European Journal of Population 10: 1–42, 1994.

Monitoring International Migration Flows in Europe
Towards a Statistical Data Base Combining Data From Different Sources

Frans Willekens

Population Research Centre, University of Groningen, The Netherlands

Received 2 September 1993; accepted 25 January 1994

Willekens, Frans, 1994. Monitoring international migration flows in Europe. Towards a statistical data base combining data from different sources. European Journal of Population / Revue Européenne de Démographie, 10: 1–42.

Abstract. The paper reviews techniques developed in demography, geography and statistics that are useful for bridging the gap between available data on international migration flows and the information required for policy making and research. The basic idea of the paper is as follows: to establish a coherent and consistent data base that contains sufficiently detailed, up-to-date and accurate information, data from several sources should be combined. That raises issues of definition and measurement, and of how to combine data from different origins properly. The issues may be tackled more easily if the statistics that are being compiled are viewed as different outcomes or manifestations of underlying stochastic processes governing migration. The link between the processes and their outcomes is described by models, the parameters of which must be estimated from the available data. That may be done within the context of socio-demographic accounting. The paper discusses the experience of the U.S. Bureau of the Census in combining migration data from several sources. It also summarizes the many efforts in Europe to establish a coherent and consistent data base on international migration.

The paper was written at IIASA. It is part of the Migration Estimation Study, which is a collaborative IIASA-University of Groningen project, funded by the Netherlands Organization for Scientific Research (NWO). The project aims at developing techniques to obtain improved estimates of international migration flows by country of origin and country of destination.

Willekens, Frans, 1994. La maîtrise des flux de migration internationale en Europe. Vers une base de données statistiques combinant différentes sources. European Journal of Population / Revue Européenne de Démographie, 10: 1–42.

Résumé. Cet article présente des techniques développées en démographie, géographie et statistique, utiles pour réduire l'écart entre les données disponibles sur les flux de migration internationale et l'information nécessaire pour élaborer des politiques et entreprendre des recherches. L'idée de base est la suivante: pour mettre en place une base de données cohérentes et conséquentes qui contienne une information suffisamment détaillée, à jour et précise, il est nécessaire de combiner des données issues de différentes sources. Cela lève des problèmes de définition et de mesure, et pose la question suivante: comment combiner de façon correcte des données d'origine différente? On peut mieux aborder ces problèmes si les statistiques à réunir sont vues comme différents résultats ou manifestations des processus stochastiques sous-jacents qui régissent les migrations. Le lien entre les processus et leurs manifestations est décrit par des modèles, dont les paramètres doivent être estimés à partir des données disponibles.

Cela peut être réalisé dans le contexte de la comptabilité démographique. L'article discute l'expérience du "Bureau of the Census" des États-Unis qui a combiné des données sur les migrations issues de différentes sources. Il présente aussi rapidement les nombreux efforts réalisés en Europe pour établir une base de données cohérente et conséquente sur la migration internationale.

1. Introduction

International migration is a politically sensitive issue. The discussions tend to be emotional, partly because facts are hard to come by. This observation led Dr. Nafis Sadiq, Executive Director of the United Nations Population Fund, to conclude, on the occasion of the Expert Group Meeting on Population Distribution and Migration held in Santa Cruz, Bolivia, in January 1993, that "[t]here is a crucial lack of information about both the size and the nature of migration flows" (Sadiq, 1993, p. 4). A similar concern was expressed by the OECD Council, held at the ministerial level on 4 and 5 June 1991: "Ministers request the Organisation, . . . , to intensify its work on assessing and comparing migratory flows and their trends; . . ." A "growing concern about worldwide migratory pressures" was also issued by the London Summit of the G-7 on 17 July 1991.

Much of the debate in Europe on international migration seems to be determined by fear, a view of international migration as threatening, instead of facts.[1] Mr. Pacini, Director of the Giovanni Agnelli Foundation, expressed the opinion of many in Europe when he states: "In the near future, the birth rates in Arab North Africa are not expected to decline sufficiently enough to contain population expansion, nor will the economic gap with Europe narrow significantly. The perpetuation of a situation in which the Maghreb market is increasingly unable to absorb the available labour force will drive vast numbers of young immigrants to Europe" (Pacini, 1992, p. 230). Fear without reason, according to Simon: "Additional immigrants raise the standard of living of native-born persons in Western countries, and have little or no negative effect upon any occupation or income class" (Simon, 1993, p. 116). Arguments pro and contra migration are many and the facts are few. Van de Kaa (1993) presents an analysis of the factors that may influence the thinking on, and reactions to, international migration. The debate and the uncertainties involved in international migration generated a great deal of attention in all spheres of public life.[2]

Adequate migration data are essential, not only to guide the political debate, but also to monitor and manage migratory movements.[3] The objective of this paper is to review a set of techniques which are used by

statisticians, demographers and geographers to bridge the gap between available data and the information required for policy making and research. Incomplete data problems occur frequently in these disciplines. Many of the solutions that have been proposed are ingenious, but tend to have an *ad hoc* character, being solutions applying to one particular type of missing data problem.[4]

The lack of adequate data on international migration in Europe was felt particularly strong relatively recently during an attempt to assess the demographic consequences of international migration. In 1991, EUROSTAT, the statistical office of the European Communities, called upon the Population Research Centre of the University of Groningen to assist in the preparation of long-term demographic scenarios for the member countries of the European Communities in preparation of the conference on 'Human resources in Europe at the dawn of the 21st century'. Particular attention should be devoted to the demographic consequences of changes in the level of migration and the direction of migration flows. The scenarios should include alternative patterns of migration between the countries of Europe and between Europe and the rest of the world. A baseline projection should be based on the most recent estimates of migration. Because of the emphasis on migration flows, a multiregional demographic projection model was selected to prepare the scenarios. The results are reported in EUROSTAT (1991) and Extercate (1993).

The preparation of the migration data base revealed data problems that were never adequately documented before. Information on migration flows between European countries is very limited and the information that exists is highly inaccurate. In a study commissioned by EUROSTAT, Poulain et al. (1990) found several anomalies in the international migration statistics. For instance, the number of migrations from Italy to Germany in 1988 was 10,926 according to the statistics from Italy and 44,549 according to German sources. The migration from Germany to France was 12,210 according to German statistics and only 1,114 according to French data.[5,6,7] It was decided not to use the migration data but to infer net migration as the difference between population growth and natural growth. Furthermore, the analysis of the changing composition of migration flows towards the European Community has been excluded "due to the general lack of reliable and comprehensive data" (Muus and Cruijsen, 1991, p. 57).

It became apparent that (i) neither the volume nor direction of migration in Europe is properly measured, (ii) the available data on migration do not permit a reliable assessment of the demographic consequences of current and

alternative migration patterns and (iii) the impact assessments that have been carried out using data from a single country may have produced erroneous results. What is urgently needed is an international migration data base that is complete, consistent and contains reliable migration estimates. Completeness is striven for by bringing together information about all main aspects of the migration process. Consistency is reached through coordination of concepts and coverage and by an integration procedure which forces the migration data to fulfill the sets of definitions and accounting equations. The proper and timely measurement and indirect estimation of migration may complement the activities devoted to the assembly of data from existing data sources. EUROSTAT, ECE (both the Statistics Division and the Population Activities Unit) and the OECD are producing data bases from data obtained form national statistical offices. International organizations try to improve the comparability and develop initiatives to harmonize the data that are being collected.

In order to be reliable, the estimates should not be based on a single source of migration data, but on several sources, each source (or even each number of a given source) being weighted by the degree of confidence in how well the source (number) represents reality. The dominant sources of migration data are censuses, surveys (labour force surveys, migration surveys, etc.) and administrative records (e.g. general population register; special-purpose registration, such as the registration of asylum seekers, refugees and/or foreign labour; drivers' licences; tax records; work permits; etc.). The integration of data from several sources raises a number of conceptual and methodological issues. These issues are not specific to international migration, but arise in the study of internal migration as well. The lessons learned in internal migration research, and in mobility research in general, are relevant to studies of international migration.[8] In this paper, it is suggested that the issues may be dealt with more adequately if the problem of integrating data from several sources is situated within the context of demographic accounting. Since Sir Richard Stone proposed in the early sixties the use of accounts as a monitoring device, most countries have developed economic, social and demographic accounts. The methodological issues involved in the construction of the accounts are relatively well understood.

The organization of the paper is as follows. Section 2 reviews the use of accounts within the context of a system of monitoring of the activities of populations. The development of accounts requires the solution of several problems of definition and measurement. In Section 3, the substantive issues

involved in the definition and measurement of migration are briefly discussed. Techniques that are used to bridge the gap between available statistical data and required information are presented in Section 4. These techniques were originally developed for internal migration and other manifestations of spatial interaction. It is shown that the very many methods that have been developed over the years belong to a particular category of regression models, the Generalized Linear Models (GLMs). This observation is important since the various methods may now be given a theoretical underpinning and the further development of these methods may benefit from the extensive literature on GLMs. In order to generate the required information, it is frequently necessary to combine data from different sources. The methods reviewed in Section 4 allow the integration of various data sets. Section 5 reviews work at the U.S. Bureau of the Census aimed at the development of a better data base on interstate migration by combining data from several sources. The available techniques and the U.S. experience are relevant to a major European need today, namely, the production of a coherent, consistent and reliable data base on migration between countries of Europe and between Europe and the rest of the world. The final section of this paper is devoted to new, potentially significant innovations in the production of information, namely, the use of expert knowledge and opinion. The use of expert judgment is the subject of a separate research project (Wils and Willekens, 1993).

2. Accounts

The integration of statistical data from several sources towards a 'synthetic' data base is a well-known problem in the compilation of statistics. Huigen et al. (1988) mention three examples: the compilation of an input-output table for the national accounts, the weighing of micro-data to make a sample representative and the construction of synthetic estimators for the combination of data from a register and a household survey. "In each of these examples data from one source contradict the data from another source, and the data from both sources should be combined in one or another way. Usually this is done by constructing an estimator which is a function of the data in both sources and which is optimal in some sense" (Huigen et al., 1988, p. 28). Much of that work has been carried out within the context of the construction of economic and socio-demographic accounts. The development of accounts has shown to be a very useful approach to integrate

data from several sources, because accounts reveal the relations between the numbers in the data base and forces the correct specification of the system of relations.

In recent years, the combination of data from several sources has become an accepted method to produce a coherent and internally consistent picture of the population structure and the various demographic processes, in particular mobility. For instance, in 1983, the International Labour Organization (ILO) opted for greater coherence in labour statistics and explored the development of a coherent Labour Accounting System (LAS). The LAS was seen as the answer to a growing desire for complete and consistent labour market information (Bos et al., 1988, p. 6; Hoffmann, 1990). In the early 1980s, Statistics Sweden responded to the growing demand for small area statistics by combining data from population registers, sample surveys and censuses (Lundström, 1984). In the mid-1980s the Netherlands Central Bureau of Statistics (NCBS) initiated the development of socio-demographic accounts to give detailed and coherent information on the structure (stock) and mobility (flows) of subgroups of the Dutch population (educational mobility, labour force participation and mobility, and household change) (Koesoebjono, 1985a, 1985b; NCBS, 1989, 1990). The accounts are constructed by integrating data sources already existing within the NCBS. Both micro- and macro-data from various sources are integrated into the accounts. Internal consistency is an important objective of accounts. Koesoebjono states: "The development of accounts will be instrumental in increasing the consistency of statistics originating from different sources, which implies the use of a common method of data collection, concepts, definitions and classifications" (Koesoebjono, 1985a, p. 3). The problem of consistency can hardly be overestimated. Apart from its statistical purpose, the account also serves as a detector of missing links in the system of statistics. From the account, it becomes immediately clear if no statistical information is available on specific flows. "This offers a guiding principle in the process of developing new statistics" (Huigen et al., 1988, p. 36; see also Koesoebjono, 1985a, p. 3).

Demographic accounts are tables of population flows that link together population stocks at two points in time (Rees, 1993, p. 28). The pioneer in developing demographic accounting was Richard Stone (e.g. Stone et al., 1942; Stone, 1971; United Nations, 1975), who saw them as part of a much wider system of social and demographic monitoring of the activities of national populations. Much of the advantage of using accounts in the production of statistics is due to the simple requirement that accounts must

balance: the number of persons in a given state at the end of a period must be equal to the number in that state at the beginning of the period plus the inflows minus the outflows.

Demographic accounts are also being used to produce coherent and internally consistent data on migration. Rees and Wilson (1977) applied and extended the system of social and demographic statistics, developed by Stone, to internal and international migration. Their work was utilized in the International Comparative Migration and Settlement Study, which was carried out by IIASA in the late 1970s and early 1980s and involved the analysis of migration and population distribution patterns in all of IIASA's member countries (Rogers and Willekens, 1986; for a discussion of migration data problems, see Rees and Willekens, 1986). The most recent review of demographic accounts is by Rees (1993). Although the work by Rees and Wilson, IIASA and others dealt mostly with internal migration, the results of that research are useful for international migration. The U.S. Bureau of the Census devoted considerable energy to combining migration data from different sources in order to produce a data base of interstate (internal) migration for state population projections. The experience is described in Section 5.

3. The identification and measurement of migration

The event of migration is generally not unambiguously defined, and even if it is, it is frequently not properly measured. In addition, its interpretation is complicated by the fact that migration is generally subordinate to other life events. Migration is not an end in itself, but serves other purposes, such as household formation, job change, search for better living conditions, amenities, accessibility to various types of facilities, etc. The issues of concept, measurement and interpretation need to be considered in establishing a data base on migration. For an extensive discussion of the issues, see Willekens (1982b, 1985), Courgeau (1988) and Zlotnik and Hovy (1990, pp. 1–4).

Migration is generally defined as a relatively permanent change in residence (usual residence) that crosses jurisdictional boundaries. It is only one of several forms of spatial movement (Morrison, 1983). Other forms include travel, commuting, local residential change and seasonal migration. They are relevant for migration because one form of spatial movement may lead to another form.[9] For instance, for different reasons commuters may opt for migration. The forms of spatial movement constitute different responses to

the fact that facilities and opportunities are geographically situated (located in space). Ideally, the complexity of spatial movements is reflected in the statistical information available. This is not feasible at the moment. It would require a coherent system of statistics on travel, commuting, residential change within jurisdictional boundaries and various types of migration.

The concept of migration involves two dimensions; namely, time and space. Migration is said to take place in a time-space framework (Rees and Willekens, 1986). The concept of usual residence generally involves some reference to an actual duration of residence or an intended duration of stay (degree of permanence). The reference to the boundary between jurisdictions introduces the spatial dimension. The boundary could be between entities as small as a municipality or as large as an entire country. Since the geographical unit used affects the level of migration, some authors advocate that migration be defined in terms of the actual distance (in kilometers or miles) between origin and destination, regardless of jurisdiction (Long, 1988).

The various definition problems are reflected in the measurement of migration. The distinction between migration concept and migration measurement is not always clear. In many countries, the definition of migration is determined by the way migration is measured. For instance, in many developing countries the census is the main source of migration information and a migrant is defined as a person who is living at the time of the census in a place different from the place of birth. In discussing the measurement problems, it is useful to distinguish between 'migration' and 'migrant' (Courgeau, 1973, 1980). Migration is the event of change of residence that involves the crossing of an administrative boundary. Frequently, migration is measured only indirectly by comparing the residence of a person at two points in time. A migrant is a person who at a given point in time is residing in a place different from the place of residence at a previous point in time. If the length of time interval is small, the difference between the number of migrants and migrations is negligible, but the number of migrations per migrant is expected to increase with the length of the interval. The period between the two points may be fixed (e.g. one year, five years) or variable (e.g. age in case of place-of-residence-by-place-of-birth data and duration of residence in case of place-of-current-residence-by-place-of-previous-residence data). Registration systems collect information on migrations, since every change of usual residence must be reported to the local authorities. In censuses and surveys, information may be recorded on either migrants or migration. When the place of

residence at some previous point in time is recorded (and compared with the place of residence at the time of the census or survey), the data refer to migrants; when the previous place of residence and the duration of current residence are recorded, they refer to migration. Rees refers to data on migration as 'movement data' and to data on migrants as 'transition data' (see e.g. Rees, 1993, pp. 36–38).

4. Indirect estimation of migration flows: Model approaches

Migration data are collected with a purpose. The information is used either to determine the levels and patterns of mobility, or to be used as input to population projections. For projection purposes, the migration data are generally required by age and sex. Long (1991, 1992) provides a nice illustration of the use of migration data in international comparative research on mobility. The use of migration data for the purpose of projection is discussed by Speare (1992).

The inference of gross migration flows from aggregate and incomplete data has a long history.[10] The methods that have been proposed over the years are described in this section in a historical perspective. Although the methods that have been developed mainly pertain to internal migration, they are very relevant to determining international migration flows by country of origin and country of destination. The 'hard' statistical data that may be used to determine the migration flows may be highly inadequate. That is particularly true in international migration. Some authors therefore supplement statistical information with expert judgment in order to determine recent or current migration patterns and predict future migration. This project builds on their work.

Methods for inferring *net* migration exist for several decades. Migration is approximated by the population change that cannot be attributed to natural increase. The data required consist of population sizes at two points in time and estimates of birth and death rates for the interval. Migration is approximated by the difference between the observed population at the second point in time and the hypothetical population that would have resulted if only natural increase would have applied. The methods are reviewed in United Nations (1970) and Courgeau (1988).

The remainder of this section is on the estimation of gross flows. In estimating gross flows from incomplete data, three directions may be distinguished:

- use of model migration profiles,
- estimation of gross migration flows from aggregate data on stocks, and
- estimation of gross migration flows from aggregate data on flows.

4.1. *Model migration profiles*

If data on migration in a particular population is incomplete or entirely lacking, information on migration in other populations that are similar may be used to obtain an estimate (guess) of the level of migration and the composition of migrants. Model migration profiles have shown to be particularly useful in providing detailed information on the age composition of migration.

Rogers (1975) approached the problem of estimating age-specific gross migration from incomplete data by generalizing techniques of indirect estimation developed in demography. One of the methods he developed is the multiregional model life table, which is a generalization of the Coale-Demeny model life table to a multiregional population (see also Rogers, 1973). The model table technique is based on empirical regularities of profiles of age-specific demographic rates (rates of mortality in model life tables, rates of fertility in model fertility tables, rates of first marriage in nuptiality tables). The single-region model life table approximates the mortality schedule of a population by resorting to the mortality experience of other populations that are presumed to be similar. A model multiregional life table approximates the mortality and migration profiles of a population disaggregated by place of residence, by resorting to the mortality and migration experience of other populations of a similar composition and in similar conditions. Using more than 500 recorded migration profiles in a considerable number of countries (mostly developed countries), Rogers found that although levels of migration vary substantially from region to region, the shape of age-specific migration curves are quite similar across a wide range of communities (Rogers and Castro, 1981, 1986). The selectivity with respect to age is therefore the dominant feature of migration. The observation motivated Rogers and Castro to summarize the regularities by means of mathematical expressions called model migration schedules and to derive 'typical' migration profiles. Castro (1985) provides some intelligible basic parameter values to produce age profiles of migration from information on the labour force population. The utility of such mathematical expressions lies in their ability to summarize and codify the fundamental regularity exhibited by age profiles of migration all over the world (Castro, 1985). Model migra-

tion schedules have been used extensively as a data reduction technique. Their use for estimating age profiles from incomplete data remained limited to the updating of migration profiles in forecasting regional populations (Bates, 1982).

4.2. *Estimation of gross migration flows from aggregate data on stocks*

Another method proposed by Rogers for estimating interregional migration, is a generalization of the procedure suggested by Eldridge and Kim (1968) for estimating net migration. Both methods use the same data: age-specific place of residence by place of birth (PRPB) data from two consecutive population censuses. Assuming that the population is stable during the intercensal period, the method yields composite probabilities of survival and migration during the interval between the two census years (see Rogers, 1975, pp. 172–185; Rogers and von Rabenau, 1971). The PRPB method derives information on intercensal migration by origin and destination from lifetime migration data from two consecutive censuses. The method was recently extended by Rogers (1991) to three instead of two consecutive PRPB data. Note that the PRPB data, when disaggregated by age, contain information on cohorts' migration histories. Using the relationship between cohort and period migration, i.e. between lifetime migration rates and various trends in period migration rates that would result under different historical conditions, Schmertmann (1992) was able to infer historical trends in gross migration from data in a single census.

The PRPB method requires observations on the residence of each individual at two points in time (census and birth). Frequently, the only data available pertains to the residence at one point in time. That is the case when we have cross-sections of the population by region of residence, i.e. when the information is restricted to data on the population distribution. Migration flows may be estimated from these data if a sequence of cross-sections is available. These data constitute a time series of aggregate stock data showing only the distribution of the population by residence at each moment of time. The estimation involves the assumption that migration is governed by a Markov process. The estimation problem is a special case of a more general problem of estimating Markov processes (transition probabilities) from aggregate data on state occupancies (proportions of people in each state at each point in time). The estimation of transition probabilities from aggregate data was extensively studied more than twenty years ago by Lee et al. (1970). The authors assumed that the transition probabilities are

constant over time. The assumption was later relaxed and improved estimators were developed (see e.g. MacRae, 1977; Kalbfleisch and Lawless, 1984). The problem remains that aggregate data do not contain a lot of information about the Markov process. As a result, large numbers of observations are needed to obtain reliable parameter estimates. Kalbfleisch and Lawless (1984, p. 178) state that if one has the choice of obtaining many observations on aggregate data, or a few observations on transition data, the latter alternative would often be preferable (see Lawless and McLeish, 1984).

4.3. *Estimation of gross migration flows from aggregate data on flows*

Most authors follow a different perspective in estimating gross migration flows from incomplete data. They approach migration as an interaction between regions and represent gross migration flows by spatial interaction models, which were originally developed in geography. In essence, a spatial interaction model relates mobility to accessibility. Both mobility and accessibility may be expressed in various ways, resulting in different types of models. The data generally available to estimate gross flows consist of two types of observations: (i) aggregate migration data that are marginal totals of the flow matrices to be determined, and (ii) some estimate of the interaction between regions. The interaction is a proxy for accessibility or, in the case of migration, residential preference. In early applications, the accessibility or spatial interaction was approximated by the inverse of the distance between the regions or a function of distance or travel cost. Distance between origin and destination may only explain a relatively small proportion of the observed migration pattern (for a recent paper, see Roy and Flood, 1992). Snickars and Weilbull (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 do 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. Roughly three periods may be distinguished in the development of estimation methods.

Period 1: The gravity model

The problem of estimating migration flows is a special case of spatial inter-

action modelling, since migrations may be viewed as transactions between regions. At the end of the 19th century, Ravenstein (1885, 1889) found that migration between any two regions is proportional to the population size of the regions and inversely related to the distance between the regions. He described migration by a gravity model. In the early days of quantitative geography, authors described social phenomena by laws borrowed from physics (see e.g. Isard, 1960). Newton's law of gravitation was particularly popular in geography and regional science. It states that any particle of matter in the universe attracts any other with a force varying directly as the product of the masses and inversely as the square of the distance between them. Newton put forward the law in 1687 and used it to explain the observed motions of the planets. The gravity model has been used often to describe all types of spatial interaction flows (e.g. transportation, trade, telephone traffic, etc.)

Period 2: The entropy method

In 1970, Wilson (1970) published a book which had a great impact on spatial modelling. Wilson showed that the most likely spatial configuration (e.g. migration) is one which maximizes the entropy. The concept of entropy was proposed in 1850 by Rudolf Clausius, a German physicist, and is closely related to the concept of disorder or chaos. It is sometimes treated as the second law of thermodynamics, which states that the entropy increases through an irreversible process such as the expansion of gas into a vacuum. The most likely state is therefore one of chaos or maximum entropy. In the statistical interpretation of the entropy, it is found that entropy is proportional to the logarithm of the number of microscopic ways in which a given macroscopic state can be realized (the proportionality factor is the Boltzmann constant). In the case of migration, the macro-state is a given migration matrix (origin-destination table); the micro-state is an assignment of *individual* migrants to the origin-destination table.[11] If information on the migration matrix is incomplete, then the most likely configuration is the one with the highest entropy, i.e. which can be produced by a maximum number of micro-states that satisfy the data. Jörnsten et al. summarize the essence of entropy in spatial interaction analysis as follows: "Entropy models use information on the aggregate level, to reflect actions of individuals taken on the disaggregate level. This information is expressed by a set of linear constraints. The entropy maximization approach generates the most probable solution consistent with these constraints" (Jörnsten et al., 1990, p. 493).

Wilson found that gravity models describe spatial interaction with

maximum entropy. The use of the entropy concept in social sciences may also be justified by means of information theory (Jaynes, 1957), by means of Bayes' theorem for conditional probabilities (Hyman, 1969), or by means of the maximum likelihood estimators (Evans, 1971). The three justifications are still visible in today's literature on spatial interaction modelling. The latter justification is of relevance to this paper. Since the entropy is the kernel of the likelihood function, entropy maximization is equivalent to maximum likelihood estimation. The relation between entropy maximization and maximum likelihood estimation was shown already more than twenty years ago by Batty and MacKie (1972).

Chilton and Poet (1973) were the first to apply the entropy method to the estimation of migration flows. A few years later, Willekens (1977) applied the method to infer age-specific migration flows from aggregate data as part of a multiregional population projection.

In 1970, another important book on the estimation of interactions was published. Bacharach (1970) dealt with the problem of estimating and updating tables describing transactions between sectors of the economy. In order to update these accounts (known as input-output tables), aggregate information on production and consumption by sector was derived from national accounts. The method, originally developed by R. A. Stone in 1962 when developing a system of economic accounts, became known as the RAS method and the biproportional adjustment method. The entropy method and the RAS method were developed independently. Although the methods address basically the same problem, it took several years to realize that they are formally almost identical (MacGill, 1977; Willekens, 1980). It has become clear that the RAS method has been used in several disciplines under various names. In transportation research, it is known as the Fratar or Furness method. Evans (1970) refers to the 'Detroit' method. Friedlander (1961) used a related technique to update the distribution of women by marital status in England and Wales to fit more recent official data on marginal totals. The method is also consistent with the method proposed by Ireland and Kullback (1968), with the iterative proportional fitting proposed by Deming and Stephan (1940) and well-known in contingency table analysis (see Bishop et al., 1975), and with the minimum information model, used by Snickars and Weibull (1977), Batty and March (1976), Plane (1982) and others. For a discussion of the similarities between the different methods, and for more history, see Willekens (1980) and De Jong (1985, pp. 4–6).

The observation that seemingly unrelated estimation techniques are in fact different representations of the same method or belong to the same family

of methods motivated the investigation of the mathematical properties of the different methods. In addition, once the basic mathematical structure of the family of estimation methods is identified, the extension to new and more complex applications becomes almost straightforward. The results are reported in Willekens et al. (1979). In the paper, a general formulation of the estimation method is derived and different special cases are presented. The special cases have to do with the types of observations on which the estimates are based. The method, which was generally referred to as the multiproportional adjustment method (MAM), was extensively applied to infer age-specific gross migration flows as part of the Comparative Migration and Settlement Project at IIASA (Rogers and Willekens, 1986). Nair (1981, 1982, 1985) applied the method to determine recent migration patterns by age and sex in India. Drewe and Willekens (1980) used the method to determine age-specific migration patterns in the Netherlands. Courgeau (1988, Chapter 10) and Rees and Woods (1986) describe the method within a broader treatment of the problem of modelling and estimating migration.

Period 3: The log-linear and generalized linear model

The observation that spatial interaction tables belong to the larger family of mobility tables, which are a type of contingency tables, opened new analytical perspectives. The models that have been developed in statistics for the analysis of cross-classifications may be applied for the description, estimation, updating and forecasting of spatial interaction. In particular, the log-linear model could be very useful. Willekens (1980, 1983a, 1983b) showed that spatial interaction models can be represented as log-linear models. Alonso (1986) made explicit the link between a more general model of spatial interaction (the systemic model of movement) and the log-linear model. The advantage is the unambiguous interpretation of the parameters of the log-linear model. The specification of spatial interaction models as log-linear models also enables the assessment of the contribution of each set of information to the predicted flow. The problem of estimating spatial interaction flows becomes a problem of including the appropriate main and interaction effects in the flow model. The log-linear modelling approach to the estimation of missing values in migration and other transition data is presented in detail in Willekens (1982a). The relation between the log-linear model parameters, the parameters of the gravity model and the Lagrange multipliers of the entropy maximization method are discussed in Willekens (1982a, 1983b). The relation between the entropy method and log-

linear models was shown much earlier by Good (1963) (see Bishop et al., 1975, pp. 345ff).

The log-linear model is itself a special case of a generalized linear model (GLM). The generalized linear model encompasses a family of models, which include the log-linear, logit and probit models. The term 'generalized linear model' is due to Nelder and Wedderburn (1972) who extended the iterative weighted least square method, which was originally used in probit analysis, to maximum likelihood estimation in exponential families (see McCullagh and Nelder, 1989). In the GLM, the model is linear in the parameters. The linearity may be assured by transforming the original dependent variable by an appropriate link function. Applications of GLMs to spatial interaction include Flowerdew (1982), who calibrated the spatial interaction model using the iterative weighted least square method presented by Nelder and Wedderburn (1972), De Jong (1985) and Willekens and Baydar (1986). The approach is consistent with the GLM approach. Willekens and Ramachandran (1993) presented the theory and applied it to the estimation of gross migration flows in India.

Since migration is treated as the outcome of a Poisson process, the migration data are Poisson count data and the regression model is referred to as Poisson regression. Poisson regression is appropriate when the dependent variable is a count, such as the number of migrants of a particular category. The dependent variable is assumed to have a Poisson distribution, which is generally true when the number of events is small compared to the population. A very good overview of the theory and applications in geography is given by Lovett and Flowerdew (1989). Since the Poisson regression belongs to the family of GLMs, the parameters may be estimated using the GLIM statistical package, which implements the theory of GLMs.

Two final observations may be made: a) First, we recently realized that the log-linear modelling approach to the estimation of missing values in migration and other transition data, followed in previous publications and in this paper, has much in common with the method presented by Laird and Olivier (1981) in the context of survival analysis. The objective of Laird and Olivier in their paper was to unite two different fields, i.e. survival and contingency analysis. They show that currently popular approaches to modelling survival data can be handled by methods for the analysis of contingency table data. The events that Laird and Olivier deal with are deaths; the total duration of exposure serves as a measure of 'accessibility'. b) Second, in this section no explicit reference was made to algorithms for computing maximum likelihood (ML) estimates from incomplete data, other than

iterative proportional fitting. The Newton-Raphson algorithm may be used when the number of parameters to be estimated is not too large. The advantage of the method is that it provides asymptotic standard errors for the ML estimates as standard output from the calculations. A very general algorithm for ML estimation in incomplete-data problems is the EM algorithm. It consists of two steps. First, missing values are replaced by a model with a given set of parameters (E-step; Expectation). Second, the parameters of the model are estimated assuming that the data are complete (M-step; Maximization). The EM algorithm is associated with Dempster et al. (1977), although the method had been proposed for many years in special contexts (for a description, see Little and Rubin, 1987, Chapter 7). An unfortunate feature of the EM algorithm is that one does not obtain an estimate of the standard errors for the ML estimates. Recently, Lang (1992) derived an explicit formula for the observed information matrix that can easily be computed.

5. Combining migration data from different sources: The U.S. experience

The establishment of a coherent and consistent data base on migration, be it internal or international migration, preoccupies statistical agencies in many countries. The approaches adopted towards the establishment of data bases on internal migration may provide information that is useful in establishing data bases on international migration. Except in Europe, few agencies can rely on population registers to generate adequate and timely migration data. In most non-European countries, the census is the major source of migration data. Since censuses are organized once every ten years, the data are rapidly out of date. Some countries supplement census data with other data. The United Kingdom, for instance, supplements census data on migration with changes of address reported to the National Health Service Central Register (NHSCR) (see e.g. Ogilvy, 1980; Bulusu, 1991). The method of combining migration data from several sources has also been suggested for developing countries (e.g. Courgeau, 1982). The experience of the U.S. Bureau of the Census is particularly useful since the Bureau combines data from three different sources to produce annual data on migration by origin *and* destination. The results are used to estimate current population and to project future population. Although the U.S. experience is with internal (interstate) migration, it is very relevant for obtaining international origin-

destination-specific migration flows by integrating data from several sources. The experience is reviewed in this section.

Since 1952, the U.S. Bureau of the Census periodically projects the resident population for the 50 states and the District of Columbia. Initially the methodology was a ratio method in which the total population projected at the national level was divided into state populations by projecting assumed ratios of state to national populations (Long, 1985a). In 1957, the component method was introduced and net migration rates were used. A major advance came in 1965 when the net migration rates were replaced by gross in- and outflows. The delay in processing the migration question from the 1980 census forced the Census Bureau to return to the net migration model for the 1982 set of projections. That experience led the Bureau to rethink the procedure for using migration data to reduce the dependence on decennial data and to benefit from recent advances in projection methodology. Since 1977, the population heterogeneity with respect to migration was accounted for by treating separately population groups that are particularly prone to migration, namely, college students and members of the Armed Forces. The projections published in 1988 are the first to use a multi-regional methodology that incorporates annual state-to-state flows of migrants (Wetrogan, 1988).

In order to prepare multiregional projections, it was felt that the migration data should meet the following conditions (Long and Wetrogan, 1986, p. 2; Wetrogan, 1988, pp. 14–15):

1. Migration should be expressed as a destination-specific, out-migration rate (i.e. migration rates by origin *and* destination).
2. The destination-specific migration flows should be disaggregated by the major demographic dimensions age, sex and race.
3. The migration data must be available for a one-year migration interval.
4. The data should be available on an annual basis for a substantial number of time periods and should be easy to update.
5. The migration data must be consistent with the population base of the projection which is a census-level estimate.

No single data source meets the above requirements (see Isserman et al., 1982 and Kasarda et al., 1988 for a review of migration statistics in the United States). It was therefore decided to combine migration data from three different sources: the decennial census, national surveys and administrative data sets (Long, 1985b). Each data set provides partial information and contains unique characteristics. The decennial census provides detailed

information on migrants by age, sex, race, place of residence at the time of the census and place of residence five years prior to the census. The migration question (residence five years ago) is coded for a sample of the enumerated population (approximately 10 percent in the 1980 census). The information becomes available a few years after the census, hence the information is often out-of-date. To obtain up-to-date information on migration, the Census Bureau relies on the annual Current Population Survey (CPS) and on administrative records. The March supplement of the CPS asks 57,000 households for the place of residence from one to five years earlier for each member of the household. The migration information is tabulated for each age, sex and race. Since the sample size is limited, geographic information is limited to four census regions with no data on state-to-state flows. In effect, the CPS-based data provide full detail on demographic dimensions but only aggregate information on interstate migration. Geographic detail on interstate migration has become available from administrative sources. To prepare for the 1986-base projections, the Census Bureau obtained individual tax returns from the Internal Revenue Service (IRS) and matched the addresses on tax returns for consecutive years to identify changes in address during a one-year period. This way, it generated a time series of annual state-to-state migration rates for the years 1975–76 onwards. The administrative data have their limitations, however. They are limited primarily to wage earners who file tax returns in two consecutive years. The process of matching tax returns is a complex procedure that may add some errors of its own. A complete discussion of the development of migration data from the tax return data is given in Current Population Reports, Series P-25, No. 957.

The three data sources are used to generate a 'synthetic' data base on interstate migration (Long and Wetrogan, 1986; Willekens et al., 1986). The 1980 Census provides the initial information on migration by state of origin, state of destination, age, sex and race. The information is updated to intercensal years using the matched tax return data and CPS data.

The method used for inferring the annual flow data from the three data sources is based on the multiproportional adjustment method (Willekens et al., 1979). The method belongs to the group of methods described in Section 4 of this paper. Suppose that to meet the needs for projection and monitoring, the required data consist of migrations by age, sex, race, state of origin and state of destination. The data may be arranged in a five-dimensional table and may be fully described by a log-linear model, including five main effect parameters and several interaction effect parameters, up to five-way inter-

action. The parameters of the model need to be estimated from the available data. The preliminary estimates of migration are obtained from the census. The census provides the necessary detailed information, but the information relates to migration over a five-year interval and is out-of-date. Using CPS data, the numbers of migrants, measured over the five-year interval is transformed into numbers of migrants over a one-year interval, required for projection. To update the census data, tax returns are used to generate annual interstate migration flows. These flows are not disaggregated by age, sex or race. The CPS data provide annual information on the age, sex and race composition of migration without geographical detail. The estimation method derives the main effects of age, sex and race and the interactions between age and sex, age and race, sex and race and the three-way interaction age-sex-race from the CPS data. The main effects of origin and destination and the interaction between origin and destination are derived from the IRS data, while the values of all other interaction effect parameters that cannot be estimated from the CPS or IRS are derived from the census (for a description, see Willekens et al., 1986). In other words, the census provides the information on the association between the cross-classified variables that cannot be derived from the recent data sources (IRS and CPS).

6. Combining data from different sources: The European challenge

Many efforts are being made in Europe to establish a coherent and consistent data base on international migration. In November 1992, EUROSTAT (1992a) reviewed the relevant work of a number of international organizations. Most of the efforts currently are devoted to

a. the inventory of migration statistics,
b. the review of concepts, definitions and measurement practices with the objective to improve the comparability of migration statistics and to harmonize migration statistics, and
c. the assembly of migration data collected by countries into an international data base.

The development of methods to estimate international migration flows indirectly or to combine migration data from different sources is lacking almost completely.

6.1. *Inventory*

The main sources of migration data are censuses, population registers and surveys, particularly labour force surveys. An extensive review of sources in European countries is given by Kuijsten (1990), Poulain (1991) and Poulain and Gisser (1993). A listing of the sources of official migration statistics in the EC and EFTA countries is provided by EUROSTAT (1993). Most countries in Europe keep population registers. In countries that keep registers, identification of international immigration or emigration is linked to recording in or deletion from the population registers. The identification criteria for immigration and emigration differ between countries. In most countries, the criteria involve some intended duration of stay either inside the country or abroad. Immigration by foreign nationals requiring a residence permit is identified by this residence permit.

Of the nineteen countries of the European Economic Area (EC and EFTA countries), five do not have a register. In the United Kingdom, the International Passenger Survey is the main source of international migration data.[12] In Ireland, statistics are compiled on the basis of the Labour Force Survey, which records information from respondents on the place of residence one year prior to the survey and from relatives on individuals residing abroad at the time of the survey while residing in Ireland one year ago. In France, the only information on immigration is the medical examination of foreign nationals at the Office of International Migration. The medical examination is compulsory only for those age 18 years and over, residing in France for more than three months and holding a residence permit. In Portugal and Greece, no statistics are compiled, since the counting of emigrants was abandoned in 1977 for Greece and in 1989 for Portugal. In these countries, residence permits are the only source of (im)migration data.

Korcelli (1993) provides a detailed review of data in Poland on international migration. He concludes that emigration was heavily underenumerated in the official population statistics in the 1970s. The data became "totally irrelevant" during the 1980s, when several hundred thousand Polish citizens emigrated unofficially to the West (Korcelli, 1993, p. 187).

Two categories of international migrants are not included in population registers: asylum seekers and illegal migrants. Asylum seekers are recorded in most countries but are considered as immigrants at different times in the course of their residence application procedure, depending on the country: either on arrival as in certain Belgian local authority areas or after a certain

period, which is one year in Switzerland and variable in Norway, or on final approval of refugee status as in Sweden (Poulain and Gisser, 1993, p. 7). Van der Erf (1993) recently completed an inventory of data available in the EC countries on asylum seekers and refugees and studied the relationship between statistics on asylum seekers/refugees and official statistics on international migration. He is currently studying the data for the EFTA countries.

Statistical offices in Europe obtain international migration statistics from a variety of data sources. Some countries with population register, such as Italy, consider residence permits better suited to measure long-term immigration (ECE, 1993, p. 4). Apart from Austria, all EFTA countries gather their data from population registers (Poulain and Gisser, 1993, p. 4). Austria scheduled for mid-1993 (postponed) to begin using its local population registers to produce migration statistics. Several countries have projects aimed at improving the migration statistics. The Netherlands Central Bureau of Statistics is considering preparing annual immigration stock statistics by immigration cohort based on statistical matching of data from the municipal population registers and/or by the introduction of unique personal identification numbers (De Beer et al., 1993, p. 13). Several other countries, particularly in Eastern Europe, are planning to improve statistical monitoring of the international migration flows. Other organizations are also contributing. For instance, the Institute of Demography of the Bulgarian Academy of Sciences proposed setting up a permanent statistical data collection system on migration.

The inventory of sources of international migration data is not completed yet. There remain sources of relevant data that are not yet identified.

6.2. *Comparability*

The issue of comparability of migration data has preoccupied statisticians and analysts for many years (for a review of earlier literature, see Willekens, 1985; for a recent discussion in the context of internal migration, see Long, 1991, 1992;[13] and for international migration, see Kelly, 1987 and Poulain, 1991).[14] In order to produce comparable data and/or to be able to harmonize migration statistics, one must first know how migration is defined and how the migration data are collected in the various countries. In other words, one must know how migration is measured (which may depend on how it is defined). That not only involves issues of definition, but also of coverage, degree of nonresponse, accuracy of response and processing of raw material

into count data. The comparability of migration data is complicated by the terminology used. In the Netherlands, for instance, immigrants are seldom referred to as foreigners or foreign citizens, as is the case, for instance, in Germany, Switzerland and Austria. In the Netherlands, most people of immigrant origin are referred to as ethnic minorities, in line with the Anglo-American use of the term (Entzinger, 1993, p. 103). Poulain (1991, pp. 125–126) lists eight questions to be answered in order to test the comparability. The data compilation process generally leads to data that are not directly comparable. The types of data that are collected may, however, be organized into a typology (Willekens, 1985; Courgeau, 1988, Part I). A further requirement to improve the comparability of migration statistics is that methods are developed to transform one type of migration data into another type of data (Courgeau, 1988, Part II). These methods are generally based on the premise that the observations (migration data) are manifestations of an underlying stochastic process, which is known and the parameters of which can be estimated from the data (see e.g. Courgeau, 1982; Willekens, 1993).

Much of the work currently underway in Europe on the comparability of international migration statistics deals with definitions, measurement procedures and comparison of numbers (Poulain, 1991; Poulain and Gisser, 1993). In principle, a migration across national boundaries is recorded twice: in the country of departure at the time of emigration and in the country of destination at the time of immigration. On the basis of immigration and emigration data, Poulain and Gisser constructed a table of migration flows between EEC and EFTA countries for the year 1988 and found that in only one out of four flows, the difference between the two recordings is less than 10 percent. For more than a quarter the difference is between 10 and 100 percent. In more than half the cases, the difference exceeds 100 percent (Poulain and Gisser, 1993, p. 13). Countries with population registers produce better statistics than other countries. Further, the pairs of data for Nordic countries are much better than the other countries. This finding may be attributed to the fact that these countries coordinate their migration data; inter-Nordic migrations are recorded on a special form common to the five Nordic countries (Inter-Nordic Migration Certificate).[15] The comparability of the immigration and emigration statistics is also studied by Wils (1993). She observed a tendency to overestimate immigration and to underestimate emigration (Wils, 1993, p. 24). The reasons for overestimation are (i) migrant categories overlap and (ii) immigrants are often defined according to shorter periods of stay than emigrants for leaving. The underestimation of emigra-

tion may be attributed to (i) departing migrants having lower motivation to deregister and (ii) rejections of applications for asylum being less publicized than the applications.

6.3. *Migration matrices*

The compilation of migration flow data from various countries into origin-destination migration matrices receives considerable interest from international organizations (for a detailed account, see Wils, 1993). The Statistical Division of the Economic Commission for Europe (ECE) has collated these matrices since 1975. Statistical offices and other official institutions of the ECE countries are sent a questionnaire, requesting information on migration flows from and to all countries of the world, and on the definition of migration. The data are later adjusted to comply with the UN definition of long-term migrant.[16] The matrices are published annually. Examples of such matrices are given in United Nations (1977), Kelly (1987), Poulain and Gisser (1993) and Wils (1993).

EUROSTAT in Luxenbourg has also been collecting migration data since 1989 in a way similar to that of the ECE, by sending questionnaires to the statistical offices of member countries (EC and EFTA countries). In 1991, the EC countries agreed to transmit to EUROSTAT data on migration statistics from 1985 onwards. A similar agreement was reached with EFTA countries in 1992. The migration data collected on stocks and flows are being stored in the ACUMEN databank in the EUROSTAT office in Luxembourg;[17] the sub-sector will be known as "MIGRAT". EUROSTAT and ECE agreed and, during the Joint ECE/EUROSTAT Work Session on Migration Statistics in Geneva in February 1993, developed plans to merge their questionnaires and to form a common one which will serve both institutions.

The flow statistics complement the statistics on stocks of migrants.[18] OECD in Paris publishes annually stocks of foreign population, inflows and outflows, as part of the Continuous Reporting System on Migration (SOPEMI). The OECD publication is based on country reports by SOPEMI correspondents. Finally, EUROSTAT is collecting stock data on population by nationality based on the 1991 censuses. The statistics will be stored in the EUROCENSUS part of the ACUMEN databank. For more information on the EUROSTAT data collection and publication program, see EUROSTAT (1992b).

Finally, the Population Activities Unit of the ECE has set up a Rapid Information System to collect and disseminate international migration data,

soon after they are released by national statistical offices. The data are collected by the ECE Population Activities Unit in collaboration with the ECE Statistical Division, by the Office of the United Nations High Commissioner for Refugees (UNHCR) and by the secretariat of the Intergovernmental Consultations (IGC). They are contained in the Bulletin "International Migration in the ECE Region". The Bulletin is published twice a year, in November and May, and covers events in the first and the second half of the year, respectively. The first issue was published in November 1992.

The foregoing overview shows that several organizations responded to the challenge that resulted from an increasing political interest in international migration and the lack of information about the size and the nature of migration flows. The types of data available from national statistical offices and other official data producers are now widely publicized and the data that exist are relatively rapidly disseminated. In addition, the data are integrated into data bases for comparison and easy retrieval later. But the gap between available statistical data and information required for policy-making and research remains large. The challenge to produce the information that is required to monitor international migration flows in Europe remains. No reliable information source should be omitted and the best techniques should be used to combine data from different sources. There exists an information source that has been left out in almost all attempts to produce adequate information on migrations. It is the large group of experts who have been observing and analyzing migration flows for years. It is the people who make the data or interpret the data. It is also the people who work with migrants either directly, such as UNHCR or indirectly, such as the persons preparing the national policies on international migration. The information that may be provided by these persons may be a very valuable addition to the generally available statistical data that are produced by the various data collection schemes. The information may complement the statistical data, if it sheds light on the data collection process and on the measurement problems that arise. The information may also have a value in its own right. The knowledge and intuition that experts have, particularly the ability to recognize patterns and to signal changes that are otherwise difficult to register, is valuable information and may, in some cases, give a more accurate picture of reality than the direct measurements. The expert opinions are of course subjective. The data are therefore generally referred to as judgmental, although interpretative data may be a more accurate qualification. The use of expert judgments in estimating migration flows is briefly discussed in

the next section. It is the subject of an entire research project (Wils and Willekens, 1993).

7. Adding judgmental data

The establishment of a data base by combining data from several sources implies the use of prior information on the association between the cross-classified variables. In the above discussion, the prior information is derived from other statistical data, generally from previous periods (historical data). The interaction effects that are not contained in the recent data source are 'borrowed' from the more detailed, but outdated source, which is generally the census. The prior information about the association between the variables is imbedded in the census data. 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 (see e.g. Albert and Gupta, 1983).

The prior beliefs on the degrees of association between the variables are not restricted to information contained in auxiliary cross-classifications, such as is the case with census data. The prior information may also be derived from established theory or from judgments by common individuals or experts. Judgmental knowledge is domain knowledge, i.e. knowledge about the processes that are being represented (measured) by the statistical data. Among experts and other people, there exists useful and usable knowledge that is never used because it is too imprecise or fragmentary to incorporate into a formal data base or a formal model.[19] The acquisition, representation and utilization of domain knowledge has been receiving attention since the development of expert systems.

The most experience to date with the quantification and utilization of expert knowledge is in the field of risk analysis and artificial intelligence (see e.g. Cooke, 1991). In forecasting, judgmental knowledge is used regularly in combination with statistical information to improve the forecasting performance. The experience of forecasters may provide some guidance on the procedure and relevance of combining statistical information and domain knowledge. Therefore, the experience is briefly reviewed in this paper. A review of the early literature on the combination of statistical and judgmental forecasting is given by Bunn and Wright (1991). Wolfe and Flores (1990)

found that judgmental adjustment of quantitative forecasts improved forecast accuracy but not always significantly (see also Flores et al., 1992, for further results). McNees (1990) found a clear tendency for modellers to overadjust their models, although the forecasting performance is better with adjustments than without. The prevailing tendency to place too much weight on the specific circumstances and too little on the model is what Kahneman and Tversky (1982, p. 416) have called "the major error of intuitive prediction." Fildes (1991) considers the use of information by a panel of expert industry forecasters, focusing on their information-processing biases. He concludes that the panel overweighed recently released information and do not understand the dynamics of the industry. Instead, the panel apparently used a simple heuristic for updating the forecast it had previously made of that year's growth. However, their forecasts are better than an alternative econometric model. Blattberg and Hoch (1990) focus on ways to combine models and experts, with the example of simple data base models and managerial intuition. They found that a combination of model and manager always outperforms either of these decision inputs in isolation. "Managers have additional insight that the model cannot incorporate such as the state of economy, fashion trends, idiosyncratic features of an item, and shifting coupon redemption patterns . . . The trick is to incorporate model consistency and managerial insight into one forecast" (Blattberg and Hoch, 1990, p. 898).

The use of expert judgment and subjective expectations are also considered in demographic forecasting. Several countries, among them Australia, the United Kingdom and the Netherlands, attempt to improve forecasting performance by combining quantitative forecasts with information on birth expectations collected in fertility surveys. This author is not aware of any country that uses expectations in migration forecasting, although some countries collect information on migration intentions as part of housing surveys (e.g. the Netherlands) or in special migration surveys. Expert opinions are also used in demographic forecasting. Some countries, among them the Netherlands, have a working group of 'experts' to discuss the predictions of the demographic parameters initially prepared by the statistical organization producing the forecast. In preparing future demographic trends in Europe and North America, Lutz et al. (1991) base their assumptions on future fertility, mortality and migration on alternative suggestions made by several expert demographers in the various chapters of the book. Alho (1992) develops a formal statistical model within which it is possible to define

and estimate the strength of expert opinion in forecasting. The model is based on the mixed-estimation approach in linear regression. He finds that too much weight is put on expert opinion overall.

In their review of the forecasting literature, Bunn and Wright (1991) conclude that judgmental knowledge and quantiative knowledge should be integrated. The *Journal of Forecasting* recently devoted two special issues (Vol. 9, No. 4, 1990 and Vol. 12, No. 2, 1993) to judgments in forecasting. The second special issue was motivated by the need to understand more fully the role judgment plays in forecasting: "It reflects attempts, especially with expert systems, to capture and use more of what people know that has predictive power although it may be too imprecise or fragmentary to incorporate into a formal model" (Ferrell, 1993). The important question is not whether human judgment is always 'good' or always 'bad'. The important questions are: "Under what circumstances is judgment most likely to incorporate information with predictive value above and beyond that which has already been incorporated in a formal model? How can the imperfect information from judgment be combined with the imperfect information in models to maximize predictive accuracy?" (McNees, 1990, p. 297). These questions are relevant not only to forecasting, but equally to the more general problem of incomplete data.

Expert judgments are being used for the forecasting of internal and international migration. Harker (1986) combined subjective judgments 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 judgmental and statistical data is based on Saaty's analytic hierarchy process (AHP). Cook et al. (1984) use the same 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.[20] George and Perreault (1992, p. 93) 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 (p. 95). In 1991, the Institute of Employment Studies in Moscow conducted a delphi study among 30 experts from government, science and business, with the aim of getting their estimates of the prospects of emigration from the territory of the former USSR during the period 1992–97. Half of the experts expected about 2–4 million emigrants, 30 percent expected 4–5 million emigrants, and 20 percent expected the number of emigrants not to exceed 2 million (Tikhonov, quoted by Vishnevsky and Zayonchkovskaya, 1993, p. 268). The uncertain opinion

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, p. 3). Particularly, the estimation and forecasting of international migration may benefit from new techniques developed in the context of judgmental forecasting, since international migration depends on many factors that are difficult to incorporate into a formal model: "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" (Van de Kaa, 1993, pp. 87–88).

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, p. 80). It turns out that the use of judgmental data in the development of accounts and in the modelling of migration is not at all a new idea. In the estimation and updating of social accounts, the Bayesian interpretation of the estimation problem has been used to incorporate subjective insights from experts (Van der Ploeg, 1984). Knudsen (1992) discusses the method of how to include *a priori* information in the context of generalized linear models. He suggests including *a priori* information into the model by treating this information as a covariate having a known parameter value of one. Note that this approach is similar to the method proposed by Laird and Olivier (1981), referred to in Section 4.

The use of judgmental data in the monitoring of international migration reflects the attempt to capture and use more of what people know although it may be too imprecise or fragmentary to incorporate into a formal data base and a formal model of migration. Many problems need to be solved before judgmental data can be incorporated into a data base and used for prediction. First, we must identify what knowledge experts have that is not contained in the available statistical data. Second, the biases in expert knowledge must be identified and removed and the expert's uncertainty must be adequately represented.[21] Third, the knowledge must be formalized, i.e. represented in a way that can be included in a data base and used in a model. Finally, the knowledge must be combined with statistical data in order to produce a coherent, consistent and accurate picture of migration that can be used in policy-making and research. With respect to the first issue, two

examples may be given of where expert knowledge could usefully be combined with quantitative data. For instance, Edmundson (1990) found expert (domain) knowledge useful because it helped to determine an appropriate decomposition of the variables and data. Armstrong and Collopy (1993) use information about causality as inputs to improve the forecasting performance of models. The latter finding supports the claim of Keyfitz, made more than twenty years ago on demographic forecasting: "The weakness of population forecast is due to our ignorance of the mechanisms by which populations grow and decline. We know much about birth rates and their differentials among statistically recognizable population subgroups, as well as about changes over time as shown in past records, but this great volume of statistical information has contributed disappointedly little to the discernment of the comprehensive causal system underlying the differentials and changes" (Keyfitz, 1972, p. 361).

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 on *how* to use expert knowledge properly remains largely open. The rules for collecting and processing 'objective' data have been developed over a great many years. Rules for the collection and processing of 'subjective data' in the form of expert knowledge and opinion, and the integration of subjective and 'objective', statistical data are much less developed.

8. Conclusion

Several institutions in Europe provide information on and analysis of international migration. The OECD provides highly informative annual reports, based on country reports from correspondents. The Statistical Division of the United Nations' Economic Commission for Europe (ECE) prepares matrices of annual international migration flows, based on data provided by the principal statistical institutions of ECE countries. The ECE's Population Activities Unit set up a Rapid Information System to disseminate up-to-date information on international migration. In the bulletin *International Migration in the ECE Region*, produced twice a year, data are published that are collected by the ECE Statistical Division, by the Office of the UNHCR and by the secretariat of the Inter-Governmental Consultations (IGC). The Council of Europe annually publishes *Recent Demographic Developments in Europe*, which includes information on international migration. Finally,

the Statistical Office of the European Communities (EUROSTAT) is integrating data provided by national statistical offices into a data base on international migration; the data include a migration matrix.

The compilation and dissemination of existing data serve an important function. They provide facts about the level and direction of international migration. The facts are desperately needed because migration is politically sensitive and the discussions tend to be emotional, partly because the facts are hard to come by. But the available statistical data do not provide an accurate picture of migration flows. Neither the volume nor the direction of migration flows in Europe are properly measured. What is needed is a coherent and consistent data base on international migration that contains sufficiently detailed, up-to-date and accurate information. Such a data base can never be created solely on the basis of data collected in the countries of Europe and provided by the principal statistical offices of the countries.

The main message of this paper is that, in order to compile coherent and internally consistent information on migration, data from several sources ought to be combined. The paper reviews techniques for integrating statistical data from different origins. The demographic account, combined with the modelling of the data in the account, provides an appropriate framework for that activity. When all the available data are entered into the account, after correction for differences in definition, several cells may remain empty and/or the reliability of some of the data may be inadequate. Estimation methods that are reviewed focus on *data structures*. The data are arranged in a multi-dimensional table and the structure is represented by the patterns of association between the cross-classified variables. A particular element in the data set (e.g. the number of migrants of a given age and sex between two countries or regions) is viewed in relation to the other elements in the set. It is part of a structure linking the individual data. The task of constructing a data base by combining data from several sources consists of the identification of the strengths and weaknesses of each data set, the representation of the structure exhibited by the data set, and the linkage of data structures in such a way that the strengths of each data set are preserved.

The monitoring of international migration in Europe calls not only for the combination of statistical data from several sources, but also for the integration of 'objective' statistical data and 'subjective' expert knowledge and opinions on the patterns of migration and the causal mechanisms. The integration raises several issues. This paper discusses the issues, without being exhaustive. For some issues, possible solutions are presented, based

on the experience of scholars in the field of migration and related disciplines and of professionals in statistical organizations.

The main conclusions are:

1. The distinction between internal and international migration becomes increasingly useless. There are no fundamental conceptual or analytical differences between internal and international migration systems. The difference is at the level of policy relevance and policy-making. The research findings in the field of internal migration may be used in the study of international migration flows.
2. The establishment of a coherent and consistent data base on international migration, which combines data from several sources, is analogous to the development of a demographic account. The account may be represented as a multidimensional table. The application of the basic accounting principles assures that the relations between the data in the data base are correctly specified.
3. Before a migration data base can be established, several conceptual and measurement issues must be resolved in order to ensure that the data measure what they are supposed to measure and that the types of data are explicitly specified.
4. In order to estimate the data that are missing from the account (data base), a data model needs to be developed that describes each number of the data base in relation to the other data. Once the data model is specified, it may be used to infer (impute) missing data or update data while preserving the data structure and/or other constraints imposed upon the data.
5. Many data models have been developed in the literature. They may be viewed as particular formulations or variants of the Generalized Linear Model (GLM), which is a family of regression models. The Poisson regression or log-linear model is appropriate to describe migration flow data. The parameters of the model denote the main and interaction effects of the cross-classified variables.
6. In combining migration data from several sources, using the log-linear model, the parameters of the model denote the contribution of each source to the final migration estimates. In other words, the model describes *how* the data originating from the various sources are integrated.
7. The modelling approach to combing data from different sources has been adopted by the U.S. Bureau of the Census. The U.S. experience is useful for Europe.

8. Many efforts are being made in Europe to establish a coherent and consistent data base on international migration. Most of the effort is devoted to the inventory of migration statistics, the evaluation of the comparability of the data, the harmonization of the data and the assembly of migration data into an international data base.
9. A migration data base that is oriented toward the monitoring of migration flows in Europe should include judgmental data, derived from expert knowledge and opinions. The experience with judgmental data is limited still in the field of migration. The lessons learned in risk analysis, forecasting, and artificial intelligence should be considered in integrating 'objective' data and 'subjective' data.

Notes

[1] Public opinion is to some extent reflected in the popular press. Wilmoth and Ball (1992) have undertaken an extensive review of American popular literature on population. They found that, although rapid population growth in less developed countries is linked to the issue of US immigration, the popular literature fails to specify why immigration itself is considered problematic (p. 653). The authors also found that professional demographers have not done much to move the debate forward. I do not know of a similar study in Europe.

[2] The increased awareness was a reason for the World Bank and the Economic Commission for Europe to review ongoing international migration research, research-related activities, and data sources for research and analysis. The first results are reported in Russell (1993).

[3] The need for better monitoring of international migration flows not only exists in Europe. The need is also expressed in the United States (see e.g. Teitelbaum, 1992, pp. 63 and 75).

[4] A collection of solutions to incomplete data problems in demography is United Nations (1983).

[5] A similar observation was made several years ago by Bretz et al. (1987) who investigated the possible causes of the differences between migration statistics on the level of migration between Italy and the Federal Republic of Germany that are compiled by Italy (ISTAT) and the FRG (StBA). The migration flows to and from Italy, as shown by the FRG, frequently were more than twice as high as those recorded by Italy. Part of the overenumeration in the FRG was due to the fact that registration/deregistration of a residence was recorded irrespective of the intended duration of stay.

[6] The anomalies in international migration statistics are not limited to Europe. Carmichael (1992) reports on idiosyncrasies of the official national databases of Australia and New Zealand, most of which became apparent as a result of trying to account for marked differences in the estimates of both migratory and short-term population movement between the two countries, and for variation over time in the patterns of differences. It is shown that the use of apparently similar broad classification principles can create an impression of comparability and continuity that may be quite misleading.

[7] Korcelli (1993) compared emigration from Poland to the Federal Republic of Germany using official German sources and data from the Polish Central Statistical Office. He found that "with regards to the 1954–1980 period, the figures listed in the German sources are roughly comparable with those published by the Polish Central Statistical Office."

[8] There is no systematic evaluation of the possible contribution of internal migration research to the study of international migration. A number of scholars suggest that the distinction between internal and international migration is no longer relevant (see e.g. Appleyard, 1991, p. 47). The systems

approach to the study of international migration, suggested by Kritz and Zlotnik (1992) was originally developed to improve the understanding of rural-urban migration. Pryor (1981) discusses the integration of international and internal migration theories. In the European context, the distinction between internal and international migration becomes flawed since, with the disappearance of frontiers between countries, international migration becomes internal migration within a unified Europe (Extercate, 1993, p. 4).

[9] For instance, Dhima (1991, quoted by Straubhaar and Fischer, 1993, p. 146) found that 66 percent of all long-term immigrants in Switzerland entered the country as seasonal migrants. Of the persons that are currently in Switzerland as seasonal migrants, 47 percent started a 'career' to obtain a permanent residence permit. Malacic (1993, p. 224) quotes Mulina et al., who, at the beginning of the 1980s, estimated that about half of the migrant workers from Yugoslavia could permanently stay in the receiving countries.

[10] In the theory of inference, a distinction is made between the *prediction* of the unknown actual outcome of a random variable and the *estimation* of the unknown expected outcome of this random variable (see e.g. De Jong, 1985, p. 3). In this context, prediction should not be confused with forecasting. In this paper, prediction and estimation are used interchangeably.

[11] An introduction to the statistical concept of entropy, as applied in the estimation of migration flows, is given by Willekens (1977). The historical development of the concept of entropy is reviewed and its relation to statistics is explained by Akaike (1985).

[12] The IPS is a survey of travellers at points of entry into the United Kingdom. It has been operational since 1964 and gathers information on both British and foreign travellers. For some details, see Coleman (1993). However, persons moving between the U.K. and Ireland are beyond the scope of the survey. Consequently, the flow of Irish citizens to the U.K., which has traditionally been important, is not reflected in the migration statistics (Zlotnik and Hovy, 1990, p. 2).

[13] Long compares residential mobility in 16 developed countries. He concludes that "annual rates of residential mobility in 12 of these countries ranged from 6% to over 19% in 1980–81 without any firm basis to explain or account for this variation."

[14] A major effort to produce comparative statistics on international migration was undertaken by the U.S. Bureau of the Census and Statistics Canada. The effort involved the development of matched tabulations of demographic, economic, and social characteristics of persons born in Canada who were counted in the U.S. census of 1980 and persons born in the United States who were counted in the Canadian census of 1981. The data base is used to assess the long-term consequences of migration (Long and Pryor, 1988).

[15] Although the data of the sending and receiving countries are comparable, the harmonization of registration practices among the Nordic countries introduced some anomalies of its own. For instance, Swedish statistics include among emigrants persons going to Denmark for an intended stay of three months or more and persons going to Norway for a period of six months or more (Zlotnik and Hovy, 1990, p. 2).

[16] The U.N. defines a long-term immigrant as a person who has never been in the country or has been away for at least one year and intends to stay for more than twelve months before departing (United Nations, 1970, pp. 5–7). Significant differences may emerge when the immigration data are revised to reflect the U.N. concept of long-term immigration. A comparison of U.S. statistics from the Immigration and Naturalization Service with estimates based on the U.N. definition revealed differences according to country of origin, age, and state of intended residence (Kraly and Warren, 1992).

[17] ACUMEN is a software package which permits the storage of data sets in multi-dimensional tables.

[18] Data on number of residents with a foreign origin must be interpreted with caution, since the definition may differ substantially. For instance, in an estimate of the size of the foreign population living in France, Tribalat (1991) includes those French nationals who have at least one parent or grandparent of foreign birth. De Beer et al. (1993) found that 15% of the Dutch population is non-native. A person is non-native if he or she is born abroad (8%) or born in the Netherlands but

with at least one parent born abroad (7%). Earlier, Entzinger (1993, p. 104) came to an estimate of 18%.

[19] Experts may also provide information on less-known sources of data about the process being studied.

[20] The analytic hierarchy process is essentially a method of measurement. The AHP is a method by which subjective weights are assigned to a set of objects or alternatives. The objects may be different models used to forecast migration. The AHP may be used by the public manager to attach weights ('trust') to the models in order to make public decisions. "It gives the public manager a tool by which he or she can combine the results of several different models in order through the use of expert judgments to arrive at a composite prediction" (Harker, 1986, p. 69). The method may also be used to add expert opinion and to summarize expert judgments, which may be necessary to predict turning points. The AHP assumes that a person attaches weights by pairwise comparison of the objects of alternatives using a ratio scale of relative magnitudes (as opposed to the interval scale in utility theory) (Saaty, 1990). Recently, the AHP has come under attack. Dyer (1990a) claims that the AHP is flawed as a procedure for ranking alternatives in that the rankings produced by this procedure are arbitrary, since the weights on a higher level of hierarchy can be determined independently of the weights on the lower levels. Dyer's paper is followed by a discussion (Saaty, 1990; Harker and Vargas, 1990; Dyer, 1990b). For applications of the AHP, see Golden et al. (1989).

[21] The reasoning behind judgements is very important, McNees points to the issue here (in the context of forecasting): "While it would be a mistake to ask forecasters to refrain from adjusting their models, it is also a mistake to accept the adjustments that are made at face value, especially when the adjustments appear without any explanation of the reasoning behind them. On occasion, forecasters may have objectives other than simply maximizing the accuracy of their forecasts" (McNees, 1990, p. 298).

References

Akaike, H., 1985. 'Prediction and entropy', in A. C. Atkinson and S. E. Fienberg (eds), A Celebration of Statistics, 1–23. The ISI Centenary Volume. Springer Verlag, New York.

Albert, J. H. and A. K. Gupta, 1983. 'Estimation in contingency tables using prior information'. Journal of the Royal Statistical Society B 45(1): 60–69.

Alho, J. M., 1992. 'Estimating the strength of expert judgment: The case of US mortality forecasts'. Journal of Forecasting 111(2): 157–167.

Alonso, W., 1986. 'Systemic and log-linear models: From here to there, then to now, and this to that'. Discussion Paper 86–10. Center for Population Studies, Harvard University, Cambridge, Mass.

Appleyard, R. T., 1991. International Migration: Challenge for the Nineties. International Organization for Migration (IOM), Geneva.

Armstrong, J. S. and Collopy, F., 1993. 'Causal forces: Structuring knowledge for time-series extrapolation'. Journal of Forecasting 12(2): 103–115.

Bacharach, M., 1970. Biproportional Matrices and Input-Output Analysis. Cambridge University Press, London.

Bates, J., 1982. 'Developing the migration component of the official sub-national population projections'. Final report prepared by Martin Voorhees Associates for DPRP3 Division of the Department of the Environment, London.

Batty, M. and MacKie, S., 1972. 'The calibration of gravity, entropy and related models of spatial interaction'. Environment and Planning 4: 205–233.

Batty, M. and March, L., 1976. 'The method of residues in urban modelling'. Environment and Planning A 8: 189–215.

Bishop, Y. M. M., Fienberg, S. E. and Holland P. W., 1975. Discrete Multivariate Analysis: Theory and Practice. M.I.T. Press, Cambridge, Mass.

Blattberg, R. C. and Hoch, S. J. 1990. 'Database models and managerial intuition: 50% model + 50% manager'. Management Science 38(8): 887–899.

Bos, W., Leunis, W. P., Sprangers, A. H., Veenstra, C. J. and Verhage, C. G., 1988. 'Towards a system of labour accounts'. Netherlands Official Statistics 3(4): 5–30.

Bretz, M., Esposito, I. and Fleischer, H., 1987. 'The precision of statistics on international migrations. A study of flows between Italy and the Federal Republic of Germany'. Statistical Journal of the United Nations ECE 5(1): 1–12.

Bulusu, L., 1991. 'A review of migration data sources'. OPCS Occasional Paper No. 39. Office of Population Censuses and Surveys, London. (See also Population Trends, No. 66, 1991).

Bunn, D. and Wright G., 1991. 'Interaction of judgmental and statistical forecasting methods: Issues and analysis'. Management Science 37(5): 501–518.

Carmichael, G. A., 1992. 'Beware the passenger card! Australian and New Zealand data on population movement between the two countries'. Working Paper in Demography, No. 3. Research School of Social Sciences, Australian National University, Canberra.

Castro, L., 1985. 'Analysis of Age-Specific Gross and Net Migration Schedules'. Paper presented at the Annual Meeting of the Population Association of America, Boston, March.

Chilton, R. and Poet, R., 1973. 'An entropy maximizing approach to the recovery of detailed migration patterns from aggregate census data'. Environment and Planning 5: 135–146.

Coleman, D., 1993. 'The United Kingdom and international migration: A changing balance', in H. Fassmann and R. Münz (eds), European Migration Since 1945. Historical Patterns, Actual Trends and Social Implications. Springer Verlag, Berlin. Forthcoming.

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, R. M., 1991. Experts in Uncertainty. Opinion and Subjective Probability in Science. Oxford University Press, Oxford.

Courgeau, D., 1973. 'Migrants et migrations'. Population 28(1): 95–129.

Courgeau, D., 1980. Analyse Quantitative des Migrations Humaines. Masson, Paris.

Courgeau, D., 1982. 'Methods of linking migration statistics collected from national sample surveys with those from population censuses', in ESCAP, National Migration Surveys Series, Manual X: Guidelines for Analysis, 312–330. United Nations, New York.

Courgeau, D., 1988. Méthodes de Mesure de la Mobilité Spatiale. Editions de l'Institut National d'Etudes Démographiques, Paris.

De Beer, J., Kuijper, H., Noordam, R., Prins, K. and Sprangers, A., 1993. 'The Linking of Immigrant Flow and Stock Data in the Netherlands: Present and Future Possibilities'. Paper presented at the Joint ECE/EUROSTAT Work Session on Migration Statistics (Working Paper No. 8), Geneva, 22–24 February 1993.

De Jong, P. M., 1985. Prediction Intervals for Missing Figures in Migration Tables. Ph.D. Dissertation. University of Groningen, The Netherlands.

Deming, W. E. and Stephan, F. E., 1940. 'On a least squares adjustment of a sampled frequency table when the expected marginal totals are known'. Annals of Mathematical Statistics 11: 427–444.

Dempster, A. P., Laird, N. M. and Rubin D. B., 1977. 'Maximum likelihood from incomplete data via the EM algorithm'. Journal of the Royal Statistical Society B 39(1): 1–38.

Drewe, P. and Willekens, F., 1980. 'Maximum likelihood estimation of age specific migration flows in the Netherlands'. Delft Progress Report 5: 92–111.

Dyer, J. S., 1990a. 'Remarks on the analytic hierarchy process'. Management Science 36(3): 249–258.

Dyer, J. S., 1990b. 'A clarification of "Remarks on the analytic hierarchy process"'. Management Science 36(3): 274–275.

ECE, 1993. Report on the ECE-EUROSTAT Joint Work Session on Migration Statistics, Geneva, 22–24

February 1993. Work Session on Migration Statistics. Statistical Commission and ECE, Conference of European Statisticians, 41st Plenary Session, Geneve 14–18 June 1993. CES/779, 26 February 1993.

Edmundson R. H., 1990. 'Decomposition: A strategy for judgmental forecasting'. Journal of Forecasting 9(4): 305–314.

Eldridge, H. T. and Kim, Y., 1968. The Estimation of Intercensal Migration from Birth-Residence Statistics: A Study of Data for the United States, 1950 and 1960. Population Studies Center, University of Pennsylvania, Philadelphia.

Entzinger, H., 1993. Shifting paradigms: The appraisal of immigration in the Netherlands, in H. Fassmann and R. Münz (eds), European Migration Since 1945. Historical Patterns, Actual Trends and Social Implications. Springer Verlag, Berlin. Forthcoming.

EUROSTAT, 1991. Two Long-Term Population Scenarios for the European Communities. Scenarios prepared for the International Conference 'Human Resources in Europe at the Dawn of the 21st Century', Luxemburg, 27–29 November 1991.

EUROSTAT, 1992a. Recent Developments of Relevance for Migration Statistics. A General Overview. Working Document DOC E3/MIG/2a/92, Working Group on Migration, Meeting 16–17 November 1992, Luxemburg.

EUROSTAT, 1992b. EUROSTAT's Data Collection and Publication Programme. Working Document DOC E3/MIG/2a/92, Working Group on Migration, Meeting 16–17 November 1992, Luxemburg.

EUROSTAT, 1993. Official Migration Statistics Publications in the EEA Countries. EUROSTAT, Luxemburg (09.02.1993)

Evans, S., 1970. 'Some properties of trip distribution methods'. Transportation Research 4: 19–36.

Evans, S., 1971. 'The calibration of trip distribution models with exponential or similar cost functions'. Transportation Research 5: 15–38.

Extercate, M., 1993. Two Long-Term Status Quo Population Projections for the EC and EFTA Countries. Population Research Centre, University of Groningen.

Ferrell, R., 1993. 'Introduction to the special issue on structuring knowledge and aiding judgment in forecasting'. Journal of Forecasting 12(2): 79.

Fildes, R., 1991. 'Efficient use of information in the information of subjective industry forecasts'. Journal of Forecasting 10(6): 597–617.

Flores, B. E., Olson, D. L. and Wolfe, C., 1992. 'Judgmental adjustments of forecasts: A comparison of methods'. International Journal of Forecasting 7(4): 421–433.

Flowerdew, R., 1982. 'Fitting the lognormal gravity model to heteroscedastic data'. Geographical Analysis 14: 263–267.

Friedlander, D., 1961. 'A technique for estimating a contingency table, given the marginal totals and some supplementary data'. Journal of the Royal Statistical Society, Series A 124: 412–420.

George, M. V. and Perreault, J., 1992. 'Methods of external migration projections and forecasts', in N. Keilman and H. Cruijsen (eds) National Population Forecasting in Industrialized Countries: 87–103. Swets and Zeitlinger, Amsterdam.

Golden, B., Wasil E. and Harker P., Eds., 1989. The Analytic Hierarchy Process: Applications and Studies. Springer Verlag, Berlin.

Good, I. J., 1963. 'Maximum entropy for hypothesis formulation especially for multidimensional contingency tables'. Annals of Mathematical Statistics 34: 911–934.

Harker, P. T., 1986. 'The use of expert judgments in predicting interregional migration patterns: An analytic hierarchy approach'. Geographical Analysis 18(1): 62–80.

Harker, P. T. and Vargas, L. G., 1990. 'Reply to "Remarks on the analytic hierarchy process" by J. S. Dyer'. Management Science 36(3): 269–273.

Hoffmann, E., 1990. 'A labour accounting system. Reflections on main concepts and principles'. Statistical Journal of the United Nations ECE 7(4): 253–263.

Huigen, R. D., van de Stadt, H. and Zeelenberg, C., 1988. 'Socio-economic accounts'. Netherlands Official Statistics 3(2): 38.

Hyman, G., 1969. 'The calibration of trip distribution models'. Environment and Planning 1: 105–112.

Ireland, C. T. and Kullback, S., 1968. 'Contingency tables with given marginals'. Biometrika 55: 179–188.

Isard, W., 1960. Methods of Regional Analysis. An Introduction to Regional Science. M.I.T. Press, Cambridge, Mass.

Isserman, A. M., Plane, D. A. and McMillen, D. B., 1982. 'Internal migration in the United States: An evaluation of federal data'. Review of Public Data Use 10: 285–311.

Jaynes, E., 1957. 'Information theory and statistical mechanics'. Physical Review 106: 620–630.

Jörnsten, K. O., Larsson, T., Lundgren, J. T. and Migdalas A., 1990. 'An entropy model with variable target'. Environment and Planning A 22(4): 493–506.

Kahnemann, D. and Tversky, A., 1982. 'Intuitive prediction: Biases and corrective procedures', in D. Kahneman, P. Slovic and A. Tversky (eds), Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, New York.

Kalbfleisch, J. D. and Lawless, J. F., 1984. 'Least-squares estimation of transition probabilities from aggregate data'. The Canadian Journal of Statistics 12(3): 169–182.

Kasarda, J. D. et al, 1988. Migration Statistics in the United States. A Report of the Subcommittee on Migration Statistics, Committee on Population Statistics, Population Association of America (J. D. Kasarda, Chair), Population Association of America, Alexandria, Va.

Kelly, J., 1987. 'Improving the comparability of international migration statistics: The contribution of the Conference of European Statisticians from 1971 to date'. International Migration Review 21(4): 1017–1037.

Keyfitz, N., 1972. 'On future population'. Journal of the American Statistical Association 67(338): 347–363.

Knudsen, D. C., 1992. 'Generalizing Poisson regression: Including a priori information using the method of offsets'. Professional Geographer 44(2): 202–208.

Koesoebjono, S., 1985a. 'The Development of Socio-Demographic Accounts in the Netherlands'. Paper presented at the IUSSP General Conference, Florence, Italy, June 1985.

Koesoebjono, S., 1985b. 'Naar stelsels van sociaal-demografische rekeningen [Towards sytems of socio-demographic accounts]'. CBS-Select 3: 39–48.

Korcelli, P., 1993. 'Emigration from Poland after 1945', in H. Fassmann and R. Münz (eds), European Migration Since 1945. Historical Patterns, Actual Trends and Social Implications. Springer Verlag, Berlin. Forthcoming.

Kraly, E. P. and Warren, R., 1992. 'Estimates of long-term immigration to the United States: Moving U.S. statistics toward the United Nations concept'. Demography 29(4): 613–626.

Kritz, M. M. and Zlotnik, H., 1992. 'Global interactions: Migration systems, processes and policies', in M. M. Kritz, L. L. Lim and H. Zlotnik (eds), International Migration Systems. A Global Approach 1–16. Clarendon Press, Oxford.

Kuijsten, A., 1990. Statistics on International Migration Flows in the European Community Member Countries. Report to EUROSTAT, Luxemburg.

Laird, N. and Oliver, D., 1981. 'Covariance analysis of censored survival data using log-linear analysis techniques'. Journal of the American Statistical Association 76(374): 231–240.

Lang, J. B., 1992. 'Obtaining the observed information matrix for the Poisson log linear model with incomplete data'. Biometrika 79(2): 405–407.

Lawless, J. F. and McLeish, D., 1984. 'The information in aggregate data from Markov chains'. Biometrika 71(3).

Lee, T. C., Judge, G. C. and Zellner, Z., 1970. Estimating the Parameters of the Markov Model from Aggregate Time Series Data. North Holland, Amsterdam.

Little, R. J. A. and Rubin, D. B., 1987. Statistical Analysis with Incomplete Data. J. Wiley & Sons, New York.

Long, J., 1985a. History of Population Projection Methods at the U.S. Census Bureau. U.S. Bureau of the Census, Washington, D.C.

Long, J. F., 1985b. 'Combining migration data from different sources in the United States', in M. Poulain (ed), Migrations internes. Collecte des données et méthodes d'analyse (Chaire Quételet 1983) 443–453. Cabay Publications, Louvain-la-Neuve, Belgium.

Long, J. and Pryor, E. T., 1988. 'Comparative demographic effects of Canadian-U.S. immigration flows'. Statistical Journal of the United Nations ECE 5(2): 135–157.

Long, J. and Wetrogan, S. I., 1986, 'Creating Annual State-to-State Migration Flows with Demographic Detail by Merging Census, Survey and Administrative Record Tabulations'. Paper presented at the Annual Meeting of the Population Association of America, San Francisco, April.

Long, L., 1988. Migration and Residential Mobility in the United States. Russell Sage Foundation, New York.

Long, L., 1991. 'Residential mobility differences among developed countries'. International Regional Science Review 14(2): 133–147.

Long, L., 1992. 'Changing residence: Comparative perspectives on its relationship to age, sex and marital status'. Population Studies 46: 141–158.

Lovett, A. and Flowerdew, R., 1989. 'Analysis of count data using Poisson regression'. The Professional Geographer 41(2): 190–198.

Lundström, S., 1984. 'Estimation for small domains: Two studies using combined data from censuses, surveys and registers'. Statistisk Tidskrift 22(2): 119–126.

Lutz, W., Prinz, C., Wils, A. B., Büttner, T. and Heilig, G., 1991. 'Alternative demographic scenarios for Europe and North America', in W. Lutz (ed), Future Demographic Trends in Europe and North America. What Can We Assume Today? 523–560. Academic Press, London.

MacGill, S. M., 1977. 'Theoretical properties of biproportional matrix adjustments'. Environment and Planning A 9: 687–701.

MacRae, E. C., 1977. 'Estimation of time-varying Markov processes with aggregate data'. Econometrica 45(1): 183–198.

Malacic, J., 1993. 'International economic migration in former Yugoslavia since 1960', in H. Fassmann and R. Münz (eds), European Migration Since 1945. Historical Patterns, Actual Trends and Social Implications. Springer Verlag, Berlin. Forthcoming.

McCullagh, P. and Nelder, J. A., 1989. Generalized Linear Models. 2nd Edition. Chapman and Hall, London.

McNees, S. K., 1990. 'The role of judgment in macroeconomic forecasting accuracy'. International Journal of Forecasting 6(3): 287–299.

Morrison, P. (Ed.) 1983. Population Movements: Their Forms and Functions in Urbanization and Development. ORDINA Publications, Liege (for IUSSP, Liege).

Muus, P. and Cruijsen, H., 1991. 'International Migration in the European Community'. Background paper on migration under two long-term population scenarios for the European Community International Conference 'Human Resources in Europe at the Dawn of the 21st Century', Luxemburg, 27–29 November 1991.

Nair, P. S., 1981. 'Estimation of Directional Migration Flows in Developing Countries: Development of a Methodology and Applications'. Ph.D. Dissertation. Interuniversity Programme in Demography, Brussels.

Nair, P. S., 1982. 'India's population: A multiregional demographic analysis'. Working Paper No. 35. Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague.

Nair, P. S., 1985. 'Estimation of period-specific gross migration flows from limited data'. Demography, 22: 133–142.

NCBS, 1989. Sociaal-demografische rekeningen. Maatschappelijke participatie en mobiliteit. Methodebeschrijving en toepassing voor 1980–85 [Socio-demographic accounts. Social participation and mobility. Method and application 1980–85]. NCBS, Voorburg, The Netherlands.

NCBS, 1990. Sociaal-demografische rekeningen. Bevolking in huishoudens. Methodebeschrijving en

toepassing voor 1985–86 [Socio-demographic accounts. Population in households. Method and application 1980–85]. NCBS, Voorburg, The Netherlands.

Nelder, J. A. and Wedderburn, R. W. M., 1972. 'Generalized linear models'. Journal of the Royal Statistical Society A 135: 370–384.

Ogilvy, A., 1980. 'Interregional migration since 1971. An appraisal of data from the National Health Service Central Register and Labour Force Survey'. Occasional Paper No. 16. Office of Population Censuses and Surveys, London.

Pacini, M., 1992. Foreword to 'The New Europe and International Migration', a special issue of International Migration Review 26: 229–232.

Plane, D. A., 1982. 'An information theoretic approach to the estimation of migration flows'. Journal of Regional Science 22: 441–456.

Ploeg, F. van der, 1984. 'Generalized least squares methods for balancing large systems and tables of national accounts'. Review of Public Data Use 12: 17–23.

Poulain, M. 1991. 'Un projet d'harmonisation des statistiques de migration internationales au sein de la Communauté Européenne'. Revue Européenne des Migrations Internationales 7(2): 115–138.

Poulain, M., Debuisson, M. and Eggericks, T., 1990. Projet d'harmonisation des statistiques de migration internationales au sein de la Communauté Européenne. Institut de Démographie. Université Catholique de Louvain, Louvain-la-Neuve.

Poulain, M. and Gisser, R., 1993. Migration Statistics for the EFTA countries. EUROSTAT, Luxemburg.

Pryor, R., 1981. 'Integration of international and internal migration theories', in M. Kritz, et al. (eds), Global Trends in Migration. Theory and Research on International Migration, 110–129. Center for Migration Studies, New York.

Ravenstein, E., 1885. 'The laws of migration I'. Journal of the Royal Statistical Society 48: 167–227.

Ravenstein, E., 1889. 'The laws of migration II'. Journal of the Royal Statistical Society 52: 241–301.

Rees, Ph., 1993. 'Counting People: Past, Present and Future'. Inaugural lecture presented at the University of Leeds, 15 February 1993'. Working Paper 93/3. School of Georgraph, University of Leeds.

Rees, Ph. and Wilson, A., 1977. Spatial Population Analysis. Edward Arnold, London.

Rees, Ph. and Willekens, F. J., 1986. 'Data and accounts', in A. Rogers and F. J. Willekens (eds), Migration and Settlement. A Multiregional Comparative Study 19–56. Reidel, Dordrecht, The Netherlands.

Rees, P. and Woods, R., 1986. 'Demographic estimation: Problems, methods and examples' in R. Woods and P. Rees (eds), Population Structures and Models. Developments in Spatial Demography 301–343. Allen and Unwin, London.

Rogers, A., 1973. 'Estimating internal migration from incomplete data using model multiregional life tables'. Demography 10: 277–287.

Rogers, A., 1975. Introduction to Multiregional Mathematical Demography. Wiley, New York.

Rogers, A., 1991. 'Indirect estimation of migration using birthplace-specific data on multiregional population distribution'. Working Paper WP–91–3. Population Program, University of Colorado, Boulder.

Rogers, A. and Castro, L., 1981. 'Model migration schedules'. Research Report RR–81–30. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and Castro, L., 1986. 'Migration', in A. Rogers and F. Willekens (eds), Migration and Settlement. A Multiregional Comparative Study, 157–208. Reidel, Dordrecht, The Netherlands.

Rogers, A. and von Rabenau, B., 1971. 'Estimation of interegional migration streams from place-of-birth-by-residence data'. Demography 8: 185–194.

Rogers, A. and Willekens, F. (Eds.) 1986. Migration and Settlement. A Multiregional Comparative Study. Reidel, Dordrecht, The Netherlands.

Roy, J. and Flood, J., 1992. 'Interregional migration modelling via entropy and information theory'. Geographical Analysis 24(1): 16–34.

Russell, S. S., 1993. International Migration in North America, Europe, Central Asia, the Middle East and North Africa: Research and Research-Related Activities. United Nations/Economic Commission for Europe, Geneva, and the World Bank, Washington.

Saaty, T. L., 1990. An exposition of the AHP in reply to the paper 'Remarks on the Analytic Hierarchy Process' by J. S. Dyer. Management Science 36(3): 259–268.

Sadiq, N., 1993. Opening statement. Expert Group Meeting on Population Distribution and Migration in Preparation of the 1994 International Conference on Population and Development, Santa Cruz, Bolivia, January.

Schmertmann, C. P., 1992. 'Estimation of historical migration rates from a single census: Interregional migration in Brazil 1900–1980'. Population Studies 46: 103–120.

Simon, J., 1993. 'The economic effect of immigration'. European Review 1(1): 109–116.

Snickars, F. and Weibull, J. W., 1977. 'A minimum information principle. Theory and practice'. Regional Science and Urban Economics 7: 137–168.

Speare, A., 1992. Preparing Migration Data for Subnational Population Projections. Department of International Economic and Social Affairs, United Nations, New York.

Stone, J. R. N., Champernowne, D. G. and Meade, J. E., 1942. 'The precision of national income estimates'. Review of Economic Studies 9: 111–125.

Stone, R., 1971. Demographic Accounting and Model-Building. Organization for Economic Cooperation and Development, Paris.

Straubhaar, T. and Fischer, P. A., 1993. 'Economic and social aspects of immigration into Switzerland', in H. Fassmann and R. Münz (eds) European Migration Since 1945. Historical Patterns, Actual Trends and Social Implications. Springer Verlag, Berlin. Forthcoming.

Teitelbaum, M. S., 1992. 'The population threat'. Foreign Affairs 72(5): 64–78.

Tribalat, M., 1991. 'Combien sont les Français d'origine étrangère? [How many French residents have a foreign origin?]'. Economie et Statistique 242: 17–29.

United Nations, 1970. 'Methods of measuring internal migration'. Population Studies, No. 47, ST/SOA/Series A/47. United Nations, New York.

United Nations, 1975. 'Towards a system of social and demographic statistics'. Studies in Methods, Series F, No. 18. ST/ESA/STAT/SER.F/18. United Nations, New York.

United Nations, 1977. Demographic Yearbook 1977. United Nations, New York.

United Nations, 1983. 'Manual X: Indirect techniques for demographic estimation'. Population Studies, No. 81. Department of International Economic and Social Affairs, United Nations, New York.

Van de Kaa, D. J., 1993. 'European migration at the end of history'. European Review 1(1): 87–108.

Van der Erf, R., 1993. Asylum Seekers and Refugees in Europe. EUROSTAT, Luxembourg.

Vishnevsky, A. and Zayonchkovskaya, Z., 1993. 'Emigration from the former USSR: The fourth wave', in H. Fassmann and R. Münz (eds), European Migration Since 1945. Historical Patterns, Actual Trends and Social Implications. Springer Verlag, Berlin. Forthcoming.

Wetrogan, S. I., 1988, 'Projections of the population of states, by age, sex and race: 1988 to 2010'. Current Population Reports, Series P–25, No. 1017. U.S. Government Printing Office, Washington, D.C.

Willekens, F., 1977. 'The recovery of detailed migration patterns from aggregate data: An entropy maximizing approach'. Research Memorandum RM–77–58. International Institute for Applied Systems Analysis, Laxenburg, Austria.

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

Willekens, F., 1982a. 'Multidimensional population analysis with incomplete data', in K. Land and A. Rogers (eds), Multidimensional Mathematical Demography, 43–111. Academic Press, New York.

Willekens, F., 1982b. 'Identification and measurement of spatial population movements', in ESCAP, National Migration Surveys Series, Manual X: Guidelines for Analysis, 74–97. United Nations, New York.

Willekens, F., 1983a. 'Log-linear modelling of spatial interaction'. Papers of the Regional Science Association. 52: 187–205.

Willekens, F., 1983b. 'Specification and calibration of spatial interaction models. A contingency table perspective and an application to intra-urban migration in Rotterdam'. Tijdschrift voor Economische en Sociale Geografie 74(4) 239–252.

Willekens, F., 1985. 'Comparability of migration data: Utopia or reality?' in M. Poulain (ed.), Migrations internes. Collecte des données et méthodes d'analyse, 409–441 (Chaire Quételet 1983). Cabay, Louvain-la-Neuve.

Willekens, F., 1993. Estimation of Transition Intensities from Census and Retrospective Survey Data. The Case of Migration. Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague.

Willekens, F. and Baydar, N. 1986. 'Forecasting place-to-place migration with generalized linear models', in R. Woods and Ph. Rees (eds.), Population Structures and Models. Developments in Spatial Demography, 203–244. Allen and Unwin, London.

Willekens, F. and Ramachandran, P., 1993. Estimation of Migration Flows from Limited Data. Theory and Application. Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague.

Willekens, F., A. Por and Raquillet, R., 1979. Entropy, multiproportional and quadratic techniques for inferring detailed migration patterns from aggregate data. Mathematical theories, algorithms, applications, and computer programs. Working Paper WP–79–88. International Institute for Applied Systems Analysis, Laxenburg, Austria (Shorter version published in IIASA Reports, Vol. 4, 1981, pp. 83–124).

Willekens, F., Wetrogan, S. I. and Long, J., 1986. 'Towards a Synthetic Data Base of Interstate Migration'. Paper presented at the Annual Meeting of the Regional Science Association, Columbus, Ohio, November.

Wilmoth, J. R. and Ball, P., 1992. 'The population debate in American popular magazines, 1946–90'. Population and Development Review 18(4): 631–668.

Wils, A. B., 1993. 'European Long-Term Migration Data: Overview and Evaluation of Existing Data Collection'. Working Paper WP–93–28. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Wils, A. B. and Willekens, F. J., 1993. 'Migration Estimation Study'. Project document. International Institute for Applied Systems Analysis, Laxenburg, Austria, and Population Research Centre. University of Groningen, The Netherlands.

Wilson, A. G., 1970. Entropy in Urban and Regional Modelling. Pion Ltd., London.

Wolfe, C. and Flores, B. E., 1990. 'An analysis of hierarchically adjusted earnings forecast'. Journal of Forecasting 4: 389–405.

Zlotnik, H. and Hovy, B., 1990. 'Trends in European Migration: What the Data Reveal'. Paper presented at the Symposium on the Demographic Consequences of International Migration, NIAS, Wassenaar, The Netherlands, September.