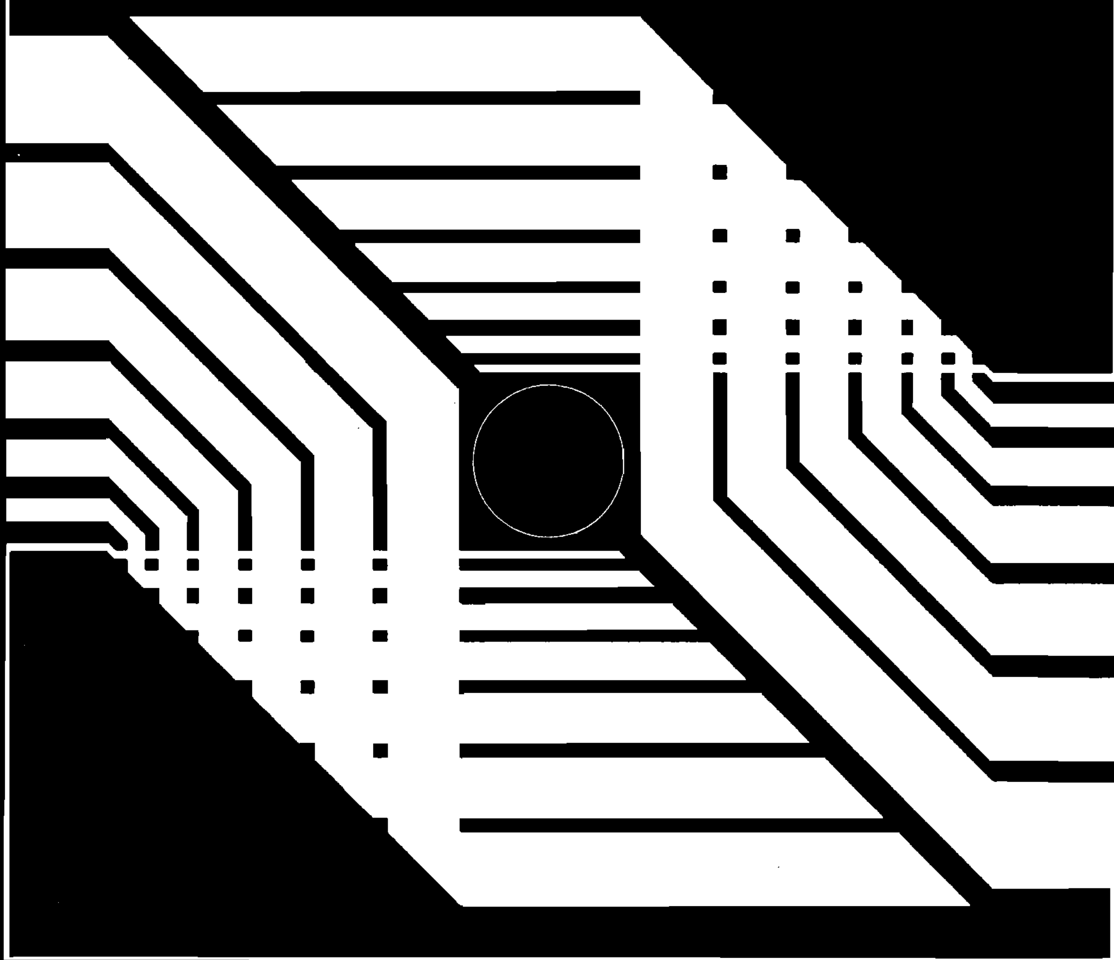


Migration and Settlement: Selected Essays

Andrei Rogers, Editor

RR-78-6
MAY 1978

reprinted from

environment & planning a

volume 10, number 5

Environment and Planning A, 1978, volume 10, number 5, May, pages 469-617

Contents

Environment and Planning A, 1978, volume 10, pages 469-474

Guest editorial

Representatives of twelve countries met in London in October 1972 to found the International Institute for Applied Systems Analysis (IIASA), situated in Laxenburg on the outskirts of Vienna, Austria. The founders put forward three goals for the Institute: (1) to strengthen international cooperation; (2) to advance the science and art of systems analysis; and (3) to apply systems analysis to a wide range of real societal problems. During the past few years, five additional countries have joined this unique nongovernmental East-West research institute, and scholars from more than two dozen nations have contributed to its research activities.

One of the principal research areas at IIASA deals with Human Settlements and Services. An important component of this research program since July 1975 has been the Migration and Settlement Study. The papers collected together in this special issue are a representative sample of the research output of that study. They complement an earlier collection of IIASA papers published in this journal three years ago (*Environment and Planning A* 7 number 7).

The Migration and Settlement Study at IIASA

Human-settlement issues and problems are the focus of increasing concern among national governments in many parts of the world. Programs to encourage the development of economically declining regions, to stem the growth of large urban centers, and to revitalize the central areas of expanding metropolises are parts of national agendas all over the globe. A notable manifestation of this growing interest in human-settlement processes was the 1976 United Nations Habitat Conference, which focused on the critical problems created by the convergence of two historical developments: unprecedentedly high rates of population growth, and massive rural-to-urban migration.

Despite the general recognition that migration processes and human-settlement growth patterns are intimately related, one nevertheless finds that the dynamics of their interrelationships are not at all well understood. An important reason for this lack of understanding is that demographers have in the past accorded migration a status subservient to fertility and mortality, and have generally neglected the spatial dimension of population growth. Thus, whereas problems of fertility and mortality long ago stimulated a rich and scholarly literature, studies of migration have only recently begun to flourish. Consequently one finds today a rather large and growing body of scholarly work on migration awaiting a systematic synthesis: for example, the recent bibliographies of Greenwood (1975), Price and Sikes (1975), and Shaw (1975). The contributions of sociologists in identifying migration differentials (the 'who' of migration), of geographers in analyzing directional migration streams (the 'where' of migration), and of economists in examining the determinants and consequences of geographical mobility (the 'why and 'so what' of migration) still have not been molded into a unified and general theory of internal migration.

Out of the growing literature in migration and human settlement, an international group of scholars at IIASA has identified and focused on four related research subtasks that are of particular relevance to the Institute's long-term broad interests in national settlement systems and strategies. They are:

(1) The study of spatial population *dynamics*;

(2) The definition and elaboration of a new research area called *demometrics* and its application to migration analysis and spatial demoeconomic forecasting;

(3) The analysis and design of migration and settlement *policy*; and
(4) A *comparative study* of national migration and settlement patterns and policies.

Dynamics

Every spatial human population evolves out of the interactions of births, deaths, and migration. Individuals are born into a population, age with the passage of time, reproduce, and ultimately depart because of death or out-migration. Thus the growth of a regional population is determined by the combined effects of natural increase (births minus deaths) and net migration (in-migrants minus out-migrants). The *dynamics* subtask focuses on such relationships in order to identify and clarify some of the more fundamental spatial population processes involved. In addition to its general concern with the expansion of our knowledge of spatial mathematical demography (Rogers and Willekens, 1976), the dynamics subtask is also focusing on problems of constructing multiregional life tables (Rogers and Ledent, 1976), of model schedules and model populations, of sensitivity analysis (Willekens, 1977), of spatial zero population growth, and of aggregation-decomposition procedures (Rogers, 1976).

The age-specific fertility, mortality, and migration schedules of most human populations exhibit remarkably persistent regularities. Consequently demographers have found it possible to summarize and codify such regularities by means of hypothetical schedules called *model* schedules. But although the development of model *fertility* and *mortality* schedules has received a considerable amount of attention, the construction of model *migration* schedules has not. The first of the seven papers in this issue shows how the techniques that have been successfully applied to treat the former problems can readily be extended to deal with the latter. In it, Rogers (USA), Raquillet (France), and Castro (Mexico) demonstrate that regularities in observed migration schedules may be summarized by means of either the 'mortality' approach of Coale and Demeny (1966) or the 'fertility' approach of Coale and Trussell (1974).

Increasing concern about the sizes and growth rates of national populations has generated a vast literature dealing with the socioeconomic and environmental consequences of a reduction of fertility to replacement levels, and the consequent evolution of national populations to a zero-growth condition called *stationarity*. Yet the ways in which stabilization of a national population is likely to affect migration and local growth have received very little attention, but do merit careful study. Rogers (USA) and Willekens (Belgium) consider in their paper some of the redistributional consequences of an immediate reduction of fertility to bare replacement levels. Using the mathematical apparatus developed by demographers to analyze the evolution of national populations to zero growth, the authors include the spatial impact of internal migration to show that the demographic redistributional effects of stabilization will depend in a very direct way on the spatial pattern of total births that is occasioned by fertility reduction.

Demometrics

In 1938 the United States National Resources Committee published a major demographic study which, after the adoption of a set of 'reasonable' assumptions with regard to future fertility, mortality, and net inmigration, projected the total United States population in 1980 to be 158 million, at which time it was also to have reached a state of stationary equilibrium. The United States population passed the 158 million mark less than fifteen years later and today exceeds 210 million.

It is difficult to fault such projections, for it is very unlikely that any competent demographer, faced with the same situation, would have come up with a radically different estimate. How then can the accuracy of such exercises in social prediction be improved? We at IIASA believe that the development of a field called *demometrics* is a necessary first step.

In economics a division is generally made between the areas of *mathematical economics* and *econometrics*. The former deals principally with abstract mathematical descriptions of economic dynamics and economic growth; the latter treats statistically estimated relationships between basic economic variables. Analogous distinctions are made to distinguish mathematical psychology from psychometrics, mathematical biology from biometrics, and mathematical sociology from sociometrics. In a similar vein, mathematical demography may be distinguished from demometrics, the development and elaboration of which forms the major focus of the second of the four principal subtasks of the Migration and Settlement Study (Rogers, 1978).

In a broad sense, demometrics is concerned with the unified application of mathematical and statistical methods to the study of demographic phenomena. Its principal aim is to establish empirically quantifiable relationships between demographic and socioeconomic variables. It is important not to confuse this activity with mathematical demography and statistical techniques (as does Winkler, 1969), or with demographic statistics (as does the layman). Demometrics is distinguished by its fusion of the deductive approach of mathematics, the inductive approach of statistics, and the causal approach of demographic theory. Its principal objective is to establish quantitative statements regarding major demographic variables that either *explain* the past behavior of such variables or *forecast* their future behavior.

Much of our work in the demometrics subtask is directed toward advancing the state of the art in consistent forecasting. This has led us to consider three related areas of research: migration theory, urban labor-force dynamics, and demoeconomic modeling.

Theory and model building in the field of migration research has both a micro and a macro dimension. The microtheory is dominated by the economist's perspective of migration as an investment in human capital—a perspective that views migration as a response to a comparison in which benefits exceed the costs of moving. The macrotheory of migration, on the other hand, focuses on aggregate movements and, like macroeconomics, generally fails to relate in any precise manner its theories with those concerning individual decisionmaking (that is, the microtheory).

Although an economic model forecasts the quantity of labor that will be demanded, it is important to know how much of that labor will be supplied by the resident population and how much by new migrants. In-migration, for example, implies a larger population, whereas a change in the labor-force participation rate of residents does not. Both sources of labor supply compete for the same jobs, and each exerts a different influence on the ultimate forecast.

Extensive literature on urban labor-force dynamics is available. In the past much of it has been concerned with proving or disproving the 'added worker' and 'discouraged worker' hypotheses (Mincer, 1966). Recently attention has been directed at the dynamics of the job-search process itself (Phelps, 1970). This body of literature may ultimately provide the connecting link between urban migration and labor-force dynamics; consequently an increasing amount of our research effort is being directed toward establishing such a connection. Building on this body of literature and on the earlier IIASA research of Cordey-Hayes (1975), John Miron (Canada) in his paper outlines a robust 'job-search' perspective for analyzing migration behavior.

A popular approach to demoeconomic modeling is the coupling of an economic model with a demographic model by means of linkages through the consumption and labor sectors (Rogers and Walz, 1973). The former linkage appears as a consumption function that demands the economy to produce a certain output for the population to consume. The latter linkage takes the form of a migration-labor-force equilibrating model that views the demographic model as the supplier of labor, and the economic model as the demander of labor. The two models operate recursively in developing forecasts of demographic and economic growth that are internally consistent. Consistency of such detail, however, is costly to obtain and requires data that often are not available. These difficulties have led Jacques Ledent (France) to explore in his paper a much simpler perspective that begins with a modest extension of the conventional economic-base model and then expands such a model into a multiequation macrodemometric model of modest size and complexity, without a corresponding expansion of data requirements.

Policy

The policy analysis and design subtask of the Migration and Settlement Study is surveying the fundamental dimensions of current national migration and settlement policies. In particular it is examining the consequences of migration, those affecting both the origin region and the destination region. And it is evaluating the utility of the 'optimal policy' paradigm of Tinbergen that recently has been usefully applied by mathematical economic planners (Fox et al, 1972).

A wide variety of countries are striving to channel urban growth to certain regions and to divert it from others. Generally such national urbanization or human-settlement policies have been defended on the grounds either of national efficiency or of regional equity. The arguments are often framed in terms of an underlying conceptual framework known as 'growth-pole theory' with migration playing a central role (Moseley, 1974).

Several scholars of internal migration have concluded that the experience of migration affects favorably the personal well-being and satisfaction of the migrant. However, the societal consequences of migration often fall unequally on different groups. By transferring labor from labor-surplus areas to labor-deficit areas, migration moves the national economy toward greater efficiency. Since it is the most productive members of the labor force that are the ones who move away, the localities they leave behind often become increasingly unattractive for industrial investment.

The diverse individual and societal consequences of internal migration have broad implications for national policies dealing with migration and settlement. The built-in conflict between the goals of national efficiency and regional equity is a fundamental 'fact of life' in the design of such policies, one that ultimately can be resolved only in the political arena. A potentially useful tool for illuminating some of the trade-offs is offered by the formal theory of economic policy, first proposed by Tinbergen (1963) in the field of economic planning.

The Tinbergen paradigm focuses on the problem of using available means to achieve desired ends in an optimal manner. It begins by adopting a quantitative empirical (econometric) model and divides variables into *endogenous* and *predetermined* variables. The policy problem, as formulated by Tinbergen, is to choose an appropriate set of values for the instrumental variables so as to render the values of the target values equal to desired values previously established by an objective function (that is, a welfare function).

The logical structure of the economic Tinbergen paradigm is formally analogous to those of decisionmaking in other fields. Spatial demography is one such field, and we therefore are actively exploring the utility of this approach for migration- and settlement-policy analysis and design. Some of our current efforts in this direction are described in the papers by Kulikowski (Poland) and by Propoi (USSR) and Willekens (Belgium). They reflect a perspective that was initiated at IIASA by Evtushenko and MacKinnon (1976) and carried forward by Willekens and Rogers (1977).

Comparative study

IIASA's comparative study of migration and settlement aims to increase our understanding of the relationships between the geographical mobility, the natural increase, and the spatial redistribution of national populations. For this task it has adopted the general framework of two recently published studies carried out in a closely related area (Keyfitz and Flieger, 1971; Berelson, 1974).

Keyfitz and Flieger focus on observed age- and sex-specific mortality and fertility schedules, and project the evolution of the populations exposed to these schedules. They examine current population trends by subjecting a data bank of population statistics from more than ninety countries to a standardized analytical process. The primary *focus* of their study is national population growth, and its principal *approach* in examining such growth is embodied in a collection of computer programs that provide the vehicle for analyzing population growth in a consistent and uniform manner. The major *contribution* of the Keyfitz and Flieger study is the uniform application of a consistent methodology to a vast amount of data in order to examine the population-growth trends of a large number of countries.

The focus, approach, and contribution of the Keyfitz and Flieger study have much in common with those of the comparative study of migration and settlement. The focus of the latter is also population growth, but *spatial* population growth. The approach also relies on a uniform set of computer programs, but these embody the models of *multiregional* mathematical demography (Rogers, 1975). And the expected contribution is also that of linking data and theory, but the data and theory that are linked are *spatial* in character.

Although chapter 4 of their book is entitled "Policy Dilemmas and the Future", the Keyfitz and Flieger study does not deal with national policies (their chapter 4 is only three pages long). The comparative study, however, explicitly considers the national migration and settlement policies of each country represented. In this respect, the study resembles more the study of population policies coordinated by Berelson (1974). This book edited by Berelson is a review of population policies in twenty-four developed countries. The individual chapters were written by collaborating scholars residing in the particular countries. The collaborators were given a common outline of topics to be addressed, but each was free to prepare his report in his own manner. It is therefore not surprising that different authors elected to emphasize different aspects of population policy, and drew on different kinds of demographic data to develop their presentations.

The comparative study of migration and settlement at IIASA is combining the Berelson approach with the Keyfitz–Flieger approach in order to capture the best features of each. Every national analysis is, as in the Berelson study, being carried out in collaboration with scholars residing in the countries being studied. However, most of the data, projections, and indicators that form the foundation of the analysis are being processed at IIASA, as in the Keyfitz–Flieger study, by a common set of computer programs. These data and programs will be published.

Dimiter Philipov's national case-study report on migration and settlement patterns and policies in Bulgaria concludes the collection of papers brought together for this special issue. In it he describes current patterns of spatial population growth in Bulgaria and traces their evolution by examining the interactions of multiregional schedules of fertility, mortality, and migration with age-specific regional distributions of the national population. The result is a detailed picture of Bulgaria's spatial demography expressed in quantitative measures that also are being developed for the sixteen other countries that are members of IIASA.

Conclusion

Spatial demographic analysis is concerned with the study of multiregional population systems, primarily with regard to their size, age composition, and regional distribution, and the changes of these over space and time. Such a perspective allows one to examine the demographic interactions between the urban and regional agglomerations that shape national human-settlement patterns. Its focus on interregional migration and on regional differences in fertility, mortality, and age composition is of particular importance in forecasting local and multiregional populations. The Migration and Settlement Study at IIASA is striving to carry out policy-relevent research concerning the geographical dimensions of present and future population growth in order to develop improved analytical tools for assessing the probable consequences of alternative population and economic policies.

A Rogers
Chairman, Human Settlements and Services Area, IIASA, Laxenburg, Austria

References

Berelson B (Ed.), 1974 *Population Policy in Developed Countries* (McGraw-Hill, New York)
Coale A J, Demeny P, 1966 *Regional Model Life Tables and Stable Populations* (Princeton University Press, Princeton, NJ)
Coale A J, Trussell J T, 1974 "Model fertility schedules: variations in the age structure of childbearing in human populations" *Population Index* **40** 185-201
Cordey-Hayes M, 1975 "Migration and the dynamics of multiregional population systems" *Environment and Planning A* 7 793-814
Evtushenko Y, MacKinnon R D, 1976 "Nonlinear programming approaches to national settlement system planning" *Environment and Planning A* 8 637-653
Fox L, Sengupta J, Thorbecke E, 1972 *The Theory of Quantitative Economic Policy* (North-Holland, Amsterdam)
Greenwood M J, 1975 "Research on internal migration in the United States: a survey" *Journal of Economic Literature* **13** 397-433
Keyfitz N, Flieger W, 1971 *Population: Facts and Methods of Demography* (Freeman, San Francisco)
Mincer J, 1966 "Labor-force participation and unemployment: a review of recent evidence" in *Prosperity and Unemployment* Eds R A Gordon, M S Gordon (John Wiley, New York) pp 73-112
Moseley M H, 1974 *Growth Centers in Spatial Planning* (Pergamon Press, Oxford)
Phelps E S (Ed.), 1970 *Microeconomic Foundations of Employment and Inflation Theory* (W W Norton, Boston, Mass)
Price D O, Sikes M M, 1975 *Rural-Urban Migration Research in the United States: Annotated Bibliography and Synthesis* (US Government Printing Office, Washington, DC)
Rogers A, 1975 *Introduction to Multiregional Mathematical Demography* (John Wiley, New York)
Rogers A, 1976 "Shrinking large-scale population-projection models by aggregation and decomposition" *Environment and Planning A* **8** 515-541
Rogers A, 1978 "Demometrics of migration and settlement" in *London Papers in Regional Science 8. Theory and Method in Urban and Regional Analysis* Ed. P W J Batey (Pion, London) pp 1-30
Rogers A, Ledent J, 1976 "Increment-decrement life tables: a comment" *Demography* **13** 287-290
Rogers A, Walz R, 1973 "Consistent forecasting of regional demographic-economic growth" in *Proceedings of the Sixth Hawaii International Conference on Systems Sciences, Second Supplement* (Western Periodicals, Los Angeles, Calif.) pp 86-89
Rogers A, Willekens F, 1976 "Spatial population dynamics" *Papers of the Regional Science Association* **36** 3-34
Shaw R P, 1975 *Migration Theory and Fact: A Review and Bibliography of Current Literature* (Regional Science Research Institute, Philadelphia, Pa)
Tinbergen J, 1963 *On the Theory of Economic Policy* (North-Holland, Amsterdam)
United States National Resources Committee, 1938 *The Problems of a Changing Population* (US Government Printing Office, Washington, DC)
Willekens F, 1977 "Sensitivity analysis in multiregional demographic models" *Environment and Planning A* 9 653-674
Willekens F, Rogers A, 1977 "Normative modeling in demo-economics" RR-77-18, International Institute for Applied Systems Analysis, Laxenburg, Austria
Winkler W, 1969 *Demometrie* (Duncker and Humboldt, Berlin)

Environment and Planning A, 1978, volume 10, pages 475-502

Model migration schedules and their applications

A Rogers, R Raquillet¶, L J Castro¶
International Institute for Applied Systems Analysis, Laxenburg, Austria
Received 17 January 1978

Abstract. This paper considers and contrasts two alternative approaches for capturing the regularities exhibited by age patterns in observed migration rates. The mortality approach is considered first and it is shown how such an approach may be used to infer migration flows from two consecutive place-of-residence-by-place-of-birth census age distributions. The fertility approach is considered next, and techniques for graduating migration age profiles are described. The advantages and disadvantages of both approaches are then briefly assessed.

1 Introduction

The evolution of a human population undisturbed by emigration or immigration is determined by the fertility and mortality schedules it has been subject to. If such a 'closed' population system is disaggregated by region of residence then its spatial evolution is largely determined by the prevailing schedules of internal migration.

The age-specific fertility, mortality, and migration schedules of most human multiregional populations exhibit remarkably persistent regularities. The age profiles of these schedules seem to be repeated, with only minor differences, in virtually all developed and developing nations of the globe. Consequently demographers have found it possible to summarize and codify such regularities by means of hypothetical schedules called *model schedules.*

Model schedules have two important applications: (1) they may be used to infer (or 'smooth') empirical schedules of populations for which the requisite data are lacking (or inaccurate), and (2) they can be applied in analytical mathematical examinations of population dynamics.

The development of model fertility and model mortality schedules, and their use in studies of the evolution of human populations, have received considerable attention (Arriaga, 1968; Coale and Demeny, 1966; Coale, 1972; Coale and Trussell, 1974; Rele, 1967). The construction of model *migration* schedules and their application to studies of the *spatial* evolution of human populations disaggregated by region of residence, however, have not. This paper addresses the latter problem and shows how techniques that have been successfully applied to treat the former problem can readily be extended to deal with the latter. We begin in section 2 by considering the regularities exhibited by observed migration schedules. We then follow this description of observed regularities with an examination, in sections 3 and 4 respectively of two alternative approaches for summarizing such regularities: the mortality approach and the fertility approach. Section 5 offers concluding remarks and points to future directions for research.

¶ L J Castro is on leave from Felipe Ochoa SC, Consultores, Avenida Revolución número 1909, piso 7, México 20, D.F., México; R Raquillet is on leave from Techniville, 61 Avenue Franklin Roosevelt, 75008 Paris, France.

2 Regularities in observed migration schedules

Demographers have long recognized that persisting regularities appear in empirical age-specific migration schedules (for example, Lowry, 1966; Long, 1973). Migration, viewed as an event, is highly selective with regard to age, with young adults generally being the most mobile group in any population. Levels of migration are also high among children, varying from a peak during the first year of life (the *initial peak*) to a *low point* around the age of sixteen. The migration age profile then turns sharply upward until it reaches a second peak (the *high peak*) in the neighborhood of 22 years, after which it declines regularly with age, except for a slight hump (the *retirement peak*) around the ages of 62 to 65.

The regularities in observed migration schedules are not surprising.

> "Young adults exhibit the highest migration rates because they are less constrained by ties to their community. Their children generally are not in school, they are more likely to be renters rather than home owners, and job seniority is not yet an important consideration. Since children move only as members of a family, their migration pattern mirrors that of their parents. Consequently, because younger children generally have younger parents, the geographical mobility of infants is higher than that of adolescents. Finally, the small hump in the age profile between ages 62 to 65 describes migration after retirement ..." (Rogers, 1975, pages 146-147).

2.1 *Migration profiles*

The shape, or *profile*, of an age-specific schedule of migration rates is a feature that may be usefully studied independently of its intensity, or *level*. This is because there is considerable empirical evidence that although the latter tends to vary significantly from place to place, the former is remarkably similar in various localities. Illustrations of this property appear in figure 1, which sets out migration rates for the USA and Sweden.

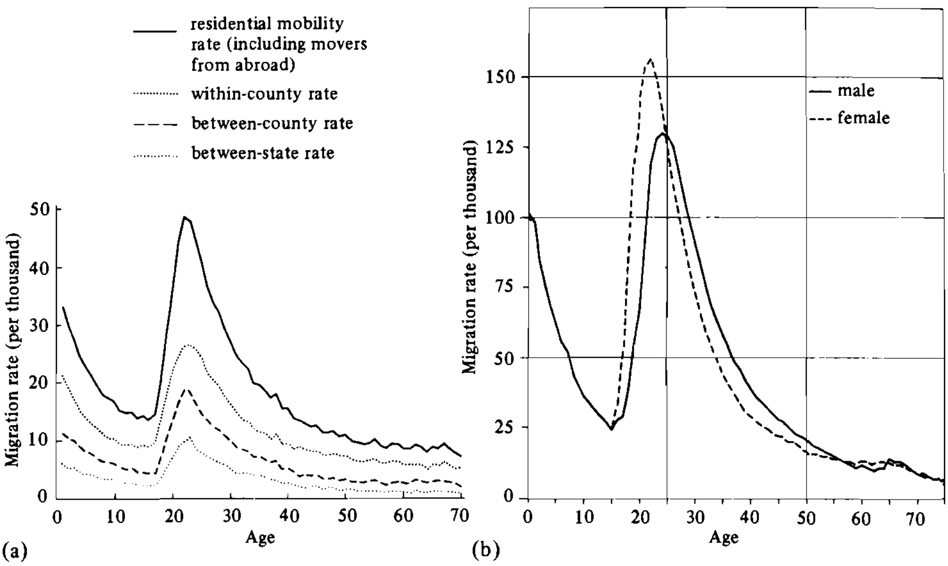


Figure 1. Age-specific annual migration rates of (a) the total United States population, by category of move, averaged over 1966-1971 (source: Long, 1973, page 38), and of (b) the Swedish population, by sex, averaged over 1968-1973 (source: Swedish Central Bureau of Statistics, 1974, page 10).

Figure 1 shows the fundamental age pattern of migration described at the start of section 2, with peaks occurring at infancy, during young adulthood, and, in one instance, at retirement. Variations in the location of the high peak and in the levels of migration at retirement ages indicate that, as in the case of mortality, age profiles of migration may be usefully disaggregated into families that are distinguished by the location and relative height of their peaks. Alternatively such a disaggregation may be carried out, in the manner of fertility schedules, by means of the *mean age*, $\bar{n}$, of the migration schedule $M(x)$,

$$\bar{n} = \sum_{x=0}^{z}(x+2\cdot5)M(x) \bigg/ \sum_{x=0}^{z} M(x) ,$$

where $x = 0, 5, 10, ..., z$ ranges over the initial years of five-year age groups and z is the initial year of the last age group. This parameter may readily be used to classify migration schedules into 'young' and 'old' categories, perhaps with suitable gradations in between.

Figure 1(a) indicates that the age profile of migration tends to be remarkably similar for residential movers, intra- and inter-county migrants, and migrants between states. What does vary is the level of migration, the level being higher for smaller territorial units.

Figure 1(b) shows that important age-specific variations exist between the migration rates of males and females. The high peak for males follows that of the female schedule by a few years, and in the Swedish case it is also lower in height.

2.2 *Migration levels*

The level of migration, like that of *mortality*, can be measured in terms of an expected duration time, for example, the fraction of a lifetime that is expected to be lived at a particular location. However, like *fertility*, migration is a potentially repetitive event, and its level therefore can be expressed in terms of an expected number of migrations per person. The summarization of observed regularities within the context of the former perspective leads to the development of a regression approach similar to the one used by Coale and Demeny (1966) to summarize regularities in mortality schedules; the latter perspective suggests an alternative procedure—one analogous to that used by Coale and Trussell (1974) to describe fertility schedules.

The most common demographic measure of level is the notion of *expectancy*. Demographers often refer to life expectancies, for example, when speaking about mortality, and to reproduction expectancies when discussing fertility. They have calculated for instance that 73 is the average number of years a female could expect to live under the mortality schedule of the United States in 1958, and 1·71 is the average number of baby girls she could expect to bear during her lifetime under the then prevailing fertility schedule. The former measure is known as the *expectation of life at birth*, $e(0)$; the latter index is called the *net reproduction rate*, R^{NR}.

A related index is the *gross reproduction rate*, R^{GR}. This measure totally ignores the effects of mortality on reproduction and may be viewed as the net reproduction rate that would arise if all individuals survived to the end of their childbearing ages. For this reason, the gross reproduction rate of a population is, of course, always larger than the corresponding net reproduction rate.

Expectancies also have been used in migration studies by Wilber (1963) and Long (1973). However, their definitions are *nonspatial* inasmuch as they view migration as an *event* in a national population rather than as a *flow* between regional populations. The study of *spatial* population dynamics can be considerably enriched by explicitly identifying the *locations* of events and flows. This permits one to define spatial expectancies such as the expectation of life at birth or the net reproduction rate of individuals born in region i [respectively ${}_ie(0)$ and ${}_iR^{\mathrm{NR}}$, say], and the expected allocation of this lifetime or rate among the various constituent regions of a multiregional population system [${}_ie_j(0)$ and ${}_iR_j^{\mathrm{NR}}$, respectively, for $j = 1, 2, ..., m$]. For example, it has been estimated (Rogers, 1975) that the expectation of life at birth of a California-born woman exposed to the 1958 schedules of mortality and migration would be 73·86 years, out of which 24·90 years would be lived outside of California. The net reproduction rate of such a woman, from 1958 fertilities, would be 1·69, with 0·50 of that total being born outside of California.

Adopting the second perspective, Wilber developed a set of migration expectancies describing the average number of migrations experienced by an individual during his remaining lifetime. The application of his formula for calculating migration expectancies for individuals just born produces the direct analog of the conventional formula for the net reproduction rate. The formula, with x starting at zero, may be expressed as

$$\sum_{x=0}^{z} L(x)M(x) , \tag{1}$$

where $L(x)$ is the stationary life-table population aged x to $x+4$ years at last birthday, $M(x)$ is the annual rate of migration among individuals in that age group, and z is the starting age of the last interval of life. The corresponding formula for the net reproduction rate is

$$R^{\mathrm{NR}} = \sum_{x=0}^{z} L(x)F(x) , \tag{2}$$

where $F(x)$ is the age-specific fertility rate. The similarity between equations (1) and (2) suggests the designation of equation (1) as the *net migraproduction rate,* a quantity we shall denote by R^{NM}. Thus R^{NR} denotes the average number of *babies* per person, and R^{NM} denotes the average number of *migrations* per person, both taken over that person's entire lifetime. Observe that both measures depict the average number of occurrences of a recurrent event over an individual's lifetime. Only the latter, however, is influenced by the spatial extent of the territorial unit.

Earlier we proposed a spatial migration expectancy based on *duration* times, specifically, the expected number of years lived in region j by individuals born in region i. The correspondence between the net migraproduction and net reproduction rates suggests an *alternative* definition of spatial migration expectancy—one reflecting a view of migration as a recurrent event. Just as R^{NR} was apportioned among the constituent regions of a multiregional system, so too can R^{NM} be similarly disaggregated by place of birth and residence. Thus the formula for the spatial net reproduction rate,

$${}_iR_j^{\mathrm{NR}} = \sum_{x=0}^{z} {}_iL_j(x)F_j(x) ,$$

suggests the following definition for the *spatial net migraproduction rate*:

$${}_iR_j^{\mathrm{NM}} = \sum_{x=0}^{z} {}_iL_j(x)M_j(x) ,$$

where ${}_iL_j(x)$ denotes the stationary life-table population of region j aged x to $x+4$ years at last birthday and born in region i, and $M_j(x)$ is the age-specific out-migration rate in region j.

The spatial net migraproduction rate ${}_iR_j^{\mathrm{NM}}$ describes the average lifetime number of migrations made out of region j by an individual born in region i. The summation of ${}_iR_j^{\mathrm{NM}}$ over all regions of destination ($j \neq i$) gives ${}_iR^{\mathrm{NM}}$, the net migraproduction rate of individuals born in region i, that is, the average number of migrations an i-born person is expected to make during his (or her) lifetime.

Associated with the concept of the *net* reproduction rate in fertility analysis is the notion of the *gross* reproduction rate,

$$R^{\mathrm{GR}} = 5 \sum_{x=0}^{z} F(x) .$$

The notion of a *gross migraproduction rate*,

$$R^{\mathrm{GM}} = 5 \sum_{x=0}^{z} M(x) ,$$

has a similarly useful interpretation in migration analysis. It measures the intensity of migration between two regions at a particular point in time. The measure therefore has basically a cross-sectional character, in contrast to the net migraproduction rate which measures the intensity of migration over a lifetime. Consequently, the *gross* migraproduction rate often may prove to be a more useful measure than the net rate in that it is a 'purer' indicator of migration (in the same sense as the gross reproduction rate is for reproduction). However, the gross rate measures the intensity of migration at a given moment and not over a lifetime. Hence, in instances where return migration is an important factor, the gross rate and the net rate may give differing indications of geographical mobility.

Table 1 presents net and gross migraproduction rates for the total United States population in 1958, disaggregated into four regions. Figure 2 plots the gross migraproduction rate against the mean age for the migration schedules of the four-region United States population system in 1958 and 1968 respectively. We find

Table 1. Net and gross migraproduction rates and mean ages of migration for the total United States population, 1958.

(a) *Net migraproduction rates*: ${}_iR_j^{\mathrm{NM}}$

Region of birth	Region of residence				Total
	Northeast	North Central	South	West	
Northeast	0·4122	0·0366	0·0589	0·0331	0·5408
North Central	0·0204	0·4923	0·0604	0·0600	0·6331
South	0·0300	0·0629	0·4397	0·0479	0·5805
West	0·0203	0·0540	0·0602	0·4181	0·5526

(b) *Gross migraproduction rates and the corresponding mean ages of migration (in parentheses)*: R_{ij}^{GM} *and* $\bar{n}_{ij}$

Region of origin	Region of destination				Total
	Northeast	North Central	South	West	
Northeast	-	0·1202 (26·99)	0·3168 (33·46)	0·1532 (29·43)	0·5902
North Central	0·0891 (28·15)	-	0·3201 (32·16)	0·3289 (30·54)	0·7381
South	0·1504 (28·59)	0·2511 (27·77)	-	0·2299 (27·27)	0·6314
West	0·0887 (27·73)	0·2167 (30·03)	0·2819 (27·61)	-	0·5873

evidence of a division into four groups:

high R^{GM}-high $\bar{n}$; high R^{GM}-low $\bar{n}$;
low R^{GM}-high $\bar{n}$; low R^{GM}-low $\bar{n}$.

Migration flows from the North Central region to the South, for example, exhibit an 'old' profile and a mean age of about 32·5 years. The reverse migration flows, on the other hand, take on the shape of a 'young' profile and show a mean age that is about five years younger. This suggests that it may be useful to develop a *family* of basic model migration schedules so that the various age profiles exhibited by empirical migration schedules can be more accurately captured and summarized.

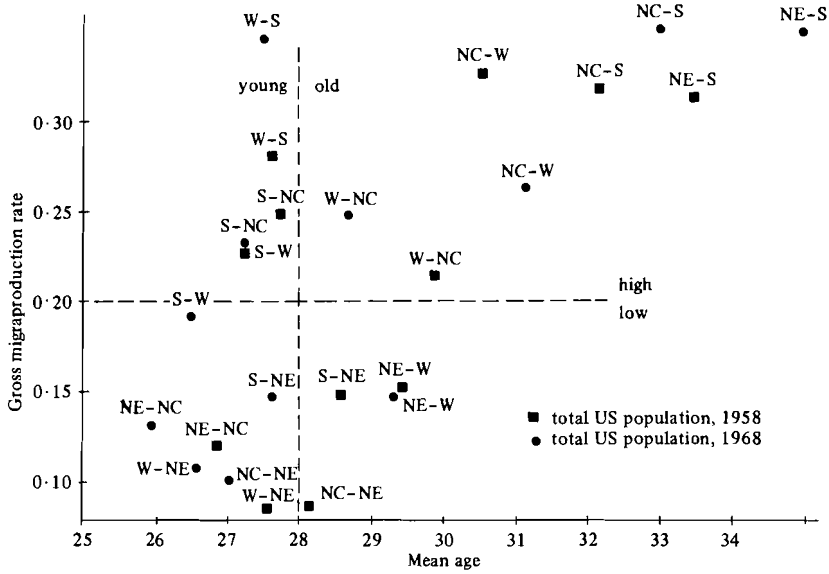


Figure 2. Relationship between observed gross migraproduction rates and mean ages of migration for total United States populations, 1958 and 1968. (NE—Northeast, NC—North Central, S—South, and W—West.)

3 Model migration schedules: the mortality approach

Regularities in the age patterns of observed death rates have fascinated demographers almost since published records of mortality first became generally available. The search for universal 'laws of mortality' gave birth to the well known Gompertz-curve graduation of the mortality schedule and, more recently, to two sets of 'model' life tables published by the United Nations (United Nations, 1955; 1967). The model life table or mortality approach for capturing regularities in observed rates may be applied in the study of migration regularities (Rogers and Castro; 1976). Such an application, however, first requires the concept of the *multiregional* life table.

3.1 *The multiregional life table*

Conventional life tables describe the evolution of a hypothetical cohort of babies born at a given moment and exposed to an unchanging age-specific schedule of mortality. For this cohort of babies, these tables exhibit a number of probabilities of dying and surviving and develop the corresponding expectations of life at various ages.

Life-table calculations are normally initiated by estimating a set of age-specific probabilities of dying within each age interval, $q(x)$ say, from observed data on age-specific death rates, $M(x)$ say. The conventional calculation that is made for an age interval of five years is (Rogers, 1975, page 12)

$$q(x) = \frac{5M(x)}{1+\frac{5}{2}M(x)} \; ,$$

or, alternatively,

$$p(x) = 1-q(x) = [1+\tfrac{5}{2}M(x)]^{-1}[1-\tfrac{5}{2}M(x)] \; , \tag{3}$$

where $p(x)$ is the age-specific probability of surviving from *exact age* x to *exact age* $x+5$ (the first year of the next interval). The latter probabilities in turn may be used to define the corresponding probabilities of survival from one *age group* to the next (Rogers, 1975, pages 16 and 85),

$$s(x) = [1+p(x+5)]p(x)[1+p(x)]^{-1} \; . \tag{4}$$

To avoid any possible confusion between the two sets of probabilities, we shall hereafter refer to $s(x)$ as a survivorship proportion, the proportion of individuals surviving from the x to $x+4$ age group to the $x+5$ to $x+9$ age group.

One of the most useful statistics provided by a life table is the average expectation of life at age x, $e(x)$, calculated by applying the probabilities of survival $p(x)$ to a hypothetical cohort of babies and then observing their average length of life beyond each age. Expectations of life at birth, $e(0)$, are particularly useful as indicators of the level of mortality in various regions and countries of the world.

Conventional life tables deal with mortality, focus on a single regional population, and ignore the effects of migration. To incorporate migration, and at the same time to extend the life-table concept to a spatial population comprised of several regions, requires the notion of a multiregional life table (Rogers, 1973). Such life tables describe the evolution of several regional cohorts of babies, all born at a given moment and exposed to an unchanging *multiregional* age-specific schedule of mortality and migration. For each regional birth cohort, these tables provide various probabilities of dying, surviving, and migrating, while simultaneously deriving regional expectations of life at various ages. These expectations of life are disaggregated both by place of birth and by place of residence; they will be denoted by ${}_ie_j(x)$, where i is the region of birth and j is the region of residence.

Multiregional-life-table calculations are greatly facilitated by the adoption of matrix algebra. This leads to a compact notation and an efficient computational procedure; it also very clearly demonstrates a simple correspondence between the single-region and the multiregional formulas. For example, equations (3) and (4) may be shown to have the following multiregional counterparts (Rogers and Ledent, 1976, Rogers, 1975, page 85):

$$\mathbf{P}(x) = [\mathbf{I}+\tfrac{5}{2}\mathbf{M}(x)]^{-1}[\mathbf{I}-\tfrac{5}{2}\mathbf{M}(x)] \tag{5}$$

and

$$\mathbf{S}(x) = [\mathbf{I}+\mathbf{P}(x+5)]\mathbf{P}(x)[\mathbf{I}+\mathbf{P}(x)]^{-1} \; . \tag{6}$$

The diagonal elements of $\mathbf{P}(x)$ and $\mathbf{S}(x)$ are the probabilities of surviving and the survivorship proportions respectively; the off-diagonal elements will be called probabilities of migrating and migration proportions respectively.

Expectations of life in the multiregional life table reflect the influences of mortality and migration. Thus they may be used as indicators of levels of internal migration, in addition to carrying out their traditional function as indicators of levels of mortality. For example, consider the regional expectations of life at birth that are set out in table 2 for the United States population with both sexes combined.

Table 2. Expectations of life at birth and migration levels (in parentheses), by region of residence and region of birth, for the total United States population, 1958.

Region of birth	Region of residence				Total
	Northeast	North Central	South	West	
Northeast	50·90 (0·7295)	4·49 (0·0643)	8·88 (0·1273)	5·50 (0·0788)	69·76 (1·00)
North Central	3·18 (0·0452)	48·45 (0·6889)	9·10 (0·1294)	9·60 (0·1365)	70·32 (1·00)
South	4·58 (0·0664)	7·52 (0·1091)	49·21 (0·7134)	7·67 (0·1111)	68·98 (1·00)
West	3·18 (0·0454)	6·60 (0·0944)	8·95 (0·1279)	51·22 (0·7322)	69·94 (1·00)

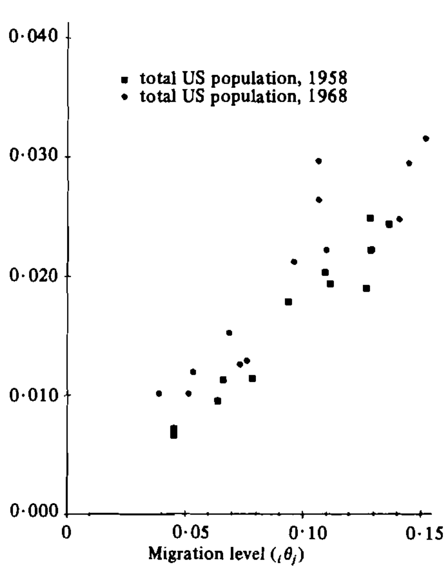


Figure 3. Relationship between initial peaks and migration levels in two observed migration schedules.

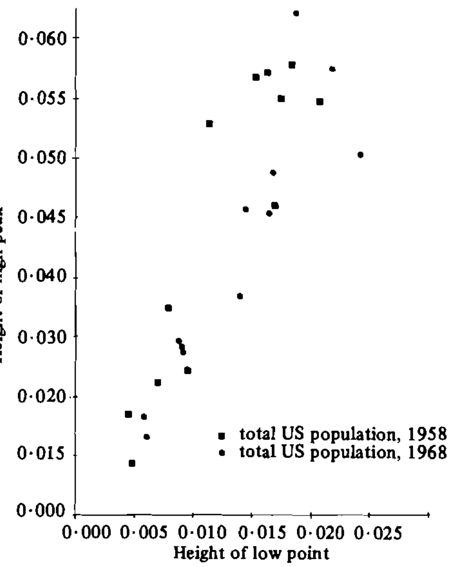


Figure 5. Relationship between high peaks and low points in migration schedules.

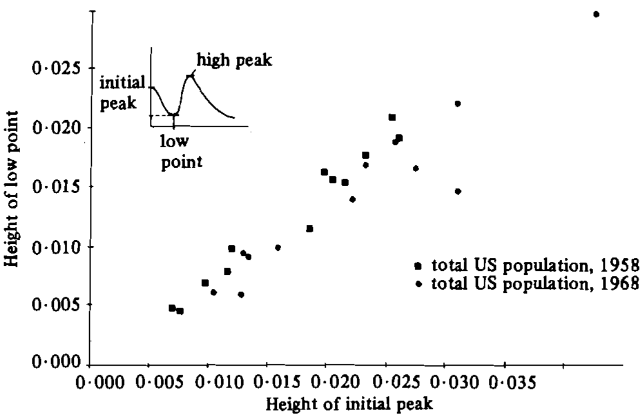


Figure 4. Relationship between low points and initial peaks in migration schedules.

A baby born in the West, and exposed to the multiregional schedule of mortality and migration that prevailed in 1958, could expect to live an average of 69·94 years, out of which total an average of 8·95 years would be lived in the South. Taking the latter as a fraction of the former, we have in 0·1279 a useful indicator of the (lifetime) migration level from the West to the South that is implied by the 1958 multiregional schedule. (Compare these migration levels with those set out earlier in table 1.)

Age-specific probabilities of migrating, $p_{ij}(x)$, in empirical multiregional life tables mirror the fundamental regularities exhibited by observed migration rates. Some of these regularities are illustrated in figures 3, 4, and 5. (We focus only on the total population and consider data for all four census regions for two points in time: 1958 and 1968.) Figure 3 shows that a strong and positive association exists between the height of the initial peak, $p_{ij}(0)$, and the level of migration as measured by, for example, ${}_i\theta_j$, the fraction of the expected lifetime of an individual born in region i that is expected to be lived in region j. Figure 4 indicates that a similarly strong and positive relationship exists between the height of the low point and the height of the initial peak. Figure 5 describes the positive association between the heights of the high peak and the low point. Thus a direct line of correlation appears to connect the general migration level between two regions to the values assumed by the corresponding age-specific probabilities of migrating. This suggests that a simple linear-regression equation may be used to associate a set of probabilities of migrating at each age x, $p_{ij}(x)$, with a single indicator of migration level, ${}_i\theta_j$ say. We explore this possibility next.

3.2 *Summarizing the regularities: regression*

The migration risks experienced by different age and sex groups of a given population are strongly interrelated, and higher (or lower) than average migration rates among one segment of a particular population normally imply higher (or lower) than average migration rates for other segments of the same population. This association stems in part from the fact that if socioeconomic conditions at a location are good or poor for one group in the population, they are also likely to be good or poor for other groups in the same population. Since migration is widely held to be a response to spatial variations in socioeconomic conditions, these high intercorrelations between age-specific migration risks are not surprising.

Figures 3, 4, and 5 support this conjecture and moreover suggest a way of summarizing the observed regularities in migration probabilities. They indicate that a relatively accurate account of the variation in the height of the initial peak (and through it in the height of the rest of the migration schedule) may be obtained by means of a straight line fitted to the scatter of points in figure 3. Thus a linear regression of the form

$$p_{ij}(0) = \gamma + \beta\, {}_i\theta_j$$

would seem to be appropriate[1]. But $p_{ij}(0)$ cannot take on negative values; a convenient way of ensuring that this possibility never arises is to force the line through the origin by adopting the model of zero-intercept simple linear regression,

$$p_{ij}(0) = \beta\, {}_i\theta_j\,. \tag{7}$$

[1] Since changes in fertility also affect the height of the initial peak, a possible further refinement of the model would be to include a variable describing the level of fertility, for example, the reproduction rate.

The least-squares fit of such an equation to the data illustrated in figure 3 gives

$$p_{ij}(0) = 0{\cdot}17392\,{}_i\theta_j$$

for the 1958 observations and

$$p_{ij}(0) = 0{\cdot}22002\,{}_i\theta_j$$

for the 1968 data points. The fit in each instance is quite satisfactory, yielding coefficients of determination (r^2) of 0·94 and 0·84 respectively.

Given estimates of β and ${}_i\theta_j$, an estimate of $p_{ij}(0)$ can be obtained. Figures 4 and 5 suggest that, with the value of $p_{ij}(0)$ fixed, the corresponding value of the low point can be found and used, in turn, to estimate the value of the high point. By generalizing this argument to all age groups beyond the first, the simple model

$$p_{ij}(x+5) = \gamma(x)p_{ij}(x)$$

may be adopted, where $p_{ij}(0)$ is estimated by equation (7). Thus

$$p_{ij}(5) = \gamma(0)p_{ij}(0) = \gamma(0)\beta\,{}_i\theta_j = \beta(5)\,{}_i\theta_j\,,$$

$$p_{ij}(10) = \gamma(5)p_{ij}(5) = \gamma(5)\beta(5)\,{}_i\theta_j = \beta(10)\,{}_i\theta_j\,,$$

and, in general,

$$p_{ij}(x) = \beta(x)\,{}_i\theta_j\,, \tag{8}$$

in which the β in equation (7) is now designated by $\beta(0)$. Note that, as a consequence of these definitions,

$$\gamma(x) = \frac{\beta(x+5)}{\beta(x)}$$

and

$$p_{ij}(x+5) = \gamma(x)\frac{\beta(x)}{\beta(0)}p_{ij}(0) = \frac{\beta(x+5)}{\beta(0)}p_{ij}(0)\,.$$

Table 3. Regression coefficients for obtaining model probabilities of migrating, for total United States populations, 1958 and 1968.

Age[a]	1958		1968	
	β	r^2	β	r^2
0	0·17392	0·94	0·22002	0·84
5	0·13460	0·95	0·15553	0·89
10	0·15736	0·86	0·15040	0·94
15	0·30757	0·93	0·29195	0·85
20	0·32271	0·72	0·26370	0·72
25	0·23251	0·96	0·20037	0·90
30	0·17897	0·95	0·17907	0·94
35	0·12912	0·95	0·14392	0·96
40	0·09790	0·93	0·10397	0·95
45	0·07522	0·86	0·07378	0·91
50	0·06838	0·73	0·06352	0·76
55	0·07347	0·63	0·07362	0·54
60	0·08254	0·47	0·08320	0·43
65	0·06086	0·50	0·06425	0·47
70	0·04488	0·58	0·04919	0·64
75	0·03019	0·67	0·03951	0·64
80	0·01342	0·18	0·02058	0·63

[a] Age denotes the initial year of the five-year age group.

Equation (8) may be treated as a simple (zero-intercept) linear-regression equation, and its coefficient $\beta(x)$ may be estimated using the conventional least-squares procedure. Table 3 presents two sets of such coefficients for the total United States population. The first set was obtained using 1958 data; the second set was estimated on the basis of 1968 data. In both instances the observed migration flows were those between the four United States census regions.

The regression coefficients in table 3 may be used in the following way. First, starting with a complete set of multiregional migration levels ${}_i\theta_j$, one calculates the matrix of migration probabilities $\mathbf{P}(x)$, for every age, by use of equation (8) and one of the two sets of regression coefficients (β) in table 3. (Figure 6 illustrates a range of such probabilities by way of example.) With $\mathbf{P}(x)$ established, one then may compute the usual life-table statistics, such as the survivorship proportions defined in equation (6) and the various region-specific expectations of life at each age. The collective results of these computations constitute a *model multiregional life table.*

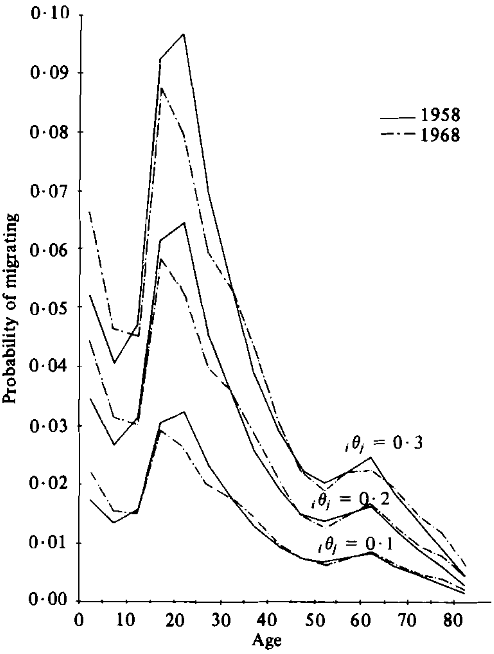


Figure 6. Age-specific model probabilities of migrating at various levels of migration.

3.3 *Families of model migration schedules*

In this section we consider the effects on the migration age profile of various disaggregations of our data on the United States population system. Specifically, we examine how the regression coefficients set out in table 3 and illustrated in figure 7 respond to various disaggregations of the empirical population on the basis of which they were estimated. First we disaggregate the total population by sex. Second we introduce a disaggregation according to mean age. Third we consider a spatial disaggregation of the four census regions into their constituent nine census divisions. Last we explore the impact of an even finer deconsolidation by mean age.

The two regression-coefficient profiles in figure 7 mirror the fundamental age profile of migrants that was analyzed earlier in this paper. The principal differences between the two coefficient profiles are the higher and older high peak in the 1958 migration schedule, and the higher and older low point of the corresponding 1968 schedule.

Beyond the mid-thirties the two profiles are quite similar, with both showing a retirement peak in the 60–64 age group.

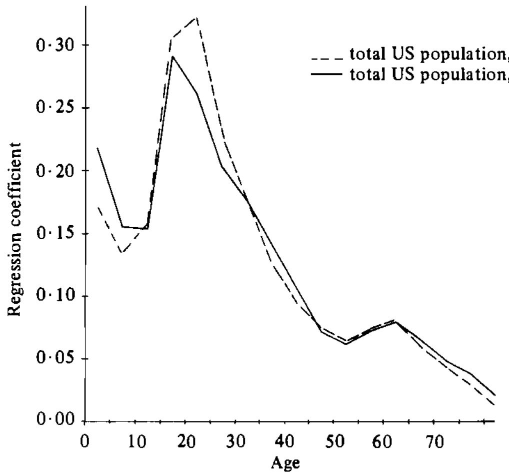


Figure 7. Regression coefficients for model migration schedules for total United States populations, 1958 and 1968.

3.3.1 *Profile differences by sex.* A disaggregation of the 1968 regression-coefficient profile introduces important variations by sex, according to figure 8. The male coefficients are higher from the very early teens to the mid-forties and are lower at all other ages. The locations of the high peak and the retirement peak are the same in both profiles, but the low point among males comes at a younger age than in females. Also the retirement peak among females is broader and starts at an earlier age.

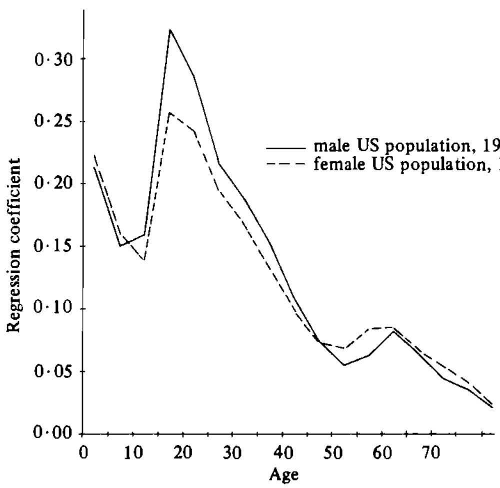


Figure 8. Regression coefficients for model migration schedules for male and female United States populations, 1968.

3.3.2 *Profile differences by mean age.* Figure 9 indicates that a division of migration schedules into 'young' and 'old' categories might be a useful way of disaggregating the regression coefficients illustrated in figures 7 and 8. It shows two basic age profiles which are distinguishable by the presence of a high retirement peak in one profile and its virtual absence in the other. We designate the former profile as a retirement profile and the latter as a labor-force profile. An alternative designation is an old and a young profile respectively.

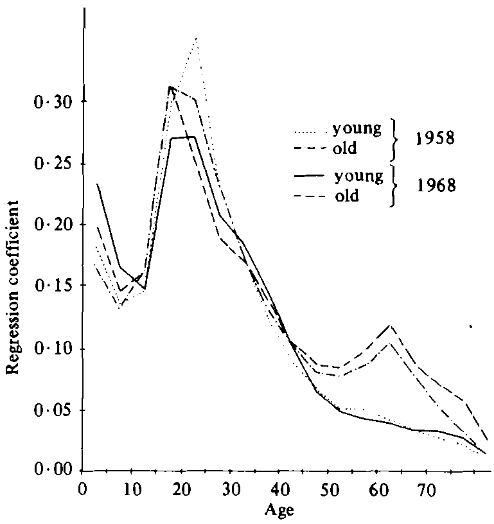


Figure 9. Regression coefficients by 'young' and 'old' classifications ($\bar{n}_{ij} \leqslant 28$ and $\bar{n}_{ij} > 28$) for total United States populations, 1958 and 1968.

3.3.3 *Profile differences by size of areal unit.* Because migration is normally defined as a crossing of a regional boundary, it is clear that reducing the size of a spatial unit should increase the level of out-migration from that unit since some of the moves that previously did not cross over the old borders will now be recorded as migrations over the new borders. But what of the age profile in each case? Should not this feature of the observed migration flows remain essentially unchanged, at least for the relatively large areal units? Figure 10 [like figure 1(a) before it] gives some evidence that this conjecture is valid. The two regression-coefficient profiles that it illustrates

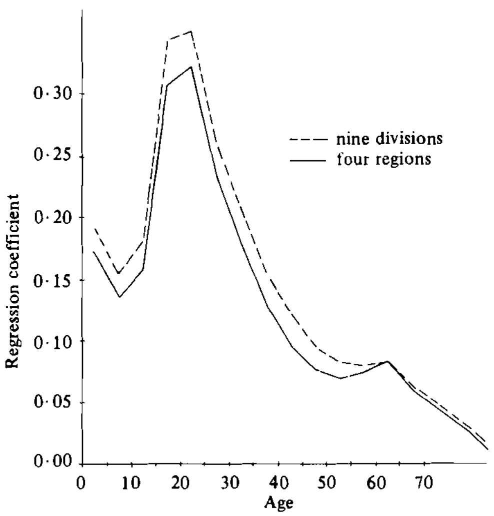


Figure 10. Regression coefficients for model migration schedules for the total United States population, 1958, by regions and divisions.

were estimated on the basis of the same data set, using first a nine-division and then a four-region spatial delineation of the total 1958 United States population. The fact that the former is always slightly higher than the latter is perhaps a consequence of some confounding of profile and level introduced by aggregation bias.

3.3.4 *Profile differences by several mean-age classes.* The spatial disaggregation of the data from four to nine areal units increases the number of observations from twelve to seventy-two and thereby affords us an opportunity to examine the impact of a finer classification by mean age. Specifically, we now consider the disaggregation of the 1958 regression-coefficient profile into four instead of two mean-age categories: 'very young' ($\bar{n}_{ij} \leqslant 26$); 'young' ($26 < \bar{n}_{ij} \leqslant 28$); 'old' ($28 < \bar{n}_{ij} \leqslant 30$); and 'very old' ($\bar{n}_{ij} > 30$).

Except for variations with respect to the retirement peak, the principal impact of the finer disaggregation by mean age appears not so much in the age profile as in the relative height of that profile for a given value of the migration level ${}_i\theta_j$. Thus, for example, the age curve of the 'very old' profile in figure 11 is almost everywhere higher than the corresponding curve of the 'very young' profile *for the same level of migration*. The reason for this is not immediately apparent and merits further study. A possible explanation may lie in the fact that ${}_i\theta_j$ is an index that combines an age-specific migration pattern with a specific (life-table) age composition. This particular confounding of schedule and composition could perhaps generate the variations in profile heights that appear in figure 11, although the underlying dynamics of this are by no means self-evident. Consequently it may well be the case that the 'fertility approach' with its focus on the gross migraproduction rate as an index of migration level has a built-in advantage over the 'mortality approach' that we have been following in this section. This possibility is considered later in this paper.

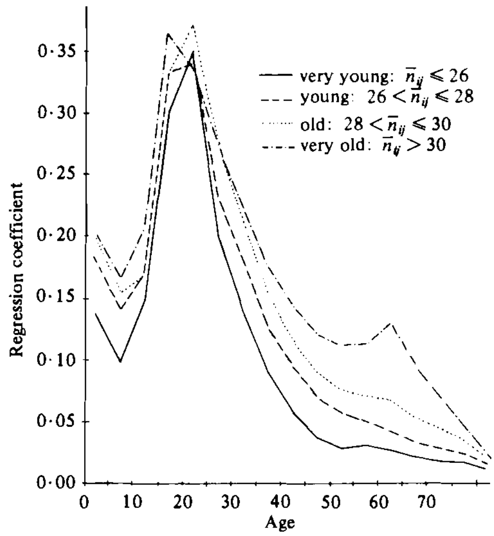


Figure 11. Regression coefficients for model migration schedules for the total United States population, 1958, by several mean age classes.

3.3.5 *Summary.* The regression coefficients illustrated in figures 7 to 11 may be said to form *families* of model migration probabilities or schedules. Those associated with different categories of mean age give 'young' and 'old' profiles; those that do not consider mean age as an index give 'average' profiles. We next illustrate an application of the female 'average' profile by constructing a specimen model multiregional life table.

3.4 *Application: inference*

We noted earlier that single-region life tables are normally computed using observed data on age-specific death rates. In countries lacking reliable data on death rates, however, recourse is often made to inferential methods that rely on model life tables such as those published by the United Nations (United Nations, 1967). These tables are entered with empirically determined survivorship proportions to obtain the particular expectation of life at birth (and corresponding life table) that best matches the levels of mortality implied by the observed proportions.

The inferential procedures of the single-region model (*the UN method*) may be extended to the multiregional case (Rogers, 1975, chapter 6). Such an extension begins with the notion of model multiregional life tables and uses a set of initial estimates of survivorship and migration proportions to identify the particular combination of regional expectations of life, disaggregated by region of birth and region of residence, that best matches the levels of mortality and migration implied by these observed proportions.

3.4.1 *Model multiregional life tables.* Model multiregional life tables approximate the mortality and migration schedules of a multiregional population system by drawing on the regularities observed in the mortality and migration experiences of comparable populations. That is, regularities exhibited by mortality and migration data, collected in regions where these data are available and accurate, are used to systematically approximate the mortality and migration patterns of populations lacking such data.

Table 4 gives the four regional expectations of life at birth and the dozen migration levels that together characterize the patterns of regional mortality and interregional mobility of the United States female population in 1968. An interpolation in the "WEST" family of model life tables developed by Coale and Demeny (1966) first gives the appropriate set of model probabilities of dying at each age for each of the four census regions. The inserting, in turn, of each of the dozen values of ${}_i\theta_j$ into equation (8), with $\beta(x)$ taking on the column of 'average' values illustrated for females in figure 8, permits the derivation of initial approximations for $p_{ij}(x)$. These probabilities of migrating may then be used in conjunction with the associated interpolated model probabilities of dying to obtain the matrix of survivorship proportions defined in equation (6). By appropriately manipulating equation (5), the associated model migration rates can also be found. And then, by following the normal computational procedures of multiregional-life table construction (Rogers, 1975, chapter 3), the corresponding matrix of expectations of life at birth, appropriately disaggregated by region of birth and region of residence, may be derived, for example. Unfortunately the latter matrix usually will not yield the same

Table 4. Expectations of life at birth and migration levels (in parentheses) by region of residence and region of birth for the United States female population, 1968.

Region of birth	Region of residence				Total
	Northeast	North Central	South	West	
Northeast	54·13 (0·7260)	5·08 (0·0681)	10·11 (0·1356)	5·25 (0·0704)	74·56 (1·00)
North Central	3·76 (0·0506)	52·14 (0·7005)	10·48 (0·1408)	8·05 (0·1081)	74·44 (1·00)
South	5·06 (0·0680)	7·88 (0·1060)	54·53 (0·7328)	6·93 (0·0931)	74·40 (1·00)
West	3·90 (0·0516)	7·94 (0·1051)	11·32 (0·1497)	52·41 (0·6936)	75·57 (1·00)

migration levels that were used to generate the $\mathbf{P}(x)$ matrix. Such inconsistencies occur in the model life tables of Coale and Demeny (1966). To eliminate them one must resort to iteration. Only in this way can one obtain a model multiregional life table whose statistics and parameters are internally consistent.

Figure 12 illustrates four sets of model migration rates which were generated in the course of constructing our specimen model multiregional life table for the United States female population. Adjoining the four model schedules are the corresponding empirical schedules observed in 1965-1970. A comparison of the two sets of schedules suggests that, although the degree of correspondence is fairly close, further improvement would be highly desirable.

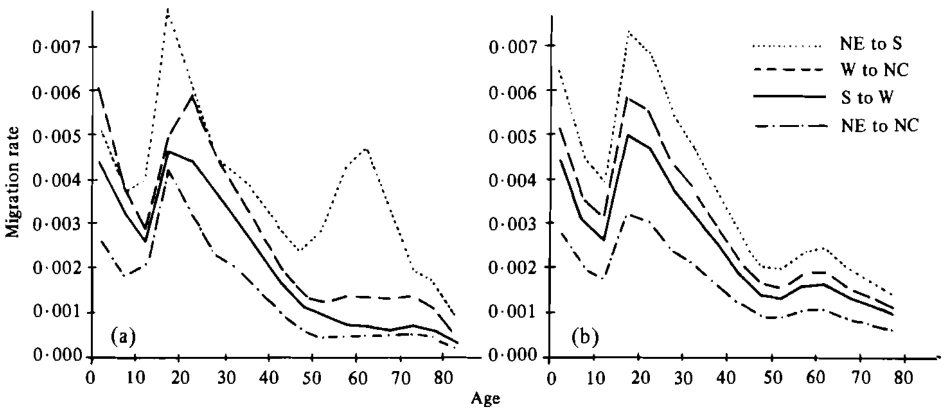


Figure 12. Some (a) observed and (b) model migration schedules for the United States female population.

3.4.2 *The UN method generalized.* The UN method of obtaining initial age-specific estimates of ten-year survivorship proportions from two consecutive decennial census-enumerated age distributions may be generalized to multiregional population systems if age-specific place-of-residence-by-place-of-birth (PRPB) data are available for both census years. This easily may be demonstrated by expressing the single-region procedure in algebraic form and reverting to matrix algebra to define the corresponding multiregional method.

First observe that the single-region procedure for estimating $s(x)$ may be expressed as follows:

$$^{10}\hat{s}(x) = K^{(t+1)}(x+10)[K^{(t)}(x)]^{-1} , \qquad (9)$$

where $K^{(t)}(x)$ denotes the number of persons aged x to $x+9$ years at time t. Next recall the multiregional demographic model that projects populations disaggregated by place of residence and place of birth (Rogers, 1975, page 172),

$$\mathbf{K}(x+5) = \mathbf{S}(x)\mathbf{K}(x) \qquad (10)$$

[where we replace $_0\mathbf{K}(x)$ in Rogers's book by $\mathbf{K}(x)$].

Normally our interest in this model centers on the determination of $\mathbf{K}(x+5)$, given particular numerical values for $\mathbf{S}(x)$ and $\mathbf{K}(x)$. Now, however, we consider the application of equation (10) to derive $\mathbf{S}(x)$ given numerical values for $\mathbf{K}(x)$ and $\mathbf{K}(x+5)$. Clearly

$$\hat{\mathbf{S}}(x) = \mathbf{K}(x+5)[\mathbf{K}(x)]^{-1} ,$$

and, for a ten-year age and time interval,

$$^{10}\hat{\mathbf{S}}(x) = \mathbf{K}^{(t+1)}(x+10)[\mathbf{K}^{(t)}(x)]^{-1} \,. \tag{11}$$

Note that equation (11) is the matrix expression of equation (9).

Having found crude initial estimates of the various regional survivorship and outmigration proportions by means of the PRPB method, one may construct the associated life table to obtain the regional expectations of life at birth that are implied by these proportions (Rogers, 1975, pages 85-88). Then, as in the single-region case, one may 'adjust' the initial probability estimates by interpolating in an appropriate set of model multiregional life tables (Rogers, 1975, pages 185-189).

4 Model migration schedules: the fertility approach

Fertility schedules have long been recognized as exhibiting a fundamental pattern that persists over a wide range of human populations. This recognition has fostered two related research efforts, one concerned with the analytic graduation of fertility curves (Keyfitz, 1968, chapter 6; Hoem and Berge, 1975) and the other focused on the construction of model fertility schedules (Coale and Demeny, 1966, page 30).

In a recent paper, Coale and Trussell (1974) combine these two lines of research to provide an analytic graduation of a standard fertility schedule from which a wide variety of fertility patterns can be derived by a simple transformation involving four parameters. Observed fertility patterns are defined with respect to the natural fertility of married women. Fertility is held to be a function of nuptiality patterns (characterized by two parameters), contraception (characterized by one parameter), and a fertility level (characterized by one parameter). Such a model seems to provide a good fit, and readily leads to methods for obtaining an appropriate fertility schedule on the basis of inadequate information regarding the fertility regime of an observed population.

In this section of the paper, we explore the potential utility of the Coale-Trussell fertility approach for constructing model migration schedules.

4.1 *The fundamental components of migration schedules*

Regularities in observed age-specific schedules of migration may be examined in a number of interesting ways. A particularly useful approach is to decompose the migration schedule into three parts, separating the migration rates of persons in the labor-force age groups from those of individuals in the pre- and post-labor-force ages respectively. Such a decomposition gives rise to the three fundamental curves illustrated in figure 13(a):

(1) the single negative-exponential curve of the pre-labor-force ages with its rate of descent, α_1;

(2) the left-skewed unimodal curve of the labor-force ages with its rates of ascent and descent, λ_2 and α_2 respectively; and

(3) the almost bell-shaped curve of the post-labor-force ages with its rates of ascent and descent, λ_3 and α_3 respectively. (When no retirement peak is exhibited by the data, this last curve is suppressed.)

For future reference, figure 13(a) also includes the constant curve c, to which we shall refer later in the paper. Its inclusion improves the quality of fit provided by the mathematical model schedule.

Figure 13(b) identifies several important points along the age profile of a migration schedule: its *low point*, x_ℓ, its *high peak*, x_h, and its *retirement peak*, x_r. Associated with the first two points is the *labor-force shift*, X, which is defined as the difference in years between the ages at the low point and the high peak, $X = x_h - x_\ell$.

Associated with this shift is *the jump*, the increase in the migration rate of individuals aged x_h over those aged x_ℓ.

Another important shift in observed migration schedules arises out of the close correspondence between the migration rates of children and those of their parents. If, for each point x on the pre-labor-force part of the migration curve, one obtains by interpolation the point, $x + A_x$, say, on the labor-force curve that yields the identical rate of migration, then the average of the values of A_x, calculated for the first fourteen years of life,

$$A = \tfrac{1}{14} \sum_{x=0}^{13} A_x ,$$

will be defined to be the observed *parental shift*.

Table 5 presents numerical approximations of the observed parental shift for eight Swedish regions (viksområden), with single-year age-specific migration and population data for 1974. [Stochastic variations in the rates were first smoothed out by Stoto (1977) by use of a method described in Tukey (1977) called 'nonlinear smoothing'.] The results indicate that the observed parental shift was roughly 26 to 28 years for females with about an additional two years for males. The last row in the table suggests that the parental shift may be closely approximated by the mean age of childbearing (Stoto, 1977).

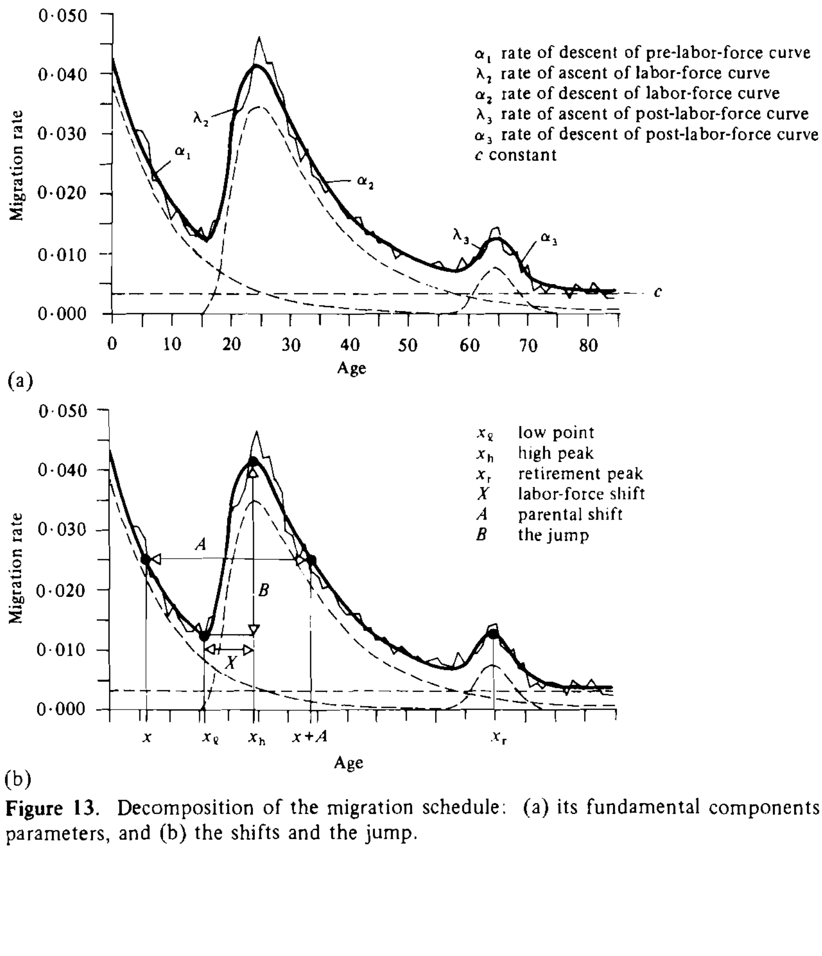


Figure 13. Decomposition of the migration schedule: (a) its fundamental components and their parameters, and (b) the shifts and the jump.

Table 5. Observed values of the parental shift for eight Swedish regions, 1974. The values in parentheses are variances. (The annual migration data by single years of age and the regional delineations were kindly provided by Arne Arvidsson of the Swedish Central Bureau of Statistics.)

Parental shift	Region of Sweden							
	Stockholm	Upper East	Lower East	South	West	Lower North	Middle North	Upper North
Observed, males	27·61	29·77	29·58	28·98	29·42	28·97	30·07	30·33
	(3·19)	(0·86)	(1·75)	(0·58)	(0·23)	(0·27)	(0·60)	(0·83)
Observed, females	25·16	26·50	27·48	27·79	27·45	26·59	27·21	29·52
	(0·50)	(0·31)	(0·96)	(0·60)	(0·23)	(0·97)	(0·82)	(2·01)
Mean age of childbearing, females	27·59	26·95	27·16	27·23	27·34	26·83	27·17	27·38

4.2 *Summarizing the regularities: curve fitting*

Our examination of the regularities in observed migration schedules suggested a decomposition into three curves: a single negative-exponential and two skewed unimodal bell-shaped functions. The obvious mathematical expression for the first is $a\exp(-\alpha x)$, for some appropriate constant a; to represent the other two we have adopted the 'double exponential' developed by Coale and McNeil (1972),

$$a\exp\{-\alpha(x-\mu)-\exp[-\lambda(x-\mu)]\}\,,$$

for some appropriate constant μ. And, because observed migration rates do not drop to zero within the range of post-labor-force ages normally recorded, an additional constant term [the c in figure 13(a)] needs to be included. We then have a model migration schedule that is the simple sum of four curves, namely,

$$\hat{M}(x) = a_1\exp(-\alpha_1 x)+a_2\exp\{-\alpha_2(x-\mu_2)-\exp[-\lambda_2(x-\mu_2)]\}$$
$$+\,a_3\exp\{-\alpha_3(x-\mu_3)-\exp[-\lambda_3(x-\mu_3)]\}+c\,, \qquad \text{for } x = 0, 1, 2, ..., \quad (12)$$

The 'full' model schedule in equation (12) has eleven parameters: a_1, α_1, a_2, μ_2, α_2, λ_2, a_3, μ_3, α_3, λ_3, and c. Migration schedules without a retirement peak may be represented by a 'reduced' model with seven parameters because in such instances the third component of equation (12) is omitted. Illustrative values for the model schedule's parameters are set out in table 6.

Having chosen the particular function in equation (12) to represent age-specific migration schedules, one then is faced with the problem of selecting a method for fitting the function to observed migration data. Previous research in the analytic graduation of fertility schedules has shown that moment-type estimators may be inconsistent and do not compare favorably with functional-minimization methods such as minimum chi-square or least-squares estimation procedures (Hoem, 1972; Hoem and Berge, 1975).

Least-squares parameter estimates are presented in table 6. Minimum chi-square estimates are also included for the United States data for purposes of comparison. The differences between the two sets of parametric estimates tend to be small, and because the latter give more weight to age groups with smaller rates of migration, we use minimum chi-square estimators in the remainder of the paper.

To assess the quality of fit that the model schedule provides when it is applied to observed data, two indices of goodness of fit have been included in table 6: the chi-square statistic, χ^2, and the 'mean absolute error as a percentage of the observed

Table 6. Fundamental parameters of the model migration schedule. Least-squares parameter estimates (LSE) are given except in the case of the United States schedule, for which minimum chi-square estimates (MCSE) are also presented. Data sources are: the United States—Long (1973) and by personal communication from Long (1976); Poland—Polish Central Bureau of Statistics (1974); and Sweden—Swedish Central Bureau of Statistics (1974). All migration schedules were first scaled to a gross migraproduction rate of unity.

Parameters and statistics	United States, 1966-1971				Poland, 1974	
	males ($\bar{n}$ = 29·63)		females ($\bar{n}$ = 28·94)		males ($\bar{n}$ = 31·97)	females ($\bar{n}$ = 33·66)
	LSE	MCSE	LSE	MCSE		
a_1	0·0192	0·0191	0·0200	0·0196	0·0272	0·0266
α_1	0·1186	0·1141	0·1242	0·1177	0·1972	0·1956
a_2	0·0437	0·0427	0·0481	0·0462	0·0771	0·0786
μ_2	19·5378	19·4613	18·4032	18·2979	22·8367	20·3397
α_2	0·1037	0·0999	0·1275	0·1216	0·1611	0·1707
λ_2	0·5354	0·5335	0·5600	0·5868	0·4460	0·4496
a_3	-	-	-	-	0·0001	0·0067
μ_3	-	-	-	-	112·5498	116·2804
α_3	-	-	-	-	0·2995	0·0945
λ_3	-	-	-	-	0·0546	0·0325
c	0·0060	0·0059	0·0066	0·0065	0·0051	0·0042
$\delta_2 = \alpha_2/\lambda_2$	0·1937	0·1872	0·2277	0·2072	0·3612	0·3797
$\delta_3 = \alpha_3/\lambda_3$	-	-	-	-	5·4866	2·9085
χ^2	0·0021	0·0020	0·0027	0·0026	0·0080	0·0017
E	3·33	3·41	3·89	3·96	4·34	2·91
R^{GM} [a]	1·00	1·00	1·00	1·00	1·00	1·00
$\bar{n}$ [a]	29·71	29·68	29·04	29·00	32·09	33·65

Parameters and statistics	Sweden, 1968-1973	
	males ($\bar{n}$ = 26·63)	females ($\bar{n}$ = 25·40)
a_1	0·0300	0·0290
α_1	0·1187	0·1261
a_2	0·0639	0·0730
μ_2	21·4133	19·2624
α_2	0·1059	0·1252
λ_2	0·3846	0·5060
a_3	-	-
μ_3	-	-
α_3	-	-
λ_3	-	-
c	0·0027	0·0032
$\delta_2 = \alpha_2/\lambda_2$	0·2754	0·2474
$\delta_3 = \alpha_3/\lambda_3$	-	-
χ^2	0·0036	0·0028
E	3·79	2·73
R^{GM} [a]	1·00	1·00
$\bar{n}$ [a]	26·71	25·48

[a] Estimated.

mean',

$$E = \left[\frac{1}{n}\sum_x |\hat{M}(x) - M(x)| \bigg/ \frac{1}{n}\sum_x M(x)\right] \times 100 \,.$$

Both measures indicate that the fit of the model to the data is remarkably good.

The numerical values in table 6 suggest possible simplifications of the model.

(1) To the extent that the migration rates of children mirror those of their parents, the parameter α_1 should be approximately equal to α_2. Table 6 indicates that this is indeed the case for the migration schedules of the United States, Poland, and Sweden. Thus a reasonable simplification of the model is to assume that $\alpha_1 = \alpha_2$.

(2) Experiments with a wide range of empirical migration schedules suggest that the ratio of the rate of descent, α, to the rate of ascent, λ, does not vary greatly, particularly for the retirement peak. We assume, therefore, that $\delta_3 = \alpha_3/\lambda_3 =$ a constant $= 5$, say.

Table 7. Goodness of fit and rates of descent of the original and the simplified model migration schedules (minimum chi-square estimates). (The definitions of and the data for the Swedish regions were provided by Arne Arvidsson of the Swedish Central Bureau of Statistics.)

Region	Original model[a]			Simplified model[b]	
	E	α_1	α_2	E	$\alpha_1 = \alpha_2$
United States, 1966-1971					
Males	3·41	0·1141	0·0999	3·60	0·1038
Females	3·96	0·1177	0·1216	3·95	0·1203
Sweden, 1968-1973					
Males	3·83	0·1169	0·1019	4·39	0·1076
Females	3·08	0·1251	0·1170	3·13	0·1195
Swedish regions, 1974					
Stockholm					
Males	6·93	0·0971	0·0776	7·48	0·0856
Females	7·29	0·0905	0·0919	7·32	0·0903
Upper East					
Males	6·42	0·0811	0·0858	6·46	0·0846
Females	7·37	0·1000	0·1030	7·23	0·1042
Lower East					
Males	12·23	0·0984	0·1046	12·45	0·1033
Females	10·82	0·1086	0·1284	11·26	0·1243
South					
Males	11·13	0·1170	0·1143	11·07	0·1153
Females	8·77	0·1043	0·1290	9·34	0·1216
West					
Males	9·40	0·0895	0·0914	9·39	0·0909
Females	9·31	0·1056	0·1140	9·28	0·1110
Lower North					
Males	10·84	0·1037	0·1032	10·83	0·1033
Females	11·66	0·0995	0·1364	12·14	0·1289
Middle North					
Males	11·78	0·1228	0·1178	11·72	0·1189
Females	11·40	0·1185	0·1480	11·94	0·1424
Upper North					
Males	14·88	0·1356	0·1140	14·85	0·1177
Females	13·28	0·1261	0·1417	13·13	0·1398

[a] The eleven-parameter full model was used for the Stockholm and Upper East regions only; in all other cases the seven-parameter reduced model was used.

[b] The nine-parameter ($\delta_3 = 5$) simplifed full model was used for the Stockholm and Upper East regions only; in all other cases the six-parameter simplified model was used.

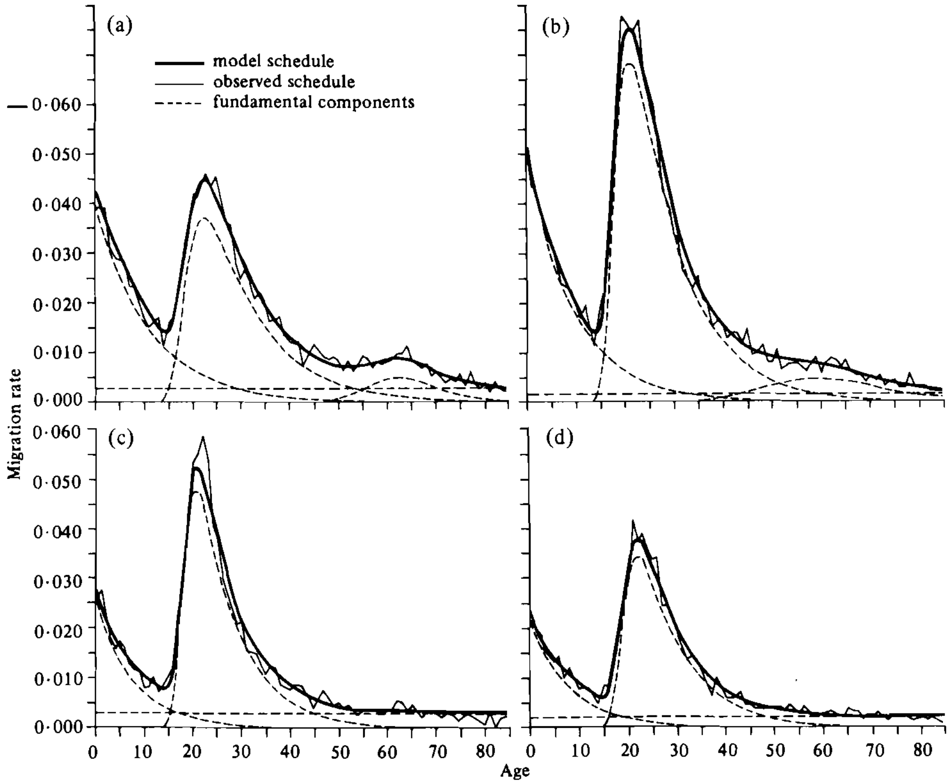


Figure 14. Observed and simplified model migration schedules for females in the Swedish regions, 1974: (a) Stockholm (nine parameters); (b) Upper East (nine parameters); (c) Lower East (six parameters); and (d) South (six parameters).

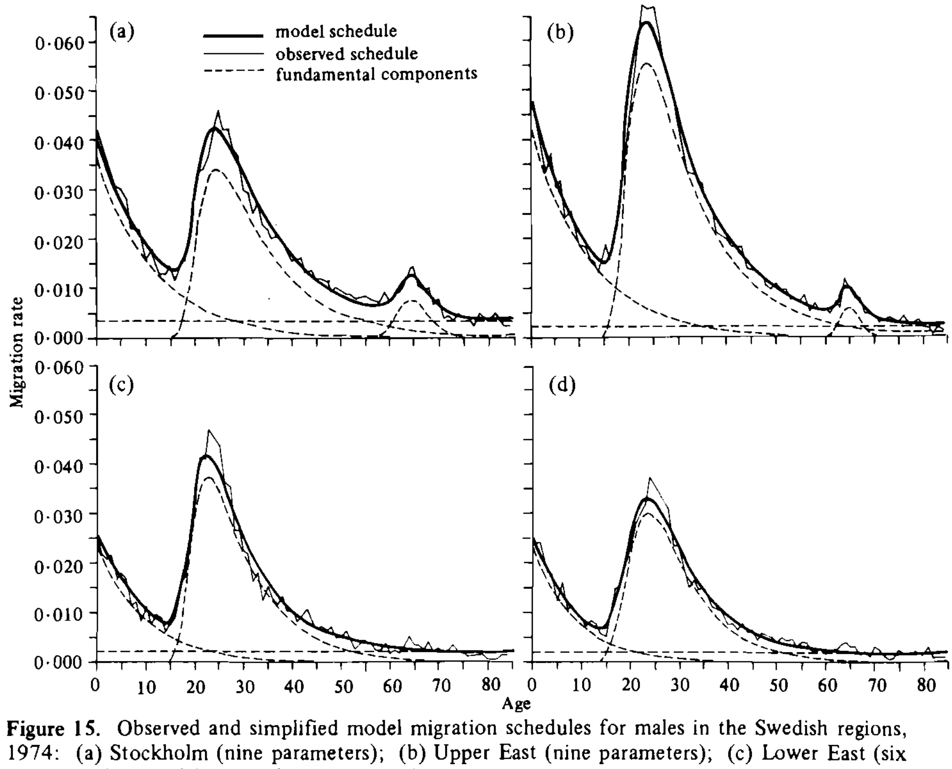


Figure 15. Observed and simplified model migration schedules for males in the Swedish regions, 1974: (a) Stockholm (nine parameters); (b) Upper East (nine parameters); (c) Lower East (six parameters); and (d) South (six parameters).

These two simplifications reduce the number of parameters in the full model to nine and in the reduced model to six. Table 7 compares the fits of the original model with those of the simplified model for data on the United States, Sweden, and eight Swedish regions. Figures 14 and 15 illustrate graphically the closeness of the fit of the simplified model migration schedule to the Swedish regional data. It appears that little information is lost by simplifying the model; and we therefore adopt the simplified full and reduced models for all analyses described in the rest of this paper.

4.3 *Families of model migration schedules*

Model migration schedules of the original form specified in equation (12), or of the simplified form just described, may be classified into *families* according to the values taken on by their principal parameters. For example, we may distinguish those schedules with a retirement peak from those without; or we may refer to schedules with relatively low or high values for the rate of descent α_2. In many applications it is also meaningful and convenient to characterize the model schedules in terms of several of the fundamental measures illustrated in figure 13, such as the low point x_ℓ, the high peak x_h, the labor-force shift X, the parental shift A, and the jump B.

The simplified model migration schedule has a built-in parental shift which can be defined analytically. Shortly after the high peak, the labor-force curve can be closely approximated by the function $a_2\exp[-\alpha_2(x_2-\mu_2)]$. Recalling that the pre-labor-force curve is given by $a_1\exp(-\alpha_2 x_1)$ when $\alpha_1 = \alpha_2$, one may equate the two functions and solve for the difference in ages, x_2-x_1, to find

$$A = x_2-x_1 = \mu_2+\frac{1}{\alpha_2}\ln\frac{a_2}{a_1} ,$$

a new analytical definition of the parental shift.

Table 8 compares the values of this analytically defined parental shift with the corresponding observed parental shifts set out earlier in table 5. The two definitions appear to produce similar numerical values, but the analytical definition has the advantage of being simpler to compute, and it is a more rigorous definition.

In addition to the parental shift, three other measures are sufficient to characterize the profile of a simplified model migration schedule without a retirement peak. They are the low point, the high peak, and the rate of descent (2). Taken together the four measures vary in a regular manner, and, by using an appropriate chain of inferences, it is possible to identify the particular age profile that they specify.

Table 8. Observed and model values of the parental shift for the Swedish regions, 1974.

Parental shift	Region of Sweden							
	Stockholm	Upper East	Lower East	South	West	Lower North	Middle North	Upper North
Observed, males[a]	27·61	29·77	29·58	28·98	29·42	28·97	30·07	30·33
Model, males	26·67	28·97	29·63	29·74	28·84	29·43	29·74	30·59
Observed, females[a]	25·16	26·50	27·48	27·79	27·45	26·59	27·21	29·52
Model, females	24·49	26·33	27·51	28·21	27·19	27·69	27·53	28·59

[a] Source: table 5.

(2) No direct analytical expression seems to exist for computing the low point and the high peak. However, their values may be calculated by means of an iterative numerical procedure that seeks the age at which the sum of two derivatives along the migration age profile is zero.

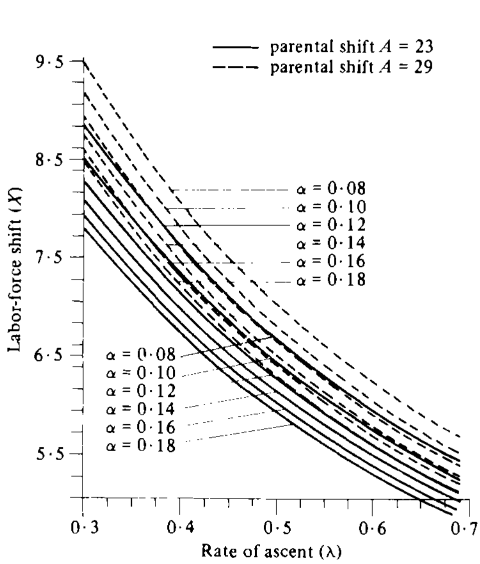


Figure 16. Graph of the labor-force shift against the rate of ascent for two values of the parental shift.

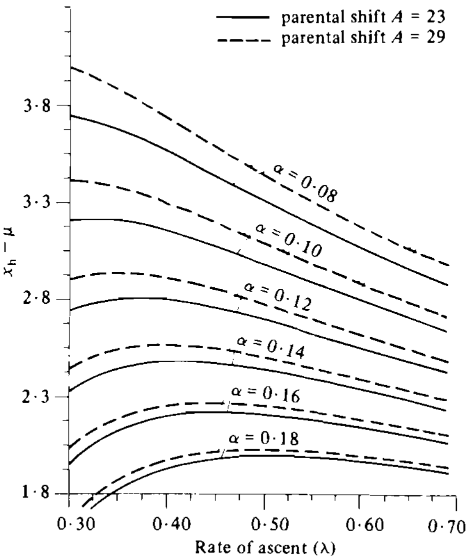


Figure 17. Graph of the high peak minus mean distance against the rate of ascent for two values of the parental shift.

Table 9. Parameter variation under geographical aggregation for three of the Swedish regions, 1974. The labels 'inter' and 'all' denote interregional migration and all migration respectively.

Parameters	Region of Sweden							
	Upper East				Lower East			
	males		females		males		females	
	inter	all	inter	all	inter	all	inter	all
R^{GM}	1·76	2·32	1·81	2·40	1·01	1·24	1·07	1·31
$\alpha_1 = \alpha_2$	0·0846	0·0900	0·1042	0·1073	0·1033	0·0958	0·1243	0·1183
λ_2	0·4050	0·4164	0·4643	0·4945	0·4061	0·4485	0·5613	0·5880
x_h	23·74	23·66	21·47	21·40	23·03	22·90	20·97	20·91
x_ℓ	15·82	15·94	14·57	14·77	15·30	15·58	14·94	15·03
X	7·92	7·72	6·90	6·63	7·72	7·32	6·03	5·88
A	28·97	28·97	26·33	26·62	29·63	29·36	27·51	27·30
B	0·0494	0·0689	0·0621	0·0836	0·0343	0·0404	0·0456	0·0525

Parameters	South			
	males		females	
	inter	all	inter	all
R^{GM}	0·88	1·47	0·85	1·47
$\alpha_1 = \alpha_2$	0·1153	0·1086	0·1216	0·1195
λ_2	0·2687	0·3113	0·4482	0·4414
x_h	24·16	23·70	22·48	21·65
x_ℓ	14·53	14·77	15·46	14·60
X	9·63	8·93	7·02	7·05
A	29·74	29·31	28·21	27·38
B	0·0264	0·0426	0·0327	0·0525

Figure 16 shows that for a given value of the parental shift, the labor-force shift, X, varies as a function only of the rate of descent, α, and the rate of ascent, λ. For a given set of values for x_ℓ, x_h, α, and A, it is therefore possible to infer the values of λ and μ. From figure 16, with known values of $x_h - x_\ell$, α, and A, λ can be obtained. With values for λ, α, and A, it is possible to use figure 17 to obtain the values of $x_h - \mu$, and therefore of μ. With values for α, λ, and μ, the profile (but not the level) of a model migration schedule has been defined. To obtain the level, values for a_1, a_2, and c are also needed.

Preliminary empirical explorations indicate that *profile* indices such as the low point, the high peak, and the two shifts are somewhat more 'robust' descriptive measures of regularities in empirical migration schedules than are *level-influenced* indices such as the fundamental model parameters α and λ. Perhaps this is because the former are 'purer' indicators of profile: they do not confound measures reflecting levels (such as, for example, the gross migraproduction rate and the jump) with measures indicating locations along the age axis. This attribute of such profile indices is suggested in table 9, in which both sets of indicators are presented for interregional migration alone and for inter- and intraregional (intercommunal) migration taken together. The results are by no means clear cut, but they do suggest a possibly fruitful direction for further study.

4.4 *Application: graduation and interpolation*

Among the various analytical and practical applications of the model-migration-schedule concept described, the most immediately obvious one is the estimation of single-year-of-age migration rates from data reported only by five-year age intervals. As a by-product of this operation one also obtains the various fundamental parameters and profile indices.

Migration rates for five-year age intervals are weighted linear combinations of the corresponding single-year rates, where each particular weight is the proportion of the population in the five-year age interval that falls within a particular single year of age inside of the interval. For purposes of interpolation, these weights may be derived from an observed age composition, or their values may be assumed to follow those of some 'standard' population composition, for example, a stable population.

Given a range of population weights by single years of age and a set of observed migration rates by five-year age intervals, one may search for the model migration schedule that reproduces best (in the least-squares or minimum-chi-square sense) the set of observed migration rates. Formally the estimation algorithm is precisely the same as before; only the criterion function to be minimized is slightly altered.

Table 10 presents the results of four such graduation-interpolation experiments using the regional data for Sweden. In it are contrasted the goodness-of-fit statistics, parameter estimates, and level and profile indices produced in the course of fitting the model schedule to migration data by single-year and five-year age intervals. The results show that, to a remarkable extent, data by five-year age intervals may be used in place of the generally scarcer migration data by single-year age intervals. Information contained in eighty-five one-year-age-group rates may be inferred quite accurately from seventeen five-year-age-group rates.

Figure 18 illustrates graphically the quality of the fit provided by the graduation-interpolation procedure to data on the migration of males out of Stockholm, and includes, for purposes of contrast, the fit provided to the same input data of a cubic spline (McNeil et al, 1977)[3]. Note that the spline interpolation is less accurate in identifying the retirement peak, and it introduces a break in the curve (an inflection point) at the age of 35.

(3) As an additional indicator of the quality of the fit, we include a plot of the residuals at the bottom of the graph. These may be reduced to zero by further interpolation, but the model schedule then loses its smooth regularities.

Table 10. Graduation interpolation with the simplified model migration schedule: goodness-of-fit and parameter estimates for Swedish males, 1974.

Parameters and statistics	Region of Sweden							
	Stockholm		Upper East		Lower East		South	
	MS[a]	GI[b]	MS[a]	GI[b]	MS[a]	GI[b]	MS[a]	GI[b]
E	7·48	7·66	6·46	6·82	12·45	12·92	11·07	11·12
R^{GM}	1·46	1·45	1·76	1·75	1·01	1·00	0·88	0·87
$\alpha_1 = \alpha_2$	0·0856	0·0824	0·0846	0·0835	0·1033	0·0963	0·1153	0·1100
λ_2	0·3529	0·3747	0·4050	0·4263	0·4061	0·5974	0·2687	0·2870
x_h	24·93	24·92	23·74	23·50	23·03	22·26	24·16	24·04
x_ℓ	16·53	16·78	15·82	15·81	15·30	16·18	14·53	14·69
X	8·40	8·15	7·92	7·69	7·72	6·08	9·63	9·35
A	26·67	26·49	28·97	29·01	29·63	29·37	29·74	29·59
B	0·0290	0·0295	0·0494	0·0505	0·0343	0·0372	0·0264	0·0264

[a] Model schedule fitted to data by single years of age (the Stockholm and Upper East regions are fitted with the nine-parameter model, the other two regions with the six-parameter model).
[b] Graduation interpolation to single years of age by use of data by five-year age groups and the simplified model migration schedule.

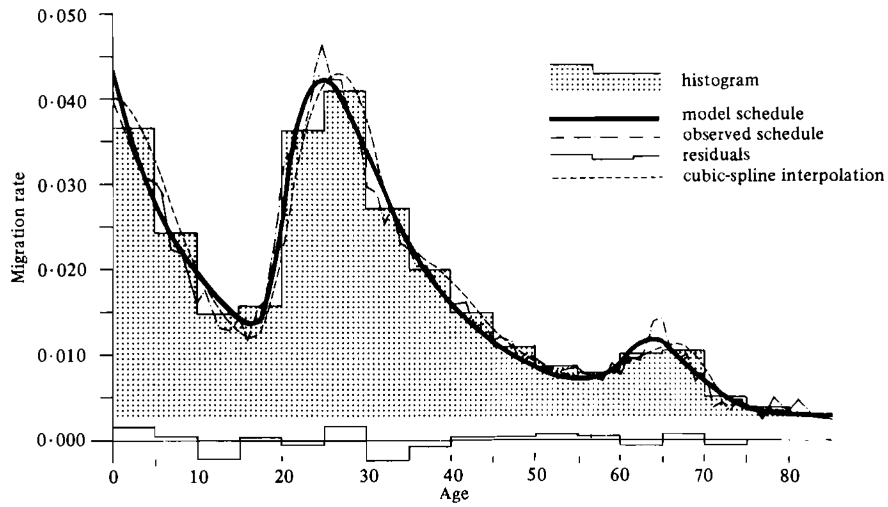


Figure 18. Two alternative graduation interpolations, the cubic spline and the model migration schedule, for males in Stockholm, 1974.

5 Conclusion

In this paper we have examined two alternative approaches for summarizing and exploiting the regularities exhibited by empirical migration schedules: the mortality approach and the fertility approach. In developing both we have elected to generalize and extend the standard paradigms put forward by Coale and his associates, specifically his early work on model mortality schedules (Coale and Demeny, 1966) and his subsequent work on model nuptiality and fertility schedules (Coale, 1971; Coale and McNeil, 1972; Coale and Trussell, 1974). Our initial exploratory efforts are not yet complete, but they do suggest several observations that will guide our future efforts.

Both approaches have their strengths and weaknesses. The strength of the mortality approach lies in its ability to infer migration flows from two consecutive censuses that contain population data disaggregated by age, region of residence, and region of birth. Its weakness is that it leads to a classification of families of schedules that may have little analytical interest.

The strength of the fertility approach lies in its ability to represent the migration schedule in terms of dimensions that are intuitively appealing and analytically robust. The approach may be used to study the fundamental properties of migration schedules, to identify and smooth out errors in observed data, and to interpolate within observed migration rates. Its weakness is that it does not suggest a ready method for inferring migration measures on the basis of distributional data alone. The method requires at least some crude estimates of age-specific interregional flows, and these are hard to come by in most developing and several developed countries.

It may well be that a combination of the two approaches will produce the most useful perspective. Such a perspective will demand the integration of concepts associated with model multiregional life tables with those defining model migration schedules. Fractions of a lifetime lived in a particular region will need to be expressed in terms of shifts, jumps, and migraproduction rates, and vice versa.

Finally, by use of an analogy argument, it seems probable that Coale's model of nuptiality and fertility can be transformed into one of labor-force participation and migration by reinterpreting (1) entry into the marriage market as entry into the job market, (2) marital search as job search, (3) first-marriage frequency as first-job frequency, and (4) proportion ever married as proportion ever active.

The menu for future research is a rich one.

Acknowledgements. The authors gratefully acknowledge the computational assistance of Frans Willekens. In carrying out this research we also have benefited from discussions with Michael Stoto, Ansley Coale, and Frans Willekens. This study could not have been carried out without the generous provision of detailed Swedish population data by Arne Arvidsson of the Swedish Central Bureau of Statistics. Our thanks also go to Larry Long of the United States Census Bureau and Kazimierz Dziewonski and Piotr Korcelli of the Geographical Institute of the Polish Academy of Sciences for providing us with United States and Polish migration data, respectively.

References

Arriaga E, 1968 "New life tables for Latin American populations in the nineteenth and twentieth centuries" Institute of International Studies, University of California, Berkeley, California

Coale A J, 1971 "Age patterns of marriage" *Population Studies* **25** 193-214

Coale A J, 1972 *The Growth and Structure of Human Populations* (Princeton University Press, Princeton, NJ)

Coale A J, Demeny P, 1966 *Regional Model Life Tables and Stable Populations* (Princeton University Press, Princeton, NJ)

Coale A J, McNeil D R, 1972 "The distribution by age of the frequency of first marriage in a female cohort" *Journal of the American Statistical Association* **67** 743-749

Coale A J, Trussell T J, 1974 "Model fertility schedules: variations in the age structure of childbearing in human populations" *Population Index* **40**(2) 185-258

Hoem J M, 1972 "On the statistical theory of analytic graduation" in *Proceedings of the Sixth Berkeley Symposium on Mathematics, Statistics, and Probability, Volume 1* Eds J Neyman, E Fix (University of California Press, Berkeley, Calif.) pp 569-600

Hoem J M, Berge E, 1975 "Some problems in Hadwiger fertility graduation" *Scandinavian Actuarial Journal* **4** 129-144

Keyfitz N, 1968 *Introduction to the Mathematics of Population* (Addison-Wesley, Reading, Mass)

Long L H, 1973 "New estimates of migration expectancy in the United States" *Journal of the American Statistical Association* **68** 37-43

Lowry I S, 1966 *Migration and Metropolitan Growth: Two Analytical Models* (Chandler, San Francisco)

McNeil D R, Trussell T J, Turner J C, 1977 "Spline interpolation of demographic data" *Demography* **14** 245-252

Polish Central Bureau of Statistics, 1974 *Rocznik Demograficzny, 1974, number 33* (Polish Central Bureau of Statistics, Warsaw)
Rele J R, 1967 "Fertility analysis through extension of stable population concepts" Institute of International Studies, University of California, Berkeley, California
Rogers A, 1973 "The multiregional life table" *Journal of Mathematical Sociology* 3(1) 127-137
Rogers A, 1975 *Introduction to Multiregional Mathematical Demography* (John Wiley, New York)
Rogers A, Castro L J, 1976 "Model multiregional life tables and stable populations" RR-76-09, International Institute for Applied Systems Analysis, Laxenburg, Austria
Rogers A, Ledent J, 1976 "Increment-decrement life tables: a comment" *Demography* 13 287-290
Stoto M A, 1977 "On the relationship of childhood to labor force migration rates" RM-77-55, International Institute for Applied Systems Analysis, Laxenburg, Austria
Swedish Central Bureau of Statistics, 1974 *Internal Migration in Sweden 1968-1973, number 9* (Swedish Central Bureau of Statistics, Stockholm)
Tukey J W, 1977 *Exploratory Data Analysis* (Addison-Wesley, Reading, Mass)
United Nations, 1955 *Age and Sex Patterns of Mortality: Model Life Tables for Underdeveloped Countries* (United Nations, New York)
United Nations, 1967 *Methods of Estimating Basic Demographic Measures from Incomplete Data* (United Nations, New York)
Wilber G W, 1963 "Migration expectancy in the United States" *Journal of the American Statistical Association* 58 444-453

Environment and Planning A, 1978, volume 10, pages 503-518

The spatial reproductive value and the spatial momentum of zero population growth†

A Rogers, F Willekens
International Institute for Applied Systems Analysis, Laxenburg, Austria
Received 31 May 1977, in revised form 17 January 1978

Abstract. This paper develops the concept of spatial reproductive value and illustrates how such a notion may be used to trace out quantitatively the geographical impacts of a sudden reduction in fertility to bare replacement level. Such a reduction does not immediately produce zero population growth in populations that previously experienced high birthrates. The built-in momentum for continued growth and its spatial dimension may be assessed with the aid of spatial reproductive values.

1 Introduction

If age-specific birth and death rates remain fixed, a population that is undisturbed by migration will ultimately evolve into a *stable* population that increases at a fixed annual rate, r say. If the birthrates are fixed at bare replacement level, then such a stable population will be a *stationary* or zero-growth population with $r = 0$. In what follows, we explore some of the *spatial* consequences of zero growth by considering how a sudden reduction of fertility to replacement level might affect the spatial evolution of a multiregional population whose constituent regional populations experience the redistributional effects of internal migration.

Mathematical analysis of spatial zero population growth can be facilitated by adopting a spatial generalization of the notion of reproductive value first set out by Fisher (1929, page 27). Keyfitz (1975) has shown how the quantitative impacts of fertility reduction, of sterilization, of mortality, and of emigration (all assumed to take place at a particular age x within a single population) can be assessed by means of reproductive values. Analogous calculations can be carried out for multiple spatially interacting populations with the aid of *spatial* reproductive values, as we demonstrate for the case of fertility reduction to replacement level.

Finally it is well-known that a sudden reduction of fertility to replacement level does not immediately produce zero population growth. Children outnumber parents in a population that maintains high birthrates. Consequently the number of potential parents in the next generation will inevitably be larger than at present, and the current population therefore has a built-in tendency to continue growing before it ultimately stabilizes into a zero-growth condition. This built-in *momentum* may be assessed with the aid of reproductive values, and the concluding parts of this paper illustrate how such a momentum may be accorded a spatial dimension.

2 Spatial zero population growth

Zero population growth for a nation implies a zero growth average for local areas. This of course still allows for the possibility of nonzero growth in particular localities. Thus *spatial* zero growth, like *temporal* zero growth, may be viewed either as a condition that ultimately prevails uniformly over space and time, or one that exists only because of a fortuitous balancing of regional rates of positive growth, of zero

† This paper is a completely revised version of "Spatial Zero Population Growth" which was presented at the Annual Meeting of the Population Association of America, 29 April-1 May 1976, Montreal, Canada.

growth, and of decline. Since no obvious advantages arise from the latter case, it is quite natural to view the attainment of temporal zero growth in the long run in terms of an indefinite continuation of temporal zero growth in the short run. We follow this tradition in this paper and view the attainment of spatial zero growth in the long run in terms of temporal zero growth within each region of a closed multiregional population system. Consequently we confine our attention to the evolution of a particular subset of stationary populations, called *spatial zero-growth populations*, that is, *stable* multiregional populations that have a zero growth rate. Thus we augment the usual twin assumptions of a fixed mortality schedule and a fixed fertility schedule, set at replacement level, with the assumption of a fixed migration schedule. Multiregional populations subjected to such regional growth regimes ultimately assume a persisting zero rate of growth in every region and exhibit zero growth both over time and over space.

2.1 *Conditions for spatial zero population growth*

Stable-population theory informs us that a stationary, say female, population arises out of the combination of the following conditions (Ryder, 1974): (1) a fixed survival function $l(a)$ which specifies the probability that a female survives to age a; (2) a fixed maternity function $m(a)$ which defines the conditional probability that a surviving female gives birth to a baby girl at age a; (3) a product sum of these two functions, the *net reproduction rate*,

$$R(0) = \int_0^\infty m(a)l(a)\,\mathrm{d}a\,,$$

that is equal to unity; and (4) an absence of migration.

If $m(a)$ and $l(a)$ are fixed for a long time, and the population is closed to migration, the number of female births in year t, $B(t)$, is given by the well-known Lotka equation,

$$B(t) = \int_0^\infty m(a)l(a)B(t-a)\,\mathrm{d}a\,.$$

A substitution of $Q\mathrm{e}^{rt}$ for $B(t)$ in the above equation gives

$$Q = \left[\int_0^\infty \mathrm{e}^{-ra}m(a)l(a)\,\mathrm{d}a\right]Q = \Psi(r)Q\,,$$

whose solution is that value of r for which $\Psi(r) = 1$, and where Q denotes the number of stable equivalent births.

For the special case of $r = 0$, then $B(t) = \hat{Q}$ and

$$\hat{Q} = \left[\int_0^\infty \hat{m}(a)l(a)\,\mathrm{d}a\right]\hat{Q} = \hat{R}(0)\hat{Q}\,,$$

where $\hat{m}(a) = \gamma m(a)$, and γ is generally taken to be equal to the reciprocal of the net reproduction rate, $R(0)$. (The circumflex will henceforth be used to distinguish the parameters of a stationary population.)

Let us now relax the last of the four conditions that generate a stationary population—the closure of the population to migration. Imagine a multiregional population whose long-run evolution follows the multiregional Lotka equation (Rogers, 1975, chapter 4)

$$\mathbf{B}(t) = \int_0^\infty \mathbf{m}(a)\mathbf{l}(a)\,\mathbf{B}(t-a)\,\mathrm{d}a\,, \tag{1}$$

where $\mathbf{m}(a)$ is a diagonal matrix of annual regional fertility rates $m_i(a)$, $\mathbf{l}(a)$ is a matrix of place-of-birth (i) by place-of-residence (j) survival probabilities ${}_il_j(a)$, and $B(t)$ is a column vector of regional births $B_i(t)$. A substitution of $Q\mathrm{e}^{rt}$ for $B(t)$ in equation (1) gives

$$Q = \left[\int_0^\infty \mathrm{e}^{-ra}\,\mathbf{m}(a)\mathbf{l}(a)\,\mathrm{d}a\right]Q = \Psi(r)Q\,, \tag{2}$$

whose solution is that value of r for which the dominant characteristic root of $\Psi(r)$ is unity. For the special case of $r = 0$, then $B(t) = \hat{Q}$ and

$$\hat{Q} = \left[\int_0^\infty \hat{\mathbf{m}}(a)\mathbf{l}(a)\,\mathrm{d}a\right]\hat{Q} = \hat{R}(0)\hat{Q}\,. \tag{3}$$

The element in the ith row and the jth column of $\hat{\mathbf{R}}(0)$ is the stationary regional net reproduction rate *in region i* of women born in region j.

$${}_j\hat{R}_i(0) = \int_0^\infty \hat{m}_i(a)_jl_i(a)\,\mathrm{d}a\,.$$

Equation (3) shows that for a spatial zero-growth population to be maintained, the dominant characteristic root of the matrix $\hat{\mathbf{R}}(0)$ must be unity. Consequently a reduction of fertility to replacement level may be interpreted as a reduction of the elements of $\mathbf{m}(a)$ to a level that reduces the dominant characteristic root of a given net reproduction matrix $\mathbf{R}(0)$ to unity. Such an operation transforms $\boldsymbol{m}(a)$ to $\hat{\boldsymbol{m}}(a)$ and $\mathbf{R}(0)$ to $\hat{\mathbf{R}}(0)$.

The column vector $\hat{Q}$ in equation (3) is the characteristic vector associated with the unit dominant characteristic root of $\hat{\mathbf{R}}(0)$, and denotes the total annual number of births in each region of a spatial zero-growth population. The proportional allocation of total births that it defines is directly dependent on the transformation that is applied to change $\mathbf{R}(0)$ to $\hat{\mathbf{R}}(0)$. Since in a spatial zero growth population the regional *stationary* equivalent population $\hat{Y}_i$ is equal to the quotient formed by $\hat{Q}_i$ and the birthrate $\hat{b}_i$, we see that the different ways in which $\mathbf{R}(0)$ is transformed into $\hat{\mathbf{R}}(0)$ become, in fact, alternative 'spatial paths' to a stationary multiregional population.

2.2 *The redistributional impacts of two alternative spatial patterns of fertility reduction*

A multiregional population system that is growing at a positive rate of growth exhibits a net-reproduction matrix $\mathbf{R}(0)$ with a dominant characteristic root $\lambda_1[\mathbf{R}(0)]$ that is greater than unity. If the rate of childbearing in each region of this population system were immediately modified such that every woman *born in that region* would now have a net reproduction rate of unity, then

$${}_i\hat{R}(0) = \sum_{j=1}^{s} {}_i\hat{R}_j(0) = 1\,,$$

where s is the number of regions, or, in matrix form,

$$[\hat{R}(0)]^{\mathrm{T}}\mathit{1} = \mathit{1}\,, \tag{4}$$

where $\mathit{1}$ is the unit column vector.

As is the normal practice in single-region exercises of this kind, assume that the fertility of each regional *cohort* of women is modified through the multiplication of each *region's* age-specific birthrates by a fixed fertility adjustment factor, γ_i say. Then

$${}_i\hat{R}(0) = \sum_{j=1}^{s} {}_i\hat{R}_j(0) = \sum_{j=1}^{s}\int_0^\infty \gamma_j m_j(a)_il_j(a)\,\mathrm{d}a = \sum_{j=1}^{s}\gamma_{j\,i}R_j(0) = 1$$

and

$$\hat{\mathbf{R}}(0) = \gamma \mathbf{R}(0) , \tag{5}$$

where γ is a diagonal matrix of fertility adjustment factors. The substitution of equation (5) into equation (4) gives

$$[R(0)]^{\mathrm{T}} \gamma I = I ,$$

whence

$$\gamma I = \{[R(0)]^{\mathrm{T}}\}^{-1} I ,$$

if the inverse exists.

The setting of the fertility of each female cohort in every region to bare replacement level, *the cohort-replacement alternative*, is but one of many possible spatial patterns of fertility reduction. One could instead, for example, consider a fertility-reduction scheme in which the aggregate system-wide net reproduction rate is reduced to unity through the multiplication of *all* age-specific birthrates by the *same* fertility adjustment factor, γ say. That is, let

$$\hat{\hat{\mathbf{R}}}(0) = \gamma \mathbf{R}(0) , \tag{6}$$

where $\gamma = 1/\lambda_1[\mathbf{R}(0)]$. This particular spatial pattern of fertility reduction may be called *the proportional-reduction alternative*, and its redistributional impacts may be quite different from those of the cohort-replacement alternative.

A numerical illustration may be instructive at this point. Assume that the net reproduction behavior of the urban and rural female populations of a national population are approximated by the net-reproduction matrix

$$\mathbf{R}(0) = \begin{bmatrix} {}_{\mathrm{u}}R_{\mathrm{u}}(0) & {}_{\mathrm{r}}R_{\mathrm{u}}(0) \\ {}_{\mathrm{u}}R_{\mathrm{r}}(0) & {}_{\mathrm{r}}R_{\mathrm{r}}(0) \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{2} \\ \frac{1}{4} & 1 \end{bmatrix} , \tag{7}$$

where, for example, ${}_{\mathrm{r}}R_{\mathrm{u}}(0)$ denotes the net reproduction rate in urban areas of rural-born women. In other words, under the observed regime of growth, each woman born in rural areas will, on the average, replace herself in the succeeding generation by $1\frac{1}{2}$ daughters, one third of whom will be born in urban areas. Urban-born women, on the other hand, have a lower net reproduction rate, ${}_{\mathrm{u}}R(0) = 1 < {}_{\mathrm{r}}R(0)$, which when combined with the net reproduction rate of rural-born women gives the national female population an overall net reproduction rate of $\lambda_1[\mathbf{R}(0)] = 1\frac{1}{4}$, where $\lambda_1[\mathbf{R}(0)]$ is the dominant characteristic root of the net-reproduction matrix in equation (7).

An immediate system-wide fertility decline to replacement level, such that each urban- or rural-born woman is followed in the next generation by exactly one daughter, implies that $\gamma_{\mathrm{u}} = \frac{6}{5}$ and $\gamma_{\mathrm{r}} = \frac{2}{5}$, whence

$$\hat{\mathbf{R}}(0) = \begin{bmatrix} \frac{9}{10} & \frac{3}{5} \\ \frac{1}{10} & \frac{2}{5} \end{bmatrix} \quad \text{and} \quad \hat{Q} = \begin{bmatrix} 6 \\ 1 \end{bmatrix} .$$

Both groups of women now exhibit unit rates of net reproduction; the dominant characteristic root of $\hat{\mathbf{R}}(0)$ is unity; and the characteristic vector associated with the unit dominant characteristic root of $\hat{\mathbf{R}}(0)$ indicates that $\frac{6}{7}$ of the total births in the spatial zero-growth population will occur in urban areas.

Consider next the redistributional implications of a fertility decline according to the proportional-reduction alternative defined in equation (6). In this instance $\gamma = \frac{4}{5}$, whence

$$\hat{\hat{\mathbf{R}}}(0) = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \quad \text{and} \quad \hat{\hat{Q}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

The fertility of urban-born women now declines to below replacement level, ${}_{\mathrm{u}}\hat{R}(0) = \frac{4}{5}$, whereas that of rural-born women continues to exceed such a level, ${}_{\mathrm{r}}\hat{R}(0) = \frac{6}{5}$. The dominant characteristic root of $\hat{\hat{\boldsymbol{R}}}(0)$ is unity, and urban and rural areas have an equal number of births.

3 The spatial reproductive value

The concept of reproductive value, as developed by Fisher (1929), revolves around the notion of regarding the offspring of a child as the repayment of a debt. Specifically if the birth of a baby is viewed as a loan of a life and if the future offspring of this child are viewed as the subsequent repayment of this loan, suitably discounted at the annual rate r and compounded momently, then the present value of the repayment may be taken to be

$$\int_0^\infty \mathrm{e}^{-ra} m(a) l(a)\, \mathrm{d}a \,.$$

Equating the loan with the discounted repayment, one gets

$$1 = \int_0^\infty \mathrm{e}^{-ra} m(a) l(a)\, \mathrm{d}a \,,$$

which is recognizable as the characteristic equation used to solve for r, the intrinsic rate of growth. Thus, as Keyfitz points out,

> "the equation can now be seen in a new light: the equating of loan and discounted repayment is what determines r, r being interpretable either as the rate of interest of an average loan or as Lotka's intrinsic rate of natural increase" (Keyfitz, 1975, page 588).

In the same paper, Keyfitz considers how much of the debt is outstanding by the time the child has reached the age x. He defines this quantity to be $v(x)$, the reproductive value at age x, where

$$v(x) = \int_x^\infty \mathrm{e}^{-r(a-x)} m(a) \frac{l(a)}{l(x)}\, \mathrm{d}a \,, \tag{8}$$

and $v(0)$ is scaled to equal unity.

3.1 *Definitions and theory*

Keyfitz's arguments have their spatial (multiregional) counterparts. To develop these it is convenient to reexpress equation (8) for arbitrary values of $v(0)$, that is,

$$v(x) = v(0) \int_x^\infty \mathrm{e}^{-r(a-x)} m(a) \frac{l(a)}{l(x)}\, \mathrm{d}a = v(0) n(x) \,,$$

where

$$v(0) = v(0) \int_0^\infty \mathrm{e}^{-ra} m(a) l(a)\, \mathrm{d}a = v(0) \Psi(r)$$

and $n(x)$ denotes the total discounted number of baby girls expected to be born to a woman now aged x. This form of the equation immediately suggests the multiregional analog, to give the (row) vector $[\boldsymbol{v}(x)]^{\mathrm{T}}$,

$$[\boldsymbol{v}(x)]^{\mathrm{T}} = [\boldsymbol{v}(0)]^{\mathrm{T}} \int_x^\infty \mathrm{e}^{-r(a-x)} \mathbf{m}(a) \mathbf{l}(a) [\mathbf{l}(x)]^{-1}\, \mathrm{d}a \tag{9}$$

$$= [\boldsymbol{v}(0)]^{\mathrm{T}} \mathbf{n}(x) \,, \tag{10}$$

where

$$[\boldsymbol{v}(0)]^{\mathrm{T}} = [\boldsymbol{v}(0)]^{\mathrm{T}} \int_0^\infty \mathrm{e}^{-ra} \mathbf{m}(a) \mathbf{l}(a)\, \mathrm{d}a = [\boldsymbol{v}(0)]^{\mathrm{T}} \boldsymbol{\Psi}(r) \,. \tag{11}$$

The matrix $\mathbf{n}(x)$ represents the expected total number of female offspring per woman at age x, discounted back to age x. The element $n_{ij}(x)$ gives the discounted number of female children to be born in region j to a woman now x years of age and a resident of region i. The vector $\boldsymbol{v}(x)$ represents the reproductive values of x-year-old women, differentiated by region of residence. Observe that the elements of $\boldsymbol{v}(x)$ depend on the scaling given to $\boldsymbol{v}(0)$, the *left* characteristic vector associated with the unit dominant characteristic root of the characteristic matrix $\boldsymbol{\Psi}(r)$[1]. Thus in the multiregional model, the reproductive value of a baby girl depends on where she is born.

Equations (9), (10), and (11) may be given the following demographic interpretation. If lives are loaned to regions according to the (column) vector $\boldsymbol{Q}$ then the amount of 'debt' outstanding x years later is given by the (row) vector $[\boldsymbol{v}(x)]^{\mathrm{T}}$, the regional expected values of subsequent offspring discounted back to age x. The elements of this vector therefore may be viewed as spatial (regional) reproductive values at age x.

A slightly modified perspective of the spatial reproductive value is adopted in this paper. Specifically we shall distinguish between the terms *number* and *value* when referring to births in the various regions of a closed multiregional system. The two expressions have identical meanings in the single-region model, but variations in regional fertility and mortality schedules give them different meanings in any multiregional model in which internal migration is represented.

Recall the definition of the multiregional characteristic matrix $\boldsymbol{\Psi}(r)$ in equation (2). An element ${}_i\Psi_j(r)$ denotes the discounted total *number* of daughters born in region j to a mother born in region i. The discounted number of female births per woman born in a particular region, ${}_i\Psi(r) = \sum\limits_j {}_i\Psi_j(r)$, may be less than unity. Although this suggests that she does not repay the full amount of her 'debt' to society, we shall show that this may be true only of *number* but not of *value*.

Consider society as an investor in a multiregional (spatial) portfolio of lives, each element of the portfolio being a region. The society distributes its investment in lives among its regions according to the vector $\boldsymbol{Q}$. The total discounted *number* of offspring attributable to the Q_i *individuals* born in region i is

$${}_i\Psi(r)Q_i = \sum_j {}_i\Psi_j(r)Q_i ,$$

out of which total ${}_i\Psi_j(r)Q_i$ will be born in region j. The total societal investment in *region i*, however, is Q_i births, which by virtue of equation (2) may be expressed as

$$Q_i = \sum_j {}_j\Psi_i(r)Q_j .$$

Hence for each region, during stable growth, the present discounted number of future births in that region must be equal to the current number of births.

The distribution of total societal births among regions is one of two related aspects of the reproductive-value problem. Associated with this *primal* aspect of *allocation* is a *dual* aspect of *valuation*[2].

The *value* of a birth reflects the capacity to produce new life. If a 0-year old in region i is worth $v_i(0)$, then, by equation (11), the reproductive value of the discounted number of offspring must also be $v_i(0)$, that is

$$v_i(0) = \sum_j v_j(0)_i\Psi_j(r) .$$

[1] Recall that in equation (2) the corresponding *right* characteristic vector of $\boldsymbol{\Psi}(r)$ was $\boldsymbol{Q}$.

[2] This primal-dual relationship resembles the one found in mathematical-programming theory and suggests the conjecture that the primal optimization problem is one of selecting an allocation of births to maximize growth, within the constraints of a given regime of fertility, mortality, and migration; and the dual optimization problem is one of valuing births in each region so as to minimize the total regionally weighted societal reproductive value.

Thus if the investment in one life in a region is viewed as a debt of an individual to society then in a stable equilibrium each individual must repay that debt to society at an annual interest rate r. The repayment does not have to take place in the region of birth, however. Part of it can occur in other regions, where births may be *worth* more (or less) than in the region of birth. Thus we may conclude that individuals pay back their debt to society in *values* $v(0)$, whereas regions pay back their debt in *numbers* Q. The former distribution is defined by equation (11); the latter derives from equation (2).

Spatial reproductive values at age x, $v_i(x)$, may be appropriately consolidated to yield *total* spatial reproductive values, v_i, by means of the relationship

$$\begin{aligned} \mathbf{v}^{\mathrm{T}} &= \int_0^{\infty} [\mathbf{v}(x)]^{\mathrm{T}}\, \mathbf{k}(x)\, \mathrm{d}x \\ &= [\mathbf{v}(0)]^{\mathrm{T}} \int_0^{\infty} \mathbf{n}(x)\, \mathbf{k}(x)\, \mathrm{d}x \\ &= [\mathbf{v}(0)]^{\mathrm{T}} \mathbf{n}\, , \end{aligned}$$

where $\mathbf{k}(x)$ is a diagonal matrix with $k_{ii}(x)$ representing the number of women at age x in region i, and $\mathbf{n}$ is a matrix of total discounted number of female offspring associated with that population. The total reproductive *value* of the multiregional population then is

$$v = V^{\mathrm{T}} I\, .$$

3.2 *Numerical evaluation*

The definition of $\mathbf{v}(x)$ in equation (9) refers to exact age x. A suitable numerical approximation, analogous to the one usually made in the single-region model, is

$$[\mathbf{v}(x)]^{\mathrm{T}} = [\mathbf{v}(0)]^{\mathrm{T}} \left[\sum_{a=x}^{\beta-5} \mathrm{e}^{-r(a+\frac{5}{2}-x)} \mathbf{M}(a)\, \mathbf{L}(a) \right] [\mathbf{l}(x)]^{-1}\, ,$$

where β is the last age of childbearing, $\mathbf{M}(a)$ and $\mathbf{L}(a)$ are respectively matrices of fertility rates and life-table populations of women aged a to $a+4$; and $\mathbf{l}(x)$ is the life-table survival matrix to exact age x (see Rogers, 1975).

The *average* spatial reproductive value for the age interval x to $x+4$ at last birthday is denoted by $V(x)$ and may be approximated by

$$\begin{aligned} [V(x)]^{\mathrm{T}} &= [\mathbf{v}(0)]^{\mathrm{T}} \tfrac{5}{2} \left[\sum_{a=x}^{\beta-5} \mathrm{e}^{-r(a-x)} \mathbf{M}(a)\, \mathbf{L}(a) + \mathrm{e}^{-r(a+5-x)} \mathbf{M}(a+5)\, \mathbf{L}(a+5) \right] [\mathbf{L}(x)]^{-1} \\ &= [\mathbf{v}(0)]^{\mathrm{T}} \mathbf{N}(x)\, , \end{aligned} \qquad (12)$$

where $\mathbf{N}(x)$ is the *average* value of $\mathbf{n}(x)$ for the age interval x to $x+4$ at last birthday(3).

Table 1 presents the values of $\mathbf{N}(x)$ for $x = 0, 5, ..., 50$, using the 1961 population data on Yugoslavian females that is published in Rogers (1975). For example, the discounted ($r = 0 \cdot 006099$) number of daughters expected to be born to a woman now living in Slovenia and 15 to 19 years old is 1·0078. Of this total 0·9417 will be born in Slovenia and 0·0661 will be born in the rest of Yugoslavia. A woman in the same age group in the rest of Yugoslavia has an expected discounted number of daughters of 1·1943, of which only 0·0068 will be born in Slovenia. This is mainly a reflection of the low level of migration from the rest of Yugoslavia to Slovenia and the low fertility in Slovenia.

(3) Equation (12) may be shown to be consistent with a somewhat different formulation set out as equation (4.37) on page 105 of Rogers (1975).

The characteristic matrix for this two-region system is

$$\Psi(0\cdot006099) = \begin{bmatrix} 0\cdot813587 & 0\cdot009081 \\ 0\cdot103594 & 0\cdot994840 \end{bmatrix};$$
$$\qquad\qquad\quad \overline{0\cdot917181} \quad \overline{1\cdot003921}$$

its dominant characteristic root is unity; and the associated left and right characteristic vectors are respectively

$$[v(0)]^{\mathrm{T}} = [1 \quad 1\cdot798369] \qquad \text{and} \qquad Q = \begin{bmatrix} 1 \\ 20\cdot515385 \end{bmatrix}.$$

Note that the discounted number of female offspring of a baby girl born in Slovenia is less than unity. Nevertheless she still repays her debt of a life to society because the 0·1036 daughters born to her in the rest of Yugoslavia have a higher value than an equivalent number born in Slovenia. The weighted discounted repayment is a single life:

$$1 = (1 \times 0\cdot813587)+(1\cdot798369 \times 0\cdot103594).$$

The combination of the numerical approximation of $\mathbf{N}(x)$ with that of $[v(0)]^{\mathrm{T}}$, as set out in equation (12), gives the set of values for $[V(x)]^{\mathrm{T}}$ in table 2. These indicate that the spatial reproductive values of Slovenian girls are, at most ages, roughly half of the corresponding values for girls living in the rest of Yugoslavia.

Finally, by weighting the age-specific values of $\mathbf{N}(x)$ in table 1 by the respective observed populations and adding, one gets the total discounted number of female offspring $\mathbf{N}$. Table 3 shows that under the 1961 regime of fertility, mortality, and migration, the total discounted *number* of daughters to be born to Yugoslavia's 1961 female population is 5528742[4]. Of these, 383133 (or 6·93%) will be born in Slovenia and 379208 (or 6·86%) will be children of the observed 1961 female *residents* of Slovenia. Of the ultimate discounted 383133 female births in Slovenia, 30404 can be attributed to women now residing in the rest of Yugoslavia and 352729 to potential mothers now living in Slovenia.

Table 1. Number of daughters born to females aged x to $x+4$ at last birthday, discounted back to age x to $x+4$, by region of birth of offspring and region of residence of the mother, for Yugoslavia, 1961.

Age group	Region of residence: Slovenia			Region of residence: the rest of Yugoslavia		
	region of birth of offspring			region of birth of offspring		
	total	Slovenia	the rest of Yugoslavia	total	Slovenia	the rest of Yugoslavia
0-4	0·944564	0·844573	0·099990	1·076658	0·009171	1·067487
5-9	0·987156	0·896680	0·090476	1·178232	0·009015	1·169217
10-14	1·017503	0·934679	0·082823	1·218201	0·008571	1·209630
15-19	1·007809	0·941719	0·066090	1·194264	0·006835	1·187429
20-24	0·819362	0·785326	0·034036	0·949871	0·003220	0·946651
25-29	0·503909	0·492094	0·011815	0·572511	0·000986	0·571526
30-34	0·254250	0·251221	0·003029	0·291552	0·000263	0·291290
35-39	0·099458	0·098904	0·000553	0·129433	0·000047	0·129386
40-44	0·024255	0·024192	0·000063	0·043597	0·000004	0·043593
45-49	0·003167	0·003156	0·000010	0·008752	0·000001	0·008751
50-54	0·000730	0·000730	0·000000	0·001785	0·000000	0·001785

[4] The slight discrepancy between this total and the one reported on page 114 of Rogers (1975) may be attributed to differences in computer hardware.

To derive the total reproductive *value* of the observed female population, one must weight the discounted *number* of offspring according to region of birth. If a value of unity is assigned to a birth in Slovenia (region 1) then 1·798369 is the corresponding value of a birth in the rest of Yugoslavia (region 2). The total reproductive value of Slovenian women is (table 3)

$$(352729 \times 1)+(26479 \times 1{\cdot}798369) = 400347\ ,$$

and the corresponding value for women residing in the rest of Yugoslavia is

$$(30404 \times 1)+(5119130 \times 1{\cdot}798369) = 9236491\ .$$

Adding the two subtotals together gives the aggregate system-wide total reproductive value,

$$V = 400347+9236491 = 9636838\ ,$$

for the case where $v_1(0)$ is set equal to unity.

Table 2. Spatial reproductive value of females aged x to $x+4$ at last birthday, by region of residence for Yugoslavia, 1961.

Age group	Region of residence		Age group	Region of residence	
	Slovenia	the rest of Yugoslavia		Slovenia	the rest of Yugoslavia
0-4	1·024392	1·928907	30-34	0·256669	0·524109
5-9	1·059390	2·111699	35-39	0·099899	0·232731
10-14	1·083626	2·183932	40-44	0·024305	0·078400
15-19	1·060574	2·142272	45-49	0·003175	0·015738
20-24	0·846536	1·705649	50-54	0·000730	0·003210
25-29	0·513342	1·028800			

Table 3. Total discounted number of daughters to the observed female population, by region of birth and residence, for Yugoslavia, 1961.

Region of birth of daughter	Region of residence of mother		
	Slovenia	the rest of Yugoslavia	Total
Slovenia	352729	30404	383133
The rest of Yugoslavia	26479	5119130	5145609
Total	379208	5149534	5528742

3.3 *Stable-population analysis*

The reproductive value may be used to establish the extent to which an individual of a given age will, on the average, contribute to the births of future generations. If κ is the mean age of childbearing in a stable population then

> "... without any change in birth rates the ultimate population birth trajectory due to $P(x)\,\mathrm{d}x$ persons at age x to $x+\mathrm{d}x$ would be $\mathrm{e}^{rt}P(x)v(x)\,\mathrm{d}x/\kappa$, and for the whole population distributed as $P(x)$ would be $\mathrm{e}^{rt}\int_0^\beta P(x)v(x)\,\mathrm{d}x/\kappa$" (Keyfitz, 1975, page 606).

This holds true for a single-region population that is closed to migration. It can be shown that the corresponding result for a multiregional population system is (Willekens, 1977)

$$Q^{(t)} = \mathrm{e}^{rt}\left[\int_0^\infty [v_1(x)]^{\mathrm{T}}k(x)\,\mathrm{d}x\right]Q_1\Big/[v_1(0)]^{\mathrm{T}}\kappa Q_1\ , \tag{13}$$

where we denote the initial population by $k(x)$ instead of $P(x)$ to maintain consistency with our earlier notation; where the unit subscripts designate vectors with *arbitrary* scalings (we shall adopt a scaling that sums the elements of each vector to unity); and where $\boldsymbol{\kappa}$, the matrix of mean ages of childbearing in the stable population, is defined by

$$\boldsymbol{\kappa} = \mathbf{R}^{(r)}(1)[\mathbf{R}^{(r)}(0)]^{-1} ,$$

with

$$\mathbf{R}^{(r)}(\epsilon) = \int_0^{\infty} x^{\epsilon} e^{-rx}\, \mathbf{m}(x)\mathbf{l}(x)\, dx , \qquad \epsilon = 0, 1 .$$

Rewriting equation (13) as

$$Q^{(t)} = \frac{v\, e^{rt} Q_1}{[v_1(0)]^{\mathrm{T}} \boldsymbol{\kappa} Q_1} = \frac{v}{\kappa} e^{rt} Q_1 ,$$

we immediately obtain an expression for stable equivalent births,

$$Q = \frac{v Q_1}{[v(0)]^{\mathrm{T}} \boldsymbol{\kappa} Q_1} = \frac{v}{\kappa} Q_1 , \qquad (14)$$

where

$$\kappa = [v_1(0)]^{\mathrm{T}} \boldsymbol{\kappa} Q_1 .$$

Goodman (1969) proves that the impact on the ultimate stable birth trajectory of a girl at age x is $v(x)e^{rt}/\kappa$. The preceding results indicate that the size of this impact depends on the spatial reproductive value of the girl and therefore on her initial region of residence. The spatial proportional *distribution* of these births, however, is independent of the initial distribution of the population and is therefore not a function of her initial region of residence.

The ultimate birth trajectory, as given by equation (13), depends on the intrinsic annual rate of growth r, the total reproductive value of the population v and its proportional distribution among the regions $[v_1(0)]^{\mathrm{T}}$, the matrix of mean ages of childbearing in the stable population $\boldsymbol{\kappa}$, and the proportional distribution of births in the stable population Q_1. For the earlier two-region illustration involving Slovenia and the rest of Yugoslavia, the following are numerical approximations for these values:

$$r = 0{\cdot}006099 , \qquad [v_1(0)]^{\mathrm{T}} = [0{\cdot}357351 \quad 0{\cdot}642649] ,$$

$$\boldsymbol{\kappa} = \begin{bmatrix} 27{\cdot}309929 & 0{\cdot}016457 \\ 0{\cdot}244865 & 27{\cdot}104288 \end{bmatrix}, \qquad \text{and} \qquad Q_1 = \begin{bmatrix} 0{\cdot}046478 \\ 0{\cdot}953522 \end{bmatrix} .$$

The substitution of these values into equation (14) gives

$$Q = \begin{bmatrix} 9374 \\ 192304 \end{bmatrix},$$

a result that can be verified by calculating the regional stable equivalent populations Y, defined in Rogers (1975), and then using the relation (Willekens, 1977)

$$Q = \mathbf{b}Y = \left[\int_0^{\infty} e^{-rx}\, \mathbf{m}(x)\mathbf{l}(x)\, dx\right] \left[\int_0^{\infty} e^{-rx}\, \mathbf{l}(x)\, dx\right]^{-1} Y = \left[\int_0^{\infty} e^{-rx}\, \mathbf{l}(x)\, dx\right]^{-1} Y .$$

The reader can verify that the earlier scalings of $[v(0)]^{\mathrm{T}}$ and Q in section 3.2 give an identical numerical result for stable equivalent births.

4 The spatial momentum of zero population growth

Differences between observed population age compositions and those of stationary populations make immediate zero growth for most national populations an unlikely condition. A closed population's birthrate and growth rate depend on its fertility schedule and its age composition. Consequently whether and how long a population continues to grow after achieving a net reproduction rate of unity depends on the age composition of that population and its degree of divergence from that of a stationary population. The ratio by which the ultimate stationary population exceeds the current population is the *momentum* of that population.

4.1 *The ultimate size and distribution of a stationary spatial population*

If fertility were to drop immediately to replacement level in a population that is closed to migration, the ultimate stationary number of births in the resulting zero-growth population would be (Keyfitz, 1975)

$$\hat{Q} = \frac{1}{\mu}\int_0^\infty \hat{v}(x)k(x)\,dx = \frac{\hat{v}}{\mu}\,, \tag{15}$$

where μ is the mean age of childbearing in the stationary population, and $\hat{v}(x)$ is the reproductive value corresponding to an intrinsic rate of growth $r = 0$, a condition we can ensure by reducing fertility to replacement level along the lines described in section 2 of this paper. The corresponding ultimate stationary total population may be found by dividing $\hat{Q}$ by the stationary birthrate $\hat{b}$ or equivalently by multiplying it by $e(0)$, the expectation of life at birth:

$$\hat{Y} = \frac{\hat{Q}}{\hat{b}} = e(0)\hat{Q}\,.$$

Such a calculation gives the same result as a full population projection carried out with the modified fertility schedule $\hat{m}(a)$.

The above results have their spatial (multiregional) counterparts. To develop these it is convenient first to recall equation (13) and to define $[\hat{v}(x)]^{\mathrm{T}}$ to be the vector of spatial reproductive values corresponding to an intrinsic rate of growth $r = 0$. (We have seen earlier that a transition to zero growth may be carried out by multiplying the fertility schedule $\boldsymbol{m}(a)$ by the fertility adjustment matrix $\boldsymbol{\gamma}$.) Then the ultimate number of stationary equivalent births must be

$$\hat{Q} = \left[\int_0^\infty [\hat{v}_1(x)]^{\mathrm{T}}\boldsymbol{k}(x)\,dx\right]\hat{Q}_1\Big/[\hat{v}_1(0)]^{\mathrm{T}}\hat{\boldsymbol{\kappa}}\hat{Q}_1 = \frac{\hat{v}\hat{Q}_1}{[\hat{v}_1(0)]^{\mathrm{T}}\boldsymbol{\mu}\hat{Q}_1} = \frac{\hat{v}}{\mu}\hat{Q}_1\,, \tag{16}$$

where $[\hat{v}_1(0)]^{\mathrm{T}}$ and $\hat{Q}_1$ are respectively the left and right characteristic vectors associated with the unit dominant characteristic root of $\boldsymbol{\gamma}\mathbf{R}(0)$, and where

$$\boldsymbol{\mu} = \hat{\boldsymbol{\kappa}} = \boldsymbol{\gamma}\mathbf{R}(1)[\mathbf{R}(0)]^{-1}\boldsymbol{\gamma}^{-1}$$

is the matrix of mean ages of childbearing in the stationary population that evolves after the decline of fertility to replacement level.

The ultimate total stationary population is

$$\hat{Y} = \hat{\mathbf{b}}^{-1}\hat{Q} = \mathbf{e}(0)\hat{Q}\,, \tag{17}$$

where

$$\hat{\mathbf{b}} = \left[\int_0^\infty \hat{\mathbf{m}}(a)\mathbf{l}(a)\,da\right]\left[\int_0^\infty \mathbf{l}(a)\,da\right]^{-1} = \boldsymbol{\gamma}\mathbf{R}(0)[\mathbf{e}(0)]^{-1}\,,$$

and $\mathbf{e}(0)$ is a matrix of expectations of life at birth, disaggregated by regions of birth and residence.

Equation (16) has a simple and intuitively appealing interpretation. Consistent with equation (15), it defines the total size of stationary equivalent births in a multiregional population to be equal to the quotient of the total reproductive value $\hat{v}$ and the weighted index $\mu = [\hat{v}_1(0)]^{\mathrm{T}}\mu\hat{Q}_1$ in that population, both evaluated after the decline in fertility to replacement level, and distributes that total according to the proportional allocation determined by the right characteristic vector associated with the unit dominant characteristic root of the modified net-reproduction-rate matrix $\gamma\mathbf{R}(0)$. The interpretation of equation (17) follows in a straightforward manner.

4.2 *The spatial momentum of an initially stable population*

An abrupt decline in fertility to bare replacement level in a single-region population that initially is experiencing stable growth leads to a computationally simpler form for equation (15). Keyfitz (1975) shows that in such an instance the ratio of the ultimate *stationary* population $\hat{Y}$ to the *stable* population of K individuals just prior to the decline in fertility is

$$\frac{\hat{Y}}{K} = \frac{be(0)}{\mu r}\left(\frac{R(0)-1}{R(0)}\right), \tag{18}$$

where b is the birth rate, r the rate of growth, $e(0)$ the expectation of life at birth, and $R(0)$ the net reproduction rate, all measured before the drop in fertility, and where μ is the mean age of childbearing afterward. Expressing equation (18) as

$$\hat{Y} = e(0)\hat{Q}, \tag{19}$$

one gets

$$\hat{Q} = \frac{bK}{\mu r}\left(\frac{R(0)-1}{R(0)}\right) = \frac{1}{\mu r}\gamma[R(0)-1]Q, \tag{20}$$

where $\gamma = 1/R(0)$ and $Q = bK$. Note that this formula for total stationary equivalent births does not require the calculation of the total reproductive value of the population; but by virtue of equation (15) it implies that, in this special situation,

$$\hat{v} = \frac{\gamma}{r}[R(0)-1]Q.$$

Conventional methods of population projection may be used to obtain the future population that evolves from any particular observed or hypothetical regime of growth. Therefore equation (18) is not needed to obtain a numerical estimate of an ultimate stationary population. However, Keyfitz's simple momentum formula gives us an understanding of the population dynamics that are hidden in the arithmetical computations of a population projection. It identifies in an unambiguous way the five parameters of a current population that determine the size of the future population.

An analogous simplification of equation (16) may be obtained in the multiregional model. If $\boldsymbol{k}(x)$ is stable, then entering

$$\boldsymbol{k}(x) = \mathrm{e}^{-rx}\mathbf{l}(x)Q$$

into equation (16), and multiplying by $\mathbf{e}(0)$ to produce the stationary equivalent population instead of stationary equivalent births, one gets

$$\begin{aligned}\hat{\boldsymbol{Y}} &= \frac{1}{\mu}\mathbf{e}(0)\left[\int_0^\infty [\hat{v}_1(x)]^{\mathrm{T}}\,\mathrm{e}^{-rx}\,\mathbf{l}(x)Q\,\mathrm{d}x\right]\hat{Q}_1 \\ &= \frac{1}{\mu}\mathbf{e}(0)\left[[\hat{v}_1(0)]^{\mathrm{T}}\gamma\int_0^\infty\int_x^\infty \mathrm{e}^{-rx}\mathbf{m}(a)\mathbf{l}(a)\,\mathrm{d}a\,\mathrm{d}xQ\right]\hat{Q}_1\,.\end{aligned}$$

Then, evaluating the double integral and simplifying, one obtains

$$\hat{Y} = \mathrm{e}(0)\frac{1}{\mu r}[\hat{\boldsymbol{v}}_1(0)]^{\mathrm{T}}\boldsymbol{\gamma}[\mathbf{R}(0) - \boldsymbol{\Psi}(r)]\boldsymbol{Q}\hat{\boldsymbol{Q}}_1 \tag{21}$$

$$= \mathrm{e}(0)\hat{Q}\hat{\boldsymbol{Q}}_1 , \tag{22}$$

where

$$\hat{Q} = \frac{1}{\mu r}[\hat{\boldsymbol{v}}_1(0)]^{\mathrm{T}}\boldsymbol{\gamma}[\mathbf{R}(0) - \boldsymbol{\Psi}(r)]\boldsymbol{Q} .$$

As with equation (20) this formula does not require the calculation of the total reproductive value of the population, but it implies that

$$\hat{v} = \frac{1}{r}[\hat{\boldsymbol{v}}_1(0)]^{\mathrm{T}}\boldsymbol{\gamma}[\mathbf{R}(0) - \boldsymbol{\Psi}(r)]\boldsymbol{Q} .$$

Equation (21) is not as practically useful as its single-region counterpart because it is much more difficult to come up with accurate guesses or estimates of the values taken on by the many parameters. Thus a more effective procedure may be to first estimate the ultimate size of the total stationary equivalent births $\hat{Q}$, by means of equation (20); then to distribute that total among the various regions according to the allocation defined by the characteristic vector associated with the unit root of $\boldsymbol{\gamma}\mathbf{R}(0)$; and finally to premultiply the resulting vector by e(0) to find $\hat{\boldsymbol{Y}}$. We shall now illustrate such a procedure with a numerical example using data for India.

4.3 *The urbanization momentum in India*

We have shown that the geographical distribution of a spatial zero-growth population depends very fundamentally on three matrices: e(0), $\mathbf{R}(0)$, and $\boldsymbol{\gamma}$. The first describes the multiregional levels of mortality and migration; the second sets out the multiregional net reproduction patterns before the decline in fertility; and the third defines the particular 'spatial path' by which fertility is reduced. The product $\boldsymbol{\gamma}\mathbf{R}(0)$ gives $\hat{\mathbf{R}}(0)$, whose characteristic vector associated with the unit root and scaled to sum to $\hat{Q}$ is $\hat{\boldsymbol{Q}}$.

Equation (22) also may be used to illustrate dramatically that *where* people choose to live in the future presents issues and problems that are potentially as serious as those posed by the number of children they choose to have. Consider, for example, the projection to zero growth of India's population that was recently carried out by Ryder on the basis of the following assumptions.

> "To simplify the task of projecting the population of India, we make the following assumptions: it is a stable population with a growth rate $r = +0{\cdot}025$ and survival functions corresponding to those labelled "West, level 13" (for which the female and male expectations of life at birth are 50 and 47·114, respectively) in the Coale/Demeny collection; the mean age of (gross) maternity $m = 29$; the ratio of male to female births $k = 1{\cdot}05$; and the current population size is 600 million" (Ryder, 1974, page 6).

From these assumptions it follows that the initial number of female births per annum is $B(t) = 12{\cdot}156$ million, that $R(0) = 2{\cdot}019$, and that $\mu = 28{\cdot}672$. By applying equation (19), Ryder finds a $\hat{Q}$ of 8·558 million for females and a zero-growth total (males plus females) population of approximately 851 million. He then shows that if survival levels in India eventually rise to $e(0) = 70$ for females and $e(0) = 66{\cdot}023$ for males, and

> "if replacement level fertility takes 40 years to achieve and the mean age of gross reproduction declines from 29 to 27, the ultimate female birth cohort size will be

... 15·029 million. Given that value, ... the consequent ultimate population size is 2·094 billion" (Ryder, 1974, page 7).

Ryder concludes that the thought of a population of 2·1 billion for India is staggering and goes on to examine in what respects the components of his projection may be modifiable.

There is no question but that the thought of a 2·1 billion population for India is staggering. What is even more mind-boggling, however, is that anywhere from one to two thirds of this total is likely to eventually be found in that nation's already teeming and over-congested urban areas (the current figure is twenty percent). To show this we need only to introduce a few additional assumptions and then apply equation (22). Specifically, assume that life expectancy in India today is 57 years in urban areas and 48 years in rural areas, with the migration pattern being such that[5]

$$\mathbf{e}(0) = \begin{bmatrix} {}_{\mathrm{u}}e_{\mathrm{u}}(0) & {}_{\mathrm{r}}e_{\mathrm{u}}(0) \\ {}_{\mathrm{u}}e_{\mathrm{r}}(0) & {}_{\mathrm{r}}e_{\mathrm{r}}(0) \end{bmatrix} = \begin{bmatrix} 43 & 9 \\ 12 & 39 \end{bmatrix} .$$

Assume further that the spatial pattern of net reproduction is given by[6]

$$\mathbf{R}(0) = \begin{bmatrix} 1\cdot 10 & 0\cdot 25 \\ 0\cdot 50 & 1\cdot 55 \end{bmatrix} .$$

Then an immediate decline in fertility to bare replacement level according to the *cohort-replacement alternative* gives

$$\hat{\mathbf{R}}(0) = \begin{bmatrix} 0\cdot 73 & 0\cdot 17 \\ 0\cdot 27 & 0\cdot 83 \end{bmatrix} \quad \text{and} \quad \hat{Q}_1 = \begin{bmatrix} 0\cdot 38 \\ 0\cdot 62 \end{bmatrix} ,$$

and an ultimate spatial zero-growth population distribution of

$$\hat{Y} = \begin{bmatrix} 369 \text{ million} \\ 482 \text{ million} \end{bmatrix} .$$

Under this projection approximately 43% of the national population will be urban, giving rise to an urbanization momentum of roughtly $1\cdot 42 \times (0\cdot 43/0\cdot 20) = 3\cdot 1$. A rough estimate of the corresponding momentum under Ryder's *gradual*-fertility-reduction alternative is $3\cdot 49 \times (0\cdot 43/0\cdot 20) = 7\cdot 6$.

The redistributional consequences of the *proportional-reduction alternative* are quite different, however. In this case

$$\hat{\hat{\mathbf{R}}}(0) = \begin{bmatrix} 0\cdot 63 & 0\cdot 14 \\ 0\cdot 29 & 0\cdot 89 \end{bmatrix} \quad \text{and} \quad \hat{\hat{Q}}_1 = \begin{bmatrix} 0\cdot 28 \\ 0\cdot 72 \end{bmatrix} .$$

if fertility declines immediately, and the spatial zero-growth population that results is

$$\hat{\hat{Y}} = \begin{bmatrix} 315 \text{ million} \\ 536 \text{ million} \end{bmatrix} .$$

(5) According to a recent census publication (Registrar General of India, 1972), the crude death rates for urban and rural areas in India in 1970 were 10·2 and 17·3 per thousand respectively. These were disaggregated into age-specific death rates using the age profiles and population data reported in this census publication and in the UN Demographic Yearbook (United Nations, 1975). The crude out-migration rates for urban and rural areas in 1970 were set at 10·0 and 6·8 per thousand respectively (Bose, 1973). A model migration age profile (Rogers, 1976) was used to disaggregate these into age-specific out-migration rates. A two-region life table, calculated using these age-specific mortality and migration rates, produced the matrix of regional expectations of life at birth, **e**(0).

(6) Crude urban and rural birthrates in India in 1970 were 29·7 and 38·8 per thousand respectively. These were disaggregated into age-specific rates using the fertility age profile set out in Ambannavar (1975). The age-specific rates then were combined with the stationary population of the two-region life table referred to in footnote (5) to obtain the matrix **R**(0) of net reproduction rates.

Here only about a third (37%) of the zero-growth population is urban. The rough estimates of the immediate- and gradual-fertility-reduction momenta now become 2·6 and 6·5 respectively[7].

But rural-to-urban migration in India is surely going to increase in the course of its development and modernization. For example, urban-rural demographic data for the Soviet Union give approximately the following regional expectations of life at birth (Rogers, 1976):

$$\mathbf{e}(0) = \begin{bmatrix} 60 & 41 \\ 10 & 29 \end{bmatrix},$$

and the net-reproduction matrix

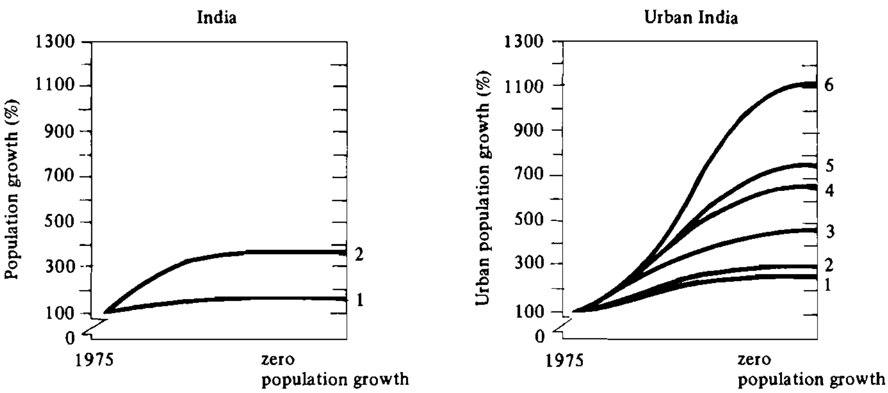
$$\mathbf{R}(0) = \begin{bmatrix} 0{\cdot}80 & 0{\cdot}64 \\ 0{\cdot}23 & 0{\cdot}46 \end{bmatrix}.$$

By way of contrast, a very crude approximation of the corresponding net-reproduction-rate matrix for the USA is (Rogers and Willekens, 1976)

$$\mathbf{R}(0) = \begin{bmatrix} 0{\cdot}85 & 0{\cdot}45 \\ 0{\cdot}25 & 0{\cdot}90 \end{bmatrix}.$$

These data suggest that reasonable assumptions for India's future life-expectancy matrix and its reduced net-reproduction-rate matrix (*after*, say, a cohort-replacement fertility decline) might be

$$\mathbf{e}(0) = \begin{bmatrix} 59 & 23 \\ 9 & 45 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{R}}(0) = \begin{bmatrix} \frac{3}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{2}{3} \end{bmatrix}.$$



1 Immediate decline
2 Gradual decline

1 Immediate-proportional-reduction alternative
2 Immediate-cohort-replacement alternative
3 Immediate-cohort-replacement alternative with increased mobility
4 Gradual-proportional-reduction alternative
5 Gradual-cohort-replacement alternative
6 Gradual-cohort-replacement alternative with increased mobility

Figure 1. Urbanization momenta of India's population.

(7) The two fertility-reduction alternatives differ not only in their redistributional impacts but also with respect to the urban and rural age compositions that they generate. The proportional-reduction alternative produces an urban population that is over five years older in mean age than the corresponding rural population (38·3 to 33·0 years). The same difference in the cohort-replacement alternative is not as pronounced, although the urban population is still older than the rural population in mean age (36·9 to 33·6 years).

The dominant characteristic vector $\hat{Q}_1$ associated with the unit root of $\hat{\mathbf{R}}(0)$ now is

$$\hat{Q}_1 = \begin{bmatrix} 0\cdot 57 \\ 0\cdot 43 \end{bmatrix},$$

and this implies the spatial zero-growth population *distribution* of

$$\hat{Y}_1 = \begin{bmatrix} 0\cdot 64 \\ 0\cdot 36 \end{bmatrix}.$$

Under these assumptions approximately 64% of India's zero-growth population would become urban, yielding an urbanization momentum of $1\cdot 42 \times (0\cdot 64/0\cdot 20) = 4\cdot 5$ for Ryder's case of immediate fertility reduction, and a corresponding momentum of $3\cdot 49 \times (0\cdot 64/0\cdot 20) = 11\cdot 2$ for the case of gradual fertility reduction. Figure 1 illustrates the various urbanization momenta calculated for India, and shows graphically the wide range of potential levels of urbanization for that country. Considering the variety of human-settlement problems that already plague cities such as Calcutta, Bombay, and Delhi, it is a foreboding view of the future.

5 Conclusion

In this paper we have illustrated that migration and redistribution may present growth issues and problems that are potentially as serious as those posed by fertility and natural increase. This troublesome feature of *spatial* population dynamics appears even in zero-growth populations, where the redistributional consequences of an immediate reduction of fertility levels can be of considerable importance.

With respect to methodological issues, this paper has demonstrated that the mathematical apparatus commonly used by demographers to examine the evolution of national populations to zero growth may be extended for application in spatial population analysis. The principle role in this extension is played by the characteristic matrix $\Psi(r)$ and its right and left characteristic vectors, Q and $[v(0)]^{\mathrm{T}}$ respectively. The former vector defines the regional allocation of stable equivalent births; the latter gives the spatial distribution of regional reproductive values at birth. This distinction is hidden in the single-region model, where stable equivalent births and the reproductive value at birth are cancelled out in each of their respective definitional equations.

References

Ambannavar J P, 1975 *Second India Studies: Population* (Macmillan, Delhi)

Bose A, 1973 *Studies in India's Urbanization 1901–1971* (Tata-McGraw-Hill, New Delhi)

Fisher R A, 1929 *The Genetical Theory of Natural Selection* 2nd revised edition (Dover, New York)

Goodman L, 1969 "The analysis of population growth when the birth and death rates depend upon several factors" *Biometrics* **25** 659–681

Keyfitz N, 1975 "Reproductive value: with applications to migration, contraception, and zero population growth" in *Quantitative Sociology* Eds H Blalock, A Aganbegian, F M Borodkin, R Boudon, V Capecchi (Academic Press, New York) pp 587–612

Registrar General of India, 1972 *Pocket Book of Population Statistics, Census Centenary 1972* (Registrar General and Census Commissioner, New Delhi, India)

Rogers A, 1975 *Introduction to Multiregional Mathematical Demography* (John Wiley, New York)

Rogers A, 1976 "Two methodological notes on spatial population dynamics in the Soviet Union" RM-76-48, International Institute for Applied Systems Analysis, Laxenburg, Austria

Rogers A, Willlekens F, 1976 "Spatial population dynamics" *Papers of the Regional Science Association* **36** 3–34

Ryder N B, 1974 "Notes on stationary populations" *Population Index* **40**(1) 3–28

United Nations, 1975 *Demographic Yearbook* (United Nations, New York)

Willekens F, 1977 "The spatial reproductive value: theory and applications" RM-77-09, International Institute for Applied Systems Analysis, Laxenburg, Austria

Environment and Planning A, 1978, volume 10, pages 519-535

Job-search perspectives on migration behaviour†

J R Miron¶
Department of Geography, Division of Social Sciences, Scarborough College, University of Toronto, West Hill, Ontario M1C 1A4, Canada
Received 31 May 1977, in revised form 17 January 1978

Abstract. Although considerable empirical work has been undertaken to estimate interregional migration models, there has not been a corresponding attempt to construct more rigorous theoretical models. For over a decade, Sjaastad's (1962) basic analysis has been unsurpassed. The more recent paper by David (1974), however, attempts to lay a firm microeconomic and risk-analytic foundation for migration models by using concepts arising from job-search theory. His analysis is still quite limited in its perspective on migration, and the present paper attempts to derive a more satisfactory theoretical view of migration. In particular, the various aspects of job-search theory as it has been applied to macroeconomic theory are reviewed and the relevance of these to migration behaviour are assessed. The paper emphasizes (1) differences between search models concerned with wage dispersion and those concerned with job finding, (2) the information possessed by job seekers about different labour markets, (3) alternative search strategies including voluntary job quitting and the migration of unemployed job seekers, and (4) different measures of job-finding probabilities. Although several theoretical models are presented, the main purpose of the paper is to provide a systematic overview of the issues in model design in this important new area of migration research.

In the past two decades, we have witnessed an explosion in the amount of research being undertaken in the area of interurban and interregional migration. Most of this research, as recently reviewed by Greenwood (1975), is empirical in nature. It has been broadly concerned with the estimation of migration flows, the identification of migratory propensities by age, sex, or other cohort groupings, the dynamics of multiregional population growth, and the explanation of migration between regions in terms of such causal variables as age structure, income differentials, distance, and intervening opportunities. On the other hand, there has been relatively little research done on the theory of migration behaviour. Some critics would go so far as to argue that nothing new has appeared since the work of Ravenstein in the late nineteenth century. Others might concede that more recent works, such as those of Sjaastad (1962), Wolpert (1965), Todaro (1969), and David (1974), have also made significant contributions. Even allowing all these, there have been relatively few advances made in the theory of migration behaviour.

What has gone on in the past is, however, not always a reasonable prediction of what can be expected in the future. Certainly the work of Todaro and of David, in their use of job-search perspectives on migration, offers a rich new avenue for theoretical exploration. It may well be that we are in fact on the verge of another explosion of research activity on this topic, this time theoretical. Before this happens, however, it is important to assess what kinds of migration questions can be usefully approached using a job-search perspective. That is the primary purpose of this paper. It is hoped that this paper will make clear what is reasonable or not reasonable to expect from future research of this type.

† This paper was presented at the Seventh Northeast Regional Science Association Meetings, Halifax, Nova Scotia, Canada, May 1977.

¶ Former IIASA research scholar.

1 The foundations of a job-search perspective

Migration has been associated with many epic moments in man's history, such as the settling of the Americas. The reasons for such migration behaviour are manifold: intolerance or persecution at home, the quest for fortune or a better standard of living, and the imperative of a religious or ideological belief, to name just a few. The theories we now have in order to explain when people will move and where they will move to are much more narrowly focussed than this. With Sjaastad's (1962) work as the clearest example in mind, it can be seen that a broadly defined economic-betterment motive is the only one to have received substantial attention in the theoretical literature. This economic emphasis is certainly consistent with contemporary North American societies wherein migration attributable to other motives is relatively limited. Nonetheless the limitations placed on this discussion by a solely economic emphasis need to be noted.

The adoption of a job-search perspective represents a slightly broader view of the migration process. Sjaastad and his empirical disciples would argue that migration tends to occur whenever there are differences in the standard of living between regions (whether this be due to wages, regional amenities, or costs of living) that are sufficient in capitalized form to exceed the costs of migration. Implicit in their view is the notion of migration as a spatial adjustment process based on some kind of disequilibrium between regions. There is, however, a curious dissonance in this approach. If migration is a dynamic disequilibrium process, should there not be more emphasis placed on understanding the behaviour of the potential migrant as an information gatherer and decision maker? Is it not unreasonable to assume, as do many researchers in this area, that the potential migrant has perfect information about the disequilibrium among regions, especially if such disequilibria are changing over time? The proponents of job-search theory would argue that their work extends the economic theory of migration specifically in this area.

Job-search theory does not have its origins specifically in the migration literature. It is principally associated with the set of papers in Phelps (1970) and with subsequent research. There it is applied principally to derive a microeconomic theory which could underpin the empirical Phillips-curve relationship of macroeconomics. A rather substantial literature now exists on that topic.

There have been relatively few attempts to link job-search models to migration theory. One of the earliest is the work of Todaro (1969) which appeared about the same time as Phelps's volume. Although not a rigorous model in microeconomic terms, it nonetheless exemplifies the job-search emphasis on the uncertainty of job finding and rational decisionmaking. The extensions to and corrections of Todaro's model by Zarembka (1970a) should also be noted. The work by David (1974) is certainly more typical of the job-search approach in its rigorous use of microeconomic principles. Also David emphasizes the dispersion of wage offers as well as the uncertainty of job finding, which is typical of the job-search-theory approach. As indicated earlier, these are among the only pieces of migration literature which use a job-search perspective.

To provide a framework for discussing job-search and migration behaviour, David's model is examined in section 2. Some major questions are raised about his model and its relevance to contemporary migration. In section 3 some alternative perspectives on job-search theory are drawn from other parts of the literature. These are used to illustrate the kinds of models which might be forthcoming in the future and what kinds of alternative assumptions would have to underlie them. A concluding assessment is presented in section 4.

2 David's model of job search and migration: a critique

David uses an analogy for what he refers to as the prospective migrant's decision problem. This section begins with an outline of that analogy and the parallels drawn. Subsequently consideration is given to the assumptions made and the restrictions this places on migration and job-search behaviour. Here the empirical relevance of some assumptions are considered in two ways: the prospective migrant's state of knowledge and his method of wage sampling. After this is a discussion of David's main findings and the significance these place on his various assumptions.

The analogy used by David is described as follows.

> "A man has been presented with a set of labeled urns, each containing many balls. Every ball has a dollar value inscribed upon it, but the value of any ball can only be obtained after it is withdrawn from the urn. Upon the label of each urn appear the parameters of the particular probability distribution to which the values of the balls therein conform. Also written upon every label is the fixed "entry fee" that must be paid just for the right to put one's hand in the urn, and, further, a schedule of "sampling charges" describing the (dollar) costs of withdrawing different numbers of balls therefrom .... He is ... required to choose an action strategy composed of two elements: He must designate (1) the single urn from which (2) a specified number of balls are to be drawn, say, in sequence, replacing each before extracting the next and recording its value. He will then be allowed, without further expense, to retrieve any ball—and obviously he will want it to be the highest valued—among those comprising the (random) sample. How should he proceed in making this strategy decision? If given a fixed budget, how should he divide it between purchasing "entry" and sampling? And how large a sum should he be willing to spend in this game?" (David, 1974, page 24).

The parallels drawn by him with the migration decision problem of an individual are clear.

> "In place of "urns" we should then quickly substitute local labor markets: rural districts, towns, cities, or even different countries. And instead of "balls", read job offers. For simplicity's sake a traditional Japanese arrangement may be imagined to prevail—permanent job tenure; the value of a job offer thus becomes the present value of an annual wage annuity received over the remainder of the worker's earning life. To what does "sampling" correspond here? Obviously to local job search, an activity which is presumed to be distinct from employment (job tenure) and can be conducted only at some (scheduled) expense to the individual concerned. Since the contemplated search is to be carried on within the local confines of a single local market, we must suppose that migration thither is a prerequisite for its conduct. Hence the fixed "entry fee" that appeared (along with the schedule of sampling charges) on each urn's label now represents the pure pecuniary and psychic costs of the migration activity necessary to effect entry into the respective local labor markets from some standard-origin place in the system—the "null urn" initially inhabited by the prospective migrant" (David, 1974, page 30).

2.1 *The prospective migrant's state of knowledge*

The prospective migrant is assumed to be in some local labour market in space. From this point he observes other labour markets elsewhere in space at which he might be employed. For each labour market, including his own, he knows the probability distribution of wage offers[1]: that is, different firms in the same labour market offer different wages for the same employee. The prospective migrant does not know beforehand which firm offers what wage or even if a firm has an opening for which he might be suited[2]. He knows only the probability that any given wage offer will turn up next. To find a specific wage offer, the job seeker must 'sample' firms.

The prospective migrant is assumed to be able to attach specific values to every job offer. Presumably each city, or local labour market, offers a bundle of net amenities to the migrant who would choose to live or work there. These in part would reflect elements entering into the job seeker's locational preferences, such as climate and residential environment. Within David's model, it can be assumed that these are all valued by the migrant and included in the 'wage offer' associated with a job opening in a particular local labour market[3].

2.2 *The method of sampling*

David envisages the process of sampling in the following way. The job seeker searches within a local labour market by collecting a set of n wage offers[4]. How many firms will have to be sampled to come up with these n offers is dependent on the frequency of occurrence of job vacancies among firms. From this sample, the job seeker is assumed to choose the highest offer[5].

How does the job seeker 'sample' firms? David has assumed that the job seeker must migrate to sample firms in any but his own present local labour market[6].

(1) How does the job seeker form his estimates of the probability distribution of wage offers in a particular local labour market? David does not consider this question. He does not suggest whether these probabilities are exact or subjectively estimated. Further he does not consider these probabilities to be changing over time either because of changes in local-labour-market conditions or because of the changing perceptions of job seekers.

(2) He thus has no information by which to orient or make systematic his search among firms within a given local labour market. The implied randomness of local search is in sharp contrast to his detailed knowledge about the general distribution of wage offers in different local labour markets and his rational search among these markets.

(3) To the extent that the migrant is uncertain about how well his locational preferences might be satisfied in a given urban area, such an approach is oversimplifying his problem.

(4) To what extent can wage offers be accumulated? As David (1974, page 70) admits, there are broad classes of job seekers who face offers which are only of the "take it or leave it" variety. Even for those fortunate enough to get an offer with a nontrivial life, the time is usually so short that the job seeker is hard-pressed to collect more than one or two additional offers. Most job seekers face a decision situation in which job alternatives are too few in number to make David's model useful.

(5) David's notion of choosing from a fixed sample is subject to some criticism. Other researchers would argue that uncertainty about the cost of generating an additional wage offer makes this a problem in sequential sampling. After each individual offer, they would argue that the job seeker should and does weigh the costs of continuing the search against the marginal expected gain in a subsequent wage offer.

(6) Since the job seeker must migrate to sample firms in other labour-market areas, David views each local labour market as distinct geographically. In fact more than this is assumed. The usual concept of a labour-market area is a fairly dense cluster of work sites surrounded by a net of residence sites lying within some maximum commuting range. The maximum commuting range is partly defined by the daily journey-to-work cost relative to wages and partly by commuting time relative to total daily disposable time. In practice such a commuting threshold is usually much smaller than the search range of a job seeker. That is, the job seeker is willing to search sporadically further afield than he is willing to commute on a daily basis. Thus David assumes both that local labour markets are nonoverlapping in spatial terms and that they are further separated by more than the maximum range of local job search. Broad-scale urban conurbations with interlinked labour-market commuting or search areas cannot be handled under these assumptions. The model is relevant only for a widely separated network of compact cities, and these are increasingly fewer in number with the passing of time.

This implies that sampling requires a physical contact with the firm. In empirical labour-market research, this most closely corresponds to the 'gate-application' strategy which has been found to be commonly used, especially by certain categories of workers[7]. There are, however, also other search methods which do not require migration or physical presence to sample distant labour markets. Letters of inquiry, written responses to regionally advertised job openings, telephone contact, and national or regional job-placement agencies all provide well-recognized substitutes for gate applications. David does not consider these alternative strategies for long-distance job search. This is a critical oversight. In his model, households can express their locational preferences only by the actual act of migrating.

David also does not consider the distinction between migration and long-distance search travel. Migration, according to census takers, is a change in the place of permanent residence from one predefined area to another. It usually involves a movement of household effects and dependents as well as the job seeker himself. David presumes that such movement is a prerequisite to job search. However, it is possible for a job seeker to be physically present for search purposes without having to move his household effects and family first. Thus, in addition to being able to search distant markets without having to be physically present, the job seeker also has a choice between migration and search travel if he wants to be physically present. Indeed such search travel may be a more efficient strategy under a wide set of conditions. Migration need only occur if, in the process of search travel, the job seeker finds and accepts a distant job opening.

At first glance this may seem a moot point. After all, one could rename this a 'search-travel' rather than a migration model and the problem could thus be simply avoided. Interest here is primarily in terms of migration and not search behaviour on its own. To examine migration with this model, it must be extended to include the possibility of both migration prior to search (as assumed by David) and migration subsequent to a successful long-distance search. His model is incomplete in failing to specify the conditions under which these alternative search strategies are each preferable.

Further David presumes an 'intensive' sampling strategy in two senses of the word. First, job seekers are assumed to be searching full-time for job offers. During the search period they must be otherwise unemployed. This does not permit, for example, part-time search where a job seeker retains a job while searching for a better position. David supports this argument by suggesting that a large proportion of workers undertake such intensive search. Second, there is no evidence presented about the proportion of total job quits associated with subsequent search unemployment. A more general search model would permit a choice between intensive and part-time search and emphasize the determinants of such choices.

The sampling strategy is also intensive in that only one local labour market is sampled. David does not consider a strategy in which the job seeker samples, either simultaneously or sequentially, a series of markets. He would undoubtedly justify his approach on the assumptions that prior migration is necessary and migration costs are too high to make extensive spatial sampling economic. However, the possibility of sampling several markets in a short period of time exists, especially when sampling is not strictly of the gate-application kind. David's simplistic view of search behaviour thus leads him to disregard another important aspect of a sampling strategy.

(7) David (1974, page 69) himself draws support from the findings of Reynolds (1951, pages 214, 215, and 240). Similar arguments with respect to young job seekers in particular have also been made by Stephenson (1976, page 108).

In conclusion, David considers a very restricted kind of search behaviour. No search occurs without prior unemployment, thus eliminating part-time search. No local-market search occurs without physical presence, thus disregarding other sampling methods. Finally, no change of location is permitted without corresponding migration, thus eliminating long-distance search travel.

2.3 *Conclusions from the model*
Given these assumptions, however flawed they might be, David is able to derive some interesting mathematical results. Perhaps the most important is an emphasis on the variance of the wage-offer distribution (denoted by σ_i^2 for local labour market i). This variance is important in two respects. First, since the prospective migrant is making a decision under uncertainty, some notion of rational risk-taking behaviour must be specified. David assumes that some degree of risk aversion is rational so that the individual is not indifferent to the variance of the wage distribution. Other things being equal, the risk-averse individual would choose to migrate to the local labour market having the lowest σ_i^2. Second, the variance is important because it helps to define the expected maximum offer in a random sample of size n. The larger the variance, other things being equal, the greater the expected maximum offer. Thus an increase in the variance of a wage distribution has an undetermined effect. On the one hand it increases the expected maximum offer, whereas on the other hand it increases the risk of having to accept a low wage.

Another contribution of David's is to make simultaneous the decision to migrate and the decision as to how long (or hard) to search in order to get optimal results. His model permits the prospective migrant to consider, in effect, the expected duration of search unemployment as one aspect in his choice of where to locate. This represents one of the first attempts in the formal theory of migration to give consideration to the duration of unemployment[8].

A simple version of David's model (for two local labour markets) serves to illustrate both of the points raised so far. Consider a prospective migrant who currently resides and works in labour market a. The present value of an earnings stream at a for this worker is

$$Y_a = Ry_a \,, \tag{1}$$

where y_a is the fixed annual income and R is the present value of a dollar flow of income over the worker's remaining working life. R reflects the discount rate and the individual's working-life horizon. Alternatively the worker could choose to migrate to b (the only other labour market in this example) and search for n wage offers. His maximum possible annual wage would be the maximum of these,

$$y_{b\,\max} = \max\{y_{b1}, y_{b2}, ..., y_{bn}\} \,, \tag{2}$$

where y_{bj} is wage offer j in labour market b. His discounted earnings stream here, Y_b, is similar to that at a except that initial search costs, S, are subtracted out,

$$Y_b = Ry_{b\,\max} - S \,. \tag{3}$$

The expected search costs,

$$S = s_0 + s_1 n \,, \tag{4}$$

include a fixed component (s_0), corresponding to the migration cost, and a local search cost (s_1) which varies with the number of offers required, or equivalently with

[8] Although the Todaro (1969) models, as will be seen shortly, also treat the probability of being unemployed in a time period, they are not formal microeconomic models.

the time required for local search at b[9]. Further, the utility, W_i, placed by an individual on these discounted earnings streams is

$$W_i = Y_i^{\gamma}\ , \qquad 0 < \gamma < 1\ , \quad i = a, b\ , \tag{5}$$

subject to risk aversion if Y_i is stochastic. Finally, the prospective migrant's decision problem is to select a local market (a or b) and, if b, to choose an optimal level of search. This is done to maximize the expected value of W_i, to find

$$\max_{i,n}\{\mathrm{E}(W_i)\}\ . \tag{6}$$

This completes the formal statement of the model, represented by equations (1) to (5) and function (6).

To make this model operational, a few additional assumptions and derivations are required. First, since y_a is deterministic (known), $\mathrm{E}(W_a) = Ry_a$. On the other hand, $y_{b\max}$ is a stochastic variable. David argues that a good second-order approximation is $\mathrm{E}(W_b) = R\bar{y}_{b\max}(1-\delta\rho_y)-S$, where $\bar{y}_{b\max} = \mathrm{E}(y_{b\max})$, $\rho_y = \sigma_y^2/\bar{y}_{b\max}$, $\sigma_y^2 = \mathrm{E}[(y_{b\max}-\bar{y}_{b\max})^2]$, and $\delta = \gamma(1-\gamma)/2$. Function (6) then reduces to

$$\max\{Ry_a\,,\, R(\bar{y}_{b\max}-\delta\sigma_y^2)-S\}\ . \tag{7}$$

To evaluate $\bar{y}_{b\max}$ and σ_y^2, the mean and variance of the extreme value of a sample, it is necessary to invoke some particular probability distribution for wage offers. David assumes wage offers are $N(\bar{y}_b, \sigma^2)$, that is, normally distributed with mean $\bar{y}_b$ and variance σ^2. In that case $\bar{y}_{b\max}$ and σ_y^2 are approximated by

$$\bar{y}_{b\max} = \bar{y}_b + B\sigma n^{\beta}, \qquad 0 < \beta < 1,\ \ B > 0 \tag{8}$$

and

$$\sigma_y^2 = A\sigma^2 n^{-\alpha}\ , \qquad \alpha > 0,\ \ A > 0\ , \tag{9}$$

where α, β, A, and B are constants. Function (7), equations (8) and (9), together with

$$S = s_0 + s_1 n\ , \qquad s_0, s_1 > 0\ , \tag{10}$$

form a complete statement of the migrant's decision problem. W_b can be solved explicitly for an optimal n, and the optimized value, W_b^*, could be used to solve separately the problem $\max\{W_a, W_b^*\}$.

Several comments are in order here. First, the model can not be analytically solved for the optimal sample size, n^*. Although the economic interpretation of an optimized expression for $\mathrm{E}(W_b)$ is clear enough (that is, the marginal cost of increasing sample size, s_1, should be equal to the marginal gains from search), the solution is tractable only at a numerical level. Second, the model can be easily extended to include search within labour market a as well. Thus the worker would choose among three strategies: retaining his current job, searching in a, and searching in b. Last, the model is readily extended to cover more than two local labour markets.

[9] Let p_1 be the probability that a firm has a vacancy. The expected number of firms a job seeker must visit to obtain n offers is $N = n/p_1$. Therefore the expected search cost is $S = s_0 + (p_1 s_2)N$, where s_2 is the marginal search cost per firm. From equation (4) it is seen that S reflects both the cost of search per firm and the likelihood of getting an offer.

3 Alternative perspectives on job search

Given the criticisms of David's (1974) approach, it is instructive to ask how alternative models of job-seeker behaviour might be formulated. Two broad sets of issues are identified. One has to do with the general setting within which job-search behaviour is examined. The other treats questions about the choices open to the job seeker and his optimal choice under different assumptions about rationality.

3.1 *The problem setting*

The 'problem setting' is made up of a grab bag of different aspects. One of these is the number of different ways in which researchers might treat space, including the number of (and separation among) local labour markets. Another concerns different possible assumptions about the knowledge (or degree of uncertainty) possessed by job seekers. A final aspect considered is the different ways in which the probability of finding a job offer in a unit of time is handled. These are now considered in more detail.

3.1.1 *The spatial setting.* In the treatment of a spatial setting, David's model is the most ambitious job-search model to date. The only other spatial models, in any sense, have been those of Todaro (1969) and Zarembka (1970a; 1972). Those are concerned with rural-urban migration and assume only two corresponding labour-market areas. David's model, by contrast, allows for any number of local (rural or urban) labour markets.

Does the multimarket aspect of David's model generate any new insights? The answer is negative because of a critical assumption made by him. By assuming spatially distinct labour-market and search areas, he is unable to consider relevant phenomena such as interarea job commuting or search. His spatial setting in this regard is more reminiscent of a nineteenth-century urban system than of a contemporary one. Although his model may be used numerically in interesting applied analyses, it is doubtful that the multimarket aspect is helpful in a theoretical model. To push this attack further, it is not clear that any useful theoretical insights can be formed which would not be available in a two- (or at most a three-) market model. Although future theoretical models should be restricted to such small urban systems, this argument of course does not deny the usefulness of a large-scale multimarket approach in empirical models.

An additional issue here is whether local labour markets should be treated as arealess points in space. Schneider (1975) considers a model which explicitly analyzes areal form in studying search in an urban market. This model emphasizes search by physical contact and the necessity of moving along paths between firms being searched. How relevant this approach is is an open question. Schneider's model does not consider the use of other search tools such as telephones, letter writing, and job vacancies, which alleviate the need for purely physical spatial search. The gains from considering simple point markets seem, in conclusion, to outweigh the restrictive complexity of the areal market models developed so far.

3.1.2 *Uncertainty and search.* In David's model it has been assumed that the job seeker is uncertain only about the wage offer, if any, of a particular firm. It has also been assumed that he knows the mean and variance of the (normal) wage distribution in each labour market. Further, the worker is assumed to know that any wage offer can be kept open until he finishes collecting his sample. Finally, it is assumed that the job seeker can assess all the additional elements intrinsic to a given job opening (for example, work environment and job stability) and include these in an overall 'wage' measure. Are these realistic assumptions for a broad sector of workers in a contemporary society? Alchian (1970) and others have attacked the assumption that the parameters of the wage distribution are known. They perceive the employed labour-force member to be cognizant only of the wage rate he himself has received in the past and is currently receiving. A less-than-average change in this rate, relative to past changes, will cause the worker to quit and search for work elsewhere in the perhaps mistaken belief that his firm is not being competitive. Such behaviour, because the worker is uninformed about average market wages, leads to a 'wage illusion' in which the worker accepts or rejects a certain wage rate without knowing how large it is relative to other wages. Parsons (1973), in an empirical study of quit rates, finds no support for such a wage illusion. His findings are consistent with the idea that workers are broadly aware of average wages. Thus, although they may not know much about the other parameters of a wage distribution, they are at least generally aware of its expected value.

As indicated earlier, it seems unreasonable that a wage offer can be kept open until a large ($n > 3$) sample is collected. Several researchers, including McCall (1970), have constructed models which relax this assumption. They find the optimal strategy is a sequential one in which, after each wage offer, the job seeker weighs the relative cost of finding an additional offer against the expected gain which that new offer would represent. A sequential strategy has been shown by McCall and others to lead to a 'reservation wage'. The individual would choose, under this strategy, to stop searching with the first offer exceeding this reservation wage. The reservation-wage approach has not yet been integrated into a model of migration behaviour.

There is, however, a more fundamental issue here. Both the fixed-sample and the reservation-wage strategies emphasize the dispersion of wages in a local labour market as the prime source of uncertainty underlying the job seeker's problem. Is such a dispersion the essential source of uncertainty underlying migration behaviour[10]? Some researchers, particularly Todaro (1969) and Zarembka (1970a; 1972), emphasize the uncertainty of job finding alone. In their models they assume that all jobs carry the same wage, thereby eliminating wage dispersion. The prospective migrant has a known and fixed probability of finding a job in any period of time and this constitutes the only uncertainty facing him. Such an approach is quite attractive because it makes the informational requirements for a migration decision much smaller. What emphasis should be placed on job finding versus wage dispersion is still an open empirical question[11].

As a final issue here, one might ask how the job seeker weighs one important intrinsic element of a given job: its expected duration. David has assumed that the job seeker believes that a job opening will last the rest of his lifetime. This, however, is a polar case. More generally the worker believes a job will have a finite life although he may be unsure about how long this will be[12].

A job may terminate for several main reasons: retirement, quitting, and layoff. The retirement effect has already been considered in David's model in terms of the discounted stream factor R. No presently available search model considers future 'voluntary' quits (for the purpose of taking or searching for a better job) in the job seeker's current decision problem. A simple discounted-earnings-stream approach, such as that of David, understates the true expected earnings stream by ignoring the expected income gain from future search. This could call for a drastically different

(10) The basic search-theory models have often been criticized for their reliance on wage dispersion. In addition to empirical doubts, Rothschild (1973, page 1288) and others have questioned the theoretical basis of such dispersion.

(11) As an example of the fragmented evidence available, Stephenson (1976, page 110) suggests that about 90% of young job seekers in the United States accept the first job offered to them.

(12) Stephenson (1976, page 109) also presents evidence suggesting that, among young job seekers in the United States, the older the job seeker the longer he expects his next job to last.

choice from the one dictated by David's model. Layoffs or 'involuntary' quits (that is, not quitting to search for or take up a better job) are not generally considered in the literature either. These too may have a significant effect on the prospective migrant's decision by making a simple discounted earnings stream overstate the expected true stream of earnings.

To incorporate the possibility of these last two kinds of job termination, more information is required by the prospective migrant. He needs to have some idea about the probability of layoff in each of the local labour markets. In addition, for each labour market, he needs to evaluate the likelihood that future search will become profitable. The latter might include, for example, an expectation about temporal shifts in the wage distribution of a local market relative to current conditions. It is unclear as yet just what sources the prospective migrant tends to use in estimating such information.

A small example illustrates the role of uncertainty about layoff on the prospective migrant's choice. Suppose that a job seeker is considering searching some particular local labour market. Assume that he knows the only wage, w, which can be obtained there. It remains fixed in real terms over the worker's life. Further assume he knows the probability p of finding a job opening and the probability q of being laid off involuntarily during a period. These are also assumed to be fixed over time. Given that the job seeker is unemployed at the outset ($t = 0$), can one find the probability e_t that he is employed at time t? Let u_t be the probability that he is unemployed and let $\mathbf{v}_t = (u_t, e_t)$, where, of course, $u_t + e_t = 1$. The earlier assumptions help to define a Markov model of the form [13]

$$\mathbf{v}_t = \mathbf{v}_{t-1}\mathbf{P}, \qquad \text{where} \quad \mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}. \tag{11}$$

From this,

$$\mathbf{v}_t = \mathbf{v}_0\mathbf{P}^t . \tag{12}$$

As Howard (1971, pages 63-64) indicates, $\mathbf{P}^t$ has a closed-form solution,

$$\mathbf{P}^t = \begin{bmatrix} \dfrac{q}{p+q} & \dfrac{p}{p+q} \\ \dfrac{q}{p+q} & \dfrac{p}{p+q} \end{bmatrix} + (1-p-q)^t \begin{bmatrix} \dfrac{p}{p+q} & \dfrac{-p}{p+q} \\ \dfrac{-q}{p+q} & \dfrac{q}{p+q} \end{bmatrix}. \tag{13}$$

Given the prior assumption that $u_0 = 1$, it is now easily seen that

$$e_t = \frac{p}{p+q}[1-(1-p-q)^t] . \tag{14}$$

The expected value of the discounted flow of future earnings can now be calculated. Assume a discount rate of r and assume that there are no unemployment benefits. The expected discounted earnings stream, E, is given by [14]

$$E = \sum_{t=0}^{T} \frac{e_t w}{(1+r)^{t+1}} . \tag{15}$$

(13) Note that the Markov model also assumes one state change per unit time (that is, from unemployed to employed or vice versa).

(14) Note that E is defined solely in terms of e_t and ignores the duration of unemployment. Thus the job seeker is indifferent between several short periods of unemployment or a few longer periods.

Substituting from equation (14) and letting the worker's time horizon, T, be large, one gets as an approximation

$$E = \frac{p}{r(p+q+r)}w \,. \tag{16}$$

Equation (16) can be expressed in a number of ways. Let us assume that the worker has a fixed level of E, say E_0, in mind. A rearrangement of equation (16) in this case yields

$$q = -r+p\left[\frac{w}{rE_0}-1\right] . \tag{17}$$

This equation ties together the three local-market variables: q, p, and w. Given that the job seeker wants to maximize E, equation (17) indicates indifference curves among these three variables[15]. For example, the job seeker is indifferent to a higher quit rate provided that it is accompanied by a higher wage or higher job-finding probability.

3.1.3 *The probability of job finding.* Whatever the emphasis on wage dispersion, some attention must be focussed on uncertainty in the search for job openings. This may be summarized by thinking of the probability that a job seeker will locate any opening either while searching a firm or during a given time period. How would one measure this probability? What are its determinants?

There are several ways in which one could measure the probability of getting a job offer during a period. These may be broadly classified according to how one would view the hiring behaviour of firms. One possibility is that firms maintain waiting lists of acceptable job applicants. In an extreme version, the firm would make available a job opening only to the first person on this list and the person's name would be dropped from the list subsequently regardless of whether he accepts or rejects the offer. An opposite view of the hiring process would be that no waiting lists are maintained. In this case, a job opening is offered to the next applicant who appears. These two polar views of the hiring process result in different kinds of job-finding probability measures.

Let us look more carefully at the 'waiting-list' or queueing model of job hiring. Suppose that a job seeker randomly visits N firms during a period of which n on average find him an acceptable applicant. Then $p_1 = n/N$ is the probability that a firm will make him a job offer at some time. This offer may come immediately but generally will come substantially after the application. The probability $p(t)$ that he gets an offer during a time interval t is dependent on p_1, on the magnitudes of the waiting lists of firms, and on the rate of occurrence of job vacancies. In general $p(t)$ will be near zero when the job seeker initially begins searching but will increase, at least to some level, with his duration of search. No application of this approach has been found in the job-search literature although this queueing model seems to be a realistic representation of some job-search problems.

The alternative view of the hiring process, in which no waiting lists are maintained, might be referred to as the 'bingo' model. Here there are at least two different measures of the probability of finding a job. One is the measure used, for example, by David (1974) and by Barron (1975). During a small interval of time, only a certain proportion of firms, say p_1, have one or more job openings of a type suitable

(15) Note that the coefficient of p in equation (17) is positive: from equation (16)

$$\frac{w}{rE_0} = \frac{p+q+r}{p} > 1 \,.$$

to the particular job seeker[16]. If a job seeker happens to find a firm with such a vacancy, this position is assumed to be offered to him. Thus, in the David-Barron view, the job seeker who engages in random search has a probability p_1 of getting an offer from a firm. Further, if it takes $m > 1$ time periods to search one firm, the probability of getting an offer in a unit time is $p = p_1/m$.

An alternative definition of this probability in the bingo model has been suggested by Todaro (1969). He imagines a labour market where all new job openings are filled within one period of time and that these positions are filled randomly from among the ranks of a large stock of unemployed[17]. Given O_t openings in the local labour market at time t and U_t unemployed job seekers, the probability that a job seeker will get an offer is $p(t) = O_t/U_t$.

These two definitions of the job-finding probability in a bingo model share certain features. They are both objectively determined by conditions in the local labour market. They both see the hiring of workers as a kind of random choice among the unemployed job seekers. They both require that the job seeker have some idea of the rate of job creation by firms in a given local labour market. Finally neither sees the duration of an individual's search as affecting p[18]. In fact if the general labour-market conditions remain constant so does p.

Are there any reasons for preferring either a queueing or a bingo model? In terms of realism, exclusive reliance on the bingo model is unsatisfactory. Firms, to varying degrees, find it helpful to maintain waiting lists of acceptable job seekers, especially where such workers infrequently approach the firm. By maintaining such lists, the firm should be able to fill its openings more quickly (and at a relatively small increase in administrative costs) than by hiring according to a bingo model. Secondly, the queueing model may make more realistic information demands on the part of the job seeker. How would the job seeker get the information required to estimate p in the bingo model? Newspaper and other reports about the expansion of economic activity in a local labour market might be a common source. The job-hunting experiences of friends who have recently searched in that market might be another frequently used source. It is doubtful in either case that the job seeker could get enough information to have more than a very vague estimate of p.

In a queueing model there are two kinds of information which the individual must acquire. First, he must know p_1, the probability of being a successful applicant. Second, he must have some idea of the expected time until he comes to the top of a waiting list. To estimate p_1 he must rely on his own experiences in job search since this is his only guide to his intrinsic qualities as generally perceived by firms. To estimate waiting times, he might reasonably rely on the personnel officers of the firm

(16) Barron assumes, more strictly, that no firm has more than one vacancy (of a given type) at a given time.

(17) This is in contrast to the David-Barron model where an opening is filled in a period only if an appropriate applicant appears.

(18) Some objections to the basic assumptions of the bingo model raise questions about the effect of search duration on p. Several researchers, including Zarembka (1972, pages 54-58), have suggested that job search involves skills which must be learned. They see the sampling of firms not as a purely random process but as one in which the job seeker becomes increasingly adept at discovering firms with vacancies. Thus the individual job seeker has his own p which increases with search duration.

An alternative argument has been developed by Salop (1973) to suggest that job seekers are relatively good at picking out high-wage and likely employers. In systematic searching, the job seeker will search these firms first and will only subsequently search low-wage, less-likely employers. In this contrasting view, p will decrease with the duration of search. It is improbable that search learning and systematic search effects exactly cancel each other out, but the net effect on p of search duration is an empirical issue if one accepts the bingo model as a realistic approximation.

who might say, for example, "you are an acceptable applicant. At current turnover and expansion rates, we would be able to hire you in six to twelve weeks." As his search gets underway, the job seeker thus has a built-in process by which to estimate and update $p(t)$ with his market experiences.

Finally, the queueing model may have an advantage in terms of its relatively subjective nature. The David–Barron and Todaro (1969) approaches assume that p is objective and known before the job seeker makes his migration decision. The queueing model, however, suggests that p is subjective. The migrant may enter a labour market with some notion about p that undergoes changes with his own job search experiences. In this case p is combined with search (sample) information to yield a posterior estimate of p.

3.2 *The choice problem*

In this subsection, the specifics of the choice problem facing the job seeker are discussed. What are the choice alternatives? Initially a general discussion of these questions is undertaken here. Subsequently a simple model illustrating some of the issues raised is presented.

Up to this point, simple models of job search have been presented. The job seeker looks at different local labour markets, makes a choice among them, migrates, and then searches for work. His choice was merely where (if at all) to migrate and how long to search. Several objections are possible about this paradigm. As indicated earlier, several researchers might criticize the notion that migration and unemployment prior to search is the only strategy open to the job seeker. The choice alternatives should be expanded to include long-distance search travel with migration only consequent to successful job finding.

Other researchers, such as Mortensen (1970), have emphasized the choice between full- and part-time search. They see the possibility, in other words, of employed workers undertaking limited search for new work while maintaining their present jobs. The presence of such part-time search in reality is quite extensive and cannot be ignored. In fact several researchers, such as Alchian (1970, page 29), are quick to point out that full-time search is justifiable only if it enables more efficient and productive search. A job seeker might engage in limited search while maintaining a job, change jobs to permit the move to search on a part-time basis with more extensive search activity, or quit to search full-time. A more realistic model of the job seeker would make the intensity of search part of his choice problem.

Finally, in treating the problem of rational choice, most researchers have assumed that the job seeker maximizes the expected value of his discounted future-earnings stream. This implies risk neutrality. David (1974) is one of the few to consider risk aversion as a basis for rational decisionmaking under uncertainty. Just what the gains are in moving from simple risk neutrality to other rationality assumptions is an open question. Although David's model, for example, places an emphasis on the variance of the wage distribution, it is unclear whether a job seeker would ever have enough information about this to behave properly. Unless alternative models of rationality can be developed which generate realistic data requirements on the part of job seekers, there would be little to gain from considering them here.

3.2.1 *Migration and search travel.* Consider the following model as an example of what can be done with the first issues raised: the choice of search strategies. Imagine two local labour markets, a and b, in which there are fixed wages, y_a and y_b, and fixed probabilities of job-finding, p_a and p_b, respectively. Assume that there is no possibility of layoff, that is, $q = 0$. The cost of migration from a to b is c_{m}, the cost of a round-trip search trip is c_{s}, and the discount rate is r. Of course, $c_{\mathrm{m}} > c_{\mathrm{s}}$. Finally assume risk neutrality on the part of the job seeker.

There are three actors whose decision might be evaluated. The employed worker at a who is considering quitting to search at b, the unemployed job seeker at a who is considering which market to search, and the unemployed job seeker at b who is considering a similar question[19]. Here attention is focussed on the first two, although the third can be handled in a similar manner.

Several earnings streams can be calculated. The person presently working at a (or b) has a discounted earnings stream of E_{ea} (or E_{eb}), where, from equations (12), (13), and (15), with $q = 0$ and $e_0 = 1$, and a subsequent approximation for large T,

$$E_{ei} = \frac{y_i}{r} , \qquad i = a, b . \tag{18}$$

Here working-life horizons have been ignored again. The unemployed job seeker at a (or b) has a similar discounted earnings stream, E_{ui}, where, from equation (16) with $q = 0$,

$$E_{ui} = \frac{p_i}{(p_i + r)r} y_i , \qquad i = a, b . \tag{19}$$

Given that a person makes a migration or search decision so as to maximize E, equations (18) and (19) can be used to identify an optimal choice. For the moment, disregard direct costs of search and migration (c_s and c_m). The minimum condition for an employed person at a to seek work at b is $E_{ub} > E_{ea}$ or[20]

$$\frac{y_b}{y_a} > 1 + \frac{r}{p_b} . \tag{20}$$

This asserts a simple relationship between the relative wage rate at b, the likelihood of finding a job there, and the discount rate. The similar minimum condition for an unemployed worker at a to search at b is $E_{ub} > E_{ua}$ or

$$\frac{y_b}{y_a} > \frac{1 + r/p_b}{1 + r/p_a} . \tag{21}$$

This is similar to equation (20) except it also includes p_a.

What about the choice of search strategy? If a job seeker at a migrates to b and then searches, his direct search cost, S_m, is an immediate outlay,

$$S_m = c_m . \tag{22}$$

If he undertakes search travel first and migrates only if a job is found, the expected search cost, S_s, when properly discounted, is

$$S_s = c_s + \left(\frac{p_b}{1+r}\right)c_m + \left(\frac{p_b}{1+r}\right)\left(\frac{1-p_b}{1+r}\right)c_m + \left(\frac{p_b}{1+r}\right)\left(\frac{1-p_b}{1+r}\right)^2 c_m + \ldots$$

$$= c_s + \frac{p_b}{r + p_b} c_m , \tag{23}$$

where c_s is the cost of the initial search travel. The advantage of this second strategy is that it allows the job seeker to defer c_m until a job is found. The disadvantage is the initial outlay c_s required. The job seeker will choose the smaller of S_m and S_s.

(19) Given the structure of the assumptions, there is no incentive for the employed person at b to want a job at a. This actor is therefore ignored.

(20) This is a minimum condition because the existence of nonzero c_m and c_s will generally require a larger y_b/y_a ratio.

He will undertake prior search if $S_s < S_m$ or

$$\frac{c_m}{c_s} > \frac{r+p_b}{r} , \tag{24}$$

and prior migration otherwise. The decision problem facing the job seeker at a is thus separable into two parts. First, how should b be searched by a job seeker at a. His optimal decision is to incur a search cost, S_*, where

$$S_* = \min\{S_s, S_m\} . \tag{25}$$

Based on this the second problem is to decide whether to search b at all. The job seeker will do this if

$$E_{ub} - S_* > E_{ka} , \qquad k = u, e , \tag{26}$$

depending of course on his current employment status. Thus it is possible to develop relatively simple models which make endogenous interesting aspects of the job seeker's decision problem.

3.2.2 *Full- and part-time search.* Although several authors, including Alchian (1970), have examined the trade-off between full- and part-time search at an aspatial scale, there has been no treatment of this problem in reference to migration.

A simple model serves to illustrate some of the issues involved. Let us consider a person at location a considering a costless move to a higher-wage location b. Suppose further that this person is currently employed at a although he is engaged in part-time search for work at b. Further assume that the probability of success in such part-time search is p_p per unit time. If e_{ti} is the probability of being employed at location i at time t, the expected utility from part-time search for this person currently employed at a is

$$E_{ea} = \sum_{t=1}^{T} \frac{e_{ta} y_a}{(1+r)^t} + \sum_{t=1}^{T} \frac{e_{tb} y_b}{(1+r)^t} , \tag{27}$$

where

$$e_{ta} = (1-p_p)^{t-1} \qquad \text{and} \qquad e_{tb} = (1-p_p)^{t-1} . \tag{28}$$

By use of equations (28) in equation (27), it can now be shown that

$$\lim_{T\to\infty} E_{ea} = \frac{y_b}{r} - \frac{y_b - y_a}{r+p_p} . \tag{29}$$

Alternatively the person who quits a job at a to search full-time at b has a utility level given by equation (19),

$$\lim_{T\to\infty} E_{ua} = \frac{p_f}{(p_f+r)r} y_b . \tag{30}$$

Here p_f is the probability of success in full-time search and it is presumed that $p_f > p_p$. The job seeker is indifferent between full- and part-time search when $E_{ua} = E_{ea}$ or when

$$\frac{y_b}{y_a} = \frac{r+p_f}{p_f - p_p} . \tag{31}$$

If the ratio y_b/y_a is larger than in equation (31), full-time search is preferable. If it is smaller, part-time search is more desirable. Thus equation (31) describes a simple relationship between the probabilities of job finding in full- versus part-time search, the discount rate, the relative wage to be earned at b, and the most desirable search strategy.

4 Priorities in future modelling of job search and migration

Throughout this paper several issues have been raised about alternative ways to model or represent job-search behaviour. In concluding this paper the most important of these need to be reidentified and suggestions made about future modelling.

The most important controversy in job-search theory is over the relative roles of wage versus job-finding uncertainty. Most models, including that of David (1974), assign a central role to the dispersion of wages within a labour market in the job-search process. Is it reasonable to assume, however, that wage dispersion is so important to prospective in-migrants? Even if it is, are not the informational requirements of a model such as David's too excessive on the migrant's part to be realistic? Finally is it realistic to think of the wage sampling that inevitably accompanies wage dispersion as a common experience of job seekers. For all these arguments, Todaro's (1969) model of wage rates which are fixed within each labour market but vary from one local market to the next seems to be an attractive alternative. His emphasis on uncertainty about finding a job seems preferable because it is more in agreement with everyday experience as well as being simpler.

The second important controversy surrounds the treatment of search behaviour over long distances. David's model is the only one to consider this and his emphasis on migration and unemployment as prior conditions for searching are objectionable. Some work must be done toward the modelling of long-distance search behaviour without prior migration. Even though certain areas are very attractive places to live or work, many prospective migrants are unwilling to take the chance of giving up a current livelihood, moving lock, stock, and barrel, and hoping for a new job upon arrival. Many of these same prospective migrants, however, are willing to engage actively from a distance in trying to find job openings in these areas. The impact of this behaviour is not lost on firms in these favoured areas, who observe constant flows of applicants or applications for job openings. The models described in this paper are only a first small step toward a better treatment of the alternative search strategies open to such prospective migrants.

References

Alchian A A, 1970 "Information costs, pricing, and resource unemployment" in *Microeconomic Foundations of Employment and Inflation Theory* Ed. E S Phelps (W W Norton, New York) pp 27-51

Barron J M, 1975 "Search in the labor market and the duration of unemployment: some empirical evidence" *American Economic Review* **65** 934-942

• Cordey-Hayes M, 1975 "Migration and the dynamics of multiregional population systems" *Environment and Planning A* 7 793-814

David P A, 1974 "Fortune, risk, and the microeconomics of migration" in *Nations and Households in Economic Growth* Eds P A David, M W Reder (Academic Press, New York) pp 21-88

• Goldfarb R S, 1973 "A model of wage setting and wage diversity in local labour markets" *Western Economic Journal* **2** 98-110

Greenwood M J, 1975 "Research on internal migration in the United States: a survey" *Journal of Economic Literature* **13** 397-433

• Gronau R, 1971 "Information and frictional unemployment" *American Economic Review* **61** 290-301

• Hartley M J, Revankar N S, 1974 "Labour supply under uncertainty and the rate of unemployment" *American Economic Review* **64** 170-175

• Holmes J, 1972 "Linkages between external commuting and out-migration" *Economic Geography* **48** 406-420

Howard R A, 1971 *Dynamic Probabilistic Systems. Volume 1: Markov Models* (John Wiley, New York)

• Kasper H, 1967 "The asking price of labor and the duration of unemployment" *Review of Economics and Statistics* **49** 165-172

McCall J J, 1970 "Economics of information and job search" *Quarterly Journal of Economics* **84** 113-126

•Mazek W F, Chang J, 1972 "The chicken or egg fowl-up in migration: comment" *Southern Economic Journal* 38 133-139

•Mirman L J, Porter W L, 1974 "A microeconomic model of the labour market under uncertainty" *Economic Inquiry* **12** 135-145

Mortensen D T, 1970 "A theory of wage and employment dynamics" in *Microeconomic Foundations of Employment and Inflation Theory* Ed. E S Phelps (W W Norton, New York) pp 167-211

Parsons D O, 1973 "Quit rates over time: a search and information approach" *American Economic Review* **63** 390-401

Phelps E S (Ed.), 1970 *Microeconomic Foundations of Employment and Inflation Theory* (W W Norton, New York)

•Raiffa H, Schlaiffer R, 1961 *Applied Statistical Decision Theory* (Harvard University Press, Cambridge, Mass)

•Rees A, 1966 "Information networks in labour markets" *American Economic Review* **76** 559-566

Reynolds L G, 1951 *The Structure of Labour Markets, Wage and Labour Mobility in Theory and Practice* (Harper and Row, New York)

Rothschild M, 1973 "Models of market organization with imperfect information: a survey" *Journal of Political Economy* **81** 1283-1308

Salop S C, 1973 "Systematic job search and unemployment" *Review of Economic Studies* **40** 191-201

Schneider C H P, 1975 "Models of space searching in urban areas" *Geographical Analysis* 7 173-185

•Siven C, 1974 "Consumption, supply of labour, and search activity in an intertemporal perspective" *Swedish Journal of Economics* **76** 44-61

Sjaastad L A, 1962 "The costs and returns of human migration" *Journal of Political Economy* **70** 80-93

Stephenson S P Jr, 1976 "The economics of youth job search behaviour" *Review of Economics and Statistics* **58** 104-111

•Stigler G J, 1962 "Information in the labor market" *Journal of Political Economy* **70** 94-105

Todaro M P, 1969 "A model of labour migration and urban development in less-developed countries" *American Economic Review* **59** 138-148

•Todaro M P, 1976 "Rural-urban migration, unemployment, and job probabilities; recent theoretical and empirical research" in *Economic Factors in Population Growth* Ed. A J Coale (Macmillan, New York)

•Toikka R S, 1974 "The economics of information: labour market aspects" *Swedish Journal of Economics* **76** 62-72

•Toikka R S, Holt C C, 1976 "Labour force participation and earnings in a demographic model of the labour market" *American Economic Review* **66** 295-301

•Wachter M L, 1972 "Wage determination in a local labour market: a case study of the Boston labour market" *Journal of Human Resources* 7 87-103

•Whipple D, 1973 "A generalized theory of job search" *Journal of Political Economy* **81** 1170-1188

Wolpert J, 1965 "Behavioural aspects of the decision to migrate" *Papers of The Regional Science Association* **15** 159-169

•Yap L, 1976 "Internal migration and economic development in Brazil" *Quarterly Journal of Economics* **90** 119-137

Zarembka P, 1970a "Labour migration and urban unemployment: comment" *American Economic Review* **60** 184-186

•Zarembka P, 1970b "Marketable surplus and growth in the dual economy" *Journal of Economic Theory* **2** 107-121

Zarembka P, 1972 *Toward a Theory of Economic Development* (Holden-Day, San Francisco)

Note: references marked • are not specifically referred to in the text but are useful background reading.

Environment and Planning A, 1978, volume 10, pages 537-560

Regional multiplier analysis: a demometric approach

J Ledent
International Institute for Applied Systems Analysis, Laxenburg, Austria
Received 17 January 1978

Abstract. This paper reports on the design and testing of an adequate framework for conducting regional multiplier studies in areas experiencing rapid population growth. It puts forward the *demometric* approach, one that applies econometric methods to the analysis of demoeconomic growth.

Two alternative models are proposed here. The first is an aggregate model presenting a demometric revision of the traditional economic-base model. The second model, an enlarged version of the first, is characterized by a breakdown of economic activities into nine major sectors. Both models are fitted to data for the rapidly growing metropolitan area of Tucson, Arizona, USA. They are then used to derive tentative impact and dynamic multipliers which substantiate the role of households as consumers and suppliers of labor in the development of the Tucson Standard Metropolitan Statistical Area (SMSA). The major finding is that, for the same level of resources, the second model yields better policy implications than the modified (and therefore also than the traditional) economic-base model.

1 Introduction

In recent years, national and regional policymakers have made a serious effort to integrate regional data bases and analytical tools into the decisionmaking process. They have placed increasing demands on public agencies to complete regional multiplier analyses aimed at determining the economic and social implications of both national policies and regional projects.

Fundamental to these studies has been the attempt to identify the growth-inducing sectors of regional economies and to understand how mechanisms of change take place and are transmitted to other regional sectors. A striking feature of these studies is their heavy reliance on a Keynesian demand approach to regional development that emphasizes the growth-inducing role of firms while according only a cursory treatment to the role of households (that is, to demographic aspects). To be sure, most of the past multiplier analyses were related to areas in which population growth was relatively moderate and the role of demographic factors in regional development was somewhat difficult to identify. However, if such analyses had been performed for rapidly growing areas such as Arizona, Florida, or their subdivisions, the results would have certainly been erroneous and misleading for policymaking purposes.

Traditional approaches for conducting such analyses rely on either input-output models or economic-base models. Generally versions of these models include the demand effects of population growth through household consumption, but neglect the effects of households as suppliers of labor. Because we expect that the larger the population growth, the more important these effects are, it is desirable to accord a better treatment to demographic variables, especially in cases of rapidly growing regions, in order to obtain a fuller picture of the mechanisms of regional development that lead to significant policy implications.

In the light of this, this paper begins by showing the inadequacy of the traditional economic-base approach[1] in determining the consequences of government intervention

[1] The use of an input-output model is here ruled out since this model requires large inputs of time and money for the preparation of an input-output table.

in rapidly growing regions. It then proposes an amended economic-base approach that presents a demometric revision of the dichotomy between basic and nonbasic sectors. A fuller demometric model that leads to the derivation of multipliers by broad industrial sectors (rather than of aggregate multipliers as in the modified economic-base model), is outlined next. Both models are fitted to data for the rapidly growing metropolitan area of Tucson, Arizona (that is, the Tucson SMSA, a political unit also known as Pima County[2]) and relevant policy implications are discussed in both cases.

Before turning to the presentation and discussion of these alternative models, the demometric philosophy that underlies this study is briefly reviewed.

2 The demometric approach in a regional setting

The demometric approach is one that applies econometric methods to the analysis of the demoeconomic growth of a region.

> "Its principal objective is to establish quantitative statements regarding major demographic variables that explain the past behavior of such variables or that forecast (i.e., predict) their future behavior" (Rogers 1976b).

Formally the demometric approach calls for the construction of regional macro-demoeconomic models covering major components of regional growth (birthrates, migration rates, employment, output, population) but emphasizing the clearing of the local labor market which provides the connection between net migration and labor force dynamics. Fundamentally such models are characterized by the coupling of an economic model and a demographic model by means of two main linkages.

> "The former linkage appears in the form of a consumption function that demands the economy to produce a certain output for the population to consume. The latter linkage takes the form of a migration–labor force equilibrating the model that views the demographic model as the supplier of labor and the economic model as the demander of labor. The two models operate recursively in developing forecasts of demographic and economic growth that are internally consistent" (Rogers, 1976a).

The traditional economic-base theory of regional development implies a demand view of economic growth that is an insufficient framework for such demoeconomic models. Indeed the implementation of such models requires the availability of a more general theory of regional development. The recent literature, in fact, displays a growing current of dissatisfaction with economic-base theory as a suitable explanation for local development. Since Borts and Stein (1964) first suggested the argument that households, rather than industries, determine the evolving spatial pattern of development through their roles as suppliers of labor, increasing consideration has been given to labor-market conditions in regional studies. This has produced an important debate, namely the identification of the sources of local growth as illustrated by the "chicken-or-egg" controversy in recent migration literature (Muth, 1971; Mazek and Chang, 1972).

What is the relationship between population growth (net in-migration) and employment growth? Are migration rates induced by differential rates of employment growth, as argued by the proponents of the aforementioned demand view of local development? Or does the path of causation go the other way around, as advocated by the supporters of the alternative supply view? Clearly the two paths of causation between migration and employment growth that these two polar views underline are not mutually exclusive but coexistent. Recent evidence, suggested by the findings of

(2) Both designations will be used interchangeably hereafter.

several empirical studies (Olvey, 1972; Greenwood, 1973; Kalindaga, 1974), indicates that migration and employment growth affect each other, with perhaps the dominant influence being that of migration on employment growth.

A large body of literature is available about labor-force dynamics. Much of it is directed toward proving and disproving the 'added worker' and 'discouraged worker' hypotheses. However, a significant shift in the direction of research that could be profitable to the development of regional demometric models has recently occurred as researchers have started directing their attention to the job-search process itself (Miron, 1978).

To summarize, the recent literature both in migration analysis and regional labor-force dynamics suggests a starting point for the construction of demoeconomic models of regional growth that could constitute an adequate framework for deriving meaningful policy implications. However, the development of such models remains difficult because of inadequate data on migration and labor-force flows on a time-series basis. To justify this paper, it should be pointed out that the only accountable demographic data for Tucson available on a time-series basis are *aggregate* data (that is, relating to the whole population); these consist of net migration, labor-force, and unemployment-rate figures. As this is clearly insufficient to formulate and test a definite connection between migration and labor-force dynamics, it becomes necessary to redirect the strategy. With it borne in mind that a simple tool such as an economic-base model strongly appeals to regional planners in spite of its weaknesses, it was decided to adopt a compromise between such an economic-base model and the demometric model that one would ideally build for the Tucson SMSA. This led to the construction of two alternate models:

(1) an amended version of the traditional economic-base model, containing explicit labor-force variables and introducing a demometric revision of the separation of the basic sector; and

(2) a fuller demometric model that takes advantage of the disaggregated data, relating to employment, the labor force, and population, which are available for the Tucson SMSA.

3 Multiplier analysis for Tucson: a demometric revision of the economic-base approach

To examine to what extent the economic-base approach to regional analysis can be adapted to the case of a rapidly growing region, first the highlights of the traditional economic-base approach are recalled and then its limitations are examined in order to uncover the sensitive elements that one has to modify to produce an amended version applicable to the Tucson SMSA.

3.1 *The traditional economic-base approach*

In general terms, the economic-base approach assumes that local economies operate on two scales: transactions either (1) take place internally, that is, involve the recycling of money already in the local economy; or (2) they concern a product that is exported or purchased by an outsider, that is, they require the importation of money from outside the considered area. The latter are called *basic* activities because the money that they bring into the local economy supposedly leads the growth and expansion of economic activity.

The conceptual basis of the analysis assumes that the amount of activity in the basic sector determines the amount of activity in the nonbasic sector, and thus in the whole economy. The general relationship between basic and nonbasic sectors can then be expressed as

$$E_{\mathrm{n}} = \alpha + \beta E_{\mathrm{b}} , \qquad (1)$$

in which E_b and E_n are respectively basic and nonbasic employment, and α and β are appropriate coefficients. Since total employment (E_t) is given by the identity

$$E_t = E_b + E_n \ , \tag{2}$$

then

$$E_t = \alpha + (1+\beta)E_b \ , \tag{3}$$

where the coefficient $1+\beta$ defines the total employment that would be generated by the creation of one employment unit in the basic sector. This coefficient is generally referred to as the economic-base multiplier. The economic-base model is thus a simple framework describing the process of local development in terms of an assumed connection between economic sectors separated into two mutually exclusive sectors. Very often an additional equation linking total population to total employment by some kind of 'activity rate' permits a translation from the economic aspect of local development—embodied in the relationship linking basic and dependent sectors—to an alternative aspect, namely, population growth. Note that, in such instances, population change, being merely a consequence of employment change, has no impact of its own on the overall development of the region.

This shortcoming can be remedied by introducing some feedback effects from population change to employment change through the explicit consideration of household consumption (Czamanski, 1964). In such instances, nonbasic employment is expressed as an increasing function of both basic employment and population:

$$E_n = \alpha + \beta E_b + \gamma P \ , \tag{4}$$

where P is the total population, and γ a suitable coefficient. The model is completed by adding an equation in which P is made dependent on E_t:

$$P = a + bE_t \ , \tag{5}$$

where a and b are suitable coefficients. This expresses the assumption that labor supply (for which P constitutes a proxy) is always forthcoming as demanded by employment growth. By solving for total employment and population as a function of basic employment, one gets

$$E_t = \frac{\alpha + \gamma a}{1 - b\gamma} + \frac{1+\beta}{1-b\gamma} E_b \ , \tag{6}$$

and

$$P = \frac{a + b\alpha}{1 - b\gamma} + \frac{b(1+\beta)}{1-b\gamma} E_b \ , \tag{7}$$

which give the following multipliers with which to estimate the consequences of job creation in the basic sector:

$$\frac{\Delta E_t}{\Delta E_b} = \frac{1+\beta}{1-b\gamma} \tag{8}$$

and

$$\frac{\Delta P}{\Delta E_b} = \frac{b(1+\beta)}{1-b\gamma} \ . \tag{9}$$

The economic-base approach involves many practical and theoretical problems. From a practical point of view, a prerequisite to the use of the economic-base approach is the identification and measurement of the economic-base sector. Such a task, however, generally cannot be performed with commonly available data.

If available resources permit, a special survey can be carried out to separate the basic and dependent segments in the major sectors. Otherwise, the identification of the basic segment must be made by using nonsurvey methods.

From a theoretical point of view, the questions raised can be classified into two broad categories. The first includes problems which stem from the simple formulation of the traditional economic-base model and which can perhaps be amended when dealing with a fast-growing area: the focus on a demand-oriented view of regional growth, and the static character of the relationships between employment and population variables. The second category of problems consists of all the questions inherent in the economic-base concept itself; questions that can be removed only by adopting an alternative approach (these problems are examined at the beginning of section 4 of this paper).

3.2 *Toward a modified economic-base model for Tucson*

The use of the economic-base model in section 3.1 to calculate multipliers for the Tucson SMSA is likely to be insufficient if not misleading. In view of the recent evolution of economic and demographic growth in that area(3), it is clear that the additional jobs that could be created as a consequence of government intervention would not only go to residents but also to new in-migrants attracted by these new prospects. Thus the creation of additional basic jobs would bring to Tucson an additional population (and thus labor force) that could exceed, at least in the short term, the population change that would result from the application of formula (9). This would undoubtedly have an impact on the area's labor market and thus affect its development.

To deal with a region in which a majority of additional jobs are likely to be taken by nonresidents therefore requires a modification of the demand-oriented view of regional growth in at least two ways. The first consists of assessing the effect of a large pool of readily available workers on the growth of labor demand; the second relates to including the consequences of relative shortages and surpluses of labor on the expectations of workers.

In view of the constraint created by the existence of only two sectors in the regional economy (basic and nonbasic), the former improvement can be accommodated by proposing a method for identifying and measuring the basic sector in Tucson and by allowing for a consideration of the role of households as suppliers of labor. The latter improvement, on the other hand, which leads to a better consideration of the response of the Tucson and United States populations to changes in economic opportunities in Tucson, can be handled by amending the structure of the traditional economic-base model. For that purpose it is suggested that the role of households as suppliers of labor be explicitly introduced by means of labor-force variables (labor-force-participation and unemployment rates) and that an alternative to population equation (5) be chosen to show explicitly the consequences of changes in employment opportunities on population change. Moreover, since some of these consequences are expected to be not contemporaneous but lagging, the resulting equation is likely to introduce a dynamic element that responds to another of the criticisms directed at the traditional economic base.

The improvements envisioned for the application of the economic-base-model approach to the case of the Tucson SMSA leads to the construction of a model slightly more complicated than the traditional economic-base model, and one which appears as a small-scale dynamic econometric model (or demometric model, since several demographic variables are to be explicitly introduced). Fortunately such a structural change permits one to derive meaningful multipliers with a time dimension.

(3) An overview of Tucson's development over the last quarter of a century appears in appendix A.

These are labelled as 'impact and dynamic multipliers'[4], and they are obtained by simple matrix calculus in the case of a linear model or by comparing a 'control' solution of the model with the 'perturbed' solutions in the case of a nonlinear model (Goldberger, 1959). Before turning to the description of the structure of the model, first an attempt is made to estimate the basic sector of Tucson, so as to implement the first suggested modification of the traditional economic-base model.

3.3 *Estimation of the basic sector in Tucson*

The two most popular methods for estimating the basic sector of a regional economy are the location-quotients method and the minimum-requirements method. Both have serious drawbacks because of their restrictive assumptions regarding what constitutes basic activity. Moreover they do not lend themselves to a simple modification to account for the specific character of Tucson's development. Here an alternative econometric method is presented, based on an extension of an idea first proposed by Mathur and Rosen (1972).

3.3.1 *The Mathur/Rosen method.* The Mathur/Rosen method hypothesizes that, in each economic sector of a regional economy, that part of employment which is basic is sensitive to changes in total employment in the nation (NEMP). The procedure used to separate basic and nonbasic employment is as follows. Assume that

$$E_i = \beta_{i0} + \beta_{i1}\,\mathrm{NEMP} + e_i\,, \tag{10}$$

where E_i is the employment in the ith industry, e_i is the stochastic disturbance term, and β_{i0} and β_{i1} are appropriate coefficients. Applying OLS (ordinary-least-squares) estimators, one can obtain a regression equation:

$$\hat{E}_i = \hat{\beta}_{i0} + \hat{\beta}_{i1}\,\mathrm{NEMP}\,. \tag{11}$$

Properties of the estimators are such that

$$\overline{E}_i = \hat{\beta}_{i0} + \hat{\beta}_{i1}\overline{\mathrm{NEMP}}\,, \tag{12}$$

where $\overline{E}_i$ and $\overline{\mathrm{NEMP}}$ are the averages (means) of E_i and NEMP, respectively, over the observed period.

Mathur and Rosen assume that the ratio of basic employment in the region's industry to total employment in the nation remains constant over the entire period, and they define the proportion of basic employment in industry i to be $\hat{\beta}_{i1}\overline{\mathrm{NEMP}}/\overline{E}_i$.

A careful examination of this procedure reveals two serious problems.

(1) The assumption of a constant ratio of basic to total employment is unrealistic in view of the processes of regional growth that typically occur in market economies.

(2) The proportion of nonbasic employment in the ith industry is then equal to

$$1 - \hat{\beta}_{i1}\frac{\overline{\mathrm{NEMP}}}{\overline{E}_i} = \frac{\hat{\beta}_{i0}}{\overline{E}_i}\,. \tag{13}$$

Thus the separation of employment into its basic and dependent components for each industrial sector depends on the sign and magnitude of $\hat{\beta}_{i0}$. The Mathur/Rosen method does not ensure that $\hat{\beta}_{i0}$ will be positive and less than $\overline{E}_i$: as a matter of fact, in the case of local industries in which employment grows faster than national total employment, it can be shown that the intercept of the above regression tends to be negative. Mathur and Rosen then recommend the plotting of $\ln E_i$ (instead of E_i)

[4] Impact multipliers reflect the *instantaneous* effects of a change in an exogenous variable on an endogenous variable, whereas dynamic multipliers relate to the *delayed* effects that would result if the initial shock to the system, imposed by the change in the exogenous variable, is sustained over time.

against NEMP. However, the intercept would tend to be positive only when $\ln E_i$ does not grow faster than NEMP.

The method proposed by Mathur and Rosen produces a separation of economic sectors that simply reflects the relative growth rates in the industry-specific employments of the local economy and in national employment in all sectors. Also the actual figures that it yields depend heavily on the choice of the equations fitted to the data (for example, whether they are linear or nonlinear). Consequently such a separation into basic and nonbasic sectors can produce arbitrary results and casts some doubts about the method's robustness. Nevertheless this method offers a starting point for an improved econometric method for identifying the basic sector.

3.3.2 *An extension of the Mathur/Rosen method.* Now the assumption of a constant ratio of basic to total employment is relaxed and an attempt is made to incorporate the earlier observation that the multiplier process is the reduced form of a process that involves an active participation of households through demand and supply effects. An obvious candidate to replace the typical Mathur/Rosen stochastic equation is

$$E_i = \beta_{i0} + \beta_{i1}\,\mathrm{NEMP} + \beta_{i2}\,\mathrm{POP} + \beta_{i3}\,\mathrm{LFPR} + e_i\,, \qquad (14)^{(5)}$$

where POP is the local population (a mixed demand/supply effect) and LFPR is the local labor-force-participation rate (supply factor).

To avoid the difficulties encountered by Mathur and Rosen with the intercept, here the focus is on *changes* in employment rather than on levels, and the percentage of employment change that is basic in nature is defined as

$$\frac{\hat{\beta}_{i1}\,[\mathrm{NEMP}(t+k) - \mathrm{NEMP}(t)]}{\hat{E}_i(t+k) - \hat{E}_i(t)} \times 100\,, \qquad (15)$$

where

$$\hat{E}_i(y) = \hat{\beta}_{i0} + \hat{\beta}_{i1}\,\mathrm{NEMP}(y) + \hat{\beta}_{i2}\,\mathrm{POP}(y) + \hat{\beta}_{i3}\,\mathrm{LFPR}(y) \qquad (16)$$

and t and $t+k$ are the first and last years of the fitting period.

If the coefficient of one of the regional variables is not statistically significant, the variable is discarded and a new regression is run without it. If the coefficient of LFPR is not significant in equation (14), the substitute form for the regression equation is

$$E_i = \beta_{i0} + \beta_{i1}\,\mathrm{NEMP} + \beta_{i2}\,\mathrm{POP} + e_i\,, \qquad (17)$$

and the percentage of employment change which is basic in nature would be given by formula (15) in which $\hat{E}_i(y)$ would now be

$$\hat{\beta}_{i0} + \hat{\beta}_{i1}\,\mathrm{NEMP}(y) + \hat{\beta}_{i2}\,\mathrm{POP}(y)\,. \qquad (18)$$

Alternatively, if the coefficient of POP is not significant in equation (14), the substitute stochastic equation is

$$E_i = \beta_{i0} + \beta_{i1}\,\mathrm{NEMP} + \beta_{i2}\,\mathrm{LFT} + e_i\,, \qquad (19)$$

where LFT is the total local labor force. The percentage of employment change that is basic would be obtained in a similar way.

If $\hat{\beta}_{i1}$ (the coefficient of the national employment variable) fails to be significant in any of the above formulations then the employment change is regarded as being nonbasic.

(5) The estimation of the corresponding regression equation is expected to yield three positive coefficients: $\hat{\beta}_{i1}$, $\hat{\beta}_{i2}$, and $\hat{\beta}_{i3}$.

If neither of the coefficients of LFPR and POP are significant, or if $\hat{\beta}_{i2}$ is not significant in equations (14), (17), and (19) ($\hat{\beta}_{i1}$ being significant), then one would simply use

$$E_i = \beta_{i0} + \beta_{i1}\,\mathrm{NEMP} + e_i\,, \tag{20}$$

and qualify the whole employment change as basic if the $\hat{\beta}_{i1}$ obtained is significant and as nonbasic otherwise.

This method has been applied to the Tucson economy, for which annual employment data were available for the period 1957–1975. In the case of the service sector (SERV), for example, by applying the OLS estimation with correction for first-order autocorrelation, the following regression equation was obtained(6):

$$\underset{}{\mathrm{SERV}} = \underset{(-23\cdot403)}{-30\cdot340} + \underset{(4\cdot4516)}{(0\cdot130 \times 10^3)}\mathrm{NEMP} + \underset{(16\cdot436)}{0\cdot059}\mathrm{POP} + \underset{(14\cdot571)}{53\cdot892}\mathrm{LFPR}\,;$$

period = 1957–1975; mean = 16·679; $\rho = -0\cdot089$;

$R^2 = 0\cdot999$; $\overline{R}^2 = 0\cdot999$; SE = 0·253; DW = 1·298; $F(3, 14) = 4296\cdot0$.

From this equation it can be established that only 16·7% of the variations in service employment are explained by variations in the national employment variable. This is consistent with the a priori expectation that the service industry is a nonbasic-oriented industry. Table 1 shows the results calculated using this approach, and indicates that industrial employment changes in the Tucson SMSA may be classified as being: totally basic in manufacturing; partially basic in mining, transportation and communication, trade, services, and various levels of government; and completely dependent (nonbasic) in agriculture, construction, and finance and real estate. In general these figures confirm a priori expectations about the basic or nonbasic character of each industry. An exception to this appears in agriculture, probably because in the Tucson SMSA this sector employs a small number of workers which has remained approximately constant over the period of observation.

Table 1. Percentages of employment change that are basic, according to sectors, for the Tucson SMSA, 1957–1975.

Sector	Percentage	Equation used	Sector	Percentage	Equation used
Agriculture	0·0	(20)	Trade	19·6	(14)
Mining	57·4	(17)	Finance, insurance, and real estate	0·0	(20)
Manufacturing	100·0	(20)			
Construction	0·0	(20)	Service	16·7	(14)
Transportation and communication	33·0	(17)	Local and state government	22·3	(19)
			Federal government	57·6	(19)

(6) The statistics in parentheses under each regression coefficient are the corresponding t statistics. The employment figures are expressed in thousands. The mean is the mean of the employment variable for the appropriate sector, in this case it is $\overline{\mathrm{SERV}}$; ρ is the first-order autocorrelation of the disturbances; R^2 and $\overline{R}^2$ are the coefficients of multiple correlation, unadjusted and adjusted for degrees of freedom; SE is the standard error; DW is the Durbin–Watson statistic; and F is the F distribution (evaluated as is appropriate).

3.4 *The modified economic-base model: description and construction*

Now the model is presented: it consists of eight equations (four identities and four stochastic equations estimated by regression analysis from time-series data for the Tucson SMSA) and includes eleven variables, three of which (change in basic employment, national unemployment rate, and a time trend) are necessarily exogenous.

Equation 1 in table 2 accounts for variations in total employment change by relating these to the two independent variables: employment change in the basic sector and the net migration level. The coefficients of both these variables should be positive[7].

Table 2. Modified economic-base model of the Tucson SMSA. All figures are expressed in thousands unless otherwise indicated. The form of equations 1-4 is explained in footnote (6). The (−1)s after certain variables denote that there is a one-year lag attached to these variables.

Regression equations

1 $$\underset{(4\cdot348)}{\text{DEMPT} = 1\cdot857\text{DBASIC}} + \underset{(2\cdot344)}{0\cdot219\text{NETMIG}}$$

mean = 4·350; ρ = 0·172; R^2 = 0·790; SE = 2·198; $F(2,15)$ = 56·43

2 $$\text{NETMIG} = \underset{(2\cdot586)}{0\cdot537\text{DEMPT}} - \underset{(-3\cdot936)}{(3\cdot729\times10^2)\text{UNR}(-1)} + \underset{(4\cdot890)}{(4\cdot530\times10^2)\text{NUNR}(-1)}$$

mean = 8·038; ρ = 0·307; R^2 = 0·834; SE = 2·915; $F(3,14)$ = 35·17

3 $$\text{NATINC} = \underset{(10\cdot930)}{0\cdot025\text{POP}(-1)} - \underset{(-4\cdot325)}{0\cdot344\text{TIME}}$$

mean = 3·918; ρ = 0·730; R^2 = 0·949; SE = 0·293; $F(2,16)$ = 276·57

4 $$\text{UNR} = \underset{(8\cdot257)}{0\cdot509\text{UNR}(-1)} + \underset{(11\cdot914)}{0\cdot599\text{NUNR}} - \underset{(-7\cdot324)}{0\cdot169\frac{\text{DEMPT}}{\text{EMPT}(-1)}}$$

mean = 4·844; ρ = 0·157; R^2 = 0·967; SE = 0·346; $F(3,14)$ = 205·93

Identities

5 $\text{EMPT} = \text{EMPT}(-1)+\text{DEMPT}$

6 $\text{POP} = \text{POP}(-1)+\text{NATINC}+\text{NETMIG}$

7 $\text{LFT} = \dfrac{\text{EMPT}}{1-\text{UNR}}$

8 $\text{LFPR} = \dfrac{\text{LFT}}{\text{POP}}$

Meaning of the variables

Endogenous (local) variables

UNR	unemployment rate
DEMPT	change in total employment
EMPT	total employment
LFPR	overall labor-force-participation rate
LFT	total labor force
NATINC	natural increase of population between 1 July of year $t-1$ and 1 July of year t
NETMIG	net in-migration between 1 July of year $t-1$ and 1 July of year t
POP	population at 1 July of year t

Exogenous variables

DBASIC	change in basic employment between year $t-1$ and year t
TIME	time trend (equal to 1 in 1956, and 20 in 1975)
NUNR	national unemployment rate

(7) Note that both employment-change variables are contemporaneous (that is, they express changes in employment observed in two successive years, $t-1$ and t). The net-migration variable is an estimate of this population component of change between 1 July in year $t-1$ and 1 July in year t. Thus no attempt is made to test for the occurrence of long delays in the responses of economic agents to changes in economic conditions.

Population growth has been broken down into its main components of change. Net migration (equation 2 in table 2) is tied to employment change and to the difference in the economic conditions that prevail at both the local and national levels as reflected by their differential unemployment rates (see figure 1 for a comparison of the evolution of these variables). Note the one-year lag attached to the local and national unemployment rates. The employment change variable in equation 2 is expected to be positively correlated with the dependent variable (the greater the job opportunities, the greater the attraction of Tucson for migrants). The local unemployment rate is expected to have a negative coefficient, and the national unemployment rate should have a positive coefficient.

Natural increase in population (equation 3 in table 2) is described by a simple regression equation in which a time trend (expected to be negative) should express the observed decreasing tendency of Tucson's rate of natural increase. (Such a treatment is justified by the greater importance of structural changes vis-à-vis changes across the business cycle in explaining fertility decline in the recent history of the United States.)

The introduction of labor-market-related variables into the model, and the characterization of the interaction of population and employment growth mainly through the impact of the labor-market surplus (unemployment), raises a consistency problem that can be summarized with the aid of figure 2. Clearly no model can independently predict all five variables in figure 2 since these variables are related by two definitional relationships: those defining labor-force-participation rates and unemployment rates. Inevitably this means that two of the five variables have to be calculated as residuals. Perhaps the obvious candidates for residuals are the unemployment rate and the labor-force-participation rate since they are not primary variables. However, when they are calculated as residuals, they often take on absurd

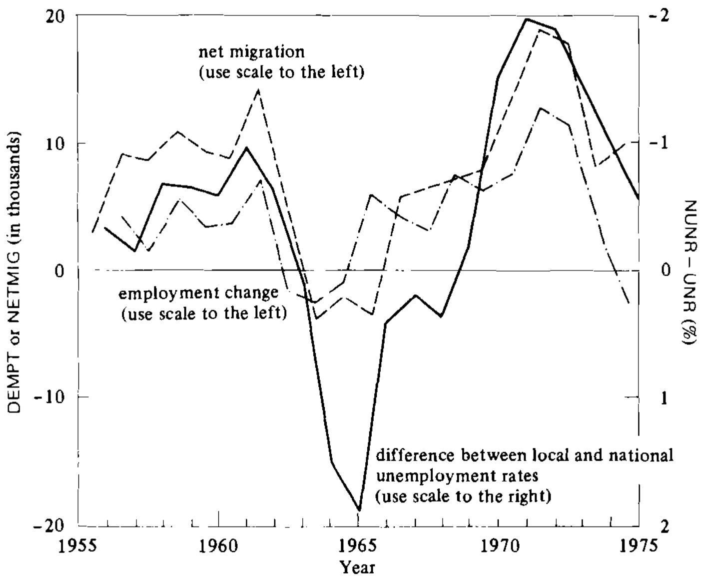


Figure 1. Comparison between net migration, employment change, and the difference between local and national unemployment rates for the Tucson SMSA, 1956-1975. (Sources: employment change and local unemployment rates—Arizona Department of Economic Security, 1976; net migration—derived by the author from data published by Arizona Department of Economic Security, 1976; national unemployment rates—various issues of *Monthly Labor Review*, published by the US Department of Labor, Washington, DC.)

values, especially the unemployment rates(8), and then the population and employment variables might not be consistent. Here the variable most likely to create problems is the unemployment rate since population and employment have been chosen as primary variables. This means that total labor force and the labor-force-participation rate should be residuals. The values of these residuals might not be plausible relative to obvious trends in labor-force participation, but it is likely that the discrepancy would be smaller than for any other choice of the dependent variables.

The fourth stochastic equation (in table 2) expresses the variations of the local unemployment rate, and relates these to variations in variables such as national unemployment, relative increase in total employment, and a lagged value of the dependent variable(9).

The model is completed by adding four identities (numbered 5–8 in table 2). The first two show that the current levels of the employment and population variables are obtained by adding their current components-of-change levels to their previous-year levels. The last two give the two residual variables: labor force and labor-force-participation rate.

The stochastic equations of the model have been fitted to annual data for the period 1956–1975 using both ordinary least squares (OLS) and two-stage least squares (2SLS). Since they both lead to similar estimates, here only the 2SLS estimates are reported. In theory these are the only appropriate ones in light of the simultaneous nature of the model.

The actual fits of the employment-change and net-migration equations show high values for the t statistics(10) (see table 2), thus indicating a large significance for the independent variables (which moreover have the expected sign in all circumstances). However, the overall performance of these equations in terms of their coefficients of determination is less satisfactory. This is not surprising when one considers the volatile character of the net-migration variable, and the aggregate nature of the relationship between basic and nonbasic sectors.

With the estimation stage complete, the next step in the construction of the model consists of carrying out the simulation. Although the model is nonlinear, the use of a Gauss–Seidel iterative method is not necessary because the final form of the model can be calculated(11). This permits one to simulate the model over the sample period

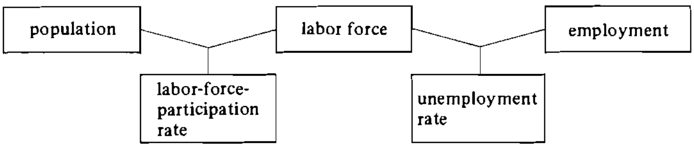


Figure 2. The basic relationship between the main demographic and economic variables in a consistent demoeconomic model.

(8) For example, if one predicts labor force with a one percent error (overestimation) and employment with also a one percent error (underestimation), the error made on the forecast of the corresponding unemployment rate is forty percent (if the unemployment rate is equal to five percent).

(9) Such a specification of the unemployment equation may be found in Chang (1976).

(10) When estimating the four stochastic equations of the model, the intercepts appeared to have very low t statistics. The corresponding equations were then refitted with zero intercepts. These versions were finally retained because they improved the overall performance of the regression equations while modifying the parameter estimates only slightly.

(11) This is so because (1) the nonlinear variables (labor force and labor-force-participation rate) appear separately from the main block and (2), in this main block, the dependent variables are linear functions of the current independent variables.

and then to compare actual versus predicted values of the endogenous variables. The precision obtained is judged sufficiently accurate given the aggregate nature of the model. It is possible to compare a perturbed simulation of the model, obtained after an exogenous increase in basic employment, with the base-run simulation, and to derive aggregate impact and dynamic multipliers.

3.5 *Impact and dynamic multipliers obtained with the modified economic-base model*
Table 3 shows impact and dynamic multipliers (derived by simulation procedures) associated with the creation of jobs in the basic sector in 1958, 1967, and 1974 respectively. It suggests the following observations.

(1) The consequences for total employment and population seem to be relatively high. For example, the corresponding five-year dynamic multipliers have maximum values of 3·16 and 6·31, respectively, for a unit exogenous change in the basic sector in 1958.

(2) Comparisons of the impact multipliers with the five-year dynamic multipliers show that the delayed effects of job creation are relatively moderate and tend to diminish as the size of the local economy increases. For example, the economic-base multiplier ΔEMPT/ΔBASIC decreases from 3·16 for job creation in 1958 to 2·57 in 1974.

(3) A comparison of the dynamic population and employment multipliers indicates that the delayed effects of job creation are proportionally higher in the case of population than in the case of employment.

(4) As expected, the effect of additional jobs in the basic sectors is to diminish the unemployment rate and to increase the labor-force-participation rate. However, if the increase in the labor-force-participation rate (as indicated by impact multipliers) is maintained well over the next few years (as indicated by the dynamic multipliers), the downward pressure on the unemployment rate will tend to decline as the dynamic multipliers for this variable tend toward zero[12].

Table 3. Impact and dynamic multipliers from the modified economic-base model. (Recall that UNR is measured as a percentage.)

Multiplier	Year of measurement	Value of multiplier for creation of basic jobs in appropriate year when *T* is		
		1958	1967	1974
$\frac{\Delta EMPT}{\Delta BASIC}$	T	2·10	2·10	2·10
	$T+2$	2·85	2·68	2·46
	$T+4$	3·16	2·87	2·57
$\frac{\Delta NETMIG}{\Delta BASIC}$	T	1·13	1·13	1·13
	$T+2$	4·53	3·76	2·74
	$T+4$	5·95	4·64	3·25
$\frac{\Delta POP}{\Delta BASIC}$	T	1·13	1·13	1·13
	$T+2$	4·63	3·85	2·82
	$T+4$	6·31	4·95	3·48
$\frac{\Delta UNR}{\Delta BASIC} \times 100$	T	−0·49	−0·38	−0·24
	$T+2$	−0·21	−0·14	−0·08
	$T+4$	−0·04	−0·02	−0·02
$\frac{\Delta LFPR}{\Delta BASIC} \times 100$	T	0·62	0·44	0·31
	$T+2$	0·50	0·40	0·31
	$T+4$	0·42	0·35	0·29

(12) This result only occurs when job expansion takes place in the later years of the sample period. For basic employment creation in the earlier years, the unemployment-rate multipliers become positive and tend to diverge. Here is a case of nonconverging dynamic effects.

(5) A comparison of the values of the multipliers, for different years of occurrence, in job creation leads one to observe that impact multipliers remain the same because of the structure of the model (simultaneous links are specified with linear equations having constant coefficients), and the magnitude of the dynamic multipliers displays a tendency to decrease as the occurrence of additional employment is retarded. This last finding is not surprising. It makes sense that the marginal effect of a given increase in basic employment diminishes as the size of Tucson increases. But which characteristics of the model account for such a result? A careful examination of the interaction between equations reveals that the specification of the unemployment-rate equation is mainly responsible for this result. Relative employment change is a determining variable with regard to the employment rate.

4 Multiplier analysis for Tucson: a fuller demometric approach

Although the preceding aggregate model has included some corrective elements not generally found in classic economic-base theory, its use for policy analysis remains questionable for reasons inherent in the nature of the dichotomy between basic and dependent sectors. The vagueness of the notion of the basic sector leaves room for a broad interpretation and this does not facilitate transferring the macro point of view of economic-base identification to the micro level. Some jobs are clearly basic (production of steel shipped outside the local area); others are clearly nonbasic (teaching in a locally oriented primary school). In many instances, however, the classification of a job as basic or nonbasic is impossible.

Another question raised by the economic-base concept is the aggregative nature of the associated multipliers which express an average of the multiplier effects induced by changes in the basic sector as a whole. As a consequence, the multipliers may not be applicable to a particular industry and thus may result in an inadequate estimation of the effects generated by the construction of a given factory.

It is therefore desirable to abandon the basic/nonbasic dichotomy and to adopt an alternative approach which is suggested by the findings of the demometric identification of the basic sector in Tucson. Since this identification requires the establishment of regression equations linking sectoral employment growth and population growth in the Tucson SMSA (the results of which were then used as exogenous information in a small-scale demometric model), it appears rational to make these equations an endogenous part of the model.

4.1 *A brief description of the structure of the demometric model*

The demometric model contains two main parts: an employment part and a demographic part. The employment part consists of an exogenous sector (agriculture) and nine endogenous sectors (mining, manufacturing, construction, transportation and communication, trade, finance and real estate, services, government, and self-employed). In accordance with the demometric philosophy, the actual equations acknowledge that external markets are not the only sources of growth, and that population growth through its demand and supply effects is a complementary growth factor.

The demographic part of the model determines actual births, deaths, and net migrants to obtain the new population every year. Although a disaggregation paralleling that of the economic side is highly desirable in determining these components of change, it cannot be implemented owing to the virtual nonexistence of time-series data on migration. This is unfortunate since the decomposition of net migration into its gross components (in- and out-migration) would have brought in useful information on the interaction of employment and population growth. A separation of retirement migration from employment-related migration also could not be implemented because of unavailable data.

Interactions between the demographic and economic parts of the model appear in both directions. The impact of employment growth on population growth occurs through economic variables lagged by one year (mainly local and national unemployment rates) with the intervening current variables being two employment-change variables (in the manufacturing and the construction sectors). In the reverse direction, a current or lagged population variable (level or change) affects most sectoral employments to generally express a mixed demand/supply situation.

Secondary feedback effects from the population part to the employment part are taken care of through per-capita income (an additional demand effect), a variable determined from population and employment changes, and the unemployment rate.

The model consists of twenty-five equations, eleven of which are identities (tables 4 and 5). The model was fitted to Tucson data for the period 1957–1975 by use of two-stage least-squares estimation with a built-in correction for first-order correlation. The regression equations thus obtained (see appendix B) display high coefficients of determination, the lowest values being observed in the cases of the unemployment-rate and net-migration equations ($\overline{R}^2 = 0{\cdot}926$ and $0{\cdot}939$ respectively).

Table 4. Structure of the demometric model: regression equations and identities.

Regression equations	
MANUF	$= f_1\left[\text{NEMP}, \dfrac{\text{EMPT}}{1-\text{UNR}(-1)}, \text{POP}, \text{DUM64}\rightarrow\right]$
MINING	$= f_2\left[\text{NEMP}, \dfrac{\text{EMPT}}{1-\text{UNR}(-1)}\right]$
CONST	$= f_3[\text{POP}(-2), \text{DPOP}(-1), \text{DPOP}, \text{MINING}, \text{DUM61}\rightarrow 62]$
TRANSP	$= f_4[\text{MANUF}+\text{MINING}, \text{WSEMP}-\text{MANUF}-\text{MINING}]$
TRADE	$= f_5[\text{POP}(-1), \text{DPOP}, \text{RPCI}]$
FIRE	$= f_6[\text{POP}, \text{UNR}]$
SERV	$= f_7[\text{POP}(-1), \text{DPOP}, \text{RPCI}]$
GOVT	$= f_8[\text{POP}, \text{POP} \times \text{TIME}]$
SELF	$= f_9[\text{DUM72}\rightarrow]$
DEATH	$= f_{10}\left[\dfrac{\text{POP}(-1)+\text{POP}}{2}, \text{TIME}\right]$
BIRTH	$= f_{11}\left[\dfrac{\text{POP}(-1)+\text{POP}}{2}, \text{TIME}\right]$
NETMIG	$= f_{12}[\text{DCONST}, \text{DMANUF}, \text{UNR}(-1), \text{NUNR}(-1), \text{H}(-1)]$
UNR	$= f_{13}\left[\text{UNR}(-1), \text{NUNR}, \dfrac{\text{DEMPT}}{\text{EMPT}(-1)}\right]$
RPCI	$= f_{14}\left[\text{RCPI}(-1), \dfrac{\text{DEMPT}}{\text{EMPT}(-1)}, \dfrac{\text{NATINC}}{\text{POP}(-1)}\right]$
Identities	
DMANUF	$= \text{MANUF}-\text{MANUF}(-1)$
DMINING	$= \text{MINING}-\text{MINING}(-1)$
WSEMP	$= \text{MANUF}+\text{MINING}+\text{CONST}+\text{TRANSP}+\text{TRADE}+\text{SERV}+\text{FIRE}+\text{GOVT}$
EMPT	$= \text{WSEMP}+\text{AGR}+\text{SELF}$
DEMPT	$= \text{EMPT}-\text{EMPT}(-1)$
LFT	$= \dfrac{\text{EMPT}}{1-\text{UNR}}$
H	$= \dfrac{\text{TRADE}+\text{SERV}+\text{FIRE}+\text{GOVT}}{\text{WSEMP}}$
NATINC	$= \text{BIRTH}-\text{DEATH}$
DPOP	$= \text{NATINC}+\text{NETMIG}$
POP	$= \text{POP}(-1)+\text{DPOP}$
LFPR	$= \dfrac{\text{LFT}}{\text{POP}}$

A frequently used indicator to measure the performance of the individual equations is the ratio of the standard error of the estimation to the mean of the dependent variable. Table B1 of appendix B indicates that standard errors are less than five percent of the mean of the dependent variable. The only exception occurs with the net-migration variable.

As its equations indicate, this model displays numerous nonlinearities, a feature that normally makes the use of an iterative (Gauss/Seidel) method necessary for the simulation stage. However, as in the modified economic-base model in section 3, the interaction between the variables of the model has been specified in a way that maintains linearity in the simultaneous links. Nonlinearities occur only in the delayed links. This permits a derivation of the final form of the model (although only with tedious analytical calculations) that leads to an easy simulation of the model.

Ex post forecasts were developed with this model to test its ability to replicate the past growth of Tucson. In addition, mean absolute percentage errors (MAPEs), which give an indication of the magnitude between the ex post forecasts obtained and the corresponding actual values, have been computed for each endogenous variable. As shown in table 6, low MAPEs were obtained for all variables except for net migration, unemployment, and employment in the construction sector. These are

Table 5. List of variables in the demometric model. All variables are expressed in thousands unless otherwise indicated.

Endogenous (local) variables	
UNR	local unemployment rate
BIRTH	number of births between 1 July of year $t-1$ and 1 July of year t
CONST	employment in the construction sector
DCONST	increase in construction employment between year $t-1$ and year t
DEATH	number of deaths between 1 July of year $t-1$ and 1 July of year t
DEMPT	increase in total employment between year $t-1$ and year t
DMANUF	increase in manufacturing employment between year $t-1$ and year t
DPOP	increase in population between 1 July of year $t-1$ and 1 July of year t
EMPT	total employment
FIRE	employment in finance, insurance, and real estate
GOVT	employment in the government sector
H	fraction of wage and salary employment in the household-serving sectors (undimensioned)
LFPR	local labor-force-participation rate (undimensioned)
LFT	total labor force
MANUF	employment in the manufacturing sector
MINING	employment in the mining sector
NATINC	natural increase of population between 1 July of year $t-1$ and 1 July of year t
NETMIG	net in-migration between 1 July of year $t-1$ and 1 July of year t
POP	local population on 1 July of year t
RPCI	real per-capita income for year t (dollars/National Consumer Price Index)
SELF	number of self-employed
SERV	employment in the service sector
TRADE	employment in the trade sector
TRANSP	employment in the transportation and communication sector
WSEMP	wage and salary employment
Exogenous variables	
AGR	agricultural employment
DUM61 → 62	dummy variable (1 in 1961 and 1962, 0 otherwise)
DUM72→	dummy variable (1 in 1972 and after, 0 otherwise)
NEMP	total national employment
TIME	time trend (1 in 1956, ..., 20 in 1975)
NUNR	national unemployment rate
DUM64→	dummy variable (1 in 1964 and after, 0 otherwise)

precisely the most volatile elements of Tucson's economy. However, a graphical comparison of the ex post forecasts with the actual data relating to these variables (figure 3) reveals that the model structure is adequate in its ability to replicate the annual variations in net migration and the unemployment rate.

Table 6. Evaluation of the ex post forecasting ability of the demometric model.

Variable	MAPE (%)	Variable	MAPE (%)	Variable	MAPE (%)
DEATH	1·89	EMPT	1·88	CONST	9·00
BIRTH	2·66	UNR	9·70	TRANSP	3·27
NETMIG	32·25	RCPI	1·73	TRADE	3·28
POP	1·47	SELF	1·68	FIRE	2·84
LFT	2·01	MINING	4·85	SERV	2·98
LFPR	1·32	MANUF	4·32	GOVT	5·57
WSEMP	1·17				

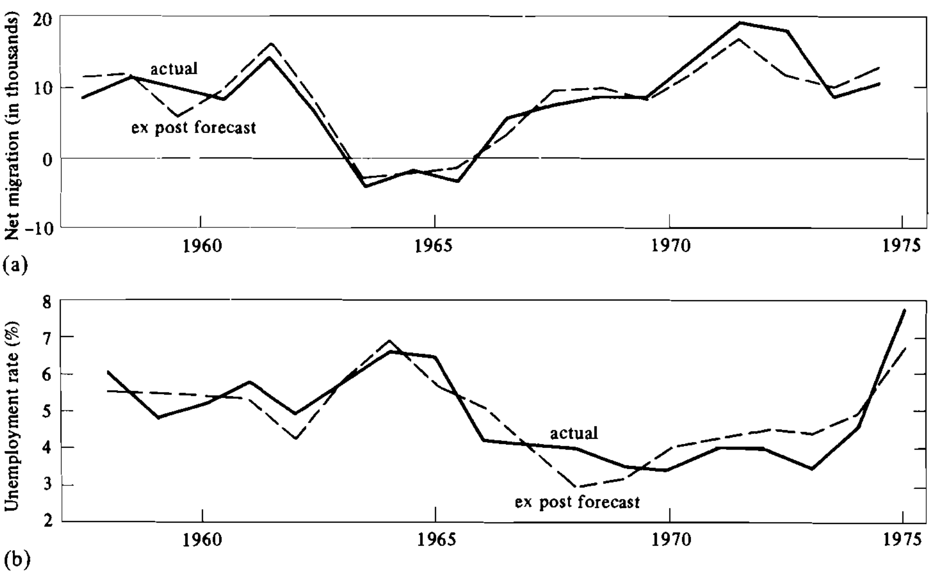


Figure 3. Demometric model of the Tucson SMSA: ex post forecasts of (a) net migration and (b) the local unemployment rate, compared with actual figures.

4.2 *Impact and dynamic multipliers obtained with the demometric model*

Table 7 summarizes the impact and dynamic multipliers resulting from exogenous increases in employment in the various sectors of Tucson's local economy in 1958 and 1976. These suggest the following observations.

1 A given exogenous increase in employment has the largest consequences on local development if it occurs in the *mining* sector. An expansion of mining employment by 1000 workers in 1976 leads to an increase in total employment of 5140 workers in 1978 and an increase in population of 7230 by the same year.

2 Exogenous employment expansions in the *manufacturing and construction* sectors also affect the growth of Tucson, but not as much as employment expansion in mining. The 1978 employment and population dynamic multipliers (for an exogenous increase taking place in 1976) are 3·62 and 5·95 respectively for job creation in manufacturing, and 3·39 and 5·95 respectively for job creation in construction (in contrast to 5·14 and 7·23 respectively for mining).

Table 7. Impact and dynamic multipliers from the disaggregate demometric model. The table shows impact and dynamic multiplier effects on the main variables (employment, net migration, population, unemployment rate, and labor-force-participation rate) from job creation in mining, manufacturing, construction, and trade. The upper of each pair of figures in each row refers to T = 1958, and the lower to T = 1976.

	Impact multiplier (year T)	Dynamic multipliers for year			
		$T+1$	$T+2$	$T+3$	$T+4$
Employment					
Mining	4·95	7·90	9·37	8·46	6·43
	4·37	4·98	5·14	4·92	4·83
Manufacturing	3·77	5·76	6·50	5·67	4·21
	3·40	3·71	3·62	3·44	3·38
Construction	3·66	5·52	6·12	5·26	3·83
	3·33	3·56	3·39	3·21	3·16
Trade	2·36	3·62	4·23	3·69	2·67
	2·03	2·32	2·46	2·31	2·21
Net migration					
Mining	2·88	7·82	6·36	2·63	−1·72
	2·76	2·83	1·32	0·41	0·12
Manufacturing	2·97	5·71	4·18	1·32	−1·55
	2·89	2·04	0·72	0·16	0·03
Construction	3·20	5·47	3·87	1·09	−1·63
	3·13	1·93	0·59	0·09	0·00
Trade	0·78	3·34	2·57	0·79	−1·17
	0·69	1·44	0·60	0·09	−0·09
Population					
Mining	2·93	10·94	17·68	20·82	19·64
	2·79	5·74	7·23	7·84	8·18
Manufacturing	3·01	8·88	13·36	15·06	13·89
	2·93	5·08	5·95	6·27	6·47
Construction	3·24	8·87	13·03	14·49	13·22
	3·17	5·22	5·95	6·20	6·38
Trade	0·79	4·20	6·92	7·90	6·93
	0·70	1·95	2·61	2·78	2·76
Unemployment rate (× 100)					
Mining	−1·19	−1·14	−0·73	0·05	0·57
	−0·50	−0·26	−0·12	−0·02	0·01
Manufacturing	−0·91	−0·80	−0·44	0·09	0·42
	−0·39	−0·19	−0·06	0·00	0·01
Construction	−0·88	−0·77	−0·40	0·11	0·42
	−0·38	−0·18	−0·05	0·00	0·01
Trade	−0·57	−0·51	−0·31	0·05	0·28
	−0·23	−0·12	−0·06	−0·08	0·01
Labor-force-participation rate (× 100)					
Mining	1·35	1·49	1·32	0·75	0·28
	0·58	0·56	0·52	0·45	0·40
Manufacturing	0·91	0·97	0·79	0·39	0·08
	0·39	0·35	0·35	0·26	0·22
Construction	0·85	0·89	0·71	0·32	0·27
	0·35	0·31	0·31	0·21	0·18
Trade	0·72	0·78	0·72	0·47	0·26
	0·32	0·32	0·32	0·29	0·27

3 Exogenous employment expansions in *all other sectors* have much lower consequences for the local economy. They are quite similar in all household-serving sectors (trade, finance and real estate, services, and government). Three-year dynamic employment and population multipliers for job creation in any of these sectors are respectively 2·46 and 2·61, and are only slightly higher for an initial expansion in the transportation and communication sector, 2·68 and 3·11 respectively.

4 In all cases, exogenous job creation sharply decreases the unemployment rate and increases the labor-force-participation rate in the year of occurrence, T. As time goes on, the effect on the unemployment rate tends to disappear, especially around years $T+2$ and $T+3$; the labor-force-participation rate tends to stabilize(13). Once again the magnitude of the multiplier varies by sectors in mining, manufacturing and construction.

5 Finally, on comparison of the values of the multipliers for different years of job creation (see table 7), it is observed that impact multipliers have a slight tendency to decrease as job creation is delayed(14), and, as in the modified economic-base model, the dynamic multipliers have a tendency to decline as job creation is retarded. In the present case, this tendency is even more pronounced(15).

Besides these five considerations suggested by the data set out in table 7, an interesting question arises from their comparison with the multipliers obtained with the modified economic-base model. Broadly speaking, tendencies displayed by the multipliers of both models are similar. This is to be expected since the structure of the demometric model is an immediate extension of that of the modified economic-base model. The disaggregated model leads to smaller impact effects and much larger dynamic effects, possibly because of an improved lag structure.

5 Conclusion

In this paper it has been argued that the use of *demometric* methods would improve the quality and accuracy of multiplier analyses for rapidly growing areas.

First, it was shown that one of the most widely used tools for constructing such analyses—namely the traditional economic-base model—presents serious limitations that makes its utilization undesirable in the case of such areas. These include

(1) a heavy reliance on a demand view of regional growth, ignoring the role of households as suppliers of labor;

(2) a quasi-static formulation of the regional-growth process; and

(3) an identification and measurement of the basic sector that is not applicable to rapidly growing urban economies.

Second, it was made clear that if the economic-base model was modified to accord a higher importance to demographic factors (the introduction of migration as a dynamic factor, the explicit consideration of labor-force and unemployment variables, the demographic revision of the measurement of the basic sector), it could lead to the derivation of more reliable impact and dynamic multipliers.

(13) As in the case of the modified economic-base model of section 3, the unemployment-rate multipliers display the same tendency toward instability when additional job creation occurs in the early years of the sample period.

(14) This slightly diminishing tendency contrasts with the constancy of impact multipliers in the aggregate model. Note that the cause of this difference lies in the different treatment of simultaneous links: although simultaneous links are expressed by means of linear equations in both models, the coefficients are constant in the modified economic-base model, but they depend on lagged variables in the demometric model.

(15) This larger diminishing tendency is the result of the larger importance accorded to the lag structure.

Third, it was demonstrated that it is possible to bypass a division into basic and nonbasic employment sectors not only by implementing an analysis having an economic-base flavor, but also by a more general application of the demometric philosophy—one that allows for the construction and application of a disaggregate model from which impact and dynamic multipliers, useful to policymakers in rapidly growing regions, can be obtained.

To summarize, the peculiarities of rapidly growing areas do not readily permit the use of simple multiplier analyses based on the traditional economic-base model. Alternative models are called for to obtain meaningful multipliers. The two models proposed in this paper, the modified economic-base model and the demometric model, require the same data and approximately similar resources. Since the disaggregate demometric model yields more and better information for the same investment of resources, this suggests that, ceteris paribus, a demometric model constitutes a more desirable alternative than a somewhat amended version of the economic-base model.

Acknowledgments. This work was initiated during my affiliation with the Division of Economic and Business Research, College of Business and Public Administration, University of Tucson, Arizona.

I wish to thank A Rogers for his generous advice and helpful comments on a preliminary version that was presented at the 17th European Conference of the Regional Science Association, held in Krakow, Poland, 23-26 August 1977. Thanks also go to T Kawashima and to R Dennis who made valuable suggestions.

References

Arizona Department of Economic Security, 1976 "Population estimates of Arizona" report number 8, Arizona Department of Economic Security, Phoenix, Arizona

Borts G H, Stein J, 1964 *Economic Growth in a Free Market* (Columbia University Press, New York)

Chang H S, 1976 "Tennessee econometric model: phase I" Center for Business and Economic Research, College of Business Administration, University of Tennessee, Knoxville, Tennessee

Czamanski S, 1964 "A model of urban growth" *Papers of the Regional Science Association* **13** 177-200

Goldberger A S, 1959 *Impact Multipliers and Dynamic Properties of Klein-Goldberger Model* (North-Holland, Amsterdam)

Greenwood M J, 1973 "Urban economic growth and migration: their interaction" *Environment and Planning* **5** 91-112

Kalindaga Y K, 1974 "Interdependence between employment growth and interregional migration: an empirical study of U.S. metropolitan areas" Ph D dissertation, Northwestern University, Evanston, Illinois

Mathur V K, Rosen H S, 1972 "An econometric export base model of regional growth: a departure from conventional techniques" in *London Papers in Regional Science 3. Patterns and Processes in Urban and Regional Systems* Ed. A G Wilson (Pion, London), pp 31-43

Mazek W F, Chang J, 1972 "The chicken or egg fowl-up in migration: comment" *Southern Economic Journal* **34** 133-142

Miron J R, 1978 "Job-search perspectives on migration behaviour" *Environment and Planning A* **10** 519-535

Muth R F, 1971 "Migration: chicken or egg?" *Southern Economic Journal* **37** 295-305

Olvey L D, 1972 "Regional growth and interregional migration: their pattern of interaction" *Review of Regional Studies* **2** 139-163

Rogers A, 1976a "Migration and settlement" in *IIASA Conference '76 Volume 2* (International Institute for Applied Systems Analysis, Laxenburg, Austria) pp 43-48

Rogers A, 1976b "Demometrics of migration and settlement" RM-76-68, International Institute for Applied Systems Analysis, Laxenburg, Austria

US Bureau of the Census, 1963a *Census of Population 1960: Migration Between State Economic Areas* (US Government Printing Office, Washington, DC)

US Bureau of the Census, 1963b *Census of Population 1960: Volume 1, Part 4, Arizona* (US Government Printing Office, Washington, DC)

US Bureau of the Census, 1972a *Census of Population 1970: Arizona, General Population Characteristics, PC(1), 84* (US Government Printing Office, Washington, DC)
US Bureau of the Census, 1972b *Census of Population 1970: Migration Between State Economic Areas, PC(2)E* (US Government Printing Office, Washington, DC)
US Bureau of the Census, 1976 "Special census for Pima County, Arizona (October 20, 1975)" US Bureau of the Census, Washington, DC (unpublished)

APPENDIX A

An overview of Tucson's development (1950-1975)

Population

The population of Pima County has experienced rapid growth over the last twenty-five years, increasing from a total of 144000 in 1950 to 441200 in 1975 (see table A1). Its growth was rapid during the 1950s but diminished significantly during the 1960s (a stabilization of the population was registered between 1963 and 1966). A sudden acceleration occurred in the early seventies and was maintained well into the economic recession that followed the 1973 oil crisis.

A large part of this population expansion results from net in-migration which accounts for roughly two-thirds of the population change over the last twenty-five years (see table A1 and figure A1). Indeed, because of its climate, Tucson attracts many retirees and workers in the older age groups. This has contributed to a continuous aging of its population. Nevertheless in 1975 the percentage of local population aged 65 and over was still slightly less than its national counterpart (see table A2). The reason for this perhaps surprising fact lies in that the continuous flow of retirement to Arizona and its metropolitan areas has been accompanied by a large flow of employment-related net in-migration, especially during the early 1970s (see table A3).

Table A1. Components of population change in the Tucson SMSA, 1950-1975, by quinquennial periods (in thousands). [Source: derived by the author from annual population estimates made by the Arizona Department of Economic Security (1976) and from annual unpublished vital-statistics data provided by the Data Analysis Section, Health Records and Statistics Division, Arizona State Department of Health, Phoenix, Arizona.]

Period	IP	Births		Deaths		Natural increase		Net migration		FP
		n	r	n	r	n	r	n	r	
1950-1955	144·0	25·2	17·5	7·0	4·9	18·2	12·6	42·8	29·7	205·0
1955-1960	205·0	32·1	15·7	9·2	4·5	22·9	11·2	41·1	20·1	269·0
1960-1965	269·0	34·3	12·7	11·7	4·3	22·6	8·4	22·1	8·2	313·7
1965-1970	313·7	28·7	9·3	13·8	4·5	14·9	4·8	26·4	8·3	355·0
1970-1975	355·0	35·0	9·9	16·5	4·7	18·5	5·2	67·7	19·1	441·2

Key: IP is population at beginning of period; FP is population at end of period; n is number; r is rate (%).

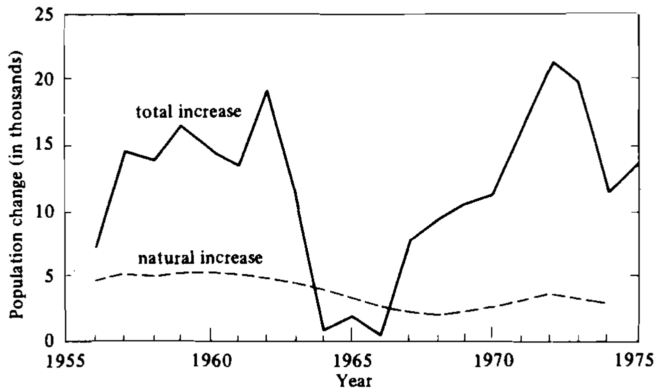


Figure A1. Components of population change for the Tucson SMSA, 1956-1975. [Source: derived by the author from unpublished vital statistics for Tucson, supplied by the Data Analysis Section, Health Records and Statistics Division, Arizona State Department of Health, Phoenix, Arizona, and from population estimates from the Arizona Department of Economic Security (1976).]

Table A2. Age and sex composition of the population of the Tucson SMSA and of the United States, 1960, 1970, and 1975. [Source: computed by the author from data from the US Bureau of Census (1963b; 1972a; 1976).]

Age	Tucson		United States		Tucson		United States	
	male	female	male	female	male	female	male	female
	1960				*1970*			
0-14	16·71	16·14	15·84	15·27	14·32	13·94	14·54	14·02
15-24	7·57	7·36	6·69	6·75	9·60	9·29	8·59	8·75
25-44	12·86	13·36	12·78	13·37	11·13	11·63	11·53	12·08
45-64	8·91	9·68	9·88	10·38	9·24	10·70	9·85	10·75
65+	3·55	3·86	4·07	4·97	4·52	5·63	4·15	5·74
Total	49·60	50·40	49·26	50·74	48·81	51·19	48·66	51·34
	1975							
0-14	12·42	11·81	12·85	12·24				
15-24	10·68	10·48	9·43	9·33				
25-44	13·05	12·75	12·33	12·83				
45-64	8·83	9·89	9·77	10·66				
65+	4·47	5·62	4·30	6·26				
Total	49·45	50·55	48·68	51·32				

Table A3. Net migration figures for selected age cohorts over three recent five-year periods for the Tucson SMSA, 1955-1960, 1965-1970, and 1970-1975. [Source: derived by the author from data from the US Bureau of the Census (1963a; 1972b; 1976).]

Period	Cohort							
	25-29		30-34		35-39		40-44	
	male	female	male	female	male	female	male	female
1955-1960	−468	617	1178	1886	1315	1304	940	1060
1965-1970	−761	15	101	391	631	312	205	494
1970-1975[a]	4625	4141	2926	2758	2274	2067	2164	1908

[a] The 1970-1975 figures represent net population gains due to migration as well as to natural increase.

Table A4. Labor force and employment figures for the Tucson SMSA, 1956-1975 (from unpublished data from the Arizona Department of Economic Security). All figures are in thousands unless otherwise indicated.

Year	Labor force by place of		Unemployment rate (%) by place of		Employment total by place of		Agricultural employment by place of		Employment[a]									Other employment[b]
	work	residence	work	residence	work	residence	work	residence	total	1	2	3	4	5	6	7	8	
1956	71·0	-	3·7	-	68·4	-	3·1	-	54·4	9·7	2·1	4·7	5·3	12·2	1·8	8·2	10·4	10·9
1957	76·0	-	4·1	-	72·9	-	2·5	-	59·1	9·9	2·4	5·4	5·3	13·2	2·1	9·1	11·7	11·3
1958	79·4	-	6·0	-	74·5	-	2·6	-	60·5	8·9	2·3	5·6	5·1	13·6	2·4	10·0	12·6	11·4
1959	84·2	-	4·8	-	80·1	-	2·3	-	66·0	9·2	2·6	6·9	5·2	14·8	2·6	11·0	13·7	11·8
1960	87·9	-	5·1	-	83·4	-	2·2	-	69·2	8·4	2·9	6·9	5·2	15·8	2·9	12·2	14·9	12·0
1961	92·4	-	5·7	-	87·1	-	2·0	-	72·8	8·4	3·1	7·8	5·1	16·3	3·1	12·9	16·1	12·3
1962	99·1	-	4·9	-	94·2	-	1·9	-	79·5	9·2	3·4	9·8	5·3	17·3	3·4	13·6	17·5	12·8
1963	98·6	-	5·8	-	92·8	-	1·8	-	78·3	9·3	3·2	6·5	5·4	17·4	3·7	13·9	18·9	12·7
1964	96·3	97·0	6·5	6·7	90·0	90·5	1·8	1·5	75·7	6·6	3·3	5·9	5·3	17·1	3·8	13·9	19·8	12·5
1965	95·3	97·3	6·3	6·5	89·0	91·0	1·7	1·3	76·0	6·3	3·4	5·5	5·2	17·3	3·6	13·7	21·0	11·3
1966	98·7	100·7	4·2	4·5	94·5	96·2	1·7	1·3	81·7	7·7	4·0	5·6	5·1	18·1	3·5	14·6	23·1	11·1
1967	102·9	104·4	4·0	4·4	98·5	99·8	1·8	1·4	86·0	8·8	4·3	5·8	5·3	18·8	3·5	15·7	23·8	10·7
1968	106·3	107·9	4·0	4·4	101·8	103·1	1·7	1·3	89·7	7·8	4·7	7·0	5·3	19·6	3·7	16·7	24·9	10·4
1969	113·8	114·6	3·3	3·5	110·1	110·6	1·8	1·3	98·0	8·3	5·5	9·4	5·5	21·5	4·2	17·8	25·8	10·3
1970	120·9	121·9	3·4	3·5	116·7	117·8	1·8	1·9	104·8	9·0	6·5	8·5	5·9	23·7	5·0	19·3	26·9	10·1
1971	130·5	132·8	4·0	4·1	125·0	127·4	1·6	1·8	113·3	8·9	6·7	10·4	6·1	25·3	5·1	20·8	30·0	10·1
1972	143·4	148·1	3·8	4·0	138·0	142·2	1·6	1·8	127·3	10·4	7·0	12·3	6·9	28·4	5·9	23·0	33·4	13·1
1973	-	158·7	-	3·4	-	153·3	-	1·8	138·0	12·1	7·9	13·3	7·6	30·2	6·5	25·4	35·0	13·5
1974	-	164·4	-	4·6	-	157·0	-	1·6	142·0	12·2	8·5	11·6	8·0	30·1	6·5	26·7	38·4	13·4
1975	-	168·9	-	8·0	-	155·4	-	1·7	140·7	11·8	7·9	10·1	7·5	30·4	6·1	26·6	40·3	13·0

[a] Includes all full- and part-time wage and salary earners who worked or received pay during the pay period ending nearest the 15th of each month.
[b] Includes self-employed, unpaid family workers, and domestics (figures before and after 1972 are not comparable).

Key: 1 manufacturing; 2 mining; 3 construction; 4 transportation and communication; 5 trade; 6 finance and real estate; 7 services; 8 government.

Employment

From an economic point of view, Tucson's rapid population growth has been paralleled by a rapid growth of local activities, thus tilting its employment structure away from the primary sector (see table A4).

Manufacturing was a relatively small proportion (8%) of total employment in 1974, a proportion which aside from Las Vegas is by far the smallest over the SMSAs in the 100000–200000 employment range. Mining (copper extraction) on the other hand, although consisting of operations accounting for a large part of the national activity in this sector, employs a relatively small part of Tucson's workforce (up to 8500 workers in 1974).

The construction sector, employing up to 10% of the work force, is comparatively large in Tucson as a consequence of continuous population growth, but most of the Tucson work force is engaged in local activities such as services (17·1%), trade (19·5%), and government (25·1%), all percentages relating to 1975.

Labor force and unemployment

Over the last quarter of a century, the growth of the labor force has more or less followed the growth of the population. In the later years, however, it has shown a tendency to surpass the growth of the population, thus leading to a significant increase of the labor-force-participation rate. Nevertheless, this rate remains lower than its national counterpart, possibly because the Tucson population includes a large proportion of students, military personnel, and people with Spanish background.

It is also important to note that the local unemployment rate has generally been lower than the national rate except in the middle 1960s when, in consequence of important layoffs in the manufacturing sector, economic development and population growth were slowed down.

APPENDIX B

Table B1. Two-stage least-squares parameter estimates for the demometric model. The meaning of each included variable is explained in table 6. Note that the various statistics attached to each equation are here reported with all the digits provided by the computer program used to perform these 2SLS regressions. Some of these statistics may not be significant.

$$\text{MANUF} = \underset{(4\cdot8166)}{(0\cdot15045\times10^{-3})}\text{NEMP} + \underset{(5\cdot1919)}{0\cdot11248}\frac{\text{EMPT}}{1-\text{UNR}(-1)} - \underset{(-2\cdot9755)}{0\cdot03914}\text{POP} - \underset{(-6\cdot0770)}{2\cdot41385}\text{DUM64}\rightarrow$$

period = 1957–1975; mean = 9·1158; $\rho = 0\cdot173$;
$R^2 = 0\cdot968$; $\bar{R}^2 = 0\cdot962$; SE = 0·54353; DW = 1·902; $F(4, 14) = 143\cdot056$

$$\text{MINING} = \underset{(-4\cdot4267)}{-10\cdot98271} + \underset{(3\cdot4026)}{(0\cdot17864\times10^{-3})}\text{NEMP} + \underset{(1\cdot7340)}{0\cdot02344}\frac{\text{EMPT}}{1-\text{UNR}(-1)}$$

period = 1957–1975; mean = 4·7158; $\rho = 0\cdot545$;
$R^2 = 0\cdot987$; $\bar{R}^2 = 0\cdot985$; SE = 0·29260; DW = 1·591; $F(2, 15) = 575\cdot496$

$$\text{CONST} = \underset{(3\cdot8962)}{0\cdot00797}\text{POP}(-2) + \underset{(2\cdot5768)}{0\cdot09080}\text{DPOP}(-1) + \underset{(3\cdot4183)}{0\cdot12337}\text{DPOP} + \underset{(4\cdot9202)}{0\cdot63949}\text{MINING} + \underset{(2\cdot4439)}{1\cdot37067}\text{DUM61}\rightarrow 62$$

period = 1957–1975; mean = 8·1211; $\rho = -0\cdot223$;
$R^2 = 0\cdot961$; $\bar{R}^2 = 0\cdot950$; SE = 0·73115; DW = 1·992; $F(5, 13) = 81\cdot080$

$$\text{TRANSP} = \underset{(5\cdot7789)}{2\cdot33883} + \underset{(3\cdot4166)}{0\cdot16548}(\text{MANUF}+\text{MINING}) + \underset{(2\cdot0759)}{0\cdot01519}(\text{WSEMP}-\text{MANUF}-\text{MINING})$$

period = 1957–1975; mean = 5·8000; $\rho = 0\cdot680$;
$R^2 = 0\cdot990$; $\bar{R}^2 = 0\cdot988$; SE = 0·18550; DW = 1·372; $F(2, 15) = 713\cdot477$

Table B1 (continued)

TRADE $= -13{\cdot}17835 + 0{\cdot}05729\text{POP}(-1) + 0{\cdot}10210\text{DPOP} + 0{\cdot}50485\text{RPCI}$
(−6·3109) (5·7156) (2·8413) (3·2723)
period = 1957–1974; mean = 19·917; $\rho = 0{\cdot}653$;
$R^2 = 0{\cdot}996$; $\bar{R}^2 = 0{\cdot}995$; SE = 0·54047; DW = 1·686; $F(3,13) = 1020{\cdot}083$

FIRE $= -2{\cdot}47205 + 0{\cdot}02098\text{POP} - (0{\cdot}06993 \times 10^2)\text{UNR}$
(−3·9859) (11·835) (−1·2626)
period = 1957–1975; mean = 4·0842; $\rho = 0{\cdot}580$;
$R^2 = 0{\cdot}982$; $\bar{R}^2 = 0{\cdot}980$; SE = 0·24607; DW = 1·541; $F(2,15) = 411{\cdot}226$

SERV $= -14{\cdot}97432 + 0{\cdot}06905\text{POP}(-1) + 0{\cdot}05475\text{DPOP} + 0{\cdot}33614\text{RPCI}$
(−6·2809) (8·3017) (2·1234) (2·6245)
period = 1957–1975; mean = 16·128; $\rho = 0{\cdot}870$;
$R^2 = 0{\cdot}997$; $\bar{R}^2 = 0{\cdot}996$; SE = 0·41218; DW = 1·605; $F(3,13) = 1376{\cdot}767$

GOVT $= 0{\cdot}04421\text{POP} + (0{\cdot}23111 \times 10^{-2})(\text{POP} \times \text{TIME})$
(32·147) (23·958)
period = 1957–1975; mean = 22·910; $\rho = 0{\cdot}421$;
$R^2 = 0{\cdot}998$; $\bar{R}^2 = 0{\cdot}998$; SE = 0·49978; DW = 2·247; $F(2,17) = 8171{\cdot}061$

SELF $= 10{\cdot}94564 + 2{\cdot}88914\text{DUM72}\rightarrow$
(16·102) (7·0422)
period = 1957–1975; mean = 11·779; $\rho = 0{\cdot}899$;
$R^2 = 0{\cdot}984$; $\bar{R}^2 = 0{\cdot}983$; SE = 0·40706; DW = 1·076; $F(1,16) = 963{\cdot}209$

DEATH $= 0{\cdot}60814 + (0{\cdot}45719 \times 10^{-2})\dfrac{\text{POP}(-1) + \text{POP}}{2} + 0{\cdot}04806\text{TIME}$
(2·0935)
period = 1957–1975; mean = 2·6090; $\rho = 0{\cdot}033$;
$R^2 = 0{\cdot}995$; $\bar{R}^2 = 0{\cdot}994$; SE = $0{\cdot}66667 \times 10^{-1}$; DW = 1·886; $F(2,15) = 1375{\cdot}479$

BIRTH $= 0{\cdot}03044\dfrac{\text{POP}(-1) + \text{POP}}{2} - 0{\cdot}29844\text{TIME}$
(18·798) (−7·1747)
period = 1957–1975; mean = 6·5128; $\rho = 0{\cdot}735$;
$R^2 = 0{\cdot}979$; $\bar{R}^2 = 0{\cdot}977$; SE = 0·24904; DW = 1·083; $F(2,17) = 780{\cdot}119$

NETMIG $= 32{\cdot}88868 + 1{\cdot}57388\text{DCONST} + 2{\cdot}21562\text{DMANUF} - (4{\cdot}56254 \times 10^2)\text{UNR}(-1)$
(5·5301) (5·6848) (4·2827) (−15·993)
$+ (3{\cdot}51779 \times 10^2)\text{NUNR}(-1) - 32{\cdot}37792\text{H}(-1)$
(10·822) (−4·2461)
period = 1957–1975; mean = 8·1207; $\rho = -0{\cdot}854$;
$R^2 = 0{\cdot}957$; $\bar{R}^2 = 0{\cdot}939$; SE = 1·6038; DW = 1·835; $F(5,12) = 53{\cdot}615$

UNR $= 0{\cdot}49260\text{UNR}(-1) + 0{\cdot}59125\text{NUNR} - 0{\cdot}12724\dfrac{\text{DEMPT}}{\text{EMPT}(-1)}$
(6·9300) (9·4148) (−5·2515)
period = 1957–1975; mean = 4·8053; $\rho = -0{\cdot}077$
$R^2 = 0{\cdot}934$; $\bar{R}^2 = 0{\cdot}926$; SE = 0·47382; DW = 1·996; $F(3,15) = 106{\cdot}708$

RPCI $= 5{\cdot}83236 + 0{\cdot}86341\text{RPCI}(-1) + 12{\cdot}64331\dfrac{\text{DEMPT}}{\text{EMPT}(-1)} - 152{\cdot}17232\dfrac{\text{NATINC}}{\text{POP}(-1)}$
(3·8830) (18·864) (3·4408) (−4·9111)
period = 1957–1975; mean = 28·266; $\rho = -0{\cdot}528$;
$R^2 = 0{\cdot}996$; $\bar{R}^2 = 0{\cdot}995$; SE = 0·57081; DW = 2·313; $F(3,13) = 1109{\cdot}180$

Environment and Planning A, 1978, volume 10, pages 561-576

A dynamic linear-programming approach to the planning of national settlement systems

A Propoi¶, F Willekens
International Institute for Applied Systems Analysis, Laxenburg, Austria
Received 31 May 1977, in revised form 21 February 1978

Abstract. In this paper the problem of planning human-settlement systems (HSS) is formulated in a dynamic linear-programming (DLP) framework. Such large time-dependent problems are treated both by optimal-control and linear-programming techniques. A multiregional population-growth model forms the state equation of the DLP problem. Budget, resource, and quality-of-life constraints are considered. The paper describes the formalization of the HSS planning problem and suggests methods for its solution.

1 Introduction

The development of human-settlement systems is becoming a public concern in most countries. Nations all over the world are adopting policies to guide the growth and the distribution of their populations (for some details, see Willekens, 1976a). This trend toward explicit national settlement policies is enhanced by the realization that land and environment are not free goods; they are scarce resources to be conserved. Settlement planning must elaborate control policies of population distribution over space and/or time in order to achieve desirable socioeconomic goals (conservation of the environment, economic efficiency, etc), and must take into account a large number of factors and constraints such as total population; age and sex structure; birth, death, and migration rates; scarceness of resources; and educational constraints. An effective way of developing optimal decisions, when a very large number of variables and constraints are involved, is by applying mathematical programming. Of the available optimization methods, linear programming is most successful in dealing with large static problems. Dynamic decision problems, on the other hand, have been treated by using optimal-control theory. National settlement systems are large in scale and dynamic in nature, and can be studied as dynamic optimization problems, more particularly as dynamic linear-programming (DLP) problems, which comprise both linear-programming (LP) and control-theory methods (Propoi, 1976).

The purpose of this paper is to discuss the applicability of the DLP approach to national settlement-system planning. It is the third IIASA paper on the application of specific mathematical-programming methods to settlement-system planning (Mehra, 1975; Evtushenko and MacKinnon, 1976), and consists of two parts—the first part describes DLP models of national settlement-system planning, and the second part is devoted to the application of DLP theory and methods in the solution of these models.

2 The planning problem

The purpose of this section is to describe in some detail the problem of national settlement-system planning. The models envisaged are in the format of a DLP problem. A DLP problem consists of three components: the state equations, the constraints imposed on the system variables, and the performance index (objective function). In DLP, all these equations and inequalities are linear in the state and the control variables.

¶ On leave from the Institute of Systems Studies, USSR Academy of Sciences, Moscow, USSR.

2.1 *The state equations*

The state equations describe the development of the multiregional population system over time. This development is determined by internal systems dynamics and by policy interventions. The internal dynamics are represented by the system's 'laws of motion'. External intervention will change the motion of the system, but the degree and the direction of the disturbance depend on the dynamic characteristics of the system.

To avoid counteractive and undesired effects of a settlement policy, we need to understand the internal dynamics governing a multiregional population system, that is, the behavior of the system over time before applying control to it. The mechanisms of demographic growth and intervention have been considered by Rogers (1968; 1971; 1975). They are explored further in this paper.

To transform the growth model into a policy model we add a sequence of vectors describing control actions distributed over time and space. A control vector defines the instruments of population-distribution policy. Population distribution does not occur in a vacuum; it is subordinate to social and economic policies. Frequently the goals of population redistribution are environmental and economic in nature. To achieve these nondemographic goals, use is made of nondemographic, economic, and legal instruments. Although the focus is on population and its distribution, policy implementation requires the consideration of socioeconomic factors. Demometrics is the study of this interdependence between spatial population growth and the socioeconomic system. This paper shows how demometrics might contribute to the formulation of models of national settlement-system planning by describing not only the internal system dynamics but also the influence of external intervention on the system.

Formally, the state equations appear as linear heterogenous equations. The homogenous part of the equation system describes the behavior of the system undisturbed by outside influences. (This behavior is defined by a multiregional demographic-growth model.) The heterogenous part describes the impact of factors exogenous to the demographic system, such as policy intervention.

The dynamics of multiregional population systems (the homogenous part of the equation) are governed by the interaction of fertility, mortality, and migration. In recent years demographers, geographers, economists, and planners have devoted their attention to modeling these dynamics in order to describe and explain the changes taking place in actual human-settlement systems. The models that have been developed have a similar underlying structure. In most instances they appear as a system of linear difference equations, or they may be transformed into such a system. The general format of the models is the matrix equation

$$\boldsymbol{x}(t+1) = \mathbf{G}(t)\boldsymbol{x}(t)\ , \qquad (1)$$

where $\boldsymbol{x}(t)$ is the vector[(1)] of population distribution at time t, and $\mathbf{G}(t)$ is the population growth matrix at time t, which in most cases is assumed to be constant over time: $\mathbf{G}(t) = \mathbf{G}$. This model does not consider exogenous contributions to population growth—these will be added later.

Depending on the aggregation level, $\boldsymbol{x}(t)$ is the population by region, or the population by age and region. Matrix models of aggregate multiregional population change are, for example, the Markov-chain model, the input–output model, and the components-of-change model. The model of disaggregate multiregional population change is known as the multiregional cohort-survival model (Rogers, 1975, chapter 5; see also Rees and Wilson, 1975).

(1) Note that all vectors are written as *column* vectors.

In the components-of-change model and the cohort-survival model, population at time t and its regional and/or age distribution depends only on the population distribution in the previous period. They are purely demographic models since they do not include other socioeconomic variables. In this closed system, the predetermined variables consist of lagged endogenous variables. The growth path of the system is completely determined by the growth matrix $\mathbf{G}$ and the initial conditions.

Both the components-of-change model and the multiregional cohort-survival model take the form of a system of homogenous first-order difference equations. They describe the dynamics of a *closed* multiregional system. The transformation of these models to open systems is straightforward. We add to system (1) a vector $e(t)$:

$$x(t+1) = \mathbf{G}(t)x(t)+e(t) , \tag{2}$$

which then describes the exogenous contributions to population growth, such as external migration.

To make the models more realistic, we extend the set of predetermined variables to include economic variables such as income, employment, housing stock, accessibility, and several types of government expenditures. Some of the predetermined variables are *controllable* by the policymaker, and are labeled *policy variables, control variables,* or *instrument variables*. Others are uncontrollable but are exogenously given. This classification of controllable and uncontrollable variables follows Tinbergen (1963). In the case of human-settlement policy, governments usually do not intervene directly in migration flows. Government control is mostly limited to socioeconomic instruments affecting migration. The modeling of such indirect population-distribution policies is discussed in Stern (1974), Wilson (1974), Cordey-Hayes (1975), MacKinnon (1975), and Willekens and Rogers (1977), among others.

The complete policy model may be expressed as[(2)]

$$x(t+1) = \mathbf{G}(t)x(t)+\mathbf{D}(t)v(t)+\mathbf{L}(t)u(t)+e(t) , \tag{3}$$

where $v(t)$ is the vector of controllable variables, $u(t)$ is the vector of uncontrollable predetermined socioeconomic variables, and $\mathbf{D}(t)$ and $\mathbf{L}(t)$ are matrix multipliers. For simplicity, and without loss of generality, we delete the uncontrollable predetermined variables. Model (3) then reduces to

$$x(t+1) = \mathbf{G}(t)x(t)+\mathbf{D}(t)v(t) . \tag{4}$$

The control vector $v(t)$ consists of socioeconomic instrument variables affecting the distribution of the population. The matrix multiplier $\mathbf{D}(t)$ is important in this setting. An element $d_{ij}(t)$ denotes the impact on the population in region–age combination i of a unit change in the jth instrument at step t. In many cases the elements of this matrix are also assumed constant over time: $\mathbf{D}(t) = \mathbf{D}$. This implies that the effects of certain policies on the population distribution are independent of the time period when the policies are implemented. This is consistent with the Markovian assumption of time homogeneity. The linearity of equation (4) implies that the effects of the various policies are additive.

Equation (4) is the state equation of a state-space model. How it may be derived from linear demometric models describing the interdependence between demographic and socioeconomic variables is described in Willekens (1976b). The rationale for using state-space model (4) as the analytical or numerical tool for population-policy analysis, is that the homogenous part of equation (4) is exactly the demographic-growth model (components-of-change or cohort-survival) that describes the population

(2) The fact that equation (3) is a first-order difference equation is by no means restrictive. Higher-order difference equations may be converted into a system of first-order difference equations (Zadeh and Desoer, 1963).

growth without intervention. The logical extension of population-growth models to policy models is therefore the addition of a heterogenous part to the growth model (see also Rogers, 1966; 1968, chapter 6; 1971, pages 98-108). The resulting model is a heterogenous system of linear first-order difference equations.

2.2 *The constraints*
Policymaking is subject to constraints. Relocating people or intervening in their residential-location decisions incurs a cost which may be political, economic, and/or social. The values that the control and state vectors $v(t)$ and $x(t)$ in equation (4) can take on are therefore restricted by political, economic, and social considerations. For example, let $v(t)$ denote the number of in-migrants, from outside the system, that have to move in in order to achieve certain population-distribution objectives. It is politically and socially unacceptable to relocate a very large number of people during a short time period. Therefore there is an upper bound to the number of in-migrants during a unit time period (Evtushenko and MacKinnon, 1976, page 639):

$$\sum_i v_i(t) \leqslant v(t) , \qquad t = 0, 1, ..., T-1 , \tag{5}$$

where the scalar $v(t)$ is the total in-migration pool available in the tth time period, and the $v_i(t)$ are the components of $\boldsymbol{v}(t)$.

Instead of restricting the control vector by defining a total in-migration pool, each component of $v(t)$ may be required to lie within a lower and an upper bound:

$$\underline{v}_i(t) \leqslant v_i(t) \leqslant \bar{v}_i(t) , \qquad i = 1, 2, ..., s . \tag{6}$$

Population-redistribution policy is not free. The imposition of controls implies the incurrence of costs. It is therefore natural to assume a budget constraint limiting the actions of the policymaker. We distinguish between a budget constraint for each period,

$$[c(t)]^{\mathrm{T}} v(t) \leqslant C(t) , \qquad t = 0, 1, ..., T-1 , \tag{7}$$

and a global budget constraint,

$$\sum_{t=0}^{T-1} [c(t)]^{\mathrm{T}} v(t) \leqslant C , \tag{8}$$

where a component $c_i(t)$ of the cost vector $c(t)$ denotes the cost of transferring a person to region i in the tth time period, where $C(t)$ is the total budget available during period t, and where C is the global budget.

Frequently the population distribution itself is constrained in addition to the control vector. For example, in a pure redistribution policy, the total population X of the system is held constant:

$$\sum_{j=1}^{n} x_j(t) = X = \sum_{j=1}^{n} x_j(0) , \qquad t = 1, 2, ..., T . \tag{9}$$

As in the case of the control vector, the policymaker may want to put lower and upper bounds on the population in each region. This would avoid the excessive growth of some regions and the depopulation of others:

$$\underline{x}_j(t) \leqslant x_j(t) \leqslant \bar{x}_j(t) , \qquad t = 1, 2, ..., T;\ j = 1, 2, ..., n . \tag{10}$$

A constraint receiving considerable attention in recent years is the resource constraint. Scarce resources include not only capital, but also raw materials, water, and the environment. As mentioned in the introduction to this paper, human-settlement policies in most countries are directed toward the conservation of those resources. This commitment must be reflected in the planning model. Therefore we introduce the resource constraint[3]:

$$\mathbf{P}(t)x(t)+\mathbf{S}(t)v(t) \leqslant r(t), \qquad t = 0, 1, ..., T, \tag{11}$$

where $r(t)$ is the vector of available resources in the tth time period; the matrices $\mathbf{P}(t)$ and $\mathbf{S}(t)$ are rectangular matrices: an entry $p_{kj}(t)$, for example, denotes the amount of resource k required by an individual in region j during time period t; an entry $s_{ki}(t)$ denotes the use of resource k per unit of control i during period t. Note that constraint (7) is a special case of constraint (11) in which a single resource, capital, is considered associated with the control.

Another constraint relates to the quality-of-life or income levels. Let $q(t)$ be the vector denoting the regional distribution of required quality-of-life levels. The quality-of-life constraint is then

$$\mathbf{M}(t)x(t)+\mathbf{N}(t)v(t) \geqslant q(t), \qquad t = 0, 1, ..., T, \tag{12}$$

where an entry $m_{lj}(t)$ of $\mathbf{M}(t)$ denotes the per-capita level of the quality-of-life index l in region j at time t, and an entry $n_{li}(t)$ of $\mathbf{N}(t)$ represents the impact of policy variable i on the level of the quality-of-life index l.

A final group of restrictions on the action span of the policymaker is represented by the boundary conditions. Since the planning of settlement systems starts from the current population distribution, we have the initial condition

$$x(0) = x^{\text{init}}. \tag{13}$$

On the other hand, the population distribution at the planning horizon, $x(T)$, may be fixed,

$$x(T) = x^{\text{hor}}, \tag{14}$$

or kept free.

2.3 *The objective function*

The objective function provides the explicit expression of the policymaker's preference system. The ultimate goal of national settlement-system planning is to increase the quality of life. There is no agreement on the factors determining the quality of life, and even less on its quantitative measurement. Therefore the usual difficulties in selecting the objective function arise. However, it should be noted that the optimization procedure cannot be considered as a final stage in the planning process, but only as a tool for analysing the connection between policy alternatives and system performance. Thus the analysis of 'optimal' solutions with different objective functions is needed. So, for example, the quality-of-life goal is replaced by a single objective involving monetary costs and benefits only. Such an objective function is given in equation (15). It is necessary to maximize the total benefit $B(v)$,

$$B(v) = \sum_{t=0}^{T} [\alpha(t)]^{\mathrm{T}}x(t)+[\beta(t)]^{\mathrm{T}}v(t), \tag{15}$$

where $\alpha(t)$ is the vector of unit benefit associated with the regional population levels at step t, and $\beta(t)$ is the vector of unit benefit associated with the controls. [If necessary, discounting is included in $\alpha(t)$ and $\beta(t)$.]

An objective function involving costs is shown in equation (16). The problem is to minimize

$$K(v) = \sum_{t=0}^{T} [\gamma(t)]^{\mathrm{T}}x(t)+[\delta(t)]^{\mathrm{T}}v(t), \tag{16}$$

(3) Note that for two vectors y and z, $y \leqslant z$ means that $y_i \leqslant z_i$ for each pair of components y_i and z_i.

where $\gamma(t)$ is the vector of unit costs associated with the regional population levels at step t, and $\delta(t)$ is the vector of unit costs associated with the controls.

Another type of objective function is connected with optimization of the terminal state, and can be formulated in the following way:

$$B = \sum_{j=1}^{n} w_j(T)x_j(T) = [w(T)]^{\mathrm{T}}x(T) , \tag{17}$$

where $w_j(T)$ is the weighting coefficient of a population group $x_j(T)$.

In some instances, the policymaker may not want to minimize the costs associated with the settlement system and with the intervention in this system. Instead he may just want to bring the population distribution as close as possible to a desired target distribution $\bar{x}(T)$ at the planning horizon. This problem has been treated by Willekens (1976b, pages 66-85) for the two cases where explicit analytical solutions can be derived: the problems of initial-period control and linear-feedback control. In DLP problems the performance index may be formulated as

$$\text{minimize}\left\{B(v) = \sum_j w_j(T)|x_j(T)-\bar{x}_j(T)|\right\} , \tag{18}$$

where | | denotes absolute value and $w_j(T)$ is the weighting coefficient.

An interesting formulation of the objective at the end of the planning horizon is as follows:

$$B(v) = \min_{1 \leqslant j \leqslant n} \left\{\frac{x_j(T)}{k_j}\right\} , \tag{19}$$

where the numbers k_j define the desired proportions of the terminal distribution.

Models with both types of nonlinear objective functions (18) and (19) can be easily reduced to linear-programming problems (Charnes and Cooper, 1961; Kantorovitch, 1965).

3 Dynamic linear-programming theory and methods

The purpose of this section is to describe DLP theory and methods in relation to models of national settlement-system planning. In the DLP framework, the dynamics of settlement systems are described by linear state equations; the feasible set of policies is delineated by linear constraints; and the preference system of the policymaker is described by a linear function. Thus we have a special type of linear-programming problem.

The impact of linear-programming models and methods on the practice of decisionmaking is well known (Dantzig, 1963; Kantorovitch, 1965). However, both LP theory itself and the basic range of its applications are of a one-stage, static nature. When the system to be optimized is changing over time, and its development is to be planned, a static approach is inadequate, and the problem of optimization becomes a dynamic multistage problem.

The principal feature of models of settlement planning is their dynamic character. On the other hand, the basic relations and conditions in such models often are linear. Hence DLP may be an efficient tool for elaborating optimal policies in large-scale, national settlement-system planning. For dynamic LP, new problems arise. For example, the basic question of the static LP problems is how to determine the optimal decision. In dynamic problems, however, knowledge of the sequence of optimal decisions is not sufficient. Of equal importance is the realization of these decisions (questions of feedback control, stability, and sensitivity analysis of optimal systems).

Our exposition of DLP theory and methods is divided into three sections. First, it is shown how demographic DLP models can be reduced to a canonical form, thus enabling the development of a unified approach for a whole range of problems of national settlement planning which arise in practice. Second, there is an analysis of duality relations. Third, DLP computational methods are described and evaluated. The exposition is kept general and introductory. For more details on the several aspects of DLP theory and methods, the reader is referred to Propoi (1976; 1977).

3.1 *The* DLP *canonical form*

The models of multiregional population policy described earlier may be reduced to a canonical form. Before doing so, however, it is useful to recall the components of the DLP problem:

the *state (development) equations* of the systems, with the distinct separation of state and control variables;

the *constraints* imposed on these variables;

the *planning period* (the number of stages during which the system is considered); and

the *performance index (objective function)* which quantifies the quality of a control.

As we have noted earlier, state equations have the following form:

$$x(t+1) = \mathbf{G}(t)x(t)+\mathbf{D}(t)v(t)+e(t)\,, \tag{20}$$

where the vector $x(t) = [x_1(t), ..., x_n(t)]$ defines the *state* of the system at stage t in the state space, the vector $u(t) = [u_1(t), ..., u_s(t)]$ defines the exogenous uncontrolled variables (known a priori in the deterministic models), $\mathbf{G}(t)$ is the $n \times n$ state-transformation matrix (in the majority of demographic problems $\mathbf{G}(t) = \mathbf{G}$ is the growth matrix), and $\mathbf{D}(t)$ is the $n \times s$ control-transformation matrix which defines the influence of a control on the state of the system.

Constraints imposed on the state and control variables may be written in general form as

$$\mathbf{P}(t)x(t)+\mathbf{S}(t)v(t) \leqslant r(t) \qquad \text{and} \qquad v(t) \geqslant 0\,, \tag{21}$$

where $r(t) = [r_1(t), ..., r_m(t)]$ is a given vector, and $\mathbf{P}(t)$ and $\mathbf{S}(t)$ are known $m \times n$ and $m \times s$ matrices.

The planning period T is assumed to be fixed. It is also assumed that the initial state of the system is given:

$$x(0) = x^{\text{init}}\,. \tag{22}$$

The objective function has the following general form:

$$B_1(v) = [\alpha(T)]^{\mathrm{T}}x(T)+\sum_{t=0}^{T-1}\{[\alpha(t)]^{\mathrm{T}}x(t)+[\beta(t)]^{\mathrm{T}}v(t)\}\,, \tag{23}$$

where $\alpha(t)$ $(t = 0, 1, ..., T)$ and $\beta(t)$ $(t = 0, 1, ..., T-1)$ are given weight coefficients.

The following definitions are used:

the vector sequence $v = \{v(0), ..., v(T-1)\}$ is a *control (policy)* of the system;

the vector sequence $x = \{x(0), ..., x(T)\}$, which corresponds to control v from state equations (20) with the initial state $x(0)$, is the *system's trajectory*;

the process $\{v, x\}$ is said to be *feasible* if it satisfies all the constraints of the problem [constraints (20)-(22) in this case];

the feasible process $\{v^*, x^*\}$, maximizing the performance index (23), is known as the *optimal* process.

Hence the canonical form of the dynamic policy problem is formulated as follows.

Problem 1 (primal): Given the initial population distribution,

$$x(0) = x^{\text{init}}\,; \tag{24}$$

the state equations,

$$x(t+1) = \mathbf{G}(t)\bar{x}(t) + \mathbf{D}(t)v(t) + e(t) , \tag{25}$$

where

$x(t)$ is the population distribution at time t (state of the system),
$v(t)$ is the vector of controls;
$\mathbf{G}(t)$ is the population-growth matrix (usually constant over time),
$\mathbf{D}(t)$ is the matrix expressing the impact of the control $v(t)$ on the population distribution $x(t)$, and
$e(t)$ describes the exogenous contributions to population growth;

and the constraints,

$$\mathbf{P}(t)x(t) + \mathbf{S}(t)v(t) \leqslant r(t) \qquad \text{and} \qquad v(t) \geqslant 0 , \tag{26}$$

where

$r(t)$ is the vector of available resources at time t,
$\mathbf{P}(t)$ is the matrix denoting the amount of resources required per individual at step t, and
$\mathbf{S}(t)$ is the matrix denoting the consumption of resources per unit of control at step t;

find a control (policy) $v = \{v(0), ..., v(T-1)\}$ and a corresponding state trajectory $x = \{x(0), ..., x(T)\}$ that maximizes the performance index

$$B_1(v) = [\alpha(T)]^{\mathrm{T}}x(T) + \sum_{t=0}^{T-1}\{[\alpha(t)]^{\mathrm{T}}x(t) + [\beta(t)]^{\mathrm{T}}v(t)\} , \tag{27}$$

where

$\alpha(t)$ is the vector of unit benefit associated with the regional population distribution $x(t)$, and
$\beta(t)$ is the vector of unit benefit associated with the control $v(t)$.

The choice of a canonical form for problem 1 is to some extent arbitrary, various modifications and particular cases of problem 1 being possible. Some of them have been considered in the first part of this paper, and a classification of these modifications is given in the appendix. It should be noted that these modifications can be transformed one into the other. For example, let us consider the problems 1 and 1a, where problem 1a is the same as problem 1 except that it has performance index (17). Introducing a new variable $x_0(t)$ ($t = 0, ..., T$), subject to

$$x_0(t+1) = x_0(t) + [\alpha(t)]^{\mathrm{T}}x(t) + [\beta(t)]^{\mathrm{T}}v(t) \qquad \text{and} \qquad x_0(0) = 0 , \tag{28}$$

one can see that

$$x_0(T) = \sum_{t=0}^{T-1}\{[\alpha(t)]^{\mathrm{T}}x(t) + [\beta(t)]^{\mathrm{T}}v(t)\} . \tag{29}$$

Thus problem 1 will have the form of problem 1a with the performance index

$$B_1 = [\tilde{\alpha}(T)]^{\mathrm{T}}\tilde{x}(T) \tag{30}$$

and the state equations

$$\tilde{x}(t+1) = \tilde{\mathbf{G}}(t)\tilde{x}(t) + \tilde{\mathbf{D}}(t)v(t) + e(t) , \tag{31}$$

where

$$\tilde{\alpha}(T) = [1, \alpha_1(T), ..., \alpha_n(T)] ;$$

$$\tilde{x}(0) = (0, x^{\mathrm{init}}) \qquad \text{and} \qquad \tilde{x}(t) = [x_0(t), x_1(t), ..., x_n(t)] ;$$

and where

$$\bar{G}(t) = \begin{bmatrix} 1 & \alpha(t) \\ \mathbf{0} & G(t) \end{bmatrix} \quad \text{and} \quad \bar{D}(t) = \begin{bmatrix} 0 & \beta(t) \\ \mathbf{0} & D(t) \end{bmatrix}$$

(where **0** is the zero vector of appropriate size).

Analogously, performance indices (18) and (19) can be reduced to (17). For example, performance index (19) can be replaced by the problem

$$\text{maximize}\{B(v) = \theta\} \tag{32}$$

with additional terminal state constraints

$$x_j(T) \leqslant \theta k_j\,, \qquad j = 1, ..., n\,. \tag{33}$$

If we consider problem 1a, with constraints (8), and introduce a variable $x_{n+1}(t)$, subject to the state equations

$$x_{n+1}(t+1) = x_{n+1}(t) + [c(t)]^{\mathrm{T}} v(t)\,, \quad t = 0, 1, ..., T-1\,, \quad \text{and} \quad x_{n+1}(0) = 0\,, \tag{34}$$

then we obtain problem 1 with equations

$$\hat{x}(t+1) = \hat{G}(t)\hat{x}(t) + \hat{D}(t)v(t) + e(t)\,, \tag{35}$$

where

$$\hat{x}(t) = [x_1(t), ..., x_n(t), x_{n+1}(t)]\,,$$

$$\hat{G}(t) = \begin{bmatrix} G(T) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad \text{and} \quad \hat{D}(t) = \begin{bmatrix} D(t) & \mathbf{0} \\ c(T) & \mathbf{I} \end{bmatrix}$$

(where **0** and **I** are the appropriate zero and identity matrices respectively), and only one terminal condition is imposed:

$$\hat{x}(T) \leqslant \hat{C}\,, \tag{36}$$

where $\hat{C} = (0, ..., 0, C)$.

These reasonings show that it is sufficient to develop solution methods only for problem 1 in order to obtain the solution methods for the whole set of DLP problems arising in case studies. But before discussing these methods let us consider some important theoretical properties of the DLP problems.

3.2 DLP *theory*

Problem 1 is an optimal-control problem with state equation (25), initial condition (24), constraints on state and control variables (26), and performance index (27). However, it may also be considered as a 'large' LP problem with constraints on variables in the form of equalities (24) and (25) and inequalities (26). In this case, problem 1 reduces to an LP problem with the staircase constraint matrix (table 1).

For the numerical solution of problem 1, one can therefore rely on a standard LP computer code. However, this straightforward approach to solving DLP problems is inefficient for two reasons. First, the 'static' LP model thus arrived at is so large in real cases that it cannot be solved effectively even by using the most up-to-date computers. Second, and more important, even if the optimal solutions of DLP problem 1 are found by conventional means, problems of implementing this solution still exist. These reasons provide the rationale for the development of DLP theory and methods.

Duality theory plays a key role in optimization methods. It permits the replacement of the original primal problem by some equivalent dual. It should be stressed that this equivalent dual can be interpreted in real terms for all real cases, thus enabling one to understand more deeply the original problem. The following results are true for DLP problem 1 (Propoi, 1977).

Problem 2 (dual): Find the *dual control* $\lambda = \{\lambda(T-1), ..., \lambda(1), \lambda(0)\}$ and the associated *dual trajectory* $\pi = \{\pi(T), ..., \pi(1), \pi(0)\}$ satisfying the *costate (dual) equation*

$$\pi(t) = [G(t)]^{T}\pi(t+1) - [P(t)]^{T}\lambda(t) + \alpha(t) \tag{37}$$

with the boundary condition

$$\pi(T) = \alpha(T)\ , \tag{38}$$

subject to the constraints

$$[D(t)]^{T}\pi(t+1) - [S(t)]^{T}\lambda(t) \leqslant -\beta(t) \qquad \text{and} \qquad \lambda(t) \geqslant 0\ , \tag{39}$$

and minimizing the objective function

$$B_2(\lambda) = [\pi(0)]^{T}x^{\text{init}} + \sum_{t=0}^{T-1}\{[p(t+1)]^{T}e(t) - [r(t)]^{T}\lambda(t)\}\ , \tag{40}$$

where $\pi(t) = [\pi_1(t), ..., \pi_n(t)]$ and $\lambda(t) = [\lambda_1(t), ..., \lambda_m(t)]$, for $\lambda_k(t) \geqslant 0$, $k = 1, ..., m$, are Lagrange multipliers for constraints (24), (25), and (26) respectively.

Dual problem 2 is also a control-type problem, as is Primal problem 1. Here the variable $\lambda(t)$ is a dual control and $\pi(t)$ is a dual or a costate of the system. The costate equations describe the dual system. Note that we have reversed time in dual problem 2: $t = T-1, ..., 1, 0$.

The dual state and control variables (shadow prices) provide valuable information on marginal costs and marginal values for problem 1. They enable one to analyse the effects of changes in the coefficients of the objective function, constraints, growth matrix, etc.

For dual problems 1 and 2, a number of duality relations hold. They are stated here as theorems without proofs—for details see Propoi (1977).

Theorem 1 (the* DLP *'global' duality conditions): (1) For any feasible primal control* v *and dual control* λ, *the inequality

$$B_1(v) \leqslant B_2(\lambda) \tag{41}$$

holds. (2) The solvability of either problem 1 or problem 2 implies the solvability of the other, with

$$B_1(v^{*}) = B_2(\lambda^{*})\ , \tag{42}$$

Table 1. Staircase control structure.

Variable											Inequality or equality of constraints	Right-hand side constant
$v(0)$	$x(1)$	...	$x(t)$	$v(t)$	$x(t+1)$	...	$x(T-1)$	$v(T-1)$	$x(T)$			
$-D(0)$	I										$=$	$e(0)+G(0)x^{\text{init}}$
$S(0)$											$\leqslant$	$r(0)+P(0)x^{\text{init}}$
		⋱									⋮	⋮
			$-G(t)$	$-D(t)$	I						$=$	$e(t)$
			$P(t)$	$S(t)$							$\leqslant$	$r(t)$
						⋱					⋮	⋮
							$-G(T-1)$	$-D(T-1)$	I		$=$	$e(T-1)$
							$P(T-1)$	$S(T-1)$			$\leqslant$	$r(T-1)$
Performance-index constant												
$\beta(0)$	$\alpha(0)$	...	$\alpha(t)$	$\beta(t)$	$\alpha(t+1)$	...	$\alpha(T-1)$	$\beta(T-1)$	$\alpha(T)$		maximize	

where v^* and λ^* are optimal controls (that is, with identical values of the performance index for problems 1 and 2).

Equality (42) shows that the solution of primal problem 1 can be replaced by the solution of dual problem 2, and inequality (41) gives the upper bound of the value of the performance index in problem 1.

The duality relations on theorem 1 are 'global' in character, and are stated for the whole planning period. The duality relations can also be formulated in a local decomposable way for each time period t, $t = 0, 1, ..., T-1$. For this purpose, let us introduce the Hamiltonian

$$H_1[\pi(t+1), v(t)] = [\beta(t)]^{\mathrm{T}}v(t) + [\pi(t+1)]^{\mathrm{T}}\mathrm{D}(t)v(t) \tag{43}$$

for primal problem 1 and

$$H_2[x(t), \lambda(t)] = [\lambda(t)]^{\mathrm{T}}r(t) - [\lambda(t)]^{\mathrm{T}}\mathrm{P}(t)x(t) \tag{44}$$

for dual problem 2.

Theorem 2 (the DLP *'local' duality conditions): (1) For any feasible processes $\{v, x\}$ and $\{\lambda, \pi\}$, the following inequalities hold:*

$$H_1[\pi(t+1), v(t)] \leqslant H_2[x(t), \lambda(t)]\,, \qquad t = 0, ..., T-1\,. \tag{45}$$

(2) For any feasible process $\{v^*, x^*\}$ of the primal and $\{\lambda^*, \pi^*\}$ of the dual to be optimal, it is necessary and sufficient that the values of Hamiltonians are equal:

$$H_1[\pi^*(t+1), v^*(t)] = H_2[x^*(t), \lambda^*(t)]\,, \qquad t = 0, ..., T-1\,. \tag{46}$$

Theorem 2 shows that in order to investigate a pair of dual dynamic problems such as problems 1 and 2 it is sufficient to consider the following pair of dual 'local' (static) problems of LP:

$$\left.\begin{array}{l} \text{maximise}\{H_1[\pi(t+1), v(t)]\} \\ \text{subject to} \\ \mathrm{P}(t)x(t) + \mathrm{S}(t)v(t) \leqslant r(t) \\ \text{and } v(t) \geqslant 0\,, \\ \text{for } t = 0, ..., T-1\,, \end{array}\right\} \tag{47}$$

and

$$\left.\begin{array}{l} \text{minimise}\{H_2[x(t), \lambda(t)]\} \\ \text{subject to} \\ [\mathrm{D}(t)]^{\mathrm{T}}\pi(t+1) - [\mathrm{S}(t)]^{\mathrm{T}}\lambda(t) \leqslant -\beta(t) \\ \text{and } \lambda(t) \geqslant 0\,, \\ \text{for } t = T-1, ..., 1, 0\,. \end{array}\right\} \tag{48}$$

Any of the 'static' duality relations of LP optimality conditions (Dantzig, 1963) for the pair of dual LP problems (47) and (48), linked by the state equations (24) and (25), and (37) and (38), determine the corresponding optimality conditions for the pair of dual DLP problems 1 and 2. Such conditions have been formulated above; in a similar manner the following important optimality conditions are obtained (Propoi, 1973; 1977):

Theorem 3 (maximum principle for primary problem 1): *For a control v^* to be optimal in primary problem 1, it is necessary and sufficient that there exists a feasible process $\{\lambda^*, \pi^*\}$ of dual problem 2 such that, for $t = 0, 1, ..., T-1$, the equality*

$$\max\{H_1[\pi^*(t+1), v(t)]\} = H_1[\pi^*(t+1), v^*(t)]$$

holds, where the maximum is taken over all $v(t)$ satisfying constraints (26), and $\lambda^*(t)$ is the optimal dual variable in LP problem (48).

The minimum principle can be formulated in a similar way for dual problem 2 (Propoi, 1977).

The Hamiltonian functions may be interpreted as 'local' objective functions for each time period t, which consist (for primal problems) of the current profit, $[\alpha(t)]^T x(t) + [\beta(t)]^T v(t)$, modified by the shadow price vector, $\pi(t+1)$—that is, with $[\pi(t+1)]^T D(t)v(t)$ instead of $[\alpha(t)]^T x(t)$—accounting for the dynamics of the process. The maximum principle for problem 1 states that if we know the optimal values of shadow prices $\pi^*(t+1)$ [or $\lambda^*(t)$] then it is possible, for each current state $x(t)$, to solve only the 'local' static LP problem (47).

3.3 DLP *computational methods*

Simple DLP problems can be handled by standard LP codes. DLP problems of a realistic size require, however, the development of special DLP methods. We shall distinguish between finite and iterative methods.

DLP *finite methods* find an optimal solution in a finite number of steps and are a further development of large-scale LP methods to dynamic problems. First, we mention the extension of the well-known simplex method to DLP problems (Propoi and Krivonozhko, 1977). The dynamic simplex method is used to obtain exact optimal solutions of DLP problems, for a finite number of steps, by treating at each step only the set of T local bases of dimension $m \times m$ [m is the number of constraint rows in inequalities (21)] instead of working with the global basis of dimension $mT \times mT$. The method is closely related to the most effective large-scale LP methods based on factorization of the constraint matrix. The second group of finite methods is based on decomposition methods of LP, especially on the Dantzig–Wolfe decomposition principle. This technique was applied to DLP problems by Glassey (1970), Ho and Manne (1974), and Krivonozhko (1977).

Iterative methods do not produce exact solutions in a finite number of iterations. But in many cases the approximation is quite adequate. In addition, the iterative methods are characterized by simplicity of computer coding, low demands on computer memory, and low sensitivity to the disturbances. A DLP iterative method is considered in Propoi and Yadykin (1975; 1976).

Unlike for static LP, the realization of the optimal solution in dynamic problems is as important as its determination. Such questions as the realization of the optimal solution as a program—that is, depending on the number of states: $v^*(t)$, $t = 0, ..., T-1$ —or as a feedback control—that is, depending on the current value of states: $v^*(t) = v_t^*[x(t)]$, $t = 0, ..., T-1$; the stability and sensitivity of the optimal system; and the connection of optimal solutions for long- and short-range models must be answered.

Naturally all the practical problems of national settlement planning cannot be kept within the format of DLP. In some cases the objective function is stated as a quadratic or nonlinear (convex) function of state and control variables (Willekens, 1976b). The extension of DLP methods to quadratic and convex dynamic-programming problems can be developed in a way similar to the static methods (see Dantzig, 1963).

When the exogenous variables cannot be given a priori, we come to DLP problems with uncertainty conditions. They can be formalized using stochastic optimization methods (Ermoljev, 1975) or max-min methods (Propoi and Yadykin, 1974a; 1974b). Such optimization techniques may be conceptually superior in some instances, but operational problems are more formidable because of data requirements and solution difficulties.

4 Conclusion

In this introductory paper, we sketched the basic idea of applying DLP to the planning of national settlement systems. The DLP approach requires a linear model. Several situations may arise in which these conditions are satisfied. For instance, most currently available multiregional population-growth models are linear. Moreover a wide variety of DLP formulations are possible. It has been shown in this paper how they may be reduced to the standard or canonical form.

Interest in dynamic linear programming started recently and the theory and algorithms are still being developed. DLP, which combines linear-programming and optimal-control theory, provides a tool for the solution of large-scale dynamic problems. The second section of this paper has given a short overview of recent findings in DLP theory and of algorithms which are relevant for formulating problems of national settlement-system planning. We did not intend to present an in-depth analysis of the theory. Our aim was merely to show the potential of DLP to the planning of human-settlement systems. Considerable research is still needed, both theoretical and applied. Interesting topics of further theoretical investigation are, for example, the interpretation of the basic dual relations in demographic terms or the investigation of linkages between demographic and economic models.

References

Charnes A, Cooper W W, 1961 *Management Models and Industrial Applications of Linear Programming* volumes 1 and 2 (John Wiley, New York)

Cordey-Hayes M, 1975 "Migration and the dynamics of multiregional population systems" *Environment and Planning A* 7 793-814

Dantzig G B, 1963 *Linear Programming and Extensions* (University Press, Princeton, NJ)

Ermoljev Yu M, 1975 "Stochastic models and optimization methods" *Kibernetika* 4 109-119

Evtushenko Y, MacKinnon R D, 1976 "Nonlinear programming approaches to national settlement system planning" *Environment and Planning A* 8 637-653

Glassey C R, 1970 "Dynamic linear programming for production scheduling" *Operations Research* **19** 45-56

Ho J K, Manne A S, 1974 "Nested decomposition for dynamic models" *Mathematical Programming* **6** 121-140

Kantorovitch L V, 1965 *The Best Use of Economic Resources* (Harvard University Press, Cambridge, Mass)

Krivonozhko V E, 1977 "Decomposition methods for dynamic linear programming" *Izvestiya Akademii Nauk SSSR, Technicheskaya Kibernetika* 3 18-25 (in Russian)

MacKinnon R D, 1975 "Controlling interregional processes of a Markovian type" *Environment and Planning A* 7 781-792

Mehra R K, 1975 "An optimal control approach to national settlement system planning" RM-75-58, International Institute for Applied Systems Analysis, Laxenburg, Austria

Propoi A, 1973 *Elementy Teorii Optimalnykh Discretnykh Protsessov* (Elements of the Theory of Optimal Discrete Systems) (Nauka, Moscow) (in Russian)

Propoi A, 1976 "Problems of dynamic linear programming" RM-76-78, International Institute for Applied Systems Analysis, Laxenburg, Austria

Propoi A, 1977 "Dual dynamic linear programming systems" RR-77-09, International Institute for Applied Systems Analysis, Laxenburg, Austria

Propoi A, Krivonozhko V E, 1977 "The simplex method for dynamic linear programming" RM-77-24, International Institute for Applied Systems Analysis, Laxenburg, Austria

Propoi A, Yadykin A B, 1974a "The planning problem under uncertain demand, I" *Avtomatika i Telemekhanika* 2 78-83 (in Russian)

Propoi A, Yadykin A B, 1974b "The planning problem under uncertain demand, II" *Avtomatika i Telemekhanika* 3 102-108 (in Russian)

Propoi A, Yadykin A B, 1975 "Parametric iterative methods for dynamic linear programming, I: non-degenerate case" *Avtomatika i Telemekhanika* 12 127-135 (in Russian)

Propoi A, Yadykin A B, 1976 "Parametric iterative methods for dynamic linear programming, II: general case" *Avtomatika i Telemekhanika* 2 136-145 (in Russian)

Rees P H, Wilson A G, 1975 "A comparison of available models of population change" *Regional Studies* 9 39-61

Rogers A, 1966 "A Markovian policy model of interregional migration" *Papers of the Regional Science Association* 17 205-224

Rogers A, 1968 *Matrix Analysis of Interregional Population Growth and Distribution* (University of California Press, Los Angeles)

Rogers A, 1971 *Matrix Methods in Urban and Regional Analysis* (Holden-Day, San Francisco)

Rogers A, 1975 *Introduction to Multiregional Mathematical Demography* (John Wiley, New York)

Stern H J, 1974 "A goal-programming approach to planning for population balance in multiregional system" *Environment and Planning A* 6 431-437

Tinbergen J, 1963 *On the Theory of Economic Policy* (North-Holland, Amsterdam)

Willekens F, 1976a "The analytics of multiregional population distribution policy" unpublished Ph D dissertation, Northwestern University, Evanston, Illinois

Willekens F, 1976b "Optimal migration policies" RM-76-50, International Institute for Applied Systems Analysis, Laxenburg, Austria

Willekens F, Rogers A, 1977 "Normative modeling in demo-economics" RR-77-23, International Institute for Applied Systems Analysis, Laxenburg, Austria

Wilson A G, 1974 *Urban and Regional Models in Geography and Planning* (John Wiley, Chichester, Sussex)

Zadeh L A, Desoer C H, 1963 *Linear Systems Theory. The State Space Approach* (McGraw-Hill, New York)

APPENDIX

Alternative formulations of the DLP problem

The purpose of this appendix is to present some modifications and variants of problem 1. For example, *state equations* may include matrices and/or vectors that are independent of the stage t—equations S2. The external disturbance $e(t)$ may vanish—equations (4). Equations S3, for example, are obtained from the difference approximation of the continuous analog of problem 1.

An important class of DLP problems is the system with delays in state and/or control variables—equation S4. For example, in a demographic system the state $x(t+1)$ at the step $t+1$ may depend on certain previous states, $x(t-n_1), x(t-n_2), \ldots, x(t-n_\nu)$, and on certain previous control actions, $v(t-s_1), v(t-s_2), \ldots, v(t-s_\mu)$.

Constraints on the state control variables can be in the form of equalities C2 [as in equations (9)] or in the form of separate inequalities C3 and C4 [as in inequalities (5)-(7)]. These variables can have additional restrictions on sign—inequalities C5 and C6. In some cases the constraints should be considered in the summarized form of inequalities C7 or C8 [as in inequalities (8)].

It is useful to single out the constraints on the left- and/or right-hand side of the trajectory (boundary conditions). For example, the left- and/or right-hand side of the trajectory can be fixed—B1 and B3—or free—B2 and B4.

The number of steps T of the *planning period* can be fixed—P1—or may be defined by some conditions on the terminal state—for example, B3. A typical problem here is to bring a demographic system to a desired population distribution in a minimal number of steps T.

The value of the *objective function* can depend only on the trajectory $\{x(t)\}$ or on the control sequence $\{v(t)\}$, or may be determined only by the terminal state $x(T)$ of the trajectory—equation I2 [as in equations (17)-(19)].

A number of variants of problem 1 are now illustrated.

State equations

S1 $\quad x(t+1) = G(t)x(t) + D(t)v(t) + e(t).$

S2 $\quad x(t+1) = Gx(t) + Dv(t) + e.$

S3 $\quad x(t+1) = x(t) + G(t)x(t) + D(t)v(t) + e(t).$

S4 $\quad x(t+1) = \sum_{j=1}^{\nu} G(t-n_j)x(t-n_j) + \sum_{i=1}^{\mu} D(t-s_i)v(t-s_i).$

Constraints

C1 $\quad P(t)x(t) + S(t)v(t) \leqslant r(t).$

C2 $\quad P(t)x(t) + S(t)v(t) = r(t).$

C3 $\quad P(t)x(t) \leqslant r^{(1)}(t).$

C4 $\quad S(t)v(t) \leqslant r^{(2)}(t).$

C5 $\quad x(t) \geqslant 0.$

C6 $\quad v(t) \geqslant 0.$

C7 $\quad \sum_{\tau=0}^{t-1} [P(\tau)x(\tau) + S(\tau)v(\tau)] \leqslant r(t), \qquad t = 1, \ldots, T.$

C8 $\quad \sum_{t=0}^{T-1} [P(t)x(t) + S(t)v(t)] \leqslant r.$

Boundary conditions

B1 $\quad x(0) = x^{\text{init}}.$

B2 $\quad x(0)$ is free.

B3 $\quad x(T) = x^{\text{hor}}.$

B4 $\quad x(T)$ is free.

Planning period

P1 T is fixed.

P2 T is free.

Performance indices

I1 $B_1(v) = [\alpha(T)]^{\mathrm{T}} x(T) + \sum_{t=0}^{T-1} \{[\alpha(T)]^{\mathrm{T}} x(t) + [\beta(t)]^{\mathrm{T}} v(t)\}.$

I2 $B_1(v) = [\alpha(T)]^{\mathrm{T}} x(T).$

Environment and Planning A, 1978, volume 10, pages 577-591

Optimization of rural-urban development and migration

R Kulikowski¶
International Institute for Applied Systems Analysis, Laxenburg, Austria
Received 17 January 1978

Abstract. This paper deals with rural-urban migration and the impact of migration on regional development. The general problem of optimal allocation of production factors in time and space is introduced in order to derive the most satisfactory migration strategy. In describing the regional economy by means of a generalized Cobb-Douglas production function, the general problem is decomposed into two levels. At the first level, optimal allocation of factors in time is solved in an explicit form using the generalized Hoelder inequality. At the second level, a spatial strategy is derived and the principle of spatial allocation of production factors is formulated. By use of the optimal strategies, the simple two-sector model (that is, agriculture and the rest of the economy) is investigated, and the labor surpluses in Polish agriculture and in an agricultural region in Poland are calculated.

1 Introduction and formulation of the basic optimization problem

Rural-urban migration usually includes agricultural labor which becomes employed in the nonagricultural sector of the economy. In this case the migration processes have a direct effect on production structure and regional development. It will be assumed in this paper that the total labor supply in the rural-urban region to be analyzed is predetermined by demographic factors. When mass migration starts—sometimes called the "mobility revolution" (Rogers, 1977)—the supply of labor is greater in rural areas than urban centers, whereas the demand for labor is greater in industrial centers, located mainly in towns.

The main question asked by regional-development planners is: what is the optimal allocation of labor to produce the greatest acceleration of regional growth? Planners feel that when they discover the optimal allocation of labor, they can also determine the labor surplus in agriculture which can be regarded as a labor reservoir for urban growth. The transfer of the labor surplus to nonagricultural sectors is, however, difficult, and it involves social costs, for example, additional housing programs, the creation of new jobs, a change in the traditional agricultural economy, and increased environment-protection programs (Herer and Sadowski, 1975).

Owing to high housing and urban-development costs in many countries, a large group of so-called 'commuting' migrants exists, who live in the country and work in the cities. They often spend two to four hours a day commuting. The lost working time and the increased transportation costs (aggravated by energy crises) represent a heavy burden to the economy.

In order to determine optimal policies for development and migration in each particular region, cost-benefit analyses should be carried out on the regional level. In particular, one would like to know whether it is better to improve transportation, and in turn increase the number of commuting migrants; to increase housing construction in the cities, and thus encourage out-migration from rural areas; or to use the capital in the region to build factories, thereby employing the labor surplus.

¶ On leave from Systems Research Institute, Polish Academy of Sciences, POB 22, 00/901 Warsaw, Poland.

To obtain an answer to this question, it is advisable to use macroeconomic concepts based on the allocation of production factors in the neoclassic two-sector models (Denison, 1967; Harris and Todaro, 1970; Rogers, 1977; Kuznets, 1960). In these models, migration depends on the labor employed and on labor productivity in agriculture and in the rest of the economy. Here, however, an attempt is made to extend these concepts in such a way that differences in all the factors, per worker, can be taken into account.

Since regional development is a complex process which takes place in time and space, one would like to find the optimal development and migration strategies as explicit functions of time and location. Such an approach requires a more general formulation of the classical problem of allocation of production factors. In the classical macroeconomy, the production function ϕ is usually assumed to be a concave (homogenous of degree one) differentiable function of production factors, say F_ν, for $\nu = 1, ..., m$ (1). The output Y can be written

$$Y = \phi(F_1, ..., F_m) .$$

The values of Y and F_ν, for $\nu = 1, ..., m$, describe the aggregated and averaged (usually within one year) processes which take place in time and space.

A more general expression for the production function,

$$y(s, t) = \phi[s, t, f_1(s, t), ..., f_m(s, t)] , \tag{1}$$

includes the time (t) and location (s) variables. In this case it is convenient to regard y and f_ν as production and factor intensities. For example, the agricultural production intensity depends on the geographical location s and time of the year t.

A typical example of equation (1) is the generalized Cobb-Douglas production function,

$$y(s, t) = A(s, t) \prod_{\nu=1}^{m} [f_\nu(s, t)]^{\beta_\nu} , \tag{2}$$

where the β_ν are constants such that $\sum_{\nu=1}^{m} \beta_\nu = 1$ and $\beta_\nu > 0$, for $\nu = 1, ..., m$; and $A(s, t) = A(s)\exp(\mu t)$, where μ is a positive coefficient representing so-called neutral progress, and $A(s)$ is a given positive function of location s.

Another widely used example is the constant elasticity of substitution (CES) production function, which in our notation can be written as

$$y(s, t) = A(s, t) \left\{ \sum_{\nu=1}^{m} \beta_\nu [f_\nu(s, t)]^\epsilon \right\}^{1/\epsilon} , \tag{3}$$

where the β_ν are coefficients with the same constraints as for equation (2) and ϵ is a constant such that $0 \leqslant \epsilon \leqslant 1$.

In the class of all production factors, it shall be assumed that at least one factor, say $f_1(s, t)$, is not mobile (exogenous), that is, cannot be transferred between the individual production processes. Capital is usually regarded as mobile, whereas land is not. Some factors, for example, labor and water, can also be regarded as mobile at additional (transportation) costs. It shall also be assumed that the total amount of each production factor, integrated over the planning interval $[0, T]$ and the region R, where R is assumed to be a closed convex set in two-dimensional Euclidean space, is bounded: that is,

$$\int_R \int_0^T \exp(-\xi_\nu t) f_\nu(s, t)\, dt\, ds \leqslant F_\nu , \qquad \text{for } \nu = 1, ..., m , \tag{4}$$

(1) The index $\nu = 1$ is usually reserved for labor, $\nu = 2$ for capital, whereas $\nu = 3, ..., m$ may represent land, education, etc.

where the bounds F_ν and the coefficients $\xi_\nu \geqslant 0$, for $\nu = 1, ..., m$, are given, and the functions $\exp(-\xi_\nu t)$ describe possible discounting of the initial values of production factors over time.

The basic regional optimization problem can now be formulated as follows: find the nonnegative strategies $f_\nu(s, t) = \hat{f}_\nu(s, t)$, for $\nu = 2, ..., m$, $s \in R$, and $t \in [0, T]$, which maximize the regional integrated product

$$Y = \int_R \int_0^T \exp(-\xi t) y(s, t)\, dt\, ds\,, \qquad \text{where } \xi = \mu + \sum_{\nu=1}^{m} \xi_\nu \beta_\nu\,, \tag{5}$$

subject to total-resource constraints (4) for $\nu = 2, ..., m$. It is possible to formulate a discrete version (in time and space) of equations (4) and (5)—this is a matter of convenience.

In order to solve this problem in time and space, the following *decomposition method* shall be applied.

(1) At the first (local) level, assume s to be fixed and find the dynamic strategies $f_\nu(s, t) = \hat{f}_\nu(s, t)$, for $\nu = 2, ..., m$ and $t \in [0, T]$, which maximize the product of $\exp(-\xi t)$ and $y(s, t)$, integrated over time,

$$Y(s, f) = \int_0^T \exp(-\xi t) y(s, t)\, dt\,, \tag{6}$$

where f is the sequence $\{f_\nu(s, t)\colon\ \nu = 2, ..., m\}$, subject to the time-integrated constraints

$$\int_0^T \exp(-\xi_\nu t) f_\nu(s, t)\, dt \leqslant F_\nu(s) \qquad \text{and} \qquad f_\nu(s, t) \geqslant 0\,. \tag{7}$$

For $\nu = 2, ..., m$; $t \in [0, T]$; and $s \in R$.

(2) Compute

$$F_\nu(s) = \int_0^T \exp(-\xi_\nu t) \hat{f}_\nu(s, t)\, dt\,, \qquad \text{for } \nu = 2, ..., m\,,$$

and

$$Y(s, \hat{f}) = \hat{Y}(s, \boldsymbol{F})\,,$$

where $\hat{f}$ is the sequence $\{\hat{f}_\nu(s, t)\colon\ \nu = 2, ..., m\}$ and $\boldsymbol{F}$ is the sequence $\{F_\nu(s)\colon\ \nu = 2, ..., m\}$.

(3) Find, at the second (spatial) level, the static strategies $F_\nu(s) = \hat{F}_\nu(s)$, for $\nu = 2, ..., m$, which maximize

$$Y = \int_R \hat{Y}(s, \boldsymbol{F})\, ds\,, \tag{8}$$

subject to the spatial constraints

$$\int_R F_\nu(s)\, ds \leqslant F_\nu \qquad \text{and} \qquad F_\nu(s) \geqslant 0\,, \tag{9}$$

for $\nu = 2, ..., m$ and $s \in R$.

In other words, the problem of allocation of resources is solved at the local level in planning interval $[0, T]$. Then by use of the local solution (in which the time variable has been suppressed) concentration is directed onto the allocation of resources over the space R. From the formal point of view, the main question to be asked is the following: will the time-space decomposition method yield the solution which is equal to the solution of the original problem of equation (5) and constraints (4)?

It can be shown (see Kulikowski, 1977b) that both solutions are equivalent when some regularity conditions are imposed on the functionals (4) and (5). For that purpose one can use the generalized Weierstrass theorem for Banach spaces. Since we shall deal with strictly concave, continuous functionals—such as that in equation (5), where $\sum_{\nu=2}^{m} \beta_\nu < 1$—on a compact set—such as that defined by inequalities (4)—the regularity conditions can be assumed to hold.

It should also be noted that factor levels $f_\nu(s, t)$ generally depend on factor endowment intensities $e_\nu(s, t)$ (usually given in monetary terms), which can be described by the following integral operators (Kulikowski, 1977b):

$$f_\nu(s, t) = \int_{-\infty}^{t} \exp[-\delta_\nu(t-\tau)][e_\nu(s, \tau - T_\nu)]^\alpha \, d\tau , \qquad \text{for } \nu = 2, ..., m , \tag{10}$$

where $0 < \alpha \leqslant 1$, T_ν, and δ_ν are positive constants. A typical relation of general form (10) is the relation between investment intensity, $e_2(s, \tau)$, and capital stock, $f_2(s, t)$, where T_2 is the construction delay and δ_2 is the capital depreciation rate.

Note that the maximization of the gross regional product (5) can hardly be regarded as the universal objective of regional development. It has been shown (see Kulikowski, 1977a), however, that the optimal allocation of resources, which follows from the solution of the problem given by equations (2) and (5) and constraints (4), also yields maximum consumption per worker in the planning interval $[0, T]$.

2 Optimal development strategies: the continuous case

Starting with production function (2), let us concentrate on solving the problem at the first level. To simplify the notation, the variable s in equations (2) and (6) and in constraints (7) shall be neglected. It shall also be assumed that labor is not mobile, so $f_1(t)$ will be regarded as a given exogenous variable.

Taking into account the analytical form of equations (2) and (6) and of constraints (7), one can use the Hoelder inequality (Korn and Korn, 1968, page 118). This becomes an equality when the factors rise in constant proportion. Hence

$$f_\nu(t) = \hat{f}_\nu(t) = \frac{F_\nu}{F_1} f_1(t) \exp[(\xi_\nu - \xi_1)t] , \qquad \text{for } \nu = 2, ..., m . \tag{11}$$

Relation (11) has been called the 'principle of factor coordination'. According to that principle, capital, education, research and development, etc should change with the exogenous factor (for example, labor) in fixed proportions. The principle holds also for the CES function (3)—Kulikowski (1977a, appendix).

Factor coordination can be used to derive the $\hat{e}_\nu(t)$ expenditures at the national level,

$$\hat{e}_\nu(t - T_\nu) = \left\{ \frac{\beta_\nu p_1}{\beta_1 p_\nu} [\delta_\nu f_1(t) + f_1'(t)] \right\}^{1/\alpha} , \qquad \text{for } \nu = 2, ..., m , \tag{12}$$

where the p_ν, for $\nu = 1, ..., m$, are prices attached to factors F_ν (Kulikowski, 1977a). It is assumed that the prices p_ν satisfy the monetary balance

$$\sum_{\nu=1}^{m} e_\nu(t - T_\nu) \leqslant y(t) .$$

Owing to delays T_ν, for $\nu = 2, ..., m$, the expenditures connected with investments, education, etc—$\hat{e}_\nu(t)$, for $\nu = 2, ..., m$—should precede the employment factor, $f_1(t)$. This is shown in figure 1 for $f_1(t) = \zeta_0 + \zeta_1 t$, where ζ_0 and ζ_1 are given constants, $\alpha = 1$, and

$$\hat{e}_\nu(t - T_\nu) = \frac{\beta_\nu p_1}{\beta_1 p_\nu} (\zeta_0 \delta_\nu + \zeta_1 + \delta_\nu \zeta_1 t) .$$

As stated in theorem 1 of Kulikowski (1977a, page 26), the optimal factor-allocation strategy, which maximizes Y, is equivalent to the strategy of maximizing consumption per head.

Now the problem can be solved at the second level—equation (8) and constraints (9). It can be shown (Kulikowski, 1977a, page 23) that, under the strategy $f = \hat{f}$, the output $\hat{Y}(s, F)$ becomes

$$Y(s, F) = [G(s)]^{\beta_1} \prod_{\nu=2}^{m} [F_\nu(s)]^{\beta_\nu} , \tag{13}$$

where $G(s) = [A(s)]^{\beta_1^{-1}} F_1(s)$. Now the values of $F_\nu(s) = \hat{F}_\nu(s)$, for $\nu = 2, ..., m$, have to be derived to maximize

$$Y = \int_R [G(s)]^{\beta_1} \prod_{\nu=2}^{m} [F_\nu(s)]^{\beta_\nu} \, ds \tag{14}$$

subject to constraints (9).

Using the Hoelder inequality, one gets

$$Y \leqslant \left[\int_R G(s)\,ds\right]^{\beta_1} \prod_{\nu=2}^{m} \left[\int_R |F_\nu(s)|\,ds\right]^{\beta_\nu} = \left[\int_R G(s)\,ds\right]^{\beta_1} \prod_{\nu=2}^{m} F_\nu^{\beta_\nu} ,$$

where the equality sign appears if and only if $F_\nu(s) = \chi_\nu G(s)$, for some constant $\chi_\nu > 0$, where $\nu = 2, ..., m$. The unknown coefficients χ_ν are found from inequalities (9),

$$\chi_\nu = \frac{F_\nu}{\int_R G(s)\,ds} , \qquad \text{for } \nu = 2, ..., m .$$

Now a theorem can be formulated, which may be called the 'principle of spatial-factors coordination'.

Theorem 1: *Let G(s) be a given integrable function. The optimum spatial allocation of production factors for the problem given by equations (8) and (13) and constraints (9) exists as*

$$\hat{F}_\nu(s) = G(s)\frac{F_\nu}{\bar{F}_1} \tag{15}$$

where $\bar{F}_1 = \int_R G(s)\,ds$ and $\nu = 2, ..., m$, *and this solution is unique.*

From equation (11) one can also find the strategies

$$\hat{f}_\nu(s, t) = [A(s)]^{\beta_1^{-1}} \exp[(\xi_\nu - \xi_1)t] f_1(s, t)\frac{F_\nu}{\bar{F}_1} , \qquad \text{for } \nu = 2, ..., m .$$

Formula (15) can also be used for the case when labor is regarded as mobile while another factor (for example, land) is immobile (exogenous). In that case, the optimum spatial allocation of labor should follow the change in exogenous factor (for example, in the productive efficiency of land).

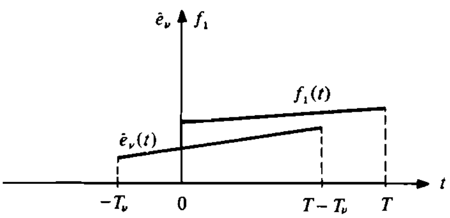


Figure 1. An example of the precedence of expenditures over the employment factor $f_1(t)$.

It should be observed that solution (15) can also be used for the model which shows how the density of population, $D(s)$, is related to the environmental factors specified by $\tilde{G}(s)$. Assuming that the utility function U, representing social preferences of inhabitants with respect to the environment, has a form similar to equation (14),

$$U = \int_R [\tilde{G}(s)]^q [D(s)]^{1-q} \, \mathrm{d}s ,$$

where $0 < q < 1$, and assuming the total population in region R to be bounded,

$$\int_R D(s) \, \mathrm{d}s \leqslant P ,$$

where P is a given positive constant, from equation (15) one gets the optimum strategy of population allocation,

$$\hat{D}(s) = \frac{\tilde{G}(s)}{\int_R \tilde{G}(s) \, \mathrm{d}s} P .$$

When, as usually happens, $G(s) \neq \tilde{G}(s)$, the existing density of population does not coincide with the best allocation of population from the economic point of view. An important problem for regional planning is to create environmental and living conditions such that $G(s) = \tilde{G}(s)$, that is the utility-maximizing strategy should coincide with the production-maximizing strategy. In general this involves many expenses, so other solutions should be considered as alternatives (for example, rapid transportation, better housing, and programs for industrial-pollution abatement).

3 Optimal regional allocation of production factors: the static discrete case

Instead of dealing with continuous strategies, it has become customary at the regional level to deal with models which are static in time and discrete in space. Therefore, the discrete version of strategy (15) shall be discussed. Labor shall be regarded as mobile at some additional transfer costs.

Generally speaking, the supply and demand for production factors in several regions may be different and may change in time. It is possible, however, to transfer some production factors between regions at additional costs. As a typical example, consider a two-sector, two-region system. The first region, R_1, represents the rural part of the country with agricultural production, whereas the second region, R_2, represents the urban part of the country and the rest of the economy. During the industrialization period, the demand for labor at R_2 is greater than at R_1, whereas the supply at R_1 is greater than at R_2. At the same time, the labor efficiency at R_2 is greater than at R_1. The migration of labor from R_1 to R_2 is hampered by high costs of housing, urbanization, etc, and in many countries an antimigration policy is adopted. In order to find out what the best government policy in migration should be, one should take into account the losses due to inefficient allocation of labor and due to migration costs. The best migration policy corresponds to the minimum value of the resulting loss function.

In order to derive the optimal allocation strategy, assume for the moment that all factors are mobile and that interregional transfer costs are not involved. The production function of a plant belonging to production sector i located in region R_j can be written in the form (Kulikowski, 1977a, page 25)

$$Y_j^{(i)} = (G_j^{(i)})^q \prod_{\nu=1}^{m} (E_{j\nu}^{(i)})^{\gamma_\nu} ,$$

where $q = 1 - \sum_{\nu=1}^{m} \gamma_\nu$, and $\gamma_\nu = \alpha\beta_\nu$, and where $Y_j^{(i)}$ is the output for sector i and region j, $G_j^{(i)}$ is a given positive coefficient, and $E_{j\nu}^{(i)}$ is the total expenditure for factor ν in sector i and region j.

In dealing with the allocation of production factors, it is convenient to introduce a three-level optimization structure. At optimization level 1, resources are allocated among sectors and it is necessary to find strategies $E_{j\nu}^{(i)} = \hat{E}_{j\nu}^{(i)}$, for $i = 1, ..., n$ and $\nu = 1, ..., m$, which maximize the regional (R_j) production

$$Y_j = \sum_{i=1}^{n} Y_j^{(i)}$$

subject to

$$\sum_{i=1}^{n} E_{j\nu}^{(i)} \leqslant E_{j\nu} , \qquad \text{for } \nu = 1, ..., m .$$

By use of the discrete version of equation (15), it can be shown that

$$\hat{E}_{j\nu}^{(i)} = \left(\frac{G_j^{(i)}}{G_j}\right) E_{j\nu} , \qquad \text{for } \nu = 1, ..., m , \quad \text{and} \quad i = 1, ..., n ,$$

where $G_j = \sum_{i=1}^{n} G_j^{(i)}$, and then

$$Y_j(\hat{E}) = G_j^q \prod_{\nu=1}^{m} E_{j\nu}^{\gamma_\nu} ,$$

where $\hat{E}$ is the sequence $\{\hat{E}_{j\nu}^{(i)}: \ i = 1, ..., n\}$. At optimization level 2, resources are allocated among regions and it is necessary to find $E_{j\nu} = \hat{E}_{j\nu}$, for $j = 1, ..., r$ and $\nu = 1, ..., m$, which maximize

$$Y = \sum_{j=1}^{r} Y_j(\hat{E})$$

subject to

$$\sum_{j=1}^{r} E_{j\nu} \leqslant E_\nu , \qquad \text{for } \nu = 1, ..., m .$$

The optimal strategies according to the discrete version of equation (15) become

$$\hat{E}_{j\nu} = \left(\frac{G_j}{G}\right) E_\nu , \qquad \text{for } \nu = 1, ..., m \quad \text{and} \quad j = 1, ..., r , \tag{16}$$

and

$$Y(\hat{E}) = G^q \prod_{\nu=1}^{m} E_\nu^{\gamma_\nu} , \tag{17}$$

where $G = \sum_{j=1}^{r} G_j$.

At optimization level 3, the optimal allocation of resources is $\hat{E}_\nu = \beta_\nu Y$, where Y can be determined from page 25 of Kulikowski (1977a).

The formula in equation (17) expresses the GNP under the assumption of full factor mobility.

In order to find the optimal allocation of labor in a two-sector, two-region system, assume the total labor supply L to be given, $L = L^{(1)} + L^{(2)}$, where $L^{(i)}$, for $i = 1, 2$, is the labor supply in production sector i, and find by equation (16) the optimum labor allocation,

$$\frac{\hat{L}^{(1)}}{\hat{L}^{(2)}} = \frac{G^{(1)}}{G^{(2)}} \,. \tag{18}$$

Assume also that

$$Y^{(i)} = (G^{(i)})^q (L^{(i)})^{\beta_1} \prod_{\nu=2}^{m} E_\nu^{(i)\beta_\nu^{(i)}} \,,$$

where $q = 1-\beta-\beta_1$, $\beta = \sum_{\nu=2}^{m} \beta_\nu^{(1)} = \sum_{\nu=2}^{m} \beta_\nu^{(2)}$, for $i = 1, 2$. Using equation (18) one finds that

$$\frac{\hat{L}^{(1)}}{L^{(1)}} : \frac{\hat{L}^{(2)}}{L^{(2)}} = \left\{ \left[\frac{Y^{(1)}}{L^{(1)}} \prod_{\nu=2}^{m} \left(\frac{E_\nu^{(2)}}{L^{(2)}} \right)^{\beta_\nu^{(2)}} \right] \Big/ \left[\frac{Y^{(2)}}{L^{(2)}} \prod_{\nu=2}^{m} \left(\frac{E_\nu^{(1)}}{L^{(1)}} \right)^{\beta_\nu^{(1)}} \right] \right\}^{1/q} \tag{19}$$

If sector 1 represents agriculture and sector 2 the rest of the economy then

$$\hat{L}^{(1)} = L^{(1)} - L^{\text{sur}} \qquad \text{and} \qquad \hat{L}^{(2)} = L^{(2)} + L^{\text{sur}} \,,$$

where L^{sur} is the surplus of labor in the agricultural sector, equal to the shortage of labor in nonagricultural sectors. One can then write

$$\frac{\hat{L}^{(1)}}{L^{(1)}} : \frac{\hat{L}^{(2)}}{L^{(2)}} = \frac{1-x_1}{1+\lambda x_1} \,,$$

where $x_1 = L^{\text{sur}}/L^{(1)}$ and $\lambda = L^{(1)}/L^{(2)}$.

In the simple case when $m = 1$ and $q = 0{\cdot}5$, one gets from equation (19),

$$\frac{1-x_1}{1+\lambda x_1} = \psi^2 \tag{20}$$

where $\psi = (Y^{(1)}/L^{(1)}) : (Y^{(2)}/L^{(2)})$ is the ratio of labor productivities in agriculture to those in the rest of the economy.

According to statistical data on the Polish economy for 1970–1975 (Central Statistical Office, Warsaw, 1976), the ratio of labor productivities ψ decreased from 0·33 to 0·24, while $L^{(1)}/L^{(2)}$ decreased from 0·625 to 0·529. In figure 2 the graph of x_1 against ψ for $q = 0{\cdot}5$ and $\lambda = 0{\cdot}529$ is shown. The surplus of labor in 1970 was around $4{\cdot}5 \times 10^6$ people.

By use of formula (20) it is also possible to find the surplus of labor at the regional level. Since labor can migrate over the whole country, it may be assumed that for each region j, $\lambda_j = \lambda$ (where $\lambda_j = L_j^{(1)}/L_j^{(2)}$), that is, that the value of λ_j for

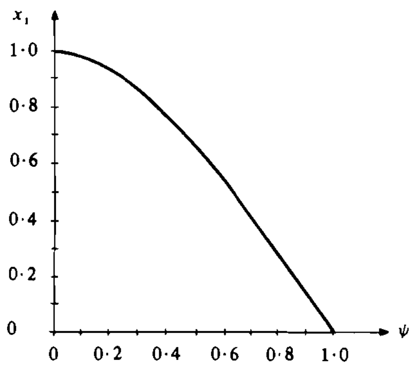


Figure 2. The graph of x_1 against ψ when $q = 0{\cdot}5$ and $\lambda = 0{\cdot}529$.

each region j is the same as the value of λ for the whole country. From this assumption and from data on the labor productivities at the micro-level, labor surpluses can be determined in a relatively simple way. In particular, such an approach has been used to determine the labor surplus in Drobin County, a typical rural area in the central part of Poland. The total number of inhabitants was 9990. The labor-productivity ratios were between 0·85 and 0·9 for the individual villages. This corresponds to a value of x_1 from 0·85 down to 0·15 and is illustrated by figure 3[2]. Differences in productivity depend largely on the size of privately owned farms. In Drobin, the great surplus of labor can be explained by the extremely small farms. The farmers, however, take advantage of bus transportation which enables them to commute to work in such industrial towns as Płock and Warszawa (Warsaw).

It should be observed that the total migration from rural to urban areas per year in Poland is around 150×10^3 to 245×10^3 persons, so it will take 20-25 years to transfer the whole of surplus labor. The explanation for the small migration figure is simple. Mass migration involves such costs as housing, urban services, training, and environmental protection. Some of these costs are connected with additional nonproductive investments. The optimal migration should therefore be chosen in such a way that the loss function Δ, which takes into account the losses due to inefficient allocation of labor and due to migration costs, is minimal—these losses are denoted by $\bar{c}L^{\mathrm{sur}}$, where $\bar{c}$ is a cost per migrant. In the simple model where $m = 1$

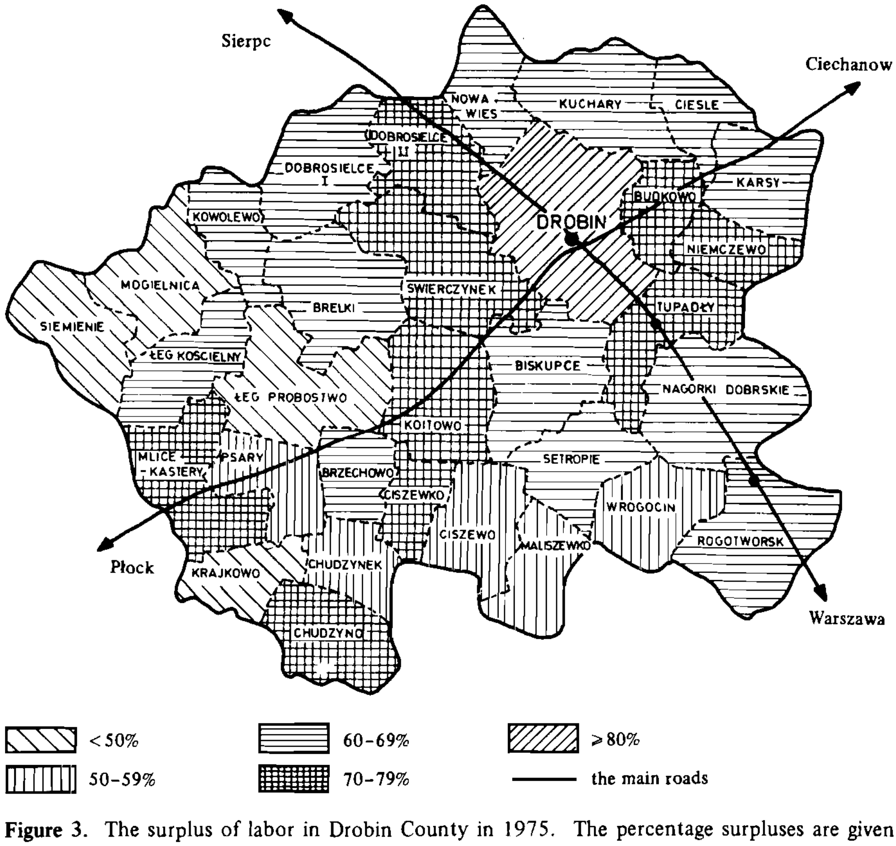


Figure 3. The surplus of labor in Drobin County in 1975. The percentage surpluses are given by $100 \times x_1$.

(2) The calculations were performed by W Kulikowski.

and $q = 1-\beta$,

$$\Delta = G^q L^\beta - (G^{(1)})^q (L^{(1)} - L^{\text{sur}})^\beta - (G^{(2)})^q (L^{(2)} + L^{\text{sur}})^\beta + \bar{c}L^{\text{sur}}$$

$$= G^q L^\beta \left[1 - \left(\frac{G^{(1)}}{G}\right)^q \left(\frac{L^{(1)}}{L} - \lambda\right)^\beta - \left(\frac{G^{(2)}}{G}\right)^q \left(\frac{L^{(2)}}{L} + \lambda\right)^\beta + cx\right],$$

where $G^q L^\beta = \hat{Y}$, $c = \bar{c}L/\hat{Y}$, and $x = L^{\text{sur}}/L$.

The loss function can be easily constructed using statistical data for the Polish economy (Central Statistical Office, Warsaw, 1976). For 1970, assuming $\beta = 0{\cdot}5$, one gets

$$\frac{L^{(1)}}{L} = 0{\cdot}3848, \quad \frac{L^{(2)}}{L} = 0{\cdot}6152,$$

$$\frac{G^{(1)}}{G^{(2)}} = \left(\frac{Y^{(1)}}{Y^{(2)}}\right)^{0\cdot 5} \left(\frac{L^{(2)}}{L^{(1)}}\right) = 0{\cdot}0687,$$

$$\frac{G^{(2)}}{G} = \left(1 + \frac{G^{(1)}}{G^{(2)}}\right)^{-1} = 0{\cdot}9357 = \frac{\hat{L}^{(2)}}{L},$$

$$\frac{G^{(1)}}{G} = \left(1 + \frac{G^{(2)}}{G^{(1)}}\right)^{-1} = 0{\cdot}0643 = \frac{\hat{L}^{(1)}}{L}.$$

Then, by use of this data, one can write

$$\frac{\Delta}{\hat{Y}} = 1 - 0{\cdot}2536(0{\cdot}3848 - x)^{0\cdot 5} - 0{\cdot}9673(0{\cdot}6152 + x)^{0\cdot 5} + cx.$$

Unfortunately, there is not much statistical information available regarding the cost, c. The investment costs estimated by Herer and Sadowski (1975) should in this case be discounted and averaged over the planning interval, and assigned to the operating costs. The cost function, $\Delta/\hat{Y}$, for $c = 0$ and for $c = 0{\cdot}15$, which yields the optimum $\hat{x} = 0{\cdot}25$, has been drawn in figure 4.

It is possible to show that in the case of commuting migration the cost function is nonlinear. A simple model can be constructed in which time lost and cost of transportation can be derived in an explicit form. Assume for this purpose that city C uses a transportation system which delivers commuting labor spread with density D over sector ABC with radius $\mathcal{R}$ (figure 5). The number of commuters X within ABC becomes

$$X = \int_0^{\mathcal{R}} D\,\mathrm{d}\sigma = \int_0^{\mathcal{R}} D\theta\rho\,\mathrm{d}\rho = d\mathcal{R}^2,$$

where $d = \theta D/2$. If the transportation cost is c_t (per person per kilometre), the

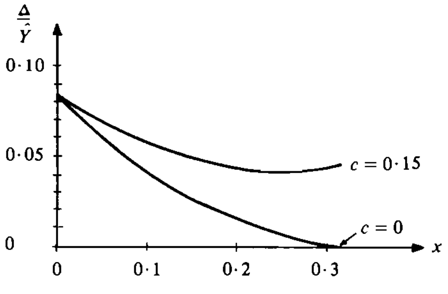


Figure 4. The function $\Delta/\hat{Y}$ for $c = 0$ and $c = 0{\cdot}15$.

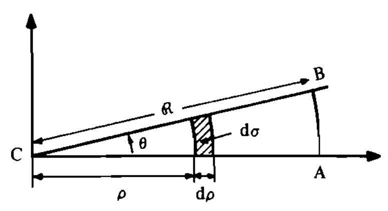


Figure 5. The sector of commuters to city C.

transportation cost C_t of X becomes

$$C_t = \int_0^{\mathcal{R}} c_t \rho D \, d\sigma = b_t \mathcal{R}^3 ,$$

where $b_t = c_t \theta D/3$. If the transport velocity is v km h^{-1}, the time taken to travel ρ km is ρ/v. Let the cost of one hour to a worker be c_h. Then the cost C_h of time lost during transportation is

$$C_h = \int_0^{\mathcal{R}} \frac{\rho}{v} c_h D \, d\sigma = b_h \mathcal{R}^3 ,$$

where $b_h = c_h \theta D/3$. Since $\mathcal{R} = (X/d)^{1/2}$, one gets the total cost C as

$$C = C_t + C_h = BX^{3/2} ,$$

where $B = b/d^{3/2}$ and $b = b_t + b_h$.

In the case where commuting migration prevails, which may happen for small X, one should deal with convex cost functions. A value X_0 exists such that for $X > X_0$ the commuting costs are greater than migration costs, and it is more profitable to let the migrants settle in urban centers.

The present statistical model does not tell us what the best migration strategy is, as a function of time. That problem is studied in the next section, under the assumption that labor resources, integrated over $[0, T]$ are given.

It follows from equation (19) that the migration necessary for maximum gain depends largely on the ratio of labor productivities, $(Y^{(1)}/L^{(1)}) : (Y^{(2)}/L^{(2)})$, and on $(K^{(1)}/L^{(1)}) : (K^{(2)}/L^{(2)})$, that is, on capital allocation (where $K^{(1)}$ and $K^{(2)}$ are the capital stocks in sectors 1 and 2). In general the allocation model of governmental expenditures on education, services, housing, and the environment (characterized by E_ν, for $\nu = 3, ..., m$) also plays an important role. Data on transfer costs of corresponding services are, however, seldom available. When relation (19) is used ex ante in the planning interval $[0, T]$, one can also find the best labor allocation, on condition that the remaining factors are allocated in an optimal fashion, that is, according to strategy (16).

4 Optimum regional allocation of resources: the dynamic continuous case

Since the labor surplus in a particular region can be determined (by the method described in section 3), one can concentrate on dynamic optimization with given labor and capital resources. In this case, one would like to achieve the fastest possible regional industrial growth, assuming the labor-supply intensity $l(t)$, to be constrained, in the integral sense, within the planning interval $[0, T]$:

$$\int_0^T w_1(t) l(t) \, dt \leqslant E_1 , \qquad (21)$$

where E_1 is the total labor cost (wages) and $w_1(t)$ is a given weight or wage function. Of interest are two production factors only, labor and capital stock, under the assumption that the investment $e(t)$ is constrained by

$$\int_0^T w_2(t) e(t - T_2) \, dt \leqslant E_2 , \qquad (22)$$

where E_2 is the total investment cost and $w_2(t)$ is a given weight function. When new investments are financed out of a bank loan, with interest rate η, it is natural to assume that $w_2(t) = (1+\eta)^{T-t}$. Since the migration cost, c_m, is partly connected with new investments (in particular, housing can be regarded as accompanying the productive investments e), one can write $c_m = ue$, where u is a constant.

The production function for the model investigated can be written in the form of equations (2) and (10):

$$y(t) = A(t)[l(t)]^{\beta}\left\{\int_0^t \exp[-\delta(t-\tau)][e(\tau-T_2)]^{\alpha}\,\mathrm{d}\tau\right\}^{1-\beta} .$$

where $A(t)$ was defined in section 1.

The problem of dynamic optimization of regional development can be formulated as follows: find the strategies $l(t) = \hat{l}(t)$ and $e(t) = \hat{e}(t)$, for $t \in [0, T]$, such that the regional contribution to the GNP,

$$Y(l, e) = \int_0^T \exp(-\xi t)y(t)\,\mathrm{d}t ,$$

attains a maximum subject to constraints (21) and (22). It should be observed that the present problem differs from the problem discussed in section 2, equations (11) and (12), where the labor supply was given.

The main idea in solving the present problem is to use the Hoelder inequality twice. Define $f(t)$ and $\phi(t)$ by

$$f(t) = A(t)\exp(-\xi t)[w_1(t)]^{-\beta}\left\{\int_0^t \exp[-\delta(t-\tau)][e(\tau-T_2)]^{\alpha}\,\mathrm{d}\tau\right\}^{1-\beta}$$

and

$$\phi(t) = [w_1(t)l(t)]^{\beta} ,$$

and observe that

$$Y(l, e) \leqslant E_1^{\beta}\left\{\int_0^T [f(t)]^{1/1-\beta}\,\mathrm{d}t\right\}^{1-\beta} ,$$

where the equality sign appears if and only if

$$\phi(t) = \kappa_1[f(t)]^{\beta/1-\beta} ,$$

where $t \in [0, T]$ and κ_1 is a positive constant. Change the order of integration and again use the Hoelder inequality to give

$$\begin{aligned}
&\int_0^T [f(t)]^{1/1-\beta}\,\mathrm{d}t \\
&= \int_0^T \left\{A(t)\exp(-\xi t)[w_1(t)]^{-\beta}\right\}^{1/1-\beta}\left\{\int_0^t \exp[-\delta(t-\tau)][e(\tau-T_2)]^{\alpha}\,\mathrm{d}\tau\right\}\mathrm{d}t \\
&= \int_0^T [w_2(\tau)e(\tau-T_2)]^{\alpha}[w_2(\tau)]^{-\alpha}\left\{\int_{\tau}^T w(\tau)\exp[-\delta(t-\tau)]\,\mathrm{d}t\right\}\mathrm{d}\tau \\
&\leqslant \left[\int_0^T w_2(\tau)e(\tau-T_2)\,\mathrm{d}\tau\right]^{\alpha}\left\langle\int_0^T\left\{[w_2(\tau)]^{-\alpha}\int_{\tau}^T w(t)\exp[-\delta(t-\tau)]\,\mathrm{d}t\right\}^{1/1-\alpha}\mathrm{d}\tau\right\rangle^{1-\alpha} \\
&= E_2^{\alpha}F^{1-\alpha} ,
\end{aligned} \tag{23}$$

where

$$F = \int_0^T\left\{[w_2(\tau)]^{-\alpha}\int_{\tau}^T w(t)\exp[-\delta(t-\tau)]\,\mathrm{d}t\right\}^{1/1-\alpha}\mathrm{d}\tau$$

and

$$w(t) = A(t)\{\exp(-\xi t)[w_1(t)]^{-\beta}\}^{1/1-\beta} .$$

The equality sign in inequality (23) appears if and only if

$$e(t) = \hat{e}(t) = \kappa_2 g(t) ,$$

for some positive constant κ_2, and where

$$g(\tau) = \left\{[w_2(\tau)]^{-\alpha}\int_\tau^T w(t)\exp[-\delta(t-\tau)]\,dt\right\}^{1/1-\alpha} .$$

Then

$$Y(l, z) \leqslant F^{(1-\beta)(1-\alpha)}E_1^\beta E_2^{\alpha(1-\beta)} . \qquad (24)$$

After finding κ_1 and κ_2 by use of inequalities (21) and (22), the optimum strategies can be written as

$$\hat{e}(t) = \frac{g(t)}{\int_0^T w_2(t)g(t)\,dt}E_2 \quad \text{and} \quad \hat{l}(t) = \frac{h(t)}{\int_0^T w_1(t)h(t)\,dt}E_1 , \qquad (25)$$

where

$$h(t) = \left\{\frac{A(t)\exp(-\xi t)}{w_1(t)}\int_0^t \exp[-\delta(t-\tau)][\hat{e}(\tau-T_2)]^\alpha\,d\tau\right\}^{1/1-\beta} .$$

Assuming that $A(t) = w_1(t) = w_2(t) = 1$, $\xi = 0$, $\alpha = \beta = \frac{1}{2}$, and $T_2 = 0$, one gets

$$g(t) = \{1-\exp[-\delta(T-t)]\}^2\delta^{-2}$$

and

$$h(t) = H\{1-\exp(-\delta t)-\exp(-\delta T)\sinh(\delta t)\}^2 ,$$

where

$$H = \frac{E_2}{G\delta^4} , \quad \text{and} \quad G = \int_0^T g(t)\,dt .$$

In figure 6 the form of optimum strategies (25) for $\delta T = 4$ are shown. It is interesting to observe that as $t \to T$ the investment intensity $e(t)$ goes down. At the same time, productive capital stock,

$$f_2(t) = \int_0^t \exp[-\delta(t-\tau)][\hat{e}(\tau)]^{1/2}\,d\tau ,$$

increases (as shown by the dotted line). Employment, that is, migration intensity $\hat{l}(t)$, increases in constant proportion to capital stock, according to the factor-coordination principle (new jobs and housing being available).

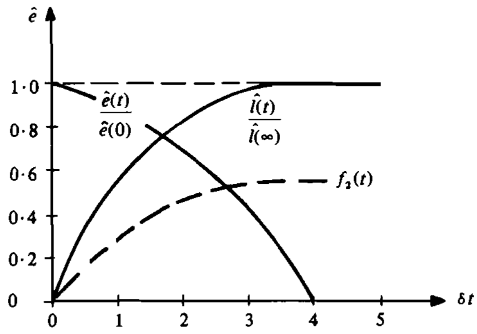


Figure 6. The optimal strategies (25) for $\delta T = 4$.

It can be shown that the maximal regional growth, specified by the right-hand side of equation (24), with the integral constraints (21) and (22), is not smaller than that which any exogenous strategy, $l(t)$ or $e(t)$, satisfying constraints (21) and (22), can produce. However, that strategy can be exercised only when the labor and capital-stock resources exist and can be effectively used.

When E_2 in equation (24) includes the migration investment cost, the productive investment $\hat{e}$ is reduced by a factor of $1-u$. Output (24) is then reduced by $(1-u)^{\alpha(1-\beta)}$.

The present dynamic model of regional growth can be easily extended to the general case with m production factors. From the point of view of optimization of migration policy, the most interesting case concerns the situation when skilled labor is needed and part of the regional budget is spent on education, research, development, services, and the environment. Obviously the factor-coordination principle can be applied here, and the corresponding strategies can be derived in an explicit form.

5 Conclusion

From the analysis carried out in section 3, the surpluses (or deficits) of labor, L_j^{sur}, for each subregion R_j within the planning interval, can be estimated. It was shown in section 4 that, for the given subregion R_j and integrated costs of labor and capital (E_{1j} and E_{2j}), one can find the optimal allocation of labor (l_j) and investments (e_j) over time.

Local labor and capital resources can in general be assigned to three alternative development strategies:

(1) the out-migration (or in-migration) of labor (l_j^1), involving the investment e_j^1;
(2) the commuting of labor (l_j^2), involving the investments e_j^2; and
(3) the in-transfer (or out-transfer) of capital (e_j^3), which yields employment for l_j^3,

where $\sum_{k=1}^{3} l_j^k = L_j^{\text{sur}}$, for $j = 1, ..., r$.

Each strategy involves different operating (or maintenance) costs, such as E_1 in constraint (21),

$$C_j^k(l_j^k) = E_{1j}^k\,, \qquad \text{for } k = 1, 2, 3\,; \quad j = 1, ..., r\,,$$

and capital costs, such as E_2 in constraint (22),

$$C_j^k(e_j^k) = E_{2j}^k\,, \qquad \text{for } k = 1, 2, 3\,; \quad j = 1, ..., r\,.$$

To the C_j^k costs one should assign also the costs of social, technological, and environmental changes.

Using the cost-benefit approach, one can investigate the ratios

$$\pi_j^k = \frac{E_{1j}^k + E_{2j}^k}{Y_j^k(l_j^k, e_j^k)}\,, \qquad \text{for } k = 1, 2, 3\,; \quad j = 1, ..., r\,,$$

and find the indices j and k which render the smallest value of π_j^k. Another possible approach is to find (by the method used in section 4) the strategies $l_j^k = \hat{l}_j^k$ and $e_j^k = \hat{e}_j^k$, for $k = 1, 2, 3$ and $j = 1, ..., r$, which maximize

$$\sum_{j=1}^{r} \sum_{k=1}^{3} Y_j^k(l_j^k, e_j^k)$$

subject to

$$\sum_{j=1}^{r} \sum_{k=1}^{3} C_j^k(l_j^k) \leqslant E_1 \qquad \text{and} \qquad \sum_{j=1}^{r} \sum_{k=1}^{3} C_j^k(e_j^k) \leqslant E_2\,.$$

The solutions to these problems can be used to determine optimal regional development and optimal interregional migration policies.

Acknowledgement. The author is indebted to Professor Roland Artle of the University of California (Berkeley), a visiting scholar at IIASA, for his valuable comments and discussions regarding the problems studied in this paper.

References

Central Statistical Office, Warsaw, 1976 *Statistical Yearbook of Poland (1976)* (Central Statistical Office, Warsaw)

Denison E F, 1967 *Why Growth Rates Differ. Postwar Experience in Nine Countries* (Brookings Institution, Washington, DC)

Harris J R, Todaro M P, 1970 "Migration, unemployment and development: a two-sector analysis" *The American Economic Review* **60** 126-146

Herer W, Sadowski W, 1975 *Migration from Agriculture* (Pan'stwowe Wydawnictwo Naukowe, Warsaw) in Polish

Korn G A, Korn T M, 1968 *Mathematical Handbook for Scientists and Engineers* 2nd edition (McGraw-Hill, New York)

Kulikowski R, 1977a "Optimization of rural-urban development and migration" RM-77-41, International Institute for Applied Systems Analysis, Laxenburg, Austria

Kulikowski R, 1977b *Systems Analysis and Applications* (Pan'stwowe Wydawnictwo Ekonomiczne, Warsaw) in Polish

Kuznets S, 1960 *Population Change and Aggregate Output in Demographic and Economic Change in Developed Countries—A Conference of the Universities* (National Bureau of Economic Research, Princeton, NJ)

Rogers A, 1977 "Migration, urbanization, resources and development" RR-77-14, International Institute for Applied Systems Analysis, Laxenburg, Austria

Environment and Planning A, 1978, volume 10, pages 593-617

Migration and settlement in Bulgaria

D Philipov¶
International Institute for Applied Systems Analysis, Laxenburg, Austria
Received 6 March 1978

Abstract. This paper examines the recent evolution of Bulgaria's population. It is part of IIASA's comparative study of migration and settlement patterns in its member nations. The paper presents a multiregional demographic analysis of fertility, mortality, and internal migration for a seven-region disaggregation of the Bulgarian state. The results give a detailed view of current spatial population dynamics in the country and offer valuable insights useful for the improvement of national population policy.

Demographers throughout the world have for a long time fixed their attention mainly on fertility (birthrate) and mortality (deathrate) patterns, neglecting to some extent migration within a given population. The reasons for this may have been a lack of efficient mathematical models and poor statistical data on migration. During the last decade these difficulties have been eased: new models have been created both for the study of migration and for the improvement of incomplete data. For the study of the spatial dynamics of a given population, the most useful models are those which analyze the joint evolution of fertility, mortality, and migration patterns. In this spatial analysis of the population of Bulgaria, the models and computer programs used were elaborated at IIASA and presented in a series of IIASA papers. The models were fitted to 1975 data.

Section 1 of this paper describes demographic changes of the Bulgarian population until 1975. Section 2 deals with the preparation of the data to fit the needs of the analysis. The following sections present the results for the models used—the multiregional life table, the population projection, and the stable equivalent population—and demonstrate the models' use in the study of spatial fertility, mortality, and migration patterns as indicated by such variables as the spatial reproduction and migraproduction rates (see section 5.1 or Rogers, 1975b), and spatial migration levels. These are examined together with other observed demographic characteristics to give a full view of the demographic structure of the Bulgarian population in 1975.

1 Overview of the demographic history of Bulgaria

In order to understand the models' results, one needs some background information about the current demographic patterns of the population of Bulgaria. In its demographic development, each nation passes through several stages closely connected with the social and economic history of the nation. Demographic studies in Bulgaria (Stefanov et al, 1974; Naoumov et al, 1974) identify three stages that have affected the present demographic structure of the Bulgarian population: (1) the period until 1920-1925; (2) the period between 1920-1925 and 1945 (the end of the Second World War); and (3) the period after 1945.

The paper begins with a brief description of the changes in fertility, mortality, and migration patterns in Bulgaria.

¶ On leave from the Institute of Statistics, 2 P. Volov Street, Sofia 6, Bulgaria.

1.1 *Fertility*

During the first stage, except for the years of the Balkan War and the First World War, crude birthrates (CBRs) were in the range of 39–42 births per thousand of the population per year. These high numbers are typical of a population that has not yet started its demographic transition. During 1925, the last year of this period, the CBR was 36·9. This marked the beginning of the demographic transition which took place during the second stage. Between 1939 and 1945 the CBRs were around 22. The second stage was characterized by the beginning, though a very slow one, of industrialization and urbanization in the country.

The last stage, the period after the Second World War, was remarkably new in the formation of the demographic structure of the Bulgarian population. This was due to the social and economic changes in the country after 1944, namely, land reforms, socialistic industrialization, collectivization and mechanization of agriculture, emancipation of women, improved health care, and urbanization.

After the fertility compensation period following the war, there appeared a decrease in fertility (figure 1). The lowest CBR observed was for 1966 (14·9), and the net reproduction rates (NRRs) for 1965–1967 were less than one (table 1)—the NRR is the number of babies born per person in a lifetime, after taking account of mortality. This trend was caused by the socioeconomic changes in Bulgaria. For example, there was a clearly identifiable migration of laborers, freed from agricultural pursuits, to urban areas, where industrialization was growing rapidly. A much improved standard of living and quality of life, plus the emancipation of women, who now had greater social and economic occupation, were factors that led to the diminishing number of children born in a given family. It must also be mentioned that, according to Bulgarian tradition, children were added working hands in an agricultural household, but were not so important in an urban household.

The fall in fertility and the increase in the average life expectancy led to an ageing of the population. To counteract this, in the fall of 1967 the government adopted laws for the encouragement of childbearing. As a result, fertility has increased since

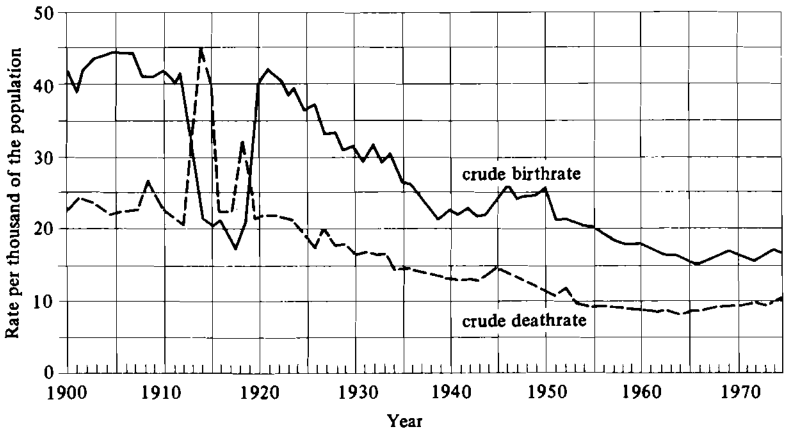


Figure 1. Crude birthrate and deathrate for Bulgaria, 1900–1975 (Central Bureau of Statistics, 1974—readjusted for 1975).

Table 1. Net reproduction rates for the population of Bulgaria, 1965–1975.

Year	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
NRR	0·969	0·943	0·946	1·064	1·078	1·030	0·967	0·963	1·017	1·084	1·055

1968. The fall in the CBR and NRR in 1971 and 1972 is due mainly to the effect of the Second World War on the 20–27 age group.

Since 1956, Bulgaria has been divided into twenty-eight administrative districts. The statistical data for this regional delineation, however, are compatible with data going back to 1947. This allows for the regional comparison of the levels of fertility during the third demographic stage.

At the beginning of the third stage and after the postwar compensation period (around 1950), fertility differed greatly among the districts, ranging between 14 and 36 (figure 2). After 1951, fertility decreased in all the districts, and the decrease was highest for districts with a previously high level of fertility. For instance, the district of Kŭrdzhali exhibited the highest levels of fertility in 1951 and 1975, but the decrease has been substantial: down from 35 to 22·3. The other extreme is the district of Vidin, which exhibited the lowest level of fertility in 1951 (between 14 and 16) and in 1975 (12·6). The same is true for each other district: a high level of fertility at the beginning of the stage leads to a high level at its end, and a low level remains a low level. However, it should be observed that the higher the level of fertility in 1951, the larger its decrease by 1975. So greater uniformity among the districts was achieved—in 1975 their CBRs ranged between 12 and 24.

The pronatalist policy adopted in 1967 has brought a uniform increase in the fertility levels in all the districts with the exception of Kŭrdzhali, which was not influenced at all.

It is clear that the traditional fertility patterns that have been historically established in the separate districts are still clear in 1975. It is to be expected that during the next five or six years the difference between the high and low levels of regional fertility will continue to diminish. The socialistic development and the planned territorial distribution of the productive forces bring further equalization of the economic and cultural quality of life among the districts of the country. Together with the overcoming of religious and other influences comes the elimination of differences in fertility levels among the regions of the state.

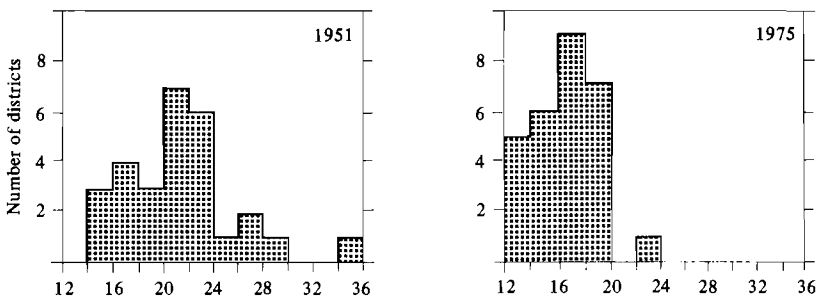


Figure 2. Distribution of the twenty-eight administrative districts of Bulgaria according to the level of fertility (Central Bureau of Statistics, 1974—readjusted for 1975).

1.2 *Mortality*

Until the end of the first stage of demographic development (1920–1925), mortality in Bulgarian was high, with a crude deathrate (CDR) of approximately 23 deaths per thousand of the population per year (figure 1). By the end of this stage and during the second stage, mortality fell together with fertility, the CDR for 1941–1945 having dropped to 13·4. An unusual feature of the Bulgarian demographic transition was the lack of any lag between the fall in fertility and the fall in mortality: the transitional population growth usually occurring in countries in which the decline in mortality appears before the decline in fertility did not occur in Bulgaria.

After the Second World War, as a result of the new conditions of living, the fall in mortality continued. Until 1965 the fall in mortality, together with the fall in fertility, led to the ageing of the population structure, causing the slight increase in the CDR after 1965.

The *expectation of life* gives a better picture of the mortality level than the CDR because it is not influenced by the age composition of the real population. It is common also to say that life expectancy is an indicator of economic development and the standard of living. Table 2 shows that this has been the case for the Bulgarian population. The life expectancy at birth is much higher during the third stage than before. This is a result of the improvement of the health-care system, as well as of the already mentioned socioeconomic changes that have taken place since the Second World War. Life expectancy in 1969–1971 is equal to 71·1 years, and it is approximately the same in 1975.

Life expectancies for Bulgaria's twenty-eight districts are unavailable, but table 3 presents the distribution of the districts during the third demographic stage, arrayed according to the magnitude of the CDRs. At the beginning of the period, mortality was higher and the CDRs varied more. In the middle of the period (1960 and 1965) the range of the CDRs was very narrow and their magnitude was lower. During the last part of the period (1970 and 1975) a rise in the CDR appears in several districts. This rise is a result of the ageing of the population structure in some districts in northern and especially in northwestern Bulgaria, caused by out migration movements during the first two decades of the third stage. Therefore it can be stated that, during the whole third stage, mortality fell, as depicted by the life expectancy for the total population, and it has been the age structure of the population that has caused the rise in some CDRs during the last ten years. It is expected that with the rapid, but uniform, social and economic development of the country, the level of mortality (if measured with age-specific rates) will continue to fall in the long run in all the twenty-eight districts, whereas the CDR, which depends on the peculiarities of the population's age structure, may continue to increase.

Table 2. Life expectancy for the population of Bulgaria during the period 1900–1970 (Central Bureau of Statistics, 1975).

Sex	Period									
	1900–1905	1921–1925	1927–1934	1935–1939	1946–1947	1956–1957	1960–1962	1965–1967	1969–1971	1973–1974
Male	42·1	44·4	47·8	51·0	53·3	64·2	67·8	68·8	68·6	68·9
Female	42·2	45·0	49·1	52·6	56·4	67·7	71·4	72·7	73·9	73·6
Both	-	44·6	48·4	51·8	54·9	65·9	69·6	70·7	71·1	-

Table 3. Distribution of the twenty-eight districts of Bulgaria according to the level of mortality (Stefanov et al, 1974—adjusted for 1975).

CDR	Year					
	1950	1955	1960	1965	1970	1975
5·0–6·9	-	1	2	4	2	2
7·0–8·9	1	13	20	16	10	2
9·0–10·9	20	20	6	8	11	13
11·0–12·9	6	3	-	-	5	6
13·0–14·9	-	1	-	-	-	4
15·0–16·9	1	-	-	-	-	1

Accordingly it cannot be stated that a high or low CDR in 1950 would lead to a high or low CDR in 1975, as was the case for the CBR. On the contrary, the district of Kŭrdzhali had the highest CDR in 1947 (26·9) and the lowest one in 1975 (6·3)! There is no other measure available for the level of regional mortality than the CDR, but it is clear that this measure is not representative because of the effect of the age structure. However, the demographic development of the populations of the districts, as depicted by fertility, suggests that where mortality is concerned there should exist greater uniformity among the districts in 1975 than in 1950, that is, the mortality levels should tend to be equalized.

1.3 *Migrations and territorial structure*

Internal migrations in any country are caused mainly by social and economic factors, but geographical, personal, and ethnical factors must not be neglected. In Bulgaria, migration rates before 1944 were very low because the industrial development of the country was very slow and agriculture was more developed than industry. Some urbanization trends were observed, but they were still not well depicted. For instance, the urban population of the country in 1900 was 19·9% of the total, and in 1934 it was 21·4%.

Table 4 gives the total number of migrations and their number per thousand of the population (the migration rate) over the period 1947–1975. As a result of the social, economic, and cultural changes after 1944, the rate of migration movements began to rise. The economic factors, being the most important reasons for migration, caused the younger part of the active population, together with pupils and students, to migrate. Because of the collectivization and the mechanization of agriculture, a large mass of the labor force moved to the urban areas, where there was a need for workers in the newly developed heavy industry. Therefore the change in the territorial structures can be best observed in the rural–urban structure resulting from the territorial changes in the structure of the social and labor product.

The urban population was 24·7% of the total in 1946, 46·5% in 1965, and 58·7% in 1975. This intensive growth appears for the first time in the demographic history of Bulgaria. The urbanization arises as a result of three main factors: migration to the urban areas, higher fertility in the urban population (insofar as its age structure is younger than that of the rural population), and the administrative reclassification of villages into towns or parts of towns. (Such reclassification involved 283 villages during the period 1945–1971, and transferred 764000 people from rural to urban status.)

The migration flow from rural to urban areas was most intensive after the Second World War, and lasted until the early sixties. Later this flow decreased because the urban population had increased and the rural one had diminished. In fact the absolute number of migrants also diminished in the period between 1960 and 1975.

Table 4. Total numbers of migrations and the migration rates for Bulgaria, 1947–1975 (Stefanov et al, 1974; Central Bureau of Statistics, 1972; 1973; 1974; 1975).

Period	Migrants (in thousands)	Migration rate (per thousand)	Period	Migrants (in thousands)	Migration rate (per thousand)
1947-1950	117·8	16·4	1970	155·7	18·4
1951-1955	138·9	18·9	1971	155·6	18·3
1956-1960	158·1	20·5	1972	151·1	17·7
1961-1965	160·4	19·9	1973	170·0	19·8
1966-1968	156·8	18·9	1974	142·1	16·4
1969	152·3	18·1	1975	124·1	14·2

This was due mainly to the direct and indirect policy of the Bulgarian government. Because of the uniform economic development of all the districts in Bulgaria, and because of the equalization of living conditions in the towns and villages, it is expected that in the next five or ten years the migration movements will drop to a low and constant level.

When only interdistrict migrations are considered, net migration flows are positive only for five districts: Sofia (city), Varna, Gabrovo, Ruse, and Stara Zagora; negative flows appear for sixteen districts; and a mixture of positive and negative net migrations appear for the remaining seven districts. Most of the last seven districts exhibit a positive flow until 1960-1965, and a negative one afterwards. Since 1965 the intensity of the flows has been decreasing or has remained constant for most of the districts. Interregional migrations are studied in greater detail in the following sections.

1.4 *Age structure of the population*

The age structures of the rural, urban, and total populations at the end of 1975 are presented in figure 3. They are a result of the changes in the fertility, mortality, and migration patterns that have been briefly explained in the preceding subsections. The inferences that might be made are as follows.

(1) The relatively low number of people in the 55-59 age group was caused by World War I (stage 1); the relatively low numbers in the 30-39 age group was caused by World War II (stage 2). The relatively low numbers in the 5-14 age group is due both to the low fertility level in the 1960-1969 period and to the low number of people of fertile age (stages 1 and 2).

(2) The size of the urban population at ages up to 54 is higher than that of the rural population (owing to the strong migration flows from the rural to urban areas in stage 3). For the older ages, the size of the rural population is larger. The urban population has a young age structure and the rural population has an old age structure.

(3) Fertility is low in the rural areas because of the small number of people of fertile age (15-44), and is especially low for the most fertile age group (15-29).

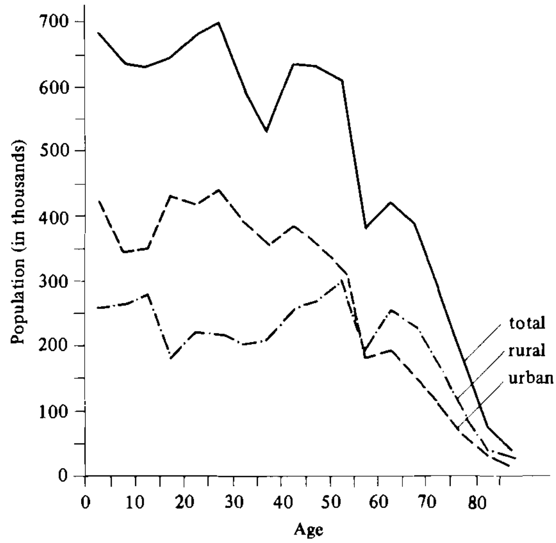


Figure 3. Age structures of the rural, urban, and total populations of Bulgaria at the end of 1975. The graph is plotted according to age groups of five-year span: 0-4, 5-9, 10-14, etc; each plot is made at the centre of each age group.

2 Preparation of the data

As mentioned earlier, since 1956 Bulgaria has been divided into twenty-eight administrative districts. They form the regional basis for the future planning of the development of the economy, and they form the smallest regional unit for which published demographic data are available. In order to simplify the multiregional analysis, it was found desirable to aggregate the twenty-eight districts into a more manageable number of regions, seven in all (figure 4).

Region 1. Northwestern Bulgaria (henceforth referred to as the N.West region), made up of four districts: Vidin, Vratsa, Mikhaylovgrad, and Sofia (district). However, the latter is to be distinguished from Sofia (city) which is an entirely different administrative district: Sofia (district) surrounds Sofia (city). Sofia (district) is included in this region because it has much the same demographic characteristics as the other three districts.

Region 2. Northern Bulgaria (the North region) made up of five administrative districts: Gabrovo, Veliko Tŭrnovo, Lovech, Pleven, and Ruse.

Region 3. Northeastern Bulgaria (the N.East region), consisting of Varna, Razgrad, Silistra, Tolbukhin, Tŭrgovishte, and Schumen.

Region 4. Southwestern Bulgaria (the S.West region), consisting of the Blagoevgrad, Kyustendil, and Pernik districts.

Region 5. Southern Bulgaria (the South region), made up of the Plovdiv, Kŭrdzhali, Pazardzhik, Smolyan, Stara Zagora, and Khaskovo districts.

Region 6. Southeastern Bulgaria (the S.East region), consisting of Burgas, Sliven, and Yambol.

Region 7. Sofia (city) (the Sofia region) forms a separate region because of its specific demographic significance. Sofia (city) has a population of about one million, and the total population of Bulgaria is about 8·5 million. It is obvious that the migration flow towards that district is very strong.

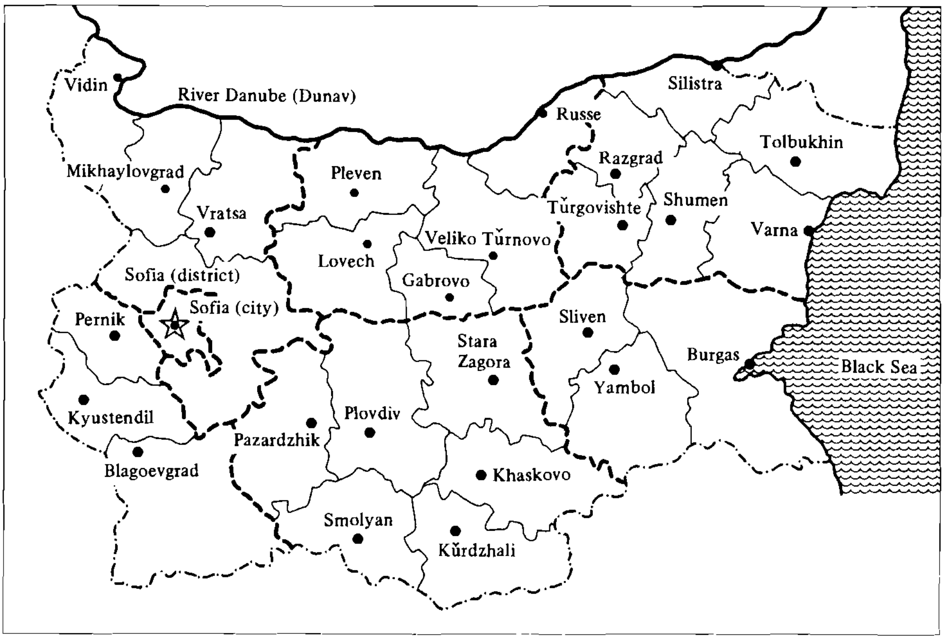


Figure 4. Map of Bulgaria showing the twenty-eight districts and the seven regions of the study (Central Bureau of Statistics, 1975).

The data for the population by age groups (total and for the twenty-eight districts) at the end of 1975, and the data for the departures and arrivals by age groups (total and for each district separately) during 1975, were gathered by the Central Bureau of Statistics in Sofia (personal communication). The data for the population at the end of 1974 were from the Central Bureau of Statistics (1975); the data for births and deaths, and the 28 x 28 migration-flow matrix, were from the Central Bureau of Statistics (1976).

Data on population were available for each of the twenty-eight districts by five-year age groups (the last one being 60+) for the end of the years 1974 and 1975, which yield the necessary midyear 1975 population data. For the analysis the population age structure was extrapolated up to 85+. This was done by following the age structure of the national population up to 100+. (Polynomial extrapolations were experimented with for different polynomial degrees but none of them were appropriate because of the low numbers in the 55–59 age group, due to the First World War and the preceding Balkan War.) It was supposed that the use of the national percentage distribution would not cause a large bias because the wars had affected uniformly the population throughout the country.

Regional data for births by age of mother were available by five-year age groups. The original data were not changed at all since they fitted exactly the needs of the analysis.

At the district level, data on deaths were available by five-year age groups up to twenty years of age, and by ten-year age groups up to seventy years and over. It was necessary to disaggregate each ten-year age group into two five-year age groups. This was done again by following the percentage distribution of the total deaths in the country. (Interpolation programs were also tried, but the results received were poor for the 50–59 age groups because the total number of deaths in the 50–54 age group exceeded the number of deaths in the 55–59 age group.)

Adjustment of the data on migration was most important to the analysis because the original data differed significantly from the input data used in the analysis. Available data for internal migration in Bulgaria (and in a number of European countries) are the departures and the arrivals for each district (given by five-year age groups), and the flow matrix (given only in total numbers) between districts. What is in fact necessary for the analysis is the flow matrix among the seven regions for each five-year age group. In the original data the total number of departures for each age group was usually less than the total number of arrivals for the same age group. This is due to the fact that migrants usually register upon arrival, but do not always register when leaving. Because of this, priority was given to the arrival data.

Table 5 presents the departures and the arrivals for each of the seven regions after the adjustments and after the removal of intradistrict moves; intraregional moves are included, however. This table was developed as follows. First, the departures and the arrivals for the twenty-eight districts were aggregated for the seven regions, and the departures were set equal to the arrivals. Second, the intradistrict moves were removed by following the proportions for each age group. This diminished the total number of moves for the whole nation (124105) to the interdistrict moves only (60782), but the numbers for the departures and the arrivals for each age group were not equal. For their equalization a two-dimensional RAS method was used (Raquillet and Willekens, 1978), with priority given to the arrivals.

Table 6 presents the flow matrix for the seven regions. It was developed by the aggregation of the 28 x 28 flow matrix, with intradistrict moves excluded but intraregional moves included (these appear along the main diagonal of table 6).

The data in table 5 were used to disaggregate the numbers in the flow matrix (table 6) into age groups. This was done by using a three-dimensional RAS method.

Table 5. Departures and arrivals for the seven regions of Bulgaria, 1975, intradistrict moves removed.

Age group	Departures from							Total departures
	N.West	North	N.East	S.West	South	S.East	Sofia	
0-4	614	669	699	244	1122	514	548	4410
5-9	417	715	644	217	875	436	305	3609
10-14	1125	1758	1660	1110	2640	1048	284	9625
15-19	2712	2827	3468	1729	5712	2352	479	19279
20-24	1306	1560	1591	634	2906	1144	757	9898
25-29	740	876	748	302	1429	548	817	5460
30-34	324	389	400	141	694	312	291	2551
35-39	177	225	244	119	457	176	320	1718
40-44	142	177	172	87	338	128	231	1275
45-49	101	119	109	55	225	90	170	869
50-54	72	84	80	34	144	53	109	576
55-59	42	49	48	21	84	28	64	336
60-64	44	55	59	23	87	30	59	357
65-69	42	44	42	23	82	27	57	317
70+	70	119	69	38	123	54	29	502
Total	7928	9666	10033	4777	16918	6940	4520	60782
Age group	**Arrivals at**							**Total arrivals**
	N.West	North	N.East	S.West	South	S.East	Sofia	
0-4	649	854	694	101	1119	333	660	4410
5-9	451	867	644	84	875	278	410	3609
10-14	1196	2130	1573	637	2783	716	590	9625
15-19	2026	3400	2918	731	5832	1135	3237	19279
20-24	917	1554	1489	185	2519	516	2718	9898
25-29	731	1009	800	129	1467	320	1004	5460
30-34	344	444	404	58	661	168	472	2551
35-39	241	252	272	37	446	116	354	1718
40-44	185	197	173	31	356	87	246	1275
45-49	131	109	115	19	251	66	178	869
50-54	98	82	71	8	166	42	109	576
55-59	43	46	51	4	93	16	83	336
60-64	33	48	53	4	84	12	123	357
65-69	21	47	36	4	80	12	117	317
70+	27	77	33	2	106	10	247	502
Total	7093	11116	9326	2034	16838	3827	10548	60782

Table 6. Flow matrix of the migrations among the seven regions of Bulgaria, 1975 (aggregated from the 28 x 28 flow matrix—Central Bureau of Statistics, 1976).

To	From							
	N.West	North	N.East	S.West	South	S.East	Sofia	Total
N.West	1896	1042	411	539	1261	271	1673	7093
North	1175	4152	2764	292	1427	559	747	11116
N.East	471	1524	4642	220	983	994	492	9326
S.West	268	146	122	823	298	67	310	2034
South	854	1107	759	813	9766	2500	1039	16838
S.East	110	249	502	103	919	1685	259	3827
Sofia	3154	1446	833	1987	2264	864	0	10548
Total	7928	9666	10033	4777	16918	6940	4520	60782

After extrapolating the numbers from the last age group (60+) up to the 85+ age group (in accordance with the percentage distribution of the arrivals for the whole state), the migration data were fitted for the analysis. Rounding-off errors were removed at each stage.

3 Multiregional life table

The concept of a life table is a basic one in demography. Such tables describe the evolution of a hypothetical cohort of babies born at a particular point in time. This evolution is expressed in a number of statistics: probabilities of dying and surviving, number of survivors, number of years to be lived, and expectations of life. The life table may be treated also as presenting a stationary population, one in which the number of births is equal to the number of deaths. This makes the life table a useful tool for the study of mortality.

The main difference between the single-region life table and the multiregional life table is that whereas the former is built for a single-region population, exposed to unchanging mortality and closed to migration, the latter focuses on several regions, and both mortality and migration schedules are accounted for. The region of residence is taken into consideration, giving the multiregional life table a spatial dimension.

In order to build a multiregional life table, one needs observed regional age-specific rates for dying and migrating. These can be computed by dividing the regional annual number of events for a given age group by the midyear population of the region in that age group. (The annual number of events in the Sofia region are presented in appendix 1.) Such input data are used to derive the basic parameters of the multiregional life table: the probabilities of dying and migrating. These are used to determine the number of survivors expected at exact age x in each region, the number of years lived in each region by the initial unit cohort, the survivorship proportions, and the life expectancies. For details regarding the construction of the multiregional life table from such input data, the reader may refer to Rogers (1975a) or Willekens and Rogers (1976).

Appendix 2 gives the observed regional age-specific rates for births and deaths for the whole of Bulgaria, and for out-migration from the Sofia region. Appendix 3 gives the part of the seven-region life table for Bulgaria that concerns the region of Sofia.

3.1 *Life history of a cohort of births*

Age-specific probabilities of dying and migrating are the basis for the construction of a multiregional life table. They permit the computation of the expected number of survivors, deaths, and migrations, for a given set of regional radices, that is, the hypothetical multiregional cohort. In this study, the radix for each region was set equal to 100000.

Of the 100000 babies that were born in the Sofia region (appendix 3.2), 97361 will be alive five years later; 94119 of them will have remained in the same region, 1283 will have moved to the N.West region, 559 to the North region, etc. After another five years the initial cohort will diminish to 97152, and only 91788 will have remained in Sofia, 2048 will be living in the N.West, etc.

The multiregional life table has a number of applications. Assume, for example, that the exact age 20 is the age of entering the labor force. Table 7 gives the probabilities that an individual born in a particular region will survive to exact age

Table 7. Probabilities of surviving to exact age 20 in the same region (from 1975 data).

Region	N.West	North	N.East	S.West	South	S.East	Sofia
Probability	0·744	0·814	0·826	0·765	0·860	0·754	0·868

20 and live in the same region. For Sofia it is derived by dividing the survivors at exact age 20 in Sofia, 86767, by the initial cohort of births of 100000. It may be noticed that this probability is high for the Sofia and South regions, and low for the N.West, S.East, and S.West regions. This suggests that the young population of the latter three regions tends to leave the region of birth before entering the labor force, whereas the natives of Sofia and the South prefer to take up employment in the same region.

3.2 *Expectations of life*

The concept of life expectancy is very important in the single-region life table, but it is perhaps even more important in the multiregional life table. Life expectancies at birth are presented in table 8. They reveal a number of interesting items of information. For example, again for the Sofia region, the life expectancy of a baby born in this region is 70·62 years, 59·49 of which will be lived in the same region, 3·80 years in the N.West, etc. For the South region, the life expectancy is 70·63 years, and 61·24 of them will be lived in the same region, a much higher proportion than in the N.West region, for example.

The expectation of life at birth in the multiregional life table is a good measure of the level of migration. The spatial migration level shall be given by ${}_i\theta_j = {}_ie_j(0)/{}_ie(0)$, where ${}_ie_j(0)$ is the number of years a person born in region i will be expected to live in region j, and ${}_ie(0)$ is the total life expectancy of a person born in region i. Thus the spatial migration level from region i to region j is the proportion of total life expectancy expected to be lived in region j by an individual born in region i (table 9).

The numbers on the main diagonal of table 9 represent the levels of 'nonmigration'. They are lowest for the N.West, S.West, and S.East regions, and highest for the South, N.East, and Sofia regions. Note that the Sofia region does not have the lowest out-migration level, contrary to what might be expected. However, its relatively high out-migration level is compensated for by an even higher in-migration level.

Table 8. Life expectancies at birth for the seven regions of Bulgaria, 1975.

Region of birth	Expected number of years spent living in region							Total life expectancy
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	52·98	3·83	1·70	0·87	2·97	0·43	8·62	71·40
North	2·32	58·57	3·46	0·38	2·69	0·60	3·15	71·17
N.East	0·92	5·19	59·37	0·26	1·69	0·92	1·74	70·09
S.West	2·27	1·45	1·06	54·98	3·44	0·47	7·22	70·89
South	1·64	1·98	1·38	0·40	61·24	1·11	2·88	70·63
S.East	1·06	2·08	3·18	0·26	7·71	53·43	2·80	70·52
Sofia	3·80	2·01	1·30	0·79	2·60	0·63	59·49	70·62

Table 9. Spatial migration levels for the seven regions of Bulgaria, 1975.

Region of birth	Proportion of life spent living in region							Total
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	0·742	0·054	0·023	0·012	0·042	0·006	0·121	1·000
North	0·033	0·823	0·049	0·005	0·038	0·008	0·044	1·000
N.East	0·013	0·074	0·847	0·004	0·024	0·013	0·025	1·000
S.West	0·032	0·020	0·015	0·775	0·049	0·007	0·102	1·000
South	0·023	0·028	0·019	0·006	0·867	0·016	0·041	1·000
S.East	0·015	0·030	0·045	0·004	0·108	0·758	0·040	1·000
Sofia	0·054	0·029	0·018	0·011	0·037	0·009	0·842	1·000

4 Population projection and stability

If a population that is closed to migration is exposed to an unchanging regime of fertility and mortality, it will reach a stable age structure which has a constant rate of natural increase through time. Achievement of stability has the property of 'forgetting' the initial age distribution, that is, the stabilization of a closed population is an ergodic process. The same is true when a multiregional population is *in addition* subjected to unchanging age-specific migration rates. This is the case of a multiregional population projected to stability. The theory on this subject can be found in Rogers (1975a).

Appendix 4 gives a population projection for the seven regions of Bulgaria to the year 2025 and the stable equivalent of the 1975 populations of the seven regions. In this projection, fertility, mortality, and the migration rate are kept constant at the level of 1975. The details of such computations are explained in Willekens and Rogers (1976). Table 10 presents some characteristics of the initial (1975) Bulgarian population, the projected population for 2025, and the stable equivalent population. The mean age, the regional share of the national population, and the (five-year) growth ratio (λ) for the stable equivalent refer also to the stable population which would be reached in the long run under the assumptions for the stability of the observed rates.

Table 10. Characteristics of the initial (1975) populations, the projected populations for 2025, equivalents of the 1975 populations for the seven regions of Bulgaria.

Variable	Population	Total	Region						
			N.West	North	N.East	S.West	South	S.East	Sofia
Total	initial	8727	1043	1400	1487	696	2164	867	1070
number	2025	10108	981	1492	1873	653	2718	881	1510
(thousands)	s.e.[a]	8748	741	1355	2123	247	2333	559	1389
Mean age	initial	35·18	38·97	37·87	33·81	34·19	33·60	34·33	34·37
	2025	36·55	37·81	37·68	35·19	38·14	35·90	36·37	36·88
	s.e.	36·42	37·37	37·46	35·51	36·36	36·09	35·51	37·23
Regional	initial	100·00	11·95	16·04	17·04	7·98	24·80	9·93	12·26
share of	2025	100·00	9·71	14·76	18·53	6·46	26·89	8·71	14·94
national	s.e.	100·00	8·47	15·49	24·26	2·83	26·67	6·39	15·88
population									
Growth	2025	1·0105	1·0006	1·0080	1·0184	0·9837	1·0154	0·9963	1·0212
ratio (λ)	s.e.	1·0119							

[a] s.e. ≡ stable equivalent

4.1 *Mean ages*

The numbers for the mean ages show that the projection brings greater uniformity among the seven regions: the difference between the highest and the lowest mean ages in 1975 is 5·37 years, in 2025 it is 2·95 years, and under stability it is 1·95 years. The greatest changes are to be observed in the S.West and Sofia regions. In the S.West region, over the fifty years of projection, the mean age will rise by four years because of the high level of out-migration. This will lead to the ageing of its population and a rise in the CDRs. In Sofia the mean age will rise continuously because of delayed childbearing and in-migration by older age groups. The older populations of the N.West and North regions yield high CDRs, and the mean age will drop a little. The mean ages in the remaining three regions rise together with the rise in the total populations.

4.2 *Regional shares*

Changes in the regional shares are not very large over the fifty-year projection period. They increase for the N.East (to end in 2025 with a rise of 1·5%), the South (2·09%), and Sofia (2·7%), and decrease for the remaining four regions. The largest decrease, 2·24%, is exhibited in the N.West.

Under stability the regional shares are strikingly different for the N.East and S.West regions. The high fertility level and the low in- and out-migration flows for the N.East contribute to the great increase in its regional share in the long run. (Although over the fifty-year projection period this increase is quite small, it increases to nearly 25% of the total at stability.) At the other extreme is the S.West, whose strong out-migration leads to a diminished share in the long run, down to less than 3% of the total.

The increase in the regional share for Sofia is a very important one. Sofia has a fertility below reproduction level, but in-migration leads to an increase in the population.

4.3 *Growth ratio*

In 2025 the growth ratio, λ, is less than one for the S.West and S.East regions, which means that their populations would decrease between 2020 and 2025. The population projection for the regions shows that, during the fifty-year period, the intrinsic rate is usually below zero for the N.West (high mortality and high out-migration) and S.West (high out-migration) regions.

The growth ratio for the stable population is 1·0119. It is the greatest positive eigenvalue of the growth matrix and can be used to derive the spatial intrinsic growth rate, $r = \frac{1}{5}\ln\lambda = 2\cdot37 \times 10^{-3}$, that is, an annual growth of 2·37 persons per thousand of the population. The value of the growth ratio for the stable population is quite close to the growth ratio for the total population in 2025. However, this is not the case for each individual region. In 2025 each region appears to be far from stable because of the peculiarities already mentioned: different levels of fertility, young or aged population structures, and large differences in the migration flows.

4.4 *Stable equivalent of the observed population*

The stable equivalent is the population that, if distributed as the stable population and growing at the stable ratio λ, would in the long run yield the same result as the observed population under projection (Willekens and Rogers, 1976; 1977). This means that the major difference between the stable equivalent population and the observed population is that the effect of the age structure is removed from the growth of the former.

Figure 5 shows the age distributions of the observed and of the stable equivalent populations of Sofia. Note that the stable equivalent population is larger than the observed population (as is shown by the larger area under the former curve), and that the dips for the 30–39 and 55–59 age groups are missing in the stable equivalent population. The curve for Sofia's stable equivalent population is not like the one that would have been obtained by a single-region analysis. There is a peak for the 25–29 age group which is due to the strong in-migration flow to Sofia by the younger age groups, especially by migrants aged 20–29.

From what has been said about the multiregional population projection and its stable growth, the following inferences can be made.

(1) The populations of the seven regions of Bulgarian are very far from stability, because of the different levels of fertility, the differences in the age structure (which cause higher or lower mortality), and the differences in the migration flows.

(2) The national population is tending to concentrate in the N.East and South regions, and to leave the N.West, the S.East, and especially the S.West regions.

(3) Despite a low fertility level, the population of Sofia has a high growth rate because of the high in-migration flow.
(4) During the next fifty years the regional share will decrease for the N.West and North regions, and increase in the South and Sofia regions. Since the mean ages for the last two regions also increase, it can be inferred that the labor force will also increase.
(5) During the next fifty years the mean age in the S.West region will increase strongly, that is, the population will be ageing rapidly. (In 1975 it was one of the youngest.)

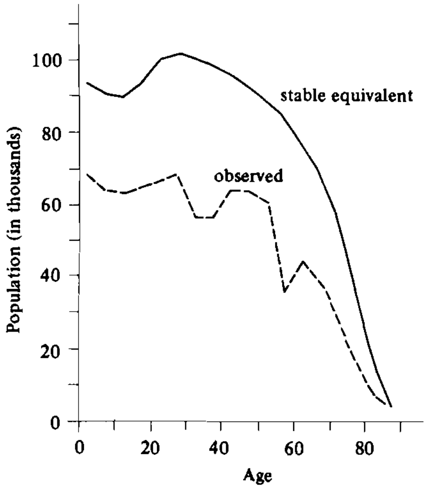


Figure 5. Number of people in each age group for the observed and the stable equivalent populations of the Sofia region, 1975. The graph is plotted as in figure 3.

5 Analysis of spatial fertility and migration patterns

Sections 3 and 4 show clearly that the multiregional life table and the population projection depend strongly on observed population characteristics. They can produce some very useful indicators for the study of the spatial demographic behavior of the population, an example of which is the spatial migration level. In this section observed fertility, mortality, and migration rates, as well as the indicators yielded from the analysis, are discussed.

5.1 *Analysis of observed population characteristics*

The study of mean ages makes it possible to follow the effect of the age composition on observed rates. Table 11 gives the mean ages of observed population characteristics, computed by the following formula:

$$\overline{m}_i = \tfrac{1}{100} \sum_x (x+2\cdot 5)c_i(x) \,, \qquad (1)$$

where $c_i(x)$ is the percentage distribution in region i at age x (where x ranges over the initial years of the five-year and 85+ age groups, $x = 0, 5, 10, ..., 85$), and $i = 1, 2, ..., 7$. The mean age therefore depends on the age structure of the population.

It can be seen that, reflecting their older age structures, the mean ages of the populations of the N.West and North regions are much higher than those of the other five regions. The mean ages of dying are also higher for the same two regions, reflecting again the older age structures of their populations.

The population age structure during the reproductive ages is similar in all the different Bulgarian regions, which is why the mean age of childbearing is at the same level in six regions. In the Sofia region it is a little higher because almost the whole population of this region is urban.

The right-hand part of table 11 is a 'flow matrix' of the mean ages of out-migrants. The mean ages are highest for departures from and arrivals in the Sofia region. The lowest mean ages for departures and arrivals can be observed for the S.West region.

In Bulgaria, students finishing their primary education can choose to continue with their obligatory secondary education in a number of specialized professional schools. In order to do this they often have to change their place of settlement. This explains the comparatively low mean ages of the migrants and the high number of moves in the 10–14 and 15–19 age groups.

Appendix 2 gives the observed age-specific, gross, and crude birth- and deathrates for the seven regions, and the out-migration rates for Sofia. The mean ages here are computed with the formula

$$\overline{m}_i = \frac{\sum_x (x+2\cdot 5) f_i(x)}{\sum_x f_i(x)} , \qquad (2)$$

where $f_i(x)$ is the age-specific rate for region i and x ranges over the initial years of the five-year and 85+ age groups. These mean ages are shown in table 12.

Formula (2) gives the mean ages with the effect of the age structure eliminated, that is, the mean ages of the schedules. The mean ages computed with formula (1) are denoted by $m_i(1)$, and those calculated with formula (2) by $m_i(2)$. A comparison of $m_i(1)$ with $m_i(2)$ reveals the effects of age composition. For example, when $m_i(2)$ is much greater than $m_i(1)$, the age structure is very young. This is most convenient for

Table 11. Mean ages of the populations, of giving birth, of death, and of out-migration for the seven regions of Bulgaria, 1975 [calculated with formula (1)].

Region	Population	Birth	Death	Out-migration to						
				N.West	North	N.East	S.West	South	S.East	Sofia
N.West	38·97	24·04	69·04	-	18·21	18·75	17·05	19·23	17·91	23·07
North	37·87	24·22	69·00	19·20	-	19·52	17·12	20·08	19·09	25·30
N.East	33·81	24·32	65·32	19·11	18·95	-	17·25	19·96	18·98	24·50
S.West	34·19	24·37	66·15	17·73	17·55	17·71	-	18·25	17·21	22·50
South	33·60	24·17	65·76	19·35	19·08	19·65	17·50	-	19·20	24·29
S.East	34·33	24·23	65·73	18·98	18·90	19·36	17·19	19·80	-	24·08
Sofia	34·37	25·35	64·71	24·93	24·49	25·30	22·68	26·53	25·15	-

Table 12. Mean ages of giving birth, of death, and of out-migration for the seven regions of Bulgaria, 1975 [calculated with formula (2)].

Region	Birth	Death	Out-migration to						
			N.West	North	N.East	S.West	South	S.East	Sofia
N.West	24·06	77·89	-	18·65	18·96	17·12	19·72	17·93	24·08
North	24·25	78·44	19·46	-	19·74	16·89	20·69	19·06	28·67
N.East	24·43	78·72	20·56	20·89	-	17·47	22·18	19·94	30·65
S.West	24·60	79·11	18·60	18·61	18·41	-	20·05	17·69	26·33
South	24·45	79·18	21·04	20·87	21·55	17·81	-	20·27	29·83
S.East	24·36	78·88	19·84	21·05	20·87	17·39	21·69	-	29·29
Sofia	25·44	80·04	26·73	27·20	27·04	22·93	30·46	26·13	-

analyses of the mortality schedules. For example, it can be inferred that the N.West region has a slightly higher mortality level and the Sofia region a slightly lower one than previously indicated.

For fertility data $m_i(1)$ and $m_i(2)$ are almost the same. For migrations, however, $m_i(2)$ is much higher than $m_i(1)$ when the Sofia region is considered. (There is almost no difference between the $m_i(1)$ and $m_i(2)$ when $i = 1$ or 2, that is, for the N.West and the North regions, because the age structures of their populations are very old.) The highest mean ages are for the migration schedules of the migrants to the Sofia region, and the lowest for those to the S.West region. The population of Sofia is young, and the mean ages of the fertility and the migration schedules are high. The reason for the higher mean age of childbearing in Sofia is delayed childbearing: its age-specific fertility for the 30–49 age groups are the largest among all the regions. The mean ages of migrations to Sofia are the highest in Bulgaria because of the large number of movers in the age groups over 20—moves caused by factors such as a change of job, students moving to secondary schools, etc.

A comparison of the crude rates (appendix 2) among the regions shows some of the features that have been outlined above: high CDRs and low CBRs in the N.West and North regions, reflecting their old age structure; and a low CDR in the Sofia region, reflecting its comparatively young age structure. The gross death rate (GDR—the sum of the age-specific death rates) for Sofia is very large, however, because of the higher age-specific death rates for the ages above 70. The GDRs are very low for the N.West and North regions because of the low age-specific death rates for the older part of the population.

The gross birthrate (GBR) is the sum of the age-specific birthrates. When multiplied by five (the width of the age groups) it gives the gross reproduction rate (GRR). When $GRR \geqslant 1$, each person (in a one-sex population) will reproduce if there is no death until the end of the last age of reproduction (50 years). When mortality is accounted for, GRR has to be greater than 1·05 (approximately) in order to ensure replacement. It is evident that the GBR (and hence, fertility) is below the replacement level in the North region (GRR = 1·01) and in Sofia (GRR = 0·96), and very high in the N.East and S.East regions (that is, in Eastern Bulgaria). The GBR for the total national population is equal to 0·22 (GRR = 1·1) which shows that the lower fertility in the North and Sofia regions is compensated for nationally by the other five regions.

The gross migration rates are the sum of the age-specific migration rates. When this sum is multiplied by five (as with the GBR), one obtains the gross migraproduction rate (GMR). These rates are presented in table 13. The GRR is a measure of fertility; in the same way the GMR is a measure of out-migration. The GRR shows how many babies each person will produce during the reproduction age period, assuming zero mortality, and the GMR shows the number of out-migrations per

Table 13. Gross migraproduction rates for the seven regions of Bulgaria, 1975.

Region	Out-migration to							Total
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	-	0·089	0·036	0·020	0·065	0·008	0·244	0·461
North	0·056	-	0·082	0·008	0·059	0·013	0·080	0·298
N.East	0·018	0·125	-	0·005	0·035	0·022	0·042	0·248
S.West	0·050	0·027	0·020	-	0·077	0·009	0·201	0·385
South	0·038	0·043	0·030	0·009	-	0·027	0·075	0·221
S.East	0·021	0·044	0·078	0·005	0·198	-	0·074	0·420
Sofia	0·106	0·048	0·031	0·020	0·068	0·016	-	0·290

person if there is no death during the entire life span. Both measures are cross-sectional; they show the levels of fertility and migration in 1975.

The last column of table 13 refers to the GMR of the total population of each one of the seven regions. The highest number of out-migrations per person is expected for the N.West region, which is the main reason for its aged population. The lowest average number of out-migrations is shown for the South region. The other columns give a clear picture of migration flows among the seven regions. $(\mathrm{GMR})_{ij}$ shall denote the GMR for the moves from region i to region j. The GMR which occupy symmetrical positions to the main diagonal—$(\mathrm{GMR})_{ij}$ and $(\mathrm{GMR})_{ji}$—can be compared to show the difference between the two opposite migration flows. For example,

(1) $(\mathrm{GMR})_{4j} > (\mathrm{GMR})_{j4}$, for all j—each migration flow from the S.West region to any other region is stronger than the corresponding counterflow. This shows that the S.West region is the most unattractive one when the rate and not the magnitude of the flow is considered.

(2) $(\mathrm{GMR})_{7j} < (\mathrm{GMR})_{j7}$, for all j—the migration flow from each region to the Sofia region is stronger than the counterflow. The same is true for the South region, except that $(\mathrm{GMR})_{57}$ is slightly greater than $(\mathrm{GMR})_{75}$. These two regions appear to be the most attractive according to this measure.

This type of analysis makes it possible to rank the regions according to their attractiveness, if the magnitude of the migration flows is not considered. (The average number of out-migrations per person has to be distinguished from the total number of migrations.) The most attractive regions are Sofia and the South; fairly attractive are the North and N.East; and the unattractive regions are the N.West, S.East, and S.West.

5.2 *Spatial reproduction and migraproduction levels*

Until now the study of fertility has been carried out on the basis of the age-specific and gross birthrates. The multiregional life table makes it possible to analyze spatial fertility levels with more refined measures such as spatial net reproduction allocations.

The spatial net reproduction rate (NRR) is defined by (Rogers, 1975b)

$$_i(\mathrm{NRR})_j = \sum_{x=0}^{z} {_iL_j(x)} F_j(x) \,,$$

where $_iL_j(x)$ is the number of persons from the multiregional-life-table population in region j that were born in region i and who are aged between x and $x+4$ (Rogers, 1975a), $F_j(x)$ is the fertility level for persons aged between x and $x+4$ in region j, x ranges over the initial years of the five-year age groups, and z is the initial year of the final fertile age group. The net-reproduction matrix is presented in table 14. The totals in the table refer to the regional net reproduction rates. These are not the conventional regional NRRs though, as these include the impact of migration. As in

Table 14. Spatial net reproduction rates for the seven regions of Bulgaria, 1975.

Region of birth	NRR for time spent living in region							Total
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	0·778	0·054	0·028	0·015	0·046	0·007	0·118	1·045
North	0·034	0·798	0·058	0·006	0·042	0·010	0·041	0·990
N.East	0·013	0·074	0·972	0·004	0·026	0·016	0·023	1·127
S.West	0·034	0·020	0·018	0·820	0·056	0·008	0·099	1·054
South	0·023	0·027	0·022	0·006	0·945	0·018	0·037	1·078
S.East	0·015	0·028	0·053	0·004	0·123	0·881	0·036	1·140
Sofia	0·044	0·023	0·017	0·011	0·032	0·008	0·804	0·938

the conventional NRRs, the totals in the table represent the number of babies to be born per person from a particular region during the individual's life span when mortality is accounted for. It can be seen that the NRRs for the North and Sofia regions are less than 1·00, that is, below replacement level. This shows the low fertility level in these two regions. The NRRs are very high for the N.East and S.East regions, reflecting the high fertility level in Eastern Bulgaria. These same inferences were evident when gross birthrates were examined.

The rows of the net reproduction matrix give the total regional NRR, distributed over the seven regions. For instance, a person who is born in the N.West region will give birth to 1·045 babies; 0·778 of the births will be in the same region, 0·118 in the Sofia region, etc.

The allocations of the net reproduction levels are presented in table 15. The spatial net reproduction allocation is

$$ {}_i\rho_j = \frac{{}_i(\mathrm{NRR})_j}{{}_i(\mathrm{NRR})} , $$

where ${}_i(\mathrm{NRR}) = \sum_j {}_i(\mathrm{NRR})_j$. Tables like table 15 were presented for $(\mathrm{GMR})_{ij}$ (table 13) and for the level of migration, ${}_i\theta_j$ (table 9). Its study can proceed in the same manner. First, the numbers along the main diagonal are high for the South, N.East, and Sofia regions, low for the S.West and S.East regions, and especially low for the N.West region. This shows that a person who was born in one of the first three regions mentioned is more likely to give birth in the same region than a person born in one of the remaining four regions. A person from the N.West region would experience only 74·4% of total lifetime births in the same region, which shows its unattractiveness for childbearing for its natives.

A comparison of the two elements symmetrical to the main diagonal shows the preference of a parent between the two appropriate regions. For instance, a person from the N.West region will make 11·3% of lifetime births in the Sofia region, whereas a person from Sofia would prefer to make only 4·6% of births in the N.West. In this way the regions can be compared for preference of childbearing. Thus the S.West region 'loses' to any other region, whereas the South region 'gains' from all the regions. When only allocation is considered, the most attractive regions for childbearing are the South and the Sofia regions, fairly attractive are the North, and N.East, and unattractive are the S.East, N.West, and S.West.

Notice that exactly the same inferences were made when the gross migraproduction rates were analyzed. This shows that the patterns of preference of childbearing follow exactly the patterns of migration among the regions, so that the allocations of births are only a result of migration. There exists no clear trend of movement which could be closely connected with childbearing.

Table 15. Net reproduction allocations for the seven regions of Bulgaria, 1975.

Region of birth	Net reproduction allocation for time spent living in region							Total
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	0·744	0·052	0·027	0·014	0·044	0·007	0·113	1·000
North	0·035	0·806	0·059	0·006	0·043	0·010	0·042	1·000
N.East	0·012	0·065	0·862	0·004	0·023	0·014	0·020	1·000
S.West	0·032	0·019	0·017	0·778	0·053	0·007	0·094	1·000
South	0·021	0·025	0·020	0·006	0·876	0·017	0·034	1·000
S.East	0·013	0·025	0·046	0·004	0·108	0·772	0·032	1·000
Sofia	0·046	0·024	0·018	0·012	0·034	0·009	0·857	1·000

The impact of migration on fertility comes from the 15–49 age interval, which is the reproductive period (with a few exceptions for ages below 15). Recall that the mean age of childbearing is 24–25 years of age, and that the largest age-specific birthrates are those in the 20–29 age interval. The impact of migration on fertility comes mainly during these ages. Since there is no special trend for moves on the occasion of childbearing (that is, the migrations in the 20–29 age interval follow one and the same schedule among the regions), the differences in the magnitudes of the fertility schedules follow the differences in the magnitudes of the migration schedules.

The allocation of a regional life expectancy (${}_i\theta_j$) is a measure of the *duration*—the number of years to be lived in a particular region. But migration, like childbearing, is also a *recurrent event*, in that one person may migrate several times during a lifetime. A measure of the recurrence of migration can be derived in a way similar to the NRR. This is the net migraproduction rate, which is computed by the formula (Rogers, 1975a)

$$ {}_i(\mathrm{NMR})_j = \sum_{x=0}^{z} {}_iL_j(x)M_j(x) \, , $$

where ${}_iL_j(x)$ is the stationary life-table population aged between x and $x+4$, living in region j and born in region i; $M_j(x)$ is the out-migration rate in region j for persons aged between x and $x+4$; x ranges over the initial years of the five-year age groups; and z is the initial year of the final age group.

The net migraproduction matrix for the Bulgarian regions is presented in table 16. The totals in the last column show the total number of migrations per person from a given region during his lifetime. The totals indicate the magnitude of out-migration per person from the regional stationary population with the effect of mortality included. The highest number of moves is to be expected from the N.West region, and the lowest from the South and N.East regions. The net migraproduction allocations presented in table 17 define each region's share of the total net migraproduction rates.

Table 16. Spatial net migraproduction rates for the seven regions of Bulgaria, 1975.

Region of birth	NMR for time spent living in region							Total
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	0·355	0·011	0·004	0·004	0·007	0·002	0·027	0·409
North	0·011	0·245	0·009	0·002	0·006	0·003	0·010	0·285
N.East	0·004	0·015	0·206	0·001	0·004	0·004	0·005	0·239
S.West	0·011	0·004	0·003	0·302	0·008	0·002	0·023	0·352
South	0·007	0·006	0·003	0·002	0·186	0·005	0·009	0·217
S.East	0·005	0·006	0·008	0·001	0·018	0·323	0·009	0·369
Sofia	0·015	0·005	0·003	0·003	0·005	0·002	0·237	0·271

Table 17. Net migraproduction allocations for the seven regions of Bulgaria, 1975.

Region of birth	Net migraproduction allocation for time spent living in region							Total
	N.West	North	N.East	S.West	South	S.East	Sofia	
N.West	0·867	0·028	0·010	0·009	0·016	0·004	0·066	1·000
North	0·039	0·860	0·030	0·005	0·022	0·009	0·034	1·000
N.East	0·017	0·064	0·861	0·004	0·015	0·017	0·022	1·000
S.West	0·031	0·012	0·007	0·857	0·023	0·006	0·064	1·000
South	0·034	0·026	0·015	0·007	0·857	0·022	0·040	1·000
S.East	0·012	0·016	0·021	0·003	0·048	0·877	0·023	1·000
Sofia	0·057	0·019	0·010	0·011	0·019	0·009	0·875	1·000

The numbers along the main diagonal of table 17 represent the preference for staying in the region of birth. For instance, a person from the Sofia region will make 5·7% of his moves to N.West, 1·9% to North, etc, and 87·5% of the time will remain in the Sofia region, a percentage which is second only to that of the S.East region. At the other extreme are the S.West and South regions, in which a person will remain 85·7% of the time.

The numbers in the positions symmetrical to the main diagonal give a picture of the preference of moving, when the magnitude of the moves is not taken into consideration. For example, a person from the N.West region will make 2·8% of his moves to the North region, whereas a person from the North region will make 3·9% of his moves to the N.West region. This comparison is most unfavorable for the S.West region, which is the most unattractive one. It is followed in unattractiveness by the S.East and N.East regions. The most attractive regions are Sofia and the N.West. This comparison ranks the N.West region as an attractive one, although its net migraproduction rate is the highest in the country. This can be explained by the high number of moves from the N.West to Sofia. When the gross migraproduction rates were compared, the N.West region was unattractive because the effect of mortality was excluded.

In the study of GMRs, NMRs, and the spatial life expectancy, the N.East region exhibits a low migration level and unattractiveness, so it can be delineated as a region that stays outside the general characteristics of migration movements among the Bulgarian population. This can be explained by its historical development as an agricultural region. The Dobrudzha area is situated here, and is known as the 'granary' of Bulgaria. Industry is developed mainly in the Varna district, and it is quite possible that, if this district were not included in the region, the migration levels would be much lower.

A comparison can be made between the NMR and θ. Recall that the NMR is the average number of migrations per person who is subject to death during his lifetime, that is, it is a measure of mobility; and θ is a measure of the level of migration in terms of duration or time spent. If for region i, the NMR is low and ${}_i\theta_i$ is high, then the migration movements out of this region are low. This is the case for the South and N.East regions (and also for Sofia). On the other hand the N.West region has the largest out-migration. When a comparison is made for the ${}_i(\mathrm{NMR})_j$ and ${}_i\theta_j$, migrational counterflows can be taken into consideration. Then the most unattractive, both in magnitude and direction of the migration flows, is the S.West region, and the most attractive are Sofia and the South. The N.East region is connected with very low in- and out-migration flows, but it is rather an unattractive region.

The patterns exhibited here can be explained in the following way. Sofia is the largest city in Bulgaria, with a population of about one million. It is a highly urbanized area and a center of national, social, and cultural life. The largest high schools are situated there. The Southern region is attractive mainly because of the city of Plovdiv, and because of the government planning for fast industrialization and social development in the southern Rodopi regions. The N.West and S.West regions exhibit a high level of out-migration because of their low industrial development. New industrialization in these regions has been planned and has already begun (for instance, a large atomic power station is under construction on the Danube in the N.West), but its effect in 1975 was still small. The comparatively high level of migration from the S.East region to the N.East region is due mainly to the city of Varna, where the maritime industry is developed, and to the seaside resorts around it. Migration levels are low for the North and N.East regions, reflecting their historically important agricultural role in the economic development of Bulgaria.

6 Conclusion

Multiregional population analysis has many advantages over single-region analysis. It allows for the study of the spatial distribution of the population by taking into consideration migration movements and their impact on regional fertility, mortality, and age composition. The Sofia region is a good example. The fertility there is below replacement level, showing a population decline in a single-region projection. In the multiregional projection, however, the population of Sofia is growing because of strong net in-migration.

The multiregional analysis generates a large number of population characteristics, some of which are examined in this paper. Together they give views of the demographic development of each region. These may be summarized as follows.

(1) The most unfavorable demographic picture is exhibited for the N.West, S.West, and S.East regions. They are losing population because of strong out-migration flows. This should lead to a fast ageing of the S.West population; the population of N.West is already aged. As for 1975, the governmental regional policy for uniform development seems not to have been very effective for these three regions. In light of the new government policy for faster industrialization of these regions, adopted during the last few years, it might be expected that the out-migration flows will lessen in the short run. This calls for a careful study of the migration flows during the years following 1975; especially of the flows towards Sofia from the N.West and the S.West regions, and towards the South and N.East from the S.East region.

(2) The North region exhibits a low fertility level (the pronatalist policy is not very effective), but to some extent this is compensated for by the in-migration flow. It also has an aged population structure, and if the in-migration flows lessen, the impact of the low fertility level will strengthen. The out-migration flow is very strong towards the South region.

(3) The N.East region exhibits a high fertility level and low migration flows. In the short run it might be expected to show slightly diminishing fertility.

(4) The Sofia region exhibits high in- and low out-migration flows. Fertility is very low. This shows that neither the pronatalist nor the migration policies (in-migrations to Sofia are discouraged by restrictions) are effective for Sofia. In the short run, major changes in fertility and migration are not expected.

It is evident from the above statements that a multiregional analysis makes it possible to direct more precisely the national population policy of a country.

Acknowledgement. The author gratefully acknowledges the assistance and advice of Frans Willekens.

References

Central Bureau of Statistics, 1972, 1973, 1974, 1975, 1976 *Demographic Yearbook of Bulgaria 1972, 1973, 1974, 1975, 1976* (Central Bureau of Statistics, Sofia, Bulgaria)

Naoumov N, Stefanov I, Sougarev Z, 1974 *La Population de la Bulgarie* CICRED series, Comité International de Coordination des Recherches Nationales en Demographie, Sofia

Raquillet R, Willekens F, 1978 "Entropy and multiproportional techniques for inferring detailed migration patterns from aggregate data" RM-78-00, International Institute for Applied Systems Analysis, Laxenburg, Austria

Rogers A, 1975a *Introduction to Multiregional Mathematical Demography* (John Wiley, New York)

Rogers A, 1975b "Spatial migration expectancies" RM-75-57, International Institute for Applied Systems Analysis, Laxenburg, Austria

Stefanov I, Sougarev Z, Naoumov N, Christov E, Atanosov A, 1974 *Demografia na Bulgaria* (Nauka i Izkustvo, Sofia) in Bulgarian

Willekens F, Rogers A, 1976 "Computer programs for spatial demographic analysis" RM-76-58, International Institute for Applied Systems Analysis, Laxenburg, Austria

Willekens F, Rogers A, 1977 "More computer programs for spatial demographic analysis" RM-77-30, International Institute for Applied Systems Analysis, Laxenburg, Austria

APPENDIX

1 Observed demographic data for the Sofia region, 1975

'Age' refers to the age group whose initial year is given.

AGE	POPULATION	BIRTHS	DEATHS	MIGRATION FROM SOFIA TO N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	80672.	0.	431.	219.	94.	58.	32.	111.	35.	0.
5	65741.	0.	28.	113.	60.	35.	19.	59.	19.	0.
10	58883.	15.	25.	100.	47.	26.	38.	57.	16.	0.
15	75198.	1986.	42.	172.	74.	45.	51.	114.	23.	0.
20	104390.	7812.	73.	265.	129.	91.	50.	182.	40.	0.
25	105771.	5762.	94.	299.	142.	90.	51.	190.	45.	0.
30	83074.	2132.	107.	117.	47.	32.	19.	61.	16.	0.
35	72132.	655.	105.	122.	48.	40.	17.	73.	21.	0.
40	78424.	126.	205.	89.	34.	23.	14.	56.	15.	0.
45	85100.	10.	339.	67.	21.	17.	9.	44.	12.	0.
50	81807.	0.	541.	47.	14.	9.	4.	28.	8.	0.
55	43266.	0.	455.	23.	9.	8.	2.	18.	3.	0.
60	53297.	0.	930.	19.	10.	8.	2.	17.	3.	0.
65	46147.	0.	966.	13.	12.	7.	2.	19.	3.	0.
70	17787.	0.	1278.	4.	3.	1.	1.	4.	1.	0.
75	10590.	0.	1287.	3.	1.	1.	0.	3.	0.	0.
80	4678.	0.	938.	1.	1.	0.	0.	2.	0.	0.
85	3018.	0.	927.	1.	1.	0.	0.	1.	0.	0.
TOTAL	1069975.	18498.	8771.	1674.	747.	491.	311.	1039.	260.	0.

2 Observed demographic rates for Bulgaria, 1975

In this section 'age' refers to the age group of five-year span whose initial year is given (except for 85 which refers to the 85+ age group). The rates are measured in thousandths rather than per thousand. The 'gross' figures are merely the sums of the respective columns. The 'crude' figures are the respective rates for the whole population of each region. The mean ages (denoted by M.AGE) are computed with formula (2).

2.1 *Observed deathrates for all of the regions*

AGE	N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	0.005287	0.005620	0.006528	0.005780	0.005696	0.006263	0.005343
5	0.000509	0.000587	0.000553	0.000440	0.000529	0.000352	0.000426
10	0.000385	0.000414	0.000366	0.000262	0.000531	0.000553	0.000425
15	0.000759	0.000592	0.000831	0.000712	0.000695	0.000765	0.000559
20	0.001142	0.000825	0.000997	0.001057	0.000870	0.000948	0.000699
25	0.001224	0.000962	0.001190	0.001273	0.001032	0.001122	0.000889
30	0.001281	0.001279	0.001287	0.001287	0.001287	0.001281	0.001288
35	0.001904	0.001802	0.002068	0.002020	0.001650	0.001823	0.001456
40	0.002616	0.002614	0.002610	0.002613	0.002611	0.002610	0.002614
45	0.004886	0.004235	0.004816	0.004025	0.003926	0.003790	0.003984
50	0.006603	0.006612	0.006608	0.006612	0.006612	0.006602	0.006613
55	0.009923	0.009548	0.011865	0.009873	0.010086	0.010048	0.010516
60	0.017458	0.017456	0.017458	0.017446	0.017459	0.017452	0.017449
65	0.033064	0.033521	0.033730	0.022848	0.029679	0.031472	0.020933
70	0.041529	0.046998	0.055309	0.054799	0.057664	0.054393	0.071850
75	0.070287	0.079505	0.093582	0.092721	0.097587	0.092053	0.121530
80	0.115978	0.131173	0.154431	0.152889	0.160989	0.151847	0.200513
85	0.177564	0.200872	0.236555	0.234248	0.246518	0.232468	0.307157
GROSS	0.492400	0.544613	0.630783	0.610905	0.645422	0.615843	0.774243
CRUDE	0.012850	0.012421	0.010122	0.009488	0.009361	0.009771	0.008197
M.AGE	77.8922	78.4360	78.7156	79.1095	79.1781	78.8792	80.0398

2.2 *Observed birthrates for all of the regions*

AGE	N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000178	0.000310	0.000522	0.000209	0.000429	0.000553	0.000255
15	0.039271	0.034266	0.040539	0.035061	0.035736	0.043050	0.026410
20	0.103274	0.093125	0.105791	0.100325	0.103499	0.105981	0.074835
25	0.052698	0.049910	0.063078	0.056285	0.058185	0.063342	0.054476
30	0.017337	0.019128	0.022896	0.021361	0.020132	0.021867	0.025664
35	0.005135	0.004494	0.006224	0.007242	0.006452	0.006908	0.009081
40	0.001169	0.000917	0.001515	0.001780	0.001558	0.001581	0.001607
45	0.000100	0.000069	0.000158	0.000166	0.000105	0.000090	0.000118
50	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
55	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
60	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
65	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
70	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
80	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
85	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
GROSS	0.219162	0.202219	0.240724	0.222429	0.226096	0.243373	0.192445
CRUDE	0.013594	0.014350	0.018504	0.016249	0.017297	0.017977	0.017288
M.AGE	24.0588	24.2499	24.4277	24.5951	24.4453	24.3576	25.4395

2.3 *Observed migration rates from the Sofia region*

		MIGRATION FROM SOFIA TO						
AGE	TOTAL	N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	0.006685	0.002715	0.001165	0.000719	0.000397	0.001376	0.000434	0.000000
5	0.004639	0.001719	0.000913	0.000532	0.000289	0.000897	0.000289	0.000000
10	0.004823	0.001698	0.000798	0.000442	0.000645	0.000968	0.000272	0.000000
15	0.006370	0.002287	0.000984	0.000598	0.000678	0.001516	0.000306	0.000000
20	0.007252	0.002539	0.001236	0.000872	0.000479	0.001743	0.000383	0.000000
25	0.007724	0.002827	0.001343	0.000851	0.000482	0.001796	0.000425	0.000000
30	0.003515	0.001408	0.000566	0.000385	0.000229	0.000734	0.000193	0.000000
35	0.004450	0.001691	0.000665	0.000555	0.000236	0.001012	0.000291	0.000000
40	0.002946	0.001135	0.000434	0.000293	0.000179	0.000714	0.000191	0.000000
45	0.001998	0.000787	0.000247	0.000200	0.000106	0.000517	0.000141	0.000000
50	0.001345	0.000575	0.000171	0.000110	0.000049	0.000342	0.000098	0.000000
55	0.001456	0.000532	0.000208	0.000185	0.000046	0.000416	0.000069	0.000000
60	0.001107	0.000356	0.000188	0.000150	0.000038	0.000319	0.000056	0.000000
65	0.001214	0.000282	0.000260	0.000152	0.000043	0.000412	0.000065	0.000000
70	0.000787	0.000225	0.000169	0.000056	0.000056	0.000225	0.000056	0.000000
75	0.000755	0.000283	0.000094	0.000094	0.000000	0.000283	0.000000	0.000000
80	0.000855	0.000214	0.000214	0.000000	0.000000	0.000428	0.000000	0.000000
85	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
GROSS	0.058041	0.021273	0.009654	0.006194	0.003951	0.013699	0.003270	0.000000
CRUDE	0.004226	0.001565	0.000698	0.000459	0.000291	0.000971	0.000243	0.000000
M.AGE	27.4275	26.7261	27.1997	27.0395	22.9256	30.4612	26.1274	0.0000

3 Multiregional life table for the Sofia region, 1975

3.1 *Probabilities of dying and migrating*

'Age' in the first column means the age group of five-year span whose initial year is given (except for 85 which refers to the 85+ age group).

AGE	DEATH	MIGRATION FROM SOFIA TO N,WEST	NORTH	N,EAST	S,WEST	SOUTH	S,EAST	SOFIA
0	0,026386	0,012829	0,005593	0,003455	0,001888	0,006593	0,002064	0,941191
5	0,002133	0,008403	0,004501	0,002641	0,001419	0,004435	0,001420	0,975049
10	0,002120	0,008213	0,003968	0,002223	0,003116	0,004829	0,001338	0,974193
15	0,002801	0,010539	0,004923	0,002988	0,003192	0,007503	0,001466	0,966588
20	0,003515	0,011948	0,006035	0,004289	0,002295	0,008474	0,001843	0,961600
25	0,004456	0,013548	0,006570	0,004179	0,002333	0,008764	0,002065	0,958083
30	0,006419	0,006872	0,002791	0,001903	0,001122	0,003615	0,000944	0,976333
35	0,007273	0,008255	0,003264	0,002721	0,001150	0,004961	0,001425	0,970950
40	0,012985	0,005543	0,002125	0,001438	0,000872	0,003497	0,000936	0,972604
45	0,019732	0,003824	0,001203	0,000973	0,000515	0,002521	0,000688	0,970545
50	0,032528	0,002767	0,000825	0,000531	0,000236	0,001650	0,000471	0,960992
55	0,051229	0,002514	0,000986	0,000872	0,000219	0,001969	0,000328	0,941883
60	0,083600	0,001630	0,000859	0,000687	0,000172	0,001460	0,000257	0,911336
65	0,099596	0,001231	0,001136	0,000663	0,000194	0,001815	0,000285	0,895081
70	0,304410	0,000861	0,000638	0,000209	0,000209	0,000831	0,000209	0,692632
75	0,465881	0,000922	0,000302	0,000293	0,000000	0,000871	0,000000	0,531732
80	0,667600	0,000551	0,000535	0,000000	0,000000	0,001012	0,000000	0,330302
85	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000

3.2 *Survivors at exact age x*

The table gives the number of persons from the initial cohort of births from Sofia surviving in each region at the *exact* age (*not* age group) given in the first column.

AGE	TOTAL	N,WEST	NORTH	N,EAST	S,WEST	SOUTH	S,EAST	SOFIA
0	100000.	0.	0.	0.	0.	0.	0.	100000.
5	97361.	1283.	559.	345.	189.	659.	286.	94119.
10	97152.	2048.	984.	598.	320.	1077.	338.	91788.
15	96947.	2713.	1358.	818.	603.	1540.	457.	89457.
20	96669.	3315.	1825.	1101.	845.	2258.	557.	86767.
25	96316.	4089.	2326.	1486.	1003.	2960.	694.	83757.
30	95872.	5058.	2869.	1843.	1173.	3674.	856.	80399.
35	95257.	5490.	3077.	1994.	1245.	3933.	919.	78599.
40	94532.	6037.	3300.	2193.	1311.	4281.	1020.	76390.
45	93305.	6347.	3417.	2276.	1352.	4489.	1077.	74346.
50	91425.	6454.	3430.	2295.	1358.	4587.	1107.	72193.
55	88451.	6429.	3376.	2259.	1327.	4556.	1104.	69400.
60	83955.	6274.	3285.	2190.	1274.	4467.	1071.	65395.
65	76936.	5839.	3064.	2051.	1175.	4184.	996.	59628.
70	68463.	5003.	2656.	1770.	1055.	3710.	866.	53403.
75	48879.	4098.	2130.	1349.	809.	2815.	668.	37012.
80	27534.	2900.	1432.	846.	502.	1740.	416.	19698.
85	10395.	1602.	733.	374.	223.	760.	185.	6518.

3.3 *Life expectancies at various ages*

The table gives the number of years a member of the initial cohort of births from Sofia can expect to spend living in each region when at the *exact* age given in column one.

AGE	TOTAL	N,WEST	NORTH	N,EAST	S,WEST	SOUTH	S,EAST	SOFIA
0	70,615601	3,799070	2,009181	1,295852	0,792122	2,596419	0,630264	59,492687
5	67,461578	3,869886	2,049269	1,322100	0,808740	2,649853	0,642044	56,120487
10	62,601379	3,791708	2,013970	1,300679	0,797375	2,610870	0,629411	51,457367
15	57,729000	3,676991	1,957849	1,266946	0,775257	2,548915	0,610236	46,892883
20	52,887444	3,531639	1,881128	1,220962	0,740826	2,457991	0,585755	42,469940
25	48,072353	3,352406	1,780279	1,158294	0,694762	2,331566	0,555426	38,199619
30	43,283165	3,129389	1,653043	1,076844	0,641222	2,169367	0,517570	34,095730
35	38,546497	2,872773	1,507666	0,983085	0,581981	1,983737	0,474321	30,143011
40	33,822903	2,589951	1,350597	0,879878	0,518785	1,781711	0,426670	26,275311
45	29,234966	2,292183	1,188405	0,771708	0,454254	1,578146	0,376099	22,582170
50	24,784821	1,989276	1,025637	0,662575	0,389474	1,354238	0,324121	19,039499
55	20,533943	1,692023	0,867769	0,556133	0,326674	1,141324	0,272515	15,677505
60	16,499767	1,404371	0,715903	0,453460	0,266732	0,933766	0,222332	12,503203
65	12,777045	1,138985	0,574927	0,357035	0,211517	0,737854	0,175455	9,581351
70	9,048856	0,803950	0,437218	0,261688	0,156272	0,540898	0,129183	6,639647
75	6,672681	0,772632	0,367604	0,207013	0,123558	0,423875	0,102498	4,675501
80	4,907485	0,736276	0,329148	0,168184	0,100335	0,338876	0,083566	3,151099
85	3,877090	0,867740	0,351087	0,152020	0,091445	0,296430	0,076745	2,041624

4 Multiregional population projection for Bulgaria

In this section 'age' refers to the age group of five-year span whose initial year is given (except for 85 which refers to the 85+ age group).

4.1 *Spatial distribution of the Bulgarian population in the initial year 1975*

AGE	TOTAL	N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	677101.	70735.	94489.	128208.	50841.	176623.	71533.	80672.
5	636856.	68705.	88645.	119420.	56775.	172151.	65419.	65741.
10	627571.	67596.	87050.	114858.	57325.	174996.	66863.	58883.
15	648115.	65875.	96363.	110758.	54790.	179789.	65342.	75198.
20	667466.	63065.	99415.	117401.	51104.	167740.	64351.	104390.
25	680723.	69433.	107113.	117679.	51062.	163737.	65928.	105771.
30	570611.	56989.	87569.	102594.	44287.	140672.	55426.	83074.
35	568310.	62510.	85446.	95274.	46536.	146625.	59787.	72132.
40	639246.	75297.	101386.	104962.	51672.	160455.	67050.	78424.
45	636179.	80024.	102018.	101531.	48194.	152826.	66486.	85100.
50	613212.	85561.	107536.	96852.	45676.	138376.	57404.	81807.
55	362410.	49280.	64098.	58489.	28663.	84676.	33938.	43266.
60	446563.	60716.	78998.	72057.	35308.	104357.	41830.	53297.
65	386710.	52595.	68375.	62408.	30594.	90368.	36223.	46147.
70	279036.	56418.	64896.	41530.	19544.	54575.	24286.	17787.
75	166137.	33591.	38639.	24727.	11637.	32494.	14459.	10590.
80	73394.	14839.	17069.	10924.	5141.	14355.	6388.	4678.
85	47350.	9574.	11012.	7047.	3317.	9261.	4121.	3018.
TOT	8726990.	1042803.	1400117.	1486719.	696466.	2164076.	866834.	1069975.

4.2 *Spatial distribution of the Bulgarian population in 2025*

AGE	TOTAL	N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	763716.	70937.	103341.	156095.	46309.	212279.	70888.	103868.
5	739083.	69934.	101074.	149473.	45894.	205078.	68468.	99162.
10	728568.	69933.	101007.	145141.	45878.	202263.	67123.	97223.
15	723682.	68436.	103502.	139762.	44307.	200243.	64091.	103340.
20	717082.	65679.	105139.	134938.	42176.	195300.	60295.	113556.
25	699665.	64146.	103563.	130894.	41195.	187923.	57681.	114262.
30	676977.	63100.	100644.	125727.	41185.	182282.	55409.	108630.
35	663899.	62375.	98530.	121504.	41617.	179969.	54539.	105364.
40	663061.	62501.	98930.	119672.	41832.	178717.	55295.	106113.
45	660642.	62662.	99689.	119073.	41505.	173801.	56277.	107635.
50	613462.	60773.	93104.	110672.	39790.	159743.	52372.	97008.
55	561989.	56716.	84774.	100289.	39987.	150685.	47390.	82149.
60	518136.	52016.	77223.	90438.	38446.	142664.	45745.	71605.
65	477925.	46596.	72761.	78921.	35521.	130621.	42022.	71484.
70	401696.	39364.	61162.	68302.	29535.	100414.	36066.	66852.
75	287481.	33246.	48334.	48089.	21344.	68228.	26801.	41438.
80	134829.	17503.	23868.	23189.	10424.	31702.	12759.	15385.
85	76021.	15249.	15636.	11252.	5838.	16098.	7407.	4541.
TOT	10107913.	981166.	1492280.	1873431.	652782.	2718010.	880627.	1509616.

4.3 *Stable equivalent of the 1975 observed Bulgarian population*

AGE	TOTAL	N.WEST	NORTH	N.EAST	S.WEST	SOUTH	S.EAST	SOFIA
0	662471.	54276.	94175.	174034.	18832.	181079.	46281.	93875.
5	644233.	53583.	92666.	168134.	18336.	175840.	44712.	90963.
10	635120.	53643.	92844.	164014.	18253.	173010.	43846.	89509.
15	625837.	52439.	94473.	156394.	18015.	169706.	41728.	93082.
20	615979.	50157.	95088.	148951.	17418.	165078.	39090.	100190.
25	605734.	49034.	94115.	145019.	16949.	160729.	37646.	102241.
30	595083.	48641.	93105.	141929.	16652.	157445.	36727.	100584.
35	583571.	48233.	91609.	138818.	16315.	154156.	35987.	98533.
40	570388.	47703.	89747.	135427.	15933.	150601.	35123.	95774.
45	554060.	46652.	87284.	131225.	15487.	146527.	34192.	92693.
50	532871.	45078.	83987.	125908.	14890.	141088.	32958.	88962.
55	504672.	42954.	79778.	118743.	14096.	133784.	31249.	84060.
60	465355.	39769.	73780.	109023.	12989.	123484.	28815.	77494.
65	409170.	34742.	64386.	94934.	11586.	108614.	25215.	69693.
70	328040.	28541.	52213.	75355.	9471.	86800.	20180.	55479.
75	225549.	21515.	37989.	51893.	6537.	58954.	13964.	34697.
80	123159.	13588.	22629.	28278.	3572.	31378.	7657.	16073.
85	66260.	10733.	14946.	14461.	1852.	15010.	4015.	5242.
TOT	8747551.	741282.	1354813.	2122531.	247183.	2333354.	559226.	1389161.