Multilevel synthesis. From the group to the individual

Courgeau, Daniel

2007

Multilevel Synthesis

THE SPRINGER SERIES ON

DEMOGRAPHIC METHODS AND POPULATION ANALYSIS

Series Editor

KENNETH C. LAND

Duke University

In recent decades, there has been a rapid development of demographic models and methods and an explosive growth in the range of applications of population analysis. This series seeks to provide a publication outlet both for high-quality textual and expository books on modern techniques of demographic analysis and for works that present exemplary applications of such techniques to various aspects of population analysis.

Topics appropriate for the series include:

- General demographic methods
- Techniques of standardization
- Life table models and methods
- Multistate and multiregional life tables, analyses and projections
- Demographic aspects of biostatistics and epidemiology
- Stable population theory and its extensions
- Methods of indirect estimation
- Stochastic population models
- Event history analysis, duration analysis, and hazard regression models
- Demographic projection methods and population forecasts
- Techniques of applied demographic analysis, regional and local population estimates and projections
- Methods of estimation and projection for business and health care applications
- Methods and estimates for unique populations such as schools and students

Volumes in the series are of interest to researchers, professionals, and students in demography, sociology, economics, statistics, geography and regional science, public health and health care management, epidemiology, biostatistics, actuarial science, business, and related fields.

The titles published in this series are listed at the end of this volume.

Multilevel Synthesis

From the Group to the Individual

by

Courgeau Daniel
INED (Institut National d'Études Démographiques)
Paris, France

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-5621-4 (HB)
ISBN-13 978-1-4020-5621-5 (HB)
ISBN-10 1-4020-5622-2 (e-book)
ISBN-13 978-1-4020-5622-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

To Hella, my wife

CONTENTS

ACKNOWLEDGMENTS

I would like to thank Statistics Norway for permission to use files extracted from their country's population register and censuses, as well as INSEE, the French National Institute of Statistics and Economic Studies, for granting me access to data from the "Young People and Careers" (Jeunes et Carrières) survey. I am most grateful to Harvey Goldstein for our many discussions on the subject of this book during his stays at INED. I also thank the two anonymous readers for their comments and suggestions, which have helped to improve the text. Lastly, I wish to express my warm thanks to Jonathan Mandelbaum for the quality of his translation from the original French, and to Nicole Berthoux and Cyril Courgeau for developing the many figures and maps as well as preparing their English versions.

GENERAL INTRODUCTION

This book aims to show how the multilevel approach successfully overcomes the divisions that emerged during the rise of the social sciences—specifically, here, demography and statistics—from the seventeenth century to the present. We begin by examining how the approach connects different branches of the social sciences that conducted their observation and analysis at different levels or addressed different aspects of human complexity. Next, we describe in greater detail the statistical methods and techniques that it can use to simultaneously identify and measure the effects of the different levels examined on the characteristics studied.

To introduce this work, we first provide an overview of the goal of social sciences, before discussing the various levels of aggregation at which they can operate.

Social sciences start from the observation of a real-life experience and then seek to structure it according to different fields, which constitute the specific objects of study for each science. As a rule, these objects are defined independently of the scale and levels of aggregation that can be chosen to observe them. For example, the object of demography is the quantitative study of human populations, their variations, and their status, without specifying whether the level chosen is the individual, the family, the population of a town, or that of a country. Likewise, economics studies the production, distribution, and consumption of wealth, without specifying if the analysis is of an individual, a market, a firm or a nation. In sum, the distinction between levels is subordinated to the object of each science, and we shall see that the distinction applies to all the social sciences.

Second, the social sciences need to discover the appropriate categories that can serve as starting points for their growth. Indeed, it may be tempting for any person, who experiences these various social facts on a routine basis, to be content with their ostensible meaning and a naive explanation of the lived experience in its immediacy. This may be because the person already grasps the significance of the facts or because (s)he feels the absence of an explanation and is preparing to search for it among similar life experiences (Granger, 1994). This is the case for the many phenomena studied in the social sciences, such as the

birth of a child or the death of a person in demography, a price rise in economics, and so on. Far from being convinced of the complexity and opaqueness of these phenomena, the naive observer views them, on the contrary, as already laden with explanation. But this explanation, which is specific to each individual, differs from those of other individuals and will thus prevent any schematization that could be adopted by all and be publicly intelligible. Hence the importance, for the social sciences, of setting aside such explanations and developing categories that will allow an even provisional objectivation of the human experience. While these fledgling sciences have not yet managed to isolate the categories with sufficient clarity, we may assume that the objectivation process is under way.

The researcher will try to observe behaviors and objectivate the environment in which they occur and their modus operandi. Only then will (s)he face two important choices: the level of aggregation in a space that is both physical and social; and the temporality to be examined. Many questions will arise concerning the choice of aggregation level: Should the observation focus on aggregate or individual behavior? To identify the relationships between the variables measured, should one use identical methods or totally distinct ones for each level? Can different aggregation levels be used simultaneously? and so on. The researcher will concurrently face the need to take account of time: Will it be the historical time in which the events studied occur, or, on the contrary, the time lived by the individual experiencing the events? Will the study select a specific point in this time frame to explain the behaviors occurring then by the conditions prevailing immediately before? Or, on the contrary, will the researcher examine an individual's entire life course, incorporating conditions that vary continuously but may also be situated in a distant past? It is this set of issues that we shall try to address throughout the volume, as we seek to find satisfactory solutions to them.

We begin by considering the opposition between society and the individual, which raises crucial issues in the social sciences. In broad terms, the problem is the following: should social movements, and the conditions that determine them, be viewed as the consequences of social facts generated by supra-individual players such as institutions, organizations, governments, interest groups, nations, etc.—or, on the contrary, as aggregates of the actions, attitudes, relationships, and specific environments of the individuals that instigate them (Nadeau, 1999)? Using this opposition as a framework, let us examine how human behaviors are taken into account and what consequences ensue.

FROM AGGREGATE LEVEL . . .

We find it preferable to start from the aggregate level, which Aristotle already regarded as primordial in some of his writings. Accordingly, the State or *polis* (πόλιζ), irrespective of its government,

> is by nature clearly prior to the family and to the individual, since the whole is of necessity prior to the part; for example, if the whole body be destroyed, there will be no foot or hand, except in an equivocal sense, as we might speak of a stone hand; for when destroyed the hand will be no better than that (Aristotle [1885], 1253 *a* 20).

Viewed as a whole, the State is not an artificial or conventional construct but originates in the requirements of human nature: a man unable to belong to a community is "either a beast or a god."

Indeed, for Aristotle, the individual could not be the object of any science. In *Rhetoric*, he clearly states that

> none of the arts theorize about individual cases. Medicine, for instance, does not theorize about what will help to cure Socrates or Callias, but only about what will help to cure any or all of a given class of patients: this alone is business: individual cases are so infinitely various that no systematic knowledge of them is possible (Aristotle [1954], 1356 b 28).

Note that Aristotle often uses the term "art" (τέχνη) as a substitute for "science" (επιστημη), while sometimes distinguishing between the two: art is more oriented toward "necessity or adornment"; science is disinterested and does not preoccupy itself with life's pleasures or necessities. We should also note that the modern concept of a science of man does not appear in Aristotle's thought (Granger, 1976).

Closer to us, it is society or the modern State, rather than the ancient *polis*, that constitutes the macro level *par excellence*. To conduct analysis at the societal level is to regard society as a perfectly defined and organized whole, clearly distinct from the sum of the individuals that compose it, and displaying a high degree of internal integration. Accordingly, we could deal with this society independently of other simultaneously existing societies; we could regard the social phenomena to be studied as external to individuals, since they are of a different nature from individual states of consciousness. By contrast, we can compare different societies and highlight their distinguishing characteristics.

Earlier, we saw that all social sciences aim to explain specific sets of behaviors—e.g., mortality, fertility, marriage, and migration for demography; production and consumption of wealth for economics, etc. They do this by constructing an abstract structure for describing the observed phenomena. If we take society as the observation level, we represent life experience by the statistical reality of the facts observed in that society. We can separate the facts into two categories: (1) those that will represent the origin of social facts and the initial conditions observed, (2) those that will represent the results obtained in these conditions. The point of the exercise here is to use a *model* to describe not only the overall results, but also the processes leading to them from the initial conditions.

We must seek the origin of social facts in the make-up of the social environment in which they occur. The initial conditions will therefore be provided by social facts, which may lead to the phenomena studied and which are observed prior to them. We can measure social facts by means of statistics describing the status and characteristics of the society that we are studying. For example, we can link the percentages of individuals exhibiting a given behavior—proportion of suicides, proportion of migrants, proportion of persons who have had a particular disease, who have exited from farmer status, and so on—to certain characteristics that may or not induce that behavior: share of Catholics and Protestants to explain suicide; percentage of management-level workers or, on the contrary, farmers to characterize migration; percentages of individuals living in unsanitary conditions or, on the contrary, in uncontaminated locations to characterize the propensity to contract a particular disease; percentages of farm laborers or, on the contrary, of farmers operating large holdings to characterize the exit from agriculture, and so on.

Thus, taking as our starting point the society as a whole organized to perform a given set of functions, we shall show how it produces a specific economic, demographic, social or other kind of effect. To be more precise, it is by connecting the observed facts with the society of which they are, in various ways, an expression that we can explain and find a basis for their reciprocal effects (Franck, 1994).

Underlying this approach is a specific historical time. As noted earlier, we shall place ourselves at a given point in time to explain the phenomena then occurring as a function of the conditions prevailing immediately before. The approach gives precedence to the analysis of coexistence and relationships at a given moment: cross-sectional analysis in demography, static analysis in sociology, structuralism in anthropology, and so on. Naturally, the situation may evolve from one period to the next, as structures will have changed and macro-effects may also be modified. But, again, these changes happen only at the aggregate level, without involving individual behaviors occurring in lived time.

All the *paradigms*—or, rather, the *research programs*[1]—that support such an approach in each social science must therefore regard the individual as a non-relevant unit; only the individual's membership in groups or categories will influence the rates of occurrence of the phenomena studied. The paradigms will, of course, contain elements specific to each social science. This defines what epistemologists call *methodological holism*, in which some of the facts studied are a function of the social science examined, whereas others may be common to several sciences.

Later in this volume we shall see the problems facing such holism when it seeks to account for all human facts. In particular, it leads to what is called

the *ecological fallacy*, if we seek to detect individual behaviors from aggregate measures (Robinson, 1950).

. . . TO INDIVIDUAL LEVEL . . .

The other approach will focus instead on the individual. However, given the diversity of meanings assigned to individualism by different social sciences (Birnbaum and Leca, 1986), it is important to state that we shall set aside sociological, economic, legal, ethical, and philosophical individualism—presented and discussed in greater detail in Valade (2001). We shall confine our examination here to *methodological individualism*, as defined by epistemologists. In this case, the goal is to explain an observed phenomenon not as if it were determined by the society studied, but on the contrary as if it resulted from individual actions or attitudes. This makes it essential to "reconstruct the motives of the individuals concerned by the phenomenon in question, and to treat it as the result of the aggregation of individual behaviors dictated by those motives" (Boudon, 1988). Such an approach is valid for all phenomena, whether they belong to sociology, demography, economics, or any other social science.

It is important to bear in mind that methodological individualism appeared in our Western societies far later than holism, as it basically derives from the ideas elaborated during the emergence of social sciences in the early seventeenth century. During this process, "the autonomous individual constitutes the ultimate unit of the social sciences, and all social phenomena are resolved into individual decisions and actions that it is useless or impossible to analyze in terms of supra-individual factors" (Valade, 2001). However, its introduction raises a host of problems, which we now need to examine in detail.

Indeed, we noted earlier the force of Aristotle's argument that the individual is unlimited and scientifically unknowable. This is because the individual is intimately linked to the real-life experience of players—composed of thoughts, feelings, intentions, and so on—that are not directly accessible to the researcher. An individual experience of this kind cannot be turned into an object of science. How, then, can we formalize it as a theoretical object amenable to an overall modeling?

We start here from the observation of individual lives, via a biographical compilation providing all the events that are of use to the social science under examination and that occur throughout the individual's life course (they need to be properly dated). This observation in no way enables us to estimate individual random processes whose probabilistic structure is assumed to be specific to each individual tracked. Indeed, it seems hard to suppose that two individuals, even if similar in many respects, necessarily follow the same process as a result. Moreover, as we can observe only one outcome of the process for each

individual—namely, his or her personal trajectory—we have no way of identifying its probabilistic structure. This is wholly consistent with Aristotle's earlier-quoted observation to the effect that an individual process is not identifiable.

We therefore need to look at the process in a totally different way. Let us continue to take as our starting point an observed reality composed of a number of individual paths. From this observation, can we estimate a probabilistic process that takes into account all the information contained in the paths? As any random process may be viewed as a distribution of probabilities across a set of paths, we can say that, in this case, we repeatedly observe the same random process. Now, we can determine the probabilistic structure of the underlying process by observing those different paths. We thus identify a collective process—whose complexity can be as great as we want—from the observation of a set of individual paths.

In our earlier example, two *observed individuals*, whose characteristics are taken to be identical, have no reason to follow the same process. Here, two *statistical individuals*, seen as units in a repeated random draw, subject to the same selection conditions and displaying the same characteristics, automatically follow the same process. We can thus see more clearly how the use of observed event histories, which constitute the statistical reality of human facts studied, can now be transformed into an abstract description of human reality by means of concepts deliberately stripped of at least a portion of the concrete circumstances. These concepts are chained together in accordance with the logical relationships of the identified process, forming an event-history model.

Thanks to such an analysis, and by observing a certain number of individual cases, we can identify a mechanism that will connect the phenomena studied to individual characteristics, whether or not they are time-dependent. We now need to show what abstract relationships exist between the elements of a process that organizes the life of the population we are studying. At this point, however, our undertaking will require a totally different approach to human societies. As shown later, we shall need to develop a new data-collection procedure and new analytical methods.

The paradigms—or, rather, research programs—that sustain such an approach in the various social sciences will now regard individual event histories (biographies) as the life experience on which they will work. However, their research objects will consist of the processes that impart meaning to the event histories. It is important to remember that, within the framework of this methodological individualism, we cannot grasp the life experience through the actors themselves, but only an abstract process. The paradigms will, of course, contain elements specific to each science examined.

Arguably, this approach enables us to take proper account of the influence of various individual characteristics on behaviors: a combination of

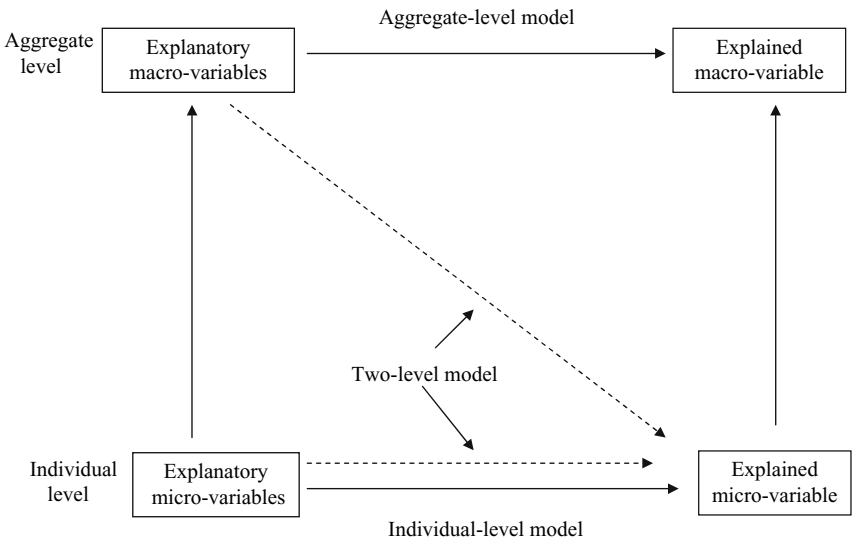
variables whose effects identified in a given generation or cohort and in a given environment explain those behaviors. This time, however, as we disregard the context in which behaviors occur, a risk of *atomistic fallacy* appears. In fact, there is no reason why the family context or the environmental context should have no influence whatsoever on observed behaviors, and it seems fallacious to examine individuals divorced from the constraints imposed by the society and environment in which they live.

. . . TOWARD RECOMPOSITION AND MULTILEVEL SYNTHESIS

Figure 1 depicts the two types of model described earlier and gives a clearer picture of the difficulties we encounter when trying to shift from one to the other.

The aggregate-level model, represented by the solid horizontal arrow at the top, shows the result obtained at the macro level—for example, the proportions of individuals experiencing the analyzed event explained by the rules and institutions governing these behaviors in the society under study; these proportions are measured, for instance, by aggregate characteristics. By contrast, the individual model—shown by the solid horizontal arrow at the bottom—connects the result obtained to the micro level: for example, the probability that

Figure 1. Links Between Different Models



a statistical individual will experience the event, explained by the individual characteristics governing this behavior.

We shall now briefly discuss the difficulties that appear when we try to move between levels. Illustrations will be provided with the aid of simple, concrete examples.

Let us begin by examining the shift from the individual level to the aggregate level, represented by the two solid vertical arrows in Figure 1. Rather than exploring all the possibilities in detail, we shall use one specific case to show the reasons for the divergence between individual models and aggregate models.

Let us assume we are working on binary individual data—i.e., the individual possesses a given characteristic or not—as regards both the dependent variable (i.e., the variable to be explained) and the explanatory variables. For example, in demography, suppose we want to link an inter-regional migration to the fact that a person is a farmer or not (Courgeau, 2002b). In this case we can regard the region in which the person lives as the aggregate level, and thus compute percentages of migrants and farmers in each region. Consequently, we see a change in the nature of the variables and therefore in the models that are applicable at the individual level and the aggregate level. There is a shift from a dichotomous model, of the logit or probit type, to a model with regression between aggregate characteristics. In such circumstances, how do we connect the parameters of the models? For example, we have been able to show that if we estimate a logit at the individual level, the model at the aggregate level is a linear regression between the percentages (Baccaïni and Courgeau, 1996). But the estimation of parameters from the individual model may differ substantially from the estimation on aggregate data and even, in some cases, contradict it (Courgeau, 2002b). Again, we cannot explain these differences except by incorporating simultaneously into the same individual model the fact of being a farmer and the percentage of farmers in the region. We are no longer dealing with an aggregate model.

Conversely, it is even less possible to recover the parameters of the individual model from those of the aggregate model, even assuming that the latter is very simple. In particular, the fact that information at the aggregate level is poorer than at the individual level will prevent the transfer. An individual migrant's occupation cannot be determined from the percentages of farmers and migrants in a given region.

As a result, we may view these two types of models—one concerning the structure of society, the other individual behaviors—as virtually independent.

So far, we have examined only two main levels at which we can position ourselves to study human phenomena: society and the individual. A more detailed examination of a society readily shows us the existence of other, intermediate

levels between them, and the need to position ourselves at these levels as well to better understand the society in which they operate and the individual behaviors that they can generate.

For instance, when the demographer or population geographer studies inter-regional migrations, instead of viewing them from an individual or national perspective, they can specify the region as the intermediate level. Various regional characteristics may exert an important, specific influence on the flows, in ways that differ from one region to another. At the national level, unemployment rates or average regional wages could be viewed as means for measuring a broader effect of unemployment or wages on inter-regional migration rates. Now, we shall examine the specific effect of these same characteristics on the individual probabilities of emigrating from each region. The same goes for migrations between municipalities or larger sub-regional territorial units such as French *départements*, which will involve levels corresponding to municipalities, *départements*, and other divisions. Other, non-geographic levels can also influence the outcome: for instance, the household, in which the individual lives, may inhibit migration, when it consists of many members engaged in different activities in the region; alternatively, the household may stimulate migration in the case of people living alone with few ties to the region.

This emergence of multiple levels is easy to generalize in all social sciences, and there is an increasing need to examine them concurrently. Indeed, it is important to grasp that these realities are not ontologically separate and that we must try to find how microstructures fit into macrostructures and vice versa.

Simultaneously, the time frames to examine will multiply and diversify. The time linked to the life of an individual may not suffice, for intermediate time frames can play an important role. Likewise, historical time may be divided into periods of unequal significance.

For instance, when demographers want to study a woman's fertility, they can focus on her age and compute age-specific fertility rates. However, one can argue that fertility is linked to couple formation. If so, it is preferable to compute rates by union duration, as this time scale will no doubt better reflect the woman's fertility. But again, for births of order greater than 1, it may be preferable to look at inter-birth intervals rather than union duration, and so on. Choosing between these time scales is not easy, and it would be better to analyze all of them simultaneously.

In other words, we need to examine a wide diversity of levels and time scales in order to better understand human phenomena. The terms "micro" and "macro" become totally relative, and a level viewed as "micro" in one analysis may become "macro" in another. For example, while the restricted family constitutes a more aggregated level than the individual, it will serve as the micro level with respect to the extended family. Not only is this reciprocal

relativity of levels now clearly visible, but, more important, it seems essential to realize how closely the levels are linked and can no longer be treated separately. We can no longer say that one of these levels is more fundamental than the others, and, even less so, that it is independent of the others. Therefore, we must now study the interrelationship between levels, and it is its recomposition that we shall now describe.

After decomposing the object of social science into its different levels and time scales—which, as we have seen, often seem mutually contradictory—we must try to reconstruct an overall object from which to arrive at a synthesis between the approaches identified here. Many researchers have long called for this synthetic approach (Alexander et al., 1987; Huber, 1991). In the next few pages we shall try to offer a very broad outline of the synthesis described in greater detail in the rest of the book. For this purpose, we shall use some concepts that have emerged in our discussion and make the reconstruction possible.

First, we believe the concept of statistical individual is crucial to grasping a more general process affecting the entire population. By properly separating actors' experiences from social-science constructs, the concept allows a linkage of the analyses performed at different aggregation levels. Nothing prevents us any longer from regarding statistical individuals as subject not only to the effect of their own characteristics but also to the imposed—or, rather, structural—constraints (Giddens, 1984) of the social system in which they live. Such constraints are exercised not independently of individual motives and reasons (as in holism), but in a form that is both empowering and restrictive. We can thus overcome the two types of fallacy noted earlier. The risk of ecological fallacy is eliminated, as the aggregate characteristics will measure a different construct than their equivalent at the individual level. They operate no longer as a substitute, but as an aggregate constraint that can influence the behavior of an individual subjected to it. At the same time, the atomistic fallacy is expunged by the proper handling of the context in which the individual lives.

That context can now involve as many aggregation levels as needed. The methods should allow the treatment of hierarchical levels (individuals situated in nuclear families, themselves situated in extended families, and so on) as well as more complex nestings (individuals classified by type of residential neighborhood and type of place of work, which, in turn, are examined in a hierarchical classification by *département* and region). It should be possible, at the same time, to generalize these contextual models, in which individual results obtained in different groups at a given level are treated as independent, as truly multilevel models in which the result for an individual in a particular group can depend on the results obtained for other members of the group.

More profoundly, by identifying a plurality of levels, we abandon the dualist approach, which pits society against the individual. In these conditions,

"it no longer makes sense to choose between holism and atomism, and, as regards social science, between holism and individualism" (Franck, 1995), for we now seek to study how these different levels will interrelate. At the same time, we want to find a way of articulating a historical time scale and several individual time scales in the same model, as noted earlier. Thus multilevel analysis effectively enables us to adopt a new approach in the social sciences.

While the multilevel model itself—as we have outlined it—operates at the level of the statistical individual, it allows us to introduce (1) the effects of characteristics measured at various aggregation levels on the individual's behavior, as well as (2) interactions between individual and aggregate characteristics. Such a model is represented in figure 1 by the two dotted lines connecting individual and aggregate characteristics to the expected result. Of course, these characteristics may interact, not only at a given aggregation level but especially between these levels, and we have the possibility of incorporating far more aggregation levels than the two shown in the diagram.

But is this approach the only solution to the problem of recomposing of the object of social science? Do we not need different models to explain the changes in structural constraints? For example, how should we analyze the birth, functioning, and death of institutions such as—in economics—the market and centrally planned systems (Lesourne, 1991)? What are the interactions between these institutions and individual behaviors? We should also examine the possibility that several types of models may exist, given the multiple time scales required to grasp the full complexity of the phenomena studied. All these questions are raised by the implementation of the synthesis consisting of multilevel models, which we shall also examine in this volume.

HOW OUR WORK WILL ADDRESS THE SYNTHESIS

To show how multilevel analysis overcomes the macro-micro dichotomy, we shall have to review the history of demography and social sciences from their origins. We need to place the holism-versus-individualism debate in a broader setting than the current observation of these sciences, and to go back to their beginnings in the seventeenth century. Retrospective examination will enable us to illustrate new relationships between these different methods and the multilevel approach, which were not clearly visible when the methods were introduced.

However, our purpose is not to provide a detailed description of the methods of demographic analysis: the reader will find their main phases and demonstrations in the manuals that have marked the history and development of demography since its birth (Graunt, 1662; Lotka, 1939; Landry, 1945; Pressat, 1966; Henry, 1972; Schryock and Siegel et al., 1973; Wunsch and Termote,

1978; Courgeau and Lelièvre, 1989, 1992; Caselli et al., 2001). Our presentation is confined to the basic elements of the methods and their application to specific examples used throughout the volume, in order to illustrate the links and dissimilarities between the various approaches.

John Graunt's book (1662) marks the beginning of the statistical study of human populations. He regarded "bills of mortality" and "bills of christenings"—which record burials and baptisms—as valid sources for measuring the changes in human populations over time and even for estimating their size. This was a revolutionary concept at a time when phenomena such as birth, illness, and death were seen as God's secret and out of bounds to scientific scrutiny. Graunt's research paved the way for demography (Vilquin, 1976), epidemiology (McMichael, 1999), and social science in general.

During the development of what William Petty called "Political Arithmetick" (1690), attention largely focused on the period analysis of social groups. In the seventeenth and eighteenth centuries, in the absence of censuses, researchers had to make assumptions about the links between observed events (births, marriages, and deaths) and the populations experiencing them, at a given moment. Investigations also began on the variations of a population as a function of the births and deaths recorded in it: by "assuming the number of all living persons in a given location remains the same, or grows or decreases in a uniform manner," Euler (1760) already articulated the concepts of stationary or stable populations, which would not be formalized until the early twentieth century (Lotka, 1939).

In the nineteenth century, Adolphe Quetelet was the first to generalize the study of populations with his theory of the average man. However, the sociologist Émile Durkheim was largely responsible for introducing a theory of the quantitative analysis of the behaviors of social groups, with his clearly stated hypotheses and his method of concomitant variations. The same methodology was presented fifty years later in Adolphe Landry's demographic treatise (1945), although the author makes no mention of his illustrious predecessor. This led to *period analysis*, also called *cross-sectional analysis*, which we discuss in Chapter 1. The main source consists of population censuses, which provide snapshots of the population measured at regular intervals.

While the distinction between "historical time" and "individual time" was not clearly perceived at the outset, some researchers already used sources that tracked the lives of individuals, such as tontine data (Deparcieux, 1746). Later, some voices suggested that period analysis, recommended by most authors, might not be the only possible approach (Delaporte, 1941). In fact, it was after World War II that demographers showed how this method—built on hypotheses that totally disregard personal lived time—yields results that are hard to interpret. These demographers countered by introducing *longitudinal analysis*,

which follows generations or cohorts over their entire life course, factoring in the length of their stay in the status examined. Vital statistics and population registers were now the preferred sources. To isolate the various phenomena, the new paradigm treated them as independent of one another in a supposedly homogeneous population. We discuss the paradigm in Chapter 2.

However, these assumptions of independence and homogeneity, which allowed the use of aggregate data, did not hold up in the face of many results obtained with surveys more detailed than population registers and vital statistics. It became necessary to develop methods postulating (1) dependence between phenomena and (2) population heterogeneity. It also became essential to make plans for the adoption of a new information-gathering method: the event-history survey. In demography, this revolution occurred in the early 1980s with *event-history analysis*. The approach required far more complex mathematics and probability theories than those in use until then. Chapter 3 describes the theory underlying these methods.

But the event-history approach was too focused on the individual. It stripped away the influence of society, with its constraints and rules, on individual behavior. A new approach maintained the individual focus (unlike period analysis) but provided for multiple aggregation levels to accommodate the effect of underlying constraints imposed or introduced by each level. This is known as *contextual* and *multilevel analysis*. It required more complex data-gathering procedures for accurately evaluating the effects of the different aggregation levels considered; it introduced new methods for estimating the random factors operating at each level. In Chapter 4, we seek to define the conditions in which this generalization is valid.

A general discussion of the proposed paradigms will enable us to better grasp their respective contributions and to see how the multilevel approach fits into the evolution of demographic hypotheses. The important point here is to realize that demography is not a set of viewpoints and options defined once and for all. The viewpoints and options are specific to the society in which the demographer lives, and they may change over time (Singleton, 1999). Hence the need to define the paradigmatic choices with precision, to show their specificity, and to discuss their foundations.

After putting the multilevel approach in its proper context in the history of demography, we shall be able to turn to the definition of aggregation levels, the different models offered by the approach, and the mathematics needed to estimate the models' new parameters.

While in some cases the definition of certain aggregation levels is self-evident, in other cases we need to examine whether the levels used are valid and necessary. For example, when we take the class as an aggregation level to study students' grades, its effect seems obvious. Students in the same class emulate

one another, and their teacher(s) influence(s) their scholastic performance. By contrast, if we consider people's place of residence as an aggregation level, its effects may seem fuzzier: can we regard it as a level that enables us to address the concept of relationship network, which is harder to measure? In Chapter 5, we shall need to take a closer look at the meaning of the aggregation levels examined.

The classical regression model is the simplest and will enable us to see in detail the hypotheses required for developing a truly multilevel model. For instance, the introduction of random factors and effects of characteristics situated at multiple aggregation levels poses problems for estimating and interpreting the results obtained. We discuss the methods proposed for solving them, which we illustrate with specific examples. The examination of estimated residuals will also be useful, allowing us to distinguish the units in a given aggregation level that are located in borderline situations. Chapter 6 will untangle these issues.

Continuous variables to be explained (dependent variables), which allow the use of classical regression methods, are generally rare in demography. In this discipline, the dependent characteristics are usually binary, "either/or" characteristics (i.e., the individual has or has not experienced the event) or polytomous characteristics (meaning that the individual may experience several competing risks). This has led to new types of models such as logit and probit. The estimation methods for these models in multilevel analysis will differ from those used in regression models, and we shall need to examine the hypotheses required. Moreover, the use of such data will raise new problems whose solution is discussed in Chapter 7.

The demographer will not be content with this instantaneous approach, even though it is restricted to individuals in a single generation or cohort. The aim is to preserve all the benefits of an event-history approach—whose richness no longer needs demonstrating—and incorporate it into a multilevel approach. We shall thus need to follow an individual over his or her entire life and identify the interactions between several demographic phenomena that will occur simultaneously in a physical space or, more generally, a social space. How will an individual's move from one geographic area to another modify his or her behavior? Conversely, how will an individual's past behavior modify his or her possibilities for relocating to another area? The implementation of a *multilevel event-history analysis* should provide answers to all or at least some of these questions. It will encounter difficulties both in the implementation of surveys designed to capture event histories in a complex space and in the analysis of the data. We address these issues in Chapter 8.

Lastly, we need to look at the problems imperfectly resolved by a multilevel approach and the broader issues that it raises, particularly regarding the definition and interconnection of levels. That will be the subject of our

conclusion, which will also emphasize the major contributions of multilevel analysis. We shall be able to glimpse the various possible perspectives on probabilities that the analysis opens up for us, and to see how we can go beyond the explanation of human behavior by seeking to discover the broader social mechanisms underlying the behavior.

This volume does not contain a detailed description of the mathematical and statistical methods used to perform the analyses reported here. Our aim, instead, is to indicate the paths followed and to illustrate them by means of a streamlined presentation of the methods used in the simplest cases. The reader interested in a rigorous statistical exposition of the methods can turn to the works cited throughout the text (Bryk and Raudenbush, 1992; Andersen et al., 1993; DiPrete and Forristal, 1994; Rice and Leyland, 1996; Kreft and de Lew, 1998; Lindsey, 1999; Snijders and Bosker, 1999; Goldstein, 2003a; etc.), which supply full details on the mathematical and statistical approaches used in social science.

ENDNOTE

1. See Appendix 1 for a fuller definition and discussion of the epistemological terms used in this work.

SUMMARY

This textbook presents a historical panorama of the evolution of demographic thought from its seventeenth-century origins up to the present day. Daniel Courgeau demonstrates how the multilevel approach can resolve some of the contradictions that have become apparent and thus achieve a synthesis of the different approaches employed. Part One guides the reader from period analysis to multilevel analysis, examining longitudinal and event-history analysis on the way. Part Two is a detailed account of multilevel analysis, its methods, and the relevant mathematical models notably as regards the type of variables used. Numerous examples, used across successive chapters, make the book clear and easy to follow.

In his theoretical and epistemological treatment of these issues, the author revisits the foundations of sociology and demography while outlining the logical development that has led to the most recent approaches. This presentation is sufficiently rigorous to satisfy social scientists yet accessible for readers new to the field. The whole adds up to a comprehensive account of progress in sociological and demographic *savoir-faire*. Courgeau offers us both a textbook and an assessment of multilevel analysis that tackles one of the major challenges in empirical sociology: how to integrate analysis at the individual and group levels.