to vary with age but the effect of the covariates on the migration rate does not vary with age. The transition rate model that results is a Cox proportional hazard model. It is written as

$$m_i(x) = m_{i0}(x)\ \exp\left[\beta_{i0} + \beta_{i1} Z_1 + \beta_{i2} Z_2 + \cdots\right], \tag{6.49}$$

where $m_{i0}(x)$ is the baseline hazard. It is the set of age-specific migration rates for the reference category. Note that if the age dependence (age structure) of migration is independent of the dependence on covariates (motivational structure), the baseline hazard may be represented by a parametric model and the two components may be estimated separately.

This brief discussion illustrates that transition rate models are ideally suited to impose age structures onto migration data. The same applies for spatial structures and motivational structures. Spatial interaction models may be used for that purpose, but also Poisson regression models (log-linear models) with offset. The migration rates in the model presented in this section depend on origin, destination and age, and they may vary over time. Several authors have found that spatial patterns of migration are stable in time (Baydar 1983; Rogers and Raymer 2001; Berentsen and Cromley 2005). It implies that the origin–destination interaction does not vary in time. Higher-order interaction effects may be omitted if empirical evidence shows that they are not significant. Van Imhoff *et al.* (1997) and van der Gaag *et al.* (2000) used a series of empirical tests to determine which combinations of the four dimensions of interregional migration (origin, destination, age and sex) were needed to describe the migration patterns adequately. Many of the simplifications that result from omitting higher-order interaction effects are studied in the context of demographic projection models.

6.4.4 From transition probabilities to transition rates

In this section, we assume that migration is measured in discrete time. Examples include the census (based on the residence at time of census and five years prior to the census). From that information, the approximate transition rates can be derived. The problem is equivalent to one in which we are given $\mathbf{P}(x, y)$ and $\mathbf{M}(x, y)$ is required. The derivation starts with the exponential expression in Equation (6.29). The exponential expression may be approximated by the linear model:

$$\mathbf{P}(x, y) = \left[\mathbf{I} + \tfrac{1}{2}\mathbf{M}(x, y)\right]^{-1} \left[\mathbf{I} - \tfrac{1}{2}\mathbf{M}(x, y)\right]. \tag{6.50}$$

The approximation is adequate when the transition rates are small or the interval is short.

The derivation of the rate of migration during an interval from information in regions of residence at two consecutive points in time is known as the inverse problem: transition rates are derived from transition probabilities (Singer and Spilerman 1979). The transition rates may be written as:

$$\mathbf{M}(x, y) = \frac{y-x}{2}\left[\mathbf{I} - \mathbf{P}(x, y)\right]\left[\mathbf{I} + \mathbf{P}(x, y\right]^{-1}. \tag{6.51}$$

This inverse relation may be used to infer transition probabilities for intervals that are different from the measurement intervals. For instance, if changes of address are recorded over a period of five years, the inverse relation may be used to infer the average migration rates $\mathbf{M}(x,y)$ and the transition probabilities over a one-year period. The expression is

$$\mathbf{P}(x, x+1) = \exp\left[-\mathbf{M}(x, x+1)\right], \tag{6.52}$$

where $\mathbf{M}(x, x+1)$ is estimated from $\mathbf{P}(x, y)$ using the inverse method. The method assumes that migration rates are constant during the (x, y) interval and that the linearity is an adequate approximation of the exponential model.

6.5 Incomplete data

6.5.1 Adding statistical data

In the previous section, it is assumed that the data are adequate to estimate the parameters of the probability models that are specified. The method applied is the maximum likelihood method. In this section the assumption is relaxed. Some data may be missing.

If data are missing, the strategy consists of two steps. The first is to *predict* the missing data and the second step is to *estimate* the parameters of the model assuming that the data are complete. This two-step procedure is the EM (expectation–maximisation) algorithm (McLachlan and Krishnan 1997). Suppose that we are interested in migration by origin and destination (N_{ij}), but the data are limited to departures and arrivals by region (n_{i+} and n_{+j}). The model is

$$E\left[N_{ij}\right] = \lambda_{ij} = \alpha_i \beta_j. \tag{6.53}$$

In the first step, the expected value of N_{ij} is determined assuming values for the parameters (*expectation*). Assume $\alpha_i = 1$ and $\beta_j = 1$. Hence $E[N_{ij}] = 1$. In the second step, the parameters of the model are estimated by maximising the probability that the model predicts the data (maximisation). If the observations are independent, the probability model is the Poisson model

$$\Pr\{N_{ij} = n_{ij}\} = \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!}\exp(-\lambda_{ij}) \qquad \text{with } E\left[N_{ij}\right] = \lambda_{ij} = \alpha_i \beta_j. \tag{6.54}$$

The maximisation of the probability is equivalent to maximising the log-likelihood:

$$l = \sum_{ij} \left[n_{ij} \ln(\alpha_i \beta_j) - \alpha_i \beta_j \right]. \tag{6.55}$$

The first-order conditions result in the following equations:

$$\hat{\alpha}_i = \frac{n_{i+}}{\sum_j \hat{\beta}_j} \quad \text{and} \quad \hat{\beta}_j = \frac{n_{+j}}{\sum_i \hat{\alpha}_i}. \tag{6.56}$$

The EM algorithm results in the well known expression,

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

which may be written as $\lambda_{ij} = n_{i+} p_j$, where n_{i+} is the number of migrations originating in i and p_j is the probability that a migration ends in j. The destination probability applies to all migrations in the pool of migrations and is independent of the origin.

The method just described imposes a spatial structure onto migration flows (Willekens 1999). The spatial structure implies that destinations are independent of regions of origin. Suppose we have reason to believe that origin matters, i.e. the choice of destination is influenced by the origin. For instance, migrants are more likely to settle in a region not too far from the region of origin. If distance matters, the prediction of λ_{ij} may be improved by the addition of a distance factor to the model:

$$E\left[N_{ij}\right] = \lambda_{ij} = \alpha_i \, \beta_j \, \gamma_{ij}, \tag{6.57}$$

where the distance factor γ_{ij} may be a function of the friction between i and j, i.e. the higher the friction, the lower the value of γ_{ij}. An expression that satisfies the relation between distance factor and spatial friction is $\gamma_{ij} = \exp\left(-c_{ij}\right)$, where c_{ij} is a measure of friction between i and j. The model that results is the gravity model

$$E\left[N_{ij}\right] = \lambda_{ij} = k \, a_i \, \beta_j \, \exp\left(-c_{ij}\right), \tag{6.58}$$

where $\alpha_{ij} = k a_{ij}$ and k is a scaling factor. Note that the model is a log-linear model,

$$\ln \lambda_{ij} = u + u_i^{\mathrm{A}} + u_j^{\mathrm{B}} + u_{ij}^{\mathrm{AB}}, \tag{6.59}$$

where u_i^{A} is the effect associated with origin (variable A) i, u_j^{B} is the effect associated with destination (variable B) j, and u_{ij}^{AB} is the interaction effect between origin and destination. The model is the conventional log-linear model (see e.g. Agresti 1996: Chapter 6). The gravity model and other spatial interaction models may be represented as log-linear models (Willekens 1980, 1983; Rogers *et al.* 2003a). Although the translation opens new perspectives and provides a thorough statistical basis for the modelling of spatial interaction, the log-linear specification of spatial interaction models has not caught on – see e.g. the review of spatial interaction modelling by Roy and Thill (2004).

Interaction effects between origin and destination may be derived from different data sources. One source is a historical migration matrix. Snickars and Weibull (1977), working on internal migration, found that migration tables of some period in the past provide much better estimates of accessibility than any distance measure. Since the publication of that article, authors started to use historical migration matrices instead of distance measures to represent the spatial interaction. Later, authors did not limit the *a priori* information to a single year in the past, but use trends over several years (see e.g. Jörnsten *et al*. 1990). The historical migration may be accurate measurements of the 'revealed' residential preferences. For a prediction of migration flows involving historical migration matrices, see Rogers *et al*. (2003a:64ff.). Table 6.1 shows one of the results. The prediction problem is to determine the interregional migration flows in the United States during the period 1980–1985 from information on the number of residents in the different regions in 1985, the number of residents in 1980 by region, and the flows during the period 1975–1980. The model is

$$E\left[N_{ij}\right] = \lambda_{ij} = \alpha_i^* \beta_j^* m_{ij}^0, \tag{6.60}$$

where m_{ij}^0 is an element of the 1975–1980 migration matrix and λ_{ij} is the predicted number of (i, j) migrants during the period 1980–1985. The historical matrix serves as an initial guess of migrations between origins and destinations. Note that the model is equivalent to the biproportional adjustment method or RAS model that is commonly used in input–output analysis.[3] The 1975–1980 migrant flow, the observed 1980–1985 flow and the predicted 1980–1985 flow are shown in Table 6.1(a)–(c). The marginal totals of the predicted migration flows are consistent with the data (1980–1985). The origin–destination interactions exhibited by the predicted values are identical to the interactions exhibited by the 1975–1980 flows. The interaction can be interpreted in terms of odds and odds ratios. The odds that, in 1975–1980, a migrant from the Midwest selects the South rather than the West is 1845/1269 = 1.454. The odds for a migrant from the Northeast is 1800/753 = 2.390. A migrant from the Northeast prefers the South over the West considerably more than a migrant from the Midwest. The odds ratio is 2.390/1.454 = 1.644. The odds of selecting the South rather than the West are 64 % higher for a migrant from the Northeast than for a migrant from the Midwest. The migration flow that is predicted for the period 1980–1985 exhibits the same odds ratio: [1614/632]/[1977/1272] =1.644. Main effects and interaction effects that are not included in the contemporary data are 'borrowed' from the historical data (Willekens 1982). For detailed

[3] In the input–output literature, the method has become known as RAS because of the notation used by Sir Richard Stone (1961), who developed the method. The method projects a matrix **A** to give it column and row sums of another matrix; the result is **RAS**, where **R** and **S** are diagonal matrices. For a history of method development, see Lahr (2004).

Table 6.1 US interregional migration flows (in thousands): 1975–1980 and 1980–1985.

Region of origin	Region of destination				
	Northeast	Midwest	South	West	Total
(a) 1975–1980 data					
Northeast	43 123	462	1 800	753	46 138
Midwest	350	51 136	1 845	1 269	54 600
South	695	1 082	67 095	1 141	70 013
West	287	677	1 120	37 902	39 986
Total	44 455	53 357	71 860	41 065	210 737
(b) 1980–1985 data					
Northeast	44 845	379	1 387	473	47 084
Midwest	326	52 311	1 954	1 144	55 735
South	651	855	68 742	1 024	71 272
West	237	669	1 085	40 028	42 019
Total	46 059	54 214	73 168	42 669	216 110
(c) 1980–1985 flows predicted based on marginal totals and 1975–1980 matrix					
Northeast	44 445	393	1 614	632	47 084
Midwest	431	52 055	1 977	1 272	55 735
South	814	1 047	68 324	1 087	71 272
West	369	719	1 253	39 678	42 019
Total	46 059	54 214	73 168	42 669	216 110
(d) 1980–1985 flows predicted based on West survey and 1975–1980 matrix					
Northeast	21 181	272	1 037	473	22 963
Midwest	247	43 135	1 526	1 144	46 052
South	488	909	55 235	1 024	57 656
West	237	669	1 085	40 028	42 019
Total	22 153	44 985	58 883	42 669	168 690
(e) 1980–1985 flows predicted based on West survey, 1975–1980 matrix and judgemental data					
Northeast	40 243	516	2 365	899	44 023
Midwest	296	51 762	2 197	1 373	55 628
South	488	909	66 282	1 024	68 703
West	237	669	1 302	40 028	42 236
Total	41 264	53 856	72 146	43 324	210 590

Source: Rogers *et al.* (2003a).

discussions on how to impose spatial structures and age structures onto migration flows, the reader is referred to Rogers *et al*. (2002a, 2003a).

6.5.2 Adding judgemental data

Combining data from several sources implies the use of prior information on the association between the cross-classified variables. In Section 6.5.1, the prior information is derived from other statistical data, such as a historical migration matrix. The interaction effects that are not contained in the contemporary data are ‘borrowed’ from the more detailed, but outdated, source. The estimation algorithm makes use of this prior information only when more reliable (recent) information is lacking. The estimation problem may be given a Bayesian interpretation, making the use of prior information more explicit (Albert and Gupta 1983; Congdon 2001).

The prior beliefs on the degrees of association between the variables are not restricted to information contained in statistical data. The prior information may also be derived from established theory or from expert judgements. Most experience to date with the quantification and utilisation of expert knowledge is in the field of risk analysis and artificial intelligence. The combination of statistical information and judgemental knowledge is common practice in forecasting to improve forecasts. Today it is generally accepted that judgemental and statistical methods each have unique strengths that they can bring to the forecasting process (Armstrong and Collopy 1998). Expert judgements are being used for the forecasting of internal and international migration (for an early review, see Willekens 1994:28ff.). Harker (1986) combined subjective judgements of experts in the migration field with quantitative data such as physical distance and wage and employment rates to make predictions of future migration patterns in the United States. His method of combining judgemental and statistical data is based on Saaty’s analytic hierarchy process (AHP). The AHP is a method by which subjective weights are assigned to a set of objects or alternatives. Cook *et al*. (1984) use the AHP method in combination with time series analysis to make predictions of intra-urban migration. The AHP is used to correct the results of a statistically based forecast. George and Perreault (1992:93–95) report that Canada uses a consensus approach based on opinions of experts and/or administrators and that the Netherlands considers discussions with experts on international migration. The uncertain opinions of experts are used often because no other reliable data exist. Expert opinion can be a very useful source of data. But, as Cooke stresses in his book *Experts in uncertainty*, proper use of this source requires new techniques (Cooke 1991:3). Particularly, the estimation and forecasting of international migration may benefit from new techniques developed in the context of judgemental forecasting, since international migration depends on many factors that are difficult to incorporate into a formal model (van de Kaa 1993:87–88):

> The flow of migrants between countries is, normally, regulated by charters, covenants, treaties and similar agreements between (groups of) states, and by such rules and practices as individual countries choose to apply . . . Thus the way international migration is shaped and is likely to develop, depends to a great extent on the nature of the relations between the countries involved.

What is relevant to forecasting, i.e. the estimation of missing information pertaining to the future, may be relevant to the estimation of missing information in general. Expert opinion may be viewed as data (Cooke 1991:80). It turns out that the use of judgemental data in the modelling of migration is not at all a new idea. Knudsen (1992) discusses the method of how to include *a priori* information in the context of generalised linear models. He suggests including *a priori* information into the model by treating this information as a covariate having a known parameter value of unity. Note that this approach is similar to the method discussed in Section 6.4.1. In order to understand and monitor migration flows, expert knowledge on the causal structure of migration and on any other feature of migration should be used in combination with statistical data. The question of how to use expert knowledge properly remains largely open.

Rogers *et al*. (2003a) use judgemental data in combination with historical statistical data to predict migration flows between the four regions of the United States (Northeast, Midwest, South and West). The estimates in Table 6.1 were obtained assuming that the arrivals and departures are given for the period 1980–1985 and that, in addition, the full migration matrix for the period 1975–1980 is given. Now we assume that the information at hand is restricted to the historical migration matrix (1975–1980) and a migration survey carried out in the West. The survey gives the arrivals in the West during the period 1980–1985 by region of origin, and the departure from the West during the same period by region of destination. The data are shown in the fourth row and fourth column of the 1980–1985 migration flows in Table 6.1. The 1980–1985 migration flows are predicted using the survey data and the historical matrix of 1975–1980. Table 6.1(d) shows the results. The migrations from the West and to the West are equal to the input data. The other migration flows are obtained assuming the interaction patterns observed in the 1975–1980 period.

Suppose that experts indicate that the attractiveness of the West diminished in the early 1980s and that the South became more attractive. In addition, assume that other studies showed an increased propensity to leave the Northeast and the Midwest. That information can be incorporated into the parameter values of the log-linear model – that is, into the odds. Suppose the odds that a migrant selects the South rather than the West is 20 % higher than revealed by the migration pattern to the West between 1980 and 1985, and suppose that the odds that a migrant into the West originated in the Northeast (rather than in the West) is 9 % higher, and that in the Midwest it is 20 % higher than revealed by the patterns of arrival in the West. The parameters of the log-linear model with offset would then change accordingly. Table 6.1(e) shows the predictions. The predictions are considerably improved, except for the migration from the Northeast to the West, which is highly overpredicted.

6.6 Conclusion

This chapter presents a unified perspective on the modelling of migration. The level and direction of migration may be represented by different data types. Three types are distinguished: counts, rates and probabilities. The different representations of migration may be reduced to one of these types or a combination. Different types of migration data may be modelled using techniques that have been developed for the analysis of life history data. The core consists of a (multistate) transition model in discrete time and continuous time. Transitions recorded in continuous time are direct transitions or events (migrations). The migration model is a transition rate model. Transitions recorded in discrete time refer to migrants who are identified by comparing the place of residence of a person at two consecutive points in time. The appropriate model is a transition probability model, which is a multinomial logistic regression model.

A major conclusion of the chapter is that models that have been developed to study life histories are perfectly suited for the study of migration. They allow the analysis of data of different types and the conversion of one data type into another. They also allow the treatment of incomplete data. Initially migration models were spatial interaction models. Today, they are increasingly considered as applications of multistate event history models.

References

Agresti A. 1996. *An introduction to categorical data analysis*. New York: John Wiley & Sons, Inc.

Albert JH and Gupta AK. 1983. Estimation in contingency tables using prior information. *Journal of the Royal Statistical Society B* 45(1):60–69.

Andersen PJ, Borgan O, Gill RD and Keiding N. 1993. *Statistical models based on counting processes*. New York: Springer.

Aoki M. 1976. *Optimal control and system theory in dynamic economic analysis*. New York: North-Holland.

Armstrong JS and Collopy F. 1998. Integration of statistical methods for judgment for time series forecasting: principles and empirical research. In *Forecasting with judgment*. Wright G and Goodwin P, eds, pp. 269–293. New York: John Wiley & Sons, Inc. Available at http://design.cwru.edu/researchArticles/ForecastIntegration.pdf (accessed 7 September 2005).

Baydar N. 1983. *Analysis of the temporal stability of migration in the context of multiregional forecasting*. Working Paper no. 38. Netherlands Interuniversity Demographic Institute, Voorburg.

Bell M, Blake M, Boyle P, Duke-Williams O, Rees P, Stillwell J and Hugo G. 2002. Cross-national comparison of internal migration: issues and measures. *Journal of the Royal Statistical Society* 165(3):435–464.

Berentsen WH and Cromley RG. 2005. Interstate migration flows in Germany since unification: temporal and spatial patterns. *Eurasian Geography and Economics* 46(3):185–201.

Blossfeld HP and Rohwer G. 2002. *Techniques of event history modeling. New approaches to causal analysis,* 2nd edn. Mawah, NJ: Lawrence Erlbaum.

Burch TK. 1999. *Computer modelling of theory: explanation for the 21st century*. Discussion Paper no. 99–4. Population Studies Center, University of Western Ontario, London, Ontario.

Burch TK. 2003. Demography in a new key. A theory of population theory. *Demographic Research* 9(11):261–284.

Çinlar E. 1975. *Introduction to stochastic processes*. Englewood Cliffs, NJ: Prentice-Hall.
Congdon P. 2001. *Bayesian statistical modelling*. Chichester: John Wiley & Sons, Ltd.
Cook T, Falchi P and Mariano R. 1984. An urban allocation model combining time series and analytic hierarchy methods. *Management Science* 30:198–208.
Cooke RM. 1991. *Experts in uncertainty. Opinion and subjective probability in science*. Oxford: Oxford University Press.
Courgeau D. 1973. Migrants et migrations. *Population* 28:95–128.
Courgeau D. 1982. Comparaison des migrations internes en France et aux Etats-Unis. *Population* 6:1184–1188.
De Mesnard L. 2004. Biproportional methods of structural change analysis: a typological survey. *Economic Systems Research* (16(2):205–230.
Director SW and Rohrer RA. 1972.*Introduction to system theory*. New York: McGraw-Hill.
Duke-Williams O and Blake M. 1999. Database fusion for the comparative study of migration data. Presented at the *Fourth International Conference of Geocomputation*, Fredricksburg, Virginia, 25–28 July. Available at http://www.geovista.psu.edu/sites/geocomp99/Gc99/068/gc_068.htm (accessed 10 September 2005).
Evans M, Hasting N and Peacock B. 2000. *Statistical distributions*, 3rd edn. New York: John Wiley & Sons, Inc.
George MV and Perreault J. 1992. Methods of external migration projections and forecasts. In *National population forecasting in industrialized countries*. Keilman N and Cruijsen H, eds, pp. 87–103. Amsterdam: Swets and Zeitlinger.
Hachen DS. 1988. The competing risk model. *Sociological methods and research*, 17(1): 21–54. Reprinted in *Readings in population research methodology*. Bogue DJ, Arriaga EE and Anderton DL, eds, pp. 21.85–21.101. Chicago: Social Development Center; and New York: UNFPA.
Harker PT. 1986. The use of expert judgments in predicting interregional migration patterns: an analytic hierarchy approach. *Geographical Analysis* 18(1):62–80.
Harris B. 1983. Positive and normative aspects of modelling large-scale social systems. In *Systems analysis in urban policy-making and planning*. Batty M and Hutchinson B, eds, pp. 475–490. New York: Plenum.
Jackson RW, Rogerson P, Plane D and O'Huallachain B. 1990. A causative matrix approach to interpreting structural change. *Economic Systems Research* 2(3):259–269.
Jörnsten KO, Larsson T, Lundgren JT and Migdalas A. 1990. An entropy model with variable target. *Environment and Planning A* 22(4):493–506.
Kitsul P and Philipov D. 1981. The one year/five year migration problem. In *Advances in multiregional demography*. Rogers A, ed. Research Report RR-81-6, pp. 1–33. Laxenburg: International Institute for Applied Systems Analysis.
Knudsen DC. 1992. Generalizing Poisson regression: including a priori information using the method of offsets. *Professional Geographer* 44(2):202–208.
Kupiszewski M and Kupiszewska D. 2003. *Internal migration component in subnational population projections in member states of the European Union*. CEFMR Working Paper 2/2003. Central European Forum for Migration Research, Warsaw.
Lahr M. 2004. Biproportional techniques in input–output analysis: table updating and structural analysis. *Economic Systems Research* 16(22):115–134.
Ledent J. 1980. Multistate life tables: movement versus transition perspectives. *Environment and Planning A* 12:533–562
Long JF and Boertlein CG. 1981. *Using migration measures having different intervals*. Washington, DC: US Bureau of the Census.

McLachlan GJ and Krishnan T. 1997. *The EM algorithm and extensions*. New York: John Wiley & Sons, Inc.

Mueser P. 1989. The spatial structure of migration: an analysis of flows between states in the USA over three decades. *Regional Studies* 23(3):185–200.

Plane DA and Mulligan GF. 1997. Measuring spatial focusing in a migration system. *Demography* 34:251–262.

Plane DA and Rogerson PA. 1986. Dynamic flow modeling with inter-regional dependency effects: an application to structural change in the US migration system. *Demography* 23(1):91–104.

Pooler J. 1992. Spatial uncertainty and spatial dominance in interaction modelling: a theoretical perspective on spatial competition. *Environment and Planning A* 24(7):5–1008.

Rajulton F. 1999. LIFEHIST: Analysis of life histories: a state-space approach. Presented at the *Workshop on Longitudinal Research in Social Science: A Canadian Focus*. Windermere Manor, London, Ontario, 25–27 October .

Rees P and Willekens FJ. 1986. Data and accounts. In *Migration and settlement: A multiregional comparative study*, Rogers A and Willekens FJ, eds, pp. 19–58. Dordrecht: Reidel.

Rogers A and Raymer J. 1998. The spatial focus of US interstate migration flows. *International Journal of Population Geography* 4:63–80.

Rogers A and Raymer J. 2001. Immigration and the regional demographics of the elderly population in the United States. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 56(1):S44–S55.

Rogers A and Sweeney S. 1998. Measuring the spatial focus of migration patterns. *Professional Geographer* 50(2):232–242.

Rogers A and Willekens FJ. 1986. A short course on multiregional demography. In *Migration and settlement: A multiregional comparative study*. Rogers A and Willekens FJ, eds, pp. 355–384. Dordrecht: Reidel.

Rogers A, Willekens FJ and Raymer J. 2001. Modeling interregional migration flows: continuity and change. *Mathematical Population Studies* 9:231–263.

Rogers A, Willekens FJ and Raymer J. 2002a. Capturing the age and spatial structures of migration. *Environment and Planning A* 34:341–359.

Rogers A, Willekens FJ, Little K and Raymer J. 2002b. Describing migration spatial structure. *Chapters in Regional Science: Journal of the Regional Science Association International* 81(1):29–48.

Rogers A, Willekens FJ and Raymer J. 2003a. Imposing age and spatial structures on inadequate migration flow data sets. *The Professional Geographer* 55(1):56–69.

Rogers A, Raymer J and Newbold KB. 2003b. Reconciling and translating migration data collected over time intervals of different widths. *The Annals of Regional Science* 37(4):581–601.

Rohwer G and Pötter U. 1999. *TDA user's manual*. Ruhr-Universität Bochum, Fakultät für Sozialwissenschaften, Bochum, Germany.

Roy JR and Thill JC. 2004. Spatial interaction modelling. *Chapters in Regional Science* 83(1):339–361.

Schoen R. 1988. *Modeling multigroup populations*. New York: Plenum Press.

Sen A and Smith T. 1995. *Gravity models of spatial interaction behavior*. Berlin: Springer.

Singer B and Spilerman S. 1979. Mathematical representations of development. theories. In *Longitudinal research in the study of behavior and development*. Nesselroade JR and Baltes PB, eds, pp. 155–177. New York: Academic Press.

Snickars F and Weibull JW. 1977. A minimum information principle. Theory and practice. *Regional Science and Urban Economics* 7:137–168.

Stillwell J. 2005. Inter-regional migration modeling: a review and assessment. Presented at the *45th Congress of the European Regional Science Association*, Amsterdam, 23–27 August.

Stone R. 1961. *Input–Output and national accounts*. Organization for Economic Cooperation, Paris

Strang G. 1980. *Linear algebra and its applications*, 2nd edn. New York: Academic Press.

Taylor HE and Karlin S. 1998. *An introduction to stochastic modeling*, 3rd edn. San Diego: Academic Press.

Tinbergen J. 1981. The use of models: experiences and prospects. *American Economic Review* 71(6):17–29.

van de Kaa DJ. 1993. European migration at the end of history. *European Review* 1(1):87–108.

van der Gaag NE, van Imhoff E and van Wissen L. 2000. Internal migration scenarios and regional population projections for the European Union. *International Journal of Population Geography* 6:1–19.

van Imhoff E. 1990. The exponential multidimensional demographic projection model. *Mathematical Population Studies* 2(3):171–182

van Imhoff E and Keilman N. 1991. *LIPRO 2.0: An application of a dynamic demographic projection model to household structure in the Netherlands*. Amsterdam: Sweets and Zeitlinger.

van Imhoff E, van Wissen L and Spiess K. 1994. *Regional population projection in the countries of the European Economic Area*. NIDI & CBGS Publication no. 31. Amsterdam: Swets & Zeitlinger.

van Imhoff E, van der Gaag N, van Wissen L and Rees P. 1997. The selection of internal migration models for European regions. *International Journal of Population Geography* 3:137–159.

van Wissen LJG, van der Gaag N, Rees P and Stillwell J. 2005. In search of a modelling strategy for projecting internal migration in European countries: demographic versus economic–geographical approaches. Presented at the *45th Congress of the European Regional Science Association*, Amsterdam, 23–27 August. Available at http://www.feweb.vu.nl/ersa2005/final_chapters/787.pdf

Willekens FJ. 1980. Entropy, multiproportional adjustment and the analysis of contingency tables. *Systemi Urbani* 2:171–201.

Willekens FJ. 1982. Multidimensional population analysis with incomplete data. In *Multidimensional mathematical demography*. Land KC and Rogers A, eds, pp. 43–112. New York: Academic Press. Reprinted in *Readings in population research methodology*. Bogue DJ, Arriaga EE and Anderton DL, eds, pp. 22.49–22.82. Chicago: Social Development Center; and New York: UNFPA.

Willekens FJ. 1983. Log-linear modelling of spatial interaction. *Chapters of the Regional Science Association* 52:187–205.

Willekens FJ. 1994. Monitoring international migration in Europe: towards a statistical data base combining data from different sources. *European Journal of Population* 10(1):1–42.

Willekens FJ. 1999. Modeling approaches to the indirect estimation of migration flows: from entropy to EM. *Mathematical Population Studies* 7(3):239–278.

Willekens FJ. 2001. Theoretical and technical orientations toward longitudinal research in the social sciences. *Canadian Journal of Population* 28(2):189–217. Revised version of The life course: models and analysis. In *Population issues: An interdisciplinary* *focus*. van Wissen LJG and Dykstra PA, eds, pp. 23–51. New York: Kluwer Academic/Plenum.

Willekens FJ and Drewe P. 1984. A multiregional model for regional demographic projection. In *Demographic research and spatial policy*. ter Heide H and Willekens FJ, eds, pp. 309–334. London: Academic Press.

Wilson T and Bell M. 2004. Comparative empirical evaluations of internal migration models in subnational population projections. *Journal of Population Research* 21(2):127–160.

Yamaguchi K. 1991. *Event history analysis*. Newbury Park, CA: Sage.